THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Integrative analysis of multi-omics data reveals links between human diseases and the gut microbiota

PEISHUN LI

CHALMERS
UNIVERSITY OF TECHNOLOGY

Systems and Synthetic Biology
Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Integrative analysis of multi-omics data reveals links between human diseases and the gut microbiota
PEISHUN LI

ISBN 978-91-7905-595-0

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5062
ISSN 0346-718X

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

Cover: Interactions between the gut microbiota, human metabolism and environment.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Integrative analysis of multi-omics data reveals links between human diseases and the gut microbiota

Peishun Li
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

The gut microbiota plays a critical role in human diseases, including type 2 diabetes (T2D) and osteoporosis. Especially, probiotics have been suggested to provide potential intervention strategies for improving human health. This thesis focuses on elucidating the interrelationships between the gut microbiota, probiotics and human diseases by integrative analysis of plasma metabolomics and gut metagenomics, using machine learning (ML) and genome-scale metabolic model (GEM). This work is mainly structured into two parts, including a systematical investigation of: (I) associations between the gut microbiota and T2D, (II) the effects of probiotic *Lactobacillus reuteri* ATCC PTA 6475 on bone metabolism of the elderly.

For the first part, a derivative of phenylalanine was identified as a potential link between the gut microbiota and T2D. It was associated with insulin resistance and might contribute to the metabolic imbalance of (pre)diabetes. By performing a systematical analysis of four metagenomic datasets, several short-chain fatty acids (SCFAs)-producing bacteria and metabolic reactions were consistently identified to be important for predicting T2D status across different studies. For the second part, this work revealed that supplementation with *L. reuteri* ATCC PTA 6475 prevented detrimental alterations in the metabolisms of both the gut microbiota and the elderly as well as increased the microbial gene richness, which might link the beneficial effects of probiotic *L. reuteri* ATCC PTA 6475 to bone metabolism. In addition, it was demonstrated that the use of ML and GEM have the potential to identify key disease-related metabolic signatures of single *L. reuteri* strain, the entire gut microbes, or the human host, based on the metabolomics and metagenomics data.

Taken together, this work provides novel insights into links between the gut microbiota and the human diseases as well as the positive effects of *L. reuteri* ATCC PTA 6475 on bone metabolism by integrating omics data using ML and GEMs.

Keywords: gut microbiota, metabolomics, multi-omics, type 2 diabetes, osteoporosis, machine learning, metabolic modeling

List of Publications

This thesis is based on the work in the following publications and manuscript:

Paper I: **Peishun Li**, Hao Luo, Boyang Ji, Jens Nielsen. Gut microbiome and machine learning potential for personalized medicine. (Manuscript for a review)

Paper II: **Peishun Li**, Boyang Ji, Dimitra Lappa, Abraham S Meijnikman, Lisa M. Olsson, Ömrüm Aydin, et.al, Thue W. Schwartz, Fredrik Bäckhed, Max Nieuwdorp, Louise E. Olofsson, Jens Nielsen. Systems analysis of metabolic responses to a mixed meal test in an obese cohort reveals links between tissue metabolism and the gut microbiota. (Under revision in Communications Medicine)

Paper III: **Peishun Li**[*], Hao Luo[*], Boyang Ji and Jens Nielsen. Metagenomic analysis of type 2 diabetes datasets identifies cross-cohort microbial and metabolic signatures. (Manuscript)

Paper IV: Hao Luo, **Peishun Li**, Hao Wang, Stefan Roos, Boyang Ji and Jens Nielsen. Genome-scale insights into the metabolic versatility of *Limosilactobacillus reuteri*. BMC Biotechnology, 2021; 21: 46.

Paper V: **Peishun Li**[*], Daniel Sundh[*], Boyang Ji[*], Dimitra Lappa, Lingqun Ye, Jens Nielsen and Mattias Lorentzon. Metabolic Alterations in Older Women With Low Bone Mineral Density Supplemented With *Lactobacillus reuteri*. JBMR Plus, 2021; 5(4):e10478.

Paper VI: **Peishun Li**[*], Boyang Ji[*], Hao Luo, Daniel Sundh, Mattias Lorentzon and Jens Nielsen. One-year supplementation with *Lactobacillus reuteri* ATCC PTA 6475 counteracts a degradation of gut microbiota in older women with low bone mineral density. (Under revision in npj Biofilms and Microbiomes)

Additional papers and manuscripts not included in this thesis:

Paper VII: Lingqun Ye, Promi Das, **Peishun Li**, Boyang Ji, Jens Nielsen. Carbohydrate active enzymes are affected by diet transition from milk to solid food in infant gut microbiota. FEMS Microbiol. Ecol. 2019 95(11): fiz159.

Paper VIII: Sebastien Fromentin[*], Sofia K. Forslund[*], Kanta Chechi[*], Judith Aron-Wisnewsky[*], Rima Chakaroun[*], Trine Nielsen[*], Valentina Tremaroli, Boyang Ji, Edi Prifti, et.al, **Peishun Li**, Maria Zimmermann-Kogadeeva, Christian Lewinter, et.al, Fredrik Bäckhed, Jean-Michel Oppert, Jens Nielsen, Jeroen Raes, Peer Bork, Michael Stumvoll, Eran Segal, Karine Clément, Marc-Emmanuel Dumas, Dusko Ehrlich, Oluf Pedersen. Microbiome and metabolome in ischaemic heart disease. Nature Medicine. (In press)

* Contributed equally

Contribution summary

Paper I. Performed the literature review and wrote the original manuscript.

Paper II. Co-designed the study, performed omics analysis, constructed predictive model and wrote the original manuscript.

Paper III. Co-designed the study, constructed predictive model and analyzed the data and wrote the original manuscript.

Paper IV. Contributed to model analysis and manuscript revision.

Paper V. Co-designed the study, constructed predictive model, performed metabolomics analysis and wrote the original manuscript.

Paper VI. Co-designed the study, performed omics data integration and wrote the original manuscript.

Paper VII. Contributed to metagenomics analysis and manuscript revision.

Paper VIII. Contributed to model construction and simulation.

Preface

This dissertation serves as partial fulfillment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Biology and Biological Engineering at Chalmers University of Technology. The PhD research was carried out between January 2018 and January 2022 at the division of Systems and Synthetic Biology under the supervision of Jens Nielsen. The project was co-supervised by Boyang Ji and Aleksej Zelezniak and examined by Verena Siewers. The project was mainly funded by the Novo Nordisk Foundation and the Knut and Alice Wallenberg Foundation.

Peishun Li
December 2021

Table of Contents

Abstract iii

List of Publications v

Contribution summary vi

Preface vii

Abbreviations xi

1. Background 1

1.1 The human gut microbiota 1

1.1.1 Multiple factors shaping the gut microbiota 1

1.1.2 Targeting the gut microbiota as a potential health-promoting strategy 2

1.2 Relationships of the gut microbiota with human diseases 3

1.2.1 Type 2 diabetes 3

1.2.2 Osteoporosis 4

1.3 Multi-omics profiling for investigating links between the gut microbiota and human diseases 5

1.3.1 Metagenomic sequencing for characterizing the human gut microbiota 5

1.3.2 Metabolomic profiling to study the metabolisms of the human host and the gut microbiota 6

1.3.3 The microbe-derived metabolites linked to human diseases revealed by omics integration 7

1.4 Genome-scale metabolic modeling 8

1.5 Machine learning 10

1.5.1 Categories of machine learning algorithms 10

1.5.2 Workflow of machine learning modeling 11

1.6 Aim and significance 13

2. Association of the human gut microbiota with T2D 15

2.1 Links between the gut microbiota and postprandial metabolic responses 15

2.1.1 Abnormally metabolic response during the MMT in individuals with T2D 15

2.1.2 Associations of metabolomic changes with insulin resistance and glucose response 17

2.1.3 The gut microbiota associated with diabetic status and glucose response 18

2.2 Prediction of postprandial glucose response based on omics data 20

2.3 Systematical investigation of T2D-related gut microbial signatures using ML and GEMs 22

2.3.1 The overall composition of the gut microbiota associated with T2D 23

2.3.2 The functional capabilities of gut microbiota simulated by community-level metabolic models 24

2.3.3 Microbiota-based machine learning models for prediction of T2D status 26

2.3.4 Consistent T2D-related microbial features identified by classifiers of the NGT versus T2D 28

2.4 Limited performance of the microbiota-based classifiers on an independent cohort 29

3. The effect of Lactobacillus reuteri ATCC PTA 6475 on human metabolism 31
3.1 Studying the metabolism of L. reuteri ATCC PTA 6475 using GEM 31
3.2 The impact of L. reuteri ATCC PTA 6475 on the metabolic profiles of older women 33
3.2.1 The dynamic changes of metabolomic profiles during one-year probiotic intake 34
3.2.2 Differential metabolic responses relating to the probiotic effects on bone metabolism 35
3.3 The effects of L. reuteri ATCC PTA 6475 intake on the gut microbiota of the elderly 36
3.3.1 Probiotic intake reduces bone loss and decreases inflammation in the good responders 37
3.3.2 Alterations of the gut microbiota after one-year probiotic supplementation 37
3.3.3 The altered gut microbiota linked to the metabolomic changes in response to the probiotic supplementation 40
4. Conclusions 43
5. Future perspectives 45
Acknowledgements 47
References 49

Abbreviations

BMI	Body mass index
SCFA	Short-chain fatty acid
BCAA	Branched-chain amino acid
T2D	Type 2 diabetes
Pre-D	Prediabetes
NGT	Normal glucose tolerance
MMT	Mixed meal test
DL	Deep learning
ML	Machine learning
BMD	Bone mineral density
RNA	Ribonucleic acid
OTU	Operational taxonomic unit
GEM	Genome-scale metabolic model
GPR	Gene-protein-reaction
KEGG	Kyoto encyclopedia of genes and genomes
KO	KEGG Orthology
CAZy	Carbohydrate-Active Enzyme
PCoA	Principal coordinate analysis
RMSE	Root mean square error
GC-MS	Gas chromatography mass spectrometry
PKP	Phosphoketolase pathway
GR	Good responder
PR	Poor responder
HOMA-IR	Homeostasis model assessment - insulin resistance
GC	Gas chromatography
HMDB	Human metabolome database
LC-MS	Liquid chromatography mass spectrometry
ROC	Receiver operating characteristic
AUC	Area under the curve
GSA	Gene set analysis
FBA	Flux balance analysis
RF	Random forest
LightGBM	Light gradient boosting machine
XGBoost	Extreme gradient boosting decision trees
DNN	Deep neural network
RCT	Randomized controlled trial
usCRP	Ultrasensitive c-reactive protein

“It is hard to fail, but it is worse never to have tried to succeed.”

– Theodore Roosevelt

1. Background

1.1 The human gut microbiota

A vast number of microorganisms (over 10^{14} bacterial cells) collectively live inside and on human bodies, such as the intestine, oral, nasal and skin, referred to as the human microbiota [1, 2]. Thus, humans are also called superorganisms comprised of both host and microbial cells. Most of the symbiotic microorganisms reside in the intestinal tract, named as the gut microbiota. The gut microbiota consists of diverse communities of bacteria, fungi and viruses, but is dominated by bacteria from major phyla Firmicutes and Bacteroidetes [3]. The weight of the bacteria colonizing human intestine reaches about 1.5 kilograms and comprises approximately half of the feces [4]. A metagenomic analysis of individuals from four studies identified more than 1000 bacterial species and over ten million of genes in the gut microbiota [5], whose collective genome contains 100 times more genes than that in the human genome. Although the common species and core genes were shared across all cohorts, the abundances of species and genes showed large inter-individual variations.

1.1.1 Multiple factors shaping the gut microbiota

The section mainly discusses various factors that shape the gut microbiota. Babies acquire the initial gut microbiota from their mother at birth or from exposure to the environmental microorganisms according to the different delivery modes [6]. Also, the early colonization of the gut microbes in an infant is strongly linked to how the infant is fed, e.g., with breast milk or formula [7]. A previous study reported that human genetics to some extent shape the composition of the gut microbiota and a number of microbial species from the phyla Firmicutes and Verrucomicrobia are heritable [8]. Moreover, several studies have suggested ethnicity is related to the inter-individual dissimilarities in the composition of gut microbiota [9, 10]. Also, the available nutrition and innate immunity in the human host could influence the gut microbiota.

Nevertheless, environmental factors such as diet, lifestyle, anthropometric measurements, particular drugs, predominantly shape the gut microbial composition over human host genetics [11]. The gut microbiota is strongly affected by long-term diets and influences human health [12-15]. Especially, drug treatment could impact the gut microbiota and confound the microbial associations with human diseases, emphasizing it is critical to adjust for the medicine treatment when identifying disease-related microbial signatures [16]. In addition, an early study of populations from three countries revealed that the microbial diversity of the infant gut elevated with age during the first 3 years of life across all populations [17]. The study also showed that the composition and functional capacity of the gut microbiota differed between the three geographically distinct populations, consistent with that geographical location showed strong associations with variations of the microbiota in a recent report [18]. The study further revealed that the gut microbiota of an individual had more similar to members from the same household than from different

families, which implies that common living environment is an important factor for shaping the microbiota. Additionally, a metagenomic study of Ukrainian population showed that the Firmicutes/Bacteroidetes (F/B) ratio increased with the raised body mass index (BMI) [19]. Thus, the environmental factors together with human host genetics have a great potential to shape the composition and functional capacity of the gut microbiota, subsequently affecting human health state (**Figure 1**).

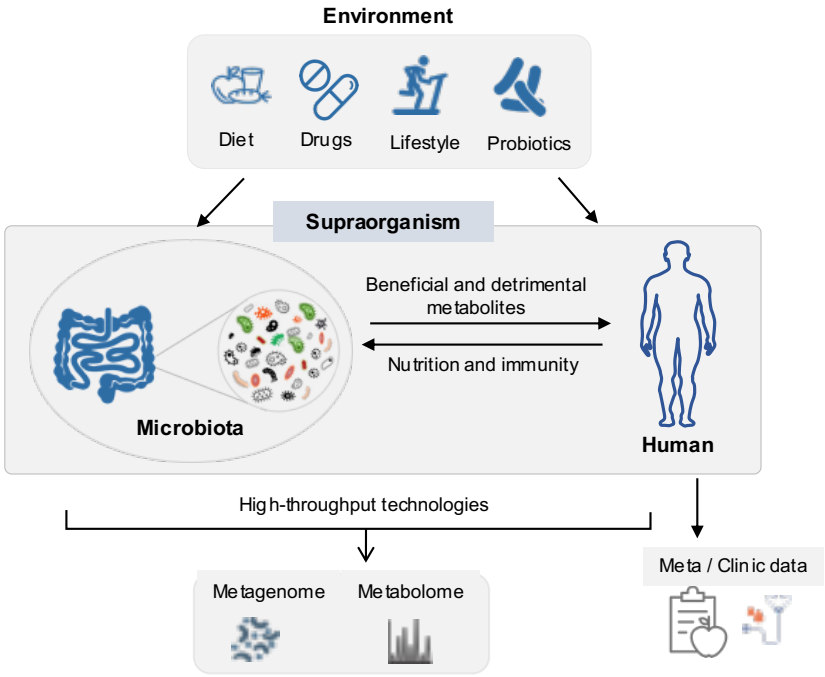

Figure 1 Complex interplays between the gut microbiota, environment and human metabolism. A number of environmental factors, such as diet, drugs, probiotics and lifestyle could influence both the gut microbiota and human host. Moreover, the gut microbiota could affect human health potentially mediated by producing the beneficial or detrimental metabolites, such as short-chain fatty acids (SCFAs), bile acids and branched-chain amino acids (BCAAs). Conversely, the human host could exert a selective pressure on the microbiota via nutritional availability and immune regulation. In order to investigate the complicated interactions related to human diseases, high-throughput technologies have been widely applied to generate multi-omics data, including the gut metagenomics and metabolomics. Integrative analysis of the multi-omics, meta or clinic data could provide more insights into the associations between the gut microbiota the human diseases.

1.1.2 Targeting the gut microbiota as a potential health-promoting strategy

As discussed above, the gut microbiota is modifiable by various environmental factors. Growing evidence has implicated that modulating the gut microbiota e.g., through dietary intervention or oral supplementation with probiotics (**Figure 1**), could be a potential intervention strategy for improvement of human heath state [20]. A metagenomic study of 49 individuals with obesity and overweight revealed that an intervention by the weight-loss diet improved gene diversity of the gut microbiota and clinical phenotypes in subjects with an initially low gene richness [21]. A recent study also showed that Mediterranean diet intervention alters the gut microbiota and improve health status in older people [22]. Furthermore, previous studies have observed high inter-person variations in postprandial glucose responses to the identical diet [23, 24], which challenges the recommendation of a standardized diet for glycemic control in individuals with cardiovascular risk. However,

the tailored meals in combination with the individual gut microbiota showed predictable of a person's metabolic response to the diets, suggesting that personalized nutrition has potential as an intervention strategy for improvement of human health.

In addition to the diet intervention, increasing studies have indicated that supplementation with probiotics (health-promoting microorganisms) or prebiotics (compounds promoting the growth of beneficial microorganisms) have positive effects on human metabolism [25-27]. In a randomized, double-blind, placebo-controlled trial, Sabico *et al* implicated that multi-strain probiotic supplementation over 6 months significantly decreased the insulin resistance and inflammation in patients with type 2 diabetes (T2D) [28]. Also, Karamali *et al* demonstrated that taking probiotic supplements in patients with gestational diabetes had positive effects on glycemic control [29]. Conversely, an early study suggested that part of subjects did not respond to the probiotic supplementation, while the responders with improved insulin sensitivity showed a higher baseline microbiota diversity [30]. Overall, personalized dietary or probiotic interventions that target the gut microbiota could be an efficient strategy for promoting human health.

1.2 Relationships of the gut microbiota with human diseases

When the commensal species are outcompeted by other pathogenic microorganisms, dysbiosis of the gut microbiota can occur. Increasing studies have demonstrated that the imbalance of the gut microbiota plays a critical role in human diseases (**Figure 1**). This thesis focuses on exploring associations between the gut microbiota and human diseases including T2D and osteoporosis. As briefly mentioned in section 1.1.2, oral supplementation with probiotics could have beneficial effects on human health. Therefore, the section further discusses the existed associations between the two diseases, gut microbiota and probiotics as well as the related mechanisms underlying the causal roles of the microbiota in the pathogenesis of the diseases.

1.2.1 Type 2 diabetes

T2D is one of the fastest increasing diseases all around the world [31, 32], characterized by hyperglycaemia. Almost all individuals with T2D have prediabetes (Pre-D), characterized by higher than normal glucose levels but not yet reaching the threshold for diabetes diagnosis [33, 34]. Moreover, 5-10% of all individuals with Pre-D will annually progress to T2D, and ~70% will ultimately develop T2D over the course of their lifespan [33]. Both the Pre-D and T2D patients undergo metabolic disorders, including the abnormal glucose and fatty acid metabolisms. In addition, they have a reduced ability to adapt to diet-triggered perturbations, e.g., the limited control for the postprandial glycemic level [35, 36]. Insulin resistance and pancreatic beta-cell dysfunction play an important role in the metabolic imbalance [37, 38]. Researchers and physicians usually apply a mixed meal test (MMT) to examine the postprandial glucose control and insulin secretion [39-42]. Also, the MMT has been used to investigate the postprandial effects on the metabolic processes in subjects with Pre-D or T2D [43-45]. Additionally, recent studies showed that

the postprandial glucose responses to a diet could be predictable using machine learning (ML) models based on the gut microbiota [23, 24].

Increasing metagenomic studies have suggested that (pre)diabetes is associated with alterations in the composition and functional capacity of the gut microbiota [46, 47]. One early study reported gut microbial dysbiosis in Chinese individuals with T2D, including a decrease in the abundance of some butyrate-producing bacteria [48]. Also, they observed an enrichment of microbial functions involved in branched-chain amino acid (BCAA) transport and oxidative stress resistance. Moreover, a previous report revealed increase in the abundance of *Lactobacillus* species and decrease in the abundance of *Clostridium* species in Swedish individuals with T2D [49]. Their results also indicated that the discriminatory microbial markers of T2D were heterogenous between the European and Chinese cohorts. Furthermore, one recent study analyzed metagenomics data from individuals with normal glucose tolerance (NGT), Pre-D and T2D [50]. Several microbial compositional changes were detected, including an enrichment of *Escherichia coli* in the Pre-D individuals and an increased abundance of *Bacteroides* spp. in the T2D patients. Additionally, a remarkable study found that the overall gut microbiota shifted in different glycemic status [51], and the butyrate-producing bacteria were depleted in the Pre-D and T2D individuals. More evidence has proved that the gut microbiota could be linked to the impaired glucose tolerance by producing detrimentally microbial metabolites [52], such as BCAA [53], imidazole propionate[54].

Nevertheless, inconsistent T2D-related gut microbial signatures have been reported across various studies. Also, different mechanisms underlying the roles of the inconsistent microbial features in T2D have been suggested. One of the main reasons is that T2D is one multi-factor disease that has an intricate interaction of human genetics, the gut microbiota and other factors [32, 55]. In addition, this might be due to other factors, such as geography, age, body mass index (BMI), diet and drugs, could affect the gut microbiota, which would possibly confound associations between the microbiota and T2D. Therefore, different types of data should be taken into account when performing analysis for the gut microbiota studies related to T2D.

1.2.2 Osteoporosis

Osteoporosis is a prevalent bone disease in the elderly, characterized by reduced bone mineral density (BMD), deteriorated bone microarchitecture and decreased bone strength. The disease increases the susceptibility to low energy or fragility fractures mainly in the older population [56, 57]. The risk of fracture could be reduced by pharmacological treatment, but treatment rates in patients with osteoporosis keep low, probably due to low osteoporosis awareness, high cost for medication and side effects of available drugs [57, 58]. Thus, there is an urgent need to develop a novel and effective intervention for the prevention and treatment of osteoporosis.

Towards this goal, the gut microbiota has been suggested to play an important role in bone metabolism, potentially by regulating the immune system and osteoclast formation in mice [59-62]. An early study of food allergic infants suggested that the probiotic *Lactobacillus rhamnosus* GG-supplemented formula could expand the butyrate-producing bacterial species [63]. Moreover, the probiotic *L. rhamnosus* GG was suggested to promote bone formation through increasing the production of the microbial butyrate, which induced T cell-produced Wnt10b in the intestine of mice [64]. These results suggest that modulation of the gut microbial composition and functional capacity by supplementation with probiotics might provide novel strategies for the prevention and treatment of osteoporosis [65].

As a lactic acid bacterium, *Lactobacillus reuteri* (also known as *Limosilactobacillus reuteri*) strains have been widely used as probiotics. Oral supplementation of the probiotic *Lactobacillus reuteri* ATCC PTA 6475 has been demonstrated to reduce bone loss and increase bone density in mice with estrogen deficiency or increased inflammation [66, 67]. Moreover, in a recent randomized controlled trial, oral supplementation of *L. reuteri* ATCC PTA 6475 could reduce bone loss by ~50% in older women with low BMD [68]. Thus, *L. reuteri* ATCC PTA 6475 may be a potential therapeutic strategy to prevent postmenopausal bone loss in the elderly. However, the mechanisms related to the effects of the probiotic *L. reuteri* ATCC PTA 6475 on bone metabolism remains unknown.

