

Electronic Thesis and Dissertation Repository

8-8-2016 12:00 AM

Utilizing untargeted metabolomics to characterize microbial communities and identify biomarkers of an unhealthy state

Amy McMillan
The University of Western Ontario

Supervisor
Gregor Reid
The University of Western Ontario

Graduate Program in Microbiology and Immunology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Amy McMillan 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>

Recommended Citation

McMillan, Amy, "Utilizing untargeted metabolomics to characterize microbial communities and identify biomarkers of an unhealthy state" (2016). *Electronic Thesis and Dissertation Repository*. 3908.
<https://ir.lib.uwo.ca/etd/3908>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

The metabolism of microbial communities is extremely complex, having contributions from multiple species as well as the host. The metabolome (the complete set of detectable small molecules in a given environment) offers a window into the culmination of these events. The goal of this thesis was to apply metabolomics to improve our understanding of the metabolism of microbial communities, with specific focus on the vaginal microbiota.

A combination of analytical chemistry techniques were employed to profile the vaginal metabolome of women with a dysbiotic vaginal microbiota, termed Bacterial Vaginosis (BV). The vaginal metabolome was closely associated with bacterial diversity and women with BV had a distinct metabolic profile compared to healthy women (N= 131). A number of novel biomarkers were identified, the most sensitive and specific being gamma-hydroxybutyrate (GHB) and 2-hydroxyisovalerate (2HV). These biomarkers were validated in three independent cohorts of diverse geographical locations and ethnicities. Correlations between the microbiota and metabolome identified putative microbe-product relationships, including production of GHB by *Gardnerella vaginalis* which was confirmed *in vitro*. Combining these data with meta-transcriptome information, metabolites could be linked to specific transcripts and microbes with increased confidence. The fibronectin binding capabilities of *Lactobacillus iners*, the most prevalent species in the vagina, was also investigated and confirmed.

To extend the tools developed during investigations of the vaginal microbiota to other systems, a study of stool and plasma samples from children with severe acute malnutrition (SAM) was conducted. Although the stool microbiota and metabolome did not discriminate children with SAM from controls, a number of metabolites differed significantly in plasma. Most of these metabolites had not been associated with SAM previously, including oxylipins, 2C6-disaccharides, truncated fibrinopeptides, and heme. These metabolic perturbations provide novel insight into the pathogenesis of SAM, and could serve as predictors of mortality/recovery and enteropathy. This study also led to the development of a novel method to filter out salt cluster artefacts in LC-MS metabolomics data using mass defect filtering.

Collectively, these studies have demonstrated how analytical chemistry, computational biology and microbiology can be integrated to advance our understanding of the metabolism of the microbiome and identify novel biomarkers of disease.

Keywords

Metabolomics, microbiome, bacterial vaginosis, LC-MS, GC-MS, 16S rRNA gene sequencing, malnutrition

Co-Authorship Statement

The experiment and data analyses within this thesis were predominantly carried out by Amy McMillan with supervision by Gregor Reid, and additional input from Mark Sumarah and Gregory Gloor. The manuscripts presented within were primarily written by Amy McMillan. Exceptions are listed below.

Chapter 2: A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women.

Amy McMillan participated in study design, supervised patient recruitment and sample collection, performed sample processing, contributed to method development for metabolite profiling, analyzed and interpreted microbiome and metabolome data, and performed *in vitro* experiments. Stephen Rulisa participated in study design and coordinated patient recruitment and ethics approval. Mark Sumarah provided platforms for metabolite profiling, and advised on method development and analysis for metabolomics. Jean Macklaim participated in study design, and advised on data analysis. Justin Renaud participated in LC-MS method development and data analysis. Jordan Bisanz participated in acquisition of Tanzanian replication cohort and blinded samples. Gregory Gloor developed methods for integration of microbiome and metabolome data, and supervised data analysis. Gregor Reid conceived and designed the study, acquired funding and contributed to the manuscript.

Chapter 3: Integration of the vaginal metabolome and bacterial meta-transcriptome of Canadian women.

Amy McMillan and Jean Macklaim equally participated in study design, patient recruitment and sample collection. Amy McMillan performed sample processing for 16S and GC-MS analyses, and analyzed 16S, GC-MS and NMR data as well as correlations between meta-transcriptomic and metabolomic data. Jean Macklaim performed sample processing and analysis for 16S and RNA-seq data, and generated all related data and figures. Mark Sumarah provided the GC-MS platform. Jonathan Swann processed and ran

samples by NMR. Gregory Gloor supervised data analysis and developed methods to integrate the metabolome and transcriptome. Gregor Reid conceived and designed the study, acquired funding and contributed to the manuscript.

Chapter 4: Adhesion of *Lactobacillus iners* AB-1 to human fibronectin: a key mediator for persistence in the vagina?

Amy McMillan performed all experiments and wrote the manuscript. Jean Macklaim provided RNA-seq supporting data and contributed to manuscript generation. Jeremy Burton contributed to manuscript generation. Gregor Reid conceived and designed the study, acquired funding, and contributed to manuscript generation.

Chapter 5: Metabolic derangements identified through untargeted metabolomics in a cross-sectional study of Nigerian children with severe acute malnutrition

Amy McMillan developed GC-MS and LC-MS methods for stool and plasma, and analyzed and interpreted GC-MS, LC-MS, ELISA and microbiota data. Adebola Orimadegun supervised the field work, collection of clinical data and samples and clinical data management. Mark Sumarah provided platforms for metabolite profiling, and supervised GC-MS and LC-MS analyses. Justin Renaud participated in LC-MS method development and assisted in identification of metabolites by LC-MS/MS. Magdalena Muc assisted in interpretation of the results. Gregory Gloor supervised analysis of microbiota data. Gregor Reid set up the collaboration, acquired funding for metabolite standards and contributed to the manuscript. Stephen Allen and, Olusegan Akinyinka conceived and designed the project.

Chapter 6: Post-acquisition artifact filtering for LC-MS based human metabolomic studies

Amy McMillan conducted LC-MS experiments, and developed the R code. Justin Renaud conceptualized project. Gregory Gloor advised on method implementation. Gregor Reid co-supervised Amy McMillan. Mark Sumarah provided LC-MS platform.

Acknowledgments

First and foremost, I would like to thank my supervisor, Gregor Reid, for your continued support and guidance. I have been given so many opportunities under your supervision that are unique among PhD students. Only you would be daring enough to send a lone student to Rwanda to co-ordinate a clinical study. Looking back, this is one of the experiences that affected me most, instilling in me a new-found confidence in my abilities. Thank you for believing in me even when I did not believe in myself.

I also owe a great deal to Mark Sumarah, who not only granted me full access to his chemistry lab, but made me feel included as part of his research group and has become a great mentor to me. I really appreciate the interest you took in my projects, even though they were far removed from your field. Thanks for being open to the idea that someone with no prior chemistry expertise could take this on.

To Greg Gloor, for his extreme patience with my lack of computational skills at onset. This was a challenge I never thought I could achieve, much less enjoy. I can now say with confidence that the computational and statistical knowledge I acquired are among the most valuable obtained during my training, and that analyzing data is actually one of the things I enjoy most about research.

There are so many people in the Reid, Sumarah, Gloor, and Burton labs that deserve acknowledgements, and you have all made a significant impact on my life in some way. To Jean Macklaim for her advice on science and life. To Justin Renaud for taking the time to share his chemistry expertise with me. And finally to Jordan Bisanz for being a great friend and host of our regular lab “bro-BQs”.

I’d also like to thank my family for their continued support, even when my area of research made you uncomfortable (you never said it, but I’m sure it did). And finally to Marc Monachese for agreeing to go on this journey with me from across the pond. You have been such a great support system to me and I couldn’t have done it without you.

Table of Contents

Abstract.....	i
Co-Authorship Statement.....	iii
Acknowledgments.....	v
Table of Contents.....	vi
List of Tables.....	xi
List of Figures.....	xii
List of Appendices.....	xv
Chapter 1.....	1
1 General Introduction.....	1
1.1 The human microbiome.....	1
1.2 Tools for studying microbial communities.....	1
1.3 The vaginal microbiome.....	4
1.3.1 Healthy vaginal microbiota.....	4
1.3.2 Bacterial vaginosis.....	5
1.4 Metabolomics.....	8
1.4.1 GC-MS.....	11
1.4.2 LC-MS.....	14
1.5 Using metabolomics to characterize microbial communities.....	17
1.6 Scope and objectives.....	18
1.7 References.....	20
Chapter 2.....	33
2 A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women.....	33
2.1 Introduction.....	33

2.2	Materials and methods	34
2.2.1	Clinical samples	34
2.2.2	Microbiome profiling.....	35
2.2.3	Sample preparation GC-MS.....	37
2.2.4	Data processing GC-MS	37
2.2.5	Global metabolomic analysis.....	38
2.3	Sample preparation LC-MS	39
2.3.1	Data processing LC-MS.....	40
2.3.2	Validation in blinded replication cohort	41
2.3.3	Identification of putative GHB dehydrogenases in <i>G. vaginalis</i>	41
2.3.4	<i>In vitro</i> extraction of GHB from vaginal isolates	42
2.4	Results.....	42
2.4.1	The vaginal metabolome is most correlated with bacterial diversity ..	42
2.4.2	Metabolites and taxa associated with diversity.....	45
2.4.3	Succinate is not associated with diversity or clinical BV, and is produced by <i>L. crispatus</i>	47
2.4.4	Metabolites associated with diversity are sensitive and specific for clinical BV	49
2.4.5	Validation of biomarkers in a blinded replication cohort from Tanzania.....	53
2.4.6	Identification of <i>G. vaginalis</i> as a producer of GHB	54
2.5	Discussion.....	56
2.6	References.....	59
	Chapter 3.....	68
3	Integration of the vaginal metabolome and bacterial meta-transcriptome of Canadian women	68
3.1	Introduction.....	68

3.2	Materials and methods	68
3.2.1	Study design and sample collection	68
3.2.2	DNA isolation, PCR amplification, and sequencing	69
3.2.3	16S sequence analysis.....	70
3.2.4	Meta-transcriptome analysis	70
3.2.5	Sample preparation GC-MS.....	71
3.2.6	Data analysis GC-MS.	72
3.2.7	Sample preparation NMR.	72
3.2.8	Correlations between transcriptome and metabolome	73
3.3	Results.....	73
3.3.1	The vaginal microbiota of Canadian women	73
3.3.2	The vaginal metabolome of Canadian women.....	70
3.3.3	The bacterial meta-transcriptome identifies subtypes of BV and health.....	73
3.3.4	Integrating the meta-transcriptome and metabolome	74
3.4	Discussion.....	79
3.5	References.....	81
Chapter 4	84
4	Adhesion of <i>Lactobacillus iners</i> AB-1 to human fibronectin: a key mediator for persistence in the vagina?	84
4.1	Introduction.....	84
4.2	Materials and methods	85
4.2.1	Strains and growth conditions.....	85
4.2.2	Fibronectin adhesion assay	87
4.2.3	Bacterial binding to fibronectin coated cover slips.....	88
4.2.4	Statistical analysis.....	88

4.3 Results.....	88
4.4 Discussion.....	92
4.5 References.....	94
Chapter 5.....	99
5 Metabolic derangements identified through untargeted metabolomics in a cross-sectional study of Nigerian children with severe acute malnutrition.....	99
5.1 Introduction.....	99
5.2 Materials and methods.....	100
5.2.1 Study design and sample collection.....	100
5.2.2 Metabolite extraction from stool and plasma.....	101
5.2.3 Untargeted GC-MS analyses.....	101
5.2.4 Untargeted LC-MS analyses.....	102
5.2.5 Statistical analysis of metabolome data.....	108
5.2.6 Calprotectin and lactoferrin ELISAs.....	108
5.2.7 Microbiome profiling.....	109
5.3 Results.....	110
5.3.1 Study population.....	110
5.3.2 The fecal microbiota of Nigerian children.....	112
5.3.3 Fecal metabolome and inflammatory markers do not distinguish SAM from controls.....	114
5.3.4 Children with SAM have a distinct plasma metabolome.....	115
5.3.5 Oxylipins, phospholipids and amino acids are depleted in the plasma of children with SAM.....	117
5.3.6 Metabolites elevated in plasma of children with SAM.....	115
5.3.7 Assessment of biomarkers to discriminate SAM from controls.....	117
5.4 Discussion.....	119

5.5	References.....	125
Chapter 6	133
6	Post-acquisition artifact filtering for LC-MS based human metabolomic studies	133
6.1	Findings	133
6.2	Materials and Methods.....	138
6.2.1	Metabolite extraction	138
6.2.2	LC-MS analyses.....	139
6.3	References.....	140
Chapter 7	143
7	General discussion	143
7.1	The vaginal metabolome in BV	143
7.2	Integration of metabolite and NGS data	145
7.3	The microbiome and metabolome in malnutrition.....	146
7.4	Improving upon metabolomics methods.....	146
7.5	Future directions	147
7.6	References.....	149

List of Tables

Table 4-1. Bacterial strains used in this study	86
Table 4-2. Bacterial adhesion to fibronectin coated cover slips cell counts	92
Table 5-1. Demographic and clinical characteristics of cases and controls.(1).....	111
Table 6-1. Percent data reduction after mass defect filtering alone or in combination with retention time and hmdb inclusion list. All features detected in plasma in the malnutrition dataset are shown in both positive and negative ionization mode. RT: retention time, md: mass defect.....	136

List of Figures

Figure 1-1. The vaginal microbiota assessed by 16S rRNA gene sequencing.	6
Figure 1-2. The morphology of lactobacilli are highly variable. Adapted from Veehelst et al 2005 ⁵³	7
Figure 1-3. An overview of the steps involved in analysis of metabolomics data ⁵⁷	10
Figure 1-4 Schematic of the Agilent MSD GC-MS system ⁷⁰	12
Figure 1-5. Example of metabolite (gamma-aminobutyrate) identification by GC-MS using the NIST database.	13
Figure 1-6. Thermo Q-Exactive Orbitrap configuration ⁷⁵	16
Figure 2-1. The vaginal metabolome is most correlated with bacterial diversity.....	44
Figure 2-2. Bacterial taxa and metabolites correlated with bacterial diversity in the vagina.	46
Figure 2-3. Relative abundance of succinate in women dominated by <i>L. crispatus</i> , <i>L. iners</i> , or Nugent BV detected by GC-MS.	48
Figure 2-4. Succinate production by vaginal isolates.	48
Figure 2-5. Comparison of biomarkers to identify Nugent BV from Nugent N.....	50
Figure 2-6. Biomarker cut points effectively group Nugent Intermediate samples as BV or N.....	52
Figure 2-7. Biomarker validation in a blinded replication cohort of 45 women from Tanzania.....	53
Figure 2-8. Correlations between metabolites and taxa which are robust to random sampling of the underlying data.....	55
Figure 2-9. GHB is produced by <i>Gardnerella vaginalis</i>	56

Figure 3-1. Taxonomic barplots and clustered heatmap of meta-transcriptome expression.	70
Figure 3-2. The vaginal metabolome of Canadian women analyzed by GC-MS.	71
Figure 3-3. The vaginal metabolome of Canadian women analyzed by NMR.	73
Figure 3-4. Correlation between metabolite abundance and microbial function.	75
Figure 3-5. Relative abundance of bacterial genus (<i>Megasphaera</i>) and transcript (lysine decarboxylase) predicted to be responsible for cadaverine synthesis in women with BV compared to health.	77
Figure 3-6. Relative abundance of bacterial species (<i>Gardnerella vaginalis</i>) and transcript (GHB dehydrogenase) predicted to be responsible for GHB synthesis in women with BV compared to health.	79
Figure 4-1. <i>L. iners</i> AB-1 grown in 10 ml MRS broth for 7 days.	86
Figure 4-2. Bacterial binding to immobilized fibronectin as measured by optical density at 600 nm (OD600).	89
Figure 4-3. Effect of proteinase K pre-treatment on adhesion to immobilized fibronectin by <i>L. iners</i> AB-1.	90
Figure 4-4. Bacterial adhesion to fibronectin coated cover slips.	91
Figure 5-1. LC-MS/MS spectra for metabolites elevated in plasma of children with SAM.	105
Figure 5-2. The stool microbiota does not discriminate SAM from controls.	113
Figure 5-3. Principal component analysis (PCA) scoreplots of stool metabolome from children with SAM and controls as determined by LC-MS using positive (top) and negative (bottom) electrospray ionization (ESI).	115

Figure 5-4. Concentration of inflammatory proteins in stool of children with SAM compared to controls as measured by ELISA.....	115
Figure 5-5. Principal component analysis (PCA) plots of plasma metabolome from children with SAM and controls as determined by LC-MS using positive (top) or negative (bottom) electrospray ionization (ESI).	116
Figure 5-6. Free amino acids/dipeptides (A), phospholipids (B), and oxylipins (C) significantly decreased in children with SAM compared to controls.	118
Figure 5-7. LC-MS/MS spectra examples for oxylpin features containing two or more species.....	114
Figure 5-8. Comparison of intact fibrinopeptide A (FPA) and B (FPB) in plasma of children with SAM compared to controls measured using positive ESI LC-MS.	116
Figure 5-9. Metabolites significantly elevated at least 2-fold in the plasma of children with SAM compared to controls (Wilcoxon test, FDR corrected $P < 0.1$).	117
Figure 5-10. Evaluation of biomarkers to identify SAM from controls.	119
Figure 6-1. Retention time and m/z of plasma metabolites differing significantly between children with severe acute malnutrition and controls	134
Figure 6-2. Mass defect as a function of m/z for all endogenous and food-derived compounds between 50-750 m/z in the human metabolome database (hmdb, $n=17957$) compared to features detected experimentally in plasma ($n=2227$).	136
Figure 6-3. Significant metabolites (Wilcoxon test, FDR corrected $P < 0.1$, > 2 fold change) remaining in the malnutrition dataset after A. no filter, B. mass defect filter only, C, mass defect filter with retention time.....	135
Figure 6-4. Effect of column type on salt cluster formation in plasma, urine and stool.	137

List of Appendices

Appendix A: Content License from Scientific Reports	158
Appendix B: Content License from Reproductive Sciences.....	159
Appendix C: Content License from PLoS ONE	160
Appendix D: Content License from BMC Microbiology	161
Appendix E: Content License from Frontiers in Bioengineering and Biotechnology	162
Appendix F: Ethics Approval for studies.....	163

Chapter 1

1 General Introduction

1.1 The human microbiome

The human microbiome can be considered our largest endocrine organ, with the abundance of approximately 10% of metabolites detected in blood significantly altered by microbial colonization¹. Indeed, the number of microbial cells by far outnumbers the number of human cells, and the microbiome possess ~150 times more genes than humans². Due to recent advances in next generation sequencing and analytical chemistry techniques, we are just now beginning to understand the contribution of microbial metabolism towards human health.

1.2 Tools for studying microbial communities

The most used tool to survey microbial community composition is 16S rRNA gene sequencing. The 16S ribosomal RNA is a component of the 30S small ribosomal subunit in bacteria and archaea. Woese and Fox³ were the first to use the rRNA gene to infer the phylogenetic relationship between organisms, using this approach to define the three domains of life. Due to its slow rate of evolution, this gene is highly conserved between species⁴ and therefore can be amplified via universal 16S primers. The presence of hypervariable regions within the 16S gene allows microbes to be distinguished from one-another, most commonly down to the genus level, with speciation possible in some cases. Thus 16S primers universally amplify bacterial and archaeal DNA, while allowing some level of taxonomic classification of the microbes in a given community.

The choice of hypervariable region depends upon the environment being surveyed. The V6 region is commonly applied to the vaginal environment as speciation of *Lactobacillus*, the most common genus in the vagina, is possible⁵⁻⁸. Conversely, the V4 region has been shown to most accurately capture the microbial diversity of the gut^{9,10}. Alternative approaches attempt to amplify two variable regions across a constant region

such as V4-V5, however these approaches are not ideal due to a high rate of chimeric reads composed of two different genomes¹¹.

There are numerous next-generation sequencing (NGS) platforms suitable for 16S sequencing purposes, with improvements being made continuously. The Illumina series of machines have recently become leaders in the field due to their unparalleled read depth, low error rate and ever increasing read length capabilities¹². The Illumina platforms amplify DNA by incorporating adaptor sequences into their primer design, which are complementary to sequences on the Illumina flow cells¹³. Once DNA has annealed to the flow cells, PCR is then conducted by bridge amplification and sequences determined by incorporation of fluorescent nucleotides. To allow a large number of communities to be sequenced in a single NGS run, sample specific barcodes are also incorporated into the primer design¹⁴. With the advancement of this multiplexing technique, the number of samples that can be analyzed in a single run is limited only by read depth (the number of DNA amplicons that can be sequenced in a single run).

Analyzing the data resulting from 16S experiments requires considerable computational expertise, and therefore a number of software packages have been developed specifically for 16S data. The most widely used is QIIME (Quantitative Insights into Microbial Ecology)¹⁵. QIIME is a Python-based tool that provides a workflow for taxonomy assignment, statistical analysis and plotting of 16S data. Alternatively, many groups choose to develop in house pipelines to suit their specific needs.

After data are acquired and sample specific barcodes have been removed, sequences are conventionally clustered by similarity into operational taxonomic units or OTUs. The percent cutoff for two sequences to be considered a single OTU is user defined, and is often set to 97% to allow speciation¹⁶. Taxonomy is then assigned based on sequence similarity using a variety of databases including the Ribosomal database project (<https://rdp.cme.msu.edu/>)¹⁷, Greengenes (<http://greengenes.lbl.gov/>)¹⁸, and/or Silva (<http://www.arb-silva.de/>)¹⁹, in addition to in-house curated databases.

To describe the composition of the microbiome, the concepts of alpha and beta diversity are commonly applied. Alpha diversity describes the diversity within a sample, and can be calculated by a number of different methods, the most common being the Shannon index, which takes into account both the number of OTUs present in a community and their evenness²⁰. Conversely, beta diversity describes the diversity between microbial communities. Again there are a variety of methods employed to calculate beta diversity. The method used most often, developed and available in QIIME, is UniFrac. UniFrac is a distance metric that relies on both the taxonomic relationships between OTUs and the relative abundance of taxa within a sample (weighted UniFrac) to calculate the distances of samples from one another²¹. These distances can be visualized in 2D space in the form of phylogenetic trees, or in 3D space through the use of principal coordinate plots (PCoA). The separation of samples based on these trees or PCoA plots can then be used to infer the relationship between microbial composition and a given parameter, for example a disease state.

Although UniFrac provides a useful way to visualize microbiome data, it is a relatively qualitative tool and provides little information about the specific organism(s) differing between a healthy and diseased state, for example. To answer this question, univariate statistics are required. This aim is non-trivial due to the compositional nature of NGS data, to which classical univariate statistics cannot be applied²²⁻²⁴. Compositional data is by definition constrained by an arbitrary constant sum, and therefore can only be expressed as proportions of this sum²². For example, a given NGS run will return a certain number of reads for each sample. This number on its own provides no biological information, as read number can be influenced by a variety of factors including the PCR efficiency and NGS platform read depth in addition to the original concentration of DNA in the sample^{24,25}. Because all OTUs can only be expressed as a proportion of the total reads, no OTU is truly independent from another, rendering classical statistical methods invalid²². Various approaches have been undertaken to circumvent the compositional issue in microbiome data. The most widely used method is rarefaction, which normalizes the per sample read count to that of the sample with the fewest reads. In addition to discarding much of the reliable information from the experiment, this method results in a

high number of false positives²⁶. Because most microbiome studies do not treat NGS as compositional, the literature is rife with irreproducible association between the microbiome and disease^{24,27}.

An alternative approach is to apply methods developed specifically for compositional data stemming from geology where much of the data is also compositional^{24,28}. Gloor *et al*²⁴ have adapted these methods specifically for NGS data to create a multi-purpose tool for statistical analysis of compositional sequencing data called ALDEx2 (ANOVA-Like Differential Expression)^{29,30}. ALDEx2 applies the centered log ratio (clr) transformation ($[\ln(x/g(x))]$), which renders values independent from one another and therefore appropriate for classical univariate statistics. It also treats the data as probability distributions rather than point estimates as the data in actuality represents only a random sampling of sequences in an environment, which vary upon replication. Finally, ALDEx2 incorporates corrections for multiple testing, which are often overlooked by other methods. The ALDEx2 approach is applied throughout this thesis to identify differences in microbial communities between disease states as well as to evaluate correlations between the metabolome and NGS data.

1.3 The vaginal microbiome

1.3.1 Healthy vaginal microbiota

A healthy vaginal microbiota is dominated by *Lactobacillus* species⁵⁻⁸, which possess a number of properties that benefit the host. Lactic acid, the main end product of anaerobic respiration by these organisms, lowers the pH of the vagina and protects against infections by pathogens^{31,32}. In addition these bacteria can produce bacteriocins³³ and hydrogen peroxide^{34,35} to further combat invading organisms, although the *in vivo* plausibility of the latter has been disputed³². Unlike the gut, where low microbial diversity is associated with diseases such as inflammatory bowel disease and obesity³⁶⁻³⁸, in the vagina low diversity is a hallmark of health⁵⁻⁷. Indeed, most women are dominated by a single *Lactobacillus* species, the most common being *L. iners* followed by *L. crispatus*. These species are uniquely adapted to the human vaginal tract and predominate irrespective of ethnicity,

geographical location, or pregnancy status⁵⁻⁸. *L. iners* is particularly unique, having the smallest genome of any *Lactobacillus* species reported to date, and possessing several components which make it ideally suited to the vaginal niche³⁹. Despite its small genomic capacity, *L. iners* is able to persist fluctuations in community composition that occur during a condition known as bacterial vaginosis (BV), while *L. crispatus* does not⁵⁻⁸.

1.3.2 Bacterial vaginosis

Bacterial vaginosis is the most common vaginal condition in women, affecting an estimated 30% at any given time⁴⁰. This condition is characterized by an increase in microbial diversity with a concomitant decrease in the proportion and absolute abundance of lactobacilli^{5-8,41}. The loss of lactic acid producing bacteria results in increased pH of the vagina along with increased risk of sexually transmitted infections such as HIV, gonorrhea, and chlamydia⁴²⁻⁴⁵. Women with BV are also more likely to experience pre-term labor, believed to be due in part to the immune response to these organisms^{46,47}. Although some but not all women experience symptoms such as malodor due to production of biogenic amines⁴⁸⁻⁵⁰, many women with BV are asymptomatic^{5,7,40,51}, making detection reliant on signs or symptoms unreliable.

BV is a polymicrobial disease, meaning no single organisms causes the condition, rather it is the complex interaction of multiple species that causes dysbiosis. The most common bacteria associated with BV belong to the genus *Gardnerella*, *Prevotella*, *Atopobium*, *Megasphaera*, *Sneathia*, and *Dialister*⁵⁻⁸. As mentioned previously, *L. iners* is also found in many patients with BV, raising questions of whether some clones may play a role in the condition, while others may align more with health. The presence of *L. iners* may be important for recovery from BV, by being the lactobacilli that helps re-create a healthy environment³⁹.

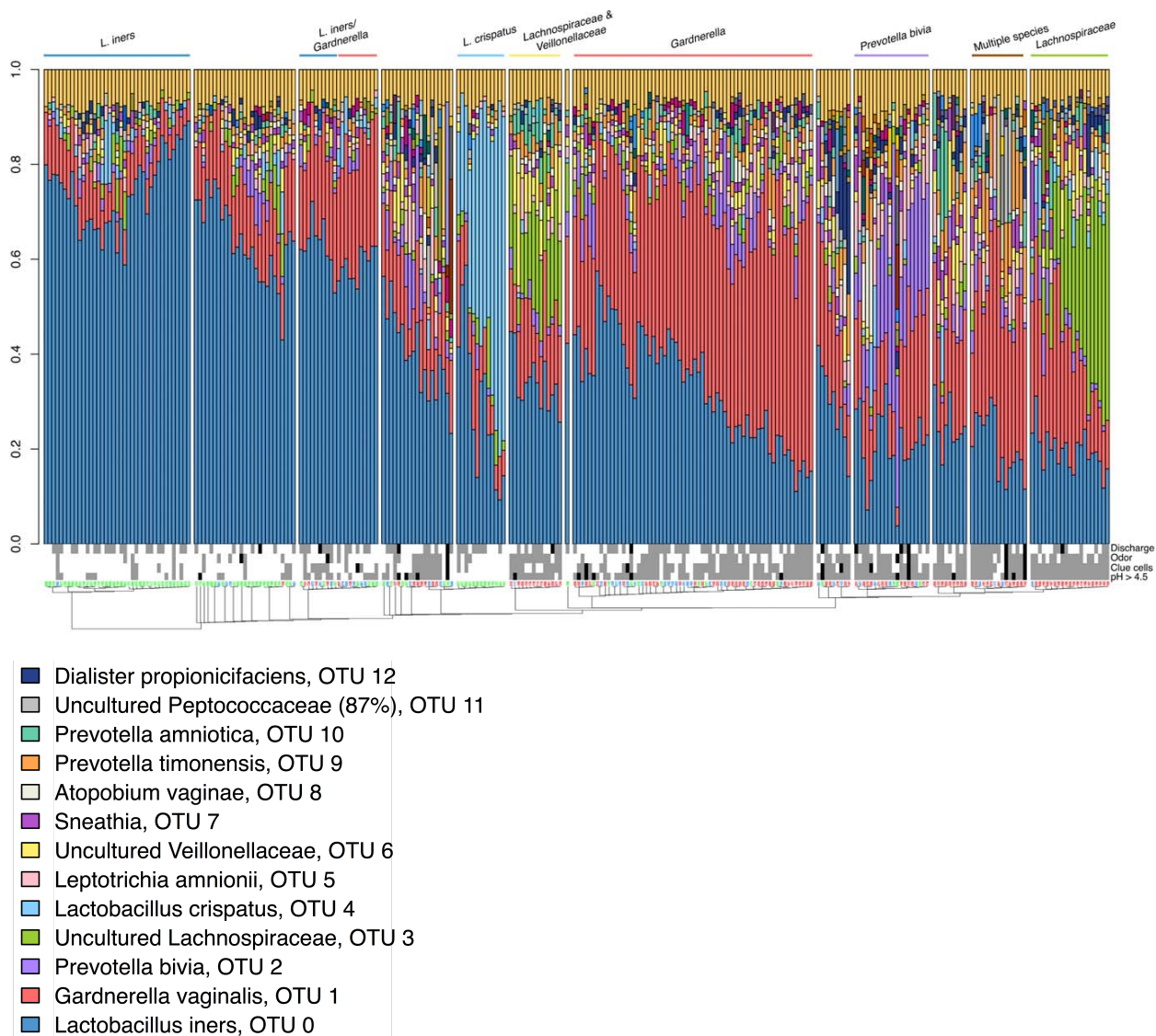


Figure 1-1. The vaginal microbiota assessed by 16S rRNA gene sequencing.

Each bar represents a single sample from a single woman and each color a different bacterial taxa. Samples were clustered by the similarity of the relative organism abundance and their similarity was visualized using a neighbor-joining tree. The major clusters are named after a taxa with a relative even abundance across samples in the cluster. The sample numbers are colored according to Nugent categories with BV = red, intermediate = green, normal = blue. Amsel criteria are shown for each sample with present = grey, absent = white and missing data = black. Figure taken from Hummelen et al 2010⁷.

Despite the prevalence of BV, the cause remains unknown. Furthermore, the diagnostic methods are imprecise, time consuming and have poor reliability. The current gold standard for BV diagnosis is known as the “Nugent Score”. This microscopy-based

technique defines BV as a score of 7-10 when low numbers of lactobacilli morphotypes are observed, and high numbers of short rods are present, which are presumed to represent BV associated bacteria⁵². Nugent Normal (N) is defined as a score of 1-3, indicating almost exclusively *Lactobacillus* morphotypes. Intermediate samples are given a score of 4-6 and do not fit into either group. As demonstrated in Figure 1-2, the high morphological variability of lactobacilli makes the identification of BV associated bacteria a non-trivial task. These analyses are also time consuming as they must be shipped to a central lab for processing.

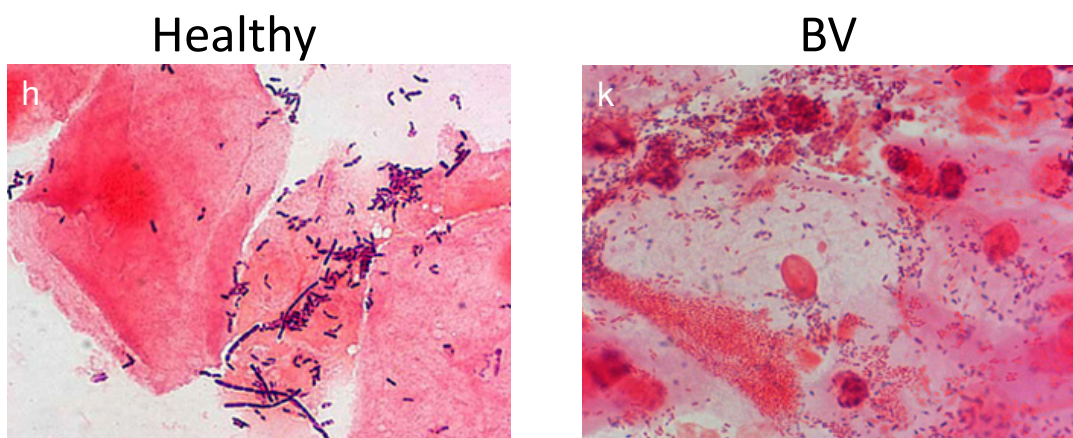


Figure 1-2. The morphology of lactobacilli are highly variable. Adapted from Veehelst et al 2005⁵³.

An alternative diagnostic approach was developed by Amsel *et al*⁵⁴ and is known as the Amsel criteria. This method relies on the presence of at least three of the following: i) vaginal pH > 4.5, ii) presence of “clue cells” (epithelial cells covered in dense biofilm of bacteria), iii) positive amine “whiff” test and iv) abnormal discharge. The Amsel criteria is equally poor in identifying BV as a large proportion of women with BV are asymptomatic^{5,7,40,51}.

Given the stark contrast in microbial composition of women with with BV compared to healthy women, one would expect the metabolic by-products of these communities to differ significantly. This has been further emphasized by meta-transcriptomic studies demonstrating significant differences in gene expression of the community in BV⁵⁵, suggesting the resulting small molecules could serve as accurate diagnostic biomarkers for the condition.

1.4 Metabolomics

The complete set of detectable small molecules in a given environment is known as the metabolome, and its study coined ‘metabolomics’ in alignment with other ‘omics’ technologies⁵⁶. There are a number of different analytical tools to survey the metabolome, the most common being nuclear magnetic resonance (NMR) and mass-spectrometry (MS)^{57,58}. The latter is usually coupled to a separation technique such as gas or liquid chromatography. Each method has its own advantages and disadvantages, and therefore a combination of techniques is ideal for most complete metabolite coverage⁵⁸.

NMR has been widely used for metabolomics due to its reproducibility and relative ease of compound identification^{57,58}. Also, unlike MS based platforms, NMR is non-destructive to the sample. However, the number of metabolites which can be resolved from a complex mixture by NMR is low in comparison to other technologies. This method also lacks the sensitivity of MS-based platforms, providing a survey of only the most abundant compounds in a sample.

Gas-chromatography mass spectrometry (GC-MS), like NMR, is also prevalent due to the fact that compound identification is fairly straightforward^{57,58}. GC-MS is also generally more sensitive than NMR. The caveat is that derivatization is usually required, which leads to degradation of samples over time. Sample preparation and data acquisition time for GC-MS is also quite long in comparison to its liquid chromatography counterpart.

The final platform employed most often in metabolomics is liquid-chromatography mass-spectrometry (LC-MS). This platform undoubtedly holds the most promise due to its unparalleled metabolite coverage and sensitivity^{57,58}. Many different classes of metabolites can be detected by this method, from small organic acids, to lipids and even small peptides. Due to advances in ultra-high performance liquid chromatography (UHPLC), the data acquisition time is also quite short in comparison to GC-MS, and derivatization is usually not required. However, the bottleneck in analysis of LC-MS data is compound identification. Due to the diversity of molecules, limited databases, and poor

reproducibility between different methods and platforms, identification remains a very challenging and time consuming task.

