# EVALUATING PREDICTIVE METABOLIC MODELLING TOOLS USING MICROBIOME SEQUENCING-METABOLOMIC DATA

by

Chris Ga-Jing Tang

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
November 2021

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all treaty people.

Dedication

To my friends and family and
everyone who lent a hand along the way.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

**ABSTRACT**

Metabolomic technologies have been utilized increasingly in tandem with sequencing technologies in microbiome studies. Combined microbiome-metabolome approaches have fueled the discovery of many convincing associations between disease-associated metabolites and the microbial taxa and genes driving their variation, and new computational tools to integrate these datasets are just emerging. A subset of these tools can mathematically predict the metabolic activity of the microbiome using the genomic content encoded in the microbiome. If these tools are accurate, they could allow researchers to circumvent the difficult and time-consuming process of performing metabolomics experiments by generating metabolite data using microbiome sequencing data alone, however few groups have validated their predictions against experimentally measured data. I sought to examine the performances of six metabolic modelling tools by calculating the correlations between predicted and experimentally measured metabolite profiles using paired sequencing and metabolomic data taken from the human gut and vaginal microbiomes. Out of all surveyed tools, MelonnPan generated predictions that were correlated best to experimentally measured data, although the observed correlations were generally poor across all tools and varied depending on sample characteristics such as donor disease status and sampling site. I did not observe any metabolites that were robustly predicted across all datasets and tools, and in addition all tools generally performed poorly in identifying differentially abundant metabolites. In this work, I have demonstrated the feasibility of predicting metabolites solely from microbiome sequencing data, while raising important limitations relating to the robustness of these predictions.

## LIST OF ABBREVIATIONS USED

| | |
|---|---|
| ACBM | Agent and Constraint-Based Modeling |
| AGORA | Assembly of gut organisms through reconstruction and analysis |
| ASD | Autism spectrum disorder |
| BV | Bacterial vaginosis |
| CD | Crohn's disease |
| COBRA | Constraint-Based Reconstruction and Analysis |
| CRC | Colorectal cancer |
| ddNTP | Dideoxynucleotide triphosphate |
| FBA | Flux balance analysis |
| GC | Gas chromatography |
| GC-MS | Gas chromatography mass-spectrometry |
| GEM | Genome-scale metabolic model |
| HILIC | Hydrophobic interaction chromatography |
| HMP | Human Microbiome Project |
| HUMAnN | HMP Unified Metabolic Analysis Network |
| IBD | Inflammatory bowel disease |
| LC | Liquid chromatography |
| LC-MS | Liquid chromatography mass-spectrometry |
| MAG | Metagenome-assembled genome |
| MAMBO | Metabolomic Analysis of Metagenomes using fBa and Optimization |
| MATLAB | MATrix LABoratory |
| MelonnPan | Model-based Genomically Informed High-dimensional Predictor of Microbial Community Metabolic Profiles |
| MetaPhlAn | Metagenomic Phylogenetic Analysis |
| MGS | Metagenomic shotgun sequencing |
| MICOM | Microbial community (Python software package) |
| MiMeNet | Microbiome-Metabolome Network |
| MIMOSA | Model-based Integration of Metabolite Observations and Species Abundances |
| MRM | Multiple reaction monitoring |

| | |
|---|---|
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| m/z | Mass-to-charge |
| NMR | Nuclear magnetic resonance |
| PCR | Polymerase chain reaction |
| PRMT | Predicted relative metabolomic turnover |
| PA | Procrustes analysis |
| RAVEN | Reconstruction, Analysis, and Visualization of Metabolic Networks |
| RP | Reverse phase |
| rRNA | Ribosomal RNA |
| SCFA | Short-chain fatty acid |
| SRA | Sequence read archive |
| TMA | Trimethylamine |
| TMAO | Trimethylamine N-oxide |

**ACKNOWLEDGEMENTS**

It has been an incredibly rewarding two years studying as a Master of Science student in the Department of Microbiology & Immunology here at Dalhousie University. Pandemic notwithstanding, the department has always been very supportive of its students and I am thankful to consider myself one of them. To my supervisor Morgan, I want to thank you for the tremendous patience and your constant encouragement of my academic work here in Halifax. It has been a very strange past two years, but now I can't imagine having done my master's anywhere else but here.

To my lab mates, I would not be where I am now without the moral and technical support you have given me. Thank you for all your guidance – I could not wish for a better group of colleagues.

A heartfelt thanks to my committee members for their guidance through these past two years: Dr. Devanand Pinto and Dr. John Rohde. I would also like to thank Dr. John Archibald for taking the time to serve as my external examiner for this work. Additionally, I am grateful to Dr. Andrew Stadnyk and his lab team at the IWK for generously providing me with a workspace to perform my experiments.

Many thanks to the departmental staff, particularly Marsha Scott-Meldrum and Joseph Behl, for keeping me on my toes with respect to program requirements.

Last but not least, I would like to extend an immense thank you to my family and my friends here, back home, and abroad, for your tireless support – and virtual company – over these challenging past two years.

# CHAPTER 1 – INTRODUCTION

Microscopic organisms, or *microbes*, are the oldest and most ubiquitous lifeforms on Earth, occupying a diverse array of habitats from the ocean depths up to the near-surface atmosphere. Microbes have adapted and thrived within these many different ecological niches, carrying out important biogeochemical processes essential towards the cycling of nutrients worldwide (Colwell, 1997; Falkowski, Fenchel, & Delong, 2008). Many trillions of microbes have also evolved to live in association with human bodies, where in exchange for a warm, moist habitat and a steady flow of nutrients they contribute key biological processes essential towards the maintenance of bodily equilibrium (Valdes, Walter, Segal, & Spector, 2018). Collectively, the entire catalogue of living microorganisms and viruses that colonize the human body are known as the *microbiota*, and together with their genomic contents and metabolic activities they make up the human *microbiome*.

Owing to the positive metabolic activities it contributes to the human body, the microbiome has been alternatively described as the "last" or an "additional" organ within the body (Berg et al., 2020). Other efforts to conceptually define the human body as a "super" or "meta"-organism further highlight the growing appreciation of the symbiotic interdependence between host and microbiota within the scientific community. Though estimates of microbial cell counts within the microbiome have been revised downwards and remain considerably uncertain, microbial cells are still estimated to outnumber human cells by a margin of 1.3:1 (Sender, Fuchs, & Milo, 2016). This figure belies the microbiome's substantial contribution to the human body's pool of genetic diversity – within the gut, the most densely colonized organ within the human body, the microbiome is estimated to encode over 100-fold more genes compared to human cells (Qin et al., 2010). The end products of these encoded genes are low molecular weight compounds, or *metabolites*. These metabolites serve important roles in the body as cell-to-cell signalling molecules, growth factors, or exist as fuel for microbial and human cells alike.

Our microbiome can modulate human health outcomes by synthesizing or degrading these metabolites, but how can we determine which microbes or microbial genes are responsible for driving their variation in the human body? Rapid advancements in microbiome and metabolome profiling technologies may bring us closer to answering this

question, although finding ways to interpret these complex data meaningfully is a continuing challenge. On the other hand, mathematical modelling strategies now enable us to infer the composition of the *metabolome* – the entire catalogue of metabolites within a given setting – from the encoded genetic information of the microbiome. Potentially, these tools could allow researchers to circumvent the complicated process of directly measuring metabolites within human-microbiome samples and provide tractable tools to assess the metabolic effects of altering the microbiota. However, few groups have sought to evaluate whether these tools generate results that are comparable to metabolite data obtained through traditional analytical techniques. In this thesis, I describe a systematic meta-analysis performed to gauge the performance of these tools regarding their relatedness to experimentally acquired metabolomics data and whether they might be suitable as proxies for these data.

### 1.1 Microbial metabolism and disease

The human body plays host to a diverse consortium of microbes of varying taxonomic identity and metabolic potential. Dense colonies of microbes are resident to the skin, the lungs, the genitals, and the entire length of the gastrointestinal tract. These microbes do not colonize the body uniformly – the density, taxonomic composition and encoded metabolic processes of the microbiota differ across the many niches of the microbiome, even within the same apparent body site. For instance, bacterial densities along the large intestine – from which most of our understanding of the microbiome has been gleaned – follow an ascending gradient from the proximal to distal colon (Sartor, 2008). Microbes resident to the large intestine are predominantly anaerobic, a reflection of the oxygen-scarce conditions within the gut lumen, and must utilize a variety of degradative pathways and cross-feeding strategies to scavenge carbon and energy from complex polysaccharides that have escaped host digestion machinery. Past studies however have uncovered important differences between luminal and mucosa-associated bacteria, notably that accumulated oxygen within the gut mucosa permits the formation of radially distributed microenvironments in which communities of aerotolerant, non-saccharolytic

bacteria may proliferate near the gut epithelium (Albenberg et al., 2014).

Beyond the topographical variation of the gut landscape, gene catalogue studies of the microbiome have also revealed staggering levels of genetic diversity at the inter-individual level. Findings gleaned from large cohorts such as the MetaHit and the Human Microbiome Project (HMP) cohorts have identified upwards of ten million non-redundant genes encoded by our intestinal microbiota (Qin et al., 2010, Li et al., 2014). A more recent study has revised this figure upwards to over 22 million genes, of which nearly one-half were found in no more than a single participant (Tierney et al., 2019). Other efforts to enumerate the microbial gene catalogue in the oral cavity (Tierney et al., 2019), the vagina (Ma et al., 2020), and the skin (Li et al., 2021) highlight the substantial cache of microbial functional diversity across the human body.

The genetic variation inherent to the microbiome underscores its tremendous capacity to modulate human health by influencing the metabolic landscape of the host (Lee-Sarwar, Lasky-Su, Kelly, Litonjua, & Weiss, 2020). A key question in many microbiome studies relates to the microbiota's potential to impact the health of their host through its metabolic activity. Below, I describe some of the best-characterized examples of microbial metabolites that have strong associations with human health outcomes.

### 1.1.1 Trimethylamine/trimethylamine N-oxide

Trimethylamine N-oxide (TMAO) and its precursor trimethylamine (TMA) are metabolites formed from the conversion of choline, L-carnitine, and lecithin (Janeiro, Ramírez, Milagro, Martínez, & Solas, 2018). Studies in the past decade have uncovered convincing links between the microbiota-dependent metabolism of TMA and increased risks for cardiovascular diseases, providing one of the most compelling examples of how the interplay between host, microbiome, and diet can impact human health through a single microbial metabolite.

Until recently, TMAO/TMA dysregulation had been mostly discussed in relation to trimethylaminuria, an otherwise benign disorder characterized by strong body odours resulting from the accumulation of TMA in the sweat and urine (Fennema, Phillips, &

Shephard, 2016). Dysfunctional TMAO/TMA metabolism begins with the diet, wherein foods such as red meats and fish – naturally abundant sources of TMAO, TMA, and its precursors – are consumed. Colonic bacteria readily convert these substrates to produce TMA which is further oxidized by host enzymes to form TMAO in the liver. Studies in the past decade have implicated TMAO accumulation in the blood plasma as a risk factor for cardiovascular disease, possibly through the inhibition of reverse cholesterol transport by foamy macrophages (Wang et al., 2011; Koeth et al., 2013; Heianza, Ma, Manson, Rexrode, & Qi, 2017).

As conversion of TMA to TMAO is dependent on the microbiota, prevailing hypotheses on how to reduce TMAO levels in the blood have centered around microbiome-targeted or dietary interventions. Several groups have demonstrated that adherence to a Mediterranean diet – which is low in red meat consumption – can reduce TMAO levels in cohort studies (De Filippis et al., 2016; Pignanelli et al., 2018). Reducing choline-dependent TMA formation in the gut through targeted inhibition of gut microbial metabolism with a choline analogue has been shown to reduce plasma TMAO levels and atherosclerotic plaque development in mouse models (Wang et al., 2015). In addition, microbial engraftment with methanogenic archaeal species in mice fed a high choline diet has also been demonstrated to reduce plasma TMAO levels (Ramezani et al., 2018), demonstrating the feasibility of a microbiome-targeted therapy to counteract the effects of dysfunctional TMAO metabolism.

### 1.1.2 Short-chain fatty acids

The principal end-products of anaerobic fermentation within the colon are short-chain fatty acids (SCFAs), which carry out a multiplicity of functions throughout the human body (Tan et al., 2014). Structurally complex and otherwise amylase-resistant carbohydrates (and to a lesser extent proteins) that reach the large intestine are fermented into these simple organic compounds, comprised of a single carboxylic acid group bonded to aliphatic carbon tails 1-6 carbons in length.

The three most abundant SCFAs in the colon – acetate, butyrate, and propionate –

play important roles in the local maintenance of gut homeostasis including but not limited to the maintenance of tight junctions between gut epithelial cells (Peng, Li, Green, Holzman, & Lin, 2009), the promotion of immunotolerance through regulatory T cell proliferation (Smith et al., 2013), and the stimulation of mucus production (Finnie, Dwarakanath, Taylor, & Rhodes, 1995). Additionally, intermediary SCFAs such as formate, succinate, and lactate help promote the maintenance of gut diversity and syntrophy by acting as substrates in cross-feeding interactions which ultimately produce butyrate and propionate as end-products (Blaak et al., 2020). Though the vast majority of SCFAs are consumed locally through cross-feeding with resident microbes or by colonic epithelial cells as energy sources, a small fraction of these SCFAs pass into the portal blood and are dispersed to peripheral tissues (Cummings, Pomare, Branch, Naylor, & Macfarlane, 1987). Butyrate and propionate are rapidly removed from the portal circulation by the liver, with the latter acting as a minor substrate in hepatic gluconeogenesis (Perry et al., 2016). Acetate readily passes the blood-brain barrier into the brain where it functions as an appetite regulator (Frost et al., 2014).

Reflecting their numerous physiological functions, dysregulation of SCFA levels in the gut has been linked to many different human maladies such as neurological disorders (Silva, Bernardi, & Frozza, 2020), obesity and other metabolic disorders (Kim, Yao, & Ju, 2019), and inflammatory bowel disease (IBD; Parada Venegas et al., 2019). Decreased abundances of SCFA-producing bacteria and SCFAs – particularly butyrate – in the stool and mucosa are a reliably reproducible finding in active cases of IBD (Huda-Faujan et al., 2010; Machiels et al., 2014). Although SCFA irrigation in clinical trials has seen mixed success in inducing disease remission for active IBD (Scheppach et al., 1992; Breuer et al., 1997), increasing SCFA production through dietary or probiotic interventions has been proposed as a potential therapeutic (Parada Venegas et al., 2019).

### 1.2 Omics surveys of the microbiome

TMA/TMAO and SCFA metabolism in the gut are just two of the best-understood mechanisms through which the microbiome has been shown to influence human health

outcomes. Recent technological developments have greatly empowered the discovery of many more associations between microbial metabolism and emergent host phenotypes such as human disease. Omics technologies permit the collection of comprehensive, high-dimensional data which encapsulate different levels of biological information – DNA, RNA, proteins, and metabolites (Abram, 2015; Mallick et al., 2017). In particular, an increasing number of studies are relying on the simultaneous profiling of both microbiome sequencing data and metabolite data to identify disease-associated metabolites and the potential microbial determinants – microbial taxa and genes – driving their variation (Noecker, Chiu, McNally, & Borenstein, 2019). To highlight one example: Lloyd-Price et al., utilized this combined omics approach in a large cohort of IBD patients to link depletions in the obligate aerobes *Faecalibacterium prausnitzii* and *Roseburia homini* to reductions in metabolites such as SCFAs and secondary bile acids (Lloyd-Price et al., 2019). This strategy has been similarly applied to study host-microbiota dynamics across various diseases such as coronary heart disease (Feng et al., 2016), bacterial vaginosis (BV; Srinivasan et al., 2015), diabetes (Pedersen et al., 2016), pulmonary disease (Bowerman et al., 2020), and colorectal cancer (CRC; Yachida et al., 2019).

Collectively, these approaches have been described as *top-down* or systems approaches, as they seek to catalogue as many constitutive parts of the microbiome as possible to generate and corroborate broad hypotheses on how these different parts interact with one another. Such studies can help create hypotheses as to how microbes behave *in situ* across different human phenotypes, which in turn may inform the development of novel microbiome-directed therapies to combat human disease (Heinken, Basile, Hertel, Thinnes, & Thiele, 2021). However, as I will discuss later, interpreting these data meaningfully in tandem remains a significant challenge.

### 1.3 Sequencing approaches for the microbiome

As mentioned, a common motivation of these studies is to identify convincing associations between the metabolic potential of the microbiota and specific human phenotypes. To this end, studies of the microbiome will often begin by cataloguing the

taxonomic *composition* – the identities and corresponding abundances of the microbes, and the *functions* – the genes and pathways that may potentially be active, within the resident microbiota.

Historically, microbiome samples were characterized by isolating, culturing, and biochemically characterizing pure isolates on permissive growth media. This approach was constrained by its low throughput, coupled with the fact that only a small fraction of microbes observed in situ could be recovered on standard laboratory media (Staley and Konopka, 1985). Culture-dependent techniques gradually gave way to culture-independent techniques in which microbial communities were profiled through DNA and RNA sequencing, rather than through the recovery of pure isolates (Escobar-Zepeda, Vera-Ponce de León, & Sanchez-Flores, 2015).

In the past two decades, rapid advancements in both sequencing platforms and bioinformatics pipelines have led to a great deal of interest in microbiome research from both the scientific and the public sphere, and a commensurate expansion in publicly available microbiome data (Stephens, Lee, Faghri, Campbell, & Zhai, 2015). Sequencing data is being deposited into the Sequence Read Archive (SRA), the world's largest depository for high-throughput sequencing data, at rates which necessitate novel strategies to cope with the growing storage needs (Kodama, Shumway, & Leinonen, 2012). Next-generation sequencing technologies – sequencing instruments with high throughput capabilities – have fueled much of this expansion by allowing researchers to simultaneously sequence multiple samples rapidly in parallel, while removing the need for tedious and time-consuming cloning experiments with *Escherichia coli* that typified earlier culture-independent approaches. Below, I provide an overview of the sequencing approaches used to probe the microbiome.

### 1.3.1 Amplicon sequencing

The most widely used sequencing approach in microbiome research is amplicon sequencing. Herein, a well-conserved segment of the genome, often referred to as a marker gene, is targeted for polymerase chain reaction (PCR) amplification. These amplicons are

subsequently sequenced to determine the taxonomic makeup of the sequenced microbes and their corresponding abundances. Marker genes are selected on the basis of how universally conserved they are across the taxonomic grouping you wish to study (e.g. across all bacteria), and for their capacity to discriminate between different lineages within that taxonomic grouping (e.g. different species of bacteria) by studying accumulated mutations within that marker gene. The principal marker genes used in amplicon sequencing are the 16S ribosomal RNA (rRNA) gene (which cover archaea and bacteria), the 18S rRNA gene (eukaryotes), and the internal ribosomal spacer (fungi).

16S rRNA gene sequencing is the most common choice for researchers aiming to study the bacterial population within the microbiome. The principles behind 16S rRNA-based phylogeny were pioneered by Carl Woese, who proposed the existence of archaea as a phylogenetically distinct kingdom of life based on studies of sequence homology in the 16S rRNA gene (Woese & Fox, 1977). Since this pivotal paper, 16S rRNA sequencing has been used to identify associations between members of the microbiota and a number of diseases with poorly-understood causes such as obesity (Turnbaugh et al., 2009), Crohn's disease (CD; Willing et al., 2009), and Parkinson's disease (Scheperjans et al., 2015).

The 16S rRNA gene encodes part of the 30S small ribosomal subunit in bacteria and archaea and spans roughly 1600 base pairs in length. Nine 'hypervariable' regions (V1 - V9) are interleaved within the otherwise highly-conserved gene sequence. A basic workflow for 16S rRNA gene sequencing typically begins with the PCR amplification of extracted DNA. Primer pairs designed to match with any two universally conserved regions of the 16S rRNA gene are used to amplify one or more hypervariable regions. Researchers can then sequence the resultant amplicon products, resolve products into individual taxonomic units or *amplicon sequence variants* (ASVs; Callahan et al., 2017) and assign taxonomy to these ASVs by examining their sequence similarities against public databases containing known 16S rRNA gene sequences with matching taxonomic identifiers. Various dedicated databases have been assembled for this purpose with corresponding trade-offs in size and taxonomic resolution (Balvočiūtė & Huson, 2017), such as GreenGenes (DeSantis et al., 2006), SILVA (Pruesse et al., 2007), EzTaxon (Chun et al., 2007), and the Ribosomal Database Project (Cole et al., 2009).

At present, several competing next-generation sequencing technologies are

available for amplicon sequencing (Buermans & den Dunnen, 2014). Most next-generation sequencers build upon the basic "sequencing-by-synthesis" approach laid out by Frederick Sanger, whereby the sequence of the DNA template is determined by detecting which nucleotides are added by DNA polymerase. In Sanger sequencing, chain-terminating dideoxynucleotides (ddNTPs) are selectively incorporated during *in vitro* DNA replication to generate fragments of different intermediary lengths (Sanger, Nicklen, & Coulson, 1977). Fluorescent tagging of ddNTPs enables researchers to iteratively determine the sequence of the DNA template by observing the fluorescent colours of the size-resolved fragments. Sanger sequencing is appropriate only for a single DNA template, requiring a labour-intensive cloning step to generate libraries from samples with mixed communities of microbes, and was gradually supplanted by next-generation technologies for microbiome sequencing projects. Next-generation approaches have altogether eliminated this cloning step using barcoded primers while massively parallelizing sequencing reactions to generate millions of fragments in a single run. The 454 Life Sciences method of pyrosequencing, considered the first next-generation technology to be made commercially available, utilizes beads to partition and immobilize barcoded DNA strands for sequencing in a highly parallel fashion (Margulies et al., 2005). Several initial rounds of solid-phase PCR occur, generating clusters of clonal template DNA ligated to beads. One of four nucleotides are iteratively incorporated, resulting in the release of pyrophosphate and the generation of detectable light in a luciferase-catalyzed reaction. Though capable of generating reads of up to 800 bp, pyrosequencing suffered from high costs of reagents and high error rate and was recently discontinued after production of the instrumentation and its reagents were ceased in 2015 by its parent company Roche (Slatko, Gardner, & Ausubel, 2018).