1.3 Multi-omics profiling for investigating links between the gut microbiota and human diseases

The fast developments of high-throughput omics technologies have enabled the quantifications of a large number of molecular features from different biological samples, such as metabolomics for the metabolites abundancies, metagenomics for the taxonomic and functional profiles of the gut microbiota and transcriptomics for the expression levels of ribonucleic acid (RNA). In this thesis, the serum metabolomics and gut metagenomics were jointly used to explore links between the gut microbiota and the studied diseases (**Figure 1**). Therefore, the following section mainly introduces metabolomics and metagenomics technologies.

1.3.1 Metagenomic sequencing for characterizing the human gut microbiota

To quantify and characterize the gut microbial communities, feces samples are first collected. Traditionally, target microbes are isolated from the feces sample and then cultured in a laboratory media. Due to the difficulty to grow most of the microorganisms within the human gut as well as the sequencing cost is decreasing dramatically, DNA based sequencing methods have been widely used, including amplicon sequencing and metagenomic shotgun sequencing. The amplicon sequencing mainly profiles 16S ribosomal RNA (16S rRNA) that contains around 1500 base pairs and are regarded as main markers for bacteria and archaea [69]. When analyzing sequences from the 16S rRNA genes, close sequences are classified into operational taxonomic units (OTUs). A subset of

bioinformatics tools and databases have been well developed and easily available for the taxonomic identification and quantification at a genus or species level. However, many human gut microbes lack reference genomes, and species with similar 16S rRNA genes could exhibit differential functional potential. Therefore, based on the taxonomic markers, it is challenging to characterize the functional capacities of the gut microbiota by 16S rRNA sequencing.

Metagenomic shotgun sequencing has enabled to profile not only the microbial composition but also its functional capacity by quantification of microbial genes or metabolic pathways. A number of computational tools for the taxonomic and functional profiles of the metagenome have been devised. To profile the composition of the microbial communities from metagenomic data, several marker gene-based computational tools have been proposed such as the tool MetaPhlAn [70] and mOTUs2 [71]. The MetaPhlAn is based on unique clade specific marker genes collected from about 17000 reference genomes, while mOTUs2 is based on phylogenetic marker genes that could profile over 7700 species.

To characterize the potential functions of the gut metagenomes, a set of bioinformatic tools also have been developed and mainly grouped into two classes. The first type of tools quantify directly the functional capacity by mapping metagenomic reads to a predefined catalogue of genes with known functions, such as the online metagenomics RAST service [72], the standalone tool HUMAnN2 [73]. Similarly, the MEDUSA tool is an integrated pipeline for analysis of metagenomic sequences, which maps reads to a global human gut microbial gene catalogue comprising over 11 million genes [5]. The second type of tools first perform a de novo assembly of metagenomic reads into a catalogue of contigs or genes, and then map the reads to the assembled genes with functional annotations. For example, by calling the de novo assembly tool SOAPdenovo [74], MOCAT has been devised as a highly modular pipeline for standardized processing, assembly and profiling of metagenomic data [75]. Usually, the assembled genes are annotated with functional information from databases, such as the Kyoto encyclopedia of genes and genomes (KEGG) [76], COG [77], eggNOG [78] and the Carbohydrate-Active Enzyme (CAZy) database [79].

1.3.2 Metabolomic profiling to study the metabolisms of the human host and the gut microbiota

Metabolomics refers to a collection of high-throughput technologies used to identify and quantify a large number of small-molecule chemicals (<1500 Da) in a biological sample. The detectable molecules mainly consist of both endogenous metabolites naturally produced in an organism (including fatty acids, amino acids, carbohydrates, nucleic acids) and exogenous metabolites (including food additives, drugs and other xenobiotics) not naturally synthesized in an organism. Thus, metabolomics offers a great opportunity to investigate the global metabolic processes in human populations [80, 81].

Metabolomics technologies are mainly categorized into targeted and untargeted methods. Targeted methods measure a predefined set of small molecules with high sensitivity, while untargeted approaches quantify a broader range of detectable compounds. In addition, liquid chromatography (LC) or gas chromatography (GC) in tandem with mass spectroscopy (LC-MS or GC-MS) and nuclear magnetic resonance (NMR) spectroscopy have been widely used to quantify metabolites in metabolomics. This thesis mainly applied the LC-MS technology that generally has a high sensitivity and broad scope of detectable metabolites. After qualification, based on in-house or public databases, all identified molecules are annotated to known metabolites with various identifiers from public databases, such as KEGG [76], the human metabolome database (HMDB) [82] and the PubChem database [83]. After identification and quantification of molecules, the missing values of metabolites' peak areas can be imputed in different ways, e.g., using the minimum value of each metabolite abundance. Due to the high variability of the metabolite abundances up to three orders of magnitude, the raw values are usually processed by the logarithmic transformation to restrict the range of values approximately fulfilling a normal distribution. For metabolomic analysis, the web based MetaboAnalyst tool has been developed and widely used to perform biomarker identification, pathways analysis and multi-omics integration [84].

1.3.3 The microbe-derived metabolites linked to human diseases revealed by omics integration

As discussed in previous reviews [52, 85-87], the gut microbiota may contribute to human diseases mediated by the microbe-derived metabolites involved in the key signaling pathways (**Figure 1**). Particularly, the integrative analysis of the metagenomics and metabolomics data not only could provide new insights into the metabolisms of the gut microbiota or human host, but also study the causal links between them. For example, Visconti *et al* found that the gut microbial metabolic pathways have over 18,000 significant associations with blood and fecal metabolites [88], whereas species show less than 3,000 associations. To examine relationships between blood metabolome and the gut microbiota, Wilmanski *et al*. predicted alpha diversity of the gut microbiota based on a set of 40 plasma metabolites [89]. Out of the 40 metabolites, 13 are microbe-derived metabolites including imidazole propionate, secondary bile acids, trimethylamine N-oxide (TMAO) and indole propionate, which are linked to cardiovascular diseases (CVD) risk and T2D. In a cross-sectional study, Kurilshikov *et al* showed that plasma levels of short-chain fatty acids (SCFAs) from the gut microbial fermentation of fibers were associated with inflammation and CVD risk [90]. Moreover, Pedersen *et al* identified *Prevotella copri* and *Bacteroides vulgatus* as the main drivers, which induced insulin resistance via the production of branched-chain amino acids (BCAAs) [53]. These demonstrated that the common cardiometabolic disorders could be regulated by the microbial metabolites. In addition, by integrating metabolomic and metagenomic data, Franzosa *et al* identified a number of associations between inflammatory bowel disease-related species and metabolites including caprylic acid, which provides an insight into possible mechanism involved in dysfunction of the gastrointestinal tract [91].

1.4 Genome-scale metabolic modeling

As discussed in section 1.3, increasing studies have accumulated tons of multi-omics data generated from high-throughput technologies, such as transcriptomics, metagenomics and metabolomics. Thus, we are facing major challenges to efficiently extract useful information by integrative analysis of these omics data. Systems biology applies mathematical models or networks to study complex biosystems that contain varied molecular components. Due to considering the intricate interactions between the different molecular components, systems biology has the potential to reveal latently novel signatures that might not be identified through analysis of a single molecular profile. Particularly, in systems biology, genome-scale metabolic models (GEMs) have been a powerful tool to study the metabolisms of an organism in detail. Therefore, this section mainly introduces the framework for GEM construction and analysis.

GEMs contain the detailed collections of biochemical reactions for all metabolic genes in an organism. To construct a GEM of an organism of interest, the gene-protein-reaction associations were first collected mainly based on the genomic content and annotation information from several genome and biochemical reaction databases such as KEGG [76], MetaCyc [92] and NCBI (**Figure 2a**). A number of widely used tools for the model construction have been developed such as Model SEED [93], COBRA[94] and RAVEN [95], which could automatize many steps of the construction and generate an initial draft model. After obtaining a draft model, several manual curation steps are required inevitably, such as biomass reaction definition, parameter optimizations for biomass growth and gap filling. Biomass composition can be determined according to literature and experimental data. Due to the manual steps, it would take most of time to refine the draft model to generate a finalized GEM with completely metabolic pathways that could convert substrates into biomass components. When performing the GEM simulation, all biochemical reactions in a GEM are formulated as a stoichiometric coefficient matrix *S*, where rows represent metabolites and columns represent reactions. Flux balance analysis (FBA) has been widely used to simulate reaction fluxes at the steady state when maximizing an objective function under a certain number of constraints. This can be formulated mathematically as shown in **Figure 2b**.

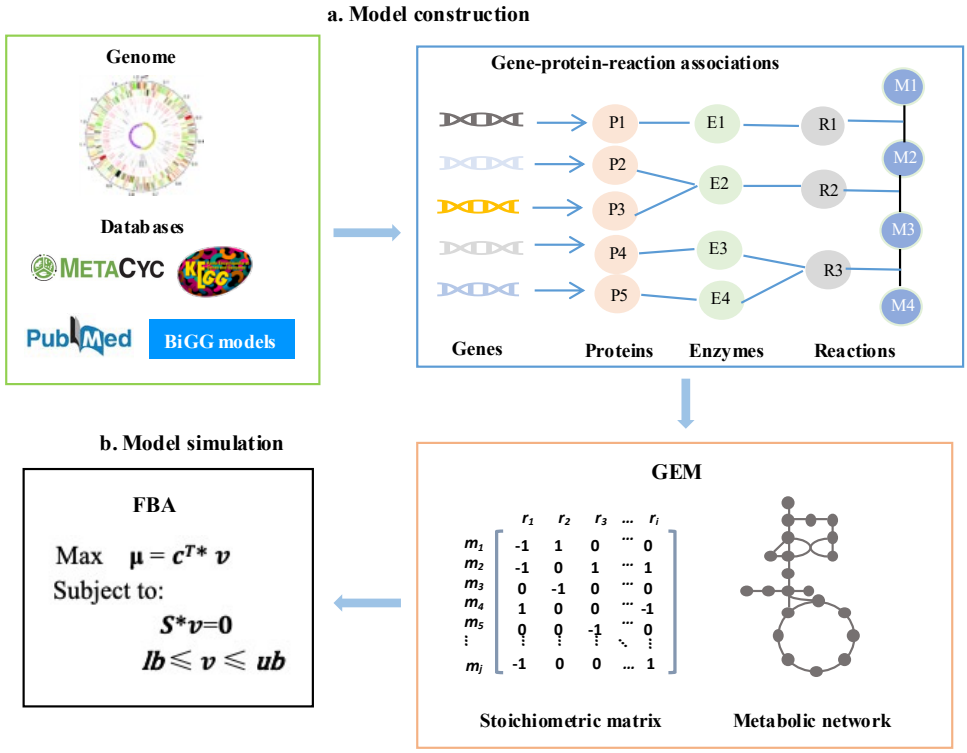

Figure 2 The framework for construction and simulation of a genome-scale metabolic model. a) Firstly, the gene-protein-reaction (GPR) associations were collected based on the genomic contents and the public databases. According to the obtained GPR associations, the stoichiometric description of metabolic reactions, genes and metabolites are integrated into an initial draft model. After manual curations, a finalized GEM is achieved with a complete metabolic network that could convert substrates into biomass components. b) For the GEM simulation, the metabolic network is defined as a stoichiometric coefficient matrix ***S***, where rows represent metabolites and columns represent reactions. Flux balance analysis (FBA) is usually used to simulate metabolic fluxes at a steady state when maximizing an objective function under given condition. c is a vector with coefficients for all reactions that specify a linear combination of all reaction fluxes to be maximized; v is a vector with fluxes of all reactions; lb (ub) denotes a vector with lower (upper) bounds for all reactions; μ indicates objective function that is maximized to simulate metabolic fluxes.

Until now, abundant GEMs have been constructed for different organisms, including the human gut related species. GEMs have been used to investigate the metabolism of a single gut microbial species, such as *L. reuteri*, *Lactobacillus casei* [96-98]. Additionally, GEMs have been applied to examine the complex interrelationships between multiple gut microbial species [99]. Particularly, GEMs provide a systematical platform to view the community-level metabolic potentials of the gut microbiota [100]. To interrogate the associations between diseases and the metabolic capabilities of the gut microbiota, the collective reaction fluxes can be simulated by optimizing some potential products or maximizing biomass growth. The simulated fluxes from metabolic models may be helpful for understanding the abnormal metabolisms or interrelations in the disease-related gut microbiota.

1.5 Machine learning

In addition to GEM as discussed in section 1.4, ML methods have been successfully applied to integrate multi-omics data for discovery of hidden patterns, phenotypic predictions and identifications of potential biomarkers in the field of human gut microbiota. ML is a branch of artificial intelligence that automatically learn and improve from input data without being explicitly programmed. The section mainly introduces the categories of ML algorithms and the general workflow for ML modeling.

1.5.1 Categories of machine learning algorithms

ML algorithms are mainly classified into two categories: unsupervised and supervised learning (**Figure 3a**). Unsupervised learning methods purely learn and discover novel hidden patterns from given datasets, therefore they are referred to as data driven prediction. Out of them, clustering algorithms, for instance k-means clustering [101], are frequently implemented to stratify a set of objects into multiple groups (clusters) based on similarities or differences. Particularly, unsupervised learning has been applied to novel pattern recognition in the gut microbiota studies, such as identifying enterotypes of the human microbiota [102, 103], co-abundance gene groups [104].

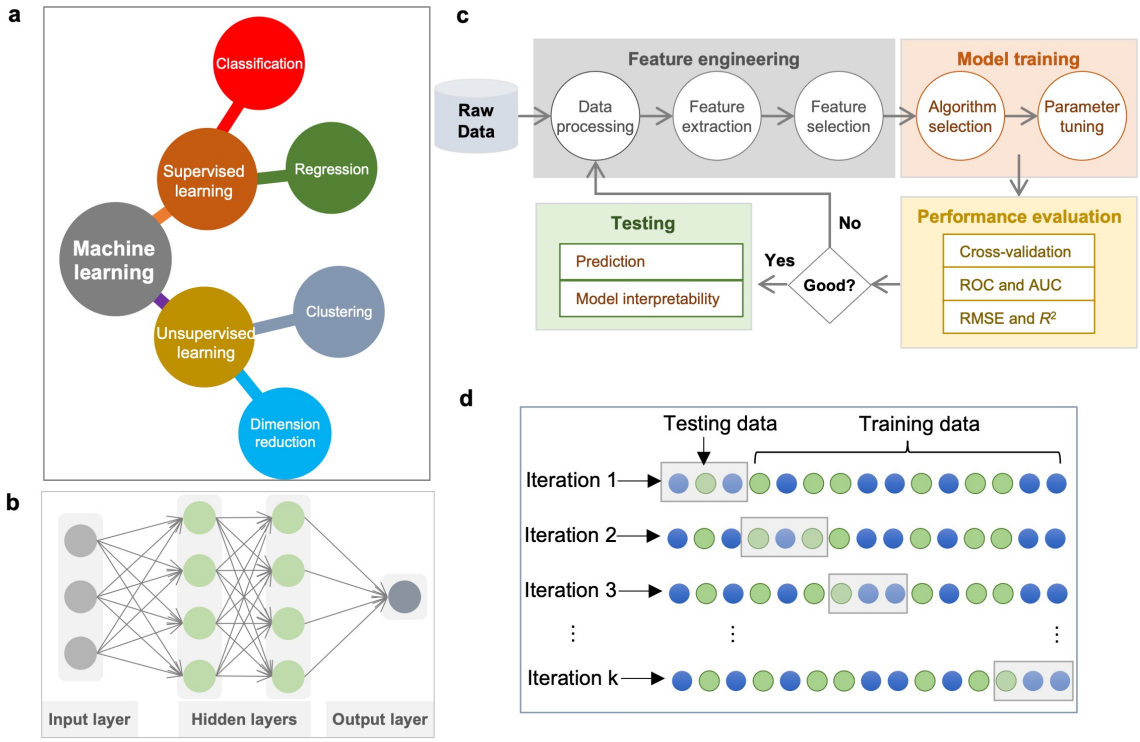

Figure 3 The categories and workflow of machine learning modelling. a) ML algorithms are mainly classified into unsupervised learning including dimension reduction and clustering methods, and supervised learning, including regression and classification approaches. b) In a deep neural network architecture, multiple (here two) hidden layers (green color) are connected in a cascade fashion between input and output layers (grey color). Each of these layers takes input from its previous layer and transforms the data into a more abstract form as an output for next layer. c) The pipeline of ML modeling commonly consists of four steps, including feature engineering, model training and optimization, performance evaluation, model application and explanation. d) The framework of a k-fold cross-validation, where the original samples are randomly split into k subsets with equal size.

In contrast to the unsupervised learning, supervised learning approaches learn and infer a function from input data, which is typically comprised of independent variables (i.e., features) and dependent variables across all samples. For supervised learning, the known dependent variables in a training dataset are used to train a ML model, which is potentially capable to predict the outcomes of new samples. This thesis focuses on the supervised ML algorithms for classification or regression problems. While the dependent variables are continuous, the ML model can be used for regression tasks [89, 105]. For instance, the generalized linear models with the penalties least absolute shrinkage and selection operator (lasso) and ridge regression [106] has been widely used in the gut microbiota studies, due to that they can efficiently process sparse microbial features. When the dependent variables are categorical, the ML model can be applied for classification tasks [49, 107]. Particularly, decision trees-based ensemble learning methods have been widely applied in the gut microbiota studies for both regression and classification tasks, such as random forest (RF) [49, 107, 108], light gradient boosting machine (LightGBM) [109, 110] and extreme gradient boosting decision trees (XGBoost) [111-113], due to their powerful performance, ease of use and model interpretability. Additionally, as a subfamily of ML methods, deep learning (DL) is a deep neural network (DNN) with multiple hidden layers [114]. In a DNN architecture as illustrated in **Figure 3b**, hidden layers are connected in a cascade fashion between input and output layers with weight representing each connection. Each of these layers takes input from its previous layer and transforms the data into a more abstract form as an output for next layer. Moreover, the backpropagation approach is utilized to adjust the weights to minimize the prediction error. Increasing studies have applied DL algorithms that could achieve a considerably accurate prediction [115, 116]. Nevertheless, DL algorithms usually need large training data sets and lack model explainablity, which limits their applications in the gut microbiota.

1.5.2 Workflow of machine learning modeling

Although a set of supervised ML algorithms have been developed, the whole pipeline of modeling commonly consists of four steps: 1, feature engineering; 2, model training and optimization; 3, performance evaluation; 4, testing of the optimal model (**Figure 3c**). Performance of a ML model lies to some extent on the quality of data used for training the model. Thus, it is essential to perform feature engineering first, which is involved in data pre-processing, feature extraction and feature selection processes. Data pre-processing includes proper cleaning, normalization and transformation. Feature extraction is intended to build a feature vector representing a reduced number of variables from raw measured data, which could contain the sufficient relevant information for the raw data. This can facilitate the subsequent training steps. Nevertheless, these extracted features in the dataset might be still uninformative and irrelevant for building the predictive model. For example, model training with extremely large amounts of variables requires extensive computing power and memory, and easily leads to overfitting. Thus, feature selection is important to obtain an optimal and non-redundant subset of the initial features, which is critical for fast model training, improved performance and even better model interpretation.

In addition, cross-validation has been frequently applied to evaluate model performance using assessment metrics such as AUC (area under ROC curve) for classification task as well as root mean square error (RMSE) and R^2 (coefficient of determination) for regression task (**Figure 3c**). In a k-fold cross-validation process, the original samples are randomly split into k subsets with equal size at first (**Figure 3d**). Then one round of cross-validation is implemented, where the predictive model is constructed using k-1 subsets (called training set) and the model is validated using the single remaining subset (called as testing dataset). This step will be iterated k times, where each of the k subsets is used successively as the testing dataset. Finally, the k validation outcomes are summarized into a single metric for assessing model performance. For most of ML methods, the training process includes iterations of model parameters tuning and feature engineering until the model performance cannot be improved further. The performance of multiple different approaches can be benchmarked and then the one or two best models can be selected. Finally, the model can be applied to make prediction on new data. Notably, disease-related biomarkers can be simultaneously identified by model interpretability in previous microbiota studies [109, 112] (**Figure 3c**), which allows us to gain biological insights into the data. Overall, the above processes can impact the model performance and thus should be taken into account when implementing a ML algorithm in the gut microbiota research.

1.6 Aim and significance

Given the heavy medical costs and increasing trend of T2D and osteoporosis around the world, there is an urgent need to find novel ways of addressing this global challenge. Evidence suggests that the gut microbiota plays a key role in the onset and progression of the human diseases. Particularly, probiotics might provide novel interventions strategies for prevention and treatment of the diseases. By investigating the metabolisms of both the gut microbiota and human host using metabolomics and metagenomics, the mechanistic effects of the gut microbiota on the human diseases have been revealed in previous studies. In addition, GEMs and ML have been successfully applied to data integration in the gut microbiota studies. With this background, this thesis aims to disentangle the associations between the gut microbiota, probiotics and human diseases (i.e., T2D and osteoporosis) by integrative analysis of plasma metabolomics and gut metagenomics, using ML and GEMs. To this end, I first reviewed the current literature about associations between ML, gut microbiota and human diseases in **Paper I**. Moreover, this thesis focuses on answering three scientific questions as follow:

How is the human gut microbiota linked with T2D?

As introduced in the background part (section 1.2.1), few studies have taken a systematical investigation on how different underlying factors, including human metabolism and the gut microbiota, contribute to the abnormal metabolic responses to a MMT in individuals with (pre)diabetes. Therefore, this work first explored the metabolic changes of both human host and the gut microbiota, and how the identified link between them might play an important role in postprandial abnormalities (**Paper II**). In addition, inconsistent T2D-related gut microbial signatures have been reported across various studies. Accordingly, the different mechanisms that the gut microbial species are involved in might contribute to the pathogenesis of T2D. Given that other factors could influence the gut microbiota and cause the inconsistent findings, this work further performed a systematical analysis of four metagenomics datasets using ML and community-level metabolic models (**Paper III**). Through the cross-cohort analysis, the common T2D-related gut microbial features and interactions across studies could be identified, which would be robust biomarkers for T2D and assist us to develop new T2D-specific interventions strategies.

How does the probiotic *L. reuteri* ATCC PTA 6475 improve bone health in the elderly?