Regardless of the platform used, the workflow for data processing is similar between technologies, with the exception of the software, which differ between platforms and vendors (Figure 1-3). Briefly, samples are collected via a standardized approach and stored preferably at -80 °C. Metabolites are then extracted with various solvents depending on the classes of metabolites of interest. After extraction, data is acquired via NMR, GC-MS or LC-MS and a spectra (NMR) or chromatogram with corresponding MS spectra (GC-MS and LC-MS) is acquired for each sample. Data processing software are then used to combine files into a feature matrix containing the abundance of every feature in every sample. The definition of a “feature” differs for each platform. For NMR, spectra are usually divided into spectral bins, so each feature represents a ‘slice’ of the NMR spectra. For GC-MS, each feature represents a unique retention time and MS spectra, which equates to a single feature for each metabolite with some exceptions. For LC-MS, each feature represents a unique ion and retention time in the MS spectra. A given metabolite will have multiple isotopic peaks and can form adducts with salts such as sodium, and therefore the number of features by far exceeds the number of metabolites in an LC-MS experiment. After spectral processing, multivariate modelling methods such as principle component analysis (PCA) are used to visualize the relationships between samples based on the metabolome. Univariate statistics are then employed to determine features of interest, for example metabolites significantly altered by a disease state. Although many forms of software are available for these purposes, the majority of statistical analyses and plotting performed in this thesis were executed using the statistical computing programming language R (www.r-project.org)⁵⁹. The final step in the pipeline is compound identification. Again identification differs by platform and usually takes advantage of several databases and software, including but not limited to the NIST (National Institute of Standards and Technology) (GC-MS)⁶⁰, METLIN (LC-MS)⁶¹, Chemomx (NMR)⁶², and many others. Of course, once metabolites have been identified they must put into biological context. Online pathway mapping tools such as Kegg (www.kegg.jp)⁶³, Metacyc (www.metacyc.org)⁶⁴, the Human Metabolome

Database (www.hmdb.ca)⁶⁵ and others can be useful for this purpose. These databases are far from complete, especially for non-model organisms, and often review of the literature is the best method of determining the biological relevance of a metabolite.

Although many variations of NMR, GC-MS and LC-MS exist, for the purposes of the thesis, the focus will be on the Agilent 7890A GC-5975 inert MSD GC-MS and Agilent 1290 Infinity HPLC-Thermo Q-Exactive Orbitrap LC-MS systems as these were the main platform employed throughout the various chapters that follow.

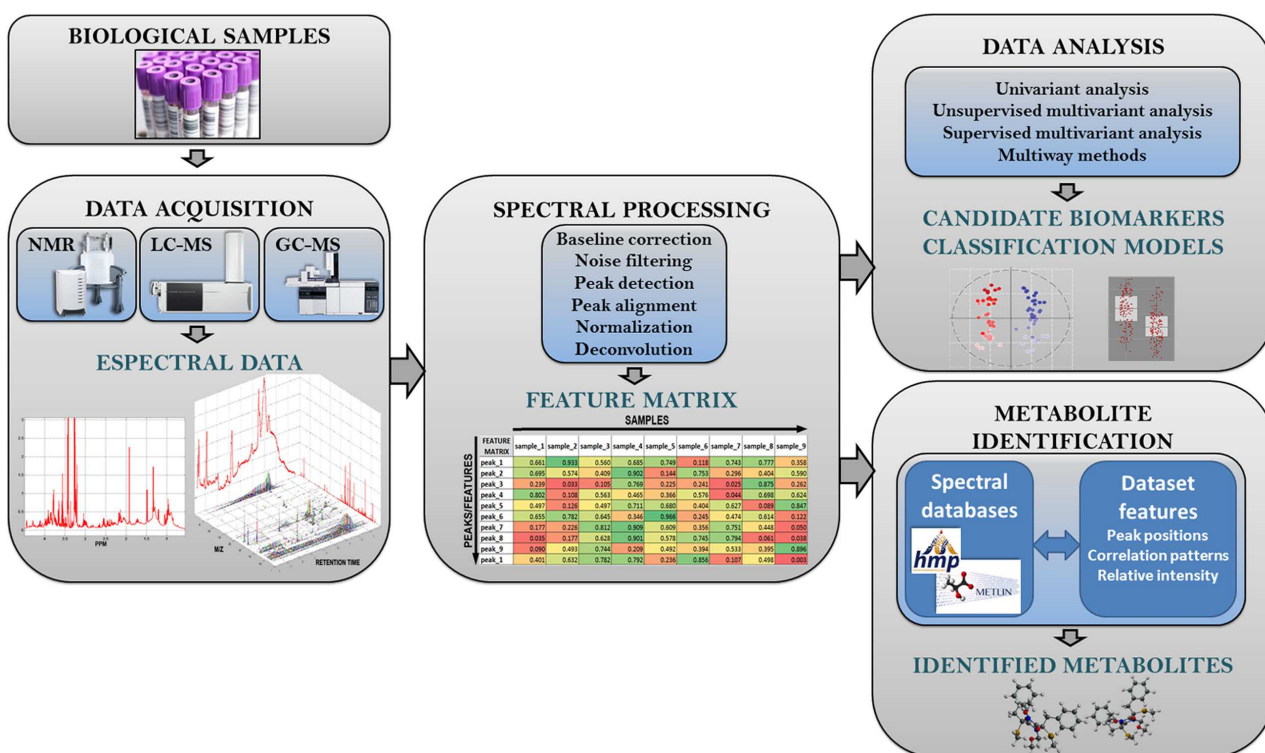


Figure 1-3. An overview of the steps involved in analysis of metabolomics data⁵⁷.

1.4.1 GC-MS

1.4.1.1 Hardware

The Agilent 5975 GC-MS series consists of a 7890A GC coupled to a 5975 MSD mass spectrometer. A schematic of the system is illustrated in Figure 1-4. Samples are usually derivatized prior to injection to allow detection of compounds which are non-volatile in their native form. The most common derivatization reagents are alkylation, acylation and silylation reagents⁶⁶. These chemicals replace hydrogens on N-H, O-H, and S-H groups with a non-polar group such as a trimethylsilyl. This reduction in polarity increases the volatility of compounds, allowing them to be detected by GC-MS. In solution, molecules with ketone or aldehydes such as carbohydrates undergo tautomerization (conversion between keto and enol form), resulting in a peak for each tautomer⁶⁷. To simplify analysis of these kinds of molecules reagents which prevent tautomerization are often used in conjunction with a silylation reagent.

After derivatization, samples are injected onto the GC column which is stored in a temperature controlled oven to allow volatilization of the molecules⁶⁸. An inert carrier gas, most often helium, serves as the mobile phase and an adsorbent silica coated column as the stationary phase. Gaseous compounds are separated based on adsorption to the column and also based on their boiling point. By applying a temperature ramp to the column, chemicals with high boiling points such as disaccharides will elute later, whereas low boilers will elute earlier. After separation, gaseous molecules enter the ion source where they are ionized by electrons emitted from a heated filament⁶⁹. This is known as electron ionization (EI), also known as electron impact. EI is a form of 'hard' ionization, in reference to the fact that fragmentation induced by this method is extensive, with the parent ion often undetectable in the resulting mass spectra. The ions then enter the quadrupole mass analyzer, which selects the mass range. The quadrupole consists of four parallel metal rods to which alternating and direct currents are applied to select a specific window of mass-to-charge ratios (m/z). Masses outside this window will collide with the quadrupole and not reach the detector. The type of detector employed by the Agilent model is known as an electron multiplier (EM). Specifically, the model used in this thesis

contains a triple axis detector which bends the ion beam at three axes to maximize the signal-to-noise ratio. When an ion hits the EM, it releases secondary electrons, amplifying (or multiplying) the original signal⁶⁹. The current created by this amplification of electrons can then be measured, allowing for ion quantification.

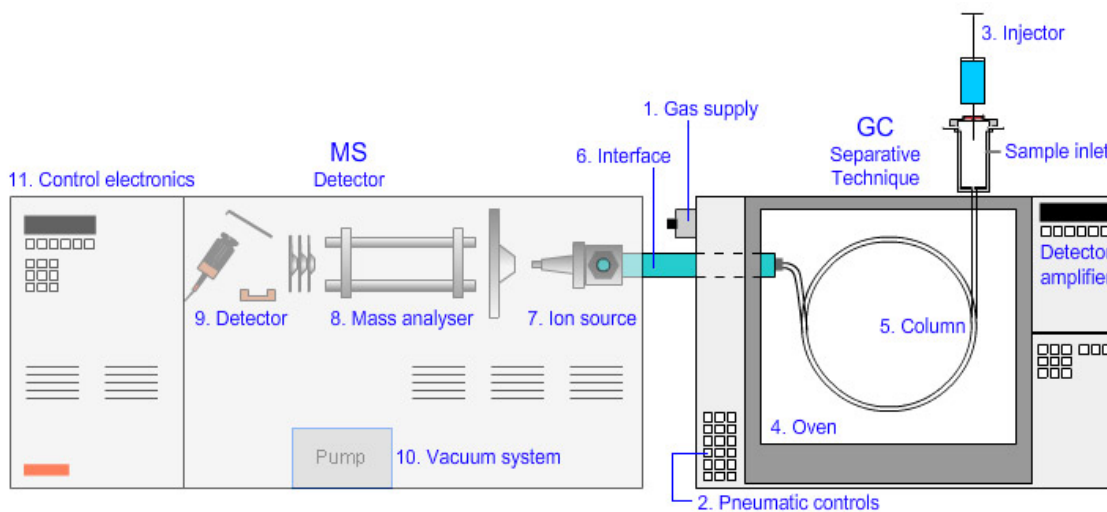


Figure 1-4 Schematic of the Agilent MSD GC-MS system⁷⁰.

1.4.1.2 Software

There are various forms of software available to view raw data and combine individual GC-MS files into a feature list. The most widely used for viewing raw chromatograms and spectra is AMDIS (Automated Mass spectral Deconvolution and Identification System)⁷¹. AMDIS is a free, vendor-independent software for viewing GC-MS chromatograms and corresponding spectra. It also integrates and extracts GC-MS features, allowing the user to define the signal-to-noise ratio as well as other aspects of feature detection. Due to integration with the NIST database, the most widely used database for metabolite identification in GC-MS, metabolites can be easily identified within the AMDIS interface.

Once features have been detected in AMDIS they need to be aligned across samples and converted into a feature list. Again there are multiple software available for this purpose. SpectConnect (<http://spectconnect.mit.edu/>)⁷² was chosen for its ease of use, and ability to be run independently through a command line interface, allowing for rapid analysis of

a large number of samples. The integration of this software with AMDIS also gives SpectConnect a considerable advantage over other software as each resulting feature represents a unique retention time and mass spectra. Other software most commonly return a separate feature for each ion within a spectra, exponentially increasing the amount of data generated and complicating feature identification.

1.4.1.3 Compound identification

Compounds detected by GC-MS are putatively identified based on MS spectra alone and then confirmed by authentic standards using both MS and retention time. The most widely used library of GC-MS spectra is the NIST Spectral Database. Due to the reproducible fragmentation of compounds by EI, and the relatively limited classes of molecules which can be ionized by this method, most compound can be confidently identified based on library comparison. Figure 1-5 shows an example of metabolite identification by comparison to the NIST database. The query spectra (red) is compared to the library spectra in blue. Attempts have been made to quantify the score of this library match, however they are often inaccurate. Currently, visual inspection of the match, combined with biological plausibility and validation with authentic standards remains the most reliable method of identification.

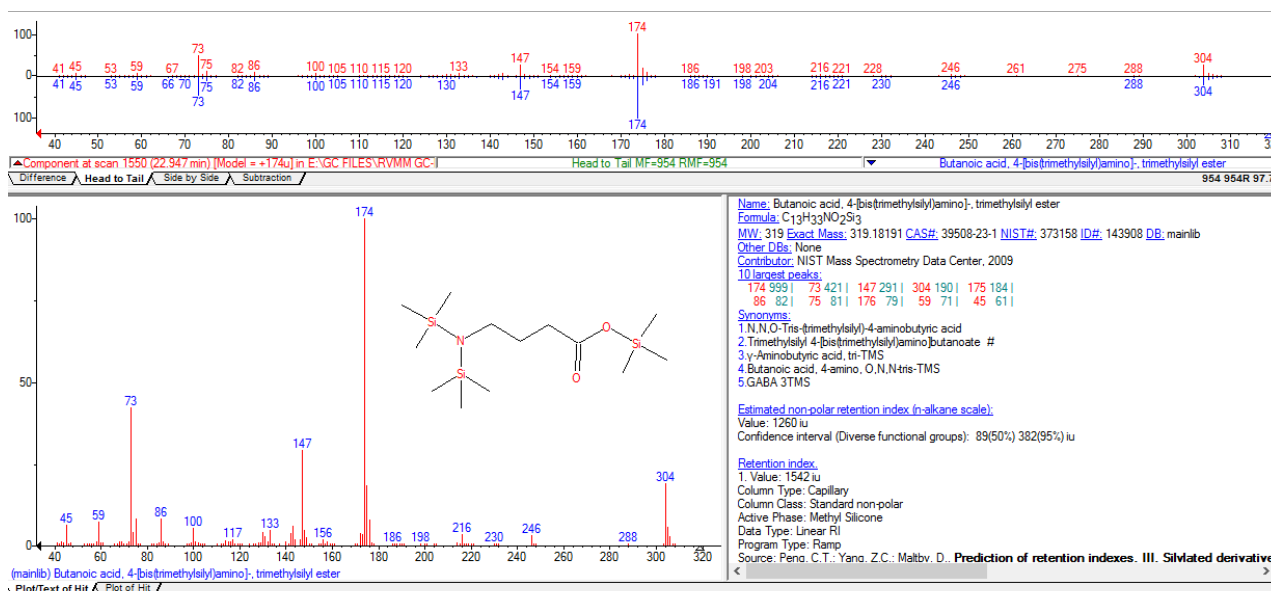


Figure 1-5. Example of metabolite (gamma-aminobutyrate) identification by GC-MS using the NIST database.

The query spectra is on top in red, with the database spectra beneath in blue. Trimethylsilyl (TMS) groups are introduced during derivatization.

1.4.2 LC-MS

1.4.2.1 Hardware

Unlike GC-MS, most molecules do not require derivatization to be detected by LC-MS, allowing for faster sample preparation and more stable extractions. Another key difference is that the LC and MS segments are manufactured independently in LC-MS, and therefore mixing segments from different vendor is common. The run times for LC-MS can also be quite short in comparison to GC-MS (5 min vs 60 min respectively for methods used in this thesis). This is due to in part to advancements in HPLC/UHPLC. By combining small particle (< 2 μ m) packing material with high pressure (> 200 bar), superior resolution can be achieved in a short amount of time⁷³. HPLC columns are coated with a packing material or stationary phase, the most widely used being octadecyl, also known as “C18”. C18 is a non-polar stationary phase, and is best suited to detect non-polar compounds, although detection of polar compounds is also possible. This is conventionally referred to as “reverse-phase” chromatography. In reverse phase, a polar mobile phase such as water is applied, followed by an organic phase, for example acetonitrile. In contrast, polar stationary phases such as HILIC (hydrophilic interaction liquid chromatography) are referred to as normal phase, but are less commonly used.

Once molecules have been separated by HPLC, they enter the source where they are ionized by electrospray ionization (ESI). Ions are formed by applying high voltage to liquid passing through a capillary tube⁶⁹. The resulting charged droplet will eventually break, creating the ‘electrospray’. By changing the voltage of the applied charge, positively or negatively charged ions are created. In contrast to EI, ESI is known as a ‘soft’ ionization technique because it does not usually induce fragmentation of the molecular ion.

The most common form of ionization are protonation [M+H] and deprotonation [M-H], however chemicals may ionize in a number of different ways, sometimes in an unpredictable manner. For positive ESI, the most common form of alternative ions are salt adducts such as [M+Na], [M+NH₄] or [M+K]. For negative ESI, acids present in the mobile phase, such as formic acid, form the most common adducts ([M-HCOO]), as well as [M-Cl]. Additionally, molecules can be multiply charged if they possess multiple ionizable sites, such is the case with most peptides. In this case the observed ion will be smaller than the parent, as only the mass-to-charge ratio (m/z) of an ion can be measured by the mass spectrometer.

Once the electrospray has been created, ions enter the mass spectrometer. A schematic of the Thermo Q-Exactive Orbitrap mass spectrometer, the model used throughout this thesis, is shown in Figure 1-6. Briefly, ions are guided into the quadrupole by a bent flatapole. The quadrupole then selects the range of m/z to be analyzed in the same fashion as was described for GC-MS. After mass selection, ions enter the C-trap which guides the ions into the orbitrap for mass detection. The orbitrap is a relatively new form of mass analyzer, having first been described in the year 2000 by Alexander Makarov⁷⁴. It is composed of a barrel-shaped outer electrode and an inner spindle-shaped electrode. The ions oscillate around this spindle at a frequency that is proportional to their m/z , enabling mass detection with high accuracy. The orbitrap's superior resolution and mass accuracy make it an ideal choice for untargeted metabolomics.

The Q Exactive is capable of performing both 'full MS' and/or 'MS/MS' analyses, the choice of which is pre-set by the user. In full MS analysis, molecules are ionized and detected intact without fragmentation. This mode maximizes the sensitivity of the instrument and is therefore most useful for relative quantification of ions, but gives little information about the structure of compounds. Alternatively, MS/MS, also known as tandem mass spectrometry, provides information about the structure of the molecule at the expense of sensitivity. For MS/MS, ions selected by the C-Trap are sent into a collision cell (HCD cell) where they collide with neutral molecules of nitrogen gas,

inducing fragmentation of the molecular ion⁶⁹. These fragments can then be used the piece together the structure of the parent molecule.

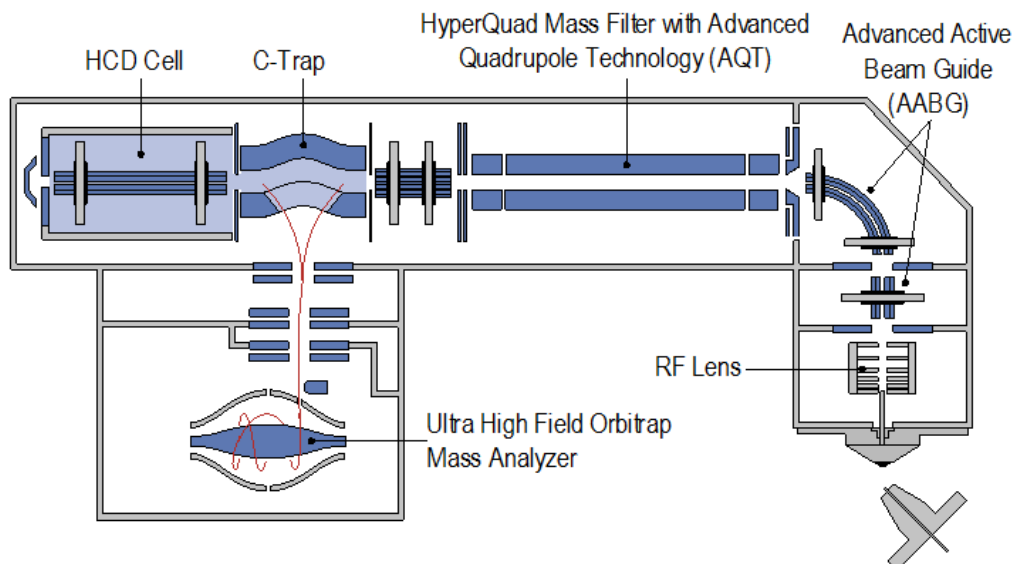


Figure 1-6. Thermo Q-Exacte Orbitrap configuration⁷⁵.

1.4.2.2 Software

As with GC-MS, a number of vendor-derived and open source software have been developed to view and analyze LC-MS metabolomics data. To view raw chromatograms and spectra, a vendor provided software is most commonly used. The Thermo Scientific software for LC-MS spectral viewing is called Xcaliber. This program allows one to view all ions in a given mass spectrum, as well as MS/MS of metabolites of interest. It is also useful for calculating molecular formulas, as well as adduct and neutral loss determination. Xcaliber is a low throughput program and is best suited to viewing a single or at most a handful of samples at a time. For this reason, other programs have been created to align and integrate multiple LC-MS files into a feature list which can then be used for multivariate modelling and statistical analysis. One of the most widely used tools for LC-MS feature alignment is XCMS^{76,77}. XCMS is an open source software, and is compatible with a wide range of LC-MS platforms. One of the main advantages of XCMS over other software is that it is available as both an online point and click interface as well as a command line R package. The R version allows for rapid processing of a large number of samples and is highly adaptable compared to the online version.

XCMS also excels at identifying signal from noise in comparison to other available programs.

1.4.2.3 Compound identification

Compound identification remains the most challenging and time consuming aspect of untargeted LC-MS analysis. Even if a molecular formula can be assigned based on the m/z alone, there can be hundreds if not thousands of molecular species with the same formula. For this reason, the mass alone provides little information about the structure of the molecule. To confidently identify a metabolite, an MS/MS spectra must be acquired. This spectra can then be searched against an MS/MS databases, the largest being METLIN⁶¹. Although these databases can be extremely useful, they are far from complete, and an MS/MS spectra for the vast majority of known molecules cannot be found in any database. In these cases, molecules must be identified ‘*de novo*’ from the fragments present in the mass spectra. To help address this issue, fragmentation prediction tools such as CFM-ID⁷⁸ have been created. These programs use machine learning algorithms to predict the MS/MS spectra of a known compound, or alternatively to predict the compound identity from an experimental MS/MS spectra. In many cases, neither of these approaches yield plausible identities and more advanced methods such as neutral loss and fragment matching must be used to piece together the putative structure. Ultimately, an MS/MS spectra of an authentic standard must be run to confirm the predicted identity, only then can one be confident in the annotation.

1.5 Using metabolomics to characterize microbial communities

The metabolome represents the end-products of both mammalian and microbial metabolism in a given community, providing a snapshot of all cellular processes in an environment. Given the complex interaction between bacterial species and with the host, metabolomics offers a unique window into the culmination of these interactions. Recent advances in processing of metabolomics and NGS data has now made it possible to combine these data to obtain a picture not only of “who is there”, but also “what they are producing”. Despite the capabilities of performing such experiments, these studies are still relatively rare due to their multidisciplinary nature, requiring biology, chemistry, and

computational expertise. Access to analytical chemistry platforms remains a formidable barrier, as metabolomics core facilities are still scarce, unlike their NGS counterpart. For this reason, most metabolomics analyses are contracted out on a fee for service basis. While this approach is attractive to groups lacking chemistry expertise, it is extremely costly. Furthermore, analysis is mainly restricted to targeted profiling as the time required to elucidate structures of unknown compounds would be far too expensive.

Despite the infancy of the field, a number of studies have successfully applied metabolomics to improve our understanding of disease mechanisms, the microbiome, and to identify diagnostic biomarkers. A prime example is the discovery of the microbial-mammalian co-metabolite trimethylamine N-oxide (TMAO) and its involvement in cardiovascular disease. In this work an untargeted study of metabolites linked to cardiovascular disease revealed a causal association between TMAO and atherosclerosis, elucidating a novel role for the microbiome in cardiovascular disease⁷⁹⁻⁸¹. Another remarkable application for metabolomics is in surgery. Work by Takats and colleagues⁸² have developed a mass spectrometry based surgical knife capable of accurately deciphering cancerous tissue from healthy in real time. Finally, metabolomics has also been used to identify biomarkers that predict future disease onset such as development of mild cognitive impairment or Alzheimer's disease⁸³.

1.6 Scope and objectives

Upon commencement of my PhD studies, the concept of using metabolomics to characterize the function of microbial communities was relatively new. Indeed, there were few if any publications attempting this feat. The purpose of this thesis was therefore to develop the tools to analyze the metabolome of biological samples and integrate this data with microbiota profiles. The chapters that follow use these tools to advance understanding of the diseases under investigation and/or identify diagnostic biomarkers, with a focus on the vaginal microbiota and BV.

Chapter 2 describes a clinical study of the vaginal metabolome and microbiota of Rwandan women. By combining mass spectrometry, NGS and *in vitro* analyses of

bacterial isolates I demonstrated that the vaginal microbiota shapes the chemical composition of the vagina, and identified highly specific biomarkers for BV. By correlating metabolites with the microbiota I was able to pinpoint the species responsible for producing one of these biomarkers and validated production *in vitro*. This work provides novel insight into the metabolism of the vaginal microbiota and identifies candidate biomarkers for BV. It is also one of the first to demonstrate that metabolomics can be used to identify previously unknown microbe-product relationships.

Chapter 3 expands upon the second chapter, exploring what the bacteria are doing in addition to which organisms are present by relating the bacterial meta-transcriptome and metabolome. Combining these data (the meta-transcriptome portion performed by Jean Macklaim, a former PhD student in our lab), we demonstrated that while both methods effectively separate women with BV from health, the information provided by each method is unique and complementary. This is the first study to combine meta-transcriptomics and metabolomics, and provides valuable information about the kinds of questions that can and cannot be addressed using these approaches.

In chapter 4, the persistence of *L. iners* is explored. Based on previous genome sequencing of an isolate of *L. iners*³⁹, I hypothesized that this species had fibronectin binding capabilities. I demonstrated that *L. iners* is able to bind to fibronectin with high affinity and that this binding is protein mediated. Fibronectin binding is unique to *L. iners* in the vaginal environment, suggesting this may be one of the factors which allows it to persist under conditions of BV.

The vaginal microbiota is a relatively simple community compared to the gut, as it is a closed system and contains fewer species. The goal of chapter 5 was therefore to apply the tools developed in the previous chapters to more complex biological matrices. This chapter describes a study of the stool microbiota and stool/plasma metabolome of Nigerian children with severe acute malnutrition (SAM). I demonstrated that the plasma metabolome, but not the stool metabolome or microbiota, discriminates children with

SAM from controls. Numerous metabolic perturbations in children with SAM were elucidated, providing novel insight into disease mechanisms and management.

The final chapter stems from data generated in chapter 5 and details a method to filter out noise resulting from salts in LC-MS based human metabolomics data. Applying this filter will allow for more accurate annotation of the metabolome and decrease reporting of unknowns.

Throughout this thesis I demonstrate the value of an untargeted metabolomics approach. Identifying novel associations between metabolites and disease, I illustrate how an untargeted metabolomics can be used to identify diagnostic biomarkers and generate new hypotheses regarding the microbiome and mechanisms of disease.

1.7 References

- (1) Wikoff, W. R.; Anfora, A. T.; Liu, J.; Schultz, P. G.; Lesley, S. a; Peters, E. C.; Siuzdak, G. Metabolomics Analysis Reveals Large Effects of Gut Microflora on Mammalian Blood Metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (10), 3698–3703.
- (2) 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. S. S.; Li, D.; Cao, J.; Wang, B.; Liang, H.; Zheng, H.; Xie, Y.; Tap, J.; Lepage, P.; Bertalan, M.; Batto, J.-M.; Hansen, T.; Le Paslier, D.; Linneberg, A.; Nielsen, H. B.; Pelletier, E.; Renault, P.; Sicheritz-Ponten, T.; Turner, K.; Zhu, H.; Yu, C.; Li, S. S. S.; Jian, M.; Zhou, Y.; Li, Y.; Zhang, X.; Li, S. S. S.; Qin, N.; Yang, H.; Wang, J. J.; Brunak, S.; Doré, J.; Guarner, F.; Kristiansen, K.; Pedersen, O.; Parkhill, J.; Weissenbach, J.; MetaHIT Consortium; Bork, P.; Ehrlich, S. D.; Wang, J. J.; Ruiqiang Li¹; 3, M. A. * J. R.; Burgdorf⁴, K. S.; Manichanh⁵, C.; Nielsen⁴, T.; Pons⁶, N.; Levenez⁶, F.; Yamada², T.; Mende², D. R.; 7, J. T. Y. X. J. X. J. L.; Li¹, S. S. S.; 8, J. C. D. L.; Wang¹, B.; Liang¹, H.; Zheng¹, H.; 7, J. T. Y. X. J. X. J. L.; Lepage⁶, P.; Bertalan⁹, M.; Batto⁶, J.-M.; Hansen⁴, T.; Le, D.; Allan Linneberg¹¹, H. B. N. P.; Eric Pelletier¹⁰, P. R.; Sicheritz-Ponten⁹, T.; Keith Turner¹², H. Z.; Yu¹, C.; Li¹, S. S. S.; Jian¹, M.; Zhou¹, Y.; Li¹, Y.; Zhang¹, X.; Li¹, S. S. S.; Qin¹, N.; Yang¹,

- H.; Wang¹, J.; Brunak⁹, S.; Dore⁶, J.; Guarner⁵, F.; MetaHIT Consortium[{], P. B. K. K. O. P. 14 J. P. J. W.; Wang, S. D. E. & J. A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing: Commentary. *Nature* **2010**, *11* (1), 28.
- (3) Woese, C. R.; Fox, G. E. Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (11), 5088–5090.
 - (4) Liu, Z.; DeSantis, T. Z.; Andersen, G. L.; Knight, R. Accurate Taxonomy Assignments from 16S rRNA Sequences Produced by Highly Parallel Pyrosequencers. *Nucleic Acids Res.* **2008**, *36* (18), e120.
 - (5) Fredricks, D. N.; Fiedler, T. L.; Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **2005**, *353* (18), 1899–1911.
 - (6) Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G. M.; Koenig, S. S. K.; McCulle, S. L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C. O.; Brotman, R. M.; Davis, C. C.; Ault, K.; Peralta, L.; Forney, L. J. Vaginal Microbiome of Reproductive-Age Women. *Proc. Natl. Acad. Sci.* **2010**, *108* (Supplement_1), 4680–4687.
 - (7) Hummelen, R.; Fernandes, A. D.; Macklaim, J. M.; Dickson, R. J.; Chantalucha, J.; Gloor, G. B.; Reid, G. Deep Sequencing of the Vaginal Microbiota of Women with HIV. *PLoS One* **2010**, *5* (8).
 - (8) McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J. E.; Gloor, G. B.; Reid, G. A Multi-Platform Metabolomics Approach Identifies Highly Specific Biomarkers of Bacterial Diversity in the Vagina of Pregnant and Non-Pregnant Women. *Sci. Rep.* **2015**, *5*, 14174.
 - (9) Claesson, M. J.; O’Sullivan, O.; Wang, Q.; Nikkil^{??}, J.; Marchesi, J. R.; Smidt, H.; de Vos, W. M.; Ross, R. P.; O’Toole, P. W. Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS One* **2009**, *4* (8).
 - (10) Claesson, M. J.; Wang, Q.; O’Sullivan, O.; Greene-Diniz, R.; Cole, J. R.; Ross, R.

- P.; O'Toole, P. W. Comparison of Two next-Generation Sequencing Technologies for Resolving Highly Complex Microbiota Composition Using Tandem Variable 16S rRNA Gene Regions. *Nucleic Acids Res.* **2010**, *38* (22).
- (11) Haas, B. J.; Gevers, D.; Earl, A. M.; Feldgarden, M.; Ward, D. V.; Giannoukos, G.; Ciulla, D.; Tabbaa, D.; Highlander, S. K.; Sodergren, E.; Meth, B.; DeSantis, T. Z.; Petrosino, J. F.; Knight, R.; Birren, B. W. Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons. *Genome Res.* **2011**, *21* (3), 494–504.
- (12) Di Bella, J. M.; Bao, Y.; Gloor, G. B.; Burton, J. P.; Reid, G. High Throughput Sequencing Methods and Analysis for Microbiome Research. *J. Microbiol. Methods* **2013**, *95* (3), 401–414.
- (13) Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J.; Barnes, C. L.; Bignell, H. R.; Boutell, J. M.; Bryant, J.; Carter, R. J.; Keira Cheetham, R.; Cox, A. J.; Ellis, D. J.; Flatbush, M. R.; Gormley, N. A.; Humphray, S. J.; Irving, L. J.; Karbelashvili, M. S.; Kirk, S. M.; Li, H.; Liu, X.; Maisinger, K. S.; Murray, L. J.; Obradovic, B.; Ost, T.; Parkinson, M. L.; Pratt, M. R.; Rasolonjatovo, I. M. J.; Reed, M. T.; Rigatti, R.; Rodighiero, C.; Ross, M. T.; Sabot, A.; Sankar, S. V.; Scally, A.; Schroth, G. P.; Smith, M. E.; Smith, V. P.; Spiridou, A.; Torrance, P. E.; Tzonev, S. S.; Vermaas, E. H.; Walter, K.; Wu, X.; Zhang, L.; Alam, M. D.; Anastasi, C.; Aniebo, I. C.; Bailey, D. M. D.; Bancarz, I. R.; Banerjee, S.; Barbour, S. G.; Baybayan, P. A.; Benoit, V. A.; Benson, K. F.; Bevis, C.; Black, P. J.; Boodhun, A.; Brennan, J. S.; Bridgham, J. A.; Brown, R. C.; Brown, A. A.; Buermann, D. H.; Bundu, A. A.; Burrows, J. C.; Carter, N. P.; Castillo, N.; Chiara E Catenazzi, M.; Chang, S.; Neil Cooley, R.; Crake, N. R.; Dada, O. O.; Diakoumakos, K. D.; Dominguez-Fernandez, B.; Earnshaw, D. J.; Egbujor, U. C.; Elmore, D. W.; Etchin, S. S.; Ewan, M. R.; Fedurco, M.; Fraser, L. J.; Fuentes Fajardo, K. V.; Scott Furey, W.; George, D.; Gietzen, K. J.; Goddard, C. P.; Golda, G. S.; Granieri, P. A.; Green, D. E.; Gustafson, D. L.; Hansen, N. F.; Harnish, K.; Haudenschild, C. D.; Heyer, N. I.; Hims, M. M.; Ho, J. T.; Horgan, A. M.; Hoschler,

- K.; Hurwitz, S.; Ivanov, D. V.; Johnson, M. Q.; James, T.; Huw Jones, T. A.; Kang, G.-D.; Kerelska, T. H.; Kersey, A. D.; Khrebtukova, I.; Kindwall, A. P.; Kingsbury, Z.; Kokko-Gonzales, P. I.; Kumar, A.; Laurent, M. A.; Lawley, C. T.; Lee, S. E.; Lee, X.; Liao, A. K.; Loch, J. A.; Lok, M.; Luo, S.; Mammen, R. M.; Martin, J. W.; McCauley, P. G.; McNitt, P.; Mehta, P.; Moon, K. W.; Mullens, J. W.; Newington, T.; Ning, Z.; Ling Ng, B.; Novo, S. M.; O'Neill, M. J.; Osborne, M. A.; Osnowski, A.; Ostadan, O.; Paraschos, L. L.; Pickering, L.; Pike, A. C.; Pike, A. C.; Chris Pinkard, D.; Pliskin, D. P.; Podhasky, J.; Quijano, V. J.; Raczky, C.; Rae, V. H.; Rawlings, S. R.; Chiva Rodriguez, A.; Roe, P. M.; Rogers, J.; Rogert Bacigalupo, M. C.; Romanov, N.; Romieu, A.; Roth, R. K.; Rourke, N. J.; Ruediger, S. T.; Rusman, E.; Sanches-Kuiper, R. M.; Schenker, M. R.; Seoane, J. M.; Shaw, R. J.; Shiver, M. K.; Short, S. W.; Sizto, N. L.; Sluis, J. P.; Smith, M. A.; Ernest Sohna Sohna, J.; Spence, E. J.; Stevens, K.; Sutton, N.; Szajkowski, L.; Tregidgo, C. L.; Turcatti, G.; Vandevondele, S.; Verhovsky, Y.; Virk, S. M.; Wakelin, S.; Walcott, G. C.; Wang, J.; Worsley, G. J.; Yan, J.; Yau, L.; Zuerlein, M.; Rogers, J.; Mullikin, J. C.; Hurler, M. E.; McCooke, N. J.; West, J. S.; Oaks, F. L.; Lundberg, P. L.; Klenerman, D.; Durbin, R.; Smith, A. J. Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry. *Nature* **2008**, *456* (7218), 53–59.
- (14) Gloor, G. B.; Hummelen, R.; Macklaim, J. M.; Dickson, R. J.; Fernandes, A. D.; MacPhee, R.; Reid, G. Microbiome Profiling by Illumina Sequencing of Combinatorial Sequence-Tagged PCR Products. *PLoS One* **2010**, *5* (10), e15406.
- (15) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunencko, T.; Zaneveld, J.; Knight, R. QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nat. Methods* **2010**, *7* (5), 335–336.

- (16) Stackebrandt, E, Goebel, B. M. Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int. J. Syst. Bacteriol.* **1994**, *44* (4), 846–849.
- (17) Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* **2007**, *73* (16), 5261–5267.
- (18) DeSantis, T. Z.; Hugenholtz, P.; Larsen, N.; Rojas, M.; Brodie, E. L.; Keller, K.; Huber, T.; Dalevi, D.; Hu, P.; Andersen, G. L. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* **2006**, *72* (7), 5069–5072.
- (19) Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F. O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2013**, *41* (D1).
- (20) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27* (July 1928), 379–423.
- (21) Lozupone, C.; Knight, R. UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2005**, *71* (12), 8228–8235.
- (22) Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B. Methodol.* **1982**, *44* (2), 139–177.
- (23) Filzmoser, P.; Hron, K.; Reimann, C. Univariate Statistical Analysis of Environmental (Compositional) Data: Problems and Possibilities. *Sci. Total Environ.* **2009**, *407* (23), 6100–6108.
- (24) Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's All Relative: Analyzing Microbiome Data as Compositions. *Ann. Epidemiol.* **2016**, 1–8.
- (25) Gloor, G. B.; Reid, G. Compositional Analysis: A Valid Approach to Analyze Microbiome High Throughput Sequencing Data. *Can. J. Microbiol.* **2016**, cjm –

2015–0821.