Today, microbiome sequencing projects are most commonly carried out on Illumina sequencers, which can capture short-reads upwards of 300 bp in length. Much like in pyrosequencing, Illumina sequencing involves several initial rounds of solid-phase PCR (termed "bridge" amplification) to generate local clusters of template DNA, a necessary step to enhance the signal-to-noise ratio. During sequencing, one of four nucleotides bonded to fluorescently labelled chain-terminating tags are incorporated, allowing cameras to iteratively capture the sequences of the growing strands through successive cycles as the

chain terminator is enzymatically cleaved off and synthesis resumes. Sequencing runs take place on flow cells containing multiple physically segregated lanes, allowing dozens of barcoded samples to run in parallel. Other competing technologies include PacBio and Nanopore, which can capture reads upwards of 100 Kbp in length (Pollard, Gurdasani, Mentzer, Porter, & Sandhu, 2018). Presently, while Illumina delivers higher base calling accuracies, greater throughput and lower costs per base compared to both PacBio and Nanopore, gradual improvements to error correction algorithms and throughput for both PacBio and Nanopore sequencers (Amarasinghe et al., 2020; Pollard et al., 2018; Rang, Kloosterman, de Ridder, 2018) make them increasingly attractive options for researchers aiming to sequence the entire length of the 16S rRNA gene.

There are several potential pitfalls that should be considered before embarking on an amplicon sequencing project. First and foremost, 16S rRNA sequences alone provide no information on the functional content (genes and pathways) of the microbiome. Bioinformatic tools such as PICRUSt2 (Douglas et al., 2020), Piphillin (Iwai et al., 2016), and Tax4Fun2 (Wemheuer et al., 2020) can map ASVs obtained through 16S rRNA sequencing to known reference genomes to infer the genes present within a sample. However, the ability of these tools to convincingly detect differentially abundant genes is still under debate (Douglas et al., 2020).

A common complaint of 16S rRNA sequencing relates to the limited taxonomic resolution it confers. Because 16S rRNA sequences from species in the same genus may share high sequence similarities despite substantial genotypic variation outside of the 16S rRNA sequence, obtaining accurate species-level classifications through amplicon sequencing is often not possible (Jovel et al., 2016; Konstantinidis et al., 2006). Some groups have demonstrated that long-read sequencing technologies can potentially achieve strain-level resolution, provided that the entire 16S rRNA gene is appraised (Johnson et al., 2019; Jeong et al., 2021). This unfortunately precludes short-read technologies (e.g. Illumina sequencers), requiring its users to make non-trivial decisions as to which regions to target for amplification. Different regions and primer pairs have been shown to systematically bias community composition owing to uneven primer specificities across different groups of bacteria (Klindworth et al., 2012). Generally, genus-level classifications are much more easily achieved compared to species-level classifications (Wang, Garrity,

Tiedje, & Cole, 2007).

Despite these shortcomings, 16S rRNA sequencing remains one of the most accessible and versatile means to characterize microbiome samples. Altogether, it is important that, for microbiome studies to be compared against one another, researchers should endeavour to standardize sample collection protocols, sequencing platforms, and processing pipelines to minimize the compounding effects of bias across studies (Nearing, Comeau, & Langille, 2021).

### 1.3.2 Metagenomic shotgun sequencing

While amplicon sequencing can be a powerful tool to study microbial diversity, 16S rRNA sequences alone do not provide any meaningful insight on the functional potential of the microbiome. An alternative approach to amplicon sequencing – metagenome sequencing – utilizes untargeted "shotgun" sequencing techniques to capture all the DNA present within an environment. Thus, unlike 16S rRNA sequencing, MGS can survey the functional potential of the microbiome, without the need for additional prediction tools. Correspondingly, metagenomic sequencing projects typically yield greater amounts of data at the cost of being more expensive to run and more computationally difficult to analyze effectively.

Modern metagenome sequencing projects for the microbiome employ the basic shotgun sequencing approach, whereby environmental DNA is extracted and randomly sheared in preparation for sequencing. The very first metagenomic shotgun sequencing (MGS) experiment was performed on marine microbes using fosmids – stably-maintained, low copy number plasmids that can tolerate exceptionally large inserts – as cloning vectors for fragments of environmental DNA in *E. coli* (Stein, Marsh, Wu, Shizuya, & DeLong, 1996). Since then, technological developments in sequencing platforms (covered in 1.3.1) have allowed us to dispense with the cloning steps during library preparation. Much like for amplicon studies, MGS projects most often utilize short-read Illumina sequencers to sequence barcoded reads accurately and rapidly (Quince, Walker, Simpson, Loman, & Segata, 2017). The Illumina HiSeq and NextSeq platforms are favoured over the mid-range MiSeq platform primarily because of the much greater sequencing depths offered – a single

HiSeq run will generate nearly two orders of magnitude more data compared to a MiSeq run (Caporaso et al., 2012). Because of the high proportions of "contaminating" host DNA within some sampling sites in the human body, higher sequencing depths are often necessary to obtain adequate coverage of the microbial DNA fraction in a sample (Pereira-Marques, et al., 2019). Although less common, long-read technologies such as PacBio and Nanopore can be used for MGS – long-read capabilities are well-suited for researchers who intend to assemble DNA fragments into longer sequences (Xie et al., 2020; Latorre-Pérez, Villalba-Bermell, Pascual, & Vilanova, 2020).

Shotgun sequences are more difficult to process compared to amplicon sequences, owing both to their highly fragmentary nature and the much greater volume of data that is generated. After sequencing has concluded, processing pipelines must make sense of hundreds of millions of short DNA fragments of indeterminate origin within a sample. Host DNA, which can comprise upwards of >90% of sequenced reads in certain human sample types (Lloyd-Price et al., 2017), must also be excluded depending on the specific research question being addressed. After quality-control and filtration, sequences can then be taxonomically and functionally characterized.

There are two basic approaches for profiling metagenome sequences: assembly-based metagenomics in which short reads are stitched together into contigs and analyzed as sets of genomes, or read-based metagenomics which forgo the assembly step entirely and characterize sequenced reads independently of one another. Assembly-based metagenomics permit the reconstruction of metagenome-assembled genomes (MAGs), through which groups have gleaned important insights into the metabolism of previously unknown microbes (van Kessel et al., 2015; Daims et al., 2015). However, genome assembly demands heavy computational resources and requires sufficient genome coverage to produce high-quality MAGs (Quince et al., 2017). Researchers who simply wish to catalogue microbial taxa and genes within environments that have already been well-studied (such as the gut) may choose to utilize a read-based approach instead, which is comparably simpler and typically less computationally intensive to execute. MetaPhlAn2 (Truong et al., 2015) is a popular read-based pipeline which estimates taxonomic abundances based on clade-specific markers. MetaPhlAn2 outputs can additionally be piped into HUMAnN2 (Franzosa et al., 2018) to quantify gene and pathway

abundances and enumerate the genic contributions of specific taxa.

The comprehensive data produced through this complex process is especially valuable for researchers aiming to study the microbiome in greater granularity than can be afforded through amplicon sequencing. MGS can reach strain-level resolution with dedicated pipelines such as ConStrains (Luo et al., 2015), StrainPhlAn (Truong, Tett, Pasolli, Huttenhower, & Segata, 2017), and PanPhlAn (Scholz et al., 2016). Additionally, when sequencing depth is sufficiently high, MGS can recover low-abundance taxa more effectively than 16S rRNA sequencing (Durazzi et al., 2021). Achieving strain-resolution is important considering that strains within the same species may differ greatly in their functional gene content and occupy metabolically distinct niches within the microbiome (Rasko et al., 2008; Tenaillon, Skurnik, Picard, & Denamur, 2010).

Functional data retrieved through MGS can provide important insights into how ecosystem-level phenotypes, such as human disease, emerge from the sum of its parts. Disease-linked functions identified through metagenomics can then be used to guide or otherwise corroborate testable hypotheses on microbiota-derived disease mechanisms. To highlight an example, multiple studies have observed dysregulation in L-arginine uptake and biosynthetic pathways in the gut microbiome of individuals with CD (Coburn et al., 2016; Vich Vila et al., 2018). These findings have been further tested through *in vivo* models of colitis in mice receiving diets enriched with L-arginine, the results of which have demonstrated the protective effects of L-arginine supplementation in active colitis (Singh et al., 2019).

Amplicon and metagenomic sequencing give only a partial snapshot of the microbiome. Sequencing data may yield biologically plausible leads as to which taxonomic groups and genes are contributing to an emergent host phenotype. However, because it cannot assay biochemical activity directly, sequencing alone provides very little insight on the potential metabolic outcomes of broad taxonomic or functional shifts in the microbiome (Aguiar-Pulido et al., 2016). In addition, a large fraction of microbial genes identified through MGS remain poorly characterized and cannot yet be functionally annotated (Prakash & Taylor, 2012). Integration of complementary approaches such as single cell genomics or additional biological information through metatranscriptomics, metaproteomics, or metabolomics may help close these gaps.

## 1.4. Metabolomics for the microbiome

*Metabolomics* – defined as the comprehensive surveying of low molecular weight compounds within a biological sample – is a particularly valuable approach for understanding how potential microbial functions might translate to realized microbial activity (Aguilar-Pulido et al., 2016). Because metabolites form the end products of all cellular processes, the composition of the metabolome can be considered as an endpoint readout of the compositional and functional structure of the microbiome (Bernini, Bertini, Luchinat, Nepi, & Saccenti, 2009; Fiehn 2002). For this reason, metabolomics can serve as an excellent complement to sequencing approaches in microbiome studies.

Approaches for studying the metabolome fall into two general categories: targeted and untargeted metabolomics. Targeted metabolomics aim to quantify the abundances of a pre-determined group of chemicals. These chemicals are selected based on *a priori* knowledge regarding their biological activity within the system under study. In contrast, untargeted metabolomics is utilized most often as a hypothesis-generating tool and focuses rather on the global detection of all metabolites in a sample. Global metabolite data collection can greatly empower the discovery of microbially-derived metabolites that may be associated with disease in microbiome studies. However, the volume and complexity of these data generated can vastly prolong data processing and complicate downstream statistical analyses (Schrimpe-Rutledge, Codreanu, Sherrod, & McLean, 2016). Key challenges relate to the identification of chemicals that are underrepresented in chemical databases, and the standardization of instrumentation and informatic tools across labs.

### 1.4.1 Metabolomics instrumentation

Metabolite measurements are most often collected through mass spectrometry (MS) or nuclear-magnetic resonance (NMR) spectroscopy. Both techniques possess highly complementary advantages and shortcomings that have been extensively reviewed elsewhere (Emwas 2015; Marshall & Powers, 2017). In brief, NMR benefits from greater reproducibility, scalability, non-destructive analysis and a simpler sample preparation

protocol – in many cases samples may be analyzed directly without an extraction step – but suffers from limited sensitivity. In contrast, MS is highly sensitive and capable of covering a much wider range of analytes, however considerable debate remains regarding best practices for chemical annotation as mentioned previously. MS has been more widely adopted compared to NMR owing to its lower costs of entry and greater sensitivity to low abundance compounds and thus will be the focus of this next section (Emwas et al., 2019).

### 1.4.2 Generating mass spectrometry data

While MS is a highly versatile technology and offers an overwhelming inventory of instruments and workflows that are highly complementary, in practice some techniques are favoured over others. This following description will focus on the most conventional instrumentation and methods – predominantly LC-MS based technologies – that are used in metabolomic studies of the microbiome.

A typical MS run begins with the physiochemical separation of extracted metabolites on a chromatography column, most commonly liquid chromatography (LC) or gas chromatography (GC). Resolved analytes are fed into a mass spectrometer where they are ionized and deflected within the apparatus to calculate their corresponding mass-to-charge (m/z) ratios and relative intensities.

Metabolites are structurally heterogeneous – no single method can be expected to detect all compounds within a sample. Thus, it is not uncommon for samples to be run more than once using different methods to capture the greatest breadth of metabolites possible. While GC is generally considered the most reproducible separation technique, microbiome studies commonly utilize LC which benefits from a wider chemical coverage and greater sensitivity compared to GC-MS without requiring a chemical derivatization step (Lamichhane, Sen, Dickens, Orešič, & Bertram, 2018; Xu, Zhang, Zheng, Yuan, & Feng, 2019). Complementary LC techniques such as hydrophilic interaction chromatography (HILIC) or reverse phase liquid chromatography (RPLC) can enhance the recovery of hydrophilic and non-polar compounds respectively and are often used in tandem (Lloyd-Price et al., 2019; Poyet et al., 2019). LC separated analytes are most

commonly ionized through electrospray ionization (ESI), in which a high voltage is applied to the solvated molecules to evaporate the solvent and ionize the molecules through charge exchange. ESI can conveniently be operated to detect both positive and negative ions within the same run if desired (Pitt, 2009).

Ionized molecules are redirected into the mass analyzer, through which ions are separated by their m/z while passing through an electromagnetic field or a time-of-flight tube. A broad range of mass analyzers are available with different performance characteristics relating to their mass accuracies, m/z scan ranges, and data acquisition speeds. Hybrid mass analyzers may also combine various aspects of these technologies to achieve different performance characteristics. (Zhou, Xiao, Tuli, & Ressom, 2012; El-Aneed, Cohen, & Banoub, 2009) Some instruments support tandem MS (MS/MS) in which two or more mass analyzers are arranged in series or in time (through ion traps). Herein, precursor ions passing through the first mass analyzer (MS1) are selectively fragmented and detected by a second mass analyzer (MS2) to generate mass spectra for product ions (Griffiths et al., 2010). The additional mass spectra obtained through MS/MS can help resolve chemical species with similar m/z ratios.

Both targeted and untargeted metabolomics share the same instrumentation, albeit high resolution MS (HRMS) techniques such as quadrupole time-of-flight (QTOF) or quadrupole orbitrap (Q-Orbitrap) are better suited for detecting chemical unknowns owing to their higher mass accuracy and ability to acquire MS1 data over a much wider m/z range. In contrast, triple quadrupole MS (QqQ) and quadrupole-ion trap (QIT) MS offer superior sensitivity and specificity when utilized for targeted analyses (Zhou et al., 2012). The choice of MS instrumentation and the data acquisition modes they were operated in will dictate the complexity of the data that is generated from an LC-MS experiment. For instance, data collected under multiple reaction monitoring (MRM) – a targeted approach in which ions with specific m/z spectra are scanned – are less complex compared to untargeted data generated by an HRMS operating on full scan mode.

### 1.4.3 Pre-processing mass spectrometry data

The desired output of an MS experiment is a table of chemical feature abundances – a *feature* being a molecular analyte with a unique m/z ratio and a chromatographic retention time – against sample identifiers. To achieve this, pre-processing of raw MS data is necessary to remove artifacts and contaminants, pick peaks from the mass spectra, correct for drifts in retention time (and to a lesser extent m/z ratios), collapse redundant peaks together, and normalize signal intensities such that features can be readily compared across samples. (Zhou et al., 2012)

Complete pre-processing workflows for targeted and/or untargeted data are available in commercial software such as Compound Discoverer (ThermoFisher) and MassProfiler Professional (Agilent), as well as popular, free alternatives like XCMS Online (Forsberg et al., 2018) and MZMine2 (Pluskal, Castillo, Villar-Briones, Orešič, 2010). While automation may save valuable person-hours, manual curation is still recommended for high-quality datasets. However, as sample sizes increase, the time spent manually curating data and optimizing tool parameters can quickly overwhelm teams without the requisite expertise (Coble & Fraga, 2014). In addition, comparisons of leading pre-processing tools for untargeted data have demonstrated variable levels of concordance between tools, which may present challenges for standardization and reproducibility across the field (Coble & Fraga, 2014; Hao et al., 2018; Hohrenk et al., 2020).

### 1.4.4 Compound identification

Compound identification is a major bottleneck in untargeted LC-MS metabolomics, particularly when identifying compounds that lack chemical standards or are not well-represented in spectral databases (Cui, Lu, & Lee, 2018). This step is extremely important to confer biological meaning and map chemical features to broader pathways (Alonso, Marsal, & Julià, 2015; Blaženović, Kind, Ji, & Fiehn, 2018). The most common approach is to match orthogonal parameters (m/z values, retention times, and fragmentation patterns) obtained from experimentally measured chemical features to those obtained from running

authenticated chemical standards or through spectral libraries, within the window of tolerance defined by the mass accuracy of the mass spectrometer. Searches can be run against vendor-specific or publicly available libraries of compounds such as the Human Metabolome Database (HMDB) (Wishart et al., 2007), BioCyc (Karp et al., 2019), METLIN (Guijas et al., 2018), and the Kyoto Encyclopedia of Genes and Genomes (KEGG). Automated database searches are usually followed by extensive and tedious manual curation to remove spurious hits (Baran & Northern, 2013).

Typically, only a small fraction of detected features can be identified through library searches, defining what has been conceptually described as *dark matter* – or features with unknown structures – in metabolomics (Blaženović et al., 2018; Silva, Dorrestein, & Quinn, 2015; Schrimpe-Rutledge et al., 2016). Difficulties relating to chemical identification may arise from the limited coverage of spectral libraries, (Matsuda, 2014; Schrimpe-Rutledge et al., 2016), the low reproducibility of LC-MS data (Zhou et al., 2012), or the lack of standardization in MS settings and instrumentation which can make database entries unsuitable or highly redundant in certain workflows (Cui et al., 2018; Dunn et al., 2013).

### 1.5 Identifying microbe-metabolite linkages in omics data

The central motivation of paired sequencing-metabolomic microbiome studies is to identify potential interactions between the constitutive parts of the microbiome that can explain different host phenotypes e.g., disease. To achieve this, integration of sequencing and metabolomic data is an indispensable step towards contextualizing how biochemical shifts (measured via metabolomics) within the environment can be induced through changes in community structure and function across the microbiome (measured via sequencing; Mallick et al., 2017).

The simplest and most popular approach is to treat sequencing and metabolite datasets as entirely independent datasets and evaluate for correlations between microbial features and metabolites (Chong & Xia, 2017) in a biologically naïve fashion. This data-driven approach benefits from its straightforwardness and relative ease of use and has

consequently been implemented to varying levels of sophistication. Strikingly, univariate correlation-based analyses have been shown to produce unacceptably high levels of false positive links between metabolites and taxa that are otherwise unrelated to their synthesis and degradation (Noecker et al., 2019). Multivariate correlation-based tools which employ dimensional reduction to alleviate false discovery, such as Procrustes analysis (Gower, 1975) or DIABLO (Singh et al., 2019), are viable alternatives to univariate statistics. However, these methods are still correlation-based and difficult to interpret meaningfully (Worley & Powers, 2013). Network analyses can readily incorporate *a priori* biological knowledge by representing taxa and metabolites as nodes and known interaction pairs as edges (Sung et al., 2017). Altogether, tool development for multi-omics analyses has accelerated rapidly over the past decade, and active research into new tools may elucidate new ways to incorporate biological knowledge. Comprehensive descriptions of these methods have been reviewed extensively in the literature (You et al., 2019, Chong & Xia, 2017; Hooft et al., 2020; Lee-Sarwar et al., 2020). Notably, many of the most popular approaches described in these reviews fail to take full advantage of the biological knowledge available on known gene-metabolite interactions in their working principles.

### 1.6 Predictive metabolic modelling of the microbiome

An alternative set of approaches seek to uncover associations between the microbiome and the metabolome by mathematically modelling these relationships from the ground up using *a priori* knowledge. These approaches have been collectively referred to as *bottom-up* approaches and work by computationally reconstructing patterns that are observed from the top-down e.g., measured through sequencing or metabolomics (Rolfsson & Palsson, 2015). This is done by integrating its constitutive parts e.g., taxa, gene, metabolite abundances, into a flexible, mathematically formulated scaffold for organizing prior knowledge (Garza & Dutilh, 2015). The requisite input for these tools is sequencing data – either taxa or gene abundances – which are then mathematically extrapolated to produce metabolite predictions that are representative of the microbiome's community-wide metabolic potential. Experimentally measured metabolomics data can be incorporated

into these tools either as parameters to improve model performance or as standards to quantify the relative contribution of the microbiome to the overall experimentally measured metabolome.

The metabolite predictions generated by these tools resemble experimentally measured metabolomics data in many ways. They are labelled with metabolite identifiers, the tables are typically high-dimensional, and they are populated with quantitative or semi-quantitative values representing the magnitude of activity associated with that specific metabolite. Some of these tools have in fact been purposefully designed to mimic and/or have been expressly compared against metabolomics data by their authors (Mallick et al., 2019, Garza, van Verk, Huynen, & Dutilh, 2018; Reiman, Layden, & Dai, 2021; Morton et al., 2019). Given the difficulty of obtaining untargeted metabolomics data relative to sequencing data, it is possible that they may be suitable alternatives for researchers who wish to biochemically characterize their microbiome samples but are not able to invest the substantial amount of time necessary to run an entire metabolomics experiment. Below, I describe some of the various approaches used by these metabolic modelling tools.

### 1.7 Flux balance analysis

Flux balance analysis (FBA) is the oldest and most developed of these metabolic modelling approaches. FBA aims to mathematically solve the optimal flow, or *flux*, of substrate through a biochemical network within defined modeling constraints (Orth, Thiele, & Palsson, 2010). These biochemical networks, or *genome-scale metabolic models* (GEMs) contain comprehensive catalogues of all the metabolic reactions taking place within an organism and are typically reconstructed from annotated genomes and other biological information gleaned from literature (Thiele and Palsson, 2010).

FBA calculates the optimal arrangement of metabolite fluxes that maximize a cellular objective within the constraints imposed by reaction stoichiometries, while operating under the steady-state assumption that metabolite concentrations cannot change over time. The cellular objective is typically growth – represented as a biomass equation that converts pre-determined ratios of macronutrients into cellular biomass – but can also

be re-defined by the user to maximize the production of a specific metabolite. Additional constraints can be applied to the system to investigate metabolite flux across different conditions, for example users can use experimentally measured metabolomics data to set upper and lower bounds for nutrient import/export or model the effects of gene knockouts or colonization with probiotic species (O'Brien, Monk, & Palsson, 2015). This makes FBA an extraordinarily flexible and tractable mechanism-driven tool with which to investigate microbe-metabolite interactions *in silico*.