In a recent randomized controlled trial, oral administration of *L. reuteri* ATCC PTA 6475 reduced bone loss in older women with low BMD [68]. However, part of older women responded poorly to the probiotic intake. In addition, the mechanisms related to the beneficial effects of *L. reuteri* ATCC PTA 6475 on human metabolism remains unknown. To systematically investigate interactions between *Lactobacillus reuteri* ATCC PTA 6475, the gut microbiota and bone metabolism in the elderly, this work first examined the metabolic properties of *L. reuteri* ATCC PTA 6475 by constructing its GEM, which can help us understand the potential benefits of the probiotics to human metabolism (**Paper**

IV). Moreover, using plasma metabolomic profiling, this work investigated the effects of *L. reuteri* ATCC PTA 6475 on global metabolisms of older women during one-year supplementation with the probiotics (**Paper V**). To interrogate whether the metabolic changes of the elderly are linked to the gut microbial changes, this work further analyzed the metagenomics data from 20 older women with good or poor responses to the probiotic supplementation (**Paper VI**). Through the integrate analysis, this work could provide new insights into the probiotic regulation of bone metabolism that might aid in the development of novel interventions strategies for osteoporosis.

Can the joint use of ML and GEM enable the identification of key gut microbial signatures related to diseases?

GEMs have served as a useful tool for studying detailed metabolism of an organism, e.g., the GEM reconstruction for *L. reuteri* ATCC PTA 6475 in **Paper IV**. The gut microbiota of one person consists of hundreds of species. Although it is challenging to construct community-level metabolic models based on GEMs of all gut microbes, this work hypothesized that the metabolic model of individual gut microbiota has the potential to reveal more detailed functional capacity as well as interspecies interactions at the metabolic level, compared to only analyzing the metagenomics data. Thus, the metabolic capacity by modeling the gut microbiota was simulated in **Paper III**.

ML has been successfully applied in the field of the gut microbiota as discussed in section 1.5. Therefore, in this thesis I predicted regression or classification questions by developing various interpretable ML models including regression models and the decision trees-based ensemble models, which could not only achieve adequate prediction accuracy, but also identify key disease-related signatures. To identify potential factors for postprandial glucose control, ML was applied to predict the glycemic responses to a meal based on multi-omics data (**Paper II**). Moreover, I used ML to predict T2D status based on different gut microbial features as well as to identify T2D-related microbial signatures (**Paper III**). Additionally, this work could also confirm whether the ML in combination with GEM could be helpful to identify novel gut microbial metabolic features related to diseases.

2. Association of the human gut microbiota with T2D

The gut microbiota plays a critical role in the pathogenesis of T2D. This chapter summarizes two studies (**Paper II & Paper III**) on the association between the gut microbiota and T2D. In chapter 2.1 and 2.2, the results from **Paper II** are presented where disfunction of the gut microbiota is linked to abnormally response to a mixed meal test (MMT) in T2D. The chapter 2.3 and 2.4 describe the work from **Paper III** where the consistent T2D-related microbial signatures across different studies were investigated by using ML and GEMs.

2.1 Links between the gut microbiota and postprandial metabolic responses

T2D is one of the fastest increasing diseases worldwide, characterized by hyperglycaemia. Before individuals develop T2D, they always have prediabetes (Pre-D) with higher than normal blood sugar levels that have not yet reached the threshold for diabetes diagnosis. Metabolic disorders in Pre-D and T2D patients lead to a decreased ability to adapt to diet-triggered perturbations [35, 36], e.g., abnormally glycemic control. Usually, a MMT can be used to assess postprandial metabolism, including glucose and insulin responses [39-42]. Previous studies have revealed postprandial effects on the multiple metabolic processes in the Pre-D and T2D patients [43-45]. In addition, two studies have predicted postprandial glucose response to a meal based on the composition of the gut microbiota using ML models based on the gut microbiota and personal features [23, 24]. However, few studies have taken a systematical view on how different underlying factors, including gut metagenomic and blood metabolomic profiles, contribute to abnormally metabolic responses to a MMT in individuals with (pre)diabetes.

In **Paper II**, 106 individuals were recruited and classified into either normal glucose tolerance (NGT, n=27), Pre-D (n = 57) or T2D (n = 22) groups according to the American Diabetes Association criteria [34]. Moreover, plasma samples for the two-hour MMT and metabolomics profiling were collected within three months before the bariatric surgery. Also, biopsies from different human organs including liver, jejunum and adipose fat tissues for RNA sequencing, and fecal samples for metagenomic shotgun sequencing were collected on the day of the surgery.

2.1.1 Abnormally metabolic response during the MMT in individuals with T2D

To evaluate the postprandial responses of glucose and insulin, individuals with different diabetic status underwent a two-hour MMT. The MMT triggered a temporary increase in plasma glucose and insulin levels in the NGT, Pre-D and T2D groups (**Figure 4a** and **b**; $P < 0.01$ by ANOVA). Especially, glucose excursions differed significantly between the three groups ($P < 0.01$ by ANOVA). This translated in significant differences in total area under the curve (tAUC) between the three groups (**Figure 4c**; $P < 0.01$ by Kruskal–Wallis test). Interestingly, a significantly positive correlation between glucose tAUC and insulin

tAUC was only found for the NGT group (**Figure 4d**; R = 0.58 and P = 0.0015). Thus, these results suggested an abnormally postprandial responses of plasma glucose and insulin in the Pre-D and T2D groups.

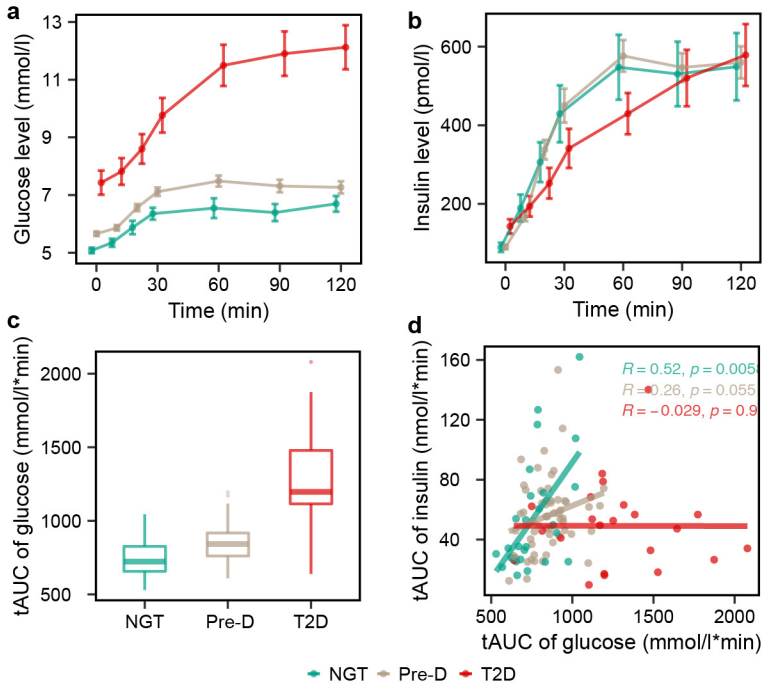

Figure 4 The postprandial responses of glucose and insulin to a mixed meal test (MMT). a) The time profiles of plasma glucose, b) insulin levels during a MMT (Mean ± SEM) in the NGT (n=27), Pre-D (n=57) and T2D (n=22) groups. c) Comparison of total area under the curve (tAUC) of glucose level between the three groups. d) The association between insulin and glucose tAUC in each group. Spearman's rank correlation analysis was performed.

By using the untargeted metabolomic profiling of plasma samples collected at fasting and two hour post MMT, this work further investigated the global metabolic responses to the MMT in individuals with different diabetic status. A total of 145 differential metabolites were identified to be associated with diabetic status (Adjusted P<0.05 by ANOVA; **Figure 5a**), mainly consisting of metabolites involved in the classes of lipids (n = 83), amino acids (n = 34), xenobiotics (n=6) and carbohydrates (n = 4). Due to the particular interest in the (pre)diabetes-related metabolites' responses to the MMT, these metabolites were further classified into three different types of response patterns by the ANOVA analyses (**Figure 5a** and **b**). Type I metabolites have no significant main effect for time and no interaction of two main effects time and groups, i.e., no response to the MMT. With this, 39 metabolites showed a response pattern with no significant difference between the two time points in each group (Type I; check details in **Paper II**). Type II metabolites have significant main effect for time but no interaction, i.e., parallel response to the MMT. Out of them, 55 metabolites showed a parallel response to the MMT, independent of diabetes status (Type II; **Figure 5b**). Type III metabolites have significant interaction of two main effects, i.e., differential response to the MMT. The remaining 51 metabolites showed a differential response to the MMT among the three groups (Type III). Therefore, in addition to glucose and insulin, other T2D-related metabolic signatures, including 1,5-

anhydroglucitol, mannose, creatinine, lysine, N−acetylaspartate (NAA), 3−hydroxydecanoate and 3−hydroxyoctanoate, showed a differentially postprandial responses among the NGT, Pre-D and T2D groups (**Figure 6**), identified by the metabolomic analysis. These metabolites with differential responses to the MMT might reflect the abnormal metabolisms in the T2D patients after diet.

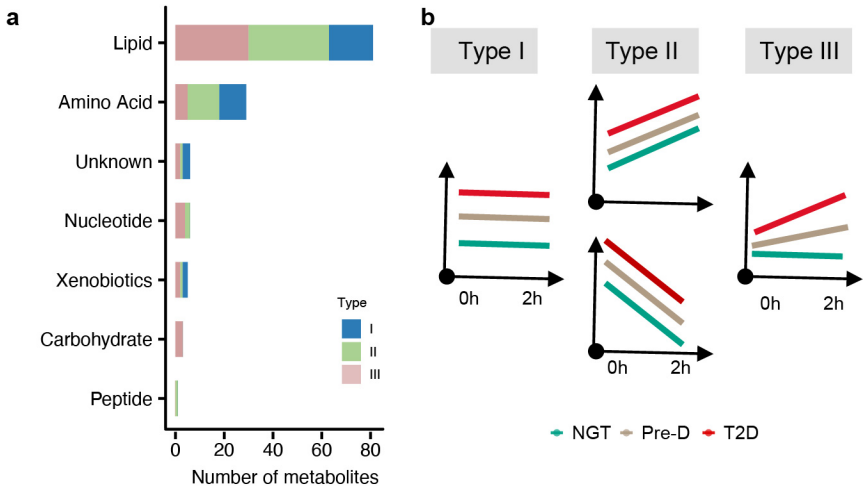

Figure 5 The global metabolic responses to the two-hour MMT. a) The (pre)diabetes-related metabolites showed three different types of response patterns. b) The three types of response patterns were classified by the ANOVA analysis. The left plot shows where the time profiles of type I metabolites have no change and are parallel for the groups (parallel means no interaction). The middle plots show where the time profiles of type II metabolites have changes but are still parallel for the groups. The last plot shows where the time profiles of type III metabolites have different changes for the three groups.

2.1.2 Associations of metabolomic changes with insulin resistance and glucose response

The correlations between the clinical variables and the T2D-related metabolites were assessed at fasting and two hour post MMT, respectively (**Figure 6**). The carboxyethyl derivatives of BCAAs and phenylalanine were positively correlated with glucose tAUC and HOMA2-IR at both time points (adjusted $P < 0.05$, R=0.29~0.55), which indicates that these metabolites might be associated with insulin resistance and glucose intolerance. Consistently, several studies in both rodents and humans have observed alterations in BCAA and amino acid metabolites in relation to insulin resistance [53, 117, 118]. In addition, previous studies have suggested that the gut microbiome of individuals with insulin resistance has an increased capacity to produce amino acids and specifically BCAA [53, 54].

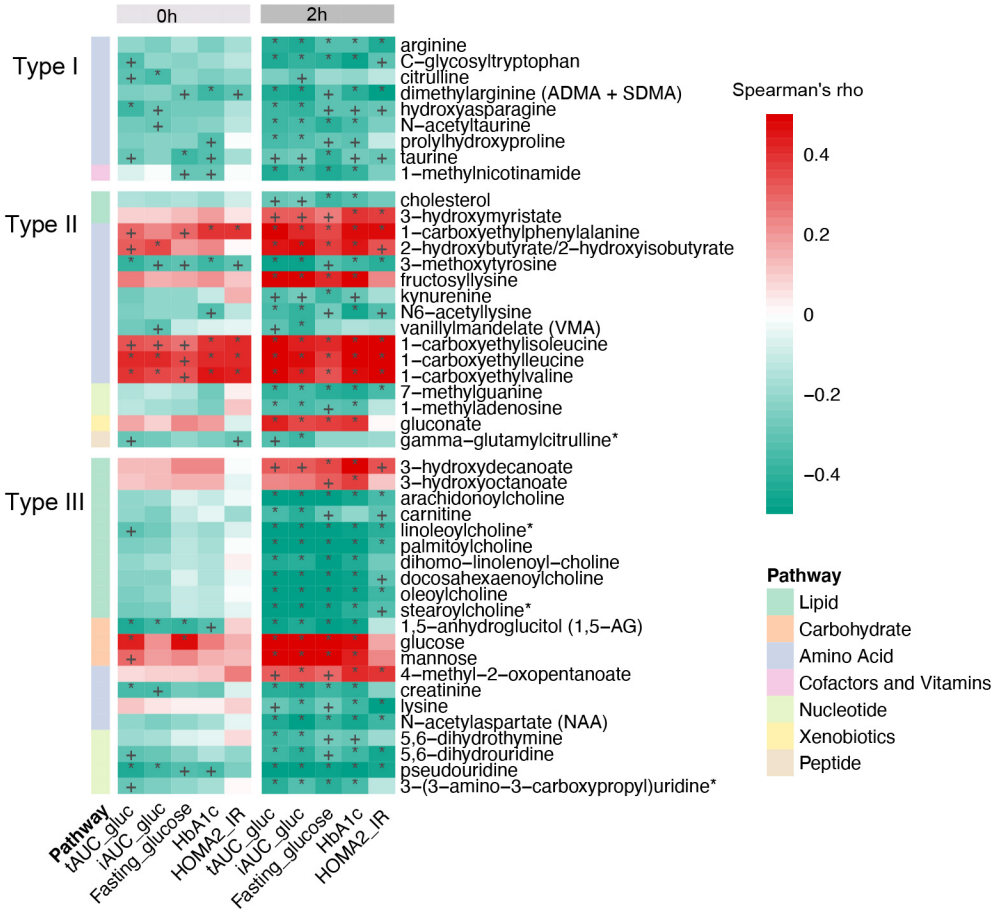

Figure 6 The associations between the metabolomic changes and the T2D-related clinic variables. Only metabolites involved in the metabolic processes, including carbohydrates, amino acids, cofactors, nucleotides, xenobiotics, peptides, acylcholines, fatty acids, carnitine and sterol metabolism are shown. Spearman's rank correlation analysis was performed. '+' denotes adjusted $P < 0.05$; '*' denotes adjusted $P < 0.01$.

2.1.3 The gut microbiota associated with diabetic status and glucose response

To investigate the association of gut microbiota with (pre)diabetes, the metagenome of all individuals was quantified using DNA shotgun sequencing. Principal coordinate analysis (PCoA) revealed that the NGT and T2D groups to some extent were separated by the second principal coordinate that accounts for 11% of the variability (**Figure 7a**). Additionally, PERMANOVA analysis showed that the diabetic status was associated with dissimilarities in gut microbiota composition (R^2= 0.027, P <0.05). Furthermore, a total of 24 species exhibited differential abundances in two or three pairwise comparisons between NGT, Pre-D and T2D groups (adjusted P <0.01; **Figure 7b)**. The abundances of nine species of genus *Streptococcus*, *Lactobacillus sanfranciscensis* and *Lactobacillus ruminis* were enriched, whereas the abundances of seven species of genus *Clostridium*, *Turicibacter sanguinis*, *Anaerococcus lactolyticus* and *Paenibacillus polymyxa* were depleted in the T2D group (adjusted P <0.01). In this study, a reduction of *Clostridium* species (*C. butyricum* and *C. novyi*, etc.) with butyrate producing capacity [119, 120] was observed, which is in accordance with a recent paper showing that a number of butyrate-producing species and the functional potential were depleted in individuals with (pre)diabetes [51]. In addition, *C. novyi* showed a significantly negative correlation with

glucose incremental AUC (iAUC, subtracting the baseline values of tAUC; R = -0.49, $P <$ 1.0e-06; **Figure 7c**), which suggests dysbiosis of the gut microbiota in T2D patients might be linked to the abnormal glycemic control. An early study reported that replenishment of butyrate producing bacteria in individuals with T2D could be a personalized approach to improve postprandial glucose control [121].

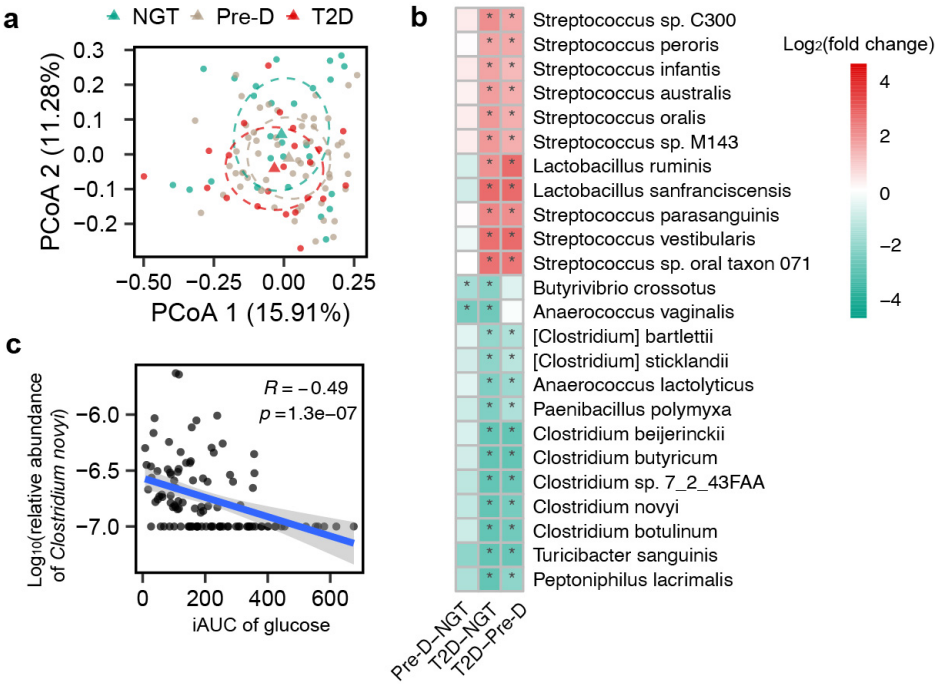

Figure 7 Alterations in the gut microbiota related to diabetes status. a) Principal coordinate analysis of microbiota community at species level based on Bray–Curtis distance (n=106). The centroid for each group is represented as a triangle and the ellipse covers the samples belonging to the group with 95% confidence. b) Heatmap showing $\log_2$ fold changes of 24 significantly differentially species between the NGT (n=27), Pre-D (n=57) and T2D (n=22). Only species exhibiting differential abundance in two or three pairwise comparisons are shown. '+' denotes adjusted $P < 0.05$; '*' denotes adjusted $P < 0.01$. c) The association between *Clostridium novyi* and glucose incremental AUC (iAUC). Spearman's rank Pearson's correlation analysis was performed.

Further, the functional capacity of the gut microbiome in the NGT, Pre-D and T2D groups was investigated. Enrichment of phenylalanine and phenylacetate metabolism capacity of the microbiome in individuals with Pre-D and T2D was observed by gene set analysis (P<0.05; **Table 1**). The microbial genes including *hcaC*, *hcaF*, *tynA*, *feaB, paaA* and *paaE* involved in phenylalanine metabolism were more abundant in the T2D group compared to the NGT or Pre-D group (P<0.01 and |$\log_2$ (fold change) |>3; **Table 1**). In line, microbial products of aromatic amino acid metabolism, in particular phenylacetic acid, have previously been linked to insulin resistance and thrombosis risk [122, 123]. Recently it was reported that phenylalanine-derived metabolites increased after autologous fecal microbiota transplantation (FMT) in individuals with liver steatosis [124]. In this study, the carboxyethyl derivatives of BCAAs and phenylalanine were also observed to be positively correlated with HOMA2-IR and glucose indexes by the metabolomic analysis. Therefore, through integrative analysis of the metabolomics and metagenomics, the

carboxyethyl derivatives of BCAAs and phenylalanine might be potential biomarkers for (pre)diabetes.

Table 1. The enriched KEGG pathways and modules in gut microbiome between the NGT (n=27), Pre-D (n=57) and T2D (n=22) groups identified by gene set analysis.

	Differential genes ($P < 0.01$)	
	T2D vs NGT	**T2D vs Pre-D**
KEGG pathway		
Phenylalanine metabolism	*tynA, feaB, paaA, paaC, paaD, paaE, paaJ*	*hcaC, hcaF, paaJ*
KEGG module		
Phenylacetate degradation	*paaA, paaC, paaD, paaE, paaJ*	–

Note: significantly enriched pathway or module comparing two groups ($P < 0.05$). '–' denotes no differential genes in the pathway or module.

In short, this section 2.1 presents the results from **Paper II** that systematically characterized the metabolic response to a MMT in individuals with different glucose tolerance. From plasma metabolomic profiling, the abnormal metabolic processes related to (pre)diabetes before and after meal intake were first identified. This work found more differential metabolites between the NGT and T2D groups after the meal intake compared to fasting condition, thus enabling us to discover abnormal metabolism related to (pre)diabetes that did not appear at fasting condition. Furthermore, this work identified three different types of response patterns in the 145 metabolites that were associated with diabetic status. The derivatives of BCAAs and phenylalanine were found to be associated with glucose control and HOMA2-IR. Further the gut microbial composition and functional capacity associated with T2D and glucose intolerance were investigated. In agreement with metabolomic analysis, the phenylalanine and phenylacetate metabolism capacity of the metagenome was enriched in individuals with T2D. Thus, through integrative analysis, the derivatives of BCAAs and phenylalanine might be a potential link between the gut microbiota and T2D, which provide a new insight into the metabolic imbalance of (pre)diabetes. However, future studies should test whether these potential biomarkers can be used for the early identification of individuals that are at risk of developing T2D.

As discussed in the section 1.5, ML approaches have been widely used for phenotypic predictions, identifications of potential biomarkers as well as data integration in the gut microbiota studies. In the section 2.2, this thesis mainly presents how ML was applied to predict postprandial glycemic responses to a diet based on multi-omics data.

2.2 Prediction of postprandial glucose response based on omics data

To investigate possible driving factors for postprandial glucose regulation in **Paper II** as introduced in section 2.1, this study predicted glucose tAUC based on multi-omics data using ridge regression models with five-fold cross-validation (**Figure 8a**). The models trained with metabolomics data (especially after 2h MMT) performed best with minimum

root mean square error (RMSE). The correlation coefficients between the predictive and actual glucose tAUC were 0.92 and 0.9 when using metabolomic profiles at fasting (n=106) and two hour post MMT (n=95) as the training sets, respectively (**Figure 8b**). Using species and KOs profiles of the gut microbiota as training sets (n=106), correlation coefficients between the predictive and actual glucose tAUC were 0.69 and 0.72, respectively (**Figure 8c**). Overall, the predictive accuracy was improved when the model was trained with metabolomic profiles compared to using other omics data.