- (26) McMurdie, P. J.; Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **2014**, *10* (4).
- (27) Sinha, R.; Abnet, C. C.; White, O.; Knight, R.; Huttenhower, C. The Microbiome Quality Control Project: Baseline Study Design and Future Directions. *Genome Biol.* **2015**, *16*, 276.
- (28) van den Boogaart, K. G.; Tolosana-Delgado. *Analyzing Compositional Data with R*; 2013.
- (29) Fernandes, A. D.; Macklaim, J. M.; Linn, T. G.; Reid, G.; Gloor, G. B. ANOVA-like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* **2013**, *8* (7), e67019.
- (30) Fernandes, A. D.; Reid, J. N.; Macklaim, J. M.; McMurrough, T. A.; Edgell, D. R.; Gloor, G. B. Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-Seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis. *Microbiome* **2014**, *2*, 15.
- (31) Juárez Tomás, M. S.; Saralegui Duhart, C. I.; De Gregorio, P. R.; Vera Pingitore, E.; Nader-Macías, M. E. Urogenital Pathogen Inhibition and Compatibility between Vaginal *Lactobacillus* Strains to Be Considered as Probiotic Candidates. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2011**, *159* (2), 399–406.
- (32) O’Hanlon, D. E.; Moench, T. R.; Cone, R. A. In Vaginal Fluid, Bacteria Associated with Bacterial Vaginosis Can Be Suppressed with Lactic Acid but Not Hydrogen Peroxide. *BMC Infect. Dis.* **2011**, *11* (1), 200.
- (33) Dover, S. E.; Aroutcheva, A. A.; Faro, S.; Chikindas, M. L. Safety Study of an Antimicrobial Peptide Lactocin 160, Produced by the Vaginal *Lactobacillus rhamnosus*. *Infect. Dis. Obstet. Gynecol.* **2007**, 2007.
- (34) Eschenbach, D. a; Davick, P. R.; Williams, B. L.; Klebanoff, S. J.; Young-Smith,

- K.; Critchlow, C. M.; Holmes, K. K. Prevalence of Hydrogen Peroxide-Producing *Lactobacillus* Species in Normal Women and Women with Bacterial Vaginosis. *J. Clin. Microbiol.* **1989**, *27* (2), 251–256.
- (35) Hillier, S. L.; Krohn, M. A.; Klebanoff, S. J.; Eschenbach, D. A. The Relationship of Hydrogen Peroxide-Producing Lactobacilli to Bacterial Vaginosis and Genital Microflora in Pregnant Women. *Obstet. Gynecol.* **1992**, *79* (3), 369–373.
- (36) Frank, D. N.; St Amand, A. L.; Feldman, R. A.; Boedeker, E. C.; Harpaz, N.; Pace, N. R. Molecular-Phylogenetic Characterization of Microbial Community Imbalances in Human Inflammatory Bowel Diseases. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (34), 13780–13785.
- (37) Manichanh, C.; Rigottier-Gois, L.; Bonnaud, E.; Gloux, K.; Pelletier, E.; Frangeul, L.; Nalin, R.; Jarrin, C.; Chardon, P.; Marteau, P.; Roca, J.; Dore, J. Reduced Diversity of Faecal Microbiota in Crohn's Disease Revealed by a Metagenomic Approach. *Gut* **2006**, *55* (2), 205–211.
- (38) Turnbaugh, P. J.; Gordon, J. I. The Core Gut Microbiome, Energy Balance and Obesity. *J. Physiol.* **2009**, *587* (Pt 17), 4153–4158.
- (39) Macklaim, J. M.; Gloor, G. B.; Anukam, K. C.; Cribby, S.; Reid, G. At the Crossroads of Vaginal Health and Disease, the Genome Sequence of *Lactobacillus iners* AB-1. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* Suppl , 4688–4695.
- (40) Allsworth, J. E.; Peipert, J. F. Prevalence of Bacterial Vaginosis: 2001-2004 National Health and Nutrition Examination Survey Data. *Obstet. Gynecol.* **2007**, *109* (1), 114–120.
- (41) Zozaya-Hinchliffe, M.; Lillis, R.; Martin, D. H.; Ferris, M. J. Quantitative PCR Assessments of Bacterial Species in Women with and without Bacterial Vaginosis. *J. Clin. Microbiol.* **2010**, *48* (5), 1812–1819.
- (42) Sha, B. E.; Zariffard, M. R.; Wang, Q. J.; Chen, H. Y.; Bremer, J.; Cohen, M. H.; Spear, G. T. Female Genital-Tract HIV Load Correlates Inversely with

Lactobacillus Species but Positively with Bacterial Vaginosis and *Mycoplasma hominis*. *J. Infect. Dis.* **2005**, *191* (1), 25–32.

- (43) Cohen, C. R.; Lingappa, J. R.; Baeten, J. M.; Ngayo, M. O.; Spiegel, C. A.; Hong, T.; Donnell, D.; Celum, C.; Kapiga, S.; Delany, S.; Bukusi, E. A. Bacterial Vaginosis Associated with Increased Risk of Female-to-Male HIV-1 Transmission: A Prospective Cohort Analysis among African Couples. *PLoS Med.* **2012**, *9* (6), e1001251.
- (44) Gallo, M. F.; Macaluso, M.; Warner, L.; Fleenor, M. E.; Hook, E. W.; Brill, I.; Weaver, M. A. Bacterial Vaginosis, Gonorrhoea, and Chlamydial Infection among Women Attending a Sexually Transmitted Disease Clinic: A Longitudinal Analysis of Possible Causal Links. *Ann. Epidemiol.* **2012**, *22* (3), 213–220.
- (45) Bautista, C. T.; Wurapa, E. K.; Sanchez, J. L. Brief Report: Associations between Antecedent Bacterial Vaginosis and Incident Chlamydia and Gonorrhoea Diagnoses, U.S. Army Females, 2006-2012. *MSMR* **2016**, *23* (2), 32–34.
- (46) Das, T. R.; Jahan, S.; Begum, S. R.; Akhtar, M. F. Association between Bacterial Vaginosis and Preterm Delivery. *Mymensingh Med. J.* **2011**, *20* (1), 115–120.
- (47) Hillier, S. L.; Nugent, R. P.; Eschenbach, D. A.; Krohn, M. A.; Gibbs, R. S.; Martin, D. H.; Cotch, M. F.; Edelman, R.; Pastorek, J. G.; Rao, A. V; McNellis, D.; Regan, J. A.; Carey, J. C.; Klebanoff, M. A. Association between Bacterial Vaginosis and Preterm Delivery of a Low-Birth-Weight Infant. *N. Engl. J. Med.* **1995**, *333* (26), 1737–1742.
- (48) Sobel, J. D.; Karpas, Z.; Lorber, A. Diagnosing Vaginal Infections through Measurement of Biogenic Amines by Ion Mobility Spectrometry. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2012**, *163* (1), 81–84.
- (49) Wolrath, H.; Forsum, U.; Larsson, P. G.; Borén, H. Analysis of Bacterial Vaginosis-Related Amines in Vaginal Fluid by Gas Chromatography and Mass Spectrometry. *J. Clin. Microbiol.* **2001**, *39* (11), 4026–4031.

- (50) Wolrath, H.; Borén, H.; Hallén, A.; Forsum, U. Trimethylamine Content in Vaginal Secretion and Its Relation to Bacterial Vaginosis. *APMIS* **2002**, *110* (11), 819–824.
- (51) Klebanoff, M. A.; Schwebke, J. R.; Zhang, J.; Nansel, T. R.; Yu, K.-F.; Andrews, W. W. Vulvovaginal Symptoms in Women with Bacterial Vaginosis. *Obstet. Gynecol.* **2004**, *104*, 267–272.
- (52) Nugent, R. P.; Krohn, M. A.; Hillier, S. L. Reliability of Diagnosing Bacterial Vaginosis Is Improved by a Standardized Method of Gram Stain Interpretation. *J. Clin. Microbiol.* **1991**, *29* (2), 297–301.
- (53) Verhelst, R.; Verstraelen, H.; Claeys, G.; Verschraegen, G.; Van Simaey, L.; De Ganck, C.; De Backer, E.; Temmerman, M.; Vanechoutte, M. Comparison between Gram Stain and Culture for the Characterization of Vaginal Microflora: Definition of a Distinct Grade That Resembles Grade I Microflora and Revised Categorization of Grade I Microflora. *BMC Microbiol.* **2005**, *5*, 61.
- (54) Amsel, R.; Totten, P. A.; Spiegel, C. A.; Chen, K. C.; Eschenbach, D.; Holmes, K. K. Nonspecific Vaginitis. Diagnostic Criteria and Microbial and Epidemiologic Associations. *Am. J. Med.* **1983**, *74* (1), 14–22.
- (55) Macklaim, J. M.; Fernandes, A. D.; Di Bella, J. M.; Hammond, J.-A.; Reid, G.; Gloor, G. B. Comparative Meta-RNA-Seq of the Vaginal Microbiota and Differential Expression by *Lactobacillus iners* in Health and Dysbiosis. *Microbiome* **2013**, *1* (1), 12.
- (56) Fiehn, O. Metabolomics--the Link between Genotypes and Phenotypes. *Plant Mol. Biol.* **2002**, *48* (1-2), 155–171.
- (57) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* **2015**, *3* (March), 23.
- (58) Wishart, D. S. Computational Approaches to Metabolomics. *Methods Mol. Biol.* **2010**, *593*, 283–313.

- (59) R Development Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **2015**, *1*, 409.
- (60) Linstrom, P. J.; Mallard, W. G. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; 2000.
- (61) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Ther Drug Monit* **2005**, *27* (6), 747–751.
- (62) Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted Profiling: Quantitative Analysis of ¹H NMR Metabolomics Data. *Anal. Chem.* **2006**, *78* (13), 4430–4442.
- (63) Kanehisa, M.; Goto, S. Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
- (64) Caspi, R.; Altman, T.; Billington, R.; Dreher, K.; Foerster, H.; Fulcher, C. A.; Holland, T. A.; Keseler, I. M.; Kothari, A.; Kubo, A.; Krummenacker, M.; Latendresse, M.; Mueller, L. A.; Ong, Q.; Paley, S.; Subhraveti, P.; Weaver, D. S.; Weerasinghe, D.; Zhang, P.; Karp, P. D. The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2014**, *42* (D1).
- (65) 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.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521–D526.
- (66) Knapp, D. R. Handbook of Analytical Derivatization Reactions. *John Wiley Sons*

New York **1979**, 741.

- (67) Ruiz-Matute, A. I.; Hernandez-Hernandez, O.; Rodriguez-Sanchez, S.; Sanz, M. L.; Martinez-Castro, I. Derivatization of Carbohydrates for GC and GC-MS Analyses. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*. 2011, pp 1226–1240.
- (68) Grob L Robert and Barry F Eugene. *Modern Practice of Gas Chromatography, Fourth Edition.*; 2004.
- (69) Hoffmann, E. De; Stroobant, V. *Mass Spectrometry Principles and Applications*; 2007.
- (70) Chromacademy. Chromacademy essentials
<http://www.chromacademy.com/essential-guide/nov2010/fig-1.jpg>.
- (71) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (8), 770–781.
- (72) Styczynski, M. P.; Moxley, J. F.; Tong, L. V.; Walther, J. L.; Jensen, K. L.; Stephanopoulos, G. N. Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery. *Anal. Chem.* **2007**, *79* (3), 966–973.
- (73) Wu, N.; Clausen, A. M. Fundamental and Practical Aspects of Ultrahigh Pressure Liquid Chromatography for Fast Separations. *J. Sep. Sci.* **2007**, *30* (8), 1167–1182.
- (74) Makarov, A. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
- (75) Thermo Scientific. Planet Orbitrap <http://planetorbitrap.com/q-exactive-hf#tab:schematic>.
- (76) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS:

- Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787.
- (77) Patti, G. J.; Tautenhahn, R.; Siuzdak, G. Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments. *Nat. Protoc.* **2012**, *7* (3), 508–516.
- (78) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* **2014**, *42* (W1), 94–99.
- (79) 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. Gut Flora Metabolism of Phosphatidylcholine Promotes Cardiovascular Disease. *Nature* **2011**, *472* (7341), 57–63.
- (80) 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.; Tang, W. H. W.; Bushman, F. D.; Lusis, A. J.; Hazen, S. L. Intestinal Microbiota Metabolism of L-Carnitine, a Nutrient in Red Meat, Promotes Atherosclerosis. *Nat. Med.* **2013**, *19* (April), 576–585.
- (81) Gregory, J. C.; Buffa, J. A.; Org, E.; Wang, Z.; Levison, B. S.; Zhu, W.; Wagner, M. A.; Bennett, B. J.; Li, L.; DiDonato, J. A.; Lusis, A. J.; Hazen, S. L. Transmission of Atherosclerosis Susceptibility with Gut Microbial Transplantation. *J. Biol. Chem.* **2015**, *290* (9), 5647–5660.
- (82) Balog, J.; Sasi-Szabo, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezsó, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takats, Z.; Sasi-Szabó, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezsó, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takáts, Z. Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Sci. Transl. Med.* **2013**, *5* (194), 194ra93.

- (83) Mapstone, M.; Cheema, A. K.; Fiandaca, M. S.; Zhong, X.; Mhyre, T. R.; MacArthur, L. H.; Hall, W. J.; Fisher, S. G.; Peterson, D. R.; Haley, J. M.; Nazar, M. D.; Rich, S. A.; Berlau, D. J.; Peltz, C. B.; Tan, M. T.; Kawas, C. H.; Federoff, H. J. Plasma Phospholipids Identify Antecedent Memory Impairment in Older Adults. *Nat. Med.* **2014**, *20* (4), 415–418.

Chapter 2

2 A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women

This chapter is reproduced with permission (Appendix A) from:

McMillan, A, Rulisa, S, Sumarah, M, Macklaim, JM, Renaud, J, Bisanz JE, Gloor JB, and Reid. (2015). A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Scientific Reports*, 5: 14174.

Supplemental material available for download at <http://www.nature.com/articles/srep14174>.

2.1 Introduction

The vaginal microbiota is dominated by *Lactobacillus* species in most women, predominantly by *L. iners* and *L. crispatus*¹⁻³. When these lactobacilli are displaced or outnumbered by a group of mixed anaerobes, belonging to the genus *Gardnerella*, *Prevotella*, *Atopobium* and others, this increase in bacterial diversity is associated with bacterial vaginosis (BV)¹⁻³. BV is the most common vaginal condition, affecting an estimated 30% of women at any given time⁴. While many women remain asymptomatic²⁻⁵, when signs and symptoms do arise they include an elevated vaginal pH>4.5, discharge, and malodor due to amines⁶⁻⁸. BV is also associated with a number of comorbidities, including increased transmission and acquisition of HIV and other sexually transmitted infections⁹, and increased risk of preterm labour¹⁰.

In most instances, diagnosis is dependent upon microscopy of vaginal samples to identify BV-like bacteria by morphology alone (Nugent Scoring¹¹), or in combination with clinical signs (Amsel Criteria¹²). The precision and accuracy of these methods are poor due to the diverse morphology of vaginal bacteria, the observation that many women with BV are asymptomatic, and subjectivity in microscopic examination¹³⁻¹⁵. Misdiagnosis creates stress for the patient, delays appropriate intervention and places a financial burden on the

health care system. A rapid test based on stable, specific biomarkers for BV would improve diagnostic accuracy and speed, and reduce costs through improved patient management.

The metabolome, defined as the complete set of small molecules in a given environment, has been studied in a variety of systems to identify biomarkers of disease^{16,17}, and advance our understanding of how the microbiota contributes to host metabolism¹⁸. Using an untargeted multiplatform metabolomics approach, combined with 16S rRNA gene sequencing, we demonstrate that the vaginal metabolome is driven by bacterial diversity, and identify biomarkers of clinical BV that can be reproduced in a blinded validation cohort. We further demonstrate that *Gardnerella vaginalis*, which has long been thought to be an important contributor to BV, is the likely source of one of the most specific compounds, GHB. This work provides a foundation for improved detection of disease and demonstrates how metabolomics can be utilized to identify validated sources of metabolites in microbial communities.

2.2 Materials and methods

2.2.1 Clinical samples

Premenopausal women between the ages of 18 and 55 were recruited at the University of Kigali Teaching Hospital (CHUK) and the Nyamata District Hospital in Rwanda. The Health Sciences Research Ethics Board at Western University, Canada, and the CHUK Ethics Committee, Rwanda granted ethical approval for all experiments involved in the study. The methods were carried out in accordance with the approved guidelines and all women provided written informed consent. Participants were excluded if they had reached menopause, had a current infection of gonorrhea, Chlamydia, genital warts, active genital herpes lesions, active syphilis, HIV, urinary tract infections, received drug therapy that may affect the vaginal microbiome, had unprotected sexual intercourse within the past 48 hours, used a vaginal douche, genital deodorant or genital wipe in past 48 hours, had taken any probiotic supplement in past 48 hours, or were menstruating at time of clinical visit. As materials for sample collection were limited, we set out to obtain an equal number of women with and without Nugent BV to ensure the study would be powered to test for BV biomarkers. To accomplish this, only women with suspected BV were recruited after the

quota of Nugent N women was met. After reviewing details of the study, participants gave their signed consent before the start of the study. For metabolome analysis, sterile Dacron polyester-tipped swabs (BD) were pre-cut with sterilized scissors and weighed in 1.5 ml microcentrifuge tubes prior to sample collection. Using sterile forceps to clasp the pre-cut swabs, a nurse obtained vaginal samples for metabolomic analysis by rolling the swab against the mid-vaginal wall. A second full-length swab was obtained for Nugent Scoring and 16S rRNA gene sequencing using the same method. Nugent Scoring was performed at CHUK by Amy McMillan. Vaginal pH was measured using pH strips. Samples were frozen within 2 hours of collection and stored at -20 °C or below until analysis.

2.2.2 Microbiome profiling

Vaginal swabs for microbiome analysis were extracted using the QIAamp DNA stool mini kit (Qiagen) with the following modifications: swabs were vortexed in 1 mL buffer ASL before removal of the swab and addition of 200 mg of 0.1 mm zirconia/silica beads (Biospec Products). Samples were mixed vigorously for 2 x 30 seconds at full speed with cooling at room temperature between (Mini-BeadBeater; Biospec Products). After heating to 95 °C for 5 minutes, 1.2 ml of supernatant was aliquoted into a 2ml tube and one-half an inhibitEx tablet (Qiagen) was added to each sample. All other steps were performed as per the manufacturers' instructions. Sample amplification for sequencing was carried out using the forward primer (ACACTCTTCCCTACACGACGCTCTTCCGATCTnnnn(8)CWACGCGARGAACC TTACC) and the reverse primer (CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTn(12)ACRACACGAGCT GACGAC) where nnnn indicates four randomly incorporated nucleotides, and (8) was a sample nucleotide specific barcode. The 5' end is the adapter sequence for the Illumina MiSeq sequencer and the sequences following the barcode are complementary to the V6 rRNA gene region. Amplification was carried out in 42 µL with each primer present at 0.8 pMol/mL, 20 µL GoTaq hot start colorless master mix (Promega) and 2 µL extracted DNA. The PCR protocol was as follows: initial activation step at 95 °C for 2 minutes and 25 cycles of 1 minute 95 °C, 1 minute 55 °C and 1 minute 72 °C.

All subsequent work was carried out at the London Regional Genomics Centre (LRGC, lrgc.ca, London, Ontario, Canada). Briefly, PCR products were quantified with a Qubit 2.0 Fluorometer and the high sensitivity dsDNA specific fluorescent probes (Life Technologies). Samples were mixed at equimolar concentrations and purified with the QIAquick PCR Purification kit (QIAGEN). Samples were paired-end sequenced on an Illumina Mi-Seq with the 600 cycle version 3 reagents with 2x220 cycles. Data was extracted from only the first read, since it spanned the entirety of the V6 region including the reverse primer and barcode

Resulting reads were extracted and de-multiplexed using modifications of in-house Perl and UNIX-shell scripts with operational taxonomic units (OTUs) clustered at 97% identity, similar to our reported protocol⁴⁸. Automated taxonomic assignments were carried out by examining best hits from comparison to the Ribosomal Database Project (rdp.cme.msu.edu) and manually curated by comparison to the Green genes database (greengenes.lbl.gov) and an in house database of vaginal sequences (Macklaim unpublished). Taxa with matches at least 95% similarity to query sequences were annotated as such. OTUs were summed to the genus level except for lactobacilli, and rare OTUs found at less than 0.5% abundance in any sample removed. Supplementary Table S1 displays the nucleotide barcodes and their corresponding samples. Reads were deposited to the Short Read Archive (BioProject ID: PRJNA289672). To control for background contaminating sequences, a no-template control was also sequenced. Barplots were constructed with R {r-project.org} using proportional values.

To avoid inappropriate statistical inferences made from compositional data, centered log-ratios (clr), a method previously described by Aitchison⁴⁹ and adapted to microbiome data was used with paired t-tests for comparisons of genus and species level data^{27,28}. The Benjamini Hochberg (False Discovery rate) method was used to control for multiple testing with a significance threshold of 0.1. All statistical analysis, unless otherwise indicated, was carried out using R (r-project.org).

2.2.3 Sample preparation GC-MS

Vaginal swabs were pre-cut into 1.5 mL tubes and weighed prior to and after sample collection to determine the mass of vaginal fluid collected. After thawing, swabs were eluted in methanol-water (1:1) in 1.5 mL microcentrifuge tubes to a final concentration of 50 mg vaginal fluid/mL, which corresponded to a volume ranging from 200-2696 μ L, depending on the mass of vaginal fluid collected. A blank swab eluted in 800 μ L methanol-water was included as a negative control. All samples were vortexed for 10 s to extract metabolites, centrifuged for 5 min at 10 621 g, vortexed again for 10 s after which time the brushes were removed from tubes. Samples were centrifuged a final time for 10 min at 10 621 g to pellet cells and 200 μ L of the supernatant was transferred to a GC-MS vial. The remaining supernatant was stored at -80 $^{\circ}$ C for LC-MS analysis. Next, 2 μ L of 1 mg/mL ribitol was added to each vial as an internal standard. Samples were then dried to completeness using a SpeedVac. After drying, 100 μ L of 2% methoxyamine-HCl in pyridine (MOX) was added to each vial for derivatization and incubated at 50 $^{\circ}$ C for 90 min. 100 μ L N- Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was then added and incubated at 50 $^{\circ}$ C for 30 min. Samples were then transferred to micro inserts before analysis by GC-MS (Agilent 7890A GC, 5975 inert MSD with triple axis detector). 1 μ L of sample was injected using pulsed splitless mode into a 30 m DB5-MS column with 10 m duraguard, diameter 0.35mm, thickness 0.25 μ m (JNW Scientific). Helium was used as the carrier gas at a constant flow rate of 1 ml/min. Oven temperature was held at 70 $^{\circ}$ C for 5 min then increased at a rate of 5 $^{\circ}$ C/min to 300 $^{\circ}$ C and held for 10 min. Solvent delay was set to 13 min to avoid solvent and a large lactate peak, and total run time was 61 min. Masses between 25 m/z and 600 m/z were selected by the detector. All samples were run in random order and a standard mix containing metabolites expected in samples was run multiple times throughout to ensure machine consistency.

2.2.4 Data processing GC-MS

Chromatogram files were deconvoluted and converted to ELU format using the AMDIS Mass Spectrometry software⁵⁰, with the resolution set to high and sensitivity to medium. Chromatograms were then aligned and integrated using Spectconnect⁵¹

(<http://spectconnect.mit.edu>), with the support threshold set to low. All metabolites found in the blank swab, or believed to have originated from derivatization reagents were removed from analysis at this time. After removal of swab metabolites, the IS matrix from Spectconnect was transformed using the additive log ratio transformation (alr)⁴⁹ and ribitol as a normalizing agent ($\log_2(x) / \log_2(\text{ribitol})$). Zeros were replaced with two thirds the minimum detected value on a per metabolite basis prior to transformation. All further metabolite analysis was performed using these alr transformed values.

Metabolites were initially identified by comparison to the NIST 11 standard reference database (<http://www.nist.gov/srd/nist1a.cfm>). Identities of metabolites of interest were then confirmed by authentic standards if available.

2.2.5 Global metabolomic analysis

In order to visualize trends in the metabolome as detected by GC-MS, principal component analysis (PCA) was performed using pareto scaling. To determine the percentage of variation in the metabolome that could be explained by a single variable we performed a series of partial least squares (PLS) regressions where each variable was used as a continuous latent variable. We tested every taxa, pH, Nugent score, pregnancy status, Shannon's diversity index and sample ID and compared the percent variation explained by the first component of each PLS. The variable with the highest value was determined to be most closely associated with the metabolome (Shannon's Diversity). Analysis was conducted in R using the PLS package and unit variance scaling. Jackknifing with 20% sample removal and 10 000 repetitions was then applied to determine 95% confidence intervals for each metabolite. Metabolites with confidence intervals that did not cross zero in both cohorts (pregnant and non-pregnant) were considered significantly associated with diversity. Heatmaps of significant metabolites were constructed using the `heatmap.2` function in R with average linkage hierarchical clustering and Manhattan distances. Unless specified otherwise, all tests for differential metabolites between groups were performed using unpaired t-test with a Benjamini-Hochberg (False Discovery Rate) significance threshold of $p < 0.01$ to account for multiple testing and multiple group comparisons.

16S rRNA microbial gene profiles generate compositional data that interferes with many

standard statistical analyses, including determining correlations²⁶⁻²⁸. We used the `aldex.corr` function from the ALDEx2 package to calculate the Spearman's rank correlation between each OTU abundance in 128 inferred technical replicates and that were transformed by center log-ratio transform^{27,28,49}. Spearman's rho values were converted to P values and corrected by the Benjamini-Hochberg procedure⁵² using the `cor.test` function in R. This approach is conceptually similar to that adopted by SPARCC²⁶, but calculates the correlation between the OTU abundances and continuous metadata variables. Heatmaps of correlation p values were constructed using the `heatmap.2` function in R with complete linkage hierarchical clustering and Euclidean distances.

Odds ratios of metabolites to identify Nugent BV from Normal were calculated from conditional logistic regressions performed on all metabolites using the `glm` function in R with 10 000 iterations and a binomial distribution. Metabolites with 95 % CI > 1 and $p < 0.01$ (unpaired t-test, Benjamini-Hochberg corrected) were determined to be significantly elevated in Nugent BV. "Nugent BV" was defined by the clinical definition of a score of 7-10, with a score of 0-3 being "Nugent Normal". ROC curves and forest plots were built in R using the `pROC` and `Gmisc` packages respectively.

2.3 Sample preparation LC-MS

To confirm GC-MS findings, samples which had at least 100 μL remaining after GC-MS were also analyzed by LC-MS (N=107). 100 μL of supernatant was transferred to vials with microinserts and directly injected into an Agilent 1290 Infinity HPLC coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific) with a HESI source. For HPLC, 2 μL of each sample was injected into a ZORBAX Eclipse plus C18 2.1 x 50mm x 1.6 micron column. Mobile phase (A) consisted of 0.1% formic acid in water and mobile phase (B) consisted of 0.1% formic acid in acetonitrile. The initial composition of 100% (A) was held constant for 30 s and decreased to 0% over 3.0 min. Mobile phase A was then held at 0% for 1.5 minutes and returned to 100% over 30s for a total run time of 5 min.

Full MS scanning between the ranges of m/z 50-750 was performed on all samples in both positive and negative mode at 140 000 resolution. The HESI source was operated under the following conditions: respectively nitrogen flow of 25 and 15 arbitrary units for the

sheath and auxiliary gas, probe temperature and capillary temperature of 425 °C and 260 °C, and spray voltage of 4.8 kV and 3.9 kV in positive and negative mode. The AGC target and maximum injection time were 3e6 and 500 ms respectively. For molecular characterization, every tenth sample was also analyzed with a data dependent MS² method where a 35 000 resolution full MS scan identified the top 10 signals above an 8.3e4 threshold which were subsequently selected at a 1.2 m/z isolation window for MS². Collision energy for MS² was 24, resolution 17 500, AGC target 1e5 and maximum injection time was 60ms. Blanks of pure methanol were run between every sample to limit carryover, and a single sample was run multiple times with every batch to account for any machine inconsistency. A blank swab extract was also run as a negative control.

For increased sensitivity, a separate LC-MS method was used for relative quantification of GHB in human samples. This was accomplished by selected ion monitoring in the mass range of 103.1 – 107.1 m/z in positive mode, and integrating the LC peak area of the [M+H⁺] ion (\pm 5 ppm).

2.3.1 Data processing LC-MS

After data acquisition Thermo .RAW files were converted to .MZML format using ProteoWizard⁵³ and imported into MZmine 2.11⁵⁴ (<http://mzmine.sourceforge.net>) for chromatogram alignment and deconvolution. Masses were detected using the Exact Mass setting and a threshold of 1E5. For Chromatogram Builder, minimum time was 0.05 min, minimum height 3E3, and m/z threshold set to 0.025 m/z or 8 ppm. Chromatogram deconvolution was achieved using the Noise Amplitude setting with the noise set to 5E4 and signal to 1E5 for negative mode. Due to an overall greater signal and noise in positive mode, the noise was adjusted to 6E5 and signal to 6.5E5 for positive mode. Join aligner was used to combine deconvoluted chromatograms into a single file with the m/z threshold set to 0.05 m/z or 10 ppm, weight for m/z and RT set to 20 and 10 respectively, and a RT tolerance of 0.4 min. After chromatograms were aligned, a single .CSV file was exported and all further analysis was carried out in R.

To confirm metabolites identified as significant by GC-MS in the LC-MS data set, the masses of metabolites of interest were searched in the LC-MS data set, and identities

confirmed by MS² using METLIN⁵⁵ and the Human Metabolome Database⁵⁶ online resources. Standards of metabolites of interest were also run to confirm identities when available. An unpaired t-test with Benjamini-Hochberg correction was used to determine metabolites significantly different between Nugent BV and Normal in the LC-MS data set. Metabolites with corrected $p < 0.05$ were considered statistically significant. Metabolites detected exclusively by LC-MS that have previously been associated with BV or health (lactate, trimethylamine) were also included in this analysis. Data was log base 10 transformed prior to data analysis and zeros replaced by two thirds the minimum detected value on a per metabolite basis. To determine optimal cut points of biomarkers for diagnostic purposes, cut points were computed from LC-MS data using the OptimalCutpoints package in R⁵⁷ and the Youden Index method.⁵⁸

2.3.2 Validation in blinded replication cohort

Women between the ages of 18 and 40 were recruited from an antenatal clinic at the Nyerere Dispensary in Mwanza, Tanzania as part of a larger study on the effect of micronutrient supplemented probiotic yogurt on pregnancy. The Medical research Coordinating Committee of the National Institute for Medical Research (NIMR), as well as the Health Sciences Research Ethics Board at Western University granted approval for all experiments involved in the study. The methods were carried out in accordance with the approved guidelines and all women provided written informed consent. The study was registered with clinicaltrials.gov (NCT02021799). Samples were collected using the methods mentioned above, and Nugent scores performed by research technicians at NIMR in Mwanza, Tanzania. A subset of samples was selected based on these Nugent scores by a third party, who ensured there was not repeated sampling of any women. Amy McMillan, who performed metabolite analysis, was blinded to the Nugent scores for the duration of sample processing and data analysis. Biomarkers were quantified in samples by LC-MS using the protocols mentioned above. The study was unblinded after the submission of BV status based on the ratio cut points established in the Rwandan data set.

2.3.3 Identification of putative GHB dehydrogenases in *G. vaginalis*

The protein sequence of a *bona fide* 4-hydroxybutyrate (GHB) dehydrogenase isolated from *Clostridium kluyveri*²⁹ (GI:347073) was blasted against all strains of *G. vaginalis* in the NCBI protein database. Blast results identified multiple isolates containing a putative protein with 44-46% identity to the GHB dehydrogenase from *C. kluyveri*. The strain used for *in vitro* experiments (*G. vaginalis* ATCC 14018) was not present in the NCBI protein database, however a nucleotide sequence in 14018 with 100% nucleotide identity to a putative 4-hydroxybutyrate dehydrogenases in strain ATCC 14019 (GI:311114893) was identified, indicating potential for GHB production by strain 14018.

2.3.4 *In vitro* extraction of GHB from vaginal isolates

Bacteria isolated from the vagina were difficult to grow consistently in liquid media. To circumvent this, a lawn of bacteria was plated and metabolites were extracted from agar punches. All strains were grown on Columbia Blood Agar (CBA) plates using 5% sheep's blood for 96h under strict anaerobic conditions, with the exception of *L. crispatus*, which was grown on de Man Rogosa Sharp (MRS) agar for 48 h. To extract metabolites, 16 agar punches 5 mm in diameter were taken from each plate and suspended in 3 mL 1:1 Me:H₂O. Samples were then sonicated in a water bath sonicator for 1h, transferred to 1.5 ml tubes after vortexing and spun in a desktop microcentrifuge for 10 min at 10 621 g to pellet cells. 200 µl of supernatant was then aliquoted for GC-MS described above. The area of each peak was integrated using ChemStation (Agilent) by selecting m/z 233 in the range of 14-16 min. Initial peak width was set to 0.042 and initial threshold at 10. An authentic standard of GHB was run with samples to confirm identification. Succinate production by vaginal isolates was measured from the same GC-MS run, and quantified using Spectconnect as described above. Un-inoculated media was used as a control and experiments were repeated three times with technical duplicates.

2.4 Results

2.4.1 The vaginal metabolome is most correlated with bacterial diversity

We completed a comprehensive untargeted metabolomic analysis of vaginal fluid in two cross-sectional cohorts of Rwandan women: pregnant (P, n=67) and non-pregnant (NP,

n=64) (Supplementary Table S1). To normalize the amount of sample collected, vaginal swabs were weighed prior to and after collection and normalized to equivalent concentrations. This enabled us to collect precise measurements of metabolites in vaginal fluid. Metabolite profiling was carried out using both gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS), and microbiota composition by 16S rRNA gene sequencing.

The metabolome determined by GC-MS contained 128 metabolites (Supplementary Table S2). We conducted a series of partial least squares (PLS) regression analyses to determine the single variable that could best explain the variation in the metabolome. In both cohorts, the diversity of the microbiota, as measured using Shannon's Diversity¹⁹, was the factor that explained the largest percent variation in the metabolome (Supplementary Table S3), demonstrating that the vaginal metabolome is most correlated with bacterial diversity (Figure 2-1). Metabolites robustly associated with this diversity (95% CI < or > 0)(Figure 2-1) were determined by jackknifing, and within this group, metabolites associated with extreme diversity tended to have less variation in the jackknife replicates, and were common to both pregnant and non-pregnant women. This identified a core set of metabolites associated with diversity.

The two cohorts overlapped by principal component analysis (PCA) (Supplementary Fig. S1), and no metabolites were significantly different between pregnant and non-pregnant women (unpaired t-test, Benjamini-Hochberg $p > 0.01$). Thus, the cohorts were combined for all further analysis.

2.4.2 Metabolites and taxa associated with diversity

A single PLS regression was performed on all samples with Shannon's diversity as a continuous latent variable (Supplementary Fig. S2). Samples were then ordered by their position on the 1st component of this PLS. The diversity indices, microbiota and metabolites associated with diversity of PLS ordered samples are shown in Figure 2-2. The vaginal microbiota of Rwandan women were similar to women from other parts of the world, with the most abundant species being *L. iners* followed by *L. crispatus*^{1-3,20} (Figure 2-2, Supplementary Table S4). Women with high bacterial diversity were dominated by a mixture of anaerobes, including *Gardnerella*, *Prevotella*, *Sneathia*, *Atopobium*, *Dialister*, and *Megasphaera* species.