GEM construction is a highly complex process – creation of a single high-quality model prior to the development of automated workflows typically took upwards of a year to complete (Thiele & Palsson, 2010). In, brief, this process begins with a complete, annotated genome of the target organism. The genes are retrieved and used to populate the metabolite space by mapping genes to gene-metabolite databases such as KEGG or HMDB. These metabolites are then assembled into broader pathways of metabolic reactions. Reactions are subsequently annotated with the appropriate stoichiometry and oriented in the correct directions. Models must also include the appropriate exchange reactions representing the organism's ability to shuttle substrates across compartments. A biomass equation must then be defined, which typically involves combing literature or conducting carefully monitored culturing experiments. Orphan and dead-end metabolites arising from unfilled gaps in the metabolic network must be appropriately linked or culled entirely from the model. Model testing and refinement occurs throughout this process, and it is not uncommon for iterative improvements to be made to GEMs after this process has long concluded when newer information is made available (Feist & Palsson, 2008). Additionally, inclusion of metabolomics data for pure isolates grown in defined media can help guide GEM validation by ensuring that predicted metabolite fluxes are within range of experimentally measured values.

Nowadays, tools such as Model SEED (Henry et al., 2010) or CarveMe (Machado, Andrejev, Tramontano, & Patil, 2018) enable the automated reconstruction of GEMs without the need for extensive manual curation. A newly released tool, metaGEM (Zorrilla, Buric, Patil, & Zelezniak, 2021), can reconstruct GEMs using sample-recovered MAGs, allowing researchers to model microbial metabolism within the FBA framework in a highly context-specific fashion. Pre-constructed GEMs may be obtained through dedicated

repositories such as BiGG Models (King et al., 2016) or the Virtual Metabolic Human database (Noronha et al., 2018). The AGORA consortium is a popular collection of over eight hundred human microbiota-specific GEMs that have been manually curated and validated by its authors (Magnúsdóttir et al., 2017). Flux balance analysis simulations of these models are typically carried out on large software suites or platforms such as the RAVEN (Wang et al., 2018) and COBRA toolboxes (Heirendt et al., 2019) which are available on MATLAB, or COBRA's python equivalent COBRApy (Ebrahim, Lerman, Palsson, & Hyduke, 2013). Notably, some platforms allow the comprehensive integration of experimentally measured metabolomics data to improve model performance (Shoaie et al., 2015).

Up until the past decade, FBA had been restricted to estimating the growth of pure cultures or simple communities of two bacteria growing on defined media (Heinken, Basile, Hertel, Thinnes, & Thiele, 2021). Modern tools have now extended the application of FBA towards multispecies communities. This is typically accomplished by integrating multiple GEMs together into a single community model comprised of individually compartmentalized organisms all sharing a single extracellular environment where nutrient exchange can occur (Heinken et al., 2021). Tools such as OptCom (Zomorrodi & Maranas, 2012), MICOM (Diener, Gibbons, & Resendis-Antonio, 2020), or the Microbiome Modelling Toolbox extension for COBRA (Baldini et al., 2019) utilize this community model approach. This approach necessitates strategies to apportion growth fairly across the community to avoid biologically implausible predictions in which growth is limited to a small subpopulation of fast-growing organisms (Diener et al., 2020). Another sophisticated multispecies approach eschews community models by modelling microbes spatiotemporally. Microbial metabolism is simulated on a two-dimensional grid that attempts to capture the spatial and temporal dynamics driving community structure. Tools utilizing this approach include BacArena (Bauer, Zimmerman, Baldini, Thiele, & Kaleta, 2017) and ACBM (Karimian & Motamedian, 2020).

MAMBO (Garza et al., 2018) utilizes a reverse-engineering approach to directly reconstruct the composition of the metabolome. Given a table of taxonomic abundances and associated GEMs for each taxon, MAMBO iteratively attempts to find an optimal set of metabolite import/export rates which yield microbial growth rates that best correlate

with observed taxonomic abundances. These import/export rates are then taken as proxies for metabolite abundances. MAMBO is the only FBA tool at present which purports to estimate metabolite abundances, rather than metabolite fluxes, and by extension cannot be used to estimate microbial growth or metabolite production. The authors were able to validate their metabolite predictions of oral, skin, stool, and vaginal samples using MS and NMR data (Garza et al., 2018), and additionally were able to apply MAMBO to predict metabolites that were important for the growth of CRC-associated taxa (Garza et al., 2020).

FBA-based metabolic modelling offers highly tractable, mechanism-driven tools for the study of microbe-metabolite interactions. Key challenges remain in the need for well-curated, experimentally validated GEMs and tools, and in the assumption of steady-state conditions which may not hold true in complex environments such as the gut microbiome (Chen, Simons, & Maranas, 2017). Validation of FBA predictions can be performed by evaluating the concordance between predicted metabolite fluxes and experimentally measured metabolomics data, as many groups have done (Hale et al., 2018; Garza et al., 2018; Kumar et al., 2018; Shoaie et al., 2015; Mardinoglu et al., 2015). However, incongruencies between measured and predicted metabolite data may not be easy to interpret as it is not always clear whether these differences arise from insufficient reference information e.g. a GEM with missing pathways, or from unaccounted-for mechanisms regulating metabolite variation e.g. metabolite uptake by human cells (Mallick et al., 2019). Future developments in the field of FBA may help elucidate answers to these challenges.

### 1.8 Gene abundance-based approaches

Gene abundance-based approaches for metabolic modelling aim to mechanistically estimate the metabolic potential of the microbiome by relating the abundance of annotated enzyme functions to the relative ability of the microbiome to synthesize or degrade metabolites. These tools are broadly based off the predicted relative metabolomic turnover (PRMT) methodology laid out by Peter Larsen (Larsen et al., 2011), whereby genes identified through metagenome sequencing are used to generate a network of metabolic

reactions and populate the metabolite space. This approach is conceptually similar to FBA, with the chief differences being that this network is formulated as a bag-of-genes and not compartmentalized by organism, and that PRMT assumes that the turnover rate of a metabolite is proportional to the number of genes encoding for its synthesis or degradation. Notably, PRMT provides only an estimate of the turnover rate of a metabolite relative to a reference sample and does not output metabolite fluxes or relative abundances. Using the PRMT methodology, Larsen et al., predicted the metabolic turnover within a coastal marine microbial community and their estimates correlated with experimentally measured metabolite abundances (Larsen et al., 2011)

To date, three tools within this category have been described: the original PRMT method (Larsen et al., 2011), MIMOSA (Noecker et al., 2016), and its successor MIMOSA2 (Noecker, Eng, & Borenstein, 2021). MIMOSA is implemented in R and utilizes the PRMT approach to identify which microbial taxa or genes are contributing to metabolite variation and has been applied to characterize metabolic interactions in the gut microbiota of mice (Sharon et al., 2019) and children with autism spectrum disorder (ASD; Snijders et al., 2016). Its successor MIMOSA2 functions similarly but greatly expands its list of compatible data types and utilizes a different statistical heuristic to evaluate data for relationships between sequencing features and metabolites (Noecker et al., 2021). Neither MIMOSA nor MIMOSA2 were designed solely to predict metabolite data solely from sequencing features. However, they are capable of generating predictions, with MIMOSA having previously been benchmarked against other metabolite modeling tools that have been expressly designed to do so (Yin et al., 2020).

The limitations associated with gene abundance-based approaches are largely shared with those utilizing FBA. Namely, they rely on complete and well-annotated genomes to build comprehensive metabolic networks. These networks are not expected to be as comprehensive as those offered in GEMs, as there is no gap-filling step to remedy orphan and dead-end metabolites and no ability to incorporate reversible reactions. In addition, the outputs generated by MIMOSA and MIMOSA2 have not endorsed by their authors as reliable alternatives for experimentally measured metabolite data. Regardless, gene abundance-based approaches are intuitive and relatively easy to implement despite its narrower field of applicability.

## 1.9 Machine learning approaches

The application of machine learning to predictive metabolic modelling is a relatively recent development. These tools aim to capture microbe-metabolite linkages by identifying and extrapolating patterns across large quantities of high dimensional data (Liebal, Phan, Sudhakar, Raman, & Blank, 2020). Machine learning metabolic modelling operates in a predominantly data-driven fashion, thus these tools do not utilize known biological information as the above approaches do. In a typical workflow, model training is first performed on samples containing paired sequencing-metabolome data, and then the trained model is applied to an external set of samples with similar taxonomic features or genes to predict metabolite relative abundances.

Three tools utilizing this general approach have been released: MelonnPan (Mallick et al., 2019), MiMeNet (Reiman et al., 2021), and mmvec (Morton et al., 2019). MelonnPan is implemented in R and applies an elastic net regularization strategy to train a model that can predict metabolic composition based on a small number of predictive features (i.e. taxa or genes). Well-predicted metabolites are identified in the training set, and the relative capacity of each predictive feature to associate with that metabolite linearly is determined. This model can then be used to predict the metabolite composition in samples for which there are only sequencing data available. The authors demonstrated the applicability of this approach beyond the human gut microbiome by successfully inferring metabolomes for vaginal microbiome samples, murine gut samples, and samples of coral-associated microbial communities (Mallick et al., 2019).

MiMeNet is implemented in python and builds on the same machine learning principles as MelonnPan. Instead of modelling metabolites linearly, MiMeNet utilizes multivariate learning to train a neural network to associate all metabolite and taxonomic abundances together rather than selecting specifically for well-predicted metabolites as MelonnPan does. Furthermore, the authors directly benchmarked MiMeNet against MelonnPan in stool, lung sputum, and soil water samples to demonstrate that their tool outperforms MelonnPan across a variety of metrics (Reiman et al., 2021).

mmvec is implemented in python and runs through the conda environment.

Similarly to MiMeNet, it utilizes neural networks to model metabolite abundances through taxonomic features. Unlike MiMeNet or MelonnPan however, mmvec does not provide its users with an output of metabolite relative abundances and is primarily used to examine paired sequencing-metabolomics data for microbe-metabolite associations (Morton et al., 2019).

Machine learning methods typically require high-quality training datasets for optimal performance, which may not always be available for environments that are not well characterized. In addition because these tools are biologically naïve, outputted results are not always easy to interpret. For instance, a strong positive association between a metabolite and a predictor (whether it be a microbe or a gene) might imply a direct mechanistic link between the two, but the conclusion reached will vary depending on whether that taxa/gene is known to be involved in the synthesis or degradation of that metabolite. Furthermore, the excitement of utilizing these sophisticated tools may lead to instances where they are applied uncritically and outside of their intended context. For instance, the prepackaged MelonnPan model was initially trained for human gut samples and is accompanied with a disclaimer that learned models are environment specific (Mallick et al., 2019). However, this same prepackaged model has been applied to predict metabolites in agricultural samples (Villarreal-Soto et al., 2020) without targeted validation of predicted metabolite abundances.

**Table 1. Summary of tools surveyed in meta-analysis.**

| Approach | Tool | Description | Input | Output |
|---|---|---|---|---|
| Machine learning | MelonnPan (Mallick et al., 2019) | An R-encoded machine learning tool that directly predicts metabolite abundances when given metagenomic feature abundances. | Metagenomic gene abundances<br>Model trained on paired metagenomic-metabolite data | Metabolite relative abundances |
| Gene-abundance | PRMT (Larsen et al., 2011) | A general method for estimating metabolite turnover using the abundances of annotated genes. Implemented by authors through Perl. | Metagenomic gene abundances | Scores estimating metabolite consumption/production |
| | MIMOSA2 (Noecker et al., 2021) | An R-encoded tool that adapts PRMT to connect the metabolic potential of a microbial community to the observed composition of the metabolome | Metagenomic gene abundances | Scores estimating metabolite consumption/production |
| FBA | MAMBO (Garza et al., 2018) | A Python-encoded FBA tool. Stochastically predicts metabolite abundances that can support microbial growth rates which best correlate with their abundances. | Microbial abundances<br>Genome-scale metabolic models | Metabolite relative abundances |
| | Microbiome Modelling Toolbox – COBRA (Baldini et al., 2019) | An extension to the popular MATLAB toolbox COBRA. Performs FBA at the community level by scaling reaction stoichiometries with microbial abundances. | Microbial abundances<br>Genome-scale metabolic models | Metabolite fluxes |
| | MICOM (Diener et al., 2020) | A Python-encoded FBA tool for modelling metabolism at the community level. | Microbial abundances<br>Genome-scale metabolic models | Metabolite fluxes |

## 1.10 Objectives

A subset of metabolic modelling tools now give researchers the ability to extrapolate the genomic content of the microbiome to produce metabolite data. In some cases, the authors have made direct comparisons of tool outputs against experimentally measured metabolomics data (Mallick et al., 2019; Reiman et al., 2021; Garza et al., 2018), however few groups have sought to compare or benchmark these tools in a systematic fashion. A recent study comparing the machine learning tool MelonnPan against MIMOSA and an in-house metabolite prediction tool demonstrated that these tools generally correlate poorly with metabolomics data (Yin et al., 2020) but made no inclusion of FBA tools in their analysis or microbiome data outside of the gut environment. The enthusiasm for applying these new bioinformatic tools may drum out important technical considerations

affecting their performances.

The main objective of this thesis project was to evaluate the performance of these different metabolic modelling approaches in their ability to predict metabolite data from sequencing data. Specifically, I sought to utilize samples containing paired sequencing-metabolomics data to test how close predicted metabolite profiles were to experimentally measured metabolite profiles in six tools: the COBRA Microbiome Modelling Toolbox (hereafter referred to simply as COBRA), MICOM, MAMBO, PRMT, MIMOSA2, and MelonnPan (Table 1). The first aim was to evaluate their performances using a simple univariate correlation-based approach and to determine whether some characteristics of the sample – namely read depth or sampling site – could affect those correlations. The second aim was to evaluate them based on their ability to recapture differential abundance patterns in metabolites. The third aim was to use multivariate statistical methods to re-examine the congruency between predicted and measured metabolomics data, and to examine inter-tool relatedness.

# CHAPTER 2 – MATERIALS AND METHODS

## 2.1 Data acquisition and pre-processing

We obtained six multi-omic datasets by performing a search on NCBI for studies relating to gut or human vaginal microbiome samples (Table 2). For inclusion into this meta-analysis, studies had to 1) include at least thirty samples, 2) contain sample-linked 16S rRNA gene or MGS sequencing data and metabolomics data, 3) contain human samples, 4) have quantitative measures for at least 50 metabolites. For the purpose of this study, I also generated an in-house metabolomics dataset from samples obtained from a cohort of study participants undergoing nutritional therapy for CD for which MGS data had previously been obtained (Jones et al., 2020). Altogether, six out of eight total datasets were sampled from the gut microbiome (N = 919), with the remaining two from the vaginal microbiome (N = 99). Sequencing data were obtained from the authors or downloaded directly from the NCBI SRA. Metabolomics data were obtained from the authors or downloaded directly from tables available in the articles. I harmonized chemical names across studies using RefMet's online "name-to-refmet" conversion tool supplemented with NCBI's online catalogue and our own in-house library of compound identifiers (Fahy and Subramaniam, 2020). Chemical feature identification across all studies had been accomplished by matching experimentally obtained parameters (m/z ratios, retention times, MS2 fragmentation data) to reference standards (level 1 annotation), parameters obtained through literature search (level 2 annotation), and through database matches (level 3 annotation). I removed all chemical features that were measured but had not been chemically identified in all metabolomics datasets.

## 2.2 Sequencing data processing

We re-processed all 16S rRNA gene and MGS sequencing data to ensure uniformity across all eight datasets. All processing was performed on a dedicated server

using the Microbiome Helper web resource (Comeau et al., 2017; https://github.com/LangilleLab/microbiome_helper/wiki/Amplicon-SOP-v2, https://github.com/LangilleLab/microbiome_helper/wiki/Metagenomics-standard-operating-procedure-v2). In brief, for MGS sequencing data I stitched together forward and reverse reads together and ran kneaddata (v0.7.2) to remove low-quality sequences (PHRED <20 within a sliding window of 4 bases) and other sequences of non-microbial origin. Filtered reads were taxonomically and functionally profiled using HUMAnN2 (v0.11.1). MGS-identified taxa, together with KEGG ortholog (KO) or enzyme commission (EC) number abundances obtained by remapping outputted UniRef90 gene families using the `humann2_regroup_table` script, were subsequently used as inputs for all metabolic modeling tools.

16S rRNA gene sequencing data were collected either through Illumina sequencing or 454 pyrosequencing platforms. I trimmed single-ended pyrosequences using cutadapt (v2.10) and subsequently filtered using trimmomatic (v0.39.0; Bolger et al., 2014) to remove contaminants, sequencing artifacts, and low-quality sequences (PHRED <20 within a sliding window of 4 bases). I imported processed pyrosequences into a QIIME2 environment (v2020.11; Bolyen et al., 2019) and denoised the reads into ASVs using the QIIME2 DADA2 plugin (Callahan et al., 2016). Similarly for 16S rRNA Illumina sequences, I imported sequences and stitched together paired-end reads using the VSEARCH plugin (Rognes et al., 2016). Joined sequences were then quality-filtered using the `q2-quality-filter q-score` script using the default settings and subsequently denoised into ASVs using the DADA2 plugin.

All steps following retrieval of ASVs from sequencing data were applied identically across both pyrosequenced or Illumina sequenced 16S rRNA gene sequences. I downloaded the full-length 16S/18S rRNA SILVA classifier from the QIIME2 website and utilized the `qiime feature-classifier classify-sklearn` script to assign taxonomy to ASVs. Rare taxa with frequencies of <0.1% mean sample depth or those or non-bacterial origin were removed. Lastly, I functionally profiled filtered taxa using PICRUSt2 to retrieve KOs and EC numbers.

## 2.3 Metabolite extractions

Stool samples were partially thawed on ice, with approximately 200mg transferred to 2mL Eppendorf tubes using a sterile spatula. Tubes were weighed and samples hydrated in three-fold 85:15 >90% HPLC-grade ethanol:20mM phosphate buffer in reverse osmosis water, then vortexed (3000RPM, 0°C) at for 3min. I reduced the speed to 500RPM and allowed the samples to shake for an additional 30min. Samples were then sonicated (Cole Parmer Ultrasonic Homogenizer 4710, CP10) at 50W on ice for 7min (1min followed by 30s rest repeated five times). I centrifuged stool homogenates (800g, 0°C) for 10min to remove insoluble debris. Samples with insufficient supernatant volumes of <200uL were re-centrifuged for an additional 10min (16900g, 0°C) to compress pellets further. Supernatants were transferred to pre-chilled 2mL cryotubes and centrifuged again (16900g, 0°C) for ten minutes to remove any remaining debris, then stored at -80°C until further processing.

## 2.4 LC-MS metabolomics analysis

Sample analysis was performed at the Dalhousie University Biological Mass Spectrometry Core Facility. Sample separation was carried out using an Agilent 1290 UPLC system equipped with a Waters Cortecs T3 RP column, (2.7 µm, 2.1 x 50 mm). 10uL of sample was injected and gradient eluted (solvent A: water + 0.1% formic acid, solvent B: acetonitrile + 0.1% formic acid, gradient from 1% to 98% solvent B over 6.5 min) at a flow rate of 400 uL/min. MS analyses were carried out on a Qtrap 5500 (Sciex) operating in positive and negative ESI MRM mode, targeting 92 positive and 80 negative mode transitions. All samples were run in duplicate with separate runs for positive and negative mode ESI. Drifts in peak intensities were monitored and normalized against peak intensities measured in pooled samples. Data were acquired using the Analyst (v1.6.2) software. Raw data pre-processing and peak area integration were carried out on Skyline (v20.2).

## 2.5 COBRA – Microbiome Modelling Toolbox

We generated sample-specific community models using the automated `mgPipe` script provided by the Microbiome Modelling Toolbox extension within the MATLAB COBRA Toolbox (v3.1; Baldini et al., 2019). Each community model contains abundance profiles of all constituent organism-specific models obtained sequencing. Community biomass equations are constrained by scaling each individual model's biomass equation to their relative abundance to ensure that represented organisms grow at rates that are proportional to their relative abundances. I name-matched taxonomic identities and abundance profiles to GEMs sourced from the AGORA consortium (v1.02; Magnúsdóttir et al., 2017) to build community models (Appendix Figure 9). To maximize the proportion of sequences that were represented by the strain-specific GEMs used to build the community models, I constructed pan-species and pan-genus GEMs by collapsing strain-specific AGORA models with the `createPanModels.m` script. Taxa that were mapped to the same species or genus-level model were summed together by their relative abundances, and any unmapped taxa were discarded. I constrained all models by the "average European diet" made available through the toolbox and took maximal outputted fluxes as final predicted metabolite profiles after converting BiGG compound IDs to chemical names. Metabolite profiles predicted using species- or genus-level models were annotated with "_spe" or "_gen" respectively. FBA calculations were performed on MATLAB 2019a (MathWorks.) paired with the IBM CPLEX Optimizer (v12.90; IBM) as a solver for all FBA calculations.

## 2.6 MICOM

We generated sample-specific community models using the QIIME2 MICOM plugin (v0.10). Similarly to COBRA, community models contain a catalogue of individual GEMs together with their relative abundances as obtained through 16S rRNA or MGS sequencing. As recommended by the authors, I opted to constrain the community growth rate to half of its maximal rate in order to redistribute growth more evenly across the

constituent organisms represented in the community model. I imported taxonomic identifiers and abundance profiles as QIIME2 artifacts to match taxa to pan-species and pan-genus level GEMs which were provided by the authors (Appendix Figure 9). Models were additionally constrained by an "average European diet" that was also author provided. I calculated metabolite fluxes using the `qiime micom grow` function and retrieved net metabolite fluxes as final predicted metabolite profiles after converting BiGG compound IDs to chemical names. Metabolite profiles predicted using species- or genus-level models were annotated with "_spe" or "_gen" respectively. Flux balance calculations were performed using the IBM CPLEX Optimizer python API.