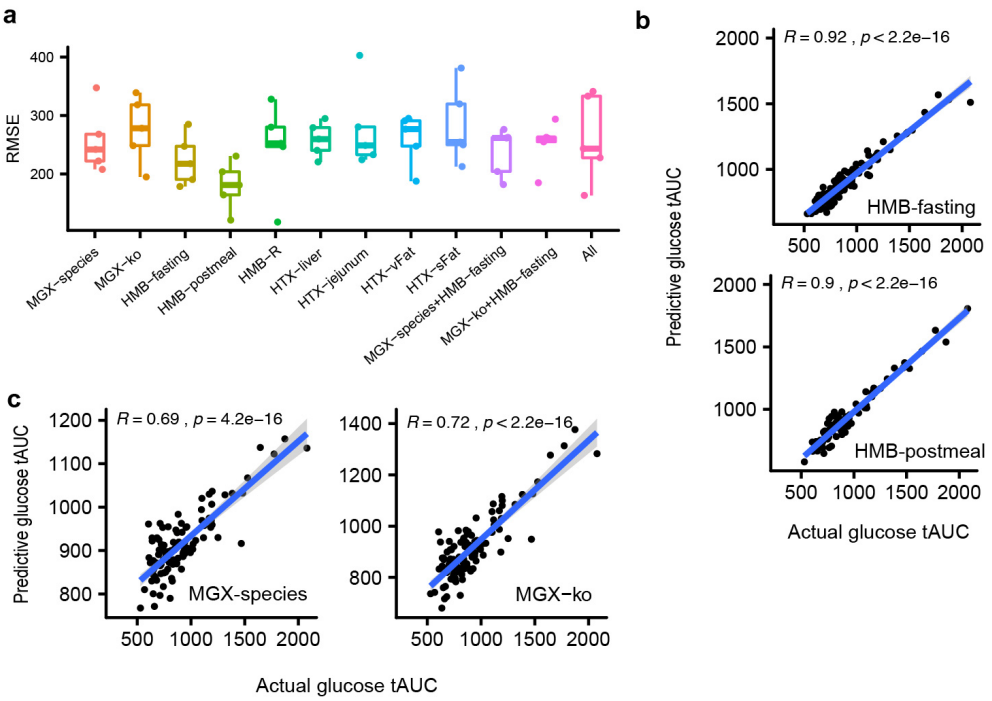

Figure 8 Predicting glucose response to a MMT by using ridge regression models based on multi-omics data. a). The performances of the ridge regression models evaluated by five-fold cross-validation and root mean square error (RMSE). b) The significant correlations between the actual glucose tAUC and the predicted glucose tAUC by ridge regression models using metabolomics profiles at fasting and two hour post meal; c) using profiles of gut microbial species and KOs. Spearman's rank correlation analysis was performed. MGX-species, microbiota composition at species level; MGX-ko, microbiota KO function profile; HMB-fasting, metabolomic profile at fasting; HMB-postmeal, metabolomic profile after 2h MMT; HMB-R, the ratios of metabolite abundance at 2h post MMT to fasting; HTX-liver, HTX-jejunum, HTX-mFat, HTX-sFat indicate human transcriptional profiles from liver, jejunum, mesenteric and subcutaneous adipose tissues, respectively; All, the integration of all multi-omics data.

In addition to the prediction of postprandial glucose response, this work could identify key signatures potentially contributing to the glycemic control based on the trained models. The important features were evaluated and ranked by the metric of regression coefficients. At fasting, glutamine, creatinine, pseudouridine, arginine, alanine, mannose, phenylalanine and lysine were identified to be the most important metabolites for prediction of glucose tAUC (check **Paper II** for detail). After two-hour MMT, mannose, allantoin, phenylalanine, 1−carboxyethylphenylalanine and NAA were predicted to be the most important metabolites (**Figure 9a**). Therefore, phenylalanine and its derived metabolites were identified important for glycemic control at both time points, which is consistent with the results from the metabolomics analysis in the section 2.1.2 (**Figure 6**).

In addition, several *Clostridium* species, such as *Clostridium sp. D5, Clostridium sp. SS2* and *Clostridium bartlettii* were identified to be correlated with glucose tAUC (**Figure 9b**), which is in line with the differential species identified by the metagenomics analysis in the section 2.1.3 (**Figure 7b**). Consequently, these results confirmed that the trained ML models enabled us to identify key signatures of both the gut microbiota and the host metabolism associated with glycemic response to a MMT.

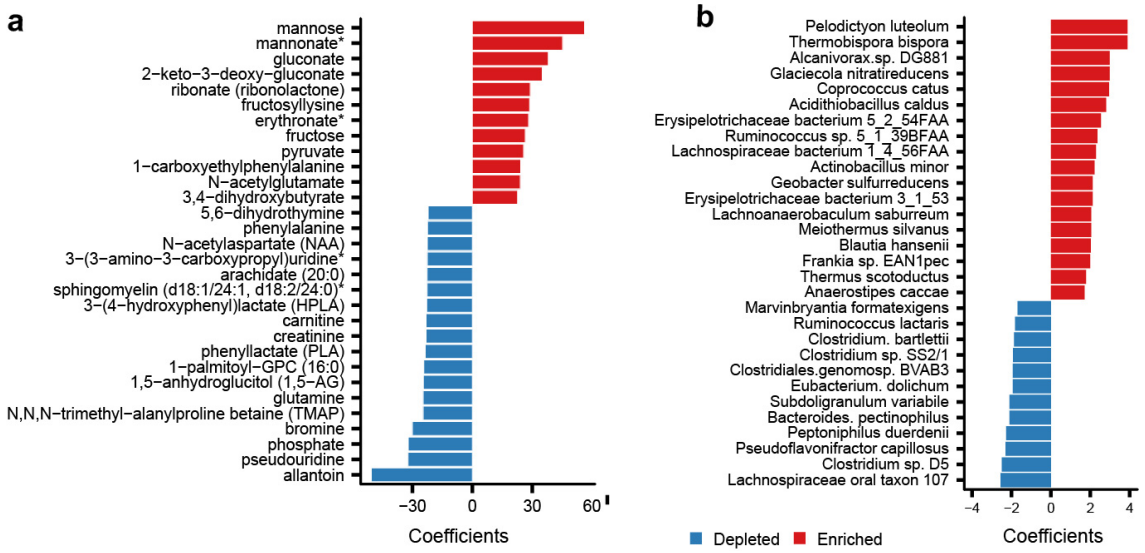

Figure 9 The important features for prediction of postprandial glucose responses. a) The regression coefficients of the top 30 metabolites based on post meal metabolomics data; b) the top 30 species for predicting glucose tAUC.

In summary, the section 2.2 mainly presents the work from **Paper II**, where regression ML models were trained for prediction of the postprandial glucose responses to a meal. These results showed that blood metabolomics-based models had better performance in comparison to other omics data. Also, the microbiota-based models showed an adequate predictive accuracy. In addition, these interpretable models had the potential to identify both the important serum metabolic and gut microbial features that might contribute to the abnormal glucose control in individuals with (pre)diabetes.

2.3 Systematical investigation of T2D-related gut microbial signatures using ML and GEMs

T2D is one multi-factor disease and has an intricate interaction of human genetics, the gut microbiota and other factors, such as ethnicity, geography, age, body mass index (BMI), diet and drugs. Due to the complexity, inconsistent findings have been reported across various gut microbiota studies related to T2D. Therefore, different types of factors should be taken into account when studying the gut microbiota related to T2D. Accumulated evidence has shown that ML holds great promise to explore and integrate diverse types of data [125-127]. Especially, the decision trees-based ensemble learning methods have been widely applied in the gut microbiota studies, such as random forest (RF) [49, 107, 108],

light gradient boosting machine (LightGBM) [109, 110] and extreme gradient boosting decision trees (XGBoost) [111, 112, 128], due to their powerful performance, ease of use and model explanation. In addition, GEMs have been widely applied to investigate the gut microbial species, e.g., *Lactobacillus reuteri*. Particularly, GEMs provide a systematical platform to investigate the community-level metabolic potentials of the gut microbiota and the interrelationships between the microbes [100]. With this background, **Paper III** performed a systematical analysis of four published metagenomic studies to identify T2D-related microbial signatures, using ML and GEMs. These four fecal metagenomic datasets and metadata were collected from the previous gut microbiota studies related to the Pre-D and T2D [48-51], which was summarized in **Table 2**. In total, 1779 individuals in these four cohorts were classified into three groups with different glycemic status, including the NGT (n = 848), Pre-D (n = 571) and T2D (n = 360) groups, according to the available metadata and disease labels in each original study.

Table 2. Characteristics of the four cohorts included in this study.

Study	NGT	Pre-D	T2D	Total	Age	BMI	Gender (female/male)	Country
Qin et al., 2012 [48]	185	-	182	367	48.0±14.4	23.4±3.4	157/210	China
Karlsson et al., 2013 [49]	43	49	53	145	70.4±0.7	27.1±4.6	145/0	Sweden
Zhong et al., 2019 [50]	97	80	79	256	62.3±9.3	24.8±3.2	149/107	China
Wu et al., 2020 [51]	523	442	46	1011	58.4±4.4	27.7±4.3	568/443	Sweden
Sum	848	571	360	1779	57.8±10.1	26.4±4.4	1019/760	

2.3.1 The overall composition of the gut microbiota associated with T2D

First the metagenomic data was consistently processed using a standardized bioinformatics pipeline, where the MetaPhlAn3 [70] and NG-meta-profiler [129] were used to obtain the taxonomic and functional profiles of the gut microbiota, respectively. To investigate the overall difference in the gut microbial composition among the NGT, Pre-D and T2D groups, the alpha-diversity at the species level using the pooled data from the four included studies (n =1779) was compared. Multiple alpha-diversity indexes showed significant differences among the three groups (P < 1e-10 by the Kruskal–Wallis test; **Figure 10a**). Moreover, the beta-diversity was investigated by using principal coordinate analysis (PCoA) with Bray–Curtis distances at the species level after pooling the data (**Figure 10b**). Consistently, the PERMANOVA result revealed a significant difference in the overall composition among the three groups (P = 0.001).

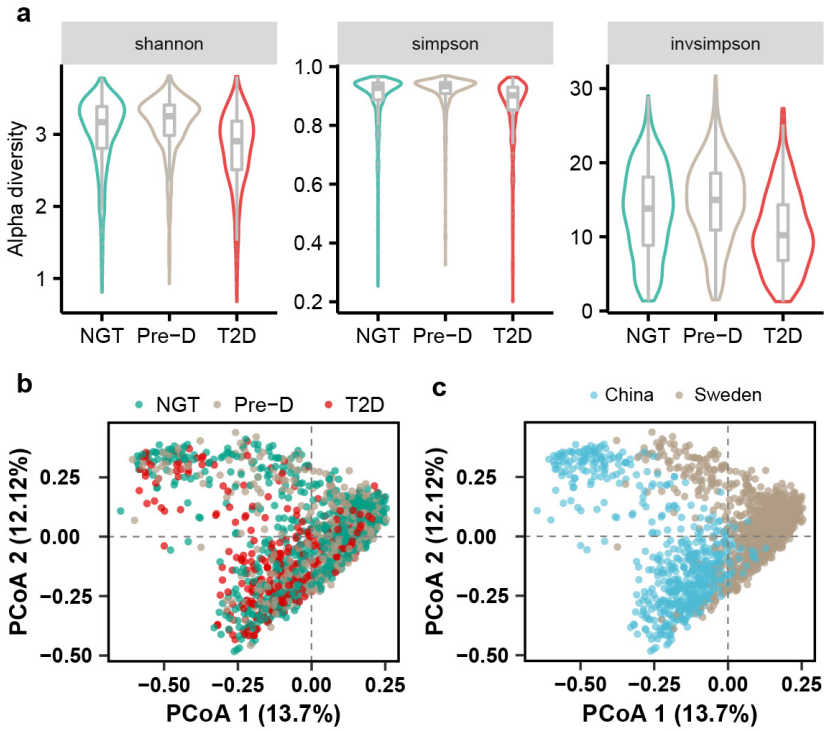

Figure 10 Alteration in the overall composition of the gut microbiota in the Pre-D and T2D groups. a) The alpha-diversity of the gut microbiota using Shannon, Simpson and Invsimpson indexes based on the species profiles after pooling the data from the four studies and grouped by different glycemic status. b) Principal coordinates analysis (PCoA) based on Bray–Curtis distances at the species level using the pooled data. c) PCoA showing difference in the compositional profiles of the gut microbiota between Chinese and Swedish.

Furthermore, the PCoA result showed a significant difference in the beta-diversity between the Chinese and Swedish cohorts (*P* values = 0.001 evaluated by PERMANOVA; **Figure 10c**). This suggests that the ethnicity or geography could have considerable impact on the gut microbiota, which has been reported in previous studies [9, 10, 18]. Also, it hints that the confounding factors, such as ethnicity (or geography), gender, age and BMI, need to be taken into account when performing data analysis of the T2D-related gut microbiota.

2.3.2 The functional capabilities of gut microbiota simulated by community-level metabolic models

In addition to studying the detailed metabolism of an organism as introduced in section 1.4, GEMs could serve as a useful tool for investigating the metabolic potential of the entire gut microbial species (**Paper III**), which could help understand the complex interactions between the microbes and human metabolism. Thus, for modeling the human gut microbiota using GEMs, a semi-automatic pipeline of constructing the microbial species GEMs was firstly developed (**Figure 11**). In the pipeline, the representative strain for each species was selected according to the strain and genome information in the NCBI database, after obtaining the taxonomic profiles from metagenomic sequence data processed by the MetaPhlAn3 tool. With the representative strains, the corresponding reference genome sequences from the NCBI database were collected for GEMs

construction. Based on the MetaCyc database, the corresponding genes, enzymes and reactions were integrated into the draft models of species. Finally, biomass, exchange and transport reactions were added into the draft models according to the Gram staining information, transporter annotations and medium composition. Using the pipeline, GEMs of 827 individual species were constructed, including 456 Gram-positive species and 331 Gram-negative species. After adding gap-filling reactions, all species GEMs were able to simulate growth under the dGMM+LAB medium, which is a mixture of the defined gut microbiota medium (GMM) and the LAB medium supporting growth of lactic acid bacteria [130].

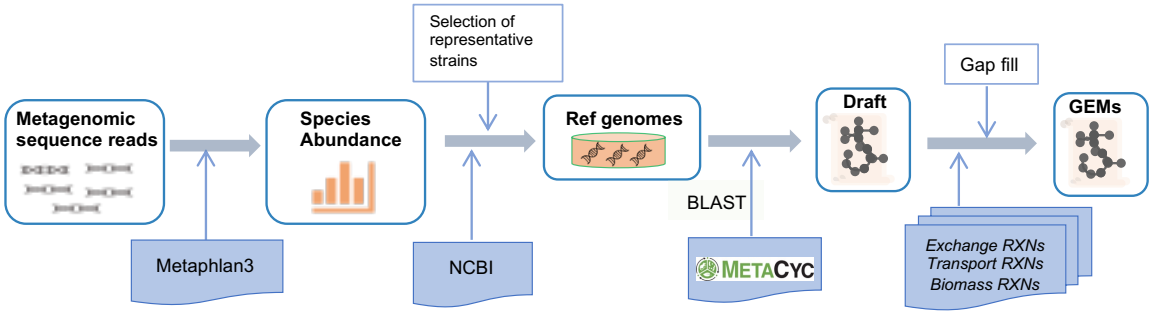

Figure 11 The semi-automatic pipeline for constructing GEMs of the gut microbial species. Firstly, the taxonomic profiles from metagenomic sequence data were extracted by the MetaPhlAn3 tool. Secondly, the representative strain for each species was selected according to information from the NCBI database. Thirdly, the corresponding reference genome sequences were collected from the NCBI database. Then, the corresponding genes, enzymes and reactions were integrated into the draft models based on the MetaCyc database. Finally, GEMs were constructed after adding biomass, exchange, transport and gap-filling reactions.

To simulate the entire metabolic capabilities of the gut microbiota in individuals with different diabetic status, the community-level metabolic models were further constructed, which considered the GEM of each microbial species as one component of the whole metabolic model (**Figure 12**). In the framework, all GEMs of the individual species in the gut microbial community were integrated into a much larger metabolic model by creating different compartments to separate intercellular metabolites from different species but allowing extracellular metabolites to be transported between species. In addition, the overall biomass reaction of the community-level metabolic model was established as the weighted combination of the biomasses of all species GEMs, where the species abundance was used as coefficients for each species biomass (**Figure 12**). Using the community-level metabolic model, the collective reaction fluxes of all species within one individual gut microbiota can be simulated under maximizing the biomass growth or optimizing the production of one microbial metabolite of interest.

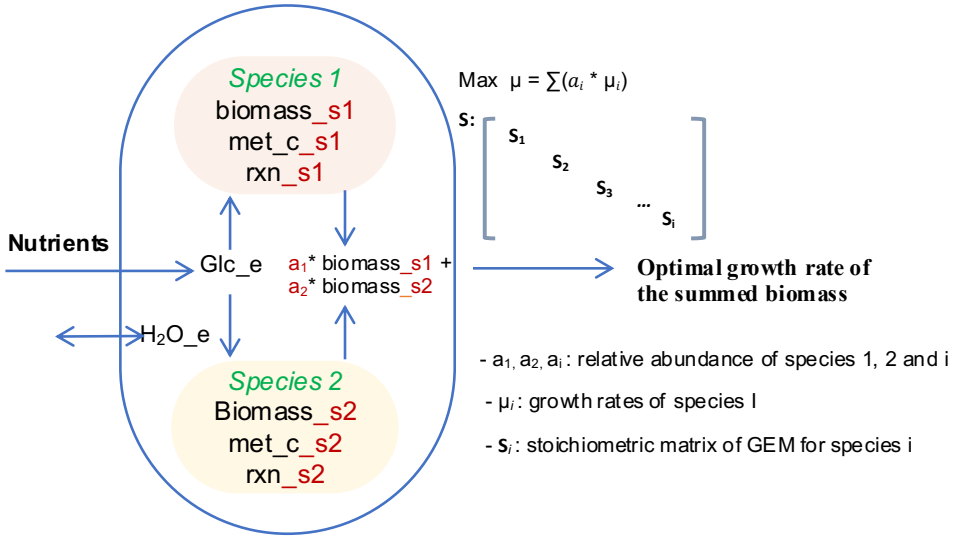

Figure 12 The framework of constructing community-level metabolic models for modeling the individual gut microbiota. The framework integrates GEMs of all microbial species within one individual gut microbiota into a large metabolic model by creating different compartments to separate intercellular metabolites from different species (e.g., the suffix '_s1' and '_s2' representing species 1 and 2) but allowing extracellular metabolites to be exchanged between species. Additionally, the overall biomass μ was set as the weighted combination of the biomasses μ_i of all species GEMs. The relative abundances of species a_i were used as coefficients for each species biomass. For the simulation, the community-level metabolic network is defined as a stoichiometric coefficient matrix ***S,*** which combines the stoichiometric coefficient matrixes $\boldsymbol{S_i}$ of all species. FBA is usually used to simulate metabolic fluxes at a steady state when maximizing an objective function under given condition.

Increasing studies have revealed the effects of the gut microbiota on the T2D probably due to the microbial-derived metabolites [52], such as SCFAs [131], BCAAs [53], imidazole propionate [54]. Therefore, the potential production capabilities of several representative metabolites, such as SCFAs, BCAAs were evaluated by optimizing the corresponding biosynthesis reaction for the individual community-level metabolic model. The simulated microbial metabolic fluxes might provide us novel insights into the abnormal metabolisms or interrelations in the T2D-related gut microbial species.

2.3.3 Microbiota-based machine learning models for prediction of T2D status

Next, this work explored whether these obtained microbial features could discriminate the T2D patients from the NGT individuals. To this end, different microbiota-based classification models that could predict T2D status were devised, using various gut microbial signatures including the taxonomic and functional profiles from the gut metagenomics data as well as the metabolic fluxes of the gut microbiota simulated by its community-level metabolic model as presented in section 2.3.2.

Using three decision trees-based ensemble learning methods including the random forest, LightGBM and XGBoost, this study first devised various microbiota-based prediction models, based on these obtained microbial features (the profiles of species, KOs, metabolic fluxes) and their combinations with ethnicity (or geography), gender, age and BMI (check **Paper III** for details).

For the pooled data from the four included studies, all samples (n=1779) were split into two parts including 70% training dataset and 30% testing dataset (**Figure 13**). Using the training datasets with various types of features, the predictive models for discriminating the NGT from the T2D were first trained and evaluated by five-fold cross-validation, and then applied to predict the T2D status of a new sample in the testing dataset.

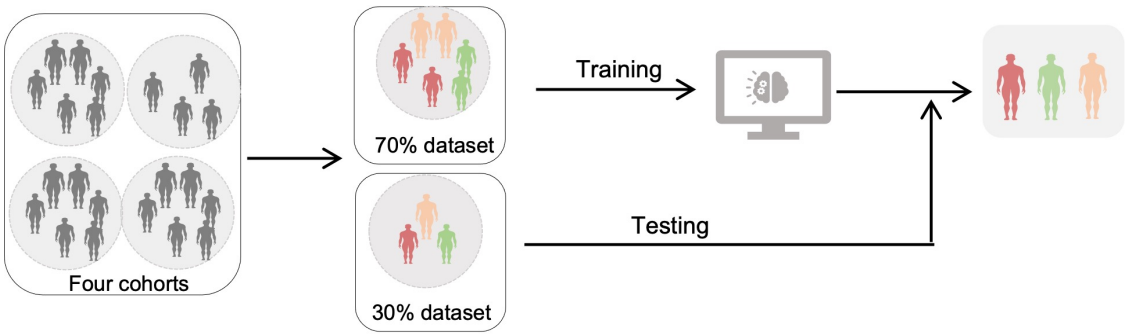

Figure 13 The pipeline of training predictive models for discriminating individuals with T2D from NGT. The pooled data from the four included studies was split into two parts including 70% training dataset and 30% testing dataset.

The accuracy of the predictive models trained by using LightGBM and XGBoost showed an increased trend with an area under the ROC curve (AUC) of ~ 0.7 compared to using the random forest (**Figure 14**). Moreover, compared to the classifiers without adjustment by covariates, the predictive models with the adjustment of variables ethnicity, gender, age and BMI showed a better classification performance for discriminating the NGT from the T2D in most cases. It is strongly suggestive that the confounding factors should be taken into consideration when studying the gut microbiota related to the diabetic disease. In addition, the classification models based on the pooled data for prediction of T2D status did not achieve a better performance than the models trained with data from one individual study. This might owe to differences in other factors across studies, such as the experimental design, medication, diet, which should be further included in the predictive model (check details in **Paper III**).