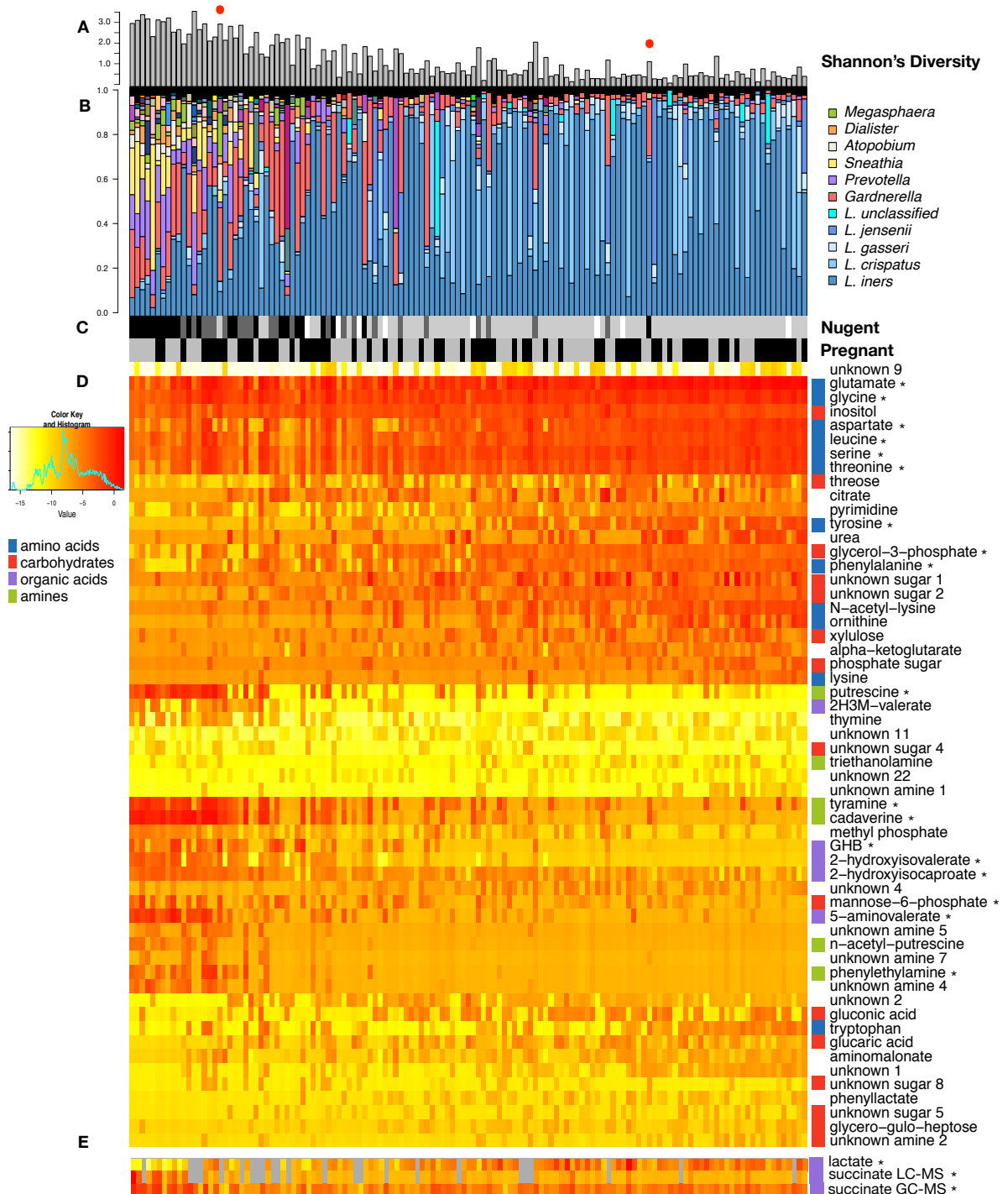


Figure 2-2. Bacterial taxa and metabolites correlated with bacterial diversity in the vagina.

Cohorts (non-pregnant and pregnant) were combined prior to analyses. Samples are ordered by their position on the first component (x-axis) of a partial least squares regression (PLS) built from metabolites using bacterial diversity as a continuous latent

variable (Supplementary Fig. S2). Diversity was calculated using Shannon's diversity (A). Red dots indicate samples clearly misclassified by Nugent. Barplots (B) display the vaginal microbiota profiled using the V6 region of the 16S rRNA gene. Each bar represents a single sample from a single woman, and each colour a different bacterial taxa. (C) Nugent Score (black=7-10 (BV), dark grey=4-6 (Int), light grey=1-3 (N), white=ND) and pregnancy status (black=P, grey=NP). (D) Heatmap of GC-MS detected metabolites which were robustly associated with diversity in both cohorts (Jackknifing, 95% CI <0>). Metabolites are clustered using average linkage hierarchical clustering. (E) Lactate and succinate abundance. Grey = ND. (*) indicates metabolites confirmed by authentic standards.

2.4.3 Succinate is not associated with diversity or clinical BV, and is produced by *L. crispatus*

Succinate and lactate abundance are shown in panel E of Figure 2-2. Succinate levels, and the succinate:lactate ratio have historically been associated with BV²¹⁻²³, and succinate has been postulated to play an immunomodulatory role²³. Here we show that succinate is not associated with bacterial diversity, nor is it significantly elevated in clinical BV as defined by Nugent scoring. This trend was independent of the detection method used. In addition, succinate was elevated in women dominated by *L. crispatus* compared to *L. iners*-dominated women (unpaired t-test, Benjamini-Hochberg $p < 0.01$) (Figure 2-3), indicating *L. crispatus* may produce succinate *in vivo*, a phenomenon that has been demonstrated *in vitro*²⁴. We extracted metabolites from vaginal isolates grown on agar plates and confirmed that succinate is produced by *L. crispatus in vitro*, but not by *L. iners* (). Succinate was also produced by *Prevotella bivia*, and *Mobiluncus curtisii*, but not by *G. vaginalis*.

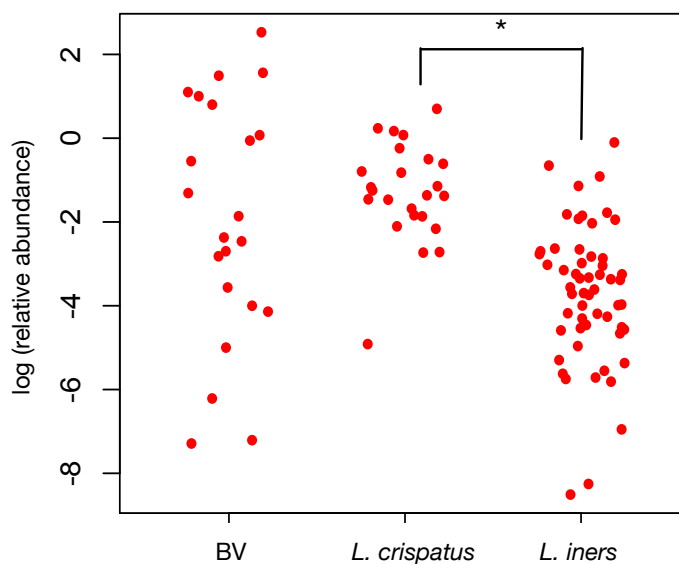


Figure 2-3. Relative abundance of succinate in women dominated by *L. crispatus*, *L. iners*, or Nugent BV detected by GC-MS.

(*) $p < 0.01$, unpaired t-test, Benjamini-Hochberg corrected.

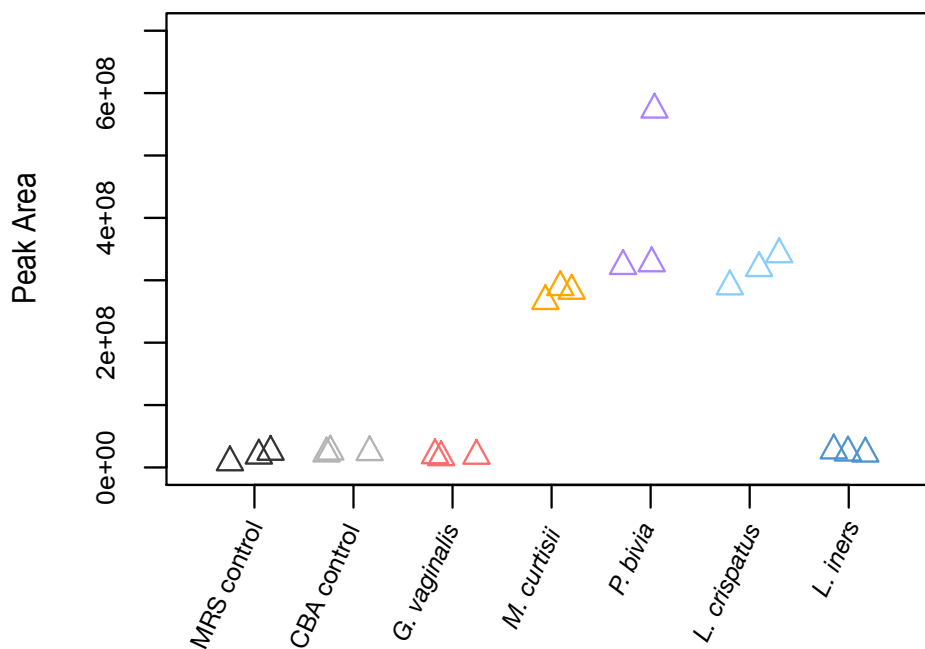


Figure 2-4. Succinate production by vaginal isolates.

Bacteria were grown on agar plates and succinate detected by GC-MS. Three biological replicates are shown in technical duplicate. CBA: Columbia Blood Agar, MRS: de Man Rogosa Sharpe Agar.

2.4.4 Metabolites associated with diversity are sensitive and specific for clinical BV

We defined clinical BV by the Nugent method, which is the current gold standard for BV diagnosis¹¹. This microscopy-based technique defines BV as a score of 7-10 when low numbers of lactobacilli morphotypes are observed, and high numbers of short rods are present, which are presumed to represent BV associated bacteria. Nugent Normal (N) is defined as a score of 1-3, indicating almost exclusively *Lactobacillus* morphotypes. Intermediate samples are given a score of 4-6 and do not fit into either group. Although Nugent scores correlated well with bacterial diversity in our study, it was apparent from the microbiota and metabolome profiles that two samples (41 and 145) had clearly been misclassified by Nugent (Figure 2-2, red dots). The Nugent status of these samples was therefore corrected prior to further analyses.

In total we identified 49 metabolites that were significantly different between clinical BV and N (unpaired t-test, Benjamini-Hochberg $p < 0.01$, Supplementary Table S2). We determined the odds ratio (OR) for BV based on conditional logistic regressions of all individual metabolites detected by GC-MS (Supplementary Table S2) to determine if the metabolites we associated with high bacterial diversity could accurately identify clinical BV as defined by Nugent scoring. Metabolites significantly elevated in Nugent BV (unpaired t-test, Benjamini-Hochberg $p < 0.01$) with $OR > 1$ are shown in Figure 2-5A. Succinate was included as a comparator, although it did not reach significance. Both GHB and 2HV were significantly higher in women with BV, and had $OR > 2.0$, demonstrating they are indicators not only of high bacterial diversity, but also clinical BV. Receiver operating characteristics (ROC) curves built from LC-MS data determined that high 2HV, high GHB, low lactate and low tyrosine were the most sensitive and specific biomarkers for BV, with the largest area under the curve (AUC) achieved using the ratio of 2HV:tyrosine (AUC=0.993)(Figure 2-5B-D). ROC curves of GC-MS data identified similar trends, with the largest AUC achieved by the ratio of GHB:tyrosine (AUC=0.968) (Supplementary Table S6).

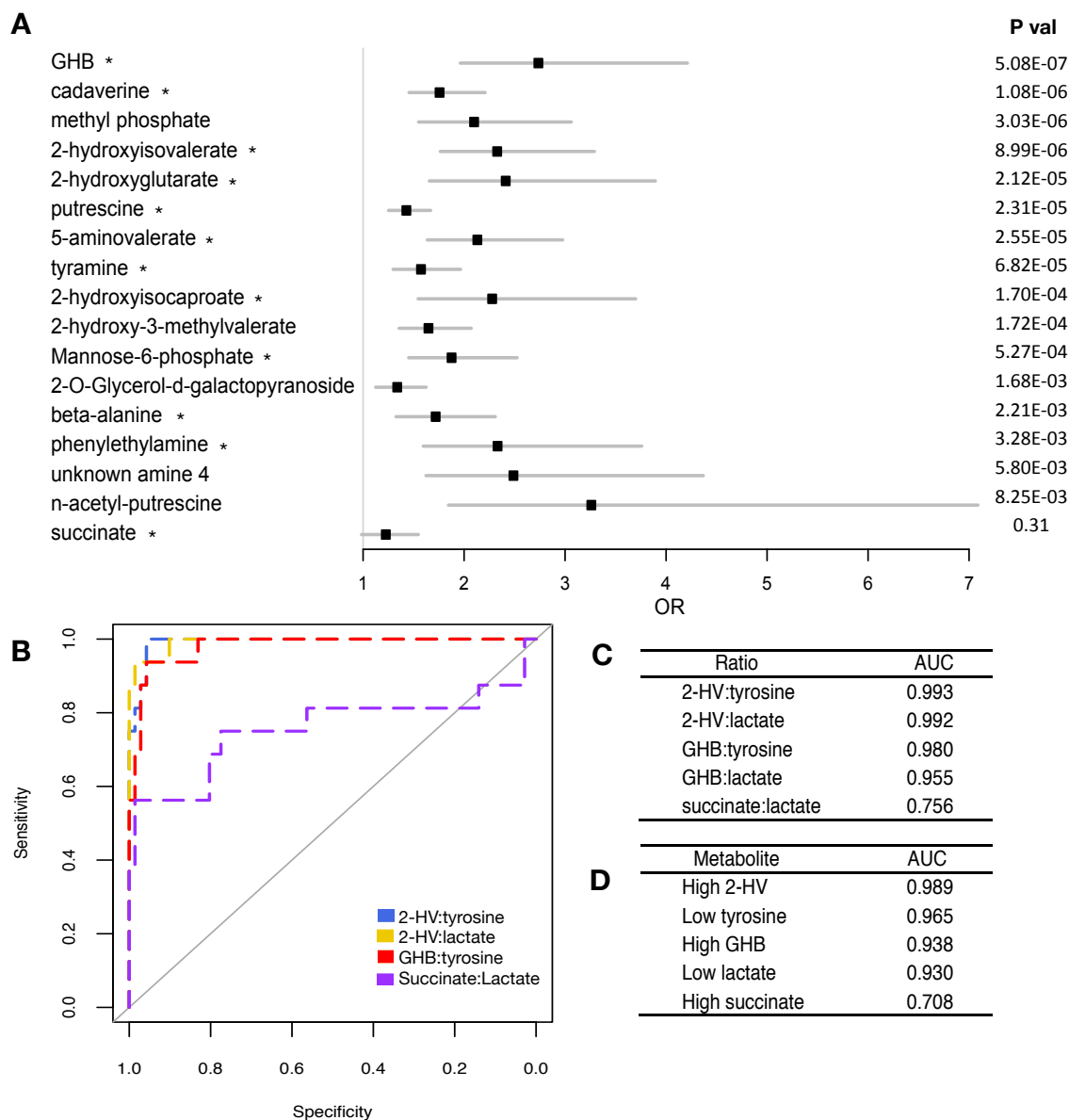


Figure 2-5. Comparison of biomarkers to identify Nugent BV from Nugent N.

(A) Odds ratios (OR) of metabolites with positive predictive value to identify Nugent BV. Bars represent 95% Confidence Intervals. Metabolites were detected by GC-MS and P values generated from unpaired t-tests with a Benjamini-Hochberg correction to account for multiple testing ($p < 0.01$). (*) indicates metabolites confirmed by authentic standards. (B) Receiver operating characteristic (ROC) curves of metabolite ratios to identify Nugent BV from Nugent N. Ratios with largest area under the curve (AUC) are shown, along with succinate:lactate as a comparator. (C) AUC of selected metabolite ratios to identify Nugent BV. (D) AUC of metabolites alone to identify Nugent BV. Panels B-D were built from LC-MS data. GHB: γ -hydroxybutyrate, 2-HV:2-hydroxyisovalerate.

The optimal cut points for the GHB:tyrosine (0.621) and 2HV:tyrosine (0.882) ratios were determined by selecting values which maximized the sensitivity and specificity for BV. These cut points were generated excluding Nugent intermediate samples, however when cut points were applied to intermediates, they grouped equally with N or BV, and samples with smaller proportions of lactobacilli tended to group with BV (Figure 2-6). PCA of metabolite data verified that Nugent intermediate samples do not cluster separately from BV or N (Supplementary Fig. S1 C), indicating they are not a metabolically distinct group.

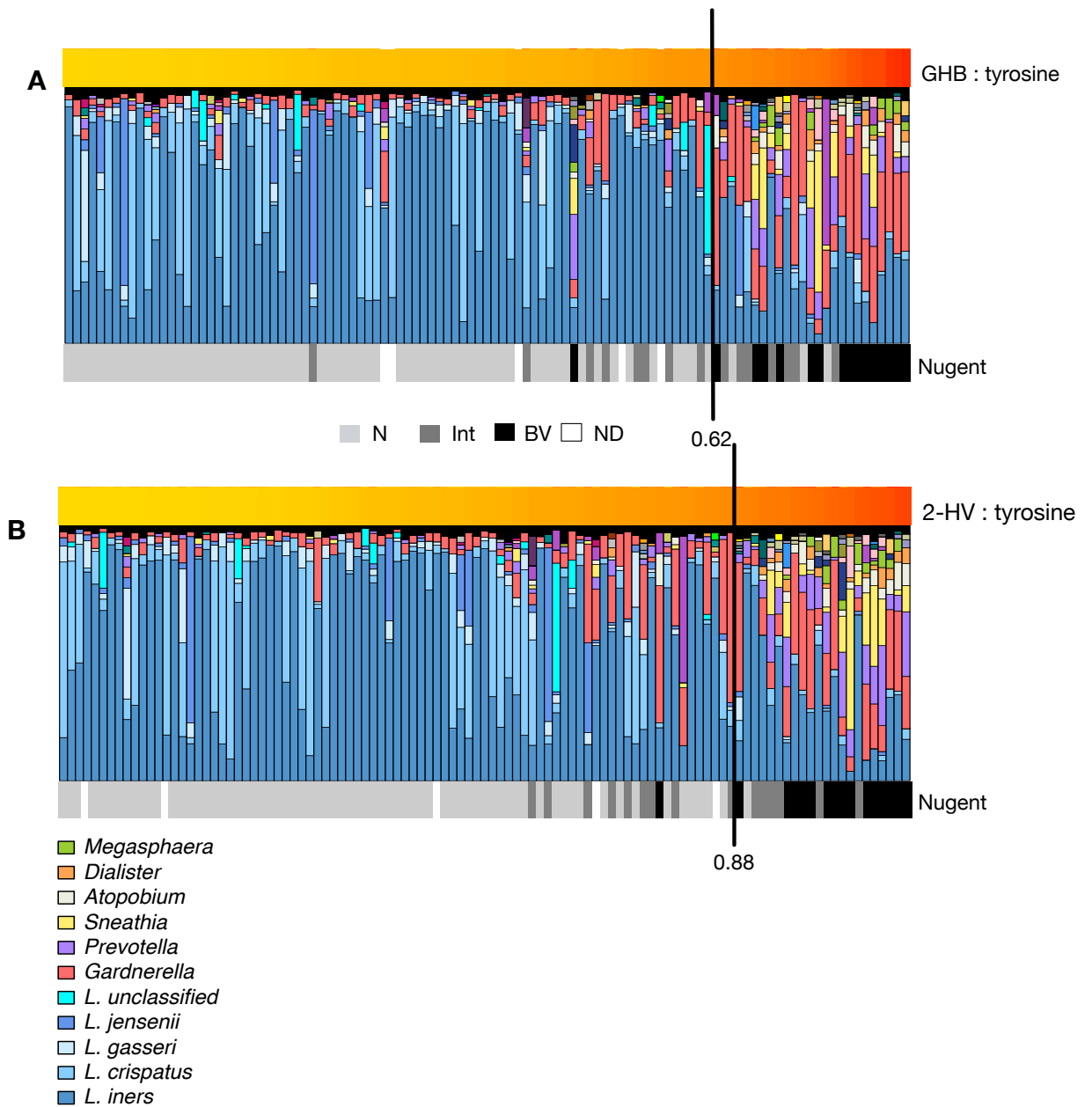


Figure 2-6. Biomarker cut points effectively group Nugent Intermediate samples as BV or N.

Barplots display the vaginal microbiota of Rwandan women sorted by (A) GHB:tyrosine or (B) 2HV:tyrosine. Each bar represents a single sample from a single woman and each colour a different bacterial taxa. Nugent scores are indicated below barplots. Black lines indicate ratio cut point for Nugent BV. Ratios were calculated from LC-MS data.

2.4.5 Validation of biomarkers in a blinded replication cohort from Tanzania

We validated these biomarkers in a blinded cohort of 45 pregnant women from Mwanza, Tanzania.²⁵ Using the 2HV:tyrosine cut point identified in the Rwanda data set, we identified Nugent BV with 89% sensitivity and 94% specificity in the validation set (AUC=0.946), demonstrating our findings are reproducible in an ethnically distinct population (Figure 2-7, Supplementary Table S7). The GHB:tyrosine ratio cut point was slightly less specific (88%), with an AUC of 0.948. We confirmed that succinate was not significantly different between Nugent N and BV in the validation set, nor was the succinate:lactate ratio.

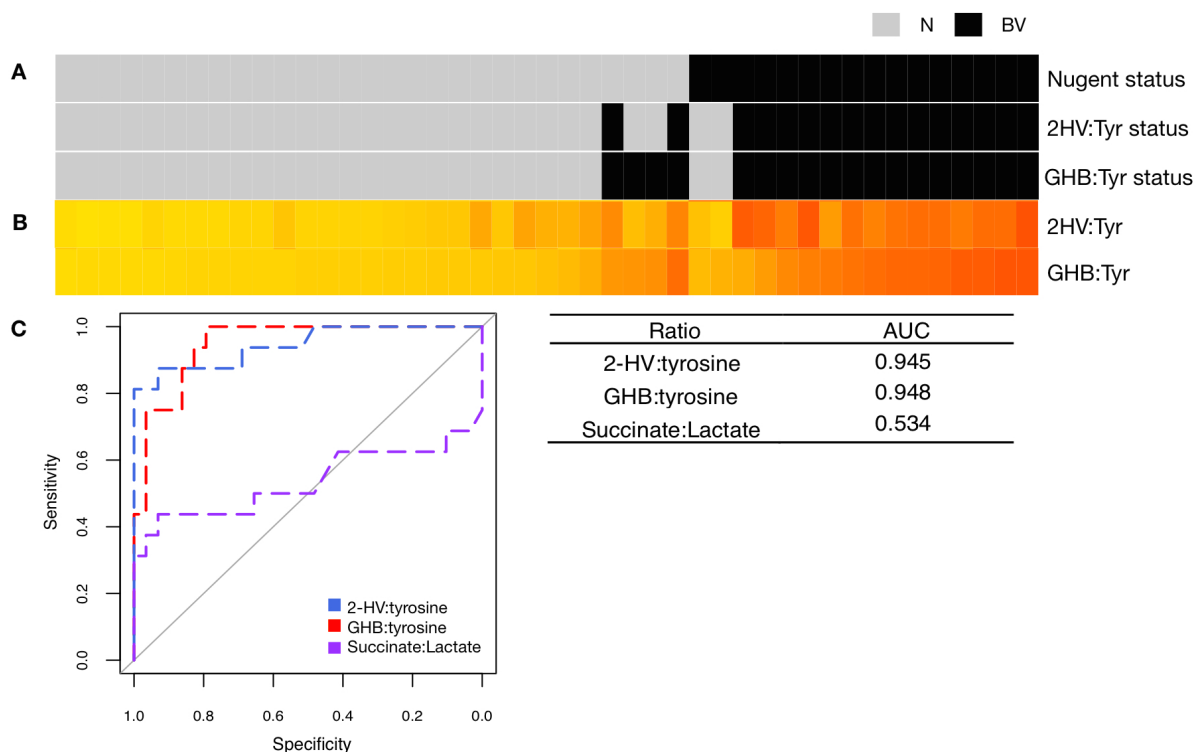


Figure 2-7. Biomarker validation in a blinded replication cohort of 45 women from Tanzania.

(A) BV status as defined by Nugent Score or ratio cut points identified in the Rwandan discovery data set. Black=BV, Gray=N. (B) Heatmap of ratio values. (C) ROC curves and AUC of ratios to identify Nugent BV from N in the validation set. 2HV: 2-hydroxyisovalerate, GHB: γ -hydroxybutyrate, Tyr: tyrosine.

2.4.6 Identification of *G. vaginalis* as a producer of GHB

Correlations between metabolites and the OTU abundances were performed using a method that took into account both the compositional nature of 16S rRNA gene survey data and the technical variation²⁶⁻²⁸. Metabolites and taxa which contained any correlation below a Benjamini-Hochberg corrected $p < 0.01$ are displayed as a heatmap in Figure 2-8. Tyramine, putrescine, and cadaverine were most correlated with *Dialister* (Spearman's $R = 0.54, 0.51, 0.61, p < 0.001$) (Supplementary Table S8), indicating this genus may contribute to malodor. GHB was most correlated with *G. vaginalis* (Spearman's $R = 0.56, p < 0.001$), while 2HV was most correlated with *Dialister*, *Prevotella*, and *Gardnerella* (Spearman's $R = 0.55, 0.48, 0.47, p < 0.001$).

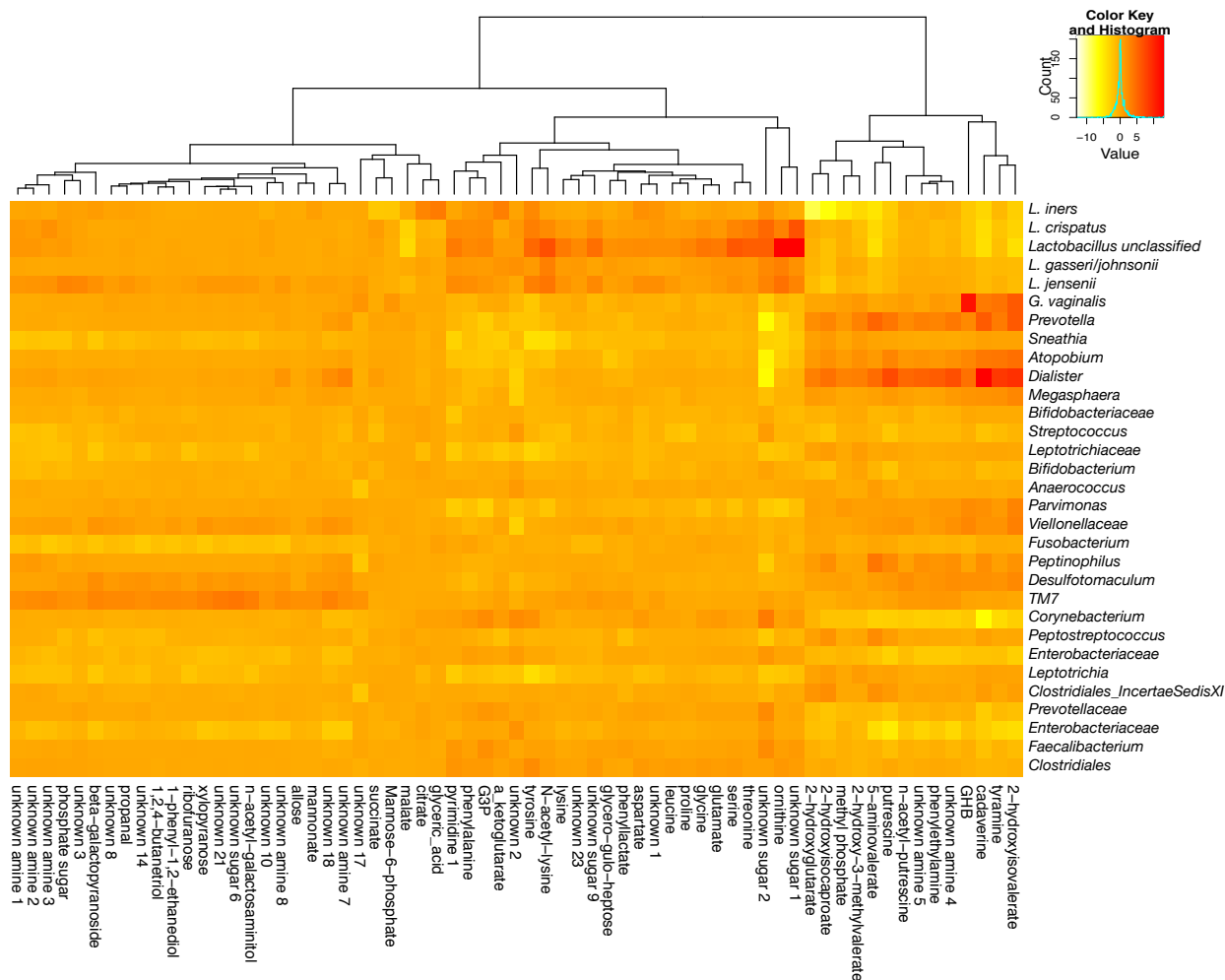


Figure 2-8. Correlations between metabolites and taxa which are robust to random sampling of the underlying data.

P values (Benjamini-Hochberg corrected) of Spearman's correlations are plotted on a log scale. The sign of each p value corresponds to the directionality of the correlation. Only metabolites and taxa for which any p values are < 0.01 are displayed.

The correlation between GHB and *G. vaginalis* was investigated since this was an unexpected metabolite that was predictive for both Shannon's diversity and Nugent BV. Examination of available genomes showed that many strains of *G. vaginalis* possess a putative GHB dehydrogenase (annotated as 4-hydroxybutyrate dehydrogenase). Metabolites were extracted from bacterial colonies grown on agar plates and reproducibly detected GHB in *G. vaginalis* extracts well above control levels (unpaired t-test, $p < 0.05$), but did not detect GHB from other species commonly associated with BV (Figure 2-9,

Supplementary Table S9). These data suggest that *G. vaginalis* is the primary source of GHB detected *in vivo*.

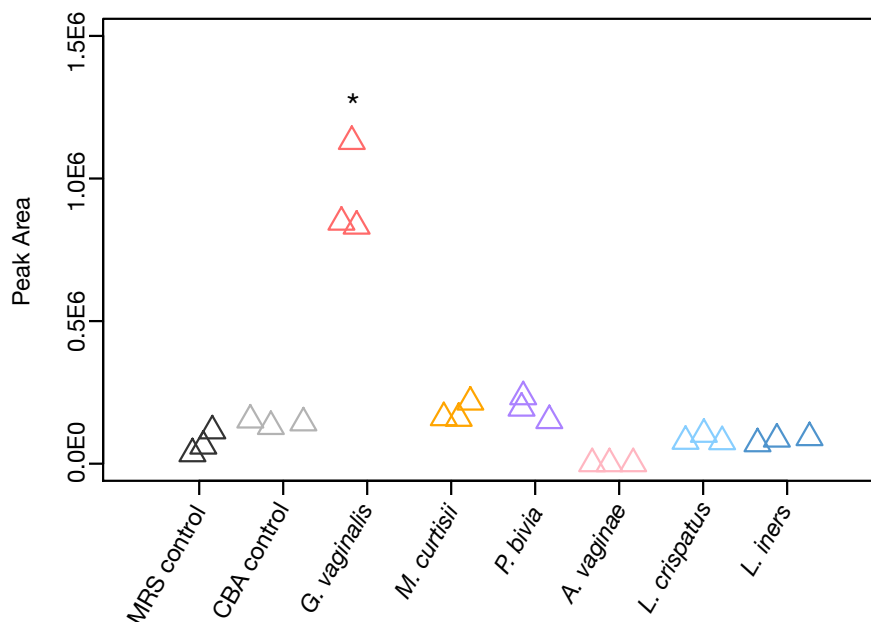


Figure 2-9. GHB is produced by *Gardnerella vaginalis*.

GHB was extracted from bacteria grown on agar plates and detected by GC-MS. Values from three independent experiments are shown where each point was generated from an average of technical duplicates. * $p < 0.05$, unpaired t-test.

2.5 Discussion

We have demonstrated that alterations in the vaginal metabolome are driven by bacterial diversity in both pregnant and non-pregnant Rwandan women, and identified 2HV and GHB as highly specific biomarkers of clinical BV, the latter of which we attribute to production by *G. vaginalis*. Extremely accurate results were obtained by controlling for the mass of vaginal fluid collected, however we recognize this may not be logistically possible in a clinical setting. To circumvent this need, biomarkers were expressed as ratios to the amino acid tyrosine, which we identified as the most differential amino acid in health (Supplementary Table S2). Using optimal cut points of these ratios we predicted 91% of Nugent BVs in a blinded replication cohort, demonstrating the reproducibility of our findings. These cut points also accurately classified Nugent intermediate samples into

groups with similar microbiota profiles dominated by either lactobacilli (N) or anaerobes (BV).

Although we demonstrated production of GHB by *G. vaginalis*, it is important to note that no single organism has been identified as the cause of BV, and *G. vaginalis* is present in many women with a lactobacilli-dominated microbiota. However, GHB is metabolized from succinate in other bacteria^{29,30}, suggesting a similar pathway could exist in *G. vaginalis*. Succinate-producing genera may therefore be required, making *G. vaginalis* essential, but not sufficient for GHB production in the vagina. This remains to be tested.

The predominant pathway for succinate production in bacteria is from pyruvate via anaerobic respiration. The genes for this pathway are expressed *in vivo* and differentiate BV from N³¹. Srinivasan et al³² recently proposed an alternate pathway whereby succinate is produced from putrescine via gamma-aminobutyric acid (GABA). Although this pathway is plausible, it is unlikely given many of the enzymes are either not expressed *in vivo* or are absent from the genomes of common vaginal organisms³¹.

Despite previous findings that succinate is elevated in BV^{21-23,32}, it was not differential in our study. This unexpected outcome is likely a result of normalizing sample weights prior to analysis, used to ensure the most consistent measurements of metabolites. Succinate was one of the most abundant metabolites detected in vaginal fluid in our study (Supplementary Table. S2), and was present in nearly all samples regardless of BV status. The universal presence of succinate make it more susceptible to dilution effects compared to GHB and 2HV, which were less abundant and below our detection limit in many non-BV samples. Other groups have reported large ranges in succinate abundance in women with BV^{21,22}, or used pooled samples²², which could account for additional disparities in results. Differences in succinate abundance may have been more pronounced in previous studies if there were a lack of *L. crispatus*-dominated women, which our data demonstrate is a succinate producer (Figure 2-3). *L. crispatus* contains all the enzymes necessary to produce succinate from malate with the exception of malate dehydrogenase (MDH). However, the pathway is annotated as complete at <http://biocyc.org/LCRI491076-HMP/missing-rxns.html>, with the closely

related enzyme lactate dehydrogenase (LDH). As there is increased expression of succinate-producing pathways during BV³⁰, it is probable that large amounts are produced initially, but then rapidly converted to other compounds, such as GHB, by the microbiota and/or host.

In addition to GHB, 2HV was identified as a highly specific biomarker for BV. 2HV is produced from breakdown of branched chain amino acids in humans³⁴ and some bacteria³⁵⁻³⁷. When the trend for amino acid depletion in BV is considered, these findings suggest increased amino acid catabolism in this condition. Some of these amino acids are converted to the amines cadaverine, tyramine, and putrescine, which are also associated with BV. These odor-causing compounds were most correlated with *Dialister*. Yeoman *et al.*³⁸ also linked amines to *Dialister* species, and the decarboxylating genes required for amine production are expressed by this genus *in vivo*³⁰. These data strongly suggest that *Dialister* is one of the genera responsible for malodor in the vagina. Given the small proportion of this genus in women with BV (0.2-8% in our study), this emphasizes the need for functional characterizations of the microbiome using metabolomic and transcriptomic approaches.

The taxa that constitute the vaginal microbiota are highly conserved across different populations^{1-3,20,39}, although prevalence of certain taxa differs between ethnicities^{1,39}. These observations lead us to believe that GHB, 2HV and their tyrosine ratios will be globally applicable for the diagnosis of BV. Our ability to replicate findings in a distinct population strongly supports this theory. Srinivasan *et al.*³² concurrently identified elevated GHB in the vaginal fluid of American women with BV³³, however they were not able to replicate this result in a second cohort. This could be due in part to the use of cervicovaginal lavages for sample collection or the use of different detection methods between cohorts. 2HV (annotated as alpha-hydroxyisovalerate) was also identified as differential in their study, but was not tested in the replication cohort. These observations further validate our findings and demonstrate these biomarkers are robust to the effects of dilution.

The exact role, if any, of GHB and 2HV in the etiology of BV is unknown. Systemically GHB has both inhibitory and excitatory effects through activation of the GABA(B) and perhaps GABA(A) receptors in the brain, resulting in stimulatory and sedative effects if

taken at high doses⁴⁰⁻⁴². The effects of GHB at other sites remain elusive. Future work should attempt to elucidate biological function of GHB and other novel metabolites to determine what effect (if any) they have on lactobacilli and the vaginal environment.