## 2.7 MAMBO

We predicted sample-specific metabolomes using MAMBO (vJul4 2019) running in tandem with the COBRApy (v0.4.1) platform in a python environment (v2.7.15). Gurobi (v9.0.2 ) python API was utilized as the solver for all FBA calculations. Taxa identified through 16S rRNA gene or MGS sequencing were name-matched to the strain-specific GEMs provided by the authors (Appendix Figure 9). As 16S rRNA sequencing data lacks the resolution necessary to classify reads down to the strain level, I mapped quality-filtered reads directly to the HMP reference genomes (https://www.hmpdacc.org/HMREFG/) which the authors used to build the strain-specific GEMs using Bowtie 2 (v1.1.2) in the `--very-sensitive` mode. Reads mapping to the same strain were summed together, and those that matched to genomes for which there was no equivalent GEM were discarded. I ran MAMBO with 100,000 Markov chain optimization steps, rather than the author-recommended one-million steps, as I observed frequent crashing the longer the tool was allowed to run. After changing SEED compound IDs to chemical names, I filtered the 100,000 total predictions generated for each sample to retrieve the top 10,000 predicted metabolite profiles with the greatest internal Pearson correlation coefficients. Those 10,000 metabolite profiles were averaged to generate a final predicted metabolite profile for each sample.

## 2.8 MelonnPan

We trained MelonnPan (v.0.99.0) models on sample metabolite data paired with KOs obtained through sequencing data. As recommended by the authors, I retained only the features that were at least 0.01% abundant and 10% prevalent across all samples per study and removed low-variance features. To mitigate the effects of overfitting, I performed a leave-one-study-out cross validation in which I trained MelonnPan models on gene-metabolite data aggregated from seven studies, then utilized the subsequent trained model to predict metabolite abundanecs on the left-out study. This process was repeated eight times in total. I trained MelonnPan models using the `melonnpan.train` function on the RStudio (v3.6.3) platform with 10-fold internal cross validation. I utilized the `melonpan.predict` function to predict metabolite abundances on left-out studies as final predicted metabolite profiles.

## 2.9 MIMOSA2

We converted metabolite names to their equivalent KEGG compound IDs using the metabolite ID conversion tool available through MetaboAnalyst (https://www.metaboanalyst.ca/MetaboAnalyst/upload/ConvertView.xhtml) supplemented with our own in-house library of compound identifiers. To standardize metabolite abundances across studies prior to inputting these data into MIMOSA2, I log- and Z-transformed their abundances and discarded metabolites that did not have an equivalent KEGG compound ID. Sample KOs and transformed metabolite abundances were input into the web application hosted by the authors of MIMOSA2 (http://elbo-spice.cs.tau.ac.il/shiny/MIMOSA2shiny/) to generate community metabolic potential scores which I took as final predicted metabolite profiles.

## 2.10 PRMT

Using the Perl scripts provided by the authors of PRMT (Larsen et al., 2011), I calculated PRMT scores using EC number abundances obtained through sequencing data. PRMT scores provide estimates for the turnover of metabolites relative to other samples being analyzed within the same batch, and I then took these scores as final predicted metabolite profiles. Author-provided scripts were run in a Perl environment (v5.22.1).

## 2.11 Spearman's correlation analysis

We assessed tool performance on a sample-by-sample basis by calculating Spearman's correlations coefficients between predicted metabolite profiles and experimentally measured metabolite profiles on a sample-by-sample basis. For metabolites occurring within a sample's metabolite profiles were matched to those identified in the sample's measured metabolite profile. Unmatched metabolites were discarded from both predicted and measured metabolite profiles to avoid penalizing metabolic modelling tools for omitting metabolites or predicting metabolites that had not been experimentally measured. I additionally calculated Spearman's correlation coefficients between the predicted metabolite abundances and measured metabolite abundances across all samples in each dataset to retrieve an estimate of tool performance on a metabolite-by-metabolite basis. I restricted these calculations to a small subset of "core" metabolites within each dataset, which I defined as a metabolite that had been measured in all samples within a dataset and had been predicted by at least seven of the eight tool configurations tested. Metabolite classes for core metabolites were assigned using RefMet's name conversion feature (https://www.metabolomicsworkbench.org/databases/refmet/index.php). Data were analyzed on the RStudio platform.

## 2.12 Procrustes correlation analysis

We utilized Procrustes Analysis (PA) to examine relatedness between metabolite predictions generated by the surveyed tools. I Hellinger transformed (Legendre and Gallagher 2001) metabolite abundances and removed zero variance metabolites from all metabolite profiles prior to calculating Euclidean distances between tools. I calculated PA correlations between tools using the R package "vegan" (v2.5-7). Procrustes correlations were also calculated between predicted metabolite profiles and measured metabolite profiles as a metric of tool performance. Significance tests on Procrustes correlations were performed using the `protest` function with 9999 permutations.

## 2.13 Differential abundance analysis

We compared overlaps in differentially abundant metabolites between predicted and measured metabolite profiles in case-control datasets (Table 2). Within each sample, metabolites were subset to only those that were shared between the predicted and the measured metabolite profiles. Differentially abundant metabolites were then identified between case and control samples by running Student's t-tests on log-normalized, mean-centered metabolite data using the R package MetaboAnalystR (v3.0.3). I evaluated overlaps in differentially abundant metabolites by appraising tools for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for all eight datasets. TPs, TNs, FPs, and FNs were then aggregated across datasets and a single F1, precision, and recall score for each tool was calculated.

# CHAPTER 3 – UNIVARIATE CORRELATION-BASED ANALYSES OF METABOLIC MODELLING TOOLS

In this study, I evaluated the performances of six metabolic modelling tools (eight tool configurations in total) across eight datasets containing paired sequencing and metabolite data. I sought to determine using univariate statistics how closely metabolite predictions generated by each tool correlated to experimentally measured metabolomics data within two different sites of the human microbiome: the gut (N = 919) and the vagina (99).

## Table 2. Studies containing paired human sequencing-metabolomics data included in meta-analysis.

| Article | Sample type | MS Platform | Detection strategy | Sequencing platform | Region targeted | Study type | Disease | Sample size | Cases/controls | Chemicals identified | SRA ID/Contact |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jones et al. 2020 (in-house cohort) | Stool | QTRAP5500 | Targeted LC-MS | NextSeq 550 | Whole genome | Longitudinal | CD | 140 | NA | 103 | j.e.vanlimbergen @amsterdamumc.nl |
| Poyet et al. 2019 | Stool | Q Exactive orbitrap | Untargeted LC-MS | HiSeq 2000 | Whole genome | Cross-sectional | N/A | 146 | NA | 458 | PRJNA544527 |
| Franzosa et al. 2019 | Stool | Q Exactive Exactive Plus orbitrap | Untargeted LC-MS | HiSeq 2500 | Whole genome | Longitudinal | IBD | 220 | 56/164 | 336 | PRJNA400072 |
| Lloyd-Price et al. 2019 | Stool | Q Exactive Exactive Plus orbitrap | Untargeted LC-MS | HiSeq 2000/2500 | Whole genome | Case-control | IBD | 286 | 105/181 | 549 | PRJNA398089 |
| Hale et al. 2018 | Intestinal biopsy | TSQ Quantum Ultra | Untargeted LC-MS | MiSeq | V3-V5 16S | Case-control | CRC | 83 | 55/28 | 624 | PRJNA284355 |
| Kang et al. 2018 | Stool | Varian Direct Drive | NMR | 454 GS FLX-Titanium | V2-V3 16S | Case-control | ASD | 44 | 21/23 | 59 | Dr.Rosy@asu.edu |
| Srinivasan et al. 2012 | Vaginal swab | QTRAP5500 | Untargeted LC-MS | 454 GS FLX-Titanium | V3-V4 16S | Case-control | BV | 30 | 8/19 | 279 | SRA051298 |
| Srinivasan et al. 2015 | Vaginal swab | QTRAP5500 | Targeted LC-MS | 454 GS FLX-Titanium | V3-V4 16S | Case-control | BV | 69 | 40/29 | 101 | SRP056030 |

### 3.1 MelonnPan outperforms other tools in predicting the presence of experimentally measured metabolites

We ran surveyed tools using the authors' recommended settings inputted with taxonomic or gene abundance data obtained from six gut and two vaginal microbiome studies (Table 2). Four of the six gut microbiome datasets contained 16S rRNA gene sequencing data with the remainder containing MGS sequences. Both vaginal microbiome datasets contained 16S rRNA gene sequences. Because I had run flux balance tools COBRA and MICOM using GSMMs aggregated at either the species (*_spe) and genus (*_gen) levels, I evaluated eight tools/tool configurations in total. To study the intersects between metabolites predicted to be present across tools, I examined shared and tool-specific metabolites across all datasets (Figure 1a, Table 3). MelonnPan predicted the greatest number of metabolites that intersected with those that were experimentally measured (336 out of 336 total) (Figure 1). 191 of these 336 overlapping metabolites predicted by MelonnPan were not predicted by any of the other tools surveyed. PRMT predicted the second greatest number of experimentally measured metabolites (255 out of 1867 total), of which 47 were not predicted by any other tool. In addition, many metabolites predicted by PRMT (1522 out of 1867 total) were uniquely predicted by that tool. The majority of directly measured metabolites (969 out of 1473) were not successfully predicted by any of the tools surveyed. In addition, MelonnPan captured the greatest mean proportion of metabolites covered in both the gut (mean: 0.3710 ± 0.0882 SD) and the vaginal microbiome (mean: 0.5470 ± 0.1239) (Figure 1b).

**Table 3. Intersects between metabolites predicted by tools and metabolites measured within dataset**

| Tool | Total identified | Uniquely identified | Intersects w/ Measured Dataset | Intersects Uniquely w/ Measured Dataset |
|---|---|---|---|---|
| Metabolite | 1473 | 969 | NA | NA |
| COBRA | 422 | 20 | 161 | 0 |
| MICOM | 752 | 392 | 149 | 0 |
| MAMBO | 335 | 42 | 152 | 2 |
| MIMOSA2 | 192 | 0 | 192 | 17 |
| PRMT | 1867 | 1522 | 255 | 47 |
| MelonnPan | 336 | 0 | 336 | 191 |

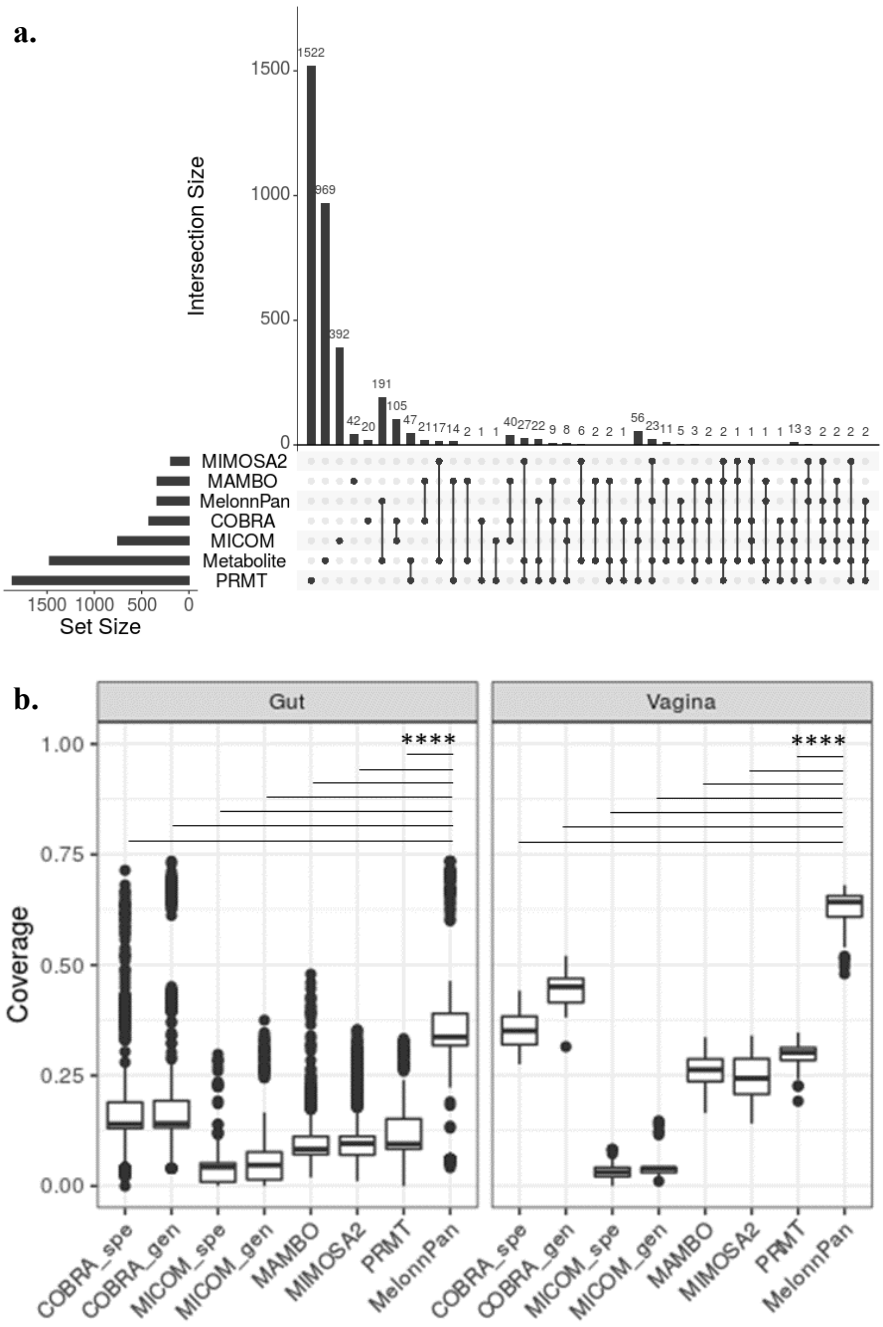



**Figure 1. MelonnPan predicts the greatest number of experimentally measured metabolites.** Overlap in metabolites were evaluated via an a. UpSet plot visualizing the intersections in metabolite occurrence across tools and experimentally measured metabolite data aggregated across all studies, and a b. box plot illustrating proportions of experimentally measured metabolites predicted by tools across the gut and vaginal microbiomes Kruskall-Wallis test followed by Dunn's post-hoc test with FDR correction. *:$p < 0.05$, **:$p < 0.005$, ***:$p < 0.0005$, ****:$p < 0.00005$

### 3.2 Correlations between tool predictions and experimentally measured metabolite data vary across body sites and by disease status

We sought to evaluate the performance of each metabolic modelling tool by examining how well each tool's predictions correlated with experimentally measured metabolite data on a sample-by-sample basis. To this end, I calculated the Spearman's correlation coefficients between predicted metabolite profiles and measured metabolite profiles for all eight studies (ρ: -0.55654 to +0.701539) (Figure 2a). Within both the gut and vaginal microbiomes, correlations between predicted measured metabolite profiles for MelonnPan, COBRA_spe and COBRA_gen were significantly different compared to correlations for metabolite profiles with randomly shuffled chemical and sample identifiers. Surprisingly, I observed that correlations for MICOM_spe were negatively significant compared to randomly shuffled metabolite profiles. In addition, correlations between predicted and measured metabolite profiles differed between the gut and vaginal microbiomes. Compared to all other surveyed tools, MelonnPan predictions correlated best with measured metabolite profiles within the gut microbiome. Within the vaginal microbiome however, MelonnPan performed equally as well compared to COBRA_spe and COBRA_gen.

Additionally, I sought to determine whether sample disease status could affect correlations between predicted and measured metabolite profiles across tools (Figure 2b). Within the gut, control samples were better correlated to measured metabolite profiles compared to case samples for MICOM_spe and MICOM_gen, whereas case samples were better correlated for PRMT. Within the vaginal microbiome, control samples were better correlated compared to case samples for both COBRA_spe and MelonnPan. This demonstrates that tool performance may vary across different sites of the human microbiome and can be affected by the disease status of the sample donor.

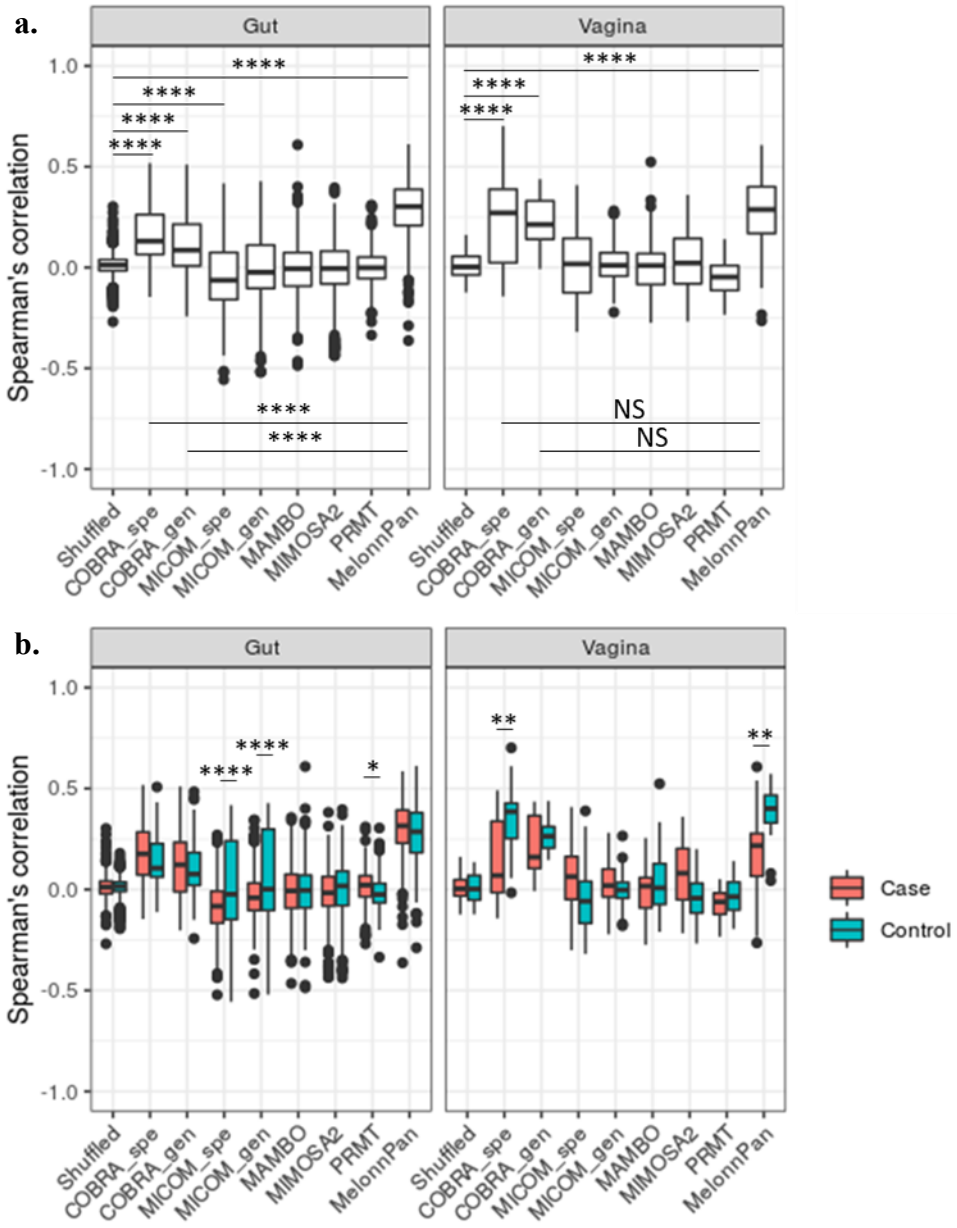



**Figure 2. Spearman's correlation coefficients between predicted and experimentally measured sample metabolite profiles across the gut and vaginal microbiomes.** Predictions were not penalized for having omitted or predicted the occurrence of a metabolite that was included or not included respectively within the measured metabolite profile. I evaluated Spearman's correlations a. across all six tools within the gut and vaginal datasets collected, and then b. further stratified correlations by cases and controls. Kruskall-Wallis test followed by Dunn's post-hoc test with FDR correction. *:$p < 0.05$, **:$p < 0.005$, ***:$p < 0.0005$, ****:$p < 0.00005$

### 3.3 Sample read depth affects tool correlations for MGS sequenced gut microbiome samples.

As read depth can affect the accurate taxonomic and functional profiling of a sample, I sought to determine whether it could by extension affect tool performances. I examined the relationship between sample read count and the Spearman's correlation coefficient between the predicted sample metabolite profile and the experimentally measured sample metabolite profile. I determined that read count does correlate positively with sample correlations between predictions and measured metabolite profiles in datasets containing gut microbiome samples with MGS sequences, albeit weakly (Pearson's ρ: 0.06740972, CI: [0.04405083, 0.09069494], P: 1.653e-08) (Figure 3). I did not observe a significant correlation for the depth of 16S rRNA gene sequenced samples within either the gut or the vaginal microbiomes (Appendix Figure 10).

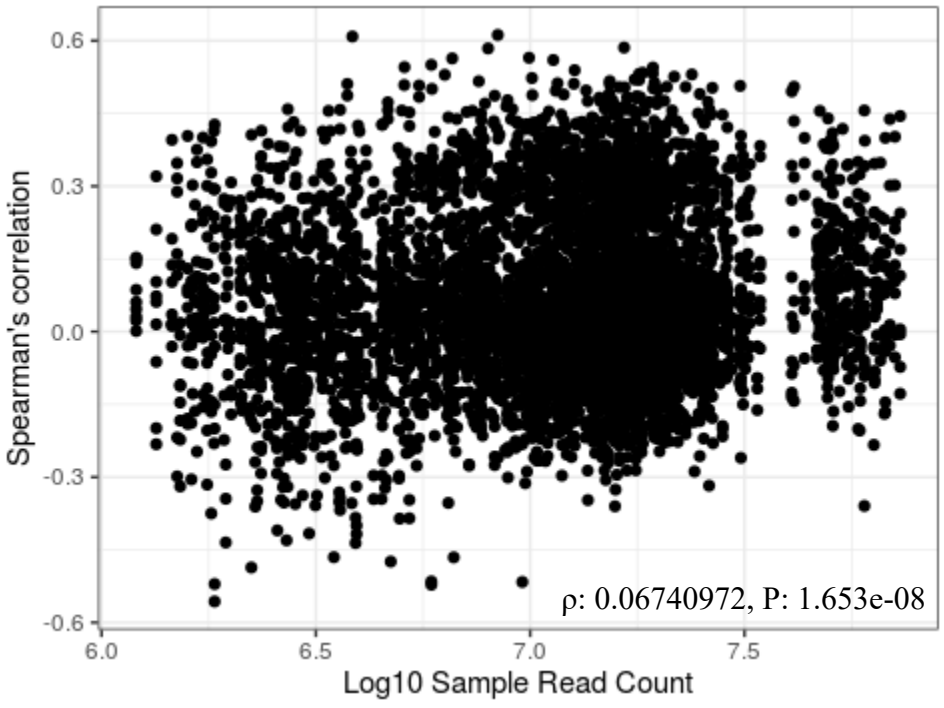



**Figure 3 Association between Spearman's correlation coefficients and sample read count.** Spearman's correlations between predicted and measured metabolite profiles correlates positively with read counts for shotgun metagenome sequenced samples from the gut.