Furthermore, the predictive models of the NGT versus Pre-D or of the Pre-D versus T2D showed a poor performance (AUC = ~ 0.5) using any type of data from single study, which implies that there might be few discriminative signatures included in the predictive models between the NGT and Pre-D groups or between the Pre-D and T2D groups through the used pipeline in this study. Also, this suggests that it is still challenging to differentiate individuals with prediabetes from the healthy people at early stage by utilizing the gut microbiota.

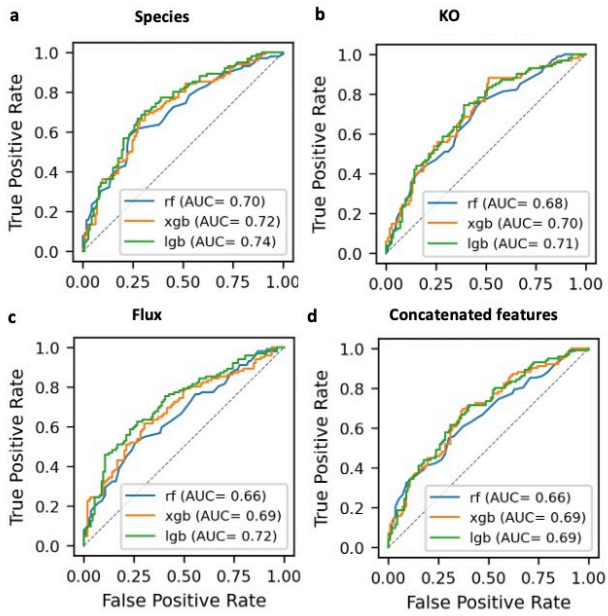

Figure 14. The classification models for prediction of T2D status based on the gut microbial features. The performance of the predictive models validated on the testing dataset when pooling all data from the four included studies a) using species abundances; b) KO profiles; c) the simulated reaction fluxes of the gut microbial community; d) the concatenated features of species, KOs and fluxes. The models were trained using three decision tree-based ML methods, including the LightGBM, XGBoost and random forest (RF) as well as were adjusted by covariates age, BMI, ethnicity and gender.

2.3.4 Consistent T2D-related microbial features identified by classifiers of the NGT versus T2D

Based on the trained models for discriminating the NGT from the T2D, the important microbial features in the classification were evaluated and ranked by the metric of mean decrease in Gini impurity. The age and BMI were identified to be two most important factors for the NGT versus T2D classification, when using the pooled species abundances and random forest (**Figure 15**). This finding is in accordance with that age and BMI could be significantly correlated with T2D and confound the relationships between the gut microbiota and T2D [19, 132]. Previous studies have demonstrated that genera *Faecalibacterium*, *Roseburia* and *Bacteroides* were negatively correlated with the T2D, whereas the genus *Ruminococcus* was positively correlated with the T2D [133]. In line with this, *Faecalibacterium prausnitzii*, three *Roseburia* species (*Roseburia intestinalis, Roseburia hominis*, *Roseburia inulinivorans*), three *Bacteroides* species (*Bacteroides uniformis, Bacteroides caccae* and *Bacteroides vulgatus*) and two *Ruminococcus* species (*Ruminococcus bromii Ruminococcus lactaris*) were identified to be important for prediction of T2D status (**Figure 15a**). Particularly, out of them, *Faecalibacterium prausnitzii* and three *Roseburia* species have been suggested to be butyrate-producing microbial species that have a beneficial effect on the T2D [131, 133].

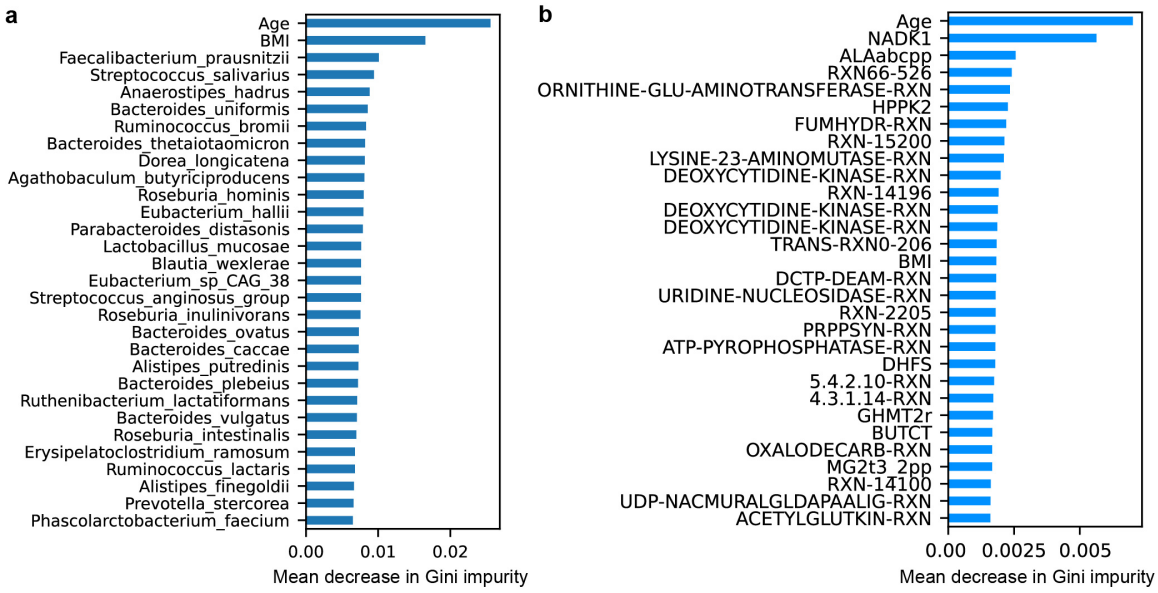

Figure 15 The important microbial features that could differentiate the NGT from the T2D. The top 30 important a) microbial species and b) reaction fluxes were identified by training random forest model based on the pooled data with adjustment by covariates age, BMI, ethnicity and gender.

In agreement, the variable age was evaluated to be the most important factor, when using the pooled reaction fluxes, age, BMI, gender and ethnicity to train model for predicting T2D risk. Moreover, reactions NADK1 (NAD kinase GTP catalyzed by NAD kinase 1), ALAabcpp (L-alanine transport via ABC system), RXN-15200 (involved in L-phenylalanine biosynthesis III pathway), two reactions LYSINE-23-AMINOMUTASE-RXN and 4.3.1.14-RXN (involved in L-lysine fermentation to acetate and butyrate), and BUTCT (catalyzed by the acetyl-CoA: butyrate-CoA transferase), GHMT2r (catalyzed by glycine hydroxymethyltransferase) were identified to be important for prediction of T2D status (**Figure 15b**). Among them, three reactions BUTCT, LYSINE-23-AMINOMUTASE-RXN and 4.3.1.14-RXN were involved in the butyrate synthesis pathway, which is in line with the above species results. Through the metabolic simulations of the gut microbiota, this work identified several important reactions, which are involved in the butyrate biosynthesis pathway and important for discriminating the NGT from the T2D. Thus, ML in combination with the community-level metabolic models have the potential to enable identification of the novel T2D-related gut microbial signatures.

2.4 Limited performance of the microbiota-based classifiers on an independent cohort

By using ML integrated with community-level metabolic models, the obtained predictive models could not only discriminate the T2D individuals from the NGT (**Figure 14**), but also identify a number of consistent T2D-related microbial features, including the SCFAs-producing microbial species and reactions (**Figure 15**). However, this work also revealed dramatical heterogeneities across the four included studies. This proposes a question about

whether the trained models using all microbial features could be accurately predictive of T2D risk on an independent cohort. Thus, out of the four studies, this work further used data from three studies as the training dataset while used data from the remaining one study as an independent testing dataset every time (**Figure 16**). This process was iteratively preformed for each individual study as an independent cohort. Notably, in most cases, the predictive models of the NGT versus T2D showed a limited performance on each independent cohort (AUC = 0.5- 0.6) using whatever type of microbial features (check details in **Paper III**). When using the dataset of species abundance from the Qin et al. or Karlsson et al., the XGBoost model reached a moderate classification performance (AUC = 0.66 and 0.64, respectively). These results imply that the microbiota-based models for predicting the T2D risk might be specific to certain types of cohorts rather than generalizable across studies, which is in accordance with the previous study [18].

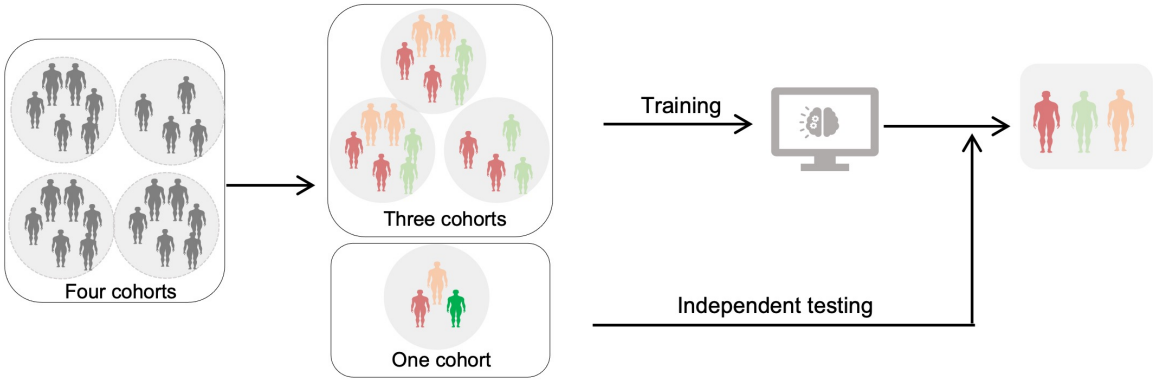

Figure 16 The pipeline for training predictive models and validation in an independent testing dataset. Out of the four datasets, three datasets were used as the training dataset while the remaining one dataset used as an independent testing dataset.

To sum up, the section 2.3 and 2.4 mainly present the work from **Paper III** where a systematical analysis of four published gut metagenomic data were performed, using ML approaches and community-level metabolic models. The microbiota-based prediction models not only showed an adequate accuracy, but also enabled us to identify important T2D-related microbial features. A number of SCFAs-producing microbial species and related metabolic reactions have been consistently identified to be important for discriminating the NGT from the T2D. This finding emphasized that alteration in the SCFAs-producing capability of the gut microbiota could play a critical role in the pathology of T2D progression. However, these results also suggest that the microbiota-based models for predicting T2D status might be specific to a population due to the heterogeneities between different cohorts. In addition, this study has proved that investigating the metabolic capabilities of the microbiota by using the metabolic models could help to interrogate the associations between T2D and the gut microbiota via targeting the key reaction fluxes and genes to specific species. This work indicates that ML in combination with GEMs has the potential to identify new microbial metabolic signatures related to T2D.

3. The effect of *Lactobacillus reuteri* ATCC PTA 6475 on human metabolism

As introduced in the background part, previous studies have suggested the positive effects of oral supplementation with probiotics on human health. This chapter mainly summarizes three studies (**Paper IV-VI**) on the complex interactions between the probiotic *Lactobacillus reuteri* ATCC PTA 6475, the gut microbiota and host metabolism in older women with low bone mineral density (BMD). The first part (chapter 3.1) introduces the detailed metabolism of *L. reuteri* ATCC PTA 6475 via the GEM reconstruction of the single probiotic strain (**Paper IV**). The second part (chapter 3.2) discusses the impact of *L. reuteri* ATCC PTA 6475 intake on the global metabolic profiles of order women with bone loss (**Paper V**). The third part (chapter 3.3) presents alterations in the gut microbiota of order women with good or poor responses to orally administered *L. reuteri* ATCC PTA 6475 (**Paper VI**).

3.1 Studying the metabolism of *L. reuteri* ATCC PTA 6475 using GEM

As mentioned in section 1.4, GEMs have served as a useful tool to elucidate the metabolism of an organism. Thus, this work investigated the metabolic capacities of *L. reuteri* ATCC PTA 6475 via reconstruction of its GEM [96]. In this study, the probiotic GEM was first reconstructed using a template-based modeling pipeline and then curated manually as shown in **Figure 17**.

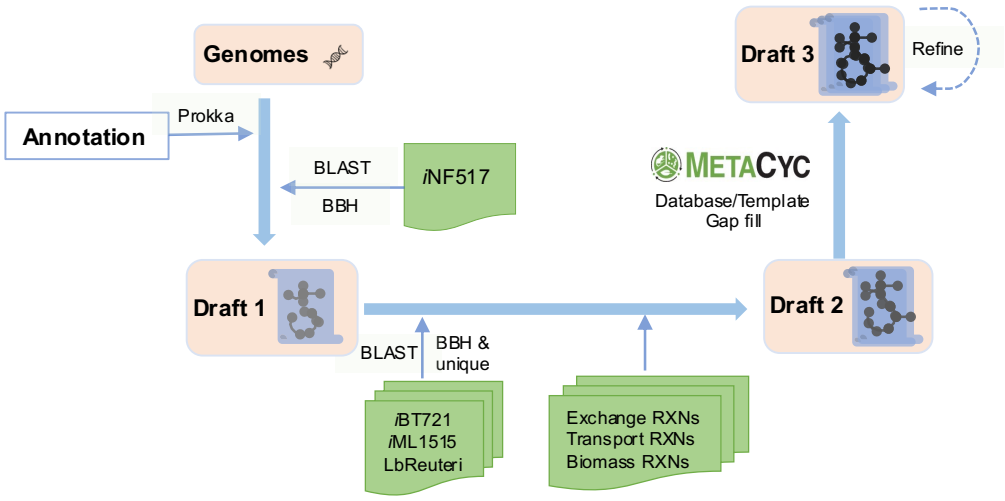

Figure 17 GEM reconstruction of *L. reuteri* ATCC PTA 6475 using a template-based pipeline. The iNF517 model was first used as a template model. The ortholog genes and reactions were extracted using the bidirectional best hits (BBH) to generate the initial draft model. After comparing to the GEMs iBT721, iML1515 and LbReuteri, the exchange and transport reactions from the template models were added according to the transporter annotations and corresponding medium composition. Then the gap-filling was performed against the template models as well as using the MetaCyc database to improve the model performance. Finally, the draft model had been manually curated during the simulation.

In the pipeline, the GEM iNF517 of *Lactobacillus casei* MG1363 was used as the main template to reconstruct the initial draft model. Then the metabolic genes and related reactions were integrated into the initial GEM in comparisons to the other three template models (more details in **Paper IV**). The exchange reactions, transport reactions and gap-filling reactions were further added. After refining the draft model, the finalized GEM of *L. reuteri* ATCC PTA 6475 includes 726 metabolites and 894 reactions mapped to 622 metabolic genes, which further was used to simulate biomass growth and microbial metabolite biosynthesis.

Previous studies have suggested that *L. reuteri* strains have capabilities to produce a number of health-related metabolites, such as acetate, lactate, reuterin (3-hydroxypropionaldehyde), histamine, vitamin B12 (cobalamin) and vitamin B9 (folate). Thus, the probiotic strain *L. reuteri* ATCC PTA 6475 might impact human metabolism by the secretion of the beneficial microbial SCFAs metabolites including acetate and lactate. As an example, the main biosynthesis pathways of microbial metabolites acetate and lactate were simulated using the GEM as illustrated in **Figure 18**. The carbohydrate metabolism mainly uses the phosphoketolase pathway (PKP) to produce lactate and acetate. In the dietary fermentation, lactate is usually the most important end-product fermented by *Lactobacillus* and acetate and ethanol are main by-products.

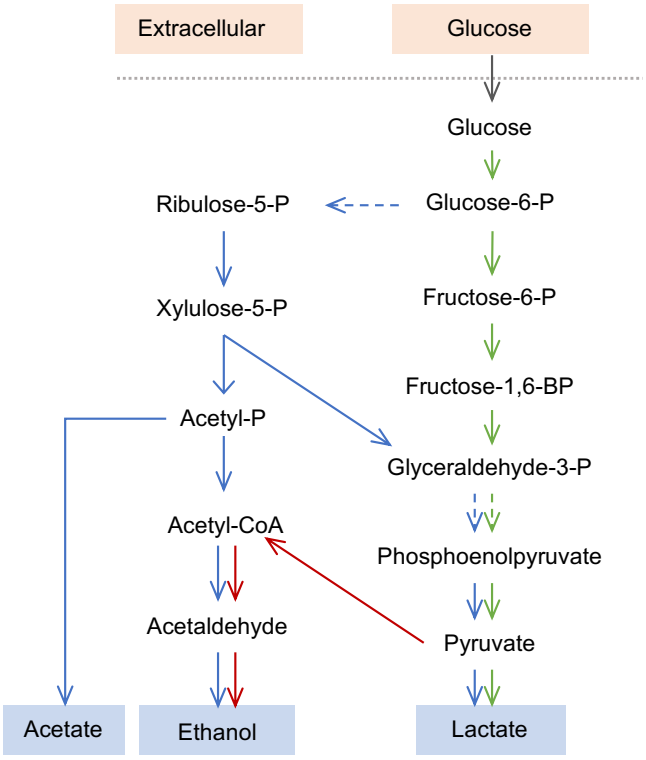

Figure 18 The biosynthesis pathways of the microbial metabolites acetate and lactate simulated by the GEM of *L. reuteri* ATCC PTA 6475. Blue arrows indicate the phosphoketolase pathway (PKP); green arrows indicate Emden-Meyerhof-Parnas pathway (EMP); red arrows indicate the extensions of EMP; and the dotted arrows indicate multiple enzymatic reactions; orange background indicates the extracellular metabolites and blue background indicates the metabolites *L. reuteri* can produce.

In general, the section 3.1 presented the results from **Paper IV** where the metabolic capabilities of the probiotic *L. reuteri* ATCC PTA 6475 were explored by using GEM. Through the model simulation, the specific biosynthesis pathways of several metabolites with potential benefits on human metabolism, including SCFAs, were investigated in detail. This suggests that the positive effect of *L. reuteri* ATCC PTA 6475 on the host metabolism is possibly mediated by the production of the microbe-derived metabolites. Therefore, the GEM could provide a reliable scaffold for studying the metabolism of the probiotic *L. reuteri* ATCC PTA 6475, which could help to understand the underlying mechanisms of its beneficial effects on older women with bone loss.

3.2 The impact of *L. reuteri* ATCC PTA 6475 on the metabolic profiles of older women

As a lactic acid bacterium, *L. reuteri* strains have been widely used as probiotics as well as applied in different food products and supplements. Oral administration with *L. reuteri* strains could have the positive effects on human health, such as reducing bone loss in the elderly and promoting immune system development. Especially, *L. reuteri* ATCC PTA 6475 has been successfully developed as a probiotic product in the market. As shown in **Figure 19**, our previous randomized controlled trial (RCT) demonstrated that supplementation of *L. reuteri* ATCC PTA 6475 led to substantially reduced bone loss in older women with BMD [68].

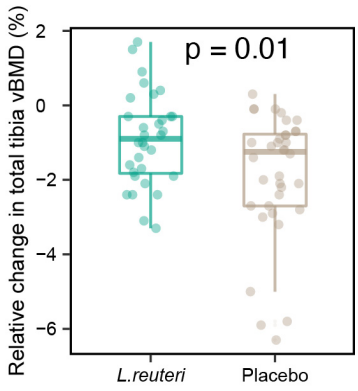

Figure 19 Relative change of total tibia volumetric BMD (vBMD) after one-year supplementation with *L. reuteri* ATCC PTA 6475 or placebo. 32 older women had supplemented with *L. reuteri* ATCC PTA 6475 (*L. reuteri* group) and 36 older women had administrated with placebo (Placebo group).

However, the mechanism underlying the beneficial effects of the probiotics on bone metabolism in the elderly is still unclear. By constructing the GEM of *L. reuteri* ATCC PTA 6475, this thesis first investigated its metabolisms as introduced in chapter 3.1 (**Paper IV**). The results suggested that the beneficial effects of the probiotic on human metabolism might be due to its biosynthesis of the beneficial metabolites, e.g., SCFAs (**Figure 18**). Therefore, the following section introduces how the probiotic *L. reuteri* ATCC PTA 6475 influences the global metabolism of older women by using the untargeted metabolomics profiling (**Paper V**).

3.2.1 The dynamic changes of metabolomic profiles during one-year probiotic intake

During one-year follow-up at four timepoints, serum samples from 32 subjects with *L. reuteri* ATCC PTA 6475 intake (*L. reuteri* group) and 36 subjects administrated with placebo (placebo group) were collected (**Figure 20a**). Then time-series metabolomic profiles of elderly women with low BMD were analyzed to investigate the metabolic changes after the probiotic supplementation. To examine the overall difference between the *L. reuteri* and placebo groups, the Euclidean distances based on metabolomic profiles between the two groups at four timepoints were calculated and illustrated in **Figure 20b**. It was clear that *L. reuteri* and placebo groups clustered together at baseline, indicating similarity of the baseline metabolism of older women in the two groups. Nevertheless, metabolic profiles showed differences between the *L. reuteri* and placebo groups in the following-up period, hinting the effects of the probiotic intake on the host metabolism.

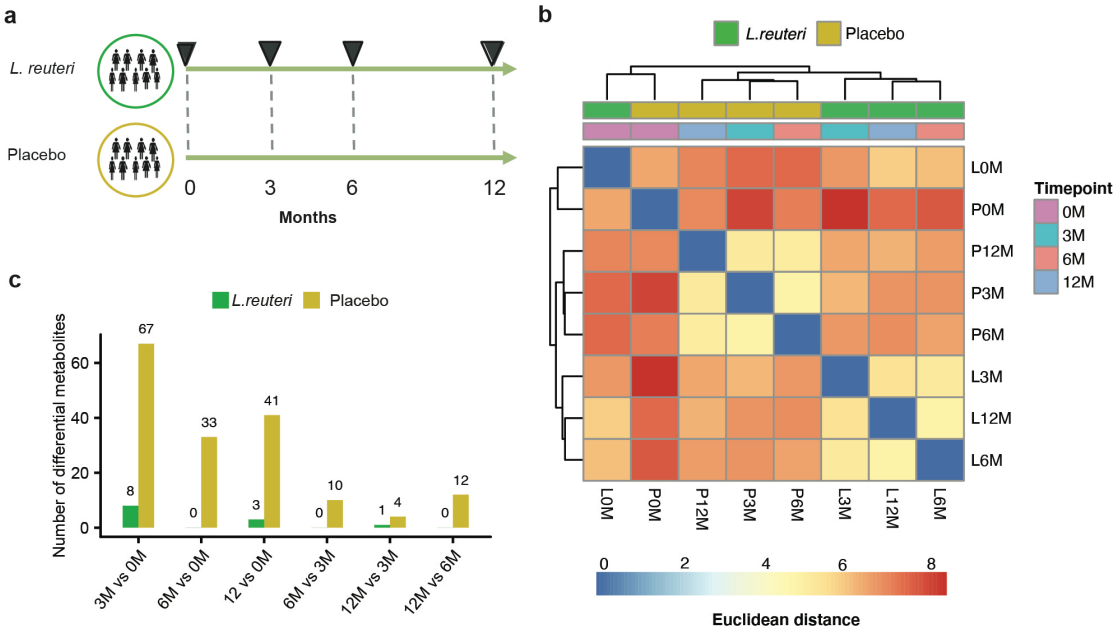

Figure 20 Alterations of the metabolomic profiles in older women supplemented with placebo or *L. reuteri* PTA 6475. a) The experimental design of metabolomics profiling. Serum samples were collected from older women with low BMD at baseline, 3, 6, and 12 months. b) The heatmap showing the hierarchical clustering of Euclidean distances between serum samples. c) Numbers of significantly differential metabolites between time points in the *L. reuteri* group or placebo group (adjusted $P < 0.1$).