Although we did not identify any metabolites that differed significantly between pregnant and non-pregnant women, it should be noted that patient selection was biased to include an even number of women with and without Nugent BV. Our study was not designed to test if the metabolome differed during pregnancy, but rather if the metabolic signatures of BV were similar between pregnant and non-pregnant women. Other groups have noted decreased bacterial diversity during pregnancy and across gestational age⁴³⁻⁴⁵. These observations suggest differences in the metabolome of pregnant women would be observable in a larger randomly sampled population, and may include elevated levels of metabolites associated with low diversity such as amino acids.

In summary, we have demonstrated using an untargeted, multiplatform approach that differences in the vaginal metabolome are driven by bacterial diversity. Other metabolomic studies have focused on symptom-associated metabolites^{32,38}, changes after treatment⁴⁶, or longitudinal changes in a few subjects⁴⁷, and included exclusively non-pregnant women. Several highly specific biomarkers were identified for clinical BV that are independent of pregnancy status, and we replicated this result in a blinded cohort. By combining high-throughput sequencing with advanced mass spectrometry techniques we have shown how *in vivo* metabolite information can be used to identify validated sources of metabolic end products in bacterial communities. These techniques can be applied to many systems where organisms may be fastidious or difficult to culture, and provide a much-needed link between microbial composition and function.

2.6 References

- (1) Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G. M.; Koenig, S. S. K.; McCulle, S. L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C. O.; Brotman, R. M.; Davis, C. C.; Ault, K.; Peralta, L.; Forney, L. J. Vaginal Microbiome of Reproductive-Age Women. *Proc. Natl. Acad. Sci.* **2010**, *108* (Supplement_1), 4680–4687.

- (2) Hummelen, R.; Fernandes, A. D.; Macklaim, J. M.; Dickson, R. J.; Chantalucha, J.; Gloor, G. B.; Reid, G. Deep Sequencing of the Vaginal Microbiota of Women with HIV. *PLoS One* **2010**, *5* (8).
- (3) Fredricks, D. N.; Fiedler, T. L.; Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **2005**, *353* (18), 1899–1911.
- (4) Koumans, E. H.; Sternberg, M.; Bruce, C.; McQuillan, G.; Kendrick, J.; Sutton, M.; Markowitz, L. E. The Prevalence of Bacterial Vaginosis in the United States, 2001-2004; Associations with Symptoms, Sexual Behaviors, and Reproductive Health. *Sex. Transm. Dis.* **2007**, *34* (11), 864–869.
- (5) Klebanoff, M. A.; Schwebke, J. R.; Zhang, J.; Nansel, T. R.; Yu, K.-F.; Andrews, W. W. Vulvovaginal Symptoms in Women with Bacterial Vaginosis. *Obstet. Gynecol.* **2004**, *104*, 267–272.
- (6) Sobel, J. D.; Karpas, Z.; Lorber, A. Diagnosing Vaginal Infections through Measurement of Biogenic Amines by Ion Mobility Spectrometry. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2012**, *163* (1), 81–84.
- (7) Wolrath, H.; Forsum, U.; Larsson, P. G.; Borén, H. Analysis of Bacterial Vaginosis-Related Amines in Vaginal Fluid by Gas Chromatography and Mass Spectrometry. *J. Clin. Microbiol.* **2001**, *39* (11), 4026–4031.
- (8) Wolrath, H.; Borén, H.; Hallén, A.; Forsum, U. Trimethylamine Content in Vaginal Secretion and Its Relation to Bacterial Vaginosis. *APMIS* **2002**, *110* (11), 819–824.
- (9) Sha, B. E.; Zariffard, M. R.; Wang, Q. J.; Chen, H. Y.; Bremer, J.; Cohen, M. H.; Spear, G. T. Female Genital-Tract HIV Load Correlates Inversely with Lactobacillus Species but Positively with Bacterial Vaginosis and Mycoplasma Hominis. *J. Infect. Dis.* **2005**, *191* (1), 25–32.
- (10) Das, T. R.; Jahan, S.; Begum, S. R.; Akhtar, M. F. Association between Bacterial Vaginosis and Preterm Delivery. *Mymensingh Med. J.* **2011**, *20* (1), 115–120.

- (11) Nugent, R. P.; Krohn, M. A.; Hillier, S. L. Reliability of Diagnosing Bacterial Vaginosis Is Improved by a Standardized Method of Gram Stain Interpretation. *J. Clin. Microbiol.* **1991**, *29* (2), 297–301.
- (12) Amsel, R.; Totten, P. A.; Spiegel, C. A.; Chen, K. C.; Eschenbach, D.; Holmes, K. K. Nonspecific Vaginitis. Diagnostic Criteria and Microbial and Epidemiologic Associations. *Am. J. Med.* **1983**, *74* (1), 14–22.
- (13) Chaijareenont, K.; Sirimai, K.; Boriboonhirunsarn, D.; Kiriwat, O. Accuracy of Nugent's Score and Each Amsel's Criteria in the Diagnosis of Bacterial Vaginosis. *J. Med. Assoc. Thail.* **2004**, *87* (11), 1270–1274.
- (14) Sha, B. E.; Chen, H. Y.; Wang, Q. J.; Zariffard, M. R.; Cohen, M. H.; Spear, G. T. Utility of Amsel Criteria, Nugent Score, and Quantitative PCR for *Gardnerella vaginalis*, *Mycoplasma hominis*, and *Lactobacillus Spp.* for Diagnosis of Bacterial Vaginosis in Human Immunodeficiency Virus-Infected Women. *J. Clin. Microbiol.* **2005**, *43* (9), 4607–4612.
- (15) Schwebke, J. R.; Hillier, S. L.; Sobel, J. D.; McGregor, J. A.; Sweet, R. L. Validity of the Vaginal Gram Stain for the Diagnosis of Bacterial Vaginosis. *Obstet. Gynecol.* **1996**, *88* (4 I), 573–576.
- (16) Mayers, J. R.; Wu, C.; Clish, C. B.; Kraft, P.; Torrence, M. E.; Fiske, B. P.; Yuan, C.; Bao, Y.; Townsend, M. K.; Tworoger, S. S.; Davidson, S. M.; Papagiannakopoulos, T.; Yang, A.; Dayton, T. L.; Ogino, S.; Stampfer, M. J.; Giovannucci, E. L.; Qian, Z. R.; Rubinson, D. A.; Ma, J.; Sesso, H. D.; Gaziano, J. M.; Cochrane, B. B.; Liu, S.; Wactawski-Wende, J.; Manson, J. E.; Pollak, M. N.; Kimmelman, A. C.; Souza, A.; Pierce, K.; Wang, T. J.; Gerszten, R. E.; Fuchs, C. S.; Vander Heiden, M. G.; Wolpin, B. M. Elevation of Circulating Branched-Chain Amino Acids Is an Early Event in Human Pancreatic Adenocarcinoma Development. *Nat. Med.* **2014**, *20* (10), 1193–1198.
- (17) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-

- Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M. Metabolomic Profiles Delineate Potential Role for Sarcosine in Prostate Cancer Progression. *Nature* **2009**, *457* (7231), 910–914.
- (18) Wikoff, W. R.; Anfora, A. T.; Liu, J.; Schultz, P. G.; Lesley, S. a; Peters, E. C.; Siuzdak, G. Metabolomics Analysis Reveals Large Effects of Gut Microflora on Mammalian Blood Metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (10), 3698–3703.
- (19) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27* (July 1928), 379–423.
- (20) Zhou, X.; Hansmann, M. A.; Davis, C. C.; Suzuki, H.; Brown, C. J.; Sch?tte, U.; Pierson, J. D.; Forney, L. J. The Vaginal Bacterial Communities of Japanese Women Resemble Those of Women in Other Racial Groups. *FEMS Immunol. Med. Microbiol.* **2010**, *58* (2), 169–181.
- (21) Ison, C. A.; Easmon, C. S.; Dawson, S. G.; Southerton, G.; Harris, J. W. Non-Volatile Fatty Acids in the Diagnosis of Non-Specific Vaginitis. *J. Clin. Pathol.* **1983**, *36* (12), 1367–1370.
- (22) Piot, P.; Van Dyck, E.; Godts, P.; Vanderheyden, J. The Vaginal Microbial Flora in Non-Specific Vaginitis. *Eur. J. Clin. Microbiol.* **1982**, *1* (5), 301–306.
- (23) Al-Mushrif, S.; Eley, A.; Jones, B. M. Inhibition of Chemotaxis by Organic Acids from Anaerobes May Prevent a Purulent Response in Bacterial Vaginosis. *J. Med. Microbiol.* **2000**, *49* (11), 1023–1030.
- (24) Kaneuchi, C.; Seki, M.; Komagata, K. Production of Succinic Acid from Citric Acid and Related Acids by *Lactobacillus* Strains. *Appl. Environ. Microbiol.* **1988**, *54* (12), 3053–3056.
- (25) Bisanz, J. E.; Enos, M. K.; PrayGod, G.; Seney, S.; Macklaim, J. M.; Chilton, S.; Willner, D.; Knight, R.; Fusch, C.; Fusch, G.; Gloor, G. B.; Burton, J. P.; Reid, G.

- Microbiota at Multiple Body Sites during Pregnancy in a Rural Tanzanian Population and Effects of Moringa-Supplemented Probiotic Yogurt. *Appl. Environ. Microbiol.* **2015**, *81* (15), 4965–4975.
- (26) Friedman, J.; Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* **2012**, *8* (9).
- (27) Fernandes, A. D.; Macklaim, J. M.; Linn, T. G.; Reid, G.; Gloor, G. B. ANOVA-like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* **2013**, *8* (7), e67019.
- (28) Fernandes, A. D.; Reid, J. N.; Macklaim, J. M.; McMurrough, T. A.; Edgell, D. R.; Gloor, G. B. Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-Seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis. *Microbiome* **2014**, *2*, 15.
- (29) Söhling, B.; Gottschalk, G. Molecular Analysis of the Anaerobic Succinate Degradation Pathway in *Clostridium kluyveri*. *J. Bacteriol.* **1996**, *178* (3), 871–880.
- (30) Scherf, U.; Söhling, B.; Gottschalk, G.; Linder, D.; Buckel, W. Succinate-Ethanol Fermentation in *Clostridium kluyveri*: Purification and Characterisation of 4-Hydroxybutyryl-CoA Dehydratase/vinylacetyl-CoA Delta 3-Delta 2-Isomerase. *Arch. Microbiol.* **1994**, *161* (3), 239–245.
- (31) Macklaim, J. M.; Fernandes, A. D.; Di Bella, J. M.; Hammond, J.-A.; Reid, G.; Gloor, G. B. Comparative Meta-RNA-Seq of the Vaginal Microbiota and Differential Expression by *Lactobacillus iners* in Health and Dysbiosis. *Microbiome* **2013**, *1* (1), 12.
- (32) Srinivasan, S.; Morgan, M. T.; Fiedler, T. L.; Djukovic, D.; Hoffman, N. G.; Raftery, D.; Marrazzo, J. M.; Fredricks, D. N. Metabolic Signatures of Bacterial Vaginosis. *MBio* **2015**, *6* (2).
- (33) McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J.; Gloor, G. B.; Reid, G. A multi-platform metabolomics approach identifies novel

biomarkers associated with bacterial diversity in the human vagina
<http://arxiv.org/abs/1504.02816> (accessed Jun 8, 2016).

- (34) Liebich, H. M.; Först, C. Hydroxycarboxylic and Oxocarboxylic Acids in Urine: Products from Branched-Chain Amino Acid Degradation and from Ketogenesis. *J. Chromatogr.* **1984**, *309* (2), 225–242.
- (35) Kawai, S.; Suzuki, H.; Yamamoto, K.; Inui, M.; Yukawa, H.; Kumagai, H. Purification and Characterization of a Malic Enzyme from the Ruminal Bacterium *Streptococcus bovis* ATCC 15352 and Cloning and Sequencing of Its Gene. *Appl. Environ. Microbiol.* **1996**, *62* (8), 2692–2700.
- (36) Pine, L.; Malcolm, G. B.; Brooks, J. B.; Daneshvar, M. I. Physiological Studies on the Growth and Utilization of Sugars by *Listeria* Species. *Can. J. Microbiol.* **1989**, *35* (2), 245–254.
- (37) Novák, L.; Loubiere, P. The Metabolic Network of *Lactococcus lactis*: Distribution of ¹⁴C- Labeled Substrates between Catabolic and Anabolic Pathways. *J. Bacteriol.* **2000**, *182* (4), 1136–1143.
- (38) Yeoman, C. J.; Thomas, S. M.; Miller, M. E. B.; Ulanov, A. V; Torralba, M.; Lucas, S.; Gillis, M.; Cregger, M.; Gomez, A.; Ho, M.; Leigh, S. R.; Stumpf, R.; Creedon, D. J.; Smith, M. A.; Weisbaum, J. S.; Nelson, K. E.; Wilson, B. A.; White, B. A. A Multi-Omic Systems-Based Approach Reveals Metabolic Markers of Bacterial Vaginosis and Insight into the Disease. *PLoS One* **2013**, *8* (2), e56111.
- (39) Fettweis, J. M.; Paul Brooks, J.; Serrano, M. G.; Sheth, N. U.; Girerd, P. H.; Edwards, D. J.; Strauss, J. F.; Jefferson, K. K.; Buck, G. A. Differences in Vaginal Microbiome in African American Women versus Women of European Ancestry. *Microbiol. (United Kingdom)* **2014**, *160*, 2272–2282.
- (40) Kohlmeier, K. A.; Vardar, B.; Christensen, M. H. γ -Hydroxybutyric Acid Induces Actions via the GABA_B Receptor in Arousal and Motor Control-Related Nuclei: Implications for Therapeutic Actions in Behavioral State Disorders. *Neuroscience*

2013, 248, 261–277.

- (41) Absalom, N.; Eghorn, L. F.; Villumsen, I. S.; Karim, N.; Bay, T.; Olsen, J. V.; Knudsen, G. M.; Bräuner-Osborne, H.; Frølund, B.; Clausen, R. P.; Chebib, M.; Wellendorph, P. $\alpha 4\beta\delta$ GABA_A Receptors Are High-Affinity Targets for γ -Hydroxybutyric Acid (GHB). *Proc. Natl. Acad. Sci. U. S. A.* **2012**, 109 (33), 13404–13409.
- (42) Connelly, W. M.; Errington, A. C.; Crunelli, V. γ -Hydroxybutyric Acid (GHB) Is Not an Agonist of Extrasynaptic GABA_A Receptors. *PLoS One* **2013**, 8 (11), e79062.
- (43) Aagaard, K.; Riehle, K.; Ma, J.; Segata, N.; Mistretta, T. A.; Coarfa, C.; Raza, S.; Rosenbaum, S.; van den Veyver, I.; Milosavljevic, A.; Gevers, D.; Huttenhower, C.; Petrosino, J.; Versalovic, J. A Metagenomic Approach to Characterization of the Vaginal Microbiome Signature in Pregnancy. *PLoS One* **2012**, 7 (6).
- (44) Romero, R.; Hassan, S. S.; Gajer, P.; Tarca, A. L.; Fadrosh, D. W.; Nikita, L.; Galuppi, M.; Lamont, R. F.; Chaemsaihong, P.; Miranda, J.; Chaiworapongsa, T.; Ravel, J. The Composition and Stability of the Vaginal Microbiota of Normal Pregnant Women Is Different from that of Non-Pregnant Women. *Microbiome* **2014**, 2 (1), 4.
- (45) MacIntyre, D. A.; Chandiramani, M.; Lee, Y. S.; Kindinger, L.; Smith, A.; Angelopoulos, N.; Lehne, B.; Arulkumaran, S.; Brown, R.; Teoh, T. G.; Holmes, E.; Nicholson, J. K.; Marchesi, J. R.; Bennett, P. R. The Vaginal Microbiome during Pregnancy and the Postpartum Period in a European Population. *Sci. Rep.* **2015**, 5, 8988.
- (46) Laghi, L.; Picone, G.; Cruciani, F.; Brigidi, P.; Calanni, F.; Donders, G.; Capozzi, F.; Vitali, B. Rifaximin Modulates the Vaginal Microbiome and Metabolome in Women Affected by Bacterial Vaginosis. *Antimicrob. Agents Chemother.* **2014**, 58 (6), 3411–3420.

- (47) Gajer, P.; Brotman, R. M.; Bai, G.; Sakamoto, J.; Schütte, U. M. E.; Zhong, X.; Koenig, S. S. K.; Fu, L.; Ma, Z. S.; Zhou, X.; Abdo, Z.; Forney, L. J.; Ravel, J. Temporal Dynamics of the Human Vaginal Microbiota. *Sci. Transl. Med.* **2012**, *4* (132), 132ra52.
- (48) Gloor, G. B.; Hummelen, R.; Macklaim, J. M.; Dickson, R. J.; Fernandes, A. D.; MacPhee, R.; Reid, G. Microbiome Profiling by Illumina Sequencing of Combinatorial Sequence-Tagged PCR Products. *PLoS One* **2010**, *5* (10), e15406.
- (49) Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B. Methodol.* **1982**, *44* (2), 139–177.
- (50) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (8), 770–781.
- (51) Styczynski, M. P.; Moxley, J. F.; Tong, L. V.; Walther, J. L.; Jensen, K. L.; Stephanopoulos, G. N. Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery. *Anal. Chem.* **2007**, *79* (3), 966–973.
- (52) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995, pp 289–300.
- (53) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536.
- (54) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* **2010**, *11*, 395.
- (55) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral

Database. *Ther Drug Monit* **2005**, 27 (6), 747–751.

- (56) 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.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, 35 (Database issue), D521–D526.
- (57) Lopez-Raton, M.; Rodríguez-Álvarez, M.X.; Cadarso-Suárez, C.; Gude, F. OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *J. Stat. Softw.* **2014**, 61 (8).
- (58) Youden, W. J. Index for Rating Diagnostic Tests. *Cancer* **1950**, 3 (1), 32–35.

Chapter 3

3 Integration of the vaginal metabolome and bacterial meta-transcriptome of Canadian women

3.1 Introduction

A number of approaches are employed to characterize the function of microbial communities. This includes classical microbiology methods where organisms are studied in isolation, as well as *in vivo* investigations of communities using 16S sequencing, metagenomics, meta-transcriptomics, proteomics and/or metabolomics. Each of these techniques provides unique and complementary information, however they are rarely examined together. We have demonstrated previously in two independent studies that the bacterial meta-transcriptome and metabolome of women with BV are distinct from healthy women^{1,2}. How these two datasets relate to one another in the context of the vaginal microbiota, or any microbial community for that matter, is not well understood.

The purpose of this chapter was two-fold. Firstly, we wished to replicate the Rwandan study in Canadian women to determine if GHB and 2HV are consistent markers of BV across geographical locations and ethnicities. Secondly, we attempted to, for the first time relate the bacterial meta-transcriptome and metabolome, and demonstrate how these very distinct types of data can be integrated to gain improved insight into microbial communities.

3.2 Materials and methods

3.2.1 Study design and sample collection

This study was approved by the Health Sciences Research Ethics Board (REB) at The University of Western Ontario, REB number 18203E. Vaginal samples were collected from two clinics in London, Ontario, Canada: Victoria Family Medicine, and the Middlesex-London Health Unit. Participants were excluded from the study if they had reached menopause, had a urogenital infection other than BV in the past 6 months, were pregnant, had a history of gonorrhoea, chlamydia, estrogen-dependent neoplasia, abnormal

renal function or pyelonephritis, were taking prednisone, immunosuppressives or antimicrobial medication, or had undiagnosed abnormal vaginal bleeding. Participants were asked to refrain from oral or vaginal intercourse and consuming probiotic supplements or foods for 48 hours prior to the clinical visit. No participants were menstruating at time of the clinical visit. After reviewing details of the study, participants gave their signed informed consent before the start of the study. Four Cytobrush swabs (Cooper Surgical) were collected from each participant by rolling against the mid-vaginal wall and used as follows: 1) immediately smeared onto a glass slide and left to dry for Gram staining and Nugent scoring, 2) immediately suspended in RNAprotect (Qiagen) containing 100 ug/ml rifampicin, 3 and 4) swabs were placed in an empty 1.5ml Eppendorf tube and frozen at -20 for later DNA and metabolite isolation.

3.2.2 DNA isolation, PCR amplification, and sequencing

Vaginal swabs for DNA microbiome analysis were extracted using the QIAamp DNA stool mini kit (Qiagen) with the following modifications: swabs were vortexed in 1 mL buffer ASL before removal of the swab and addition of 200 mg of 0.1 mm zirconia/silica beads (Biospec Products). Samples were shaken vigorously (Mini-BeadBeater; Biospec Products) for 2 x 30 seconds at full speed with cooling at room temperature between. After heating to 95 °C for 5 minutes, 1.2 ml of supernatant was aliquoted into a 2ml tube and one-half an inhibitEx tablet (Qiagen) was added to each sample. All other steps were performed as per the manufacturer's instructions. DNA amplification for sequencing was carried out using the forward primer 5' **C CATCTCATCCCTGCGTGTCTCCGACTCAGn** nnnnC[W]ACGCGA[R]GAACCTTACC 3' and the reverse primer 5' **CTCTCTATGGGCAGTCGGTGATA C[R]ACACGAGCTGACGAC** 3'. The forward primer contained an Ion Torrent adapter (bolded), followed by a 5-mer unique sequence barcode (specific to each sample), and then sequence complementary to the V6 rRNA gene region. The reverse primer contains the Ion Torrent adapter (bolded), followed by sequence complementary to the V6 rRNA gene region. Amplification was carried out in 42 µL with each primer present at 0.8 pMol/mL, 20 µL GoTaq hot start colorless master mix (Promega) and 2 µL extracted DNA. The PCR protocol was as follows: initial activation of Taq

mastermix at 95°C for 2 minutes, followed by the addition of template and primers, then 25 cycles of 1 minute 95°C, 1 minute 55°C and 1 minute 72°C.

PCR products were quantified with a Qubit 2.0 Fluorometer and the high sensitivity dsDNA specific fluorescent probes (Life Technologies). Samples were mixed at equimolar concentrations and purified with the QIAquick PCR Purification kit (QIAGEN). Library preparation and sequencing was performed at the London Regional Genomics Institute (LRCG), in London, Canada. Samples were sequenced across 3 different runs (316 chips), and computationally pooled for analysis.

3.2.3 16S sequence analysis

De-multiplexed and pooled reads (using in-house BASH and Perl scripts) were clustered into operational taxonomic units (OTUs) at a 97% ID using UClust (uclust3.0.617_i86darwin32). OTUs were kept if they represented >1% of the reads in any one sample. Taxonomic assignment was completed by comparing the representative sequence for each OTU to our curated V6 vaginal database via BLAST, and confirmed by the ribosomal database project (RDP) seqmatch tool³.

3.2.4 Meta-transcriptome analysis

Sample preparation and data analyses were performed according to the methods of Macklaim et al 2013¹. Briefly, RNA was extracted by TRIzol (Invitrogen), and rRNA depleted using MICROBEExpress (Ambion). Extracted RNA was then DNase treated and sent to the Toronto Center for Applied Genomics (TCAG) for library preparation and sequencing by Illumina Hi-seq.

Sequences were clustered into representative sequences (refseqs) by sequence identity (95% nucleotide identity over 90% sequence length) and mapped to a reference database which included all partial or complete genomes from vaginal isolates in the NCBI database. Amino acid translations of predicted coding sequences were then used to assign SEED subsystems or Kegg Orthology (KO) functions. Differential expression was assessed by ALDEx2 Wilcoxon tests. A Benjamini-Hochberg corrected (FDR) p value < 0.05 was

considered statistically significant. Compositional biplots were constructed as reported by Gloor et al 2016^{4,5}.

3.2.5 Sample preparation GC-MS.

Samples were collected from the mid-vaginal wall using the Cytobrush vaginal brush and sterile forceps, and stored at -80 C until analysis. Vaginal brushes were pre-cut into 1.5 mL tubes and weighed prior to and after sample collection to determine the mass of vaginal fluid collected. After thawing, brushes were eluted in methanol-water (1:1) to a final concentration of 0.05 g/mL. This corresponded to a volume of 200-1500 μ L, depending on the mass of vaginal fluid collected. Four control swabs were included which consisted of blank swabs eluted in 200, 400, 800, or 1500 μ L of methanol-water. Samples and controls were vortexed for 10 sec to extract metabolites, centrifuged for 5 min at 10 000 rpm, vortexed again for 10 sec after which time the brushes were removed from tubes. Samples were centrifuged a final time to pellet cells and 150 μ L of supernatant transferred to GC-MS vials. Remaining supernatant was transferred to a new 1.5 mL tube, frozen at -80 C and shipped to the University of Reading for NMR analysis. Next, 25 μ L of 0.2 mg/ml ribitol standard was added to each GC-MS vial. Samples were then dried to completeness using a SpeedVac. After drying 100 μ L of 2% methoxyamine•HCl in pyridine (MOX) was added to each sample for derivatization and samples were incubated at 50⁰C for 90 min. 100 μ L N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) was then added to each vial and incubated at 50⁰C for 30 min. After derivatization, an equal aliquot of each sample was combined to make the quality control (QC). Samples were then transferred to micro inserts before analysis by GC-MS (Agilent 7890A GC, 5975 inert MSD with triple axis detector). 1 μ L of sample was injected using pulsed splitless mode into a 30 m DB5-MS column with 10 m duraguard, diameter 0.35mm, thickness 0.25 μ m (JNW Scientific). Helium was used as the carrier gas at a constant flow rate of 1 mL/min. Oven temperature was held at 70 °C for 5 min then increased at a rate of 5 °C/min to 300 °C and held for 10 min. Solvent delay was set to 13 min to avoid solvent and a large lactate peak, and total run time was 61 min. Masses between 25 m/z and 600 m/z were selected by the detector. All samples were run in random order and the QC was run multiple times throughout the run to ensure machine consistency.

3.2.6 Data analysis GC-MS.

Chromatogram files were de-convoluted and converted to ELU format using the AMDIS⁶ mass spectrometry software with the sensitivity set to medium. Chromatograms were then aligned and integrated using Spectconnect⁷ software with the support threshold set to low. All metabolites found in the blank swab, or believed to have originated from derivatization reagents were removed from analysis at this time. After removal of swab metabolites, the IS matrix from Spectconnect was transformed using the additive log ratio transformation (alr)⁸ and ribitol as a normalizing agent ($\log_2(x) / \log_2(\text{ribitol})$). Zeros were replaced with two thirds the minimum detected value on a per metabolite basis prior to transformation. All further metabolite analysis was performed using these alr transformed values.

A total of 90 metabolites were detected by GC-MS. Upon manual inspection it was determined that 50 of these metabolites were either redundant, background noise or present in controls, and therefore they were removed from analysis. For redundant peaks, the sum of each peak was combined resulting in a single value for each metabolite. Metabolites were initially identified by comparison to the NIST 11 standard reference database (<http://www.nist.gov/srd/nist1a.cfm>). Identities of metabolites of interest were then confirmed by authentic standards if available.

Independent Wilcoxon tests with a Benjamini-Hochberg false discovery rate (FDR)⁹ correction to account for multiple testing were used to determine metabolites that differed significantly between healthy and BV ($p < 0.10$). Groups were defined as healthy or BV using a percentage *Lactobacillus* cutoff of 75 %.

3.2.7 Sample preparation NMR.

Spectra were processed according to the methods of Swann et al 2011¹⁰ with the following modifications. ¹H NMR spectra were manually corrected for phase and baseline distortions and then referenced to the TSP resonance δ 0.0. Spectra were digitized using an in-house MATLAB(version R2009b, The Mathworks, Inc.; Natwick, MA) script. To prevent baseline effects that arise from imperfect water saturation the region containing the water resonance was excised. An in-house peak alignment algorithm was then performed on each

spectrum in MATLAB to adjust for shifts in peak position due to small pH differences between samples and then each spectrum was normalized using a sum normalization approach. Principal components analysis (PCA) using pareto scaling was applied in SIMCA (Umetrics, Umea). Orthogonal projection to latent structure discriminant analysis (OPLS-DA) models were constructed using unit variance scaling to aid the interpretation of the model and distinguish the metabolites that differed between the groups. Here, ¹H NMR spectroscopic data were used as the descriptor matrix and class information (N or BV) as the response variable. The contribution of each variable (metabolite) to sample classification was visualized by back-scaling transformation, generating a correlation coefficient plot. These coefficient plots are colored according to the significance of correlation to “class” (e.g. N or BV), with red indicating high significance and blue indicating low significance. The direction and degree of the signals relate to covariation of the metabolites with the classes in the model. For all models, one orthogonal component was used to remove systematic variation unrelated to class. Predictive performance was assessed using the Q²^Y parameter.

3.2.8 Correlations between transcriptome and metabolome

The expected value of Kendall’s Tau was used when reporting the correlation between transcript abundance and metabolite abundance since it is not expected that the metabolome and meta-transcriptome tables share even passingly similar units or scaling¹¹. Spearman’s correlations were used when reporting correlations between between 16S proportions and transcripts (refseqs or KOs). All analyses were conducted in R using clr transformed values for compositional data (16S and transcriptome) with Benjamini-Hochberg corrections to account for multiple hypothesis testing.

3.3 Results

3.3.1 The vaginal microbiota of Canadian women

Forty-one women were recruited to the study initially, for which 16S and metabolome data was obtained (data not shown). From these 41 original samples, 24 were selected for meta-transcriptome analysis. Selection aimed to obtain an equal representation of biota types and

also included several unique profiles, for example subject 013B who was dominated nearly completely by *Atopobium vaginae* (**Error! Reference source not found.**A). Four samples from previously published work, for which transcriptome but not 16S data was obtained¹, were also included (Samples 30S, 4S, 31S and 27S, Figure 3-1B).

Consistent with previous studies, the vaginal microbiota of healthy Canadian woman was dominated by either *L. iners* or *L. crispatus* (**Error! Reference source not found.**A). There was also one women dominated by *L. jensenii*. Women with BV were dominated by mixture of anaerobes including members of the genera *Gardnerella*, *Prevotella*, *Megasphaera*, *Atopobium*,

Dialister, *Sneathia*, as well as a recently identified taxa putatively designated BV associated bacteria 1 (BVAB1)^{12,13} (**Error! Reference source not found.**A). Although all of these taxa were identified in Rwandan dataset, women with BV from Rwanda tended to have a larger proportion of *Sneathia* compared to Canadian women, while women with BV from Canada tended to be dominated more often by *Megasphaera*. BVAB1 was also more prevalent in Canadian women compared to Rwandan (data not shown).

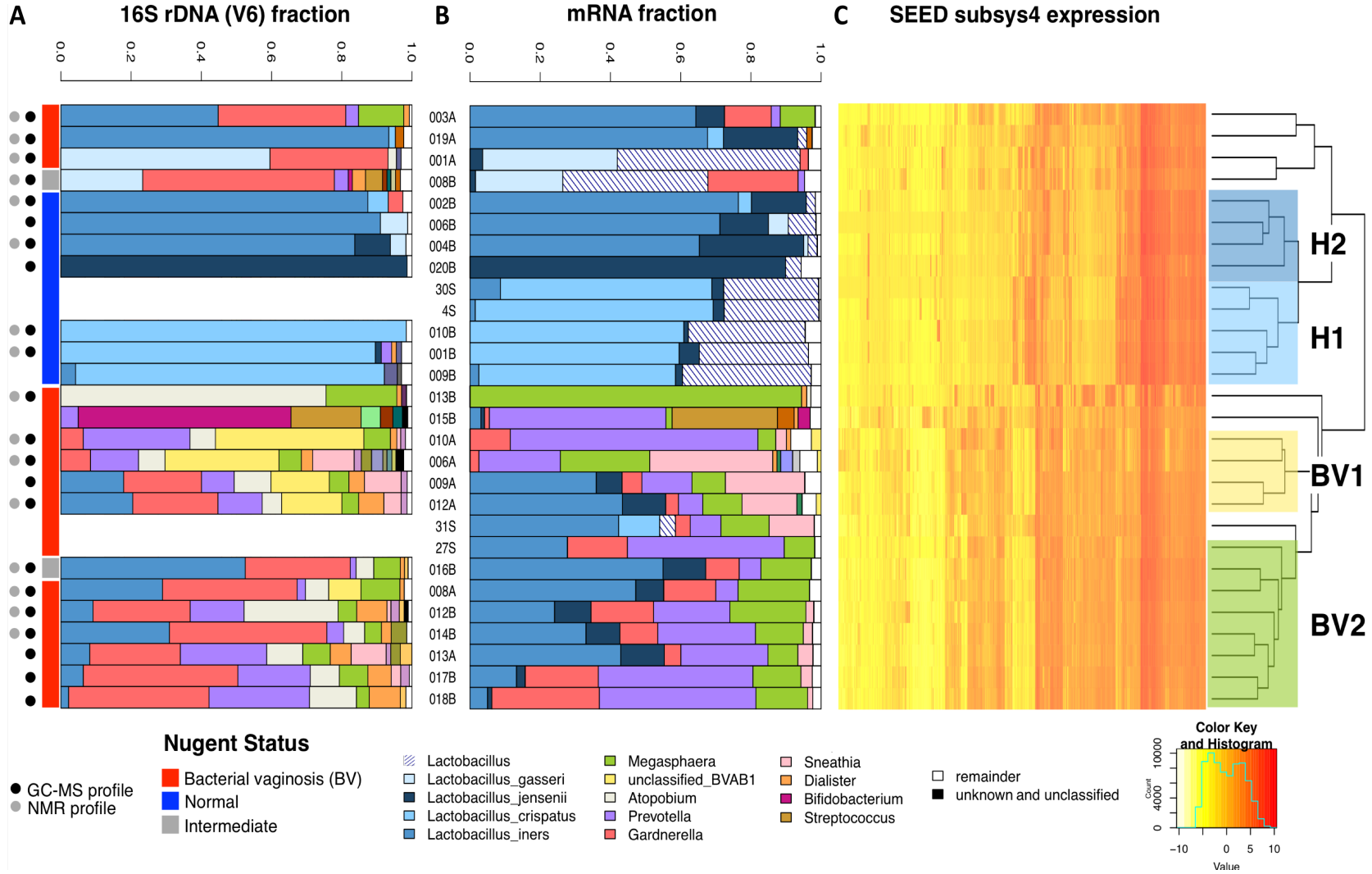


Figure 3-1. Taxonomic barplots and clustered heatmap of meta-transcriptome expression.

Taxonomic distribution by genus of each sample as a fraction of total reads are shown for V6 16S rDNA (panel **A**) and for the mapped mRNA reads (panel **B**). Reads belonging to the *Lactobacillus* genus are plotted by species-level assignment where possible. Nugent status and presence of corresponding metabolite samples (used for later analysis) are indicated. 16S sequence data are not available for the samples from the previous dataset, indicated with an “S” designation in the sample ID¹. Samples were clustered based on their relative SEED subsystem 4 (subsys4) expression, where a subsystem represents a set of functional roles that together implement a specific biological process or structural complex (panel **C**, heatmap and dendrogram). Complete linkage clustering was then used to draw the dendrogram based on Euclidean distances between the zero-replaced, clr-transformed subsys4 counts (the relative expression). The four major groups are identified on the plot as health-associated group 1 (H1) and group 2 (H2), and BV-associated group 1 (BV1) and group 2 (BV2). Data and figure generated by Jean Macklaim.

3.3.2 The vaginal metabolome of Canadian women

To determine if the biomarkers of health and BV identified in the Rwandan dataset could be replicated in Canadian women, we performed an untargeted metabolomic analysis of vaginal fluid using GC-MS (**Error! Reference source not found.A**) and NMR (Figure 3-3). LC-MS was not available at the time of study commencement. The majority of metabolites discriminating BV from health were consistent with the Rwandan study, including GHB, 2HV, 2-hydroxyisocaproate, 2-hydroxyglutarate, cadaverine, trimethylamine, and tyramine. Succinate was weakly associated with BV by NMR (Figure 3-3), but was not significant by GC-MS (**Error! Reference source not found.A**). Metabolites associated with health included amino acids and lactate, as was found in Rwandan women. Due to differences in sampling methods between studies, samples were less concentrated in the Canadian study, and therefore some of the low level biomarkers identified in Rwandan women, such as tyrosine, were not detected in Canadian samples.

Although the majority of the metabolites differing significantly between women with BV and health were consistent between studies, there were some inconsistencies. Mainly, a number of sugars were significantly increased in women with BV from Canada, including glucose, maltose, and fructose, and were confirmed by both GC-MS (Figure 4-1A) and

NMR (Figure 3-3). This was not the case for Rwandan women.