### 3.4 Correlations between predicted and experimentally measured metabolite abundances for core metabolites are poor across surveyed tools

Next, I sought to examine the correlations between predicted metabolite abundances and experimentally measured metabolite abundances across tools. I calculated Spearman's correlation coefficients for the shared sets of "core" metabolites that were commonly predicted by at least seven of eight tool configurations. These core sets were comprised primarily of amino acids, nucleotides and short-chain acids involved in glycolysis and the TCA cycle. Correlations for these core metabolites were generally poor across all tools in the gut microbiome datasets (Figure 4a). Correlations were stronger for core metabolites in the vaginal (Figure 4b) compared to the gut microbiome datasets, though there were no metabolites that were overtly well-predicted across all tools in both datasets. I observed that predictions made by the same tool but using models grouped at different taxonomic levels could result in very divergent results, as was the case when I used COBRA to predict the vaginal metabolome using models aggregated at the species (COBRA_spe) vs. genus level (COBRA_gen) (Figure 4b).

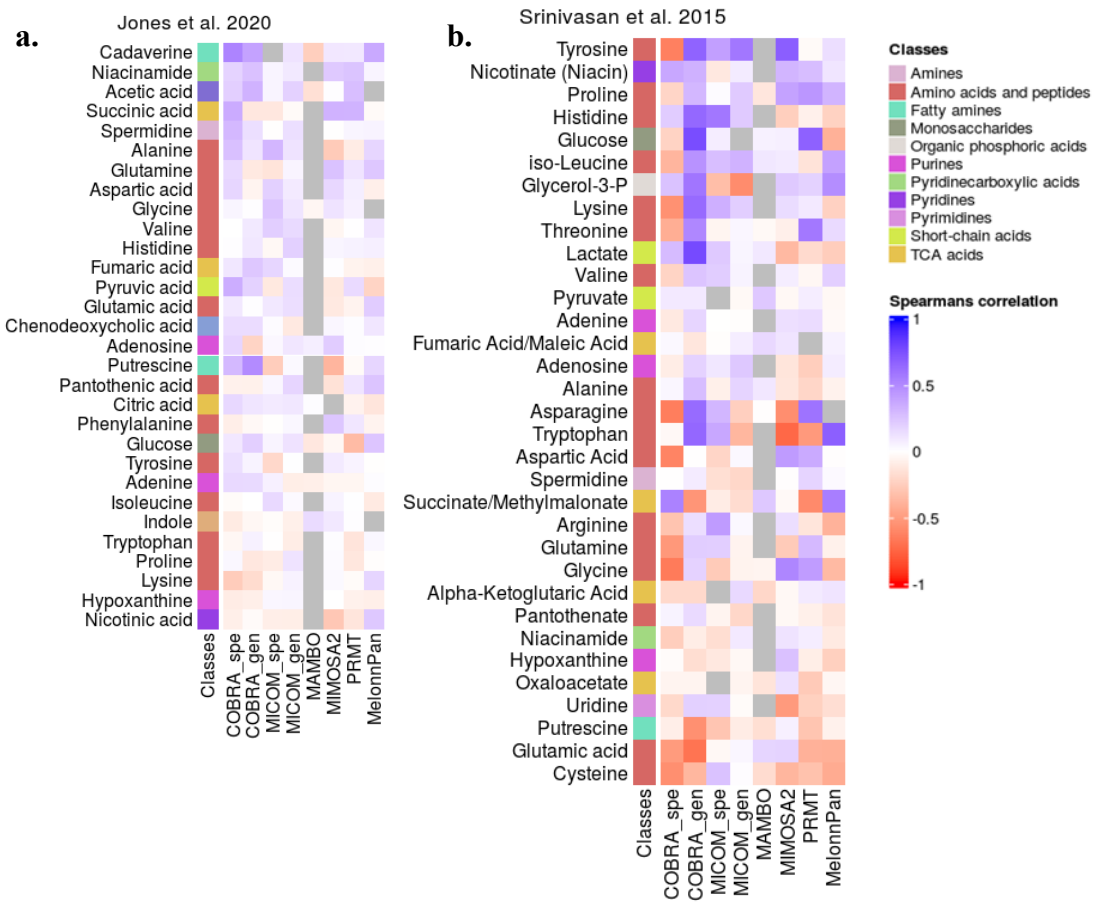



**Figure 4. Heat maps visualizing Spearman's correlations between predicted and measured core metabolites across two representative datasets.** Only metabolites present within the measured metabolite profiles and also predicted to occur within at least seven of eight tool configurations were included. Core metabolites are arranged from highest to lowest mean Spearman's correlation across all tools. Correlations represented for a single dataset containing a. gut microbiome 16S rRNA gene sequences paired with targeted LC-MS metabolite data, and b. vaginal microbiome 16S rRNA sequences paired with targeted LC-MS metabolite data. Grey boxes indicate metabolites that were not predicted by the tool.

### 3.5 Summary

We proceeded with a univariate correlation-based approach to assess tool performance. I found that the machine learning tool MelonnPan possessed the best coverage of metabolites that were experimentally observed across both the gut and the vaginal microbiomes. MelonnPan's metabolite predictions correlated best with the experimentally measured data on a sample-by-sample basis within the gut and was tied for best against the FBA tool COBRA within the vaginal microbiome, demonstrating that tool performances may vary depending on which environment the microbiome was sampled from. Additionally, I determined that case-control status and sample read depth could have an effect on tool performance. MelonnPan and COBRA's predictions for control samples correlated better to experimentally measured metabolite profiles compared to case samples in the vaginal microbiome, but not the gut microbiome. Read count was positively correlated with Spearman's correlations coefficients between predicted and measured metabolite profiles for gut microbiome samples sequenced through MGS. However, this correlation was weak, which may indicate an unsubstantial relationship between read counts and tool performance. Lastly, I found that predicted metabolite abundances for metabolites that were shared across surveyed tools were poorly correlated to measured metabolite abundances.

# CHAPTER 4 – DIFFERENTIAL METABOLITE ANALYSES

A common motivation in metabolomics studies is to identify chemical features that are differentially abundant between two or more populations e.g. cases versus control. As researchers are likely to use these modelling tools to identify discriminating features between different experimental groups, I sought to examine whether the predictions generated by the surveyed tools could approximate differential abundance patterns identified in the experimentally measured metabolite data.

## 4.1 Surveyed tools detect differentially abundant metabolites poorly

We calculated a single F1, precision, and recall score for each tool to summarize their performances across all eight datasets (Figure 5). MelonnPan possessed the best F1 and recall scores (0.3224, 0.3971 respectively) across all eight tool configurations. MIMOSA2 possessed the best precision score (0.4615). MAMBO received scores of zero across all metrics. Altogether, summary scores were low across all tools (<0.5), demonstrating a key limitation in the ability of these tools to capture differential abundance patterns in experimentally measured metabolite data.

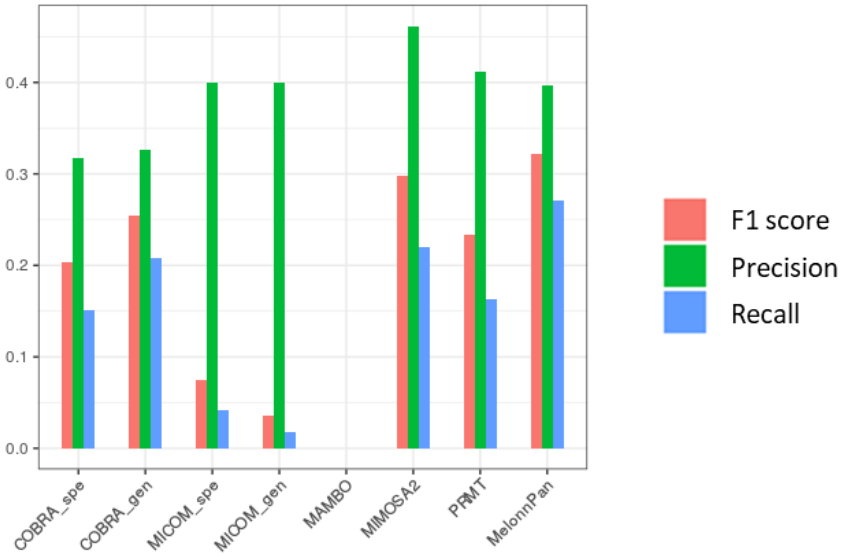



**Figure 5. Differential abundance analysis of tool predictions.** Tool predictions were evaluated for differentially abundant metabolites by Student's t-test on log-normalized, mean centered metabolite data. a. F1, precision, and recall scores were calculated for each tool.

### 4.2 Differentially abundant metabolites identified through MelonnPan intersect in the greatest number with ground-truth differentially abundant metabolites

We examined the overlaps between differential metabolites identified across all tools for all eight datasets (Fig. 6, Table 4). PRMT predicted the greatest number of unique differentially abundant metabolites (506 out of 614 total). Differential metabolites identified via MelonnPan intersected uniquely in the greatest number with those identified in the measured metabolite profiles (72 out of 244 total) (14 out of 614 total). I did not find any shared differential metabolites identified across all tools for any of the datasets tested.

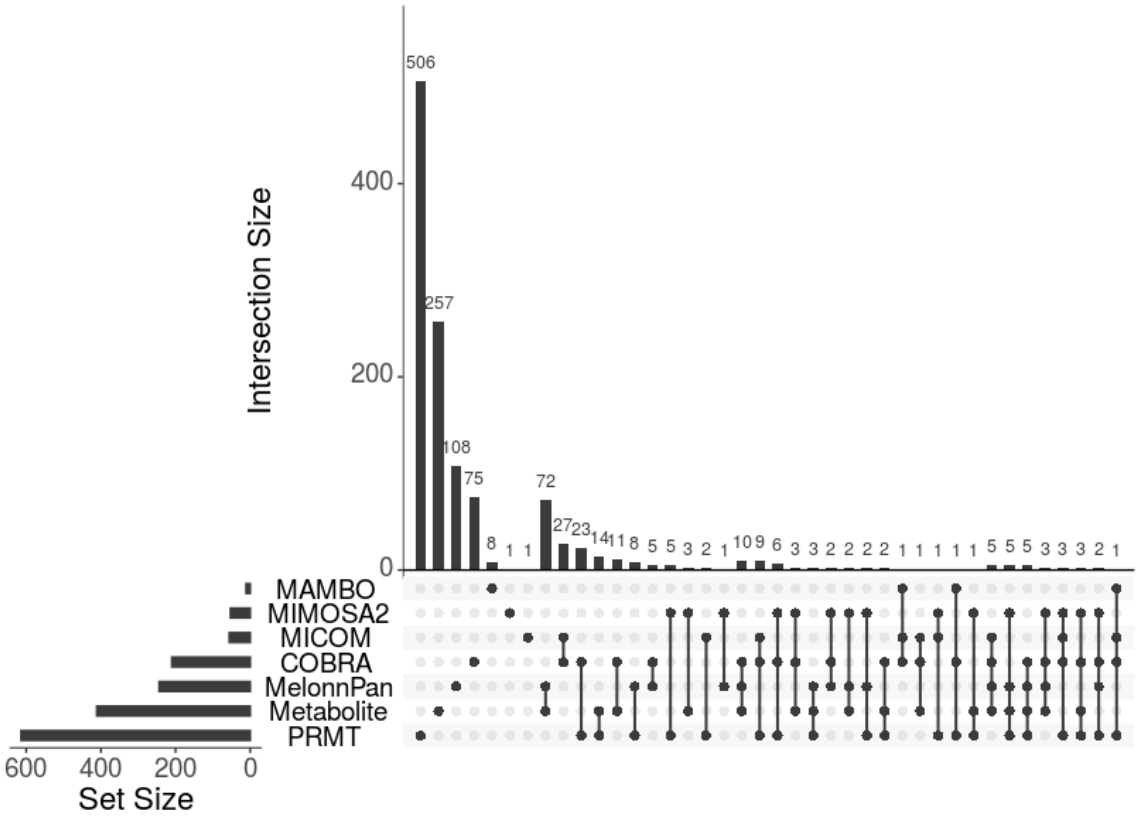



**Figure 6. UpSet plot visualizing the intersections in differential metabolites across tools and aggregated experimentally measured metabolite data.**

**Table 4. Intersects between differentially abundant metabolites within tool predictions and differentially abundant metabolites identified in measured metabolite datasets.**

| Tool | Differential Metabolites | Intersects with Measured Dataset | Intersects Uniquely with Measured Dataset |
|---|---|---|---|
| Metabolite | 411 | NA | NA |
| COBRA | 210 | 54 | 11 |
| MICOM | 56 | 10 | 0 |
| MAMBO | 11 | 0 | 0 |
| MIMOSA2 | 53 | 29 | 3 |
| PRMT | 614 | 44 | 14 |
| MelonnPan | 244 | 115 | 72 |

## 4.3 Summary

We observed low F1, precision, and recall scores of below 0.5 across all surveyed tools, demonstrating that all tools exhibit limitations in the detection of differentially abundant metabolites. MelonnPan possessed the highest F1 score (0.3224) and the greatest number of overlapping differential metabolites with the experimentally measured metabolomics data. Consistent with my earlier finding in Chapter 3 that PRMT predicts the greatest number of metabolites, I found that PRMT identified the greatest number of differentially abundant metabolites. Disturbingly, MAMBO failed to reproduce a single true positive outcome resulting in zero value scores, indicating that this tool may not be suitable for the analysis of case-control studies.

# CHAPTER 5 – MULTIVARIATE ANALYSES OF TOOL PERFORMANCE

Multivariate statistical methods are popular in the analysis of metabolite data, owing to their ability to compare the large matrices of quantitative and semi-quantitative data which are generated through metabolomics experiments. I used principal component analysis (PCA) and Procrustes analysis (PA) to examine the overall congruency between tool predictions and experimentally measured metabolomics data, and to evaluate for inter-tool relatedness.

## 5.1 PA reveals similarities between experimentally measured and predicted metabolomics data from all tools with the exception of MAMBO

PA can be utilized to examine the structural relatedness between two separate matrices of data. Having previously examined the correlations between predictive and measured metabolite profiles with univariate statistical methods (Chapter 3), I sought to re-examine their relatedness using PA (Figure 7). With the exception of MAMBO, each tool queried showed significantly similar Procrustes correlations to at least one experimentally measured metabolite dataset. Predicted metabolite profiles generated by MAMBO did not show significant Procrustes correlations to any of the eight measured metabolite datasets tested. I additionally observed that tool predictions showed greater Procrustes correlations to vaginal datasets than to gut datasets, which echoes my findings in Chapter 3 that tool performance may be affected by the environment under study.

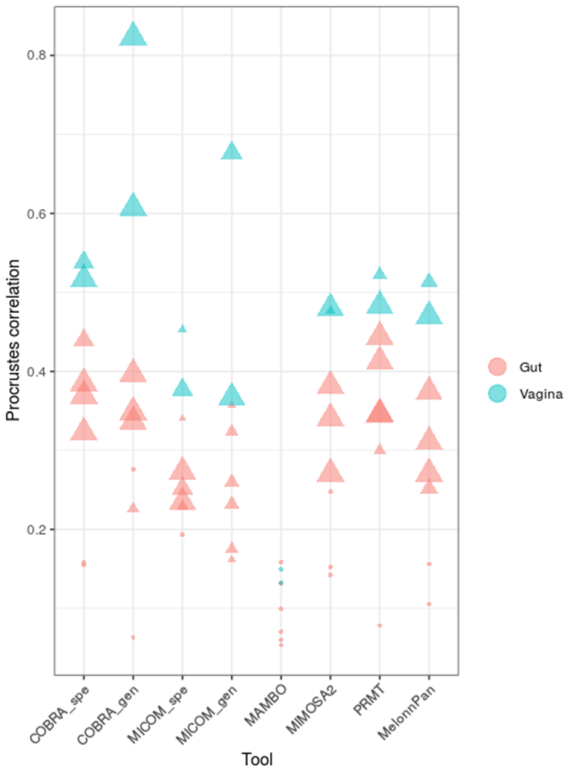



**Figure 7. Procrustes correlations between predicted and experimentally measured metabolomic datasets.** Each point represents the Procrustes correlation between an entire set of predicted and measured metabolite profiles within a single dataset. Point sizes are inversely proportional to p-values. Significantly similar correlations (p-value < 0.05) between tool predictions and measured metabolite profiles are represented by triangles.

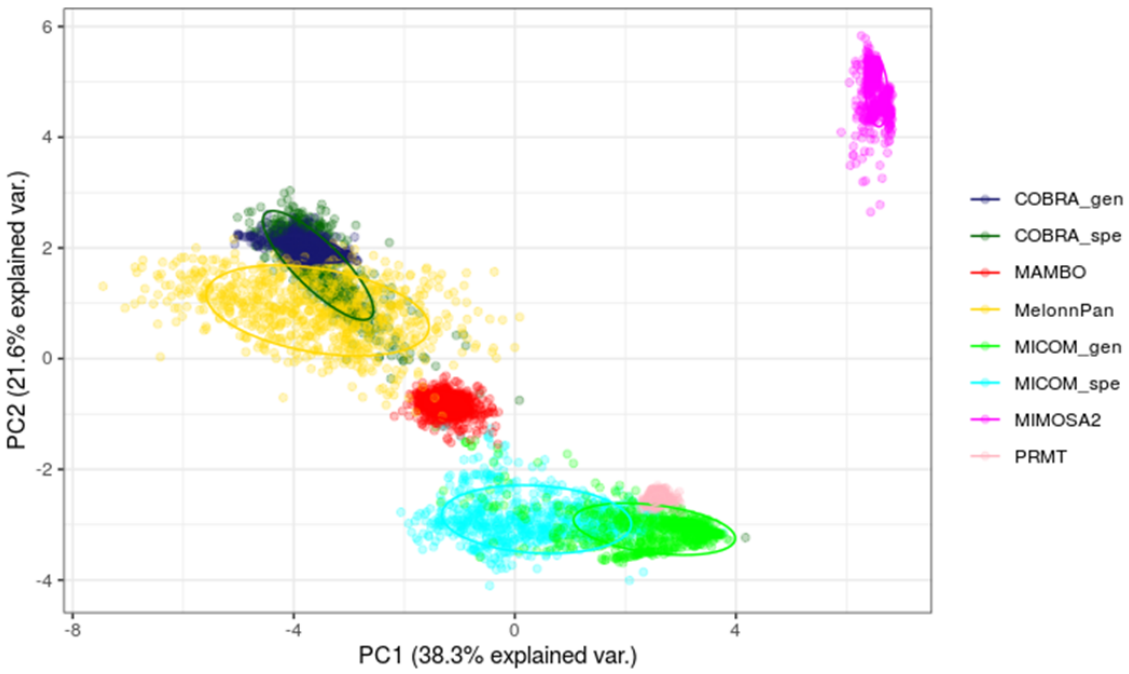
### 5.2 Multivariate analyses of inter-tool relatedness demonstrate lack of reproducibility in MAMBO's predictions

We performed PCA to visualize the underlying structure of the tool predictions. To facilitate this analysis, predicted metabolite profiles were subsetted to a common set of metabolites predicted by at least seven of the eight tool configurations tested before plotting the principal components of these predictions to better interpret the underlying structure of these data (Figure 8a). PCA of the predicted metabolite profiles revealed that, within this core subset of metabolites, predicted metabolite profiles tended to cluster by tool. I leveraged PA to further evaluate the structural similarities between different pairwise

comparisons of tools (Fig. 8b-c). I found significantly similar Procrustes correlations between nearly all pairwise comparisons of tools apart from MAMBO, which had the fewest number of significant similar correlations to other tools within both the gut (one out of seven total pairings) and the vaginal microbiome (two out of seven total pairings).

To investigate the robustness of MAMBO's prediction heuristic, I generated five new metabolomes each for five randomly-selected well-predicted samples (having a final internal Pearson correlation coefficient of >0.95) and calculated pairwise Spearman's correlation coefficients between the first metabolome and the four other metabolomes for all five samples. Concerningly, I calculated a Pearson's correlation coefficient of 0.0245 ± 0.0484 SD, indicating that repeated predictions using MAMBO were not in agreement with one another even though the tool was provided identical inputs.