Further, the differential metabolites between any two timepoints in the *L. reuteri* or placebo group were identified respectively (**Figure 20c**; adjusted $P < 0.1$ by the Wilcoxon signed-rank test). During one-year probiotic intervention, 67 (7 up-regulated and 60 down-regulated), 33 (2 up-regulated and 31 down-regulated), and 41 (10 up-regulated and 31 down-regulated) metabolites showed significantly differential at 3, 6 and 12 months, compared to the baseline in the placebo group, while a few metabolites were identified to be differential in the *L. reuteri* group. Thus, much less metabolic variations were observed in the *L. reuteri* group than the placebo group. In other words, older women in the placebo group underwent the significant alterations of the metabolic profiles, which might be associated with bone loss. Interestingly, the adverse alterations, including the increased

bone loss and metabolic changes in the placebo group, were prevented to some extent by the oral administration of *L. reuteri* ATCC PTA 6475.

3.2.2 Differential metabolic responses relating to the probiotic effects on bone metabolism

Next the probiotics-specific responses during the one-year supplementation were investigated, which could help to understand the mechanistic effects of *L. reuteri* ATCC PTA 6475 on bone metabolism. The metabolite levels at follow-up timepoints were first calculated as ratios of the baseline values, which were referred to as metabolic response to the probiotic intervention. This work further identified metabolites that showed differences in relative changes from baseline between the *L. reuteri* and placebo groups by the Wilcoxon rank-sum test. There were 30, 54 and 31 metabolites that changed from baseline differentially between the *L. reuteri* and placebo groups at 3, 6 and 12 months, respectively (**Figure 21a**; VIP score > 1 and *P* value < 0.05). Particularly, two metabolites butyrylcarnitine (C4) and 1−methyl−4−imidazoleacetate responded differentially at all follow-up timepoints and showed a robust increase in the *L. reuteri* group (**Figure 21b**). Butyrylcarnitine (C4), a butyrate ester of carnitine, could act as the pool and transporter of butyrate [134], which was previously reported to inhibit bone resorption and stimulate bone formation in mice through a signaling pathway involving regulatory T-cells and Wnt10b [62, 64]. Moreover, butyrate supplementation was recently suggested to increase bone mass in wild type mice and to prevent ovariectomy induced bone loss [135]. Thus, the robust increase in this metabolite may indicate the involvement of butyrate signaling in the effects of *L. reuteri* ATCC PTA 6475 on the reduced bone loss as shown in **Figure 19**.

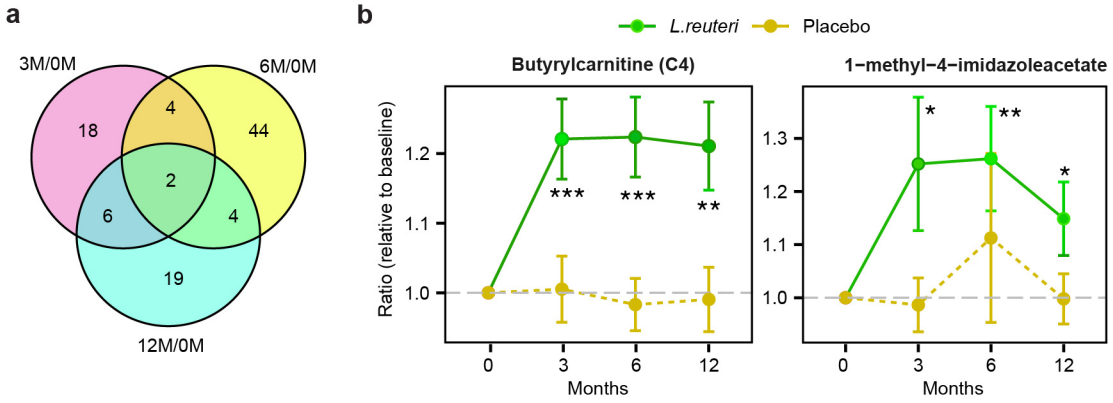

Figure 21 The differential metabolic responses to the supplementation with *L. reuteri* ATCC PTA 6475 or placebo. a) The metabolites differed in changing from baseline between the *L. reuteri* and placebo groups at 3, 6 and 12 months (VIP score > 1 and P value < 0.05). b) The relative change from baseline (Mean ± SE) of butyrylcarnitine (C4) and 1−methyl−4−imidazoleacetate that responded differentially between the *L. reuteri* and placebo groups at three timepoints. '*', $P < 0.05$; '**', $P < 0.01$; '***', $P < 0.001$.

In conclusion, the section 3.2 presents the results from **Paper V** where the effects of *L. reuteri* ATCC PTA 6475 on the global metabolism of older women with bone loss were explored by time-series metabolomic analysis. This work found that the alterations,

including the deteriorated bone loss and metabolic changes in the placebo group, were alleviated by one-year supplementation with *L. reuteri* ATCC PTA 6475. Interestingly, butyrylcarnitine (C4) level was increased at all follow-up timepoints in the *L. reuteri* group compared to the placebo group, indicating that the effects of *L. reuteri* ATCC PTA 6475 on bone metabolism might be mediated through the butyrate signaling. However, further studies are needed to identify the mechanisms and determine how gut microbiota changes are caused by supplementation with probiotic *L. reuteri* and how such changes are linked to human metabolomic dynamics.

3.3 The effects of *L. reuteri* ATCC PTA 6475 intake on the gut microbiota of the elderly

In section 3.2, the impact of *L. reuteri* ATCC PTA 6475 on the global metabolism of older women is discussed. However, it is still unknown whether the alterations in the gut microbiota of the elderly occurred due to the probiotic supplementation. To this end, the following section introduces the effects of *L. reuteri* ATCC PTA 6475 on the gut microbiota of older women with good or poor responses to the probiotic intake (**Paper VI**). In addition, the section presents whether the microbial alterations could be linked to the metabolomic changes observed in **Figure 20 and 21**.

As observed in a recent study [68], part of older women responded poorly to the oral supplementation with *L. reuteri* ATCC PTA 6475 (i.e., poor responders still had severe bone loss). To investigate the differential effects of the probiotic intake on the host, 20 elderly women by identifying 10 women with a good response (GR group) and 10 women with a poor response (PR group) were selected (**Figure 22**; more details in **Paper VI**). In addition, serum samples for metabolomic profiling and fecal samples for metagenomic sequencing were collected from the older women at baseline and 12 months.

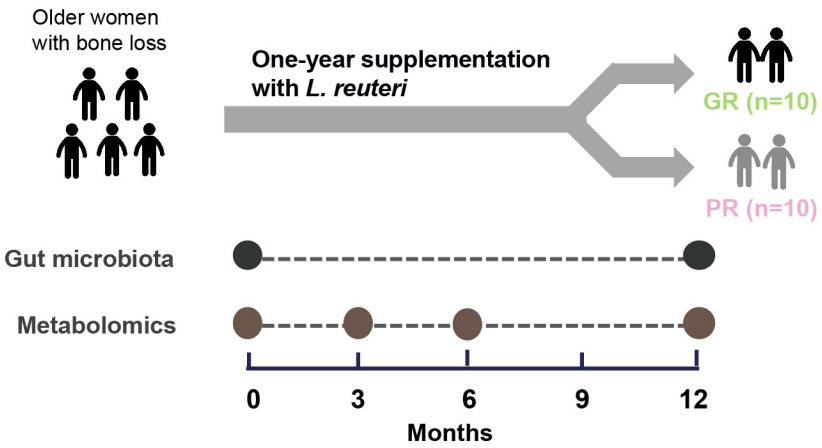

Figure 22 The experimental design for elderly women with differential responses to the supplementation with *L. reuteri* ATCC PTA 6475. Women with a good response (GR group, n=10) and with a poor response (PR group, n=10) were selected. Serum samples and fecal samples were collected from the older women at baseline and 12 months.

3.3.1 Probiotic intake reduces bone loss and decreases inflammation in the good responders

After one-year supplementation with *L. reuteri* ATCC PTA 6475, the relative change in tibia total volumetric BMD showed significantly increased in the GR group (0.39 ± 0.77) compared to the PR group (-2.22 ± 0.58; $P < 0.001$ by the t-test). In line with this, the significantly decreased level of tibia total volumetric BMD was only observed in the PR group at 12 months (**Table 3**; $P < 0.05$), indicating that the probiotic intake prevents bone loss in the GR group. Moreover, ultrasensitive c-reactive protein (usCRP) showed a significantly reduced level in the GR group at 12 months (**Table 3**; $P < 0.05$), which suggested that inflammation in the GR group was alleviated by one-year supplementation with the probiotics. In addition, older women in the GR group had higher BMI than the PR group at both baseline and 12 months ($P < 0.05$). The differences in BMI or weight might influence the baseline gut microbiota, which would further contribute to the differential responses to the probiotic intake in older women.

Table 3. Comparisons of characteristics in the GR and PR groups at baseline and 12 months.

Characteristics	Baseline		12 months		P^a	P^b
	GR (n = 10)	PR (n = 10)	GR (n = 10)	PR (n = 10)		
Total tibia vBMD (mg/cm^3)	247 ± 39.2	231 ± 44.9	248 ± 39.8	**226 ± 44.2** #	0.42	0.26
Weight (kg)	72.4 ± 8.2	63.9 ± 7.7	73.3 ± 8.5	64.0 ± 8.0	**0.03**	**0.02**
BMI (kg/m^2)	27.5 ± 3.6	24.0 ± 2.8	27.7 ± 3.7	24.1 ± 3.1	**0.03**	**0.03**
usCRP (mg/L)	2.14 (1.53-3.68)	0.98 (0.8-2.47)	**1.57 (1.13-1.90)** *	1.36 (0.67-3.19)	0.25	0.91
Total fat mass (kg)	28.9 (27.1-32.0)	20.8 (19.3-24.5)	28.2 (24.3-31.7)	20.0 (18.5-23.7)	**0.04**	**0.04**

Note: Mean ± SD. Non-normally distributed variables are presented as median with interquartile range. The t-test or Wilcoxon test were used as appropriate. '*' and '#' denote significant difference ($P < 0.05$) between baseline and 12 months in the GR and PR groups, respectively. P^a and P^b values are from comparisons between the GR and PR groups at baseline and 12 months, respectively. The significant differences ($P < 0.05$) are highlighted in bold. vBMD: volumetric bone mineral density; usCRP: ultrasensitive c-reactive protein; BMI: body mass index.

3.3.2 Alterations of the gut microbiota after one-year probiotic supplementation

By comparing the composition of the gut microbiota between the GR and PR groups, four species, including *Prevotella buccae*, *Clostridium acetobutylicum*, *Bacteroides sp.* 2_1_56FAA, *Acidaminococcus fermentans*, were identified to be differential at 12 months, while three species, including *Streptococcus australis*, *Lactobacillus antri*, *Lachnospiraceae bacterium* 4_1_37FAA, showed differential at baseline (**Figure 23**; $P < 0.01$ by the Wilcoxon rank-sum test). The differences in the endogenous baseline microbiota might be important for a good response to the probiotic intake. In other words, the differential baseline signatures have the potential to discriminate the good responders from the poor responders. *Lactobacillus antri* was less abundant in the PR group than the GR group, while *Lachnospiraceae bacterium* 4_1_37FAA was more abundant in the PR groups at baseline ($P < 0.01$). Interestingly, 11 species including *Escherichia coli* showed differential between the two timepoints in the PR group while only one differential species (i.e., *L. reuteri* due to the probiotic intake) in the GR group ($P < 0.01$ by the Wilcoxon

signed-rank test). This indicates that dramatic shifts of the gut microbiota occurred in the PR group but almost not in the GR group at 12 months, in agreement with that much less metabolic variations observed in older women with the oral administration of *L. reuteri* ATCC PTA 6475 by the metabolomic analysis (**Figure 20**). Particularly, *E. coli* was only enriched in the PR group and meanwhile showed differential between the GR and PR groups at 12 months (**Figure 23**). Additionally, *Akkermansia muciniphila*, *Ruminococcus bicirculans*, *Eubacterium* sp_CAG_38 and *Butyricimonas virosa* were depleted in the PR groups at 12 months, compared to the GR group ($P < 0.05$ by the Wilcoxon rank-sum test; check **Paper VI** for details), when analyzing the taxonomic profiles calculated by the MetaPhlAn2 tool [70].

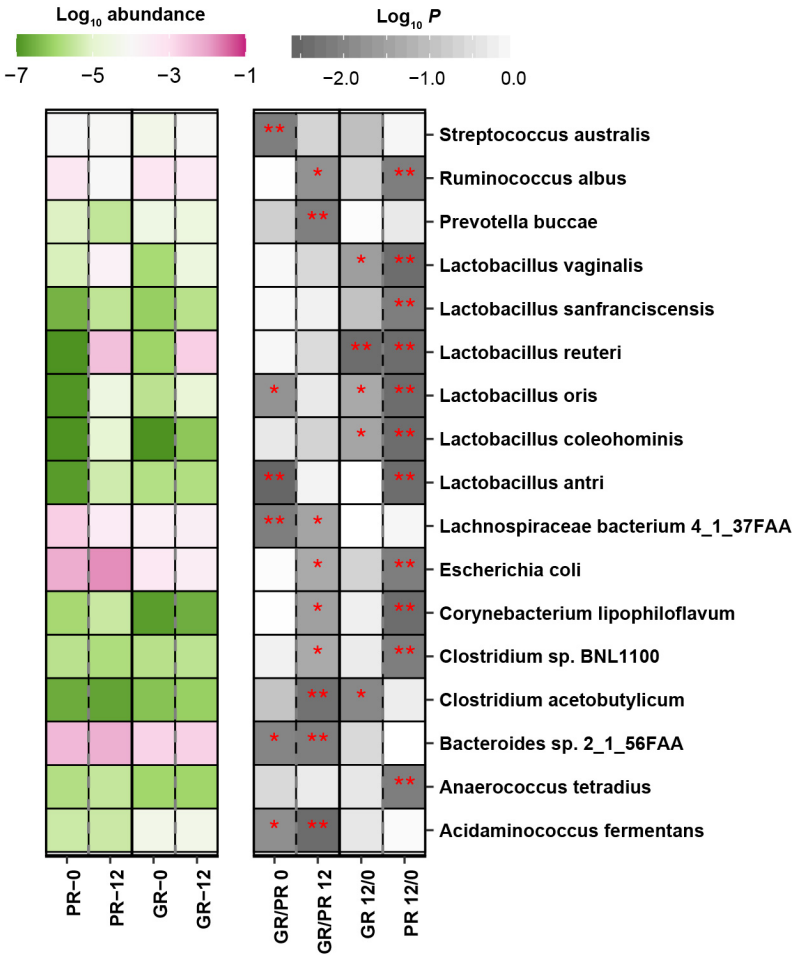

Figure 23 Alterations in the gut microbial composition after oral supplementation with *L. reuteri* ATCC PTA 6475. The left heatmap shows log-transformed mean abundances of differential species in the GR and PR groups at baseline and 12 months. The grey color in the right heatmap indicates *P* value of comparative analysis; '*' denotes $P < 0.05$; '**' denotes $P < 0.01$.

Through investigating the functional capabilities of the gut microbiome, gene richness had an increased trend in the GR group at baseline but not statistically significant, compared to the PR group (**Figure 24a**). Interestingly, gene richness was significantly higher in the GR group than the PR group at 12 months ($P < 0.01$ by the Wilcoxon rank-sum test). Moreover, there were 11 and 157 significantly differential KOs identified between the GR

and PR groups at baseline and 12 months, respectively (**Figure 24b**; Adjusted $P < 0.1$). The differences in the baseline functional potential of the gut microbiota might be important for the positive effects of the probiotic *L. reuteri* ATCC PTA 6475 on the elderly. In addition, 152 differential KOs were identified between the two time points in the PR group, while only 14 differential KOs in the GR group (**Figure 24b**; Adjusted $P < 0.1$). Thus, more significant alterations of the functional capacities in the PR group at 12 months were observed, which is consistent with the results from both microbial composition and metabolomics analysis.

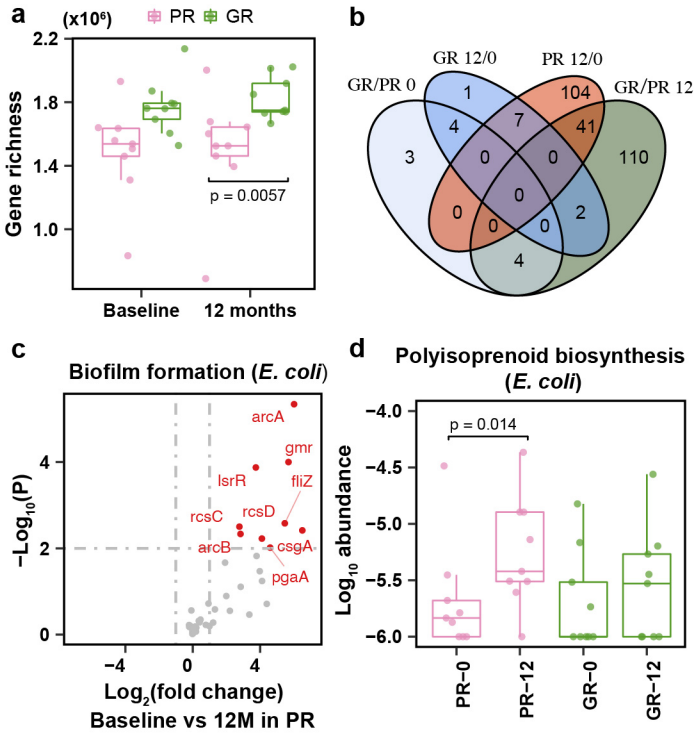

Figure 24 Alterations in the gut microbial function potential after oral supplementation with *L. reuteri* ATCC PTA 6475. a) Gene numbers of the gut microbiota in the PR (n=9) and GR (n=9) groups at the two time points. 20 million reads from each sample were sampled in order to rarefy the reads to the same depth of sequencing. b) The Venn diagram shows differential KOs between the GR and PR groups or between baseline and 12 months (Adjusted *P* value < 0.1). c) The volcano plot displays the differential genes involved in biofilm formation (*Escherichia coli*) between the two time points in the PR group. The horizontal and vertical dashed lines indicate *P* value < 0.01 and $|\log_2$ fold change$|$ >1. d) Abundances of the polyisoprenoid biosynthesis (*E. coli*) pathway in the PR and GR groups at the two time points.

In addition, by the gene set analysis (GSA), the microbial metabolism related to biofilm formation (*E. coli*) was enriched in the PR group at 12 months compared to baseline ($P < 0.05$). The microbial genes involved in the biofilm formation, including *gmr*, *arcA*, *lsrR*, *rcsC* and *rcsD*, showed an increased abundance in the PR group (**Figure 24c**; $P < 0.01$ and $|\log_2$ fold change $> 1|$). Meanwhile, the difference in the metabolism related to biofilm formation (*E. coli*) between the GR and PR groups was observed at 12 months ($P < 0.05$). In agreement with results of the GSA, the functional capacity of polyisoprenoid biosynthesis (*E. coli*) was elevated in the PR group at 12 months compared to baseline (**Figure 24d**; $P < 0.05$ by the Wilcoxon signed-rank test), when using the relative profiles of the MetaCyc pathways calculated by the metagenomic tool HUMAnN2 [73].

3.3.3 The altered gut microbiota linked to the metabolomic changes in response to the probiotic supplementation

As introduced in section 3.2.2, metabolite butyrylcarnitine (C4) responded differentially at all follow-up time points and had a robust increased level in older women supplemented with the probiotic *L. reuteri* ATCC PTA 6475 [136], compared to the placebo group (**Figure 21b**). In line with this, butyrylcarnitine (C4) showed an elevated level in the GR group after one-year probiotic supplementation (**Figure 25a**; $P < 0.05$). Additionally, a positive correlation between species *Akkermansia muciniphila* and butyrylcarnitine (C4) (**Figure 25b**; $R = 0.54$, $P = 0.017$) was observed at 12 months. As previously reported [134], butyrylcarnitine (C4) could act as the pool and transporter of butyrate, which has been shown to promote bone formation in mice [64, 135, 137]. Simultaneously, earlier studies have revealed that *A. muciniphila* has the ability to degrade mucins in the intestine mainly into SCFAs [138-140]. Thus, this correlation result indicates a link between the SCFAs-producing species and the butyrate derivative, which might contribute to the reduction of bone loss in good responders after the probiotic supplementation.

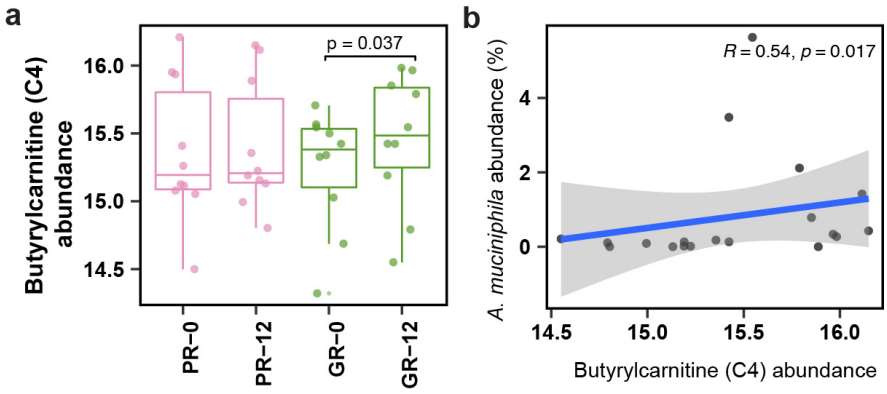

Figure 25 The metabolic changes in older women linked to the gut microbial species. a) Abundances of butyrylcarnitine (C4) in the GR and PR groups at baseline and 12 months. b) Association between butyrylcarnitine (C4) and species *Akkermansia muciniphila*. The blue line and grey shade indicate the regression line and 95% confidence interval; *R* denotes Spearman's correlation coefficient.