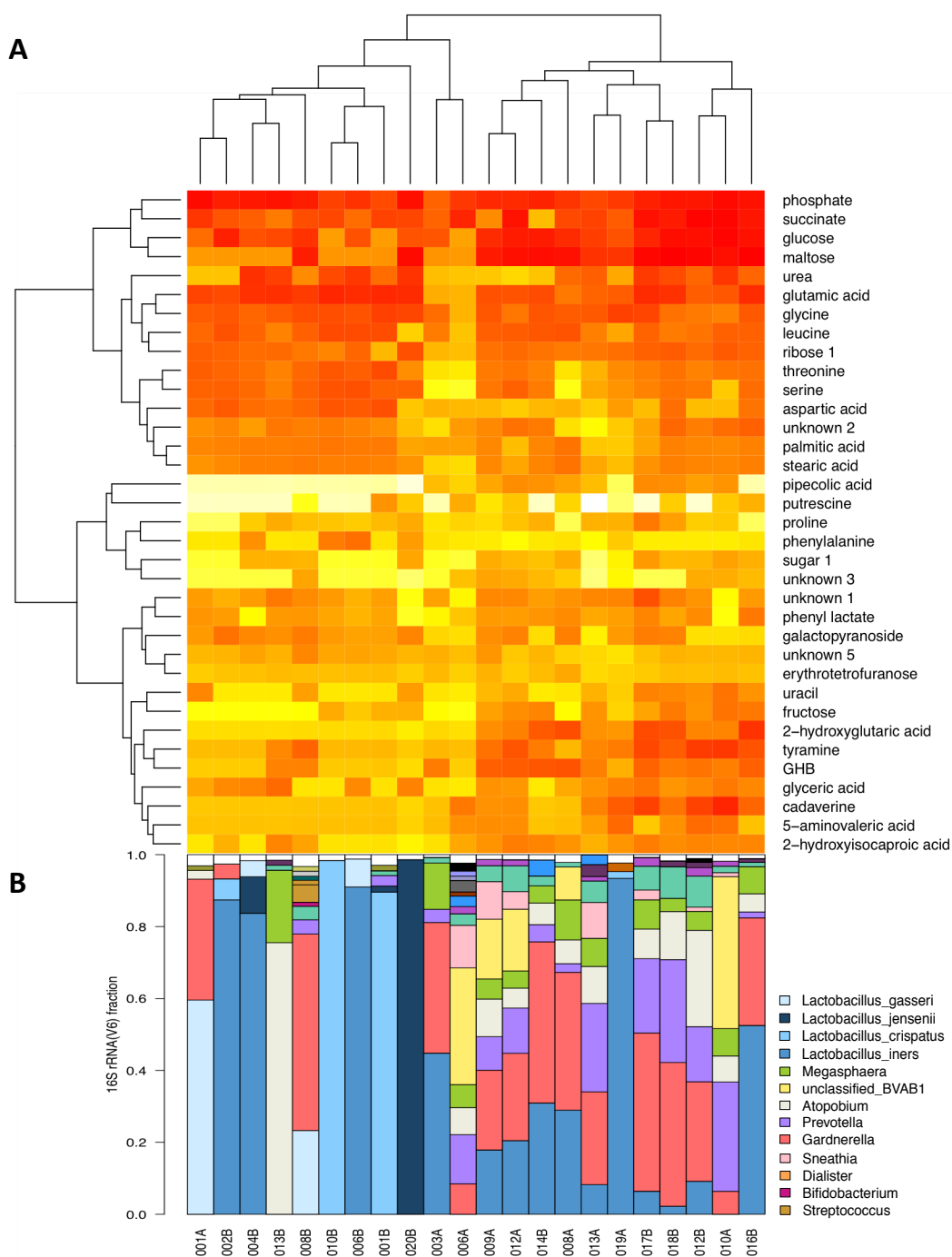


Figure 3-2. The vaginal metabolome of Canadian women analyzed by GC-MS.

A. Heatmap depicting abundance of metabolites in samples. Dendrogram above shows distances (dissimilarity) between samples based on complete linkage hierarchical clustering of metabolites. **B.** Barplots of relative abundance of bacterial species, genera, or families in each corresponding sample as determined by 16S rRNA gene sequencing.

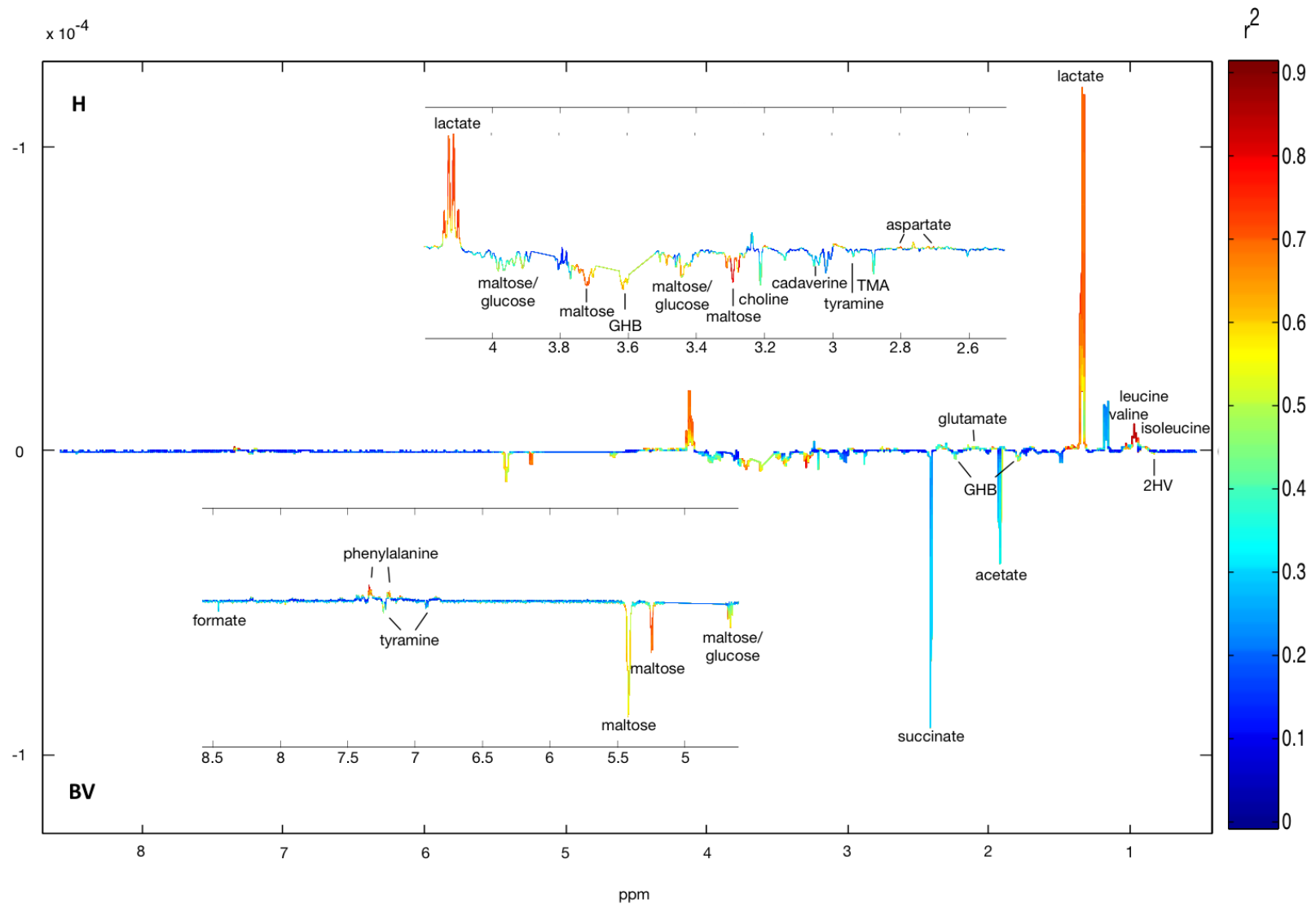


Figure 3-3. The vaginal metabolome of Canadian women analyzed by NMR.

Correlation coefficient plot was extracted from orthogonal projection to latent structures-discriminatory analysis (OPLS-DA). Metabolites pointing upwards or downwards are elevated in BV or H respectively. Peak color reflects the correlation between each peak and the condition, with red indicating stronger correlations. The direction and degree of the signals relate to covariation of the metabolites with the classes (BV or H) in the model. ($Q^2\hat{Y}=0.82$).

3.3.3 The bacterial meta-transcriptome identifies subtypes of BV and health

To visualize the relationship between samples based on the meta-transcriptome, samples were clustered based on SEED subsystem 4 (subsys4) expression profiles, where a subsystem represents a set of functional roles that together implement a specific biological process or structural complex¹⁴. Analysis of expression profiles (performed by Jean Macklaim) identified four distinct microbiota subtypes, two associated with health and two associated with BV (Figure 3-1C). The first H1 group was composed largely of *L. crispatus* by both the 16S rRNA gene sequencing and the mRNA fraction mapping. The second group, H2, was composed largely of a mixed set of *Lactobacillus* species, often dominated by *L. iners* in total abundance, but with a substantial gene expression contribution from *L. jensenii*, *L. gasseri* or unknown *Lactobacillus* sp. The BV1 group contained a substantial amount of the unclassified BVAB1 organism by 16S rRNA gene sequencing and had a large gene expression contribution from *Sneathia* sp, and also contained the largest contribution of expression from de-novo assembled contigs. These contigs were assigned to the BVAB1 group color in the figure. The final group, BV2, had only a small amount of BVAB1, and a generally larger amount of *Atopobium* sp. by 16S rRNA gene profiling, and a very small or absent contribution to gene expression by *Sneathia* and BVAB1. Samples that did not fit into these groups mostly exhibited atypical profiles and were therefore removed from further meta-transcriptome and metabolome analysis. One of these outliers, sample 019A, was particularly interesting as this subject was classified as clinically BV by Nugent scoring despite being completely dominated by *L. iners* (Figure 3-1). The metabolome of this sample also clustered with BV samples by both NMR and GC-MS (data not shown).

3.3.4 Integrating the meta-transcriptome and metabolome

To determine if microbiota subtypes identified through meta-transcriptome analysis were also metabolically distinct, the relationship between samples based on principle component analysis (PCA) of the metabolome were examined (Figure 3-4A). A clear separation was observed between BV and H by both methods, confirming the metabolome of women with BV is distinct from healthy women. Unlike the SEED subsys4 meta-transcriptome data, the microbiota subtypes did not separate clearly by metabolome, nor were there any metabolites that differed significantly between BV subtypes as measured by GC-MS (Wilcox test, Benjamini-Hochberg $p > 0.1$).

To visualize the relationship between the meta-transcriptome and metabolome, correlations between metabolite abundance and the position of samples on biplots built from SEED subsys4 meta-transcriptomic data were calculated. These correlations were then scaled by the axis of the biplots and metabolites plotted according to their scaled positions (Figure 3-4B). As expected, amino acids and lactate were correlated with the positions of healthy samples, while organic acids, amines and monosaccharides were correlated with positions of BV samples.

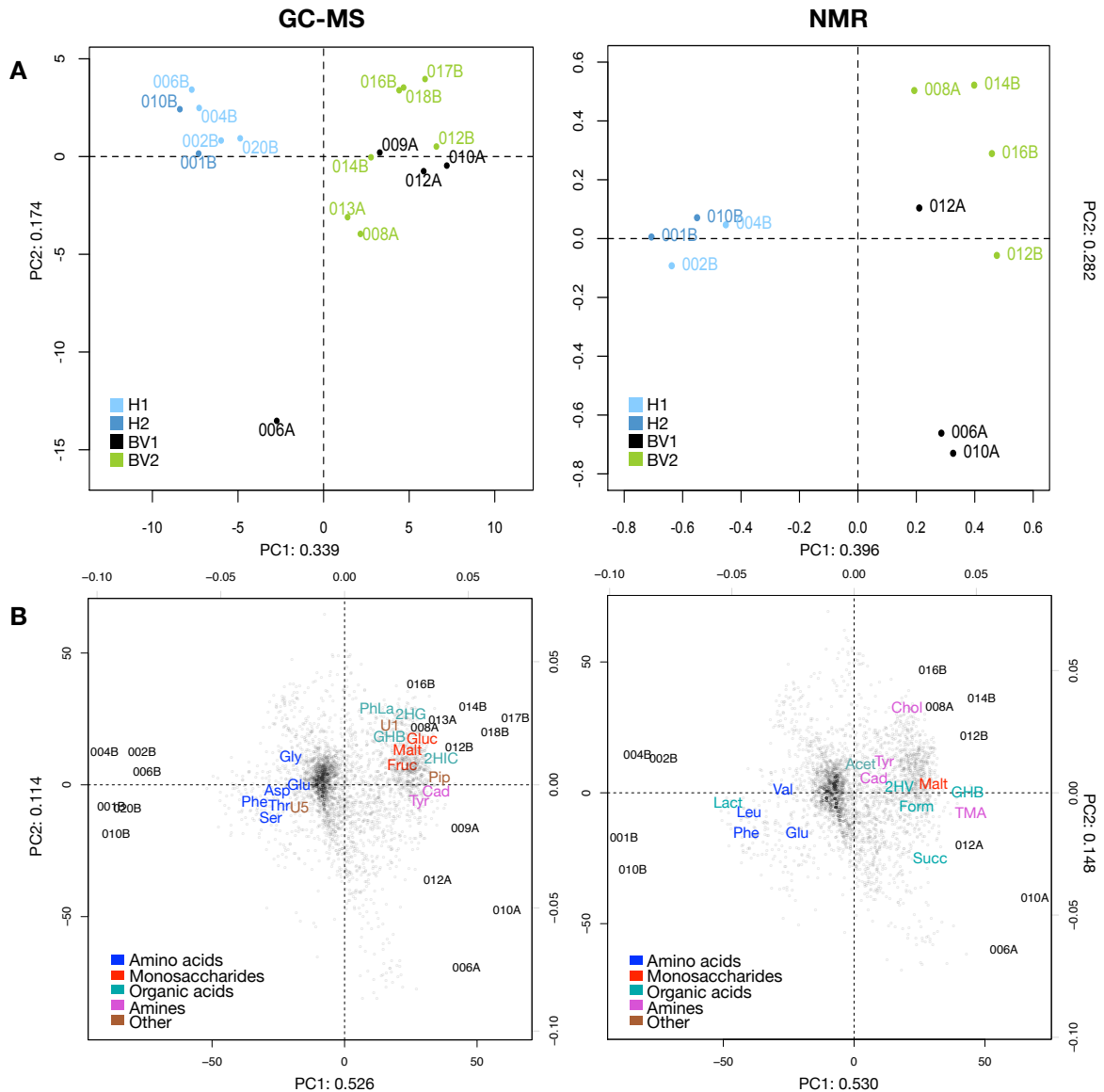


Figure 3-4. Correlation between metabolite abundance and microbial function.

A. Principal component analysis (PCA) scoreplots of vaginal metabolome as detected by GC-MS (left) or NMR (right). Each point represents a single sample from a single women colored according to their microbiota subtype. Positions of points display dissimilarities in the metabolome, with points furthest from one another being most dissimilar. **B.** Correlations between metabolite abundance and positions of samples on compositional biplots built from meta-transcriptomic data. Each point represents a single SEED subsystem4 function, with the positions of samples based on the relative abundance of these functions within samples. Correlations between metabolites and sample positions were scaled by the first and second component of the biplots, and metabolites plotted according to their scaled position. Only metabolites that differed significantly between BV and health are shown. GHB: gamma-hydroxybutyrate, 2HG: 2-hydroxyglutarate, 2HIC: 2-hydroxyisocaproate, PhLa: phenyllactate, Tyr: tyramine,

Cad: cadaverine, Pip: pipercolate, Gluc: glucose, Malt: maltose, Fruc: fructose, U: unknown, TMA: trimethylamine, Lact: lactate, Acet: acetate, Form: formate, 2HV: 2-hydroxyisovalerate, Chol: choline, Succ: succinate.

We next attempted to correlate individual metabolites and transcripts (Kegg Orthology (KO))¹⁵ to identify transcripts and taxa which could be involved in metabolite production. Cadaverine was chosen to investigate further as the bacterial gene responsible for production (lysine decarboxylase) has been characterized extensively^{16,17}, although not in vaginal organisms specifically. As expected, the relative abundance of lysine decarboxylase (K01582) differed significantly between BV and health (Wilcoxon test, FDR corrected $p = 7.2E-4$, Figure 3-5 middle panel). The majority of transcripts for this KO mapped to *Megasphaera* species, indicating this organism may be the main producer of cadaverine in the vagina (Figure 3-5 top panel). Despite the co-occurrence of lysine decarboxylase and cadaverine in BV, the correlation between lysine decarboxylase and cadaverine was not amongst the strongest (Spearman's $R=0.53$, FDR corrected $p=0.07$, Figure 3-5 bottom panel). In fact, there were 442 KOs with larger Spearman's rho values than that of lysine decarboxylase. The KO correlating most strongly with cadaverine was a sulfotransferase, an enzyme with no clear link to cadaverine metabolism.

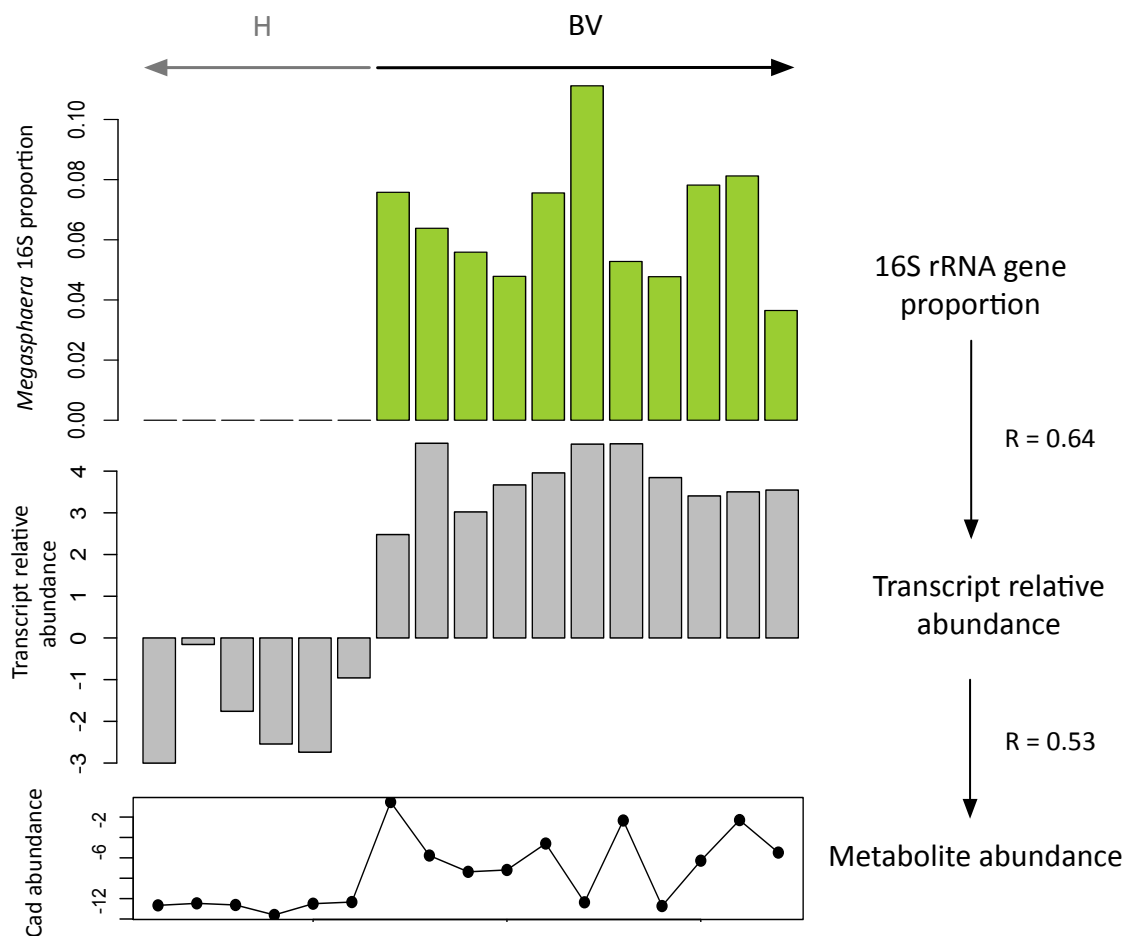


Figure 3-5. Relative abundance of bacterial genus (*Megasphaera*) and transcript (lysine decarboxylase) predicted to be responsible for cadaverine synthesis in women with BV compared to health.

The top, middle, and bottom panels display the proportion of *Megasphaera*, relative abundance of lysine decarboxylase, and absolute abundance of cadaverine within each sample respectively. “R” refers to the Spearman’s rho value between *Megasphaera* and lysine decarboxylase, as well as lysine decarboxylase and cadaverine. All correlations were performed using clr transformed sequencing data. Metabolite data was acquired by GC-MS.

GHB was a secondary metabolite of interest as it is one of the most sensitive and specific biomarkers for BV, and has been shown to be produced by *G. vaginalis*². The bacterial enzyme involved in GHB synthesis (GHB dehydrogenase¹⁸) is not in the Kegg database, and therefore the relationship between GHB and its putative transcript could not be investigated using the KO system. To circumvent this, we performed a protein BLAST

search of a validated GHB dehydrogenase from *Clostridium kluyveri*¹⁸, specifically targeting sequences in the meta-transcriptome belonging to *G. vaginalis*. This analysis identified two translated refseqs (reference sequences, defined as sequences with >90% nucleotide similarity) from *G. vaginalis* with ~46% amino acid similarity to the GHB dehydrogenase of *C. kluyveri*. Both were annotated as GHB dehydrogenases. The correlation between these refseqs was strong (Spearman's $R=0.88$, $p=3.22 \times 10^{-6}$), and the translated amino acid sequences differed by only a single Serine-Alanine repeat. Due to limitations in computing power stemming from the large number of variables in the refseq table ($n=49\,885$), an FDR correction could not be computed for the comparison of GHB dehydrogenase abundance in BV and health. However, the uncorrected p values for the two GHB dehydrogenase refseqs were low (Wilcoxon test, $p=0.001$ and 0.003), and effect sizes were large (1.77 and 1.90 respectively), indicating the between group variation was larger than that of within.

When the 16S proportion of *G. vaginalis*, relative abundance of *G. vaginalis* GHB dehydrogenase refseq1, and measured abundance of GHB were plotted across samples, it is clear they are highly correlated (Figure 3-6). Despite this fact, GHB was not most correlated with the GHB dehydrogenase refseqs (refseq 1 $R=0.69$, $p=0.002$, refseq 2 $R=0.73$, $p=0.0009$), and there were >500 refseqs with greater correlation values than that of the putative GHB dehydrogenases (data not shown).

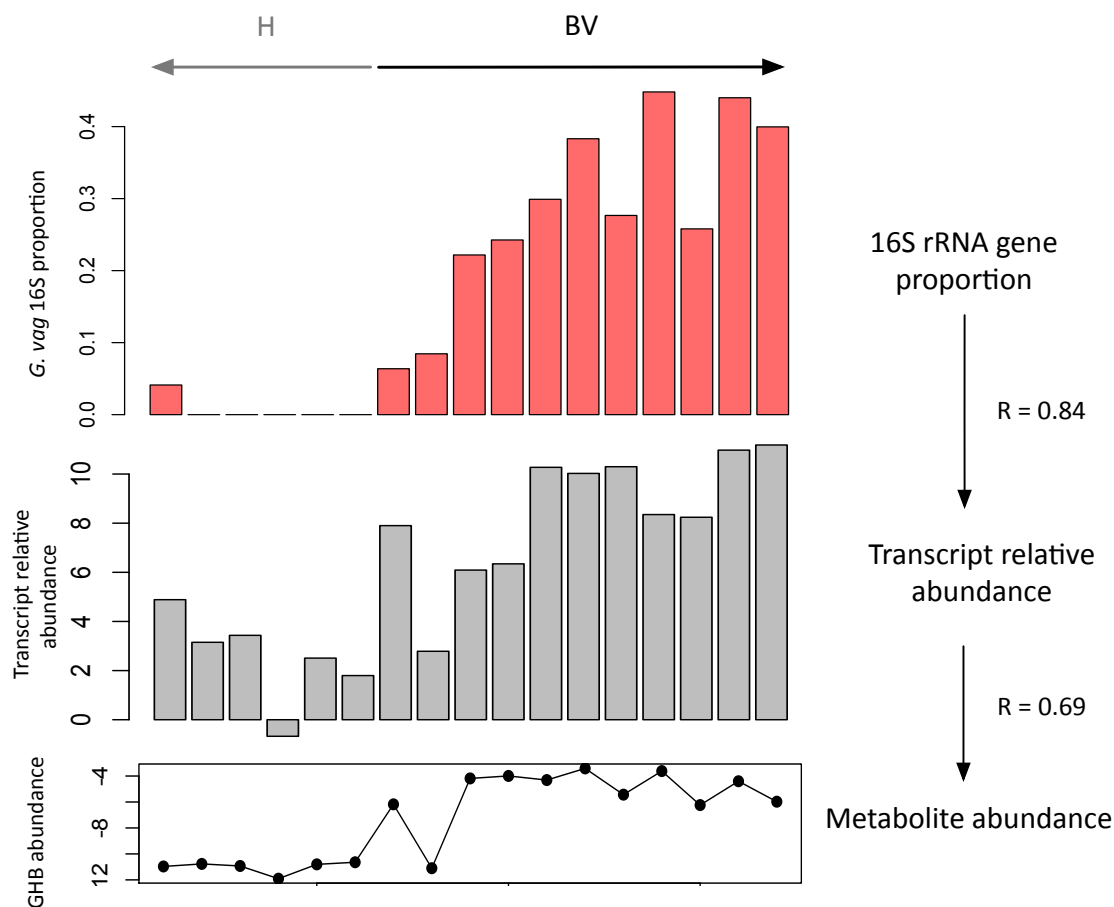


Figure 3-6. Relative abundance of bacterial species (*Gardnerella vaginalis*) and transcript (GHB dehydrogenase) predicted to be responsible for GHB synthesis in women with BV compared to health.

The top, middle, and bottom panels display the proportion of *G. vaginalis*, relative abundance of GHB dehydrogenase, and absolute abundance of GHB within each sample respectively. “R” refers to the Spearman’s rho value between *G. vaginalis* and GHB dehydrogenase, as well as GHB dehydrogenase and GHB. All correlations were performed using clr transformed 16S and transcript data. Metabolite data was acquired by GC-MS.

3.4 Discussion

This work confirms that GHB and 2HV are reproducible biomarkers of BV, irrespective of geographical location or ethnicity. It also represents the first study to relate the bacterial meta-transcriptome and metabolome. Through the examples of cadaverine and GHB, we have demonstrated how the metabolome and transcriptome can be leveraged to identify

putative genes and taxa involved in metabolite production within a community. In the case of GHB, production by *G. vaginalis* was first reported by us recently². Unfortunately, we were unable to grow *Megasphaera* isolates *in vitro* and therefore could not confirm cadaverine production by this genus. However, the near complete dominance of the lysine decarboxylase KO by *Megasphaera* provides good circumstantial evidence for production by this genus.

Despite the stark contrast in the transcriptome and metabolome of women with BV compared to health, we found that correlations between these datasets were not sufficient to identify clear gene-product relationships without prior knowledge of the enzymes involved. Multiple factors likely contribute to this. Firstly, in contrast to sequence-based methods, metabolites may be of bacterial or host origin or both. Many metabolites may therefore be generated fully or partially by mammalian metabolism, which is not captured by the bacterial meta-transcriptome. The strong correlation between BV associated bacteria, and therefore their transcripts and products, may also be a large contributor. This leads to spurious correlations, whereby transcripts and products involved in unrelated pathways are highly correlated. Whether this observation holds true for communities with less microbial co-occurrence, such as the gut, remains to be determined.

Another contributing factor is the non-specific nature of many enzymes. For example, although the KO for lysine decarboxylase is specific to lysine, the *Megasphaera* refseq assigned to this KO is annotated as a Orn/Lys/Arg decarboxylase (gi|335049912|ref|NZ_AFIJ01000038.1|c21078-19624). This enzyme could therefore perform decarboxylation of substrates other than lysine, or may not act on lysine at all. This is true of many enzymes, which may contribute to the weaker than expected correlation between transcripts and products.

Finally, the most overlooked factor may be the fact that metabolomic data, unlike sequencing data, is not compositional in nature. Metabolites are measured in absolute abundance; that is, a change in the abundance of one metabolite will not affect that of another. As the absolute abundance of gene expression cannot be measured by meta-transcriptomics, there is a considerable amount of disconnect between sequence-based and

metabolomic data, which has only just begun to be addressed. Our work highlights this fact and suggests using metabolomics to guide transcriptome analysis, as measured by RNA-seq or more quantitative methods such as RT-qPCR, may be a more fruitful approach for characterization of microbial communities and their relation to disease processes. Of course, this approach is only useful when the transcripts responsible from metabolite production are known. As a large proportion of RNA-seq reads have no known function and/or cannot be mapped to any sequenced genome^{19,20}, identifying the source of metabolites of unknown origin, such as 2HV, remains a formidable challenge.

3.5 References

- (1) Macklaim, J. M.; Fernandes, A. D.; Di Bella, J. M.; Hammond, J.-A.; Reid, G.; Gloor, G. B. Comparative Meta-RNA-Seq of the Vaginal Microbiota and Differential Expression by *Lactobacillus iners* in Health and Dysbiosis. *Microbiome* **2013**, *1* (1), 12.
- (2) McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J. E.; Gloor, G. B.; Reid, G. A Multi-Platform Metabolomics Approach Identifies Highly Specific Biomarkers of Bacterial Diversity in the Vagina of Pregnant and Non-Pregnant Women. *Sci. Rep.* **2015**, *5*, 14174.
- (3) Wang, Q.; Garrity, G. M.; Tiedje, J. M.; Cole, J. R. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* **2007**, *73* (16), 5261–5267.
- (4) Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's All Relative: Analyzing Microbiome Data as Compositions. *Ann. Epidemiol.* **2016**, 1–8.
- (5) Gloor, G. B.; Reid, G. Compositional Analysis: A Valid Approach to Analyze Microbiome High Throughput Sequencing Data. *Can. J. Microbiol.* **2016**, cjm – 2015–0821.
- (6) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J. Am. Soc.*

Mass Spectrom. **1999**, *10* (8), 770–781.

- (7) Styczynski, M. P.; Moxley, J. F.; Tong, L. V; Walther, J. L.; Jensen, K. L.; Stephanopoulos, G. N. Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery. *Anal. Chem.* **2007**, *79* (3), 966–973.
- (8) Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B. Methodol.* **1982**, *44* (2), 139–177.
- (9) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995, pp 289–300.
- (10) Swann, J. R.; Tuohy, K. M.; Lindfors, P.; Brown, D. T.; Gibson, G. R.; Wilson, I. D.; Sidaway, J.; Nicholson, J. K.; Holmes, E. Variation in Antibiotic-Induced Microbial Recolonization Impacts on the Host Metabolic Phenotypes of Rats. *J. Proteome Res.* **2011**, *10* (8), 3590–3603.
- (11) Kendall, M. G. A New Measure of Rank Correlation. *Biometrika* **1938**, *30* (1/2), 81–93.
- (12) Fredricks, D. N.; Fiedler, T. L.; Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **2005**, *353* (18), 1899–1911.
- (13) Muzny, C. A.; Sunesara, I. R.; Griswold, M. E.; Kumar, R.; Lefkowitz, E. J.; Mena, L. A.; Schwebke, J. R.; Martin, D. H.; Swiatlo, E. Association between BVAB1 and High Nugent Scores among Women with Bacterial Vaginosis. *Diagn. Microbiol. Infect. Dis.* **2014**, *80* (4), 321–323.
- (14) Overbeek, R.; Begley, T.; Butler, R. M.; Choudhuri, J. V; Chuang, H.-Y.; Cohoon, M.; de Crécy-Lagard, V.; Diaz, N.; Disz, T.; Edwards, R.; Fonstein, M.; Frank, E. D.; Gerdes, S.; Glass, E. M.; Goesmann, A.; Hanson, A.; Iwata-Reuyl, D.; Jensen, R.; Jamshidi, N.; Krause, L.; Kubal, M.; Larsen, N.; Linke, B.; McHardy, A. C.;

- Meyer, F.; Neuweger, H.; Olsen, G.; Olson, R.; Osterman, A.; Portnoy, V.; Pusch, G. D.; Rodionov, D. A.; Rückert, C.; Steiner, J.; Stevens, R.; Thiele, I.; Vassieva, O.; Ye, Y.; Zagnitko, O.; Vonstein, V. The Subsystems Approach to Genome Annotation and Its Use in the Project to Annotate 1000 Genomes. *Nucleic Acids Res.* **2005**, *33* (17), 5691–5702.
- (15) Kanehisa, M.; Goto, S. Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.
- (16) Gale, E. F.; Epps, H. M. Studies on Bacterial Amino-Acid Decarboxylases: 1. L(+)-Lysine Decarboxylase. *Biochem. J.* **1944**, *38* (3), 232–242.
- (17) Fothergill, J. C.; Guest, J. R. Catabolism of L-Lysine by *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **1977**, *99* (1), 139–155.
- (18) Söhling, B.; Gottschalk, G. Molecular Analysis of the Anaerobic Succinate Degradation Pathway in *Clostridium kluyveri*. *J. Bacteriol.* **1996**, *178* (3), 871–880.
- (19) Gill, S. R.; Pop, M.; Deboy, R. T.; Eckburg, P. B.; Turnbaugh, P. J.; Samuel, B. S.; Gordon, J. I.; Relman, D. a; Fraser-Liggett, C. M.; Nelson, K. E. Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* (80-.). **2006**, *312* (5778), 1355–1359.
- (20) Prakash, T.; Taylor, T. D. Functional Assignment of Metagenomic Data: Challenges and Applications. *Brief. Bioinform.* **2012**, *13* (6), 711–727.

Chapter 4

4 Adhesion of *Lactobacillus iners* AB-1 to human fibronectin: a key mediator for persistence in the vagina?

This chapter is reproduced with permissions (Appendix B) from:

McMillan, A, Macklaim, JM, Burton, JP, and G Reid. (2012). Adhesion of *Lactobacillus iners* AB-1 to human fibronectin: a key mediator for persistence in the vagina? *Reproductive Sciences* 20 (7), 791-796.

4.1 Introduction

The vaginal microbiota plays an important role in maintaining female and fetal health, with over 250 bacterial species identified¹⁻³. It has been known for some time that a healthy vagina is dominated by *Lactobacillus* species, in particular *Lactobacillus iners* and *Lactobacillus crispatus*. A shift from this state can result in aberrant conditions such as bacterial vaginosis (BV)¹⁻⁴ and aerobic vaginitis (AV)⁵, as well as urinary tract infection (UTI)^{6,7}.

The predominant species in the vagina as determined by 16S rRNA gene sequencing is *Lactobacillus iners*^{2,8-10}. Due to its demanding nutritional and growth requirements, this species has been largely overlooked in past bacteriologically based studies. In a deep sequencing study of women with BV, *L. iners* was the only species to increase in numbers following metronidazole treatment, demonstrating its unique ability to persist in fluctuating vaginal environments². *L. iners* has also been shown *in vitro* to displace biofilms of *Gardnerella vaginalis*¹¹, a bacterium associated with BV^{2,10}. Others have suggested that because *L. iners* can be found in women with infection, the species is not protective^{4,12}, but recent studies show that *L. iners* changes its gene expression to adapt to BV conditions by increasing metabolic activity against carbohydrates (manuscript being submitted). This suggests that *L. iners* may provide a nidus for recovery from BV and restoration of homeostasis.

Adhesion is the first step in bacterial colonization of a surface, and in the gastrointestinal tract *Lactobacillus* adhesion is commonly mediated by mucus-binding protein¹³. Adhesion

is also believed to play a role in pathogen exclusion via blockage of their binding sites on the mucosa¹⁴. The ability to produce adhesion proteins has been shown in the genome of *L. iners* AB-1, potentially providing it with a means to colonize the vagina¹⁵. Genomic sequencing of the strain *L. iners* AB-1 revealed that it has the smallest *Lactobacillus* genome known to date (1.3Mbp), and lacks better known adhesion molecules common to other *Lactobacillus* species, none of which have the persistence capability of *L. iners*. This suggests that *L. iners* contains somewhat of a unique mechanism to persist in the host.

A predicted fibronectin binding protein in *L. iners* AB-1 was of interest as it contains a motif (fbpA) common to pathogenic strains of *Staphylococcus aureus*, where it mediates adhesion and internalization into host cells¹⁶. Although there is no evidence that *L. iners* internalizes into host cells¹⁵, protein predictions suggest *L. iners* may adhere to the vaginal epithelium via fibronectin. Fibronectin (Fn) is a 454 kDa glycoprotein found in a soluble form in plasma, and an insoluble form in the extracellular matrix and attached to the surface of host cells. We demonstrate here that *L. iners* AB-1 binds strongly to human fibronectin and that this process is protein mediated.