**a.**



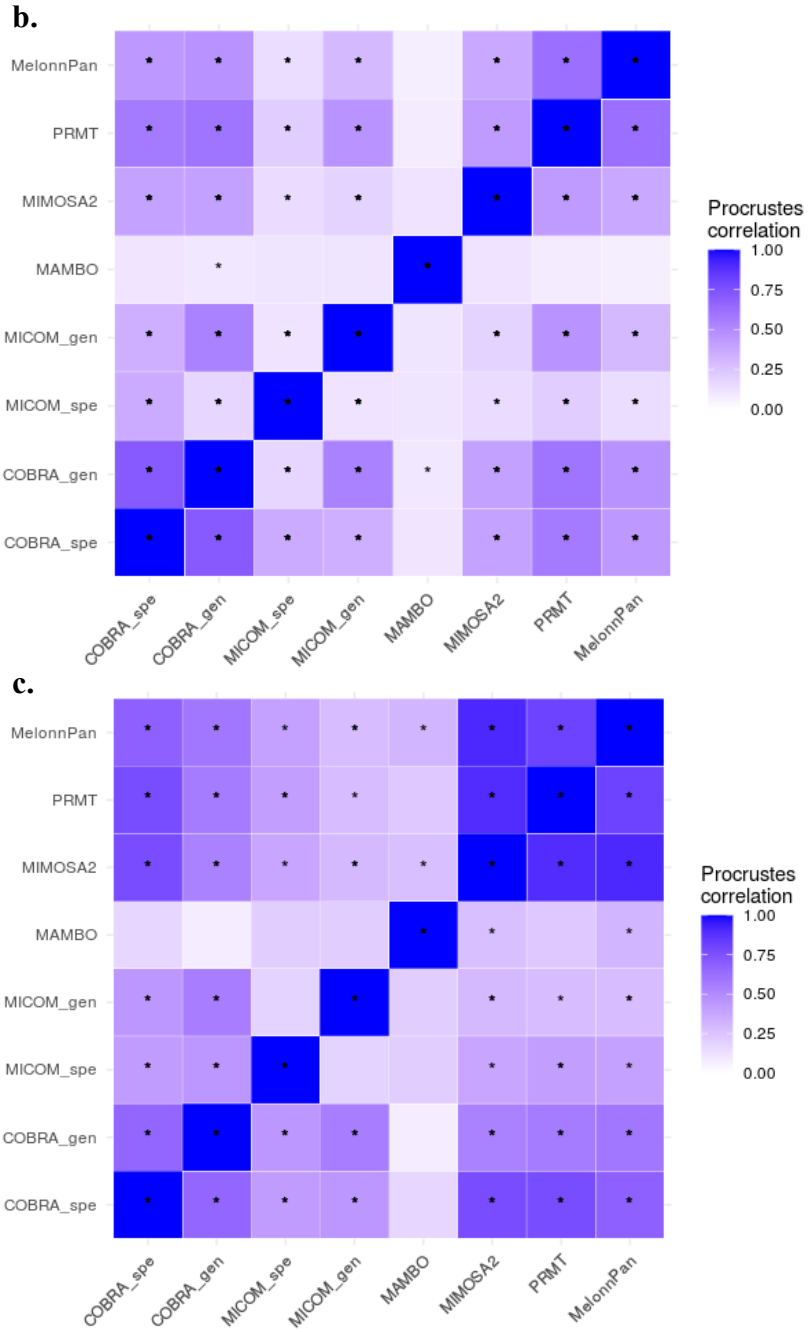



**Figure 8. Multivariate analysis of inter-tool relatedness.** Metabolites within predicted sample metabolite profiles were subsetted to a core set of metabolites that were shared by at least seven of eight tool configurations to build this a. principal component analysis (PCA) plot of predicted sample metabolomes. PA correlations between pairs of tool metabolite predictions for samples from b. the gut or c. the vaginal microbiome. *:$p < 0.05$ indicating a significant similarity as determined by the Procrustes permutational test from the `vegan` R package.

## 5.3 Summary

Here, I sought to explore the similarities between the predicted and experimentally measured metabolomic datasets using the multivariate statistical methods. PA demonstrated that all tools, with the exception of MAMBO, showed structural similarities to measured metabolomics data. Similarly, MAMBO showed the least number of significant pairwise Procrustes correlations between its predictions and predictions made by other tools, suggesting that MAMBO's predictions are structurally dissimilar to other tools. As MAMBO's prediction heuristic is partly stochastic, I sought to determine whether repeated predictions from the same sets of samples would agree with one another. I determined that predicted metabolite profiles generated by MAMBO were not reproducible, suggesting that MAMBO may not be a robust proxy for experimentally measured metabolomics data.

# CHAPTER 6 – DISCUSSION

Rapid improvements in sequencing technologies have facilitated the exponential expansion of public repositories dedicated to hosting microbial sequencing data (Kodama et al., 2012). There is a great deal of interest in extending the use of these data beyond examinations of taxonomic and gene abundances (Dorrestein, Mazmanian, & Knight, 2014; Mallick et al., 2019). To meet this interest, several predictive metabolic modelling tools have been developed that can mathematically extrapolate the encoded biosynthetic machinery within these microbial genomes to predict community-level metabolite data. These tools offer expedient and cost-effective alternatives to global metabolite profiling and may enable the rapid generation of predicted metabolite data for samples in which only sequencing data is available. In this thesis, I sought to evaluate the performances of these tools across a variety of metrics using more than a thousand samples containing paired microbiome sequencing-metabolomics data. Predicted metabolite profiles were evaluated for their congruency to experimentally measured metabolite data in order to gauge tool performance across the gut and the vaginal microbiomes.

## 6.1 Univariate correlation-based analyses of metabolic modelling tools

Using a correlation-based approach to appraise tool predictions against metabolomics data, I demonstrated that MelonnPan may be the most suitable choice for researchers aiming for an accurate and intuitive proxy for metabolomics data. On a sample-by-sample basis, MelonnPan predicted the greatest proportion of metabolites that were directly measured within both the gut and the vaginal microbiome. Sample predictions generated by MelonnPan were also better correlated to measured metabolite profiles within the gut microbiome, whereas in the vaginal microbiome it was tied for first place with COBRA.

MelonnPan's machine learning approach constrains the prediction heuristic from calling metabolites that have not been present in the training dataset. Thus, it may not be surprising that MelonnPan was able to capture greater proportions of experimentally

measured metabolites compared to other pipelines tested. MelonnPan may be sufficient for researchers simply looking for an approximation of metabolomics data. However, because the quality of MelonnPan's predictions is dependent on the quality and comprehensiveness of the training data it is given, MelonnPan's metabolite coverage is necessarily limited to the maximal coverage afforded by MS and NMR which both possess their own biases and technical limitations (Christians, Klawitter, Hornberger, & Klawitter, 2011). Knowledge-based tools, which require no training data to generate predictions, could potentially be used to predict metabolites that are below the detection limit of metabolomics instrumentation. Thus, the much greater total number of predicted metabolites I observed using PRMT, which does not constrain its outputted predictions to chemical features found in training datasets or to features that are explicitly named in exchange reactions (as FBA tools do), suggests that PRMT may still hold value as a *de novo* metabolite predictor.

We determined through a correlation-based analysis that MelonnPan produced sample metabolite profiles that were better correlated to experimentally measured metabolite profiles compared to all other tools within the gut. The lower performances observed in all other tools could have arisen through several different reasons. The chief assumption made by PRMT and MIMOSA2 – namely that gene abundances are directly related to the rate of enzymatic activity – may not hold true in the gut and vaginal microbiomes. FBA assumptions of a well-mixed system operating under steady-state conditions may not be met in the gut where microbial composition can be spatially heterogeneous (Lu et al., 2014). Changes in gene regulation arising from environmental changes – which cannot be captured through FBA or gene abundance-based approaches – could effect changes in metabolite abundances (Shlomi et al., 2008). Lastly, because FBA and gene-based abundance tools predominantly aim to reconstruct bacterial metabolism, other mechanisms controlling for metabolite variation such as phage predation (Hsu et al., 2019), eukaryotic metabolism, or human uptake of microbial metabolites (Van Treuren & Dodd, 2020) are simply unaccounted for in these tools. Machine learning can detect complex associations involving multiple factors, thus MelonnPan may be capturing these alternative mechanisms during model training. Nevertheless, my results demonstrated that the FBA tool COBRA can generate predictions that correlate well with measured metabolite profiles within both the gut and the vaginal microbiome, indicating that a

mechanism-driven approach to metabolome prediction is feasible.

There are important caveats and limitations to consider for these tools. I observed that the choice of microbiome i.e. gut or vaginal, under study as well as the health status of the donor could affect tool performance, which may be important caveats when choosing an appropriate study design if predictive metabolic modelling is to be a large component of the study. I determined that sample-by-sample correlations between predicted and measured metabolite profiles were generally low across all tools. Additionally, I observed that correlations between predicted and measured metabolite abundances were poor for subsets of metabolites that appeared in common across all surveyed tools. I did not identify any metabolites that were consistently well predicted across tools and studies, which may indicate study-specific effects on microbe-metabolite linkages. A recent study described large variations in certain subsets of metabolites as well as their identified microbial contributors across studies (Muller, Algavi, & Borenstein, 2021). Ideally, greater numbers of paired sequencing-metabolomics studies would be needed to more thoroughly elucidate the limitations associated with metabolic modelling across cohorts. Dedicated repositories for paired metabolomics and microbiome data are making effective meta-analyses of these data more accessible (Schorn et al., 2021), and I hope that my evaluation framework will be able to take advantage of these resources as they become more widely adopted.

### 6.2 Differential metabolite analyses

Often, the principal goal of a metabolomics experiment is to identify a discriminating metabolite for a human phenotype. A metabolite that is differentially abundant between a healthy and disease population could be a useful biomarker for disease onset or activity, or alternatively a potential therapeutic target. I sought to determine whether these tools could recover differential abundance patterns identified in measured metabolomics datasets by calculating F1, precision, and recall scores for each tool. Although I lacked the appropriate number of studies to evaluate for significance, I observed that MelonnPan had the highest F1 (0.3224) and recall (0.3971) scores, whereas MIMOSA2 possessed the highest precision score (0.4615). Surveyed scores were

altogether quite low (<0.5), which is consistent with earlier findings from Yin et al., that demonstrated low scores between MelonnPan, MIMOSA, and an unreleased tool of their making (Yin et al., 2020).

This is a noteworthy caveat for researchers aiming to leverage these predictive tools as replacements for experimentally obtained metabolite data. The low F1 scores I observed may result from several possibilities. The microbe-metabolite associations that are modelled using these metabolic modelling tools may not necessarily be extendible across case and control populations. Muller et al., recently characterized datasets containing paired data from case-control cohorts and determined that these associations were generally robust across healthy and disease populations (Muller et al., 2021). On the other hand, the human microbiome is not expected to be the sole determinant of metabolite variation. Paired sequencing-metabolomics surveys of the gut have indicated that, although the composition of the gut microbiome can explain a large proportion of the variation within the fecal metabolome, other factors relating to the host are also expected to play a role (Zierer et al., 2018; Visconti et al., 2019). In this context, dysregulation of host metabolite uptake, through impaired barrier function for instance, may alter the metabolome independently of the microbiome. Greater incorporation of host metabolism in these metabolic modelling tools may help improve their ability to capture differential abundance patterns if this is the case.

Alternatively, the low scores may simply indicate that the tools themselves cannot fully capture metabolite abundances accurately. This would be consistent with the low Spearman's correlations I observed between predicted and experimentally measured metabolite profiles for all tools, and further highlights the need for researchers to use the predictions generated by these tools conservatively.

### 6.3 Multivariate metabolite analyses

Multivariate statistical methods are an appropriate means to compare multiple matrices of high-dimensional data. I sought to evaluate tool performances by utilizing PA to examine the structural congruency between predicted and experimentally measured

metabolite profiles. I observed significant Procrustes correlations for at least one dataset each for all tools with the exception of MAMBO. Further investigation through pairwise comparisons of tool predictions led us to the observation that MAMBO possessed the fewest significant Procrustes correlations to any other tool within both the gut and the vaginal microbiome, and that repeated metabolomes generated from the same samples correlated poorly with one another. This may indicate that the stochastic prediction algorithm utilized by MAMBO was not robust enough to approximate high-dimensional metabolomics data. An important caveat is that, owing to issues running the tool, the number of optimization steps used to parameterize MAMBO was restricted to one-tenth of the authors' recommended one-million steps. I mitigated this limitation by selecting samples that had been identified as well-predicted (internal Pearson's correlation of >0.95) by the tool for repeated analyses, though there may be other negative effects on the outputted predictions that are masked by a high internal Pearson's correlation.

In line with my findings in Chapter 3, I found that predictions generated for the vaginal metabolome were better correlated to experimentally measured metabolome compared to the gut metabolome. This may suggest that the vaginal microbiome is a greater determinant of the surrounding metabolic landscape compared to the gut microbiome. The vaginal microbiome is dominated by a small number of microbial species – predominantly *Lactobacillus* spp. – that may induce a much greater effect on the vaginal metabolome compared to the gut microbiome, which is diffusely represented by over a hundred microbial species per person on average (Qin et al., 2010). Other determinants of metabolome composition within the gut, including the uptake of microbial metabolites by gut epithelial cells and the secretion of both host and microbially-derived metabolites through the biliary tree (Van Treuren & Dodd, 2020), may confound tool predictions if those tools cannot account for these mechanisms. However, given the limited number of vaginal datasets obtained for this study, compounded with the fact that both datasets were collected by the same authors and used 16S rRNA sequencing exclusively, future meta-analyses of metabolite prediction tools will need to revisit this question more thoroughly.

## 6.4 Limitations and future directions

There are several limitations associated with this evaluation framework. To assess tool performance, I considered the experimentally measured metabolomics data to be a "gold standard" against which metabolite predictions could be validated. This is the standard approach through which metabolic modelling tools are typically validated (Mallick et al., 2019; Garza et al., 2018; Yin et al., 2020). In practice however, confident metabolite identification of untargeted metabolomics data is technically challenging (reasons are outlined in section 1.4.4), especially in complex biospecimens such as stool. Within this context, the large numbers of metabolites that were uniquely predicted by PRMT (Figure 1) may in fact be present in the environment but simply could not be captured through metabolomics. Efforts to improve compound identification and enhance the global detection of metabolites are ongoing, and I anticipate that this will make fairer assessments of metabolite coverage across tools possible.

Quantification of untargeted peak data is equally difficult because peak intensities represent relative approximations of metabolite concentrations (Kapoore & Vaidyanathan, 2016). Without chemical standards it is difficult to compare concentrations of different metabolites against one another within the same sample. Though I included studies containing metabolomics data collected through a diverse array of instruments, an overwhelming number of samples were metabolically profiled via untargeted approaches (880 out of 1089 samples total). Thus, my sample-by-sample evaluation approach may favour machine learning modelling tools over FBA or gene abundance-based approaches because machine learning tools can learn to mimic the semi-quantative data structure of untargeted MS data. In the future, an alternative approach in which predicted metabolites are subsequently validated through targeted MS could help answer my research questions in a more quantitative fashion.

It is important to note that metabolomics is not directly analogous to metabolite fluxes, which are outputted by FBA tools. The former provides only a snapshot of metabolite concentrations at a single point in time whereas the latter provides additional information as to how those concentrations are changing temporally (Srivastava, Kowalski, Callahan, Meikle, & Creek, 2016). Thus, a large flux of metabolite directed towards the

environment may not necessarily translate to a greater relative abundance of that metabolite as measured strictly through metabolomics. Provided that the rate in which that metabolite is removed from the environment e.g., by excretion of stool or uptake by epithelial cells, is equal or greater than the rate that metabolite is added into the environment, an increased abundance of that metabolite may not be apparent. *In vivo* methods to quantify flux, such as through isotope tracing, are difficult to carry out in large cohort studies but may generate data that are more readily comparable as a "gold standard" to flux data generated from COBRA and MICOM (Berry & Loy, 2018).

While MelonnPan was generally superior compared to the mechanism-driven tools outlined in my meta-analysis, future tools incorporating human metabolism and more granular nutritional information into their working principles may help bridge this observed performance gap. To illustrate, FBA simulations incorporating gut microbe models together with human epithelial cells have uncovered novel microbe-microbe interactions arising through competition over enterocyte-derived glucose (Heinken & Thiele, 2015). Whole-body models of metabolism parameterized by dietary and microbiome sequencing data are now able to capture metabolic interactions at the organ level (Thiele et al., 2020). For the FBA tools COBRA and MICOM, I parameterized all community models using the author-provided nutritional constraints. These constraints represented the "average" European diet and are not necessarily applicable to all study participants e.g. those undergoing nutritional therapy for CD (Jones et al., 2020). Further efforts to evaluate human and microbial metabolism in tandem, and more accurately model the nutritional constraints of study populations, could improve tool performances and enhance the application of these tools beyond predicting metabolites.

## 6.5 Conclusions

Microbiome-level predictive metabolic modelling tools are a promising informatic development with the potential to greatly expand the utility of sequencing data and enhance the accessibility of metabolomic data collection. In this work, I have demonstrated the feasibility of these tools to predict metabolomics data from sequencing data alone, while

highlighting important limitations relating to their working principles and their ability to identify discriminating metabolites. The evaluation framework presented in this work is an important initiative through which future metabolic modelling tools can be appraised.

## REFERENCES


Abram, F. (2015). Systems-based approaches to unravel multi-species microbial community functioning. *Computational and Structural Biotechnology Journal*, *13*, 24–32. https://doi.org/10.1016/j.csbj.2014.11.009

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics Online*, *12*(Suppl 1), 5–16. https://doi.org/10.4137/EBO.S36436

Albenberg, L., Esipova, T. V., Judge, C. P., Bittinger, K., Chen, J., Laughlin, A., Grunberg, S., Baldassano, R. N., Lewis, J. D., Li, H., Thom, S. R., Bushman, F. D., Vinogradov, S. A., & Wu, G. D. (2014). Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota. *Gastroenterology*, *147*(5), 1055-1063.e8. https://doi.org/10.1053/j.gastro.2014.07.020

Alonso, A., Marsal, S., & Julià, A. (2015). Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Frontiers in Bioengineering and Biotechnology*, *3*, 23. https://doi.org/10.3389/fbioe.2015.00023

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, *21*, 30. https://doi.org/10.1186/s13059-020-1935-5

Baldini, F., Heinken, A., Heirendt, L., Magnusdottir, S., Fleming, R. M. T., & Thiele, I. (2019). The Microbiome Modeling Toolbox: From microbial interactions to personalized microbial communities. *Bioinformatics*, *35*(13), 2332–2334. https://doi.org/10.1093/bioinformatics/bty941

Balvočiūtė, M., & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT — how do these taxonomies compare? *BMC Genomics*, *18*(2), 114. https://doi.org/10.1186/s12864-017-3501-4

Baran, R., & Northen, T. R. (2013). Robust Automated Mass Spectra Interpretation and Chemical Formula Calculation Using Mixed Integer Linear Programming. *Analytical Chemistry*, *85*(20), 9777–9784. https://doi.org/10.1021/ac402180c

Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., & Kaleta, C. (2017). BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLOS Computational Biology*, *13*(5), e1005544. https://doi.org/10.1371/journal.pcbi.1005544

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., … Schloter, M. (2020). Microbiome definition re-visited: Old concepts and new challenges. *Microbiome*, *8*(1), 103. https://doi.org/10.1186/s40168-020-00875-0

Bernini, P., Bertini, I., Luchinat, C., Nepi, S., Saccenti, E., Schäfer, H., Schütz, B., Spraul, M., & Tenori, L. (2009). Individual human phenotypes in metabolic space and time. *Journal of Proteome Research*, *8*(9), 4264–4271. https://doi.org/10.1021/pr900344m

Berry, D., & Loy, A. (2018). Stable-Isotope Probing of Human and Animal Microbiome Function. *Trends in Microbiology*, *26*(12), 999–1007. https://doi.org/10.1016/j.tim.2018.06.004

Blaak, E. e., Canfora, E. e., Theis, S., Frost, G., Groen, A. k., Mithieux, G., Nauta, A., Scott, K., Stahl, B., van Harsselaar, J., van Tol, R., Vaughan, E. e., & Verbeke, K. (2020). Short chain fatty acids in human gut and metabolic health. *Beneficial Microbes*, *11*(5), 411–455. https://doi.org/10.3920/BM2020.0057

Blaženović, I., Kind, T., Ji, J., & Fiehn, O. (2018). Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites*, *8*(2), 31. https://doi.org/10.3390/metabo8020031

Bowerman, K. L., Rehman, S. F., Vaughan, A., Lachner, N., Budden, K. F., Kim, R. Y., Wood, D. L. A., Gellatly, S. L., Shukla, S. D., Wood, L. G., Yang, I. A., Wark, P. A., Hugenholtz, P., & Hansbro, P. M. (2020). Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nature Communications*, *11*(1), 5886. https://doi.org/10.1038/s41467-020-19701-0

Breuer, R. I., Soergel, K. H., Lashner, B. A., Christ, M. L., Hanauer, S. B., Vanagunas, A., Harig, J. M., Keshavarzian, A., Robinson, M., Sellin, J. H., Weinberg, D., Vidican, D. E., Flemal, K. L., & Rademaker, A. W. (1997). Short chain fatty acid rectal irrigation for left-sided ulcerative colitis: A randomised, placebo controlled trial. *Gut*, *40*(4), 485–491. https://doi.org/10.1136/gut.40.4.485

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta*, *1842*(10), 1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Chan, S. H. J., Simons, M. N., & Maranas, C. D. (2017). SteadyCom: Predicting microbial abundances while ensuring community stability. *PLOS Computational Biology*, *13*(5), e1005539. https://doi.org/10.1371/journal.pcbi.1005539

Chong, J., & Xia, J. (2017). Computational Approaches for Integrative Analysis of the Metabolome and Microbiome. *Metabolites*, *7*(4), 62. https://doi.org/10.3390/metabo7040062

Christians, U., Klawitter, J., Hornberger, A., & Klawitter, J. (2011). How unbiased is non-targeted metabolomics and is targeted pathway screening the solution? *Current Pharmaceutical Biotechnology*, *12*(7), 1053–1066. https://doi.org/10.2174/138920111795909078

Chun, J., Lee, J.-H., Jung, Y., Kim, M., Kim, S., Kim, B. K., & Lim, Y.-W. 2007. (n.d.). EzTaxon: A web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, *57*(10), 2259–2261. https://doi.org/10.1099/ijs.0.64915-0

Coble, J. B., & Fraga, C. G. (2014). Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *Journal of Chromatography A*, *1358*, 155–164. https://doi.org/10.1016/j.chroma.2014.06.100

Coburn, L. A., Horst, S. N., Allaman, M. M., Brown, C. T., Williams, C. S., Hodges, M. E., Druce, J. P., Beaulieu, D. B., Schwartz, D. A., & Wilson, K. T. (2016). L-arginine Availability and Metabolism Is Altered in Ulcerative Colitis. *Inflammatory Bowel Diseases*, *22*(8), 1847–1858. https://doi.org/10.1097/MIB.0000000000000790