In summary, the section 3.3 mainly presents the work from **Paper VI**, where the composition and functional capacity of gut metagenome as well as serum metabolome in good or poor responders to the probiotic intake were investigated. After one-year supplementation with *L. reuteri* ATCC PTA 6475, the decreased inflammation and significantly increased gene richness of the gut microbiota in the good responders were revealed. Moreover, detrimental changes including the enrichment of *E. coli* and its biofilm formation observed in the poor responders were alleviated in the good responders after the probiotic intake. In addition, a potential link between the SCFAs-producing species *A. muciniphila* and the butyrate derivative butyrylcarnitine (C4) were observed, confirming that the effects of *L. reuteri* ATCC PTA 6475 on bone metabolism might be regulated through the butyrate signaling. Overall, by integrative analysis of the metabolomics and gut metagenomics, oral supplementation with *L. reuteri* ATCC PTA 6475 has been suggested to have the potential to prevent a deterioration of the gut microbiota and the host

metabolism in elderly women with low BMD, which might contribute to the beneficial effects on bone loss in the elderly. This study provides a new insight into the regulation of bone metabolism and could be crucial for the development of novel osteoporosis treatments. However, further studies are needed to validate the links between the gut microbial alterations and the host metabolic changes triggered by supplementation with probiotic *L. reuteri* ATCC PTA 6475.

4. Conclusions

This thesis mainly interrogates the associations between the gut microbiota, probiotics and human metabolism by using ML and GEMs to integrate gut metagenomics with serum metabolomics. For this, I first discussed the current literature about associations between gut microbiota, human diseases and applications of ML in **Paper I**. Additionally, this thesis mainly answers the three questions that are raised in the background part (section 1.6).

In **Paper II** and **Paper III**, associations between the gut microbiota and T2D were mainly investigated. I first explored how different underlying factors, including the host metabolism and the gut microbiota, contributed to the abnormally postprandial responses in individuals with (pre)diabetes (**Paper II**). By integrative analysis of metabolomics and metagenomics, the derivatives of BCAAs and phenylalanine were identified to be a potential link between the gut microbiota and T2D, which were associated with insulin resistance and might contribute to the metabolic imbalance of (pre)diabetes. Further, using ML and community-level metabolic models, I performed a systematical analysis of four metagenomics data sets related to (pre)diabetes (**Paper III**). A number of SCFAs-producing bacterial species and metabolic reactions were consistently identified to be important for predicting T2D status across studies, which is in line with a reduction of species with butyrate producing capacity (**Paper II**). These findings suggest that alteration in the SCFAs-producing capabilities of the gut microbiota might play a critical role in the pathology of T2D progression.

Furthermore, this thesis focuses on the effects of probiotic *L. reuteri* ATCC PTA 6475 on bone metabolism of older women with low BMD. Using the GEM of *L. reuteri* ATCC PTA 6475, this work investigated the biosynthesis pathways of a number of beneficial metabolites e.g., SCFAs, which helps to understand the potential benefits of the probiotics to human metabolism (**Paper VI**). By metabolomic profiling, this work revealed that one-year supplementation with the probiotic alleviated the significantly metabolic changes of older women with bone loss as occurred in the placebo group (**Paper V**). By integrative analysis of metabolomics and metagenomics, this thesis further found that, in good responders, *L. reuteri* ATCC PTA 6475 had the potential to prevent detrimental changes of the gut microbiota (**Paper VI**). In addition, a potential link between the SCFAs-producing species *A. muciniphila* and the butyrate derivative butyrylcarnitine (C4) was observed, suggesting that the effects of *L. reuteri* ATCC PTA 6475 on bone metabolism might be regulated through butyrate signaling. These findings provide new insights into the regulation of bone metabolism and could be crucial for the development of novel osteoporosis treatments.

GEMs serve as a useful tool to study metabolic questions. In **Paper VI**, this work demonstrated that GEMs could enable us to examine the metabolism of single probiotic strain *L. reuteri* ATCC PTA 6475 in detail. In **Paper III**, GEMs were used to construct the

community-level metabolic model of individual gut microbiota, which helps us to investigate its holistic functional capacities via simulating metabolic reactions. Therefore, modeling the metabolisms by the GEMs could provide a reliable basis for identifying metabolic signatures related to diseases and even interrogating relationships between the gut microbial species.

In this thesis, the trained regression or classification models not only showed an adequate predictive accuracy, but also identified important disease-related features. In **Paper II**, the regression models were applied to predictions of the postprandial glucose responses to a meal using multi-omics data. This work found that blood metabolomics-based models had better performance in comparison to other omics data. Also, the obtained interpretable models had the potential to identify both the important serum metabolic and gut microbial features that might contribute to the abnormal glucose control in individuals with (pre)diabetes. In **Paper III**, different classification models were trained to predict T2D status based on various gut microbial features. Through this, several key SCFAs-producing microbial species and the metabolic reactions were consistently identified to be critical for detecting T2D risk. In addition, this work suggests that the gut microbiota-based models for predicting T2D status might be specific to the studied population or region and challenging to be generalized across multiple cohorts.

To sum up, this thesis contributes to knowledge on associations between the gut microbiota and the human diseases as well as the beneficial effects of *L. reuteri* ATCC PTA 6475 on bone metabolism. In addition, this work suggests that ML in combination with GEMs has the potential to identify new microbial signatures related to diseases.

5. Future perspectives

With the fast development of sequencing technologies, the genome sequences of more and more species of the human gut microbiota have been uncovered. The considerable gene reservoir can enable us to better understand the composition and potential functions of the gut microbiota. Increasing studies have used metagenomics to reveal associations between the gut microbial species and human diseases. Nevertheless, this is just a first step to set up a correlation, which further need to be verified as a causal relationship, e.g., through the use of germ-free mice. Additionally, serum metabolomics has been widely used to study the metabolism of the human host as well as to identify the gut microbe-derived metabolites. Therefore, integrative analysis of the metagenomics and metabolomics has a potential to reveal robust links between the gut microbiota and its human host, e.g., the signaling pathways the microbe-derived metabolites are involved in, which could provide new insights into the causal roles of key species in the human diseases.

Identification of the causality from metagenomic studies enables us to develop efficient intervention strategies for improving human health by targeting the gut microbiota, such as probiotics, prebiotics and personalized nutrition. Especially, oral supplementation with probiotics has been evaluated to be a safe and efficient intervention in a number of double blind, randomized placebo-controlled trials. However, subjects have differential responses to the oral administration of probiotics, i.e., some responded poorly. Therefore, further studies need to be performed for validation of the probiotic efficacy. Identification of the factors that contribute to the good or poor responses might help us to design personalized intervention. Additionally, the intake of prebiotics, referred to as chemicals that induce the growth of commensal microorganisms, could be a good choice to improve human health.

Associations between the gut microbiota and a unique disease have often been revealed to be inconsistent across different metagenomic studies. This might be due to other factors, such as drugs, age, diet, geography, lifestyle, could influence associations between the gut microbiota and the disease. Therefore, these factors need to be taken into account when researchers design experiments and analyze the metagenomics data. Particularly, integrative analysis of complex data including multi-omics, dietary composition and clinic data, has been required to elaborate the underlying links between human diseases and the gut microbiota. However, it’s challenging to efficiently extract disease-specific signatures from the complicated interactions between the environmental factors, gut microbiota and host metabolism.

ML holds great promise to integrate these heterogeneous data for generating interpretable models that could not only predict phenotypes but also identify potential biomarkers related to human diseases, thus allowing us to gain novel insights into disease pathogenesis and further propose potential intervention strategies. Nevertheless, due to high-dimensional data including extremely large amounts of molecular variables with relatively small samples, it is challenging to develop robust and reliable prediction models, and easily

leads to overfitting problem. To mitigate this, a range of techniques could be useful such as using feature selection, reducing model complexity and utilizing data augmentation. In addition, a set of autoencoder-based deep learning methods have been devised to transform high-dimensional features into low-dimensional latent representations, which could be used for further analysis and prediction. However, the development of gut microbial predictive models and diagnostic biomarkers would possibly be specific to the studied population, and difficult to be generalized across multiple ethnicities or geographies.

GEMs are a powerful tool for studying metabolisms of a single gut microbe or entire microbial community, which could provide new knowledge about the gut microbiota via simulating its metabolic capabilities such as biomass growth, target metabolite synthesis. However, it is indispensable to address a number of challenges, such as considerable manual curations of a draft model, part of microbial species with little gene annotations or limited experimental data. In order to model metabolisms of individual gut microbiota consisting of hundreds of species, a community-level metabolic model is usually constructed based on many GEMs. This process is time-consuming to refine the draft model and needs to be accelerated.

Taking together, the joint use of the ML and GEMs for integration of complex data provides a great opportunity to elucidate the causal roles of key microbes in human diseases. Also, it has the potential to assist in developing gut microbiota-targeted intervention strategies for prevention and treatment of human diseases, which would be promising solutions for precision medicine.

Acknowledgements

First of all, I would like to thank my supervisor Jens Nielsen for giving me the opportunity to join SysBio, for supporting me with constructive suggestions and freedom to explore those projects, for providing advice to my personal development. I would benefit a lot from these experience for the rest of my life. Thank my co-supervisor Boyang Ji for all your discussions and suggestions during my studies, and for reviews and comments of my manuscripts. Thanks to my co-supervisor Aleksej Zelezniak, examiners Verena Siewers and Dina Petranovic for all your helps and discussions we had.

I have had the honor to work with many kind collaborators during my PhD studies. Thanks to Fredrik Bäckhed, Max Nieuwdorp, Thue W. Schwartz and Louise E. Olofsson for your valuable discussions and constructive suggestions to my first project. Thanks to Valentina Tremaroli, Lisa M. Olsson, Annika Lundqvist for sample preparation and omics preprocessing. Thank Abraham S Meijnikman, Ömrüm Aydin, Arnold van de Laar and Sjoerd C. Bruin for collecting medical data and biopsies. Additionally, I would like to thank Mattias Lorentzon and Daniel Sundh at Sahlgrenska University Hospital for your valuable suggestions and contributions to the probiotic projects. I have learnt lots of knowledge about bone disease and probiotics from all the discussions we had. Thanks to Hao Luo for your valuable contributions to my studies, especially in respect of metabolic modeling. Thanks to Dimitra Lappa for your kind help and critical input to my project. Thank Linqun Ye for allowing me to join your interesting project as well as your helps on metagenomics analysis.

I also would like to thank many colleagues and friends for your helps during these years. Thank Angelo and Hao Wang for your valuable suggestions to my projects. Thanks to Ivan, Yun Chen, Joakim, Eduard, Lei Shi, Xin Chen, Gang Li, Jun Geng, Yating, Xiaowei, Behnaz, Johan, Jing Fu, Le Yuan, Carl, Rasool, Filip, Naghmeh, Fariba, Xiang Jiao, Iván, Xiaozhi, Yanyan Chen, Veronica, Avlant, Zhengming and Chunjun. Thanks to Feiran, Qi Qi, Yanyan Wang, Jiwei and Yu Chen for afterwork cooking and games, which brought me lots of joy. Thanks to research engineers and administrators, Anne-Lise, Erica, Martina, Gunilla, Elin and Mihail for your professional assistance.

Finally, my special thanks to my family. My parents and sister for your unconditional support. My wife for your encouragement and accompany all the time. My kids for giving me power to go forward. Hope to share my life with you forever!

References

[1] S. R. Gill *et al.*, "Metagenomic analysis of the human distal gut microbiome," *Science,* vol. 312, no. 5778, pp. 1355-9, Jun 2 2006, doi: 10.1126/science.1124234.

[2] A. Almeida *et al.*, "A unified catalog of 204,938 reference genomes from the human gut microbiome," *Nature biotechnology,* vol. 39, no. 1, pp. 105-114, Jan 2021, doi: 10.1038/s41587-020-0603-3.

[3] R. E. Ley, F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon, "Obesity alters gut microbial ecology," *Proc Natl Acad Sci U S A,* vol. 102, no. 31, pp. 11070-5, Aug 2 2005, doi: 10.1073/pnas.0504978102.

[4] L. Zhao, "The gut microbiota and obesity: from correlation to causality," *Nat Rev Microbiol,* vol. 11, no. 9, pp. 639-47, Sep 2013, doi: 10.1038/nrmicro3089.

[5] F. H. Karlsson, I. Nookaew, and J. Nielsen, "Metagenomic data utilization and analysis (MEDUSA) and construction of a global gut microbial gene catalogue," *PLoS Comput Biol,* vol. 10, no. 7, p. e1003706, Jul 2014, doi: 10.1371/journal.pcbi.1003706.

[6] R. Wall *et al.*, "Role of gut microbiota in early infant development," *Clin Med Pediatr,* vol. 3, pp. 45-54, 2009, doi: 10.4137/cmped.s2008.

[7] J. Penders *et al.*, "Factors influencing the composition of the intestinal microbiota in early infancy," *Pediatrics,* vol. 118, no. 2, pp. 511-21, Aug 2006, doi: 10.1542/peds.2005-2824.

[8] A. B. Hall, A. C. Tolonen, and R. J. Xavier, "Human genetic variation and the gut microbiome in disease," *Nat Rev Genet,* vol. 18, no. 11, pp. 690-699, Nov 2017, doi: 10.1038/nrg.2017.63.

[9] A. W. Brooks, S. Priya, R. Blekhman, and S. R. Bordenstein, "Gut microbiota diversity across ethnicities in the United States," *PLoS Biol,* vol. 16, no. 12, p. e2006842, Dec 2018, doi: 10.1371/journal.pbio.2006842.

[10] M. Deschasaux *et al.*, "Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography," *Nat Med,* vol. 24, no. 10, pp. 1526-1531, Oct 2018, doi: 10.1038/s41591-018-0160-1.

[11] D. Rothschild *et al.*, "Environment dominates over host genetics in shaping human gut microbiota," *Nature,* vol. 555, no. 7695, pp. 210-215, Mar 8 2018, doi: 10.1038/nature25973.

[12] N. Zmora, J. Suez, and E. Elinav, "You are what you eat: diet, health and the gut microbiota," *Nat Rev Gastroenterol Hepatol,* vol. 16, no. 1, pp. 35-56, Jan 2019, doi: 10.1038/s41575-018-0061-2.

[13] F. Asnicar *et al.*, "Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals," *Nat Med,* vol. 27, no. 2, pp. 321-332, Feb 2021, doi: 10.1038/s41591-020-01183-8.

[14] G. D. Wu *et al.*, "Linking long-term dietary patterns with gut microbial enterotypes," *Science,* vol. 334, no. 6052, pp. 105-8, Oct 7 2011, doi: 10.1126/science.1208344.

[15] L. Ye, P. Das, P. Li, B. Ji, and J. Nielsen, "Carbohydrate active enzymes are affected by diet transition from milk to solid food in infant gut microbiota," *FEMS Microbiol Ecol,* vol. 95, no. 11, Nov 1 2019, doi: 10.1093/femsec/fiz159.

[16] A. Mardinoglu, J. Boren, and U. Smith, "Confounding Effects of Metformin on the Human Gut Microbiome in Type 2 Diabetes," *Cell metabolism,* vol. 23, no. 1, pp. 10-2, Jan 12 2016, doi: 10.1016/j.cmet.2015.12.012.

[17] T. Yatsunenko *et al.*, "Human gut microbiome viewed across age and geography," *Nature,* vol. 486, no. 7402, pp. 222-7, May 9 2012, doi: 10.1038/nature11053.

[18] Y. He *et al.*, "Regional variation limits applications of healthy gut microbiome reference ranges and disease models," *Nat Med,* vol. 24, no. 10, pp. 1532-1535, Oct 2018, doi: 10.1038/s41591-018-0164-x.

[19] A. Koliada *et al.*, "Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population," (in English), *Bmc Microbiol,* vol. 17, May 22 2017, doi: ARTN 120 10.1186/s12866-017-1027-1.

[20] W. Jia, H. Li, L. Zhao, and J. K. Nicholson, "Gut microbiota: a potential new territory for drug targeting," *Nat Rev Drug Discov,* vol. 7, no. 2, pp. 123-9, Feb 2008, doi: 10.1038/nrd2505.

[21] A. Cotillard *et al.*, "Dietary intervention impact on gut microbial gene richness," *Nature,* vol. 500, no. 7464, pp. 585-8, Aug 29 2013, doi: 10.1038/nature12480.

[22] T. S. Ghosh *et al.*, "Mediterranean diet intervention alters the gut microbiome in older people reducing frailty and improving health status: the NU-AGE 1-year dietary intervention across five European countries," *Gut,* vol. 69, no. 7, pp. 1218-1228, Jul 2020, doi: 10.1136/gutjnl-2019-319654.

[23] S. E. Berry *et al.*, "Human postprandial responses to food and potential for precision nutrition," *Nat Med,* vol. 26, no. 6, pp. 964-973, Jun 2020, doi: 10.1038/s41591-020-0934-0.
[24] D. Zeevi *et al.*, "Personalized Nutrition by Prediction of Glycemic Responses," *Cell,* vol. 163, no. 5, pp. 1079-1094, Nov 19 2015, doi: 10.1016/j.cell.2015.11.001.
[25] M. Le Barz *et al.*, "Probiotics as Complementary Treatment for Metabolic Disorders," *Diabetes Metab J,* vol. 39, no. 4, pp. 291-303, Aug 2015, doi: 10.4093/dmj.2015.39.4.291.
[26] R. A. Rastall and G. R. Gibson, "Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health," *Curr Opin Biotechnol,* vol. 32, pp. 42-6, Apr 2015, doi: 10.1016/j.copbio.2014.11.002.
[27] P. W. O'Toole, J. R. Marchesi, and C. Hill, "Next-generation probiotics: the spectrum from probiotics to live biotherapeutics," *Nat Microbiol,* vol. 2, p. 17057, Apr 25 2017, doi: 10.1038/nmicrobiol.2017.57.
[28] S. Sabico *et al.*, "Effects of a 6-month multi-strain probiotics supplementation in endotoxemic, inflammatory and cardiometabolic status of T2DM patients: A randomized, double-blind, placebo-controlled trial," *Clin Nutr,* vol. 38, no. 4, pp. 1561-1569, Aug 2019, doi: 10.1016/j.clnu.2018.08.009.
[29] M. Karamali *et al.*, "Effects of probiotic supplementation on glycaemic control and lipid profiles in gestational diabetes: A randomized, double-blind, placebo-controlled trial," *Diabetes Metab,* vol. 42, no. 4, pp. 234-41, Sep 2016, doi: 10.1016/j.diabet.2016.04.009.
[30] R. Mobini *et al.*, "Metabolic effects of Lactobacillus reuteri DSM 17938 in people with type 2 diabetes: A randomized controlled trial," *Diabetes Obes Metab,* vol. 19, no. 4, pp. 579-589, Apr 2017, doi: 10.1111/dom.12861.
[31] N. C. D. R. F. Collaboration, "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants," *Lancet,* vol. 387, no. 10027, pp. 1513-1530, Apr 9 2016, doi: 10.1016/S0140-6736(16)00618-8.
[32] P. Z. Zimmet, "Diabetes and its drivers: the largest epidemic in human history?," *Clin Diabetes Endocrinol,* vol. 3, p. 1, 2017, doi: 10.1186/s40842-016-0039-3.
[33] A. G. Tabak, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimaki, "Prediabetes: a high-risk state for diabetes development," *Lancet,* vol. 379, no. 9833, pp. 2279-90, Jun 16 2012, doi: 10.1016/S0140-6736(12)60283-9.
[34] A. American Diabetes, "Diagnosis and classification of diabetes mellitus," *Diabetes Care,* vol. 34 Suppl 1, pp. S62-9, Jan 2011, doi: 10.2337/dc11-S062.
[35] E. Corpeleijn, W. H. Saris, and E. E. Blaak, "Metabolic flexibility in the development of insulin resistance and type 2 diabetes: effects of lifestyle," *Obes Rev,* vol. 10, no. 2, pp. 178-93, Mar 2009, doi: 10.1111/j.1467-789X.2008.00544.x.
[36] B. H. Goodpaster and L. M. Sparks, "Metabolic Flexibility in Health and Disease," *Cell metabolism,* vol. 25, no. 5, pp. 1027-1036, May 2 2017, doi: 10.1016/j.cmet.2017.04.015.
[37] R. A. DeFronzo and D. Tripathy, "Skeletal muscle insulin resistance is the primary defect in type 2 diabetes," *Diabetes Care,* vol. 32 Suppl 2, pp. S157-63, Nov 2009, doi: 10.2337/dc09-S302.
[38] M. Cnop, N. Welsh, J. C. Jonas, A. Jorns, S. Lenzen, and D. L. Eizirik, "Mechanisms of pancreatic beta-cell death in type 1 and type 2 diabetes: many differences, few similarities," *Diabetes,* vol. 54 Suppl 2, pp. S97-107, Dec 2005, doi: 10.2337/diabetes.54.suppl_2.s97.
[39] C. J. Greenbaum *et al.*, "Mixed-Meal Tolerance Test Versus Glucagon Stimulation Test for the Assessment of beta-Cell Function in Therapeutic Trials in Type 1 Diabetes," (in English), *Diabetes Care,* vol. 31, no. 10, pp. 1966-1971, Oct 2008, doi: 10.2337/dc07-2451.
[40] R. E. J. Besser, A. G. Jones, T. J. McDonald, B. M. Shields, B. A. Knight, and A. T. Hattersley, "The impact of insulin administration during the mixed meal tolerance test," (in English), *Diabetic Med,* vol. 29, no. 10, pp. 1279-1284, Oct 2012, doi: 10.1111/j.1464-5491.2012.03649.x.
[41] M. Ahmed, M. C. Gannon, and F. Q. Nuttall, "Postprandial plasma glucose, insulin, glucagon and triglyceride responses to a standard diet in normal subjects," *Diabetologia,* vol. 12, no. 1, pp. 61-7, Mar 1976, doi: 10.1007/bf01221966.
[42] R. E. Besser, B. M. Shields, R. Casas, A. T. Hattersley, and J. Ludvigsson, "Lessons from the mixed-meal tolerance test: use of 90-minute and fasting C-peptide in pediatric diabetes," *Diabetes Care,* vol. 36, no. 2, pp. 195-201, Feb 2013, doi: 10.2337/dc12-0836.
[43] S. Wopereis *et al.*, "Multi-parameter comparison of a standardized mixed meal tolerance test in healthy and type 2 diabetic subjects: the PhenFlex challenge," *Genes Nutr,* vol. 12, p. 21, 2017, doi: 10.1186/s12263-017-0570-6.

[44] L. Pellis *et al.*, "Plasma metabolomics and proteomics profiling after a postprandial challenge reveal subtle diet effects on human metabolic status," *Metabolomics,* vol. 8, no. 2, pp. 347-359, Apr 2012, doi: 10.1007/s11306-011-0320-5.

[45] S. S. Shankar *et al.*, "Standardized Mixed-Meal Tolerance and Arginine Stimulation Tests Provide Reproducible and Complementary Measures of beta-Cell Function: Results From the Foundation for the National Institutes of Health Biomarkers Consortium Investigative Series," *Diabetes Care,* vol. 39, no. 9, pp. 1602-13, Sep 2016, doi: 10.2337/dc15-0931.