4.2 Materials and methods

The overall design of the experiments was to determine the extent of binding of *L. iners* to fibronectin, to compare it to a well known binder, *S. aureus*, in order to better understand how the lactobacilli might utilize the glycoprotein for persistence in the vagina.

4.2.1 Strains and growth conditions

The strains used in this study are listed in Table 4-1. *Lactobacillus crispatus* was selected as it is the second most common vaginal lactobacilli after *L. iners*^{1-3,30}. *Lactobacillus rhamnosus* GR-1 and *L. reuteri* RC-14 are used in a combination probiotic therapy for Bacterial Vaginosis, but do not themselves colonize the vagina²⁸. *Lactobacillus rhamnosus* GG is a common gut probiotic and was chosen to represent a species not native to the vagina²⁹. *Staphylococcus aureus* was used as a positive control as its fibronectin binding properties have been well documented^{16,31}.

L. rhamnosus GR-1, *L. rhamnosus* GG, *L. reuteri* RC-14, and *L. crispatus* ATCC 33820 were grown in de Man, Rogosa and Sharpe (MRS) agar and broth (Difco, MD), anaerobically at 37°C under static conditions. *L. iners* AB-1 was grown on Columbia blood agar (Difco), then sub cultured in 10ml MRS broth and grown anaerobically at 37°C overnight under static conditions. Growth of *L. iners* AB-1 in MRS broth is shown in Figure 4-1. *Staphylococcus aureus* MN8 was grown in brain heart infusion (BHI, Difco) agar and broth at 37°C aerobically with agitation.

Table 4-1. Bacterial strains used in this study

Strain	Source/relevant characteristics
<i>Lactobacillus iners</i> AB-1	Vaginal isolate; from healthy woman ¹⁵
<i>Lactobacillus rhamnosus</i> GR-1	Distal urethral isolate; from a healthy woman ²⁸
<i>Lactobacillus rhamnosus</i> GG	Human feces isolate ²⁹
<i>Lactobacillus reuteri</i> RC-14	Vaginal isolate; from a healthy woman ²⁸
<i>Lactobacillus crispatus</i> ATCC 33820	Vaginal isolate ³⁰
<i>Staphylococcus aureus</i> MN8	Clinical isolate from patient with menstrual toxic shock syndrome (TSS) ³¹

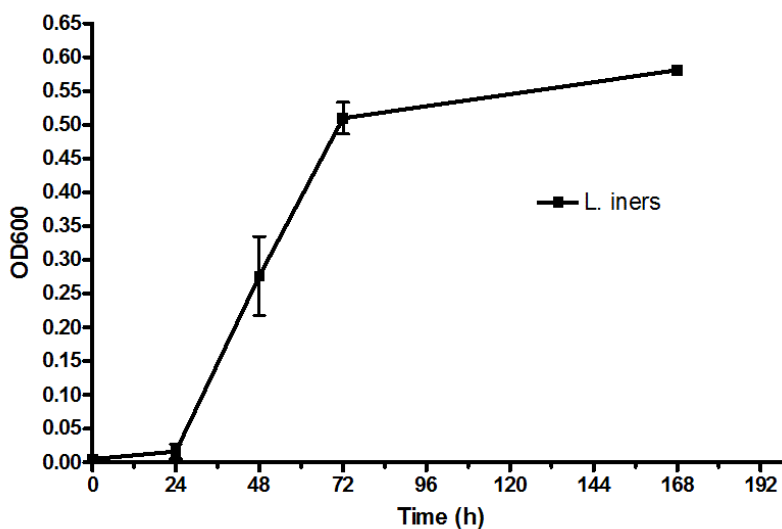


Figure 4-1. *L. iners* AB-1 grown in 10 ml MRS broth for 7 days.

The bars represent standard deviation. The OD600 is a measurement of the density of the *L. iners* growth.

4.2.2 Fibronectin adhesion assay

Binding of bacteria to immobilized fibronectin was measured according to the methods of Munoz-Provencio *et al.*, 2009¹⁷. Briefly, 96-well Polysorp plates (Nunc, Rochester, NY) were coated with 50 µg/ml human fibronectin (Fn) (Sigma Chem. Co, St Louis, MO) in carbonate/bicarbonate buffer 50 mmol/l pH 9.6 and incubated at 4°C overnight. Wells were washed three times with 100 µl PBS (pH 7.4), then blocked with 100 µl PBS + 1% Tween 20 for 1 h. All strains were grown to late exponential phase (54 h for *L. iners* AB-1, 5-6 h for all other strains tested), washed three times with PBS, and re-suspended in PBS to an optical density that represented 10⁷ CFU/ml as determined previously by growth curves. 100 µl of each strain was then added to each well and incubated at 37°C for 2 hours. After incubation non-adherent cells were removed by three washes with 200 µl PBS + 0.05% Tween 20. Adhered cells were stained with 100 µl of 1 mg/ml crystal violet for 45 minutes. After washing, colorant was released with 100 µl citrate buffer (50 mmol/l, pH 4.0), and the absorbance at 600 nm was measured in a Multiskan Ascent plate reader (Thermo-Scientific, Rockford, U. S.). Drop dilution plates using 10 ul of culture and series of 10 times dilutions were performed simultaneously with binding assays to enumerate bacterial numbers upon addition to Fn-coated plates. Experiments were repeated three times and in triplicate.

The effect of protease treatment was assayed by incubating cells with 100 µg/ml proteinase K for 1 h at 37 °C followed by 1 h incubation with 1 mmol/L phenylmethanesulfonylfluoride (PMSF) to inactivate proteinase K. Cells were then transferred to Fn-coated plate. Experiments were repeated twice and in duplicate. Wells without Fn were run as controls in all experiments and their absorbance values subtracted from Fn-coated wells. PBS alone was included in all experiments to account for any Fn binding by PBS.

4.2.3 Bacterial binding to fibronectin coated cover slips

Binding of bacteria to fibronectin coated cover slips was determined by coating glass cover slips overnight with 50 ug/mL fibronectin in PBS + 0.05 % Tween 20 at 4°C. Slips (VWR 18 mm glass) were washed three times with 1 mL PBS, and blocked with 1 mL PBS + 1 % Tween 20 for 1 h. Cells were grown to late exponential phase, washed twice with PBS and diluted to a concentration of 10^7 CFU/mL with PBS. Numbers were confirmed by drop plating. One mL of these cultures was added to separate cover slips and incubated for 2 h at 37°C. Non-adherent cells were removed by three washes with PBS + 0.05 % Tween 20. Cover slips were then Gram-stained and adherent cells viewed under a Zeiss Axioscop HBO50 light microscope under 100 x objective. Cover slips without fibronectin were run as controls in all experiments. Five fields of view were recorded for each cover slip and biological replicates were repeated three times to confirm results.

4.2.4 Statistical analysis

Statistical analysis was performed on all data from binding assays using the GraphPad Prism[®] 4 program. The data fitted a normal distribution and was therefore appropriate for ANOVA testing. Differences between the normally distributed means were compared using a one-way ANOVA analysis with a Tukeys post hoc test. Differences in binding were considered significant with p values of < 0.05.

4.3 Results

Lactobacillus iners AB-1 bound to fibronectin coated polystyrene plates in significantly greater amounts than all strains tested (Fig. 2). The effect of proteinase K, a serine protease, was determined by incubating bacteria with proteinase K followed by the protease inhibitor PMSF before addition of cells to fibronectin coated plates. Proteinase K treatment significantly reduced binding of *L. iners* AB-1 to fibronectin ($p < 0.05$) (Figure 4-3).

Photomicrograph images of cells adhering to fibronectin coated and uncoated cover slips are shown in Figure 4-4, and bacterial cells counts in Table 4-2. Adhesion of *L. iners* AB-

1 and *S. aureus* MN8 was enhanced by coating slips with fibronectin. There was no difference in adhesion to Fn of *L. iners* AB-1 compared with *S. aureus* ($p > 0.05$). The addition of Fn had no effect on adhesion of *L. crispatus* ($p > 0.05$).

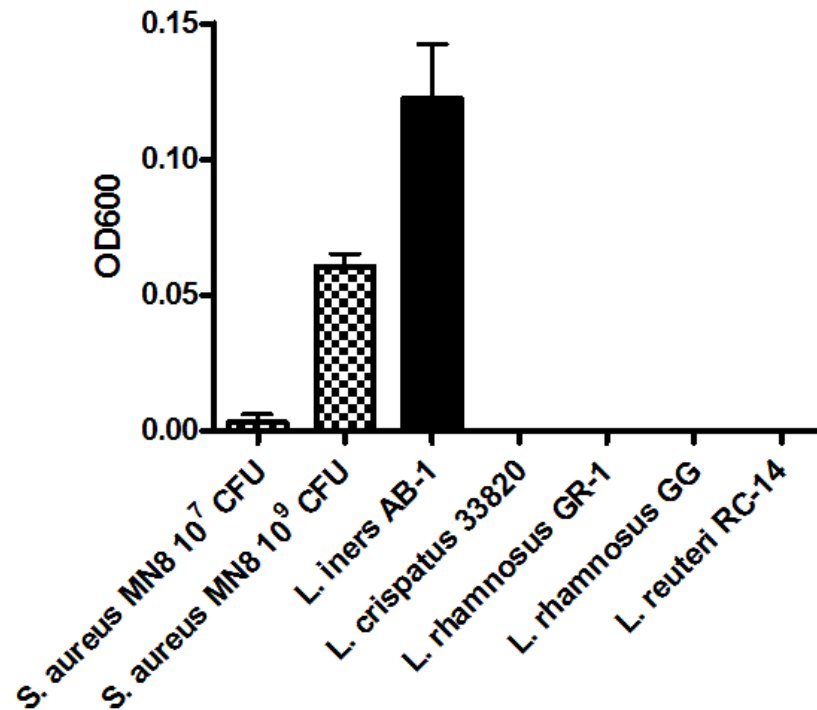


Figure 4-2. Bacterial binding to immobilized fibronectin as measured by optical density at 600 nm (OD600).

All strains were added to fibronectin-coated plates at a concentration of 10^7 CFU/mL unless indicated and bacterial numbers were confirmed by drop plating. Binding by *L. iners* AB-1 is significantly greater than all strains ($p < 0.05$). $n = 9$. The bars represent standard deviation.

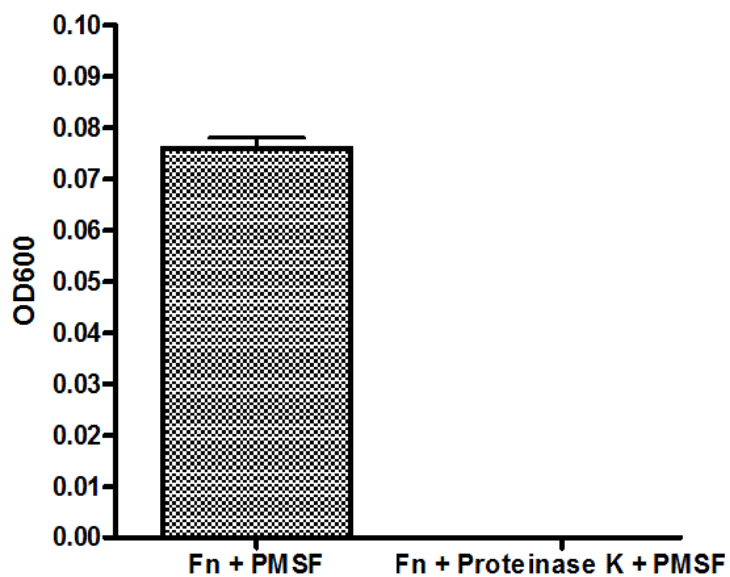


Figure 4-3. Effect of proteinase K pre-treatment on adhesion to immobilized fibronectin by *L. iners* AB-1.

PMSF (phenylmethanesulfonylfluoride) is a proteinase K inhibitor. *L. iners* AB-1 binding was significantly reduced by proteinase K pre-treatment ($p < 0.05$). $n = 4$. The bars represent standard deviation.

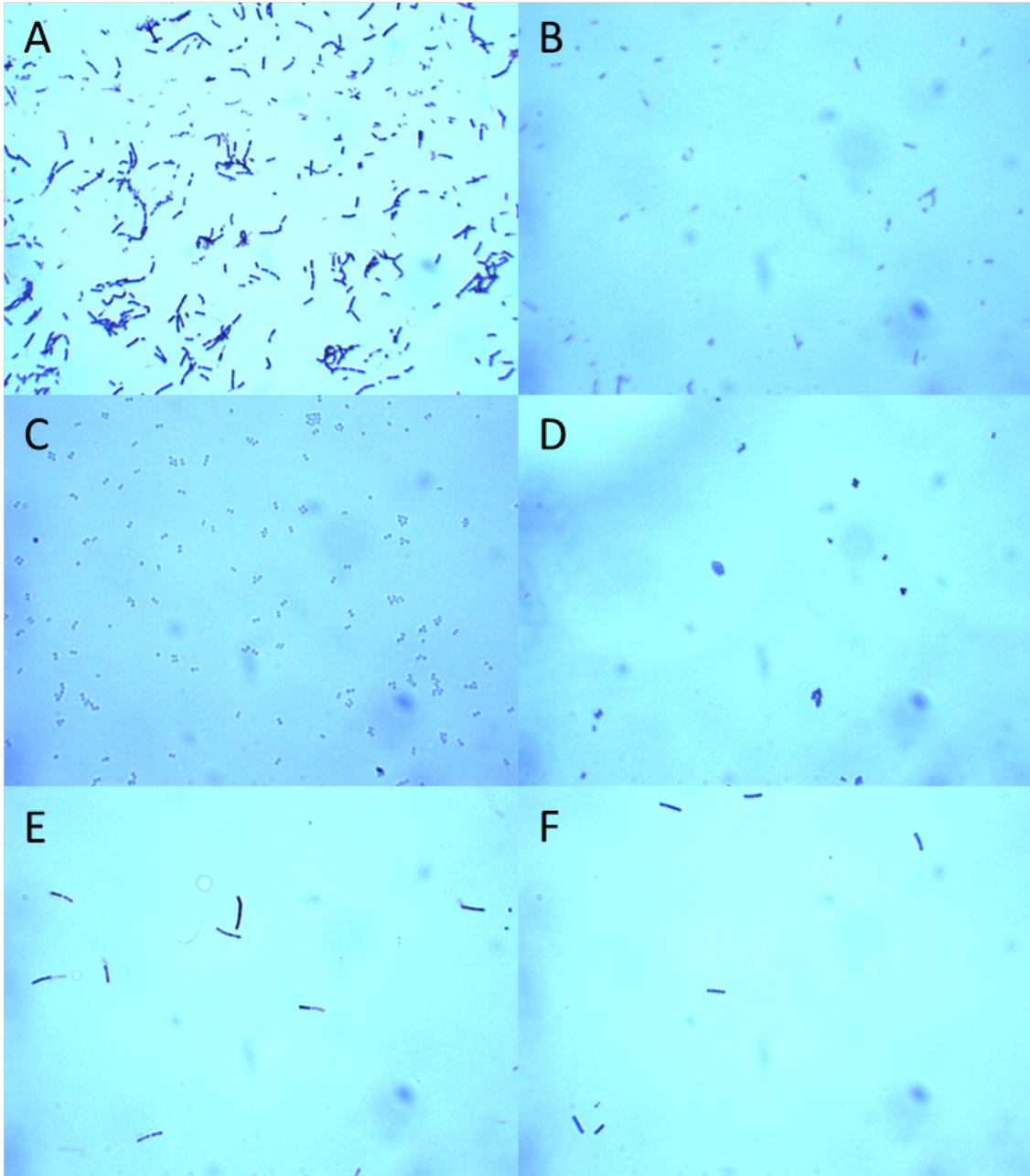


Figure 4-4. Bacterial adhesion to fibronectin coated cover slips.

A. *L. iners* + fibronectin B. *L. iners* no fibronectin C. *S. aureus* + fibronectin D. *S. aureus* no fibronectin. E. *L. crispatus* + fibronectin. F. *L. crispatus* no fibronectin.

Table 4-2. Bacterial adhesion to fibronectin coated cover slips cell counts.

Counts represent an average of 3 fields of view \pm SD. Adhesion of MN8 and AB-1 to fibronectin was statistically higher ($p < 0.01$), while that of 33820 was not altered ($p > 0.05$).

Strain	Fibronectin	No Fibronectin
<i>S. aureus</i> MN8	299 \pm 28.2	54 \pm 19.7
<i>L. iners</i> AB-1	341 \pm 27.7	33 \pm 19.2
<i>L. crispatus</i> 33820	8 \pm 3.1	20 \pm 11.0

4.4 Discussion

The concentration of Fn in the vagina has been reported from as high as 90-140 pg/ml during menses to 3-5 pg/ml at the end of the menstrual cycle¹⁸. As *L. iners* possesses a predicted fibronectin binding protein and adhesion is the first step in colonization of a surface, we hypothesized that strong adhesion to Fn could be contributing to its persistence in the human vagina¹⁵. *L. iners* AB-1 bound significantly greater to human Fn compared to other *Lactobacillus* strains tested, including vaginal probiotic *L. rhamnosus* GR-1 and *L. reuteri* RC-14, and gut probiotic *L. rhamnosus* GG (Figure 4-2), none of which are able to persist long term in the vagina. However, since strains *L. rhamnosus* GR-1 and *L. reuteri* RC-14 help restore the vaginal microbiota back to a healthy state^{19,20} they must use different adhesion mechanisms, such as extracellular polysaccharides, or metabolic pathways, in the vagina. *L. iners* bound Fn significantly better than the *S. aureus* positive control as assessed by crystal violet microplate assay, however this binding was more comparable when assessed on a per cell basis via light microscopy. These results indicate that differences in cell surface area and/or size may account for the apparent difference in Fn binding between *S. aureus* and *L. iners* as detected by crystal violet staining alone.

In relation to pregnancy, fetal fibronectin found in the amniotic fluid, placental tissue, and the decidua basalis, but its presence in the cervicovaginal secretions between 22 and 37 weeks may be due to disruption of the chorionic-decidual interface secondary to infection²¹. Hypothetically, *L. iners* could bind to fibronectin and potentially counter pathogenesis thereby extending the duration of the pregnancy. Since *L. iners* has only

relatively recently been discovered and detected using molecular methodologies, no studies have looked for its presence during elevated fibronectin and infection. Investigations of this nature perhaps warrant undertaking.

This is the first report of *L. iners* having such strong adhesion to a mucosal protein. This trend was confirmed by viewing cells bound to Fn coated cover slips under a light microscope (Figure 4-4). Interestingly, *L. iners* AB-1 bound equally or stronger to Fn compared to *S. aureus* MN8, a species well studied for its Fn binding properties¹⁶. *L. crispatus*, another member of the core vaginal microbiota, lacks Fn binding properties, indicating that *L. iners* and *L. crispatus* use different adhesion mechanisms to colonize the same environment. This is particularly interesting since *L. crispatus* strains do not appear to persist to the extent of *L. iners*, and their displacement coincides with onset of BV¹.

When candidate probiotic strain *L. crispatus* CTV-05 did colonize hosts who lacked indigenous *L. crispatus*, a healthy state was maintained²². It is not clear how BV organisms induce *L. crispatus* and not *L. iners* displacement, but the avidity of adhesion to Fn of the latter likely plays a role. A recent study using atomic force microscopy highlighted the weaker adhesion of *L. crispatus* and *L. jensenii* and suggested that for *Lactobacillus* species to displace pathogens, they need to display strong adhesion forces²³. This is not ubiquitous across *Lactobacillus*, and Nagy *et al.*,²⁴ found only 9/54 (16.7%) of vaginal *Lactobacillus* isolates bound Fn at pH 7.2, however, *L. iners* was not included as it had not yet been discovered. They also observed that adhesion to Fn increased when the pH was lowered to 4.0. This may help explain the loss of some *Lactobacillus* species during BV, where the pH is elevated. It also suggests that *L. iners* is able to persist during incidences of BV due in part to its ability to bind Fn at a more neutral pH as shown in our study. The fluctuations in vaginal Fn during episodes of BV have not yet been elucidated. Future work should determine if free Fn increases during BV, which might offer *L. iners* a competitive advantage over other vaginal lactobacilli.

To determine if Fn adhesion by *L. iners* was protein-mediated we pre-treated cells with the serine protease, proteinase K. This treatment significantly reduced binding to Fn ($p < 0.05$), confirming that Fn adherence by *L. iners* AB-1 is protein-mediated. A predicted Fn binding

protein (LINAB1_0564) in *L. iners* AB-1 could be contributing to this interaction. This protein has 34% amino acid similarity to a Fn binding protein of *S. aureus* MN8. Furthermore, analysis of the transcriptome of *L. iners* has revealed that this gene is expressed in both healthy patients and those with BV (manuscript submitted), indicating the protein may be used during both conditions to adhere to the vaginal mucosa and colonize the vagina.

In summary, the persistence of *L. iners* in the vagina despite the presence of other lactobacilli and pathogens or antibiotics could be due to its strong adhesion to Fn. If clones of this species do indeed form an environment that allows restitution of a *Lactobacillus* microbiota that protects against disease, future probiotic strains, such as those used against HIV^{25,26} or ones devised to compete with Fn binding *Trichomonas*²⁷ could benefit from cloning of the *L. iners* Fn binding protein.

4.5 References

- (1) Fredricks, D. N.; Fiedler, T. L.; Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **2005**, 353 (18), 1899–1911.
- (2) Hummelen, R.; Fernandes, A. D.; Macklaim, J. M.; Dickson, R. J.; Changalucha, J.; Gloor, G. B.; Reid, G. Deep Sequencing of the Vaginal Microbiota of Women with HIV. *PLoS One* **2010**, 5 (8).
- (3) Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G. M.; Koenig, S. S. K.; McCulle, S. L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C. O.; Brotman, R. M.; Davis, C. C.; Ault, K.; Peralta, L.; Forney, L. J. Vaginal Microbiome of Reproductive-Age Women. *Proc. Natl. Acad. Sci.* **2010**, 108 (Supplement_1), 4680–4687.
- (4) Tamrakar, R.; Yamada, T.; Furuta, I.; Cho, K.; Morikawa, M.; Yamada, H.; Sakuragi, N.; Minakami, H. Association between *Lactobacillus* Species and Bacterial Vaginosis-Related Bacteria, and Bacterial Vaginosis Scores in Pregnant Japanese Women. *BMC Infect. Dis.* **2007**, 7, 128.
- (5) Donders, G. G. G.; Vereecken, A.; Bosmans, E.; Dekeersmaecker, A.; Salembier,

- G.; Spitz, B. Definition of a Type of Abnormal Vaginal Flora That Is Distinct from Bacterial Vaginosis: Aerobic Vaginitis. *BJOG* **2002**, *109* (1), 34–43.
- (6) Cadieux, P. A.; Burton, J. P.; Devillard, E.; Reid, G. *Lactobacillus* by-Products Inhibit the Growth and Virulence of Uropathogenic *Escherichia coli*. *J. Physiol. Pharmacol.* **2009**, *60* (SUPPL.6), 13–18.
- (7) Sharami, S. H.; Afrakhteh, M.; Shakiba, M. Urinary Tract Infections in Pregnant Women with Bacterial Vaginosis. *J. Obstet. Gynaecol.* **2007**, *27* (3), 252–254.
- (8) Burton, J. P.; Cadieux, P. A.; Reid, G. Improved Understanding of the Bacterial Vaginal Microbiota of Women before and after Probiotic Instillation. *Appl. Environ. Microbiol.* **2003**, *69* (1), 97–101.
- (9) Sobel, J. D. Bacterial Vaginosis. *Annu. Rev. Med.* **2000**, *51*, 349–356.
- (10) Zhou, X.; Hansmann, M. A.; Davis, C. C.; Suzuki, H.; Brown, C. J.; Sch??tte, U.; Pierson, J. D.; Forney, L. J. The Vaginal Bacterial Communities of Japanese Women Resemble Those of Women in Other Racial Groups. *FEMS Immunol. Med. Microbiol.* **2010**, *58* (2), 169–181.
- (11) Saunders, S.; Bocking, A.; Challis, J.; Reid, G. Effect of *Lactobacillus* Challenge on *Gardnerella vaginalis* Biofilms. *Colloids Surf. B. Biointerfaces* **2007**, *55* (2), 138–142.
- (12) Yan, D. H.; Lü, Z.; Su, J. R. Comparison of Main *Lactobacillus* Species between Healthy Women and Women with Bacterial Vaginosis. *Chin. Med. J. (Engl)*. **2009**, *122* (22), 2748–2751.
- (13) Vélez, M. P.; De Keersmaecker, S. C. J.; Vanderleyden, J. Adherence Factors of *Lactobacillus* in the Human Gastrointestinal Tract. *FEMS Microbiology Letters*. 2007, pp 140–148.
- (14) Irvine, S. L.; Hummelen, R.; Hekmat, S.; Looman, C. W. N.; Habbema, J. D. F.; Reid, G. Probiotic Yogurt Consumption Is Associated with an Increase of CD4

- Count among People Living with HIV/AIDS. *J. Clin. Gastroenterol.* **2010**, *44* (9), e201–e205.
- (15) Macklaim, J. M.; Gloor, G. B.; Anukam, K. C.; Cribby, S.; Reid, G. At the Crossroads of Vaginal Health and Disease, the Genome Sequence of *Lactobacillus iners* AB-1. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108 Suppl*, 4688–4695.
- (16) Sinha, B.; François, P. P.; Nüsse, O.; Foti, M.; Hartford, O. M.; Vaudaux, P.; Foster, T. J.; Lew, D. P.; Herrmann, M.; Krause, K. H. Fibronectin-Binding Protein Acts as *Staphylococcus aureus* Invasin via Fibronectin Bridging to Integrin alpha5beta1. *Cell. Microbiol.* **1999**, *1* (2), 101–117.
- (17) Muñoz-Provencio, D.; Pérez-Martínez, G.; Monedero, V. Characterization of a Fibronectin-Binding Protein from *Lactobacillus casei* BL23. *J. Appl. Microbiol.* **2010**, *108* (3), 1050–1059.
- (18) Antonas K, Yajko D, H. W. Assay of Fibronectin in Vaginal Fluid and the Role of Fibronectin in Bacterial Colonization. *Annu. Meet. Am. Soc. Microbiol.* **1983**, *Abstract n*.
- (19) Martinez, R. C.; Franceschini, S. A.; Patta, M. C.; Quintana, S. M.; Gomes, B. C.; De Martinis, E. C.; Reid, G. Improved Cure of Bacterial Vaginosis with Single Dose of Tinidazole (2 G), *Lactobacillus rhamnosus* GR-1, and *Lactobacillus reuteri* RC-14: A Randomized, Double-Blind, Placebo-Controlled Trial. *Can J Microbiol* **2009**, *55* (2), 133–138.
- (20) Reid, G.; Charbonneau, D.; Erb, J.; Kochanowski, B.; Beuerman, D.; Poehner, R.; Bruce, A. W. Oral Use of *Lactobacillus rhamnosus* GR-1 and *L. fermentum* RC-14 Significantly Alters Vaginal Flora: Randomized, Placebo-Controlled Trial in 64 Healthy Women. *FEMS Immunol. Med. Microbiol.* **2003**, *35* (2), 131–134.
- (21) Dutta, D.; Norman, J. E. The Efficacy of Fetal Fibronectin Testing in Minimising Hospital Admissions, Length of Hospital Stay and Cost Savings in Women Presenting with Symptoms of Pre-Term Labour. *J. Obstet. Gynaecol.* **2010**, *30* (8),

768–773.

- (22) Antonio, M. A. D.; Meyn, L. A.; Murray, P. J.; Busse, B.; Hillier, S. L. Vaginal Colonization by Probiotic *Lactobacillus crispatus* CTV-05 Is Decreased by Sexual Activity and Endogenous Lactobacilli. *J. Infect. Dis.* **2009**, *199* (10), 1506–1513.
- (23) Younes, J. A.; van der Mei, H. C.; van den Heuvel, E.; Busscher, H. J.; Reid, G. Adhesion Forces and Coaggregation between Vaginal Staphylococci and Lactobacilli. *PLoS One* **2012**, *7* (5), e36917.
- (24) Nagy, E.; Fröman, G.; Mårdh, P. A. Fibronectin Binding of *Lactobacillus* Species Isolated from Women with and without Bacterial Vaginosis. *J. Med. Microbiol.* **1992**, *37* (1), 38–42.
- (25) Pusch, O.; Kalyanaraman, R.; Tucker, L. D.; Wells, J. M.; Ramratnam, B.; Boden, D. An Anti-HIV Microbicide Engineered in Commensal Bacteria: Secretion of HIV-1 Fusion Inhibitors by Lactobacilli. *AIDS* **2006**, *20* (15), 1917–1922.
- (26) Liu, J. J.; Reid, G.; Jiang, Y.; Turner, M. S.; Tsai, C.-C. Activity of HIV Entry and Fusion Inhibitors Expressed by the Human Vaginal Colonizing Probiotic *Lactobacillus reuteri* RC-14. *Cell. Microbiol.* **2007**, *9* (1), 120–130.
- (27) Alderete, J. F.; Benchimol, M.; Lehker, M. W.; Crouch, M. L. The Complex Fibronectin-*Trichomonas vaginalis* Interactions and Trichomonosis. *Parasitology International*. 2002, pp 285–292.
- (28) Reid, G.; Cook, R. L.; Bruce, A. W. Examination of Strains of Lactobacilli for Properties That May Influence Bacterial Interference in the Urinary Tract. *J. Urol.* **1987**, *138* (2), 330–335.
- (29) Silva, M.; Jacobus, N. V.; Deneke, C.; Gorbach, S. L. Antimicrobial Substance from a Human *Lactobacillus* Strain. *Antimicrob. Agents Chemother.* **1987**, *31* (8), 1231–1233.
- (30) Skerman, V.; McGowan, V.; Sneath, P. *Approved Lists of Bacterial Names*

(Amended); ASM Press, 1989.

- (31) Schlievert, P. M.; Blomster, D. A. Production of Staphylococcal Pyrogenic Exotoxin Type C: Influence of Physical and Chemical Factors. *J. Infect. Dis.* **1983**, *147* (2), 236–242.

Chapter 5

5 Metabolic derangements identified through untargeted metabolomics in a cross-sectional study of Nigerian children with severe acute malnutrition

A. McMillan, A.E. Orimadegun, M.W. Sumarah, J. Renaud, M. Muc, G.B. Gloor, O.O. Akinyinka, G. Reid, and S.J. Allen.

Supplemental Tables 5-(1-6) are available for download as additional files (.xlsx)

5.1 Introduction

Severe acute malnutrition (SAM) is a leading cause of global child mortality, is associated with growth faltering, and results in impaired cognitive development¹. SAM is most frequently defined by a weight-for-length/height Z-score (WHZ) < -3 and/or presence of edema². Middle upper arm circumference (MUAC) < 11.5 cm is also commonly used in children aged 6-59 months². Despite its prevalence, the metabolic changes that occur during malnutrition are poorly understood. Identification of these metabolites could i) lead to improved interventions through identification of nutrient deficiencies, ii) advance our understanding of the metabolic adaptations to malnutrition and shed light on pathophysiology, and iii) identify biomarkers of nutritional status for diagnostic and monitoring purposes.

Amino acids are known to be depleted in children with SAM³, but untargeted studies pertaining to other metabolites are limited. Bartz et al⁴ applied a targeted and untargeted approach to identify biomarkers associated with recovery and mortality in Ugandan children with SAM. The concentrations of a number of metabolites were significantly altered upon rehabilitation, including amino acids, acylcarnitines and leptin. However, non-malnourished controls were not included, and the untargeted analysis was limited to gas chromatography-mass spectrometry (GC-MS) which covers only a small portion of the metabolome. Malnutrition in juvenile pigs has also revealed alterations in the metabolome, including changes in amino acids, choline metabolites, and products of microbial-mammalian co-metabolism⁵, suggesting the gut microbiota may be playing a role.

Differences in gut microbial composition of children with SAM compared to healthy co-twins and unrelated controls have been described in Malawian and Bangladeshi children^{6,7}. Additionally, specific taxa capable of causing enteropathy when combined with a low protein diet have been identified in mice⁸. Enteropathy resulting in malabsorption and increased intestinal permeability is associated with SAM in humans, as indicated by intestinal biopsies and dual sugar permeability tests⁹. This phenomenon may be a more severe form of environmental enteric dysfunction (EED), previously referred to as “tropical enteropathy”, which occurs almost universally in people exposed to poor hygiene and sanitation. The etiology of enteropathy in SAM is not well understood, but may involve repeated exposure to fecal enteropathogens combined with a low protein diet^{8,10}. There is currently no widely accepted non-administered biomarker for enteropathy, and therefore the role of enteropathy in malnutrition is difficult to determine.

The aim of this study was to identify conserved changes in the metabolome of children with SAM, both to identify biomarkers and to improve our understanding of the pathophysiology of the condition. We aimed to undertake a pragmatic study of unselected children admitted to a health facility in a region with endemic malnutrition, to maximise the relevance of our findings to clinical practice. Given the association with enteropathy and recent studies linking changes in the microbiota to SAM^{6,7}, we also measured biomarkers of intestinal inflammation and profiled the gut microbiota of these children.

5.2 Materials and methods

5.2.1 Study design and sample collection

Children aged 6 – 59 months with SAM (WHZ <-3 or MUAC <11.5 cms and/or nutritional edema) admitted to the Federal Medical Centre, Gusau, Zamfara State, Nigeria during July-September 2012 were enrolled in the study. The hospital serves both urban and rural populations and recruitment occurred during the rainy season when malnutrition is most common. Written information about the study and a verbal explanation in the appropriate local language was provided to parents/carers and signed or thumb-printed

consent secured. Clinical management was performed according to usual practice based on WHO guidelines including those for the management of SAM (WHO 1999). Well-nourished children (MUAC >12.5 cms or WHZ score \geq -1 and no nutritional edema) were recruited from the paediatric ward or outpatient clinics as controls. HIV positive individuals were excluded from the study. In both groups, baseline demographic and clinical information was collected on standard forms by one of five clinicians trained in the research methods (Supplemental Table 5-1). The first available stool sample was collected and its consistency and the presence of visible blood and mucus were recorded. At the time of clinical sampling, approximately 2.5 mL of venous blood was collected into EDTA tubes for the purposes of the study. Ethical approval was provided by the Joint Ethical Review Committees of the University of Ibadan / University College Hospital, Ibadan, Nigeria.

5.2.2 Metabolite extraction from stool and plasma

Approximately 250 mg of wet stool was lyophilized overnight. After drying, 40 mg was weighed into microcentrifuge tubes and extracted with 8:2 methanol:H₂O to a final concentration of 40 mg/mL. Samples were then vortexed for 30 sec, followed by centrifugation for 15 min at 10 000 rpm. 50 μ L of this supernatant was dried in a speedvac for GC-MS analysis and the remaining stored at -80°C for LC-MS.

Metabolites were extracted from plasma according to the methods of Dunn et al, 2011¹¹. Briefly, plasma samples were thawed on ice for 30 min. Once thawed, 805 μ L of 8:2 methanol:H₂O was added to 230 μ L of plasma to make a 4.5 fold dilution. Samples were vortexed for 15 sec and centrifuged at 15 000 rpm for 15 min to pellet precipitated proteins. 370 μ L of supernatant was then transferred to separate vials and dried down for GC-MS and LC-MS using a speedvac with no heat.

5.2.3 Untargeted GC-MS analyses

For stool, samples were derivatized with 40 μ L of 2% methoxyamine-HCl in pyridine (MOX) incubated at 50 °C for 90 min, followed by 20 μ L of N- Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) for 30 min at 50°C. Samples were then transferred to micro-

inserts before analysis by GC-MS. For plasma, 50 μL of MOX was added to dried samples and incubated for 90 min at 50°C. Fifty microlites of MSTFA was then added and incubated for 30 min at 50°C. Samples were then transferred to 1.5 ml microcentrifuge tubes and centrifuged at 15 000 rpm for 5 min to pellet debris. This supernatant was then transferred to HPLC vials containing inserts for analysis.

One μL of sample was injected into an Agilent 7890A GC, 5975 inert MSD with triple axis detector. Samples were injected using pulsed splitless mode using a 30 m DB5-MS column with 10 m duraguard, diameter 0.35mm, thickness 0.25 μm (J&W Scientific, Folsom, USA). Helium was used as the carrier gas at a constant flow rate of 1 mL/min. Oven temperature was held at 70°C for 5 min then increased at a rate of 5 °C/min to 300°C and held for 10 min. Solvent delay was set to 7 min, and total run time was 61 min. Masses between 25 m/z and 600 m/z were selected by the detector. All samples were run in random order and a single sample was run with every batch as a quality control to ensure machine consistency.