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., & Tiedje, J. M. (2009). The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, *37*(suppl_1), D141–D145. https://doi.org/10.1093/nar/gkn879

Colwell, R. R. (1997). Microbial diversity: The importance of exploration and conservation. *Journal of Industrial Microbiology and Biotechnology*, *18*(5), 302–307. https://doi.org/10.1038/sj.jim.2900390

Cui, L., Lu, H., & Lee, Y. H. (2018). Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases. *Mass Spectrometry Reviews*, *37*(6), 772–792. https://doi.org/10.1002/mas.21562

Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P., & Macfarlane, G. T. (1987). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut*, *28*(10), 1221–1227. https://doi.org/10.1136/gut.28.10.1221

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R. H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P. H., & Wagner, M. (2015). Complete nitrification by Nitrospira bacteria. *Nature*, *528*(7583), 504–509. https://doi.org/10.1038/nature16461

De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I. B., La Storia, A., Laghi, L., Serrazanetti, D. I., Di Cagno, R., Ferrocino, I., Lazzi, C., Turroni, S., Cocolin, L., Brigidi, P., Neviani, E., Gobbetti, M., O'Toole, P. W., & Ercolini, D. (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut*, *65*(11), 1812–1821. https://doi.org/10.1136/gutjnl-2015-309957

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Diener, C., Gibbons, S. M., & Resendis-Antonio, O. (2020). MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *MSystems*, *5*(1), e00606-19. https://doi.org/10.1128/mSystems.00606-19

Dorrestein, P. C., Mazmanian, S. K., & Knight, R. (2014). From microbiomess to metabolomes to function during host-microbial interactions. *Immunity*, *40*(6), 824–832. https://doi.org/10.1016/j.immuni.2014.05.015

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., & Viant, M. R. (2013). Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics*, *9*(1), 44–66. https://doi.org/10.1007/s11306-012-0434-4

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Scientific Reports*, *11*(1), 3030. https://doi.org/10.1038/s41598-021-82726-y

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*, 74. https://doi.org/10.1186/1752-0509-7-74

El-Aneed, A., Cohen, A., & Banoub, J. (2009). Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews*, *44*(3), 210–230. https://doi.org/10.1080/05704920902717872

Emwas, A.-H. M. (2015). The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In J. T. Bjerrum (Ed.), *Metabonomics: Methods and Protocols* (pp. 161–193). Springer. https://doi.org/10.1007/978-1-4939-2377-9_13

Emwas, A.-H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Gowda, G. A. N., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites*, *9*(7), 123. https://doi.org/10.3390/metabo9070123

Escobar-Zepeda, A., Vera-Ponce de León, A., & Sanchez-Flores, A. (2015). The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Frontiers in Genetics*, *6*, 348. https://doi.org/10.3389/fgene.2015.00348

Fahy, E., & Subramaniam, S. (2020). RefMet: A reference nomenclature for metabolomics. *Nature Methods*, *17*(12), 1173–1174. https://doi.org/10.1038/s41592-020-01009-y

Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science*, *320*(5879), 1034–1039. https://doi.org/10.1126/science.1153213

Feist, A. M., & Palsson, B. Ø. (2008). The Growing Scope of Applications of Genome-scale Metabolic Reconstructions: The case of E. coli. *Nature Biotechnology*, *26*(6), 659–667. https://doi.org/10.1038/nbt1401

Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., Wen, B., Chen, X., Yan, W., Fan, Y., Guo, Z., Meng, N., Chen, J., Yu, X., Zhang, Z., Kristiansen, K., Wang, J., Xu, X., He, K., & Li, G. (2016). Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific Reports*, *6*, 22525. https://doi.org/10.1038/srep22525

Fennema, D., Phillips, I. R., & Shephard, E. A. (2016). Trimethylamine and Trimethylamine N-Oxide, a Flavin-Containing Monooxygenase 3 (FMO3)-Mediated Host-Microbiome Metabolic Axis Implicated in Health and Disease. *Drug Metabolism and Disposition*, *44*(11), 1839–1850. https://doi.org/10.1124/dmd.116.070615

Fiehn, O. (2002). Metabolomics—The link between genotypes and phenotypes. *Plant Molecular Biology*, *48*(1–2), 155–171.

Finnie, I. A., Dwarakanath, A. D., Taylor, B. A., & Rhodes, J. M. (1995). Colonic mucin synthesis is increased by sodium butyrate. *Gut*, *36*(1), 93–99. https://doi.org/10.1136/gut.36.1.93

Forsberg, E. M., Huan, T., Rinehart, D., Benton, H. P., Warth, B., Hilmers, B., & Siuzdak, G. (2018). Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nature Protocols*, *13*(4), 633–651. https://doi.org/10.1038/nprot.2017.151

Frost, G., Sleeth, M. L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., Anastasovska, J., Ghourab, S., Hankir, M., Zhang, S., Carling, D., Swann, J. R., Gibson, G., Viardot, A., Morrison, D., Louise Thomas, E., & Bell, J. D. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nature Communications*, *5*(1), 3611. https://doi.org/10.1038/ncomms4611

Garza, D. R., & Dutilh, B. E. (2015). From cultured to uncultured genome sequences: Metagenomics and modeling microbial ecosystems. *Cellular and Molecular Life Sciences*, *72*, 4287–4308. https://doi.org/10.1007/s00018-015-2004-1

Garza, D. R., Taddese, R., Wirbel, J., Zeller, G., Boleij, A., Huynen, M. A., & Dutilh, B. E. (2020). Metabolic models predict bacterial passengers in colorectal cancer. *Cancer & Metabolism*, *8*, 3. https://doi.org/10.1186/s40170-020-0208-9

Garza, D. R., van Verk, M. C., Huynen, M. A., & Dutilh, B. E. (2018). Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nature Microbiology*, *3*(4), 456–460. https://doi.org/10.1038/s41564-018-0124-8

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*(1), 33–51. https://doi.org/10.1007/BF02291478

Griffiths, W. J., Koal, T., Wang, Y., Kohl, M., Enot, D. P., & Deigner, H.-P. (2010). Targeted Metabolomics for Biomarker Discovery. *Angewandte Chemie International Edition*, *49*(32), 5426–5445. https://doi.org/10.1002/anie.200905579

Guijas, C., Montenegro-Burke, J. R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., Koellensperger, G., Huan, T., Uritboonthai, W., Aisporna, A. E., Wolan, D. W., Spilker, M. E., Benton, H. P., & Siuzdak, G. (2018). METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Analytical Chemistry*, *90*(5), 3156–3164. https://doi.org/10.1021/acs.analchem.7b04424

Hale, V. L., Jeraldo, P., Mundy, M., Yao, J., Keeney, G., Scott, N., Cheek, E. H., Davidson, J., Green, M., Martinez, C., Lehman, J., Pettry, C., Reed, E., Lyke, K., White, B. A., Diener, C., Resendis-Antonio, O., Gransee, J., Dutta, T., … Chia, N. (2018). Synthesis of multi-omic data and community metabolic models reveals insights into the role of hydrogen sulfide in colon cancer. *Methods (San Diego, Calif.)*, *149*, 59–68. https://doi.org/10.1016/j.ymeth.2018.04.024

Hao, L., Wang, J., Page, D., Asthana, S., Zetterberg, H., Carlsson, C., Okonkwo, O. C., & Li, L. (2018). Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease. *Scientific Reports*, *8*(1), 9291. https://doi.org/10.1038/s41598-018-27031-x

Heianza, Y., Ma, W., Manson, J. E., Rexrode, K. M., & Qi, L. (2017). Gut Microbiota Metabolites and Risk of Major Adverse Cardiovascular Disease Events and Death: A Systematic Review and Meta-Analysis of Prospective Studies. *Journal of the American Heart Association*, *6*(7), e004947. https://doi.org/10.1161/JAHA.116.004947

Heinken, A., Basile, A., Hertel, J., Thinnes, C., & Thiele, I. (2021). Genome-Scale Metabolic Modeling of the Human Microbiome in the Era of Personalized Medicine. *Annual Review of Microbiology*, *75*, 199–222. https://doi.org/10.1146/annurev-micro-060221-012134

Heinken, A., & Thiele, I. (2015). Anoxic Conditions Promote Species-Specific Mutualism between Gut Microbes In Silico. *Applied and Environmental Microbiology*, *81*(12), 4049–4061. https://doi.org/10.1128/AEM.00101-15

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., Magnusdóttir, S., Ng, C. Y., Preciat, G., Žagare, A., Chan, S. H. J., Aurich, M. K., Clancy, C. M., Modamio, J., Sauls, J. T., … Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, *14*(3), 639–702. https://doi.org/10.1038/s41596-018-0098-2

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., & Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, *28*(9), 977–982. https://doi.org/10.1038/nbt.1672

Hohrenk, L. L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., & Schmidt, T. C. (2020). Comparison of Software Tools for Liquid Chromatography–High-Resolution Mass Spectrometry Data Processing in Nontarget Screening of Environmental Samples. *Analytical Chemistry*, *92*(2), 1898–1907. https://doi.org/10.1021/acs.analchem.9b04095

Hooft, J. J. J. van der, Mohimani, H., Bauermeister, A., C. Dorrestein, P., R. Duncan, K., & H. Medema, M. (2020). Linking genomics and metabolomics to chart specialized metabolic diversity. *Chemical Society Reviews*, *49*(11), 3297–3314. https://doi.org/10.1039/D0CS00162G

Hsu, B. B., Gibson, T. E., Yeliseyev, V., Liu, Q., Lyon, L., Bry, L., Silver, P. A., & Gerber, G. K. (2019). Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe*, *25*(6), 803-814.e5. https://doi.org/10.1016/j.chom.2019.05.001

Huda-Faujan, N., Abdulamir, A. S., Fatimah, A. B., Anas, O. M., Shuhaimi, M., Yazid, A. M., & Loong, Y. Y. (2010). The impact of the level of the intestinal short chain Fatty acids in inflammatory bowel disease patients versus healthy subjects. *The Open Biochemistry Journal*, *4*, 53–58. https://doi.org/10.2174/1874091X01004010053

Iwai, S., Weinmaier, T., Schmidt, B. L., Albertson, D. G., Poloso, N. J., Dabbagh, K., & DeSantis, T. Z. (2016). Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. *PLoS ONE*, *11*(11), e0166104. https://doi.org/10.1371/journal.pone.0166104

Janeiro, M. H., Ramírez, M. J., Milagro, F. I., Martínez, J. A., & Solas, M. (2018). Implication of Trimethylamine N-Oxide (TMAO) in Disease: Potential Biomarker or New Therapeutic Target. *Nutrients*, *10*(10), 1398. https://doi.org/10.3390/nu10101398

Jeong, J., Yun, K., Mun, S., Chung, W.-H., Choi, S.-Y., Nam, Y., Lim, M. Y., Hong, C. P., Park, C., Ahn, Y. J., & Han, K. (2021). The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. *Scientific Reports*, *11*, 1727. https://doi.org/10.1038/s41598-020-80826-9

Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, *10*(1), 5029. https://doi.org/10.1038/s41467-019-13036-1

Jones, C. M. A., Connors, J., Dunn, K. A., Bielawski, J. P., Comeau, A. M., Langille, M. G. I., & Van Limbergen, J. (2020). Bacterial Taxa and Functions Are Predictive of Sustained Remission Following Exclusive Enteral Nutrition in Pediatric Crohn's Disease. *Inflammatory Bowel Diseases*, *26*(7), 1026–1037. https://doi.org/10.1093/ibd/izaa001

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A. L., Madsen, K. L., & Wong, G. K.-S. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. *Frontiers in Microbiology*, *7*, 459. https://doi.org/10.3389/fmicb.2016.00459

Kapoore, R. V., & Vaidyanathan, S. (2016). Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2079), 20150363. https://doi.org/10.1098/rsta.2015.0363

Karimian, E., & Motamedian, E. (2020). ACBM: An Integrated Agent and Constraint Based Modeling Framework for Simulation of Microbial Communities. *Scientific Reports*, *10*(1), 8695. https://doi.org/10.1038/s41598-020-65659-w

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., & Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, *20*(4), 1085–1093. https://doi.org/10.1093/bib/bbx085

Kim, K. N., Yao, Y., & Ju, S. Y. (2019). Short Chain Fatty Acids and Fecal Microbiota Abundance in Humans with Obesity: A Systematic Review and Meta-Analysis. *Nutrients*, *11*(10), E2512. https://doi.org/10.3390/nu11102512

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(Database issue), D515–D522. https://doi.org/10.1093/nar/gkv1049

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), e1. https://doi.org/10.1093/nar/gks808

Kodama, Y., Shumway, M., Leinonen, R., & International Nucleotide Sequence Database Collaboration. (2012). The Sequence Read Archive: Explosive growth of sequencing data. *Nucleic Acids Research*, *40*(Database issue), D54-56. https://doi.org/10.1093/nar/gkr854

Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., Britt, E. B., Fu, X., Wu, Y., Li, L., Smith, J. D., DiDonato, J. A., Chen, J., Li, H., Wu, G. D., Lewis, J. D., Warrier, M., Brown, J. M., Krauss, R. M., … Hazen, S. L. (2013). Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Medicine*, *19*(5), 576–585. https://doi.org/10.1038/nm.3145

Konstantinidis, K. T., Ramette, A., & Tiedje, J. M. (2006). The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1475), 1929–1940. https://doi.org/10.1098/rstb.2006.1920

Kumar, M., Ji, B., Babaei, P., Das, P., Lappa, D., Ramakrishnan, G., Fox, T. E., Haque, R., Petri, W. A., Bäckhed, F., & Nielsen, J. (2018). Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling. *Metabolic Engineering*, *49*, 128–142. https://doi.org/10.1016/j.ymben.2018.07.018

Lamichhane, S., Sen, P., Dickens, A. M., Orešič, M., & Bertram, H. C. (2018). Gut metabolome meets microbiome: A methodological perspective to understand the relationship between host and microbe. *Methods*, *149*, 3–12. https://doi.org/10.1016/j.ymeth.2018.04.029

Larsen, P. E., Collart, F. R., Field, D., Meyer, F., Keegan, K. P., Henry, C. S., McGrath, J., Quinn, J., & Gilbert, J. A. (2011). Predicted Relative Metabolomic Turnover (PRMT): Determining metabolic turnover from a coastal marine metagenomic dataset. *Microbial Informatics and Experimentation*, *1*, 4. https://doi.org/10.1186/2042-5783-1-4

Latorre-Pérez, A., Villalba-Bermell, P., Pascual, J., & Vilanova, C. (2020). Assembly methods for nanopore-based metagenomic sequencing: A comparative study. *Scientific Reports*, *10*(1), 13588. https://doi.org/10.1038/s41598-020-70491-3

Lee-Sarwar, K. A., Lasky-Su, J., Kelly, R. S., Litonjua, A. A., & Weiss, S. T. (2020). Metabolome–Microbiome Crosstalk and Human Disease. *Metabolites*, *10*(5), 181. https://doi.org/10.3390/metabo10050181

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., … Wang, J. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology*, *32*(8), 834–841. https://doi.org/10.1038/nbt.2942

Li, Z., Xia, J., Jiang, L., Tan, Y., An, Y., Zhu, X., Ruan, J., Chen, Z., Zhen, H., Ma, Y., Jie, Z., Xiao, L., Yang, H., Wang, J., Kristiansen, K., Xu, X., Jin, L., Nie, C., Krutmann, J., … Wang, J. (2021). Characterization of the human skin resistome and identification of two microbiota cutotypes. *Microbiome*, *9*(1), 47. https://doi.org/10.1186/s40168-020-00995-7

Liebal, U. W., Phan, A. N. T., Sudhakar, M., Raman, K., & Blank, L. M. (2020). Machine Learning Applications for Mass Spectrometry-Based Metabolomics. *Metabolites*, *10*(6), 243. https://doi.org/10.3390/metabo10060243

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., … Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, *569*(7758), 655–662. https://doi.org/10.1038/s41586-019-1237-9

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, *550*(7674), 61–66. https://doi.org/10.1038/nature23889

Lu, H.-P., Lai, Y.-C., Huang, S.-W., Chen, H.-C., Hsieh, C., & Yu, H.-T. (2014). Spatial heterogeneity of gut microbiota reveals multiple bacterial communities with distinct characteristics. *Scientific Reports*, *4*(1), 6185. https://doi.org/10.1038/srep06185

Luo, C., Knight, R., Siljander, H., Knip, M., Xavier, R. J., & Gevers, D. (2015). ConStrains identifies microbial strains in metagenomic datasets. *Nature Biotechnology*, *33*(10), 1045–1052. https://doi.org/10.1038/nbt.3319

Ma, B., France, M. T., Crabtree, J., Holm, J. B., Humphrys, M. S., Brotman, R. M., & Ravel, J. (2020). A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nature Communications*, *11*(1), 940. https://doi.org/10.1038/s41467-020-14677-3

Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, *46*(15), 7542–7553. https://doi.org/10.1093/nar/gky537

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P., & Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut*, *63*(8), 1275–1283. https://doi.org/10.1136/gutjnl-2013-304833

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., Greenhalgh, K., Jäger, C., Baginska, J., Wilmes, P., Fleming, R. M. T., & Thiele, I. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, *35*(1), 81–89. https://doi.org/10.1038/nbt.3703

Mallick, H., Franzosa, E. A., Mclver, L. J., Banerjee, S., Sirota-Madi, A., Kostic, A. D., Clish, C. B., Vlamakis, H., Xavier, R. J., & Huttenhower, C. (2019). Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. *Nature Communications*, *10*, 3136. https://doi.org/10.1038/s41467-019-10927-1

Mallick, H., Ma, S., Franzosa, E. A., Vatanen, T., Morgan, X. C., & Huttenhower, C. (2017). Experimental design and quantitative analysis of microbial community multiomics. *Genome Biology*, *18*(1), 228. https://doi.org/10.1186/s13059-017-1359-z

Mardinoglu, A., Shoaie, S., Bergentall, M., Ghaffari, P., Zhang, C., Larsson, E., Bäckhed, F., & Nielsen, J. (2015). The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Molecular Systems Biology*, *11*(10), 834. https://doi.org/10.15252/msb.20156487

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Goodwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature*, *437*(7057), 376–380. https://doi.org/10.1038/nature03959

Marshall, D. D., & Powers, R. (2017). Beyond the Paradigm: Combining Mass Spectrometry and Nuclear Magnetic Resonance for Metabolomics. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *100*, 1–16. https://doi.org/10.1016/j.pnmrs.2017.01.001

Matsuda, F. (2014). Rethinking Mass Spectrometry-Based Small Molecule Identification Strategies in Metabolomics. *Mass Spectrometry*, *3*(Spec Iss 2), S0038. https://doi.org/10.5702/massspectrometry.S0038

Morton, J. T., Aksenov, A. A., Nothias, L. F., Foulds, J. R., Quinn, R. A., Badri, M. H., Swenson, T. L., Van Goethem, M. W., Northen, T. R., Vazquez-Baeza, Y., Wang, M., Bokulich, N. A., Watters, A., Song, S. J., Bonneau, R., Dorrestein, P. C., & Knight, R. (2019). Learning representations of microbe–metabolite interactions. *Nature Methods*, *16*(12), 1306–1314. https://doi.org/10.1038/s41592-019-0616-3

Muller, E., Algavi, Y. M., & Borenstein, E. (2021). A meta-analysis study of the robustness and universality of gut microbiome-metabolome associations. *Microbiome*, *9*, 203. https://doi.org/10.1186/s40168-021-01149-z

Nearing, J. T., Comeau, A. M., & Langille, M. G. I. (2021). Identifying biases and their potential solutions in human microbiome studies. *Microbiome*, *9*, 113. https://doi.org/10.1186/s40168-021-01059-0

Noecker, C., Chiu, H.-C., McNally, C. P., & Borenstein, E. (2019). Defining and Evaluating Microbial Contributions to Metabolite Variation in Microbiome-Metabolome Association Studies. *MSystems*, *4*(6), e00579-19. https://doi.org/10.1128/mSystems.00579-19

Noecker, C., Eng, A., & Borenstein, E. (2021). *MIMOSA2: A metabolic network-based tool for inferring mechanism-supported relationships in microbiome-metabolome data* (p. 2021.09.14.459910). https://doi.org/10.1101/2021.09.14.459910

Noecker, C., Eng, A., Srinivasan, S., Theriot, C. M., Young, V. B., Jansson, J. K., Fredricks, D. N., & Borenstein, E. (2016). Metabolic Model-Based Integration of Microbiome Taxonomic and Metabolomic Profiles Elucidates Mechanistic Links between Ecological and Metabolic Variation. *MSystems*, *1*(1), e00013-15. https://doi.org/10.1128/mSystems.00013-15

Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A. D., Krecke, M., Merten, D., Haraldsdóttir, H. S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D. A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., … Thiele, I. (2019). The Virtual Metabolic Human database: Integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, *47*(D1), D614–D624. https://doi.org/10.1093/nar/gky992

O'Brien, E. J., Monk, J. M., & Palsson, B. O. (2015). Using Genome-Scale Models to Predict Biological Capabilities. *Cell*, *161*(5), 971–987. https://doi.org/10.1016/j.cell.2015.05.019

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Parada Venegas, D., De la Fuente, M. K., Landskron, G., González, M. J., Quera, R., Dijkstra, G., Harmsen, H. J. M., Faber, K. N., & Hermoso, M. A. (2019). Short Chain Fatty Acids (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Frontiers in Immunology*, *10*, 277. https://doi.org/10.3389/fimmu.2019.00277

Pedersen, H. K., Gudmundsdottir, V., Nielsen, H. B., Hyotylainen, T., Nielsen, T., Jensen, B. A. H., Forslund, K., Hildebrand, F., Prifti, E., Falony, G., Le Chatelier, E., Levenez, F., Doré, J., Mattila, I., Plichta, D. R., Pöhö, P., Hellgren, L. I., Arumugam, M., Sunagawa, S., … Pedersen, O. (2016). Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature*, *535*(7612), 376–381. https://doi.org/10.1038/nature18646

Peng, L., Li, Z.-R., Green, R. S., Holzman, I. R., & Lin, J. (2009). Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *The Journal of Nutrition*, *139*(9), 1619–1625. https://doi.org/10.3945/jn.109.104638

Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., van Doorn, L.-J., Knetsch, C. W., & Figueiredo, C. (2019). Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. *Frontiers in Microbiology*, *10*, 1277. https://doi.org/10.3389/fmicb.2019.01277

Perry, R. J., Borders, C. B., Cline, G. W., Zhang, X.-M., Alves, T. C., Petersen, K. F., Rothman, D. L., Kibbey, R. G., & Shulman, G. I. (2016). Propionate Increases Hepatic Pyruvate Cycling and Anaplerosis and Alters Mitochondrial Metabolism. *The Journal of Biological Chemistry*, *291*(23), 12161–12170. https://doi.org/10.1074/jbc.M116.720631

Pignanelli, M., Just, C., Bogiatzi, C., Dinculescu, V., Gloor, G. B., Allen-Vercoe, E., Reid, G., Urquhart, B. L., Ruetz, K. N., Velenosi, T. J., & Spence, J. D. (2018). Mediterranean Diet Score: Associations with Metabolic Products of the Intestinal Microbiome, Carotid Plaque Burden, and Renal Function. *Nutrients*, *10*(6), 779. https://doi.org/10.3390/nu10060779

Pitt, J. J. (2009). Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews*, *30*(1), 19–34.