[46] L. B. Thingholm *et al.*, "Obese Individuals with and without Type 2 Diabetes Show Different Gut Microbial Functional Capacity and Composition," (in English), *Cell Host & Microbe,* vol. 26, no. 2, pp. 252-+, Aug 14 2019, doi: 10.1016/j.chom.2019.07.004.

[47] K. Forslund *et al.*, "Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota," *Nature,* vol. 528, no. 7581, pp. 262-266, Dec 10 2015, doi: 10.1038/nature15766.

[48] J. Qin *et al.*, "A metagenome-wide association study of gut microbiota in type 2 diabetes," *Nature,* vol. 490, no. 7418, pp. 55-60, Oct 4 2012, doi: 10.1038/nature11450.

[49] F. H. Karlsson *et al.*, "Gut metagenome in European women with normal, impaired and diabetic glucose control," *Nature,* vol. 498, no. 7452, pp. 99-103, Jun 06 2013, doi: 10.1038/nature12198.

[50] H. Zhong *et al.*, "Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naive type 2 diabetics," *EBioMedicine,* vol. 47, pp. 373-383, Sep 2019, doi: 10.1016/j.ebiom.2019.08.048.

[51] H. Wu *et al.*, "The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study," *Cell metabolism,* Jul 2 2020, doi: 10.1016/j.cmet.2020.06.011.

[52] A. Koh and F. Backhed, "From Association to Causality: the Role of the Gut Microbiota and Its Functional Products on Host Metabolism," *Mol Cell,* vol. 78, no. 4, pp. 584-596, May 21 2020, doi: 10.1016/j.molcel.2020.03.005.

[53] H. K. Pedersen *et al.*, "Human gut microbes impact host serum metabolome and insulin sensitivity," *Nature,* vol. 535, no. 7612, pp. 376-81, Jul 21 2016, doi: 10.1038/nature18646.

[54] A. Koh *et al.*, "Microbially Produced Imidazole Propionate Impairs Insulin Signaling through mTORC1," *Cell,* vol. 175, no. 4, pp. 947-961 e17, Nov 1 2018, doi: 10.1016/j.cell.2018.09.055.

[55] S. G. Wannamethee, A. G. Shaper, I. J. Perry, and S. British Regional Heart, "Smoking as a modifiable risk factor for type 2 diabetes in middle-aged men," *Diabetes Care,* vol. 24, no. 9, pp. 1590-5, Sep 2001. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/11522704.

[56] M. Lorentzon and S. R. Cummings, "Osteoporosis: the evolution of a diagnosis," *Journal of internal medicine,* vol. 277, no. 6, pp. 650-61, Jun 2015, doi: 10.1111/joim.12369.

[57] M. Lorentzon, "Treating osteoporosis to prevent fractures: current concepts and future developments," *Journal of internal medicine,* vol. 285, no. 4, pp. 381-394, Apr 2019, doi: 10.1111/joim.12873.

[58] S. Khosla and L. C. Hofbauer, "Osteoporosis treatment: recent developments and ongoing challenges," *Lancet Diabetes Endocrinol,* vol. 5, no. 11, pp. 898-907, Nov 2017, doi: 10.1016/S2213-8587(17)30188-2.

[59] C. Ohlsson and K. Sjogren, "Effects of the gut microbiota on bone mass," *Trends in endocrinology and metabolism: TEM,* vol. 26, no. 2, pp. 69-74, Feb 2015, doi: 10.1016/j.tem.2014.11.004.

[60] C. Medina-Gomez, "Bone and the gut microbiome: a new dimension," *Journal of Laboratory and Precision Medicine,* 2018.

[61] J. Zhang, Y. Lu, Y. Wang, X. Ren, and J. Han, "The impact of the intestinal microbiome on bone health," *Intractable Rare Dis Res,* vol. 7, no. 3, pp. 148-155, Aug 2018, doi: 10.5582/irdr.2018.01055.

[62] M. M. Zaiss, R. M. Jones, G. Schett, and R. Pacifici, "The gut-bone axis: how bacterial metabolites bridge the distance," (in English), *Journal of Clinical Investigation,* vol. 129, no. 8, pp. 3018-3028, Aug 1 2019, doi: 10.1172/Jci128521.

[63] R. Berni Canani *et al.*, "Lactobacillus rhamnosus GG-supplemented formula expands butyrate-producing bacterial strains in food allergic infants," *ISME J,* vol. 10, no. 3, pp. 742-50, Mar 2016, doi: 10.1038/ismej.2015.151.

[64] A. M. Tyagi *et al.*, "The Microbial Metabolite Butyrate Stimulates Bone Formation via T Regulatory Cell-Mediated Regulation of WNT10B Expression," *Immunity,* vol. 49, no. 6, pp. 1116-1131 e7, Dec 18 2018, doi: 10.1016/j.immuni.2018.10.013.

[65] L. McCabe, R. A. Britton, and N. Parameswaran, "Prebiotic and Probiotic Regulation of Bone Health: Role of the Intestine and its Microbiome," (in English), *Current Osteoporosis Reports,* vol. 13, no. 6, pp. 363-371, Dec 2015, doi: 10.1007/s11914-015-0292-x.

[66] R. A. Britton *et al.*, "Probiotic L. reuteri treatment prevents bone loss in a menopausal ovariectomized mouse model," *J Cell Physiol,* vol. 229, no. 11, pp. 1822-30, Nov 2014, doi: 10.1002/jcp.24636.
[67] F. L. Collins *et al.*, "Lactobacillus reuteri 6475 Increases Bone Density in Intact Females Only under an Inflammatory Setting," *PLoS One,* vol. 11, no. 4, p. e0153180, 2016, doi: 10.1371/journal.pone.0153180.
[68] A. G. Nilsson, D. Sundh, F. Backhed, and M. Lorentzon, "Lactobacillus reuteri reduces bone loss in older women with low bone mineral density: a randomized, placebo-controlled, double-blind, clinical trial," *Journal of internal medicine,* vol. 284, no. 3, pp. 307-317, Sep 2018, doi: 10.1111/joim.12805.
[69] X. C. Morgan and C. Huttenhower, "Chapter 12: Human microbiome analysis," *PLoS Comput Biol,* vol. 8, no. 12, p. e1002808, 2012, doi: 10.1371/journal.pcbi.1002808.
[70] N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, and C. Huttenhower, "Metagenomic microbial community profiling using unique clade-specific marker genes," *Nat Methods,* vol. 9, no. 8, pp. 811-4, Jun 10 2012, doi: 10.1038/nmeth.2066.
[71] A. Milanese *et al.*, "Microbial abundance, activity and population genomic profiling with mOTUs2," *Nat Commun,* vol. 10, no. 1, p. 1014, Mar 4 2019, doi: 10.1038/s41467-019-08844-4.
[72] F. Meyer *et al.*, "The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes," *BMC Bioinformatics,* vol. 9, p. 386, Sep 19 2008, doi: 10.1186/1471-2105-9-386.
[73] E. A. Franzosa *et al.*, "Species-level functional profiling of metagenomes and metatranscriptomes," *Nat Methods,* vol. 15, no. 11, pp. 962-968, Nov 2018, doi: 10.1038/s41592-018-0176-y.
[74] Y. Li, Y. Hu, L. Bolund, and J. Wang, "State of the art de novo assembly of human genomes from massively parallel sequencing data," *Hum Genomics,* vol. 4, no. 4, pp. 271-7, Apr 2010, doi: 10.1186/1479-7364-4-4-271.
[75] J. R. Kultima *et al.*, "MOCAT: a metagenomics assembly and gene prediction toolkit," *PLoS One,* vol. 7, no. 10, p. e47656, 2012, doi: 10.1371/journal.pone.0047656.
[76] M. Kanehisa, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, "The KEGG resource for deciphering the genome," *Nucleic Acids Res,* vol. 32, no. Database issue, pp. D277-80, Jan 1 2004, doi: 10.1093/nar/gkh063.
[77] R. L. Tatusov *et al.*, "The COG database: an updated version includes eukaryotes," *BMC Bioinformatics,* vol. 4, p. 41, Sep 11 2003, doi: 10.1186/1471-2105-4-41.
[78] L. J. Jensen *et al.*, "eggNOG: automated construction and annotation of orthologous groups of genes," *Nucleic Acids Res,* vol. 36, no. Database issue, pp. D250-4, Jan 2008, doi: 10.1093/nar/gkm796.
[79] B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics," *Nucleic Acids Res,* vol. 37, no. Database issue, pp. D233-8, Jan 2009, doi: 10.1093/nar/gkn663.
[80] C. H. Johnson, J. Ivanisevic, and G. Siuzdak, "Metabolomics: beyond biomarkers and towards mechanisms," *Nat Rev Mol Cell Biol,* vol. 17, no. 7, pp. 451-9, Jul 2016, doi: 10.1038/nrm.2016.25.
[81] J. K. Nicholson and J. C. Lindon, "Systems biology: Metabonomics," *Nature,* vol. 455, no. 7216, pp. 1054-6, Oct 23 2008, doi: 10.1038/4551054a.
[82] D. S. Wishart *et al.*, "HMDB: the Human Metabolome Database," *Nucleic Acids Res,* vol. 35, no. Database issue, pp. D521-6, Jan 2007, doi: 10.1093/nar/gkl923.
[83] S. Kim *et al.*, "PubChem Substance and Compound databases," *Nucleic Acids Res,* vol. 44, no. D1, pp. D1202-13, Jan 4 2016, doi: 10.1093/nar/gkv951.
[84] J. Chong *et al.*, "MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis," *Nucleic Acids Res,* vol. 46, no. W1, pp. W486-W494, Jul 2 2018, doi: 10.1093/nar/gky310.
[85] H. Chu, Y. Duan, L. Yang, and B. Schnabl, "Small metabolites, possible big changes: a microbiota-centered view of non-alcoholic fatty liver disease," *Gut,* vol. 68, no. 2, pp. 359-370, Feb 2019, doi: 10.1136/gutjnl-2018-316307.
[86] T. Hendrikx and B. Schnabl, "Indoles: metabolites produced by intestinal bacteria capable of controlling liver disease manifestation," *Journal of internal medicine,* vol. 286, no. 1, pp. 32-40, Jul 2019, doi: 10.1111/joim.12892.
[87] A. Wahlstrom, S. I. Sayin, H. U. Marschall, and F. Backhed, "Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism," *Cell metabolism,* vol. 24, no. 1, pp. 41-50, Jul 12 2016, doi: 10.1016/j.cmet.2016.05.005.

[88] A. Visconti *et al.*, "Interplay between the human gut microbiome and host metabolism," *Nat Commun,* vol. 10, no. 1, p. 4505, Oct 3 2019, doi: 10.1038/s41467-019-12476-z.
[89] T. Wilmanski *et al.*, "Blood metabolome predicts gut microbiome alpha-diversity in humans," *Nature biotechnology,* vol. 37, no. 10, pp. 1217-1228, Oct 2019, doi: 10.1038/s41587-019-0233-9.
[90] A. Kurilshikov *et al.*, "Gut Microbial Associations to Plasma Metabolites Linked to Cardiovascular Phenotypes and Risk A Cross-Sectional Study," (in English), *Circ Res,* vol. 124, no. 12, pp. 1808-1820, Jun 7 2019, doi: 10.1161/Circresaha.118.314642.
[91] E. A. Franzosa *et al.*, "Gut microbiome structure and metabolic activity in inflammatory bowel disease," (in English), *Nature Microbiology,* vol. 4, no. 2, pp. 293-305, Feb 2019, doi: 10.1038/s41564-018-0306-4.
[92] R. Caspi *et al.*, "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases," *Nucleic Acids Res,* vol. 44, no. D1, pp. D471-80, Jan 4 2016, doi: 10.1093/nar/gkv1164.
[93] C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, "High-throughput generation, optimization and analysis of genome-scale metabolic models," *Nature biotechnology,* vol. 28, no. 9, pp. 977-82, Sep 2010, doi: 10.1038/nbt.1672.
[94] L. Heirendt *et al.*, "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0," *Nat Protoc,* vol. 14, no. 3, pp. 639-702, Mar 2019, doi: 10.1038/s41596-018-0098-2.
[95] H. Wang *et al.*, "RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor," *PLoS Comput Biol,* vol. 14, no. 10, p. e1006541, Oct 2018, doi: 10.1371/journal.pcbi.1006541.
[96] H. Luo, P. Li, H. Wang, S. Roos, B. Ji, and J. Nielsen, "Genome-scale insights into the metabolic versatility of Limosilactobacillus reuteri," *BMC Biotechnol,* vol. 21, no. 1, p. 46, Jul 30 2021, doi: 10.1186/s12896-021-00702-w.
[97] T. Kristjansdottir *et al.*, "A metabolic reconstruction of Lactobacillus reuteri JCM 1112 and analysis of its potential as a cell factory," *Microb Cell Fact,* vol. 18, no. 1, p. 186, Oct 29 2019, doi: 10.1186/s12934-019-1229-3.
[98] E. Vinay-Lara, J. J. Hamilton, B. Stahl, J. R. Broadbent, J. L. Reed, and J. L. Steele, "Genome-scale reconstruction of metabolic networks of Lactobacillus casei ATCC 334 and 12A," *PLoS One,* vol. 9, no. 11, p. e110785, 2014, doi: 10.1371/journal.pone.0110785.
[99] S. Shoaie, F. Karlsson, A. Mardinoglu, I. Nookaew, S. Bordel, and J. Nielsen, "Understanding the interactions between bacteria in the human gut through metabolic modeling," *Scientific reports,* vol. 3, p. 2532, 2013, doi: 10.1038/srep02532.
[100] M. Kumar *et al.*, "Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling," *Metab Eng,* vol. 49, pp. 128-142, Sep 2018, doi: 10.1016/j.ymben.2018.07.018.
[101] J. A. Hartigan and M. A. Wong, "A k-means clustering algorithm," *JSTOR: Applied Statistics,* vol. 28, no. 1, pp. 100--108, 1979.
[102] M. Arumugam *et al.*, "Enterotypes of the human gut microbiome," (in English), *Nature,* vol. 473, no. 7346, pp. 174-180, May 12 2011, doi: 10.1038/nature09944.
[103] F. H. Karlsson *et al.*, "Symptomatic atherosclerosis is associated with an altered gut metagenome," (in English), *Nat Commun,* vol. 3, Dec 2012, doi: Artn 1245
10.1038/Ncomms2266.
[104] H. B. Nielsen *et al.*, "Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes," *Nature biotechnology,* vol. 32, no. 8, pp. 822-8, Aug 2014, doi: 10.1038/nbt.2939.
[105] C. Menni *et al.*, "Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes," *Gut microbes,* vol. 11, no. 6, pp. 1632-1642, Nov 1 2020, doi: 10.1080/19490976.2020.1778261.
[106] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *J Stat Softw,* vol. 33, no. 1, pp. 1-22, 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/20808728.
[107] Y. Zhou *et al.*, "Gut Microbiota Offers Universal Biomarkers across Ethnicity in Inflammatory Bowel Disease Diagnosis and Infliximab Response Prediction," *mSystems,* vol. 3, no. 1, Jan-Feb 2018, doi: 10.1128/mSystems.00188-17.
[108] L. Breiman, "Random Forests," *Machine learning,* vol. 45, no. 3, pp. 5-32, 2001, doi: https://doi.org/10.1023/A:1010933404324.

[109] W. Gou *et al.*, "Interpretable Machine Learning Framework Reveals Robust Gut Microbiome Features Associated With Type 2 Diabetes," *Diabetes Care,* vol. 44, no. 2, pp. 358-366, Feb 2021, doi: 10.2337/dc20-1536.
[110] G. L. Ke, Q. Meng, T. Finley, T. Wang, and W. Chen, "LightGBM: a highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems 30 (NIPS 2017),* pp. pp. 3149-3157, 2017.
[111] T. Chen and C. Guestrin, " XGBoost: A scalable tree boosting system," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining- KDD '16,* pp. pp. 785–794, 2016.
[112] A. P. Carrieri *et al.*, "Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences," *Scientific reports,* vol. 11, no. 1, p. 4565, Feb 25 2021, doi: 10.1038/s41598-021-83922-6.
[113] X. W. Wang and Y. Y. Liu, "Comparative study of classifiers for human microbiome data," *Med Microecol,* vol. 4, Jun 2020, doi: 10.1016/j.medmic.2020.100013.
[114] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics,* vol. 5, no. 4, pp. 115--133, 1943.
[115] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature,* vol. 529, no. 7587, pp. 484-9, Jan 28 2016, doi: 10.1038/nature16961.
[116] A. W. Senior *et al.*, "Improved protein structure prediction using potentials from deep learning," *Nature,* vol. 577, no. 7792, pp. 706-710, Jan 2020, doi: 10.1038/s41586-019-1923-7.
[117] C. B. Newgard *et al.*, "A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance," *Cell metabolism,* vol. 9, no. 4, pp. 311-26, Apr 2009, doi: 10.1016/j.cmet.2009.02.002.
[118] P. J. White and C. B. Newgard, "Branched-chain amino acids in disease," *Science,* vol. 363, no. 6427, pp. 582-583, Feb 8 2019, doi: 10.1126/science.aav0558.
[119] D. Chen *et al.*, "Clostridium butyricum, a butyrate-producing probiotic, inhibits intestinal tumor development through modulating Wnt signaling and gut microbiota," *Cancer Lett,* vol. 469, pp. 456-467, Jan 28 2020, doi: 10.1016/j.canlet.2019.11.019.
[120] M. Vital, A. C. Howe, and J. M. Tiedje, "Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data," *mBio,* vol. 5, no. 2, p. e00889, Apr 22 2014, doi: 10.1128/mBio.00889-14.
[121] F. Perraudeau *et al.*, "Improvements to postprandial glucose control in subjects with type 2 diabetes: a multicenter, double blind, randomized placebo-controlled trial of a novel probiotic formulation," *BMJ Open Diabetes Res Care,* vol. 8, no. 1, Jul 2020, doi: 10.1136/bmjdrc-2020-001319.
[122] L. Hoyles *et al.*, "Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women," *Nat Med,* vol. 24, no. 7, pp. 1070-1080, Jul 2018, doi: 10.1038/s41591-018-0061-3.
[123] I. Nemet *et al.*, "A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors," *Cell,* vol. 180, no. 5, pp. 862-877 e22, Mar 5 2020, doi: 10.1016/j.cell.2020.02.016.
[124] J. J. Witjes, L. P. Smits, C. T. Pekmez, A. Prodan, and A. S. Meijnikman, "Donor Fecal Microbiota Transplantation Alters Gut Microbiota and Metabolites in Obese Individuals With Steatohepatitis," *Hepatology communications,* 2020, doi: 10.1002/hep4.1601.
[125] L. J. Marcos-Zambrano *et al.*, "Applications of Machine Learning in Human Microbiome Studies: A Review on Feature Selection, Biomarker Identification, Disease Prediction and Treatment," *Front Microbiol,* vol. 12, p. 634511, 2021, doi: 10.3389/fmicb.2021.634511.
[126] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities," *Inf Fusion,* vol. 50, pp. 71-91, Oct 2019, doi: 10.1016/j.inffus.2018.09.012.
[127] A. Singh *et al.*, "DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays," *Bioinformatics,* vol. 35, no. 17, pp. 3055-3062, Sep 1 2019, doi: 10.1093/bioinformatics/bty1054.
[128] M. Baranwal, A. Magner, P. Elvati, J. Saldinger, A. Violi, and A. O. Hero, "A deep learning architecture for metabolic pathway prediction," *Bioinformatics,* vol. 36, no. 8, pp. 2547-2553, Apr 15 2020, doi: 10.1093/bioinformatics/btz954.
[129] L. P. Coelho, R. Alves, P. Monteiro, J. Huerta-Cepas, A. T. Freitas, and P. Bork, "NG-meta-profiler: fast processing of metagenomes using NGLess, a domain-specific language," *Microbiome,* vol. 7, no. 1, p. 84, Jun 3 2019, doi: 10.1186/s40168-019-0684-8.
[130] M. Tramontano *et al.*, "Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies," *Nat Microbiol,* vol. 3, no. 4, pp. 514-522, Apr 2018, doi: 10.1038/s41564-018-0123-9.

[131] D. Salamone, A. A. Rivellese, and C. Vetrani, "The relationship between gut microbiota, short-chain fatty acids and type 2 diabetes mellitus: the possible role of dietary fibre," *Acta Diabetol,* vol. 58, no. 9, pp. 1131-1138, Sep 2021, doi: 10.1007/s00592-021-01727-5.
[132] M. Kumar, P. Babaei, B. Ji, and J. Nielsen, "Human gut microbiota and healthy aging: Recent developments and future prospective," *Nutr Healthy Aging,* vol. 4, no. 1, pp. 3-16, Oct 27 2016, doi: 10.3233/NHA-150002.
[133] M. Gurung *et al.*, "Role of gut microbiota in type 2 diabetes pathophysiology," *EBioMedicine,* vol. 51, p. 102590, Jan 2020, doi: 10.1016/j.ebiom.2019.11.051.
[134] S. R. Srinivas, P. D. Prasad, N. S. Umapathy, V. Ganapathy, and P. S. Shekhawat, "Transport of butyryl-L-carnitine, a potential prodrug, via the carnitine transporter OCTN2 and the amino acid transporter ATB(0,+)," *Am J Physiol Gastrointest Liver Physiol,* vol. 293, no. 5, pp. G1046-53, Nov 2007, doi: 10.1152/ajpgi.00233.2007.
[135] S. Lucas *et al.*, "Short-chain fatty acids regulate systemic bone mass and protect from pathological bone loss," *Nat Commun,* vol. 9, no. 1, p. 55, Jan 4 2018, doi: 10.1038/s41467-017-02490-4.
[136] P. Li *et al.*, "Metabolic Alterations in Older Women With Low Bone Mineral Density Supplemented With Lactobacillus reuteri," *JBMR Plus,* vol. 5, no. 4, 2021, doi: 10.1002/jbm4.10478.
[137] J. Yan *et al.*, "Gut microbiota induce IGF-1 and promote bone formation and growth," *Proc Natl Acad Sci U S A,* vol. 113, no. 47, pp. E7554-E7563, Nov 22 2016, doi: 10.1073/pnas.1607235113.
[138] M. Derrien, M. C. Collado, K. Ben-Amor, S. Salminen, and W. M. de Vos, "The Mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract," *Appl Environ Microbiol,* vol. 74, no. 5, pp. 1646-8, Mar 2008, doi: 10.1128/AEM.01226-07.
[139] P. Louis and H. J. Flint, "Formation of propionate and butyrate by the human colonic microbiota," *Environ Microbiol,* vol. 19, no. 1, pp. 29-41, Jan 2017, doi: 10.1111/1462-2920.13589.
[140] K. Zhou, "Strategies to promote abundance of Akkermansia muciniphila, an emerging probiotics in the gut, evidence from dietary intervention studies," *J Funct Foods,* vol. 33, pp. 194-201, Jun 2017, doi: 10.1016/j.jff.2017.03.045.