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*, 395. https://doi.org/10.1186/1471-2105-11-395

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: Their purpose and place. *Human Molecular Genetics*, *27*(R2), R234–R241. https://doi.org/10.1093/hmg/ddy177

Poyet, M., Groussin, M., Gibbons, S. M., Avila-Pacheco, J., Jiang, X., Kearney, S. M., Perrotta, A. R., Berdy, B., Zhao, S., Lieberman, T. D., Swanson, P. K., Smith, M., Roesemann, S., Alexander, J. E., Rich, S. A., Livny, J., Vlamakis, H., Clish, C., Bullock, K., … Alm, E. J. (2019). A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nature Medicine*, *25*(9), 1442–1452. https://doi.org/10.1038/s41591-019-0559-3

Prakash, T., & Taylor, T. D. (2012). Functional assignment of metagenomic data: Challenges and applications. *Briefings in Bioinformatics*, *13*(6), 711–727. https://doi.org/10.1093/bib/bbs033

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188–7196. https://doi.org/10.1093/nar/gkm864

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Ramezani, A., Nolin, T. D., Barrows, I. R., Serrano, M. G., Buck, G. A., Regunathan-Shenk, R., West, R. E., Latham, P. S., Amdur, R., & Raj, D. S. (2018). Gut Colonization with Methanogenic Archaea Lowers Plasma Trimethylamine N-oxide Concentrations in Apolipoprotein e−/− Mice. *Scientific Reports*, *8*, 14752. https://doi.org/10.1038/s41598-018-33018-5

Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, *19*, 90. https://doi.org/10.1186/s13059-018-1462-9

Rasko, D. A., Rosovitz, M. J., Myers, G. S. A., Mongodin, E. F., Fricke, W. F., Gajer, P., Crabtree, J., Sebaihia, M., Thomson, N. R., Chaudhuri, R., Henderson, I. R., Sperandio, V., & Ravel, J. (2008). The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates. *Journal of Bacteriology*, *190*(20), 6881–6893. https://doi.org/10.1128/JB.00619-08

Reiman, D., Layden, B. T., & Dai, Y. (2021). MiMeNet: Exploring microbiome-metabolome relationships using neural networks. *PLOS Computational Biology*, *17*(5), e1009021. https://doi.org/10.1371/journal.pcbi.1009021

Rolfsson, Ó., & Palsson, B. O. (2015). Decoding the jargon of bottom-up metabolic systems biology. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *37*(6), 588–591. https://doi.org/10.1002/bies.201400187

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467.

Sartor, R. B. (2008). Microbial Influences in Inflammatory Bowel Diseases. *Gastroenterology*, *134*(2), 577–594. https://doi.org/10.1053/j.gastro.2007.11.059

Scheperjans, F., Aho, V., Pereira, P. A. B., Koskinen, K., Paulin, L., Pekkonen, E., Haapaniemi, E., Kaakkola, S., Eerola-Rautio, J., Pohja, M., Kinnunen, E., Murros, K., & Auvinen, P. (2015). Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders: Official Journal of the Movement Disorder Society*, *30*(3), 350–358. https://doi.org/10.1002/mds.26069

Scheppach, W., Sommer, H., Kirchner, T., Paganelli, G.-M., Bartram, P., Christl, S., Richter, F., Dusel, G., & Kasper, H. (1992). Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. *Gastroenterology*, *103*(1), 51–56. https://doi.org/10.1016/0016-5085(92)91094-K

Scholz, M., Ward, D. V., Pasolli, E., Tolio, T., Zolfo, M., Asnicar, F., Truong, D. T., Tett, A., Morrow, A. L., & Segata, N. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nature Methods*, *13*(5), 435–438. https://doi.org/10.1038/nmeth.3802

Schorn, M. A., Verhoeven, S., Ridder, L., Huber, F., Acharya, D. D., Aksenov, A. A., Aleti, G., Moghaddam, J. A., Aron, A. T., Aziz, S., Bauermeister, A., Bauman, K. D., Baunach, M., Beemelmanns, C., Beman, J. M., Berlanga-Clavero, M. V., Blacutt, A. A., Bode, H. B., Boullie, A., … van der Hooft, J. J. J. (2021). A community resource for paired genomic and metabolomic data mining. *Nature Chemical Biology*, *17*(4), 363–368. https://doi.org/10.1038/s41589-020-00724-z

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted metabolomics strategies – Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Sharon, G., Cruz, N. J., Kang, D.-W., Gandal, M. J., Wang, B., Kim, Y.-M., Zink, E. M., Casey, C. P., Taylor, B. C., Lane, C. J., Bramer, L. M., Isern, N. G., Hoyt, D. W., Noecker, C., Sweredoski, M. J., Moradian, A., Borenstein, E., Jansson, J. K., Knight, R., … Mazmanian, S. K. (2019). Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice. *Cell*, *177*(6), 1600-1618.e17. https://doi.org/10.1016/j.cell.2019.05.004

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., & Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology*, *26*(9), 1003–1010. https://doi.org/10.1038/nbt.1487

Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., de Wouters, T., Juste, C., Rizkalla, S., Chilloux, J., Hoyles, L., Nicholson, J. K., Dore, J., Dumas, M. E., Clement, K., Bäckhed, F., & Nielsen, J. (2015). Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metabolism*, *22*(2), 320–331. https://doi.org/10.1016/j.cmet.2015.07.001

Silva, R. R. da, Dorrestein, P. C., & Quinn, R. A. (2015). Illuminating the dark matter in metabolomics. *Proceedings of the National Academy of Sciences*, *112*(41), 12549–12550. https://doi.org/10.1073/pnas.1516878112

Silva, Y. P., Bernardi, A., & Frozza, R. L. (2020). The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Frontiers in Endocrinology*, *11*, 25. https://doi.org/10.3389/fendo.2020.00025

Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., & Lê Cao, K.-A. (2019). DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*, *35*(17), 3055–3062. https://doi.org/10.1093/bioinformatics/bty1054

Singh, K., Gobert, A. P., Coburn, L. A., Barry, D. P., Allaman, M., Asim, M., Luis, P. B., Schneider, C., Milne, G. L., Boone, H. H., Shilts, M. H., Washington, M. K., Das, S. R., Piazuelo, M. B., & Wilson, K. T. (2019). Dietary Arginine Regulates Severity of Experimental Colitis and Affects the Colonic Microbiome. *Frontiers in Cellular and Infection Microbiology*, *9*, 66. https://doi.org/10.3389/fcimb.2019.00066

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59. https://doi.org/10.1002/cpmb.59

Smith, P. M., Howitt, M. R., Panikov, N., Michaud, M., Gallini, C. A., Bohlooly-Y, M., Glickman, J. N., & Garrett, W. S. (2013). The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science*, *341*(6145), 569–573. https://doi.org/10.1126/science.1241165

Snijders, A. M., Langley, S. A., Kim, Y.-M., Brislawn, C. J., Noecker, C., Zink, E. M., Fansler, S. J., Casey, C. P., Miller, D. R., Huang, Y., Karpen, G. H., Celniker, S. E., Brown, J. B., Borenstein, E., Jansson, J. K., Metz, T. O., & Mao, J.-H. (2016). Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nature Microbiology*, *2*, 16221. https://doi.org/10.1038/nmicrobiol.2016.221

Srinivasan, S., Morgan, M. T., Fiedler, T. L., Djukovic, D., Hoffman, N. G., Raftery, D., Marrazzo, J. M., & Fredricks, D. N. (2015). Metabolic signatures of bacterial vaginosis. *MBio*, *6*(2), e00204-15. https://doi.org/10.1128/mBio.00204-15

Srivastava, A., Kowalski, G. M., Callahan, D. L., Meikle, P. J., & Creek, D. J. (2016). Strategies for Extending Metabolomics Studies with Stable Isotope Labelling and Fluxomics. *Metabolites*, *6*(4), 32. https://doi.org/10.3390/metabo6040032

Staley, J. T., & Konopka, A. (1985). Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats. *Annual Review of Microbiology*, *39*(1), 321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*, *178*(3), 591–599.

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLOS Biology*, *13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Sung, J., Kim, S., Cabatbat, J. J. T., Jang, S., Jin, Y.-S., Jung, G. Y., Chia, N., & Kim, P.-J. (2017). Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis. *Nature Communications*, *8*(1), 15393. https://doi.org/10.1038/ncomms15393

Tan, J., McKenzie, C., Potamitis, M., Thorburn, A. N., Mackay, C. R., & Macia, L. (2014). The role of short-chain fatty acids in health and disease. *Advances in Immunology*, *121*, 91–119. https://doi.org/10.1016/B978-0-12-800100-4.00003-9

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010). The population genetics of commensal Escherichia coli. *Nature Reviews. Microbiology*, *8*(3), 207–217. https://doi.org/10.1038/nrmicro2298

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121. https://doi.org/10.1038/nprot.2009.203

Thiele, I., Sahoo, S., Heinken, A., Hertel, J., Heirendt, L., Aurich, M. K., & Fleming, R. M. (2020). Personalized whole-body models integrate metabolism, physiology, and the gut microbiome. *Molecular Systems Biology*, *16*(5), e8982. https://doi.org/10.15252/msb.20198982

Tierney, B. T., Yang, Z., Luber, J. M., Beaudin, M., Wibowo, M. C., Baek, C., Mehlenbacher, E., Patel, C. J., & Kostic, A. D. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. *Cell Host & Microbe*, *26*(2), 283-295.e8. https://doi.org/10.1016/j.chom.2019.07.008

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., & Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, *12*(10), 902–903. https://doi.org/10.1038/nmeth.3589

Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., & Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Research*, *27*(4), 626–638. https://doi.org/10.1101/gr.216242.116

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ*, *361*, k2179. https://doi.org/10.1136/bmj.k2179

van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op den Camp, H. J. M., Kartal, B., Jetten, M. S. M., & Lücker, S. (2015). Complete nitrification by a single microorganism. *Nature*, *528*(7583), 555–559. https://doi.org/10.1038/nature16459

Van Treuren, W., & Dodd, D. (2020). Microbial Contribution to the Human Metabolome: Implications for Health and Disease. *Annual Review of Pathology: Mechanisms of Disease*, *15*(1), 345–369. https://doi.org/10.1146/annurev-pathol-020117-043559

Vich Vila, A., Imhann, F., Collij, V., Jankipersadsing, S. A., Gurry, T., Mujagic, Z., Kurilshikov, A., Bonder, M. J., Jiang, X., Tigchelaar, E. F., Dekens, J., Peters, V., Voskuil, M. D., Visschedijk, M. C., van Dullemen, H. M., Keszthelyi, D., Swertz, M. A., Franke, L., Alberts, R., … Weersma, R. K. (2018). Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. *Science Translational Medicine*, *10*(472), eaap8914. https://doi.org/10.1126/scitranslmed.aap8914

Visconti, A., Le Roy, C. I., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., Li, W., de Rinaldis, E., Bell, J. T., Venter, J. C., Nelson, K. E., Spector, T. D., & Falchi, M. (2019). Interplay between the human gut microbiome and host metabolism. *Nature Communications*, *10*(1), 4505. https://doi.org/10.1038/s41467-019-12476-z

Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., & Kerkhoven, E. J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLOS Computational Biology*, *14*(10), e1006541. https://doi.org/10.1371/journal.pcbi.1006541

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wang, Z., Klipfell, E., Bennett, B. J., Koeth, R., Levison, B. S., DuGar, B., Feldstein, A. E., Britt, E. B., Fu, X., Chung, Y.-M., Wu, Y., Schauer, P., Smith, J. D., Allayee, H., Tang, W. H. W., DiDonato, J. A., Lusis, A. J., & Hazen, S. L. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*, *472*(7341), 57–63. https://doi.org/10.1038/nature09922

Wang, Z., Roberts, A. B., Buffa, J. A., Levison, B. S., Zhu, W., Org, E., Gu, X., Huang, Y., Zamanian-Daryoush, M., Culley, M. K., DiDonato, A. J., Fu, X., Hazen, J. E., Krajcik, D., DiDonato, J. A., Lusis, A. J., & Hazen, S. L. (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. *Cell*, *163*(7), 1585–1595. https://doi.org/10.1016/j.cell.2015.11.055

Wemheuer, F., Taylor, J. A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., & Wemheuer, B. (2020). Tax4Fun2: Prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environmental Microbiome*, *15*, 11. https://doi.org/10.1186/s40793-020-00358-7

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Järnerot, G., Engstrand, L., Tysk, C., & Jansson, J. K. (2009). Twin Studies Reveal Specific Imbalances in the Mucosaassociated Microbiota of Patients with Ileal Crohn's Disease. *Inflammatory Bowel Diseases*, *15*(5), 653–660. https://doi.org/10.1002/ibd.20783

Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.-A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., … Querengesser, L. (2007). HMDB: The Human Metabolome Database. *Nucleic Acids Research*, *35*(suppl_1), D521–D526. https://doi.org/10.1093/nar/gkl923

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics*, *1*(1), 92–107.

Xie, H., Yang, C., Sun, Y., Igarashi, Y., Jin, T., & Luo, F. (2020). PacBio Long Reads Improve Metagenomic Assemblies, Gene Catalogs, and Genome Binning. *Frontiers in Genetics*, *11*, 1077. https://doi.org/10.3389/fgene.2020.516269

Xu, J., Zhang, Q.-F., Zheng, J., Yuan, B.-F., & Feng, Y.-Q. (2019). Mass spectrometry-based fecal metabolome analysis. *TrAC Trends in Analytical Chemistry*, *112*, 161–174. https://doi.org/10.1016/j.trac.2018.12.027

Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., Hosoda, F., Rokutan, H., Matsumoto, M., Takamaru, H., Yamada, M., Matsuda, T., Iwasaki, M., Yamaji, T., Yachida, T., … Yamada, T. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*, *25*(6), 968–976. https://doi.org/10.1038/s41591-019-0458-7

You, Y., Liang, D., Wei, R., Li, M., Li, Y., Wang, J., Wang, X., Zheng, X., Jia, W., & Chen, T. (2019). Evaluation of metabolite-microbe correlation detection methods. *Analytical Biochemistry*, *567*, 106–111. https://doi.org/10.1016/j.ab.2018.12.008

Zhou, B., Xiao, J. F., Tuli, L., & Ressom, H. W. (2012). LC-MS-based metabolomics. *Molecular BioSystems*, *8*(2), 470–481. https://doi.org/10.1039/c1mb05350g

Zierer, J., Jackson, M. A., Kastenmüller, G., Mangino, M., Long, T., Telenti, A., Mohney, R. P., Small, K. S., Bell, J. T., Steves, C. J., Valdes, A. M., Spector, T. D., & Menni, C. (2018). The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics*, *50*(6), 790–795. https://doi.org/10.1038/s41588-018-0135-7

Zomorrodi, A. R., & Maranas, C. D. (2012). OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. *PLOS Computational Biology*, *8*(2), e1002363. https://doi.org/10.1371/journal.pcbi.1002363

Zorrilla, F., Buric, F., Patil, K. R., & Zelezniak, A. (2021). metaGEM: Reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Research*, *gkab815*. https://doi.org/10.1093/nar/gkab815

## APPENDIX

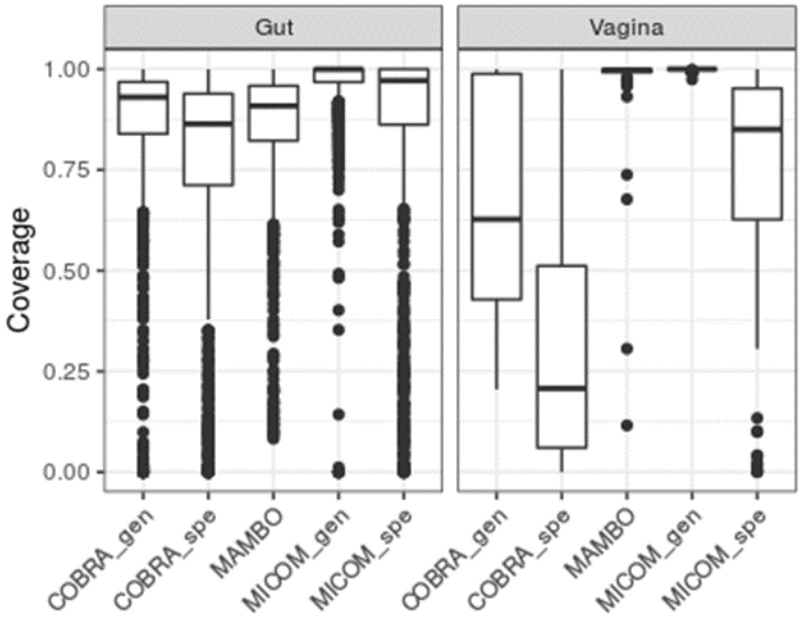



**Figure 9. Proportion of reads in samples represented by GEMs across tools.**

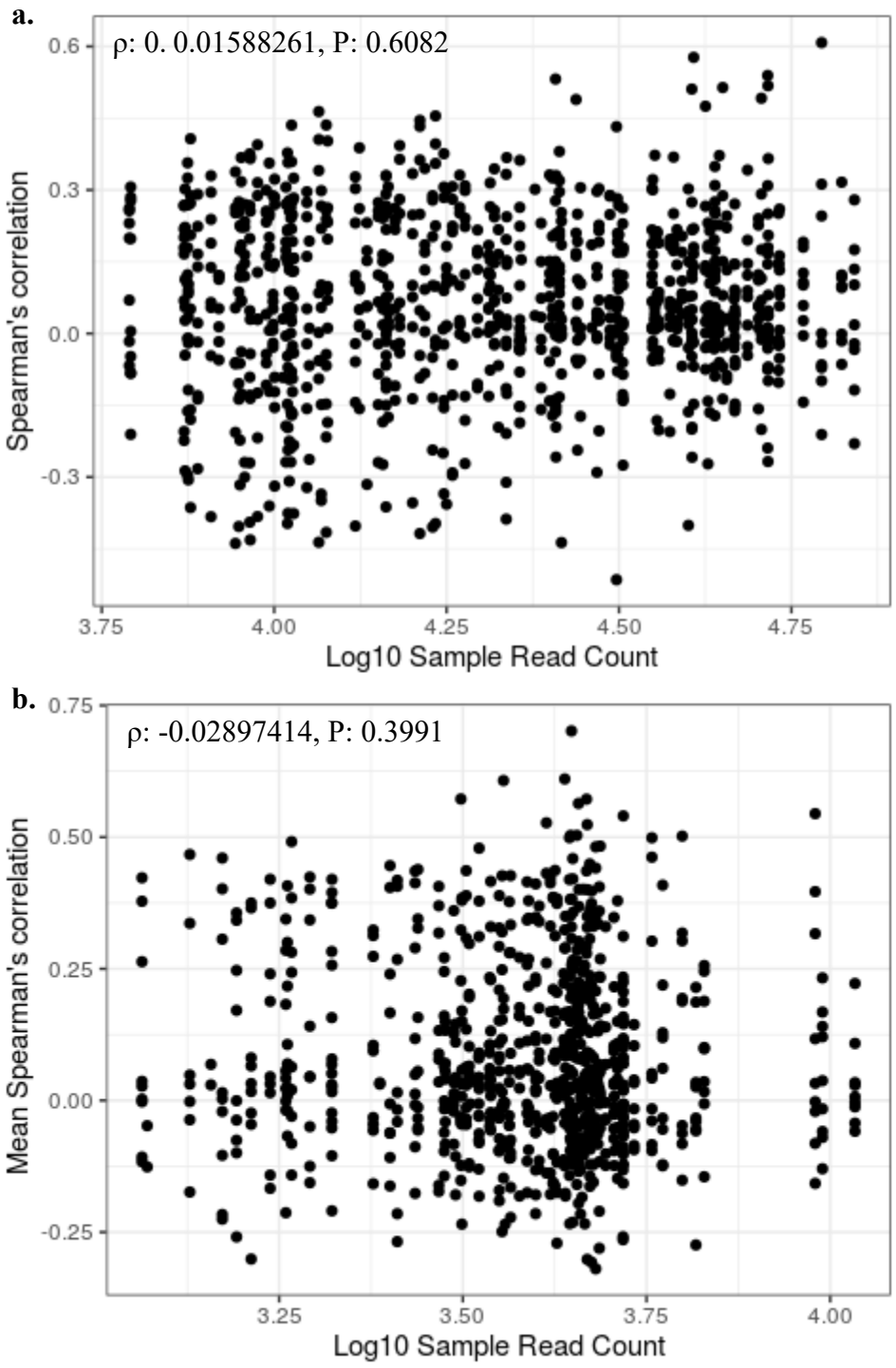



**Figure 10. Non-significant associations between Spearman's correlation coefficients and sample read count.** No significant correlation was found between these two variables for 16S rRNA gene sequenced samples from a. the gut or b. vaginal microbiomes.