Chromatogram files were deconvoluted and converted to ELU format using AMDIS Mass Spectrometry software¹² with the sensitivity set to low, resolution to medium, and support threshold to high. Chromatograms were aligned using Spectconnect (<http://spectconnect.mit.edu>)¹³ with the support threshold set to low. The integrated signal (IS) matrix output was used for all further analysis. Zeros were replaced with two thirds the minimum detected value on a per metabolite basis¹⁴, followed by a log base 2 transformation. All further analyses were performed using these log-transformed values.

Metabolites were initially identified by comparison to the NIST 11 standard reference database (<http://www.nist.gov/srd/nist1a.cfm>). Identities of metabolites of interest were then confirmed by authentic standards if available.

5.2.4 Untargeted LC-MS analyses

For stool, 2.5 μL of 1 $\mu\text{g}/\text{mL}$ $^{13}\text{C}_6$ phenylalanine internal standard (Cambridge Isotopes, Tewksbury, USA) was added to 47.5 μL of extracted stool. For plasma, dried samples were

reconstituted in 85.5 μ L of ddH₂O. 4.5 μ L of 1 μ g/mL ¹³C₆ phenylalanine in ddH₂O was then added to each vial as an internal standard. Samples were vortexed for 15 sec, then transferred to microinserts and directly injected into an Agilent 1290 Infinity HPLC coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, USA) with a HESI (heated electrospray ionization) source. For HPLC, 2 μ L of each sample was injected into a ZORBAX Eclipse plus C18 2.1 x 50mm x 1.8 micron column. Mobile phase (A) consisted of 0.1% formic acid in water and mobile phase (B) consisted of 0.1% formic acid in acetonitrile. The initial composition of 0% (B) was held constant for 30s and increased to 100% over 3.0 min. For stool, mobile phase B was held at 100% for 1 minute and returned to 0% over 30s for a total run time of 5 min. For plasma, mobile phase B was held at 100% for 2 minutes and returned to 0% over 30s for a total run time of 6 min.

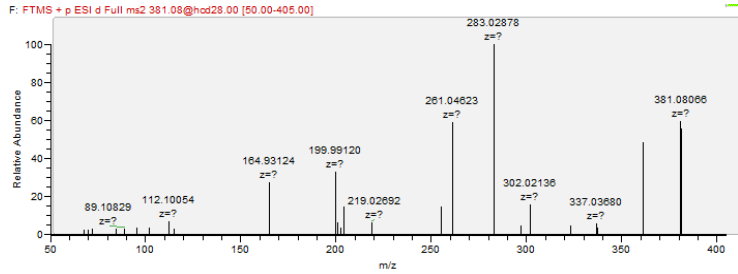
Full MS scanning between the ranges of m/z 50-750 was performed on all samples in both positive and negative mode at 140 000 resolution. The HESI source was operated under the following conditions: nitrogen flow of 30 and 8 arbitrary units for the sheath and auxiliary gas respectively, probe temperature and capillary temperature of 450°C and 250°C respectively and spray voltage of 3.9 kV and 3.5 kV in positive and negative mode respectively. The automatic gain control (AGC) target and maximum injection time were 1e6 and 500 ms respectively. For molecular characterization, every tenth sample was also analyzed with a data dependent MS/MS method where a 35 000 resolution full MS scan identified the top 12 signals above a 1e5 threshold which were subsequently selected at a 1.2 m/z isolation window for MS/MS. Normalized collision energy for MS/MS was 28, resolution 17 500, AGC target 1E5 and maximum injection time was 60ms. Blanks of pure methanol were run between every sample to limit carryover, and a single sample was run multiple times with every batch to account for any machine inconsistency. After data acquisition Thermo .RAW files were converted to .MZML format and centroided using `jjjProteoWizard`¹⁵. Files were then imported into R using the XCMS package¹⁶ for chromatogram alignment and deconvolution. Features were detected with the “`xcmsSet`” function using the “`centWave`” method and a ppm tolerance of 1. Prefilter was set to 3-5000, noise 1E5, and signal to noise threshold was set to 5. Due to a lower overall noise and signal in negative mode, noise was set to 1E3 for this mode. Retention time correction

was conducted using the “obiwarp” method, grouping included features present in at least 25% of samples, allowable retention time deviation was 5 seconds, and m/z width set to 0.015. Areas of features below the signal to noise threshold in the data were integrated using the “fillPeaks” function with default settings. Any remaining zeros in the data were then replaced with two-thirds the minimum value on a per mass basis¹⁴ before log base 2 transformation. The log-transformed mass list was then exported as a single .txt file and used for all further analyses. All further analyses were carried out in R unless otherwise specified. Positive and negative mode data were treated as two independent datasets for all analyses.

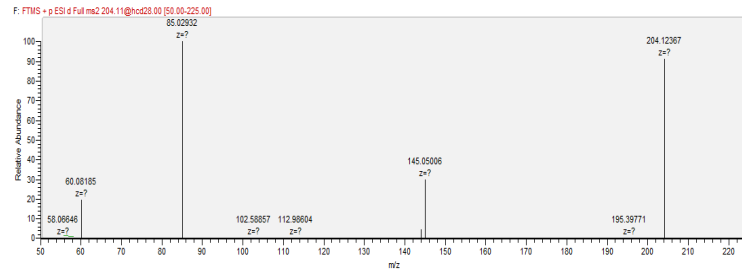
Metabolites were putatively identified based on accurate mass and LC-MS/MS fragmentation patterns (Error! Reference source not found.). Predictions were made mainly by *de novo* compound identification from in depth investigation of individual MS/MS spectra, utilizing the METLIN database (<http://metlin.scripps.edu>)¹⁷, Human Metabolome Database (www.hmdb.ca)¹⁸, and CFM-ID (<http://cfmid.wishartlab.com>)¹⁹ whenever possible to aid in identification. Metabolites of interest were confirmed by authentic standards when available based on accurate mass, retention time and MS/MS spectra (Supplemental Table 5-2).

Figure 5-1. LC-MS/MS spectra for metabolites elevated in plasma of children with SAM.

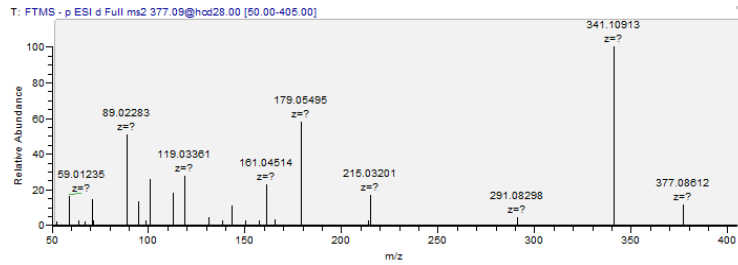
2C6 Disaccharides M+K



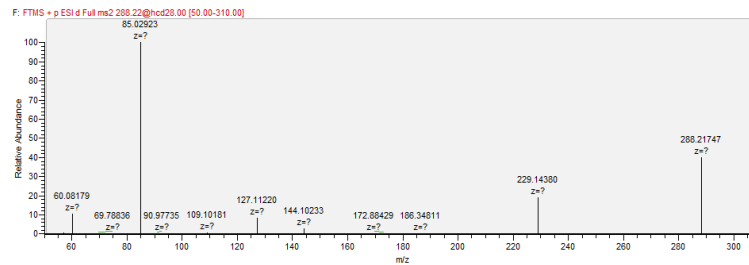
Acetylcarnitine M+H



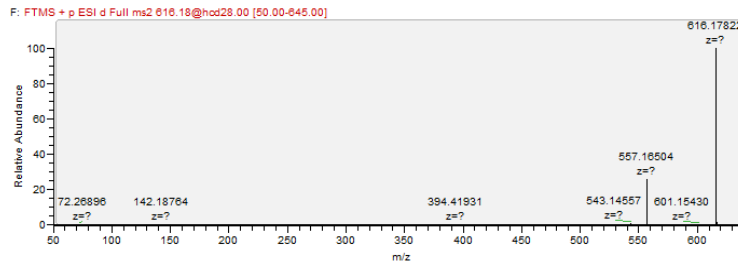
2C6 Disaccharides M+Cl-



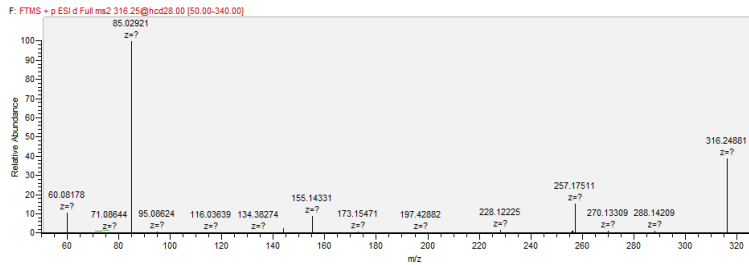
Octanoylcarnitine M+H



Heme M+

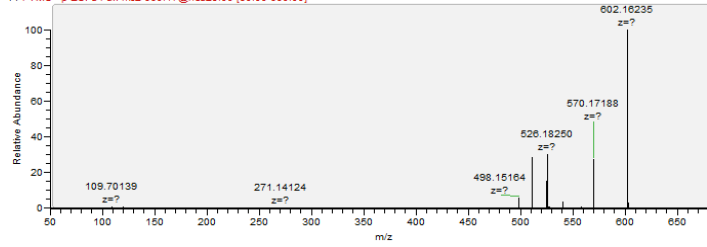


Decanoylcarnitine M+H



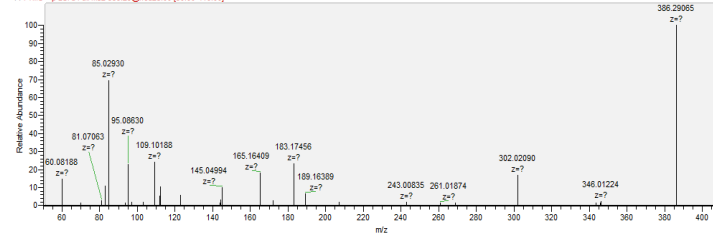
Heme M+FA-H

F: FTMS - p ESI d Full ms2 660.17@hcd28.00 [50.00-690.00]



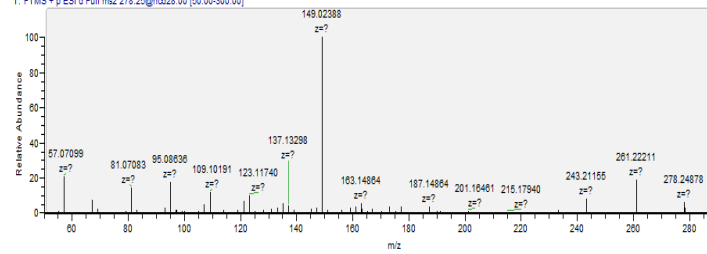
3-Hydroxy-5-tetradecenylcarnitine M+H

F: FTMS - p ESI d Full ms2 386.29@hcd28.00 [50.00-410.00]



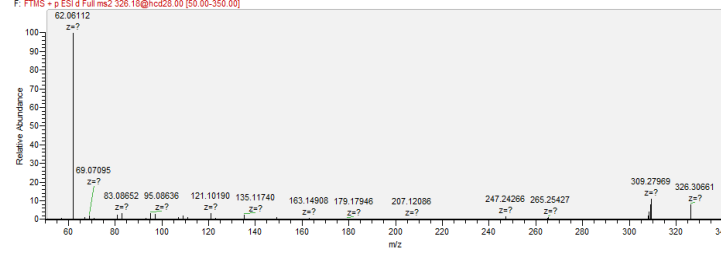
C18H31ON Sphingoid base M+H

T: FTMS + p ESI d Full ms2 278.25@hcd28.00 [50.00-300.00]



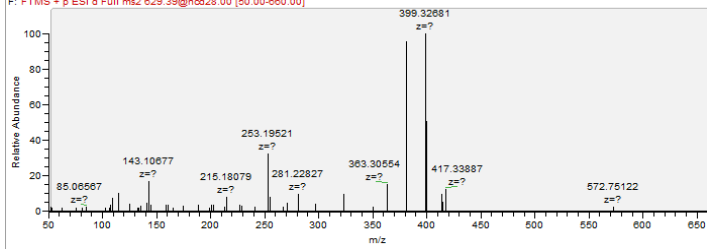
Oleoyl ethanolamide M+H

F: FTMS - p ESI d Full ms2 326.18@hcd28.00 [50.00-350.00]



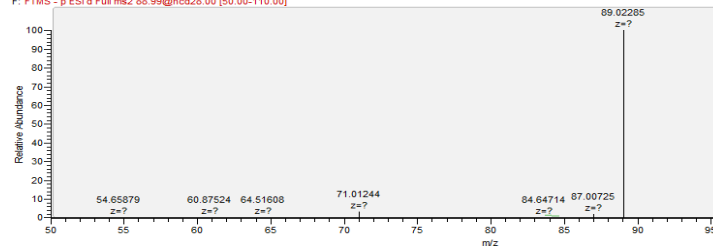
C34H52N4O7 hydroxyvitaminD3 derivative M+H

F: FTMS + p ESI d Full ms2 629.39@hcd28.00 [50.00-660.00]



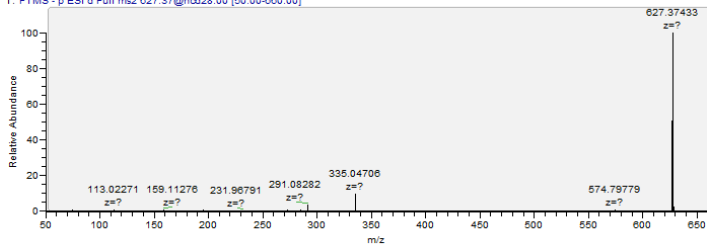
Lactate M-H

F: FTMS - p ESI d Full ms2 89.99@hcd28.00 [50.00-110.00]



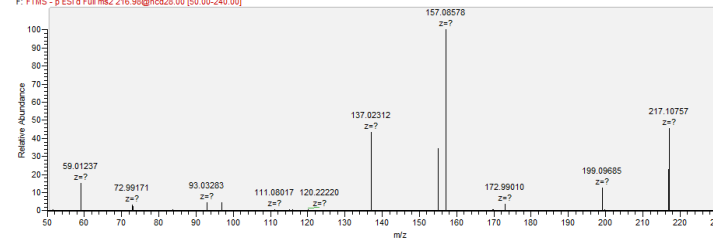
C34H52N4O7 hydroxyvitaminD3 derivative M-H

T: FTMS - p ESI d Full ms2 627.37@hcd28.00 [50.00-660.00]

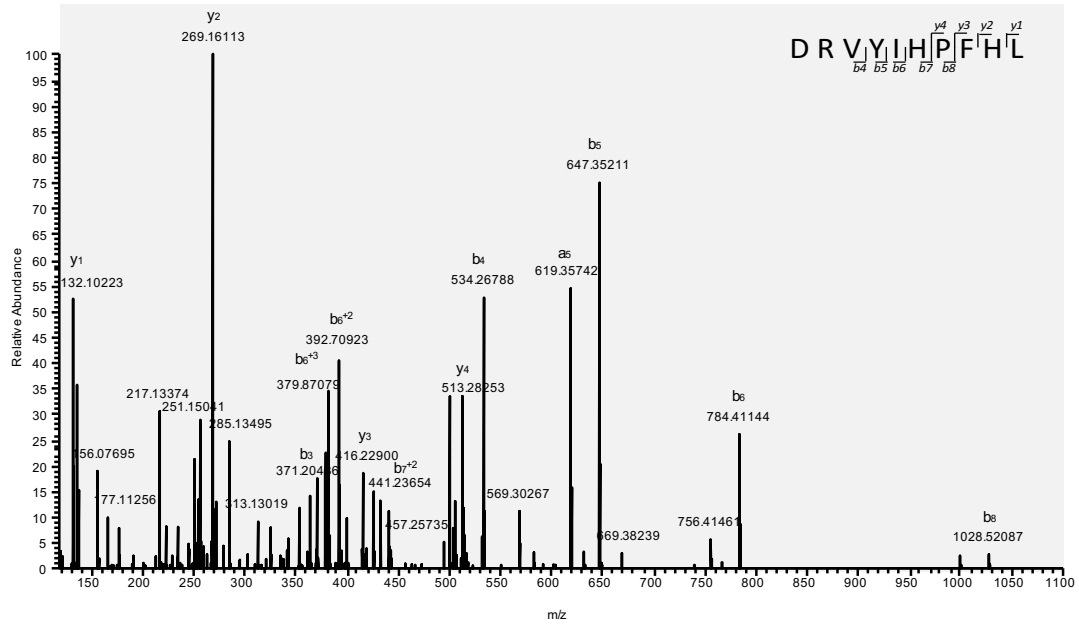


3-Hydroxysebacic acid M-H

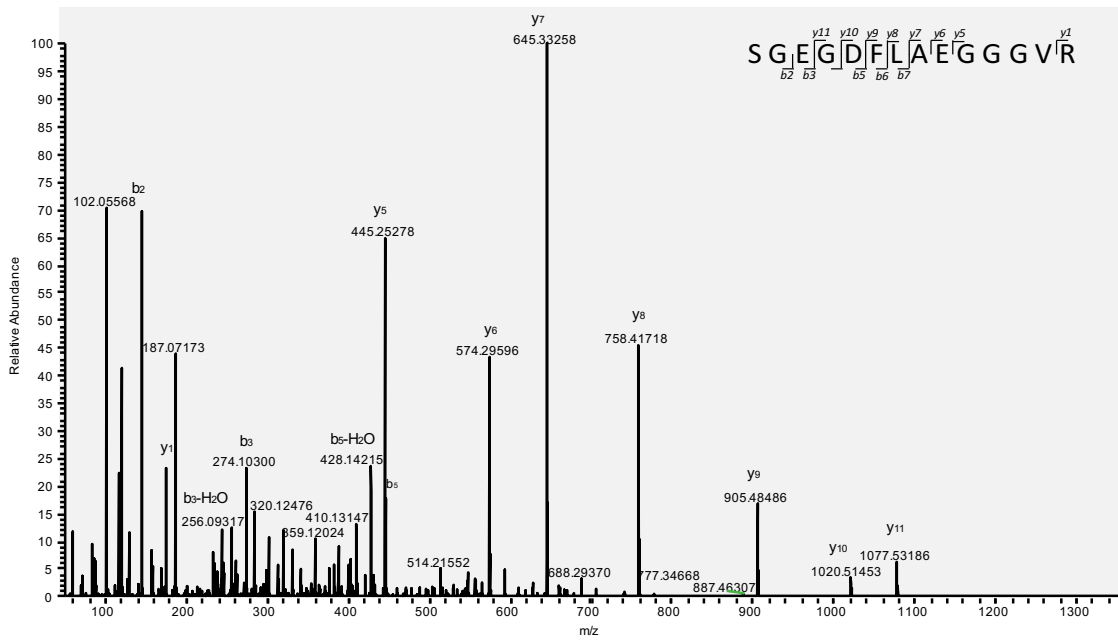
F: FTMS - p ESI d Full ms2 216.90@hcd28.00 [50.00-240.00]



Angiotensin I



desAD-FPA



5.2.5 Statistical analysis of metabolome data

Principal Component Analysis (PCA) was conducted in R using the “FactoMineR” package with pareto scaling. Components 1 and 2, representing approximately 15% and 9% of the variation respectively, could not be explained by any of the metadata collected. We therefore selected the lowest components which maximized separation between SAM and controls (components 3 and 4) for qualitative analysis of the metabolome.

Metabolites that differed between SAM and controls were determined independently of PCA analysis using unpaired Wilcoxon tests with Benjamini-Hochberg (False Discovery Rate (FDR)) corrections to account for multiple hypothesis testing²⁰. Metabolites with a corrected $P < 0.1$ and average fold change > 2 in either group (SAM or control) were selected for further investigation. Fold changes were calculated using the geometric mean to limit inflation of fold change values due to outliers. Boxplots and stripcharts were constructed in R using the ggplot2 package. Odds ratios of metabolites to identify SAM from controls were calculated from conditional logistic regressions performed on all metabolites using the glm function in R with 10 000 iterations and a binomial distribution. Receiver operating characteristic (ROC) curves and forest plots were built in R using the pROC and forestplot packages respectively. In the case of Des ADS-FPA, the ¹³C isotope was used for statistical analyses and plotting due to a large singly-charged interfering peak with mass similar to the ¹²C Des ADS-FPA. All raw data files were manually inspected to ensure the data was accurately represented by the ¹³C peak.

5.2.6 Calprotectin and lactoferrin ELISAs

Calprotectin was measured using the IDK® Calprotectin ELISA Kit along with the IDK Extract® stool extraction kit (Immundiagnostik, Bensheim, Germany). Lactoferrin was measured using the Lactoferrin Scan™ ELISA kit (Techlab®, Blacksburg, USA). All protocols were followed as per the manufacturer’s instructions.

5.2.7 Microbiome profiling

DNA was extracted from stool samples using the PowerSoil-htp 96 Well Soil DNA isolation kit from MoBio (Carlsbad, CA) according to the manufacturer's protocol, with modifications as outlined by the Earth Microbiome Project (version 4_13). Approximately 250 mg of fecal sample was used for the extractions. Samples were sequenced by amplifying the V4 hypervariable region of the 16S rRNA gene. Sample amplification for sequencing was carried out using the forward primer (ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN(8)GTGCCAGCMGCCG CGGTAA) and the reverse primer (CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNN(8)GGACTACHV GGGTWTCTAAT) where nnnn indicates four randomly incorporated nucleotides, and (8) was a sample-specific nucleotide barcode where the barcodes differed by an edit distance of at least 4. The 5' end is the adapter sequence for the Illumina MiSeq sequencer and the sequences following the barcode are complementary to the V4 rRNA gene region. Amplification was carried out in 42µL with each primer present at 3.2 pmol/µL, 20µL GoTaq hot start colorless master mix (Promega) and 2µL extracted DNA. The PCR protocol was as follows: initial activation step at 95°C for 2 minutes and 25 cycles of 1 minute 95°C, 1 minute 50°C and 1 minute 72°C.

All subsequent work was carried out at the London Regional Genomics Centre (LRGC, lrgc.ca, London, Ontario, Canada). Briefly, PCR products were quantified with a Qubit 2.0 Fluorometer and the high sensitivity dsDNA specific fluorescent probes (Life Technologies). Samples were mixed at equimolar concentrations and purified with the QIAquick PCR Purification kit (QIAGEN). Samples were paired-end sequenced on an Illumina Mi-Seq with the 600 cycle version 3 reagents with 2x220 cycles.

Resulting reads were extracted and de-multiplexed using modifications of in-house Perl and UNIX-shell scripts with operational taxonomic units (OTUs) clustered at 97% identity, similar to our reported protocol²¹. Automated taxonomic assignments were carried out by comparison to the SILVA database (<http://www.arb-silva.de/>). Supplemental Table 5-1

displays the nucleotide barcodes and their corresponding samples. To control for background contaminating sequences, a no-template control was also sequenced. Barplots were constructed with R (r-project.org) using proportional values. Rare OTUs found at less than 1% abundance in all samples were grouped in the remainder for barplots only. All other analyses were conducted with all OTUs. Principal Coordinate Analysis (PCoA) plots were constructed in Qiime²² (qiime.org) using weighted UniFrac distances.

To avoid inappropriate statistical inferences made from compositional data, centred log-ratios (clr), a method previously described by Aitchison²³, and adapted to microbiome data was used with unpaired Wilcoxon tests for comparisons of OTU level data^{24,25}. The Benjamini Hochberg (FDR) method was used to control for multiple testing with a significance threshold of 0.1. All statistical analysis, unless otherwise indicated, was carried out using R (r-project.org).

5.3 Results

5.3.1 Study population

We recruited 58 children aged 6 -59 months: 47 children with SAM (WHZ < -3 and/or MUAC < 11.5 and/or edema) and 11 well-nourished hospital controls. Demographic and clinical characteristics at recruitment are shown in Table 5-1. The distribution of ethnicity and area of residence differed significantly between cases and controls.

Table 5-1. Demographic and clinical characteristics of cases and controls.(1)

Variable	Severe acute malnutrition (N=47)	Non-malnourished controls (N=11)	P value
Male: No (%)	30 (63.8)	7 (63.6)	1.00
Age (months; median, range)	22.0 (6 to 48)	14.0 (6 to 44)	0.76
Ethnicity: No. (%)			
Hausa		9 (81.8)	
Fulani	43 (91.5)	0 (0)	
Ibo	3 (6.4)	0 (0)	0.048
Yoruba	1 (2.1)	1 (9.1)	
Bugaje	0 (0)	1 (9.1)	
	0 (0)		
Residence: No. (%)			
Urban		7 (63.6)	
Peri-urban	21 (44.7)	2 (18.2)	0.024
Rural	1 (2.1)	2 (18.2)	
	25 (53.2)		
Weight-for-length/height z score (median; IQR)	-5.08 (-10.74 to -2.32)	-0.61 (-1.64 to 0.73)	<0.001
Mid-upper arm circumference (cms: median; IQR)	10.0 (7.0 to 12.0)	14.0 (13.5 to 18.0)	<0.001
Pedal edema: No. (%)	21 (44.7)	0 (0)	-
Stool appearance (Bristol Stool Form Scale):			
2		3 (27.3)	
3	0 (0)	4 (36.4)	
4	9 (19.1)	0 (0)	0.003
5	2 (4.3)	2 (18.2)	
6	9 (19.1)	2 (18.2)	
7	18 (38.3)	0 (0)	
	9 (19.1)		
Ever received formula milk: No. (%)	23 (48.9)	2 (18.2)	0.093

(1) P values were calculated using the Wilcoxon test with the exception of ethnicity, residence, ever received formula, for which the Chi squared or Fisher's exact test was used.

Amongst the cases, 26 (55.3%) were diagnosed with kwashiorkor. Apathy was present in 34 (72.3%), anorexia in 42 (89.4%), thin hair in 42 (89.4%), glossitis in 27 (57.4%), one or more Bitot spots in 9 (19.1%), dermatitis in 22 (46.8%), stomatitis in 27 (57.4%) and 18 (38.3%) had oral ulceration. Abdominal distension was present in 11 (23.4%), hepatomegaly in 28 (59.6%) and splenomegaly in 26 (55.3%). Loose or watery stools (Bristol Stool Form Scale 5-7) occurred in 36 (76.6%) of the SAM cases compared with 4 (36.4%) controls (Table 1). None of the children had bloody stools.

5.3.2 The fecal microbiota of Nigerian children

To determine if the composition of the gut microbiota differed between Nigerian children with SAM and controls, we profiled the microbiota of stool samples by amplifying the V4 region of the 16S rRNA gene. Two samples did not have enough material for DNA extraction, leaving 45 SAM and 11 controls remaining. The fecal microbiota of Nigerian children was dominated by *Enterobacteriaceae*, *Bifidobacterium*, *Enterococcus*, *Pediococcus*, *Lactobacillus* and *Streptococcus* species (Figure 5-2A). There were no differences in alpha or beta diversity between SAM and controls (Figure 5-2B,C), nor were there any individual taxa that differed significantly between groups (Wilcoxon test, FDR corrected $P > 0.1$, Supplemental Table 5-3). Although there was a trend for lower bacterial diversity in children with SAM compared to controls, the difference was not significant, even when age was taken into account (Figure 5-2C, ANCOVA $P > 0.05$).C, $A > 0.05$).

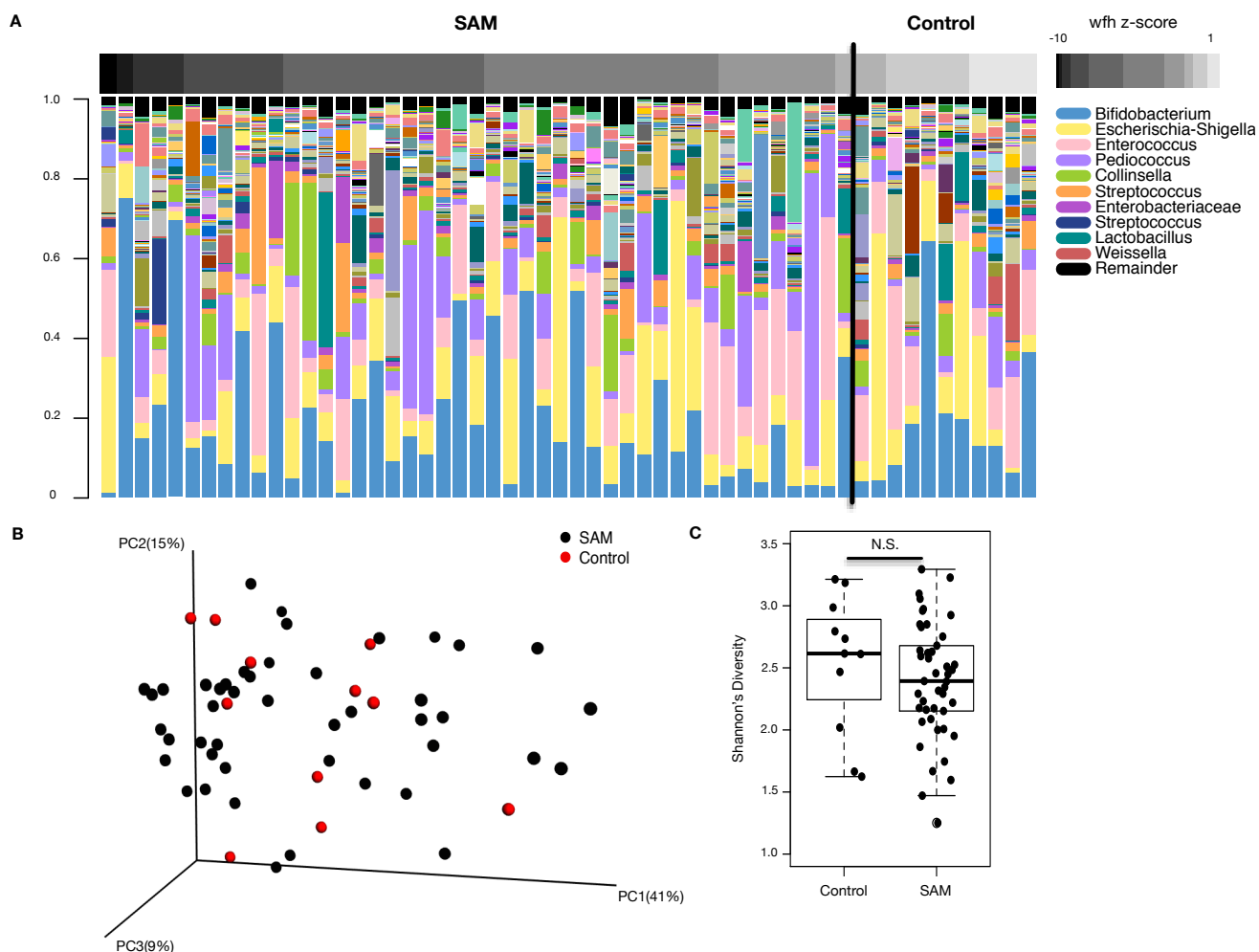


Figure 5-2. The stool microbiota does not discriminate SAM from controls.

(A). Stool microbiota profiled using the V4 region of the 16S rRNA gene. Each bar in the barplot represents a single sample from a single child and each color a different bacterial taxa (OTU). OTUs present at less than one percent in every sample were placed in the remainder displayed in black. Samples are ordered by their WHZ displayed in the heatmap above. (B). Principal Coordinate Analysis (PCoA) plot of microbiota profiles built from weighted UniFrac distances. Each point represents a single sample from a single child. Positions of points display dissimilarities in the microbiota, with points further from one another being more dissimilar. (C). Shannon's diversity of the stool microbiota in children with SAM compared with controls. The boxes represent the 25th and 75th quartiles, and the line displays the median value within each group. Points extending beyond the lines are outliers defined as values greater or less than 1.5 times the interquartile range.

5.3.3 Fecal metabolome and inflammatory markers do not distinguish SAM from controls

To obtain a global view of nutrient deficiency in SAM and gain insight into pathophysiology, we performed a comprehensive untargeted analysis of metabolites in stool from all 58 children using two different methods; gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS). Surprisingly, both methods yielded no metabolites significantly affected by SAM (Wilcoxon test, FDR corrected $P > 0.1$, Figure 5-3, Supplemental Table 5-4). There were also no significant differences in the inflammatory proteins calprotectin or lactoferrin, as measured by ELISA (Figure 5-4).

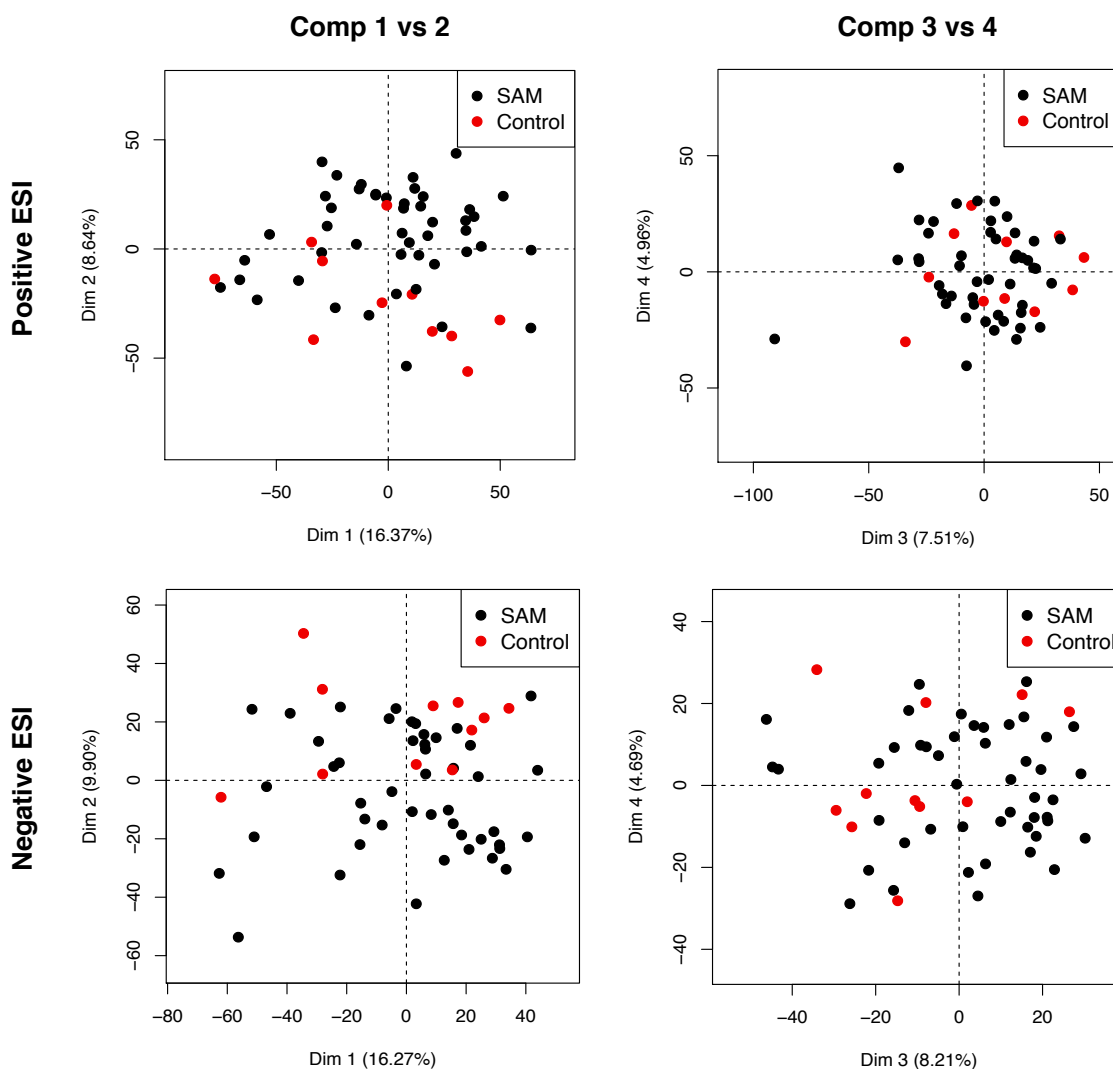


Figure 5-3. Principal component analysis (PCA) scoreplots of stool metabolome from children with SAM and controls as determined by LC-MS using positive (top) and negative (bottom) electrospray ionization (ESI).

Components 1 and 2 are shown on the left with components 3 and 4 on the right. Each point represents a single sample from a single child. Positions of points display dissimilarities in the metabolome, with points furthest from one another being most dissimilar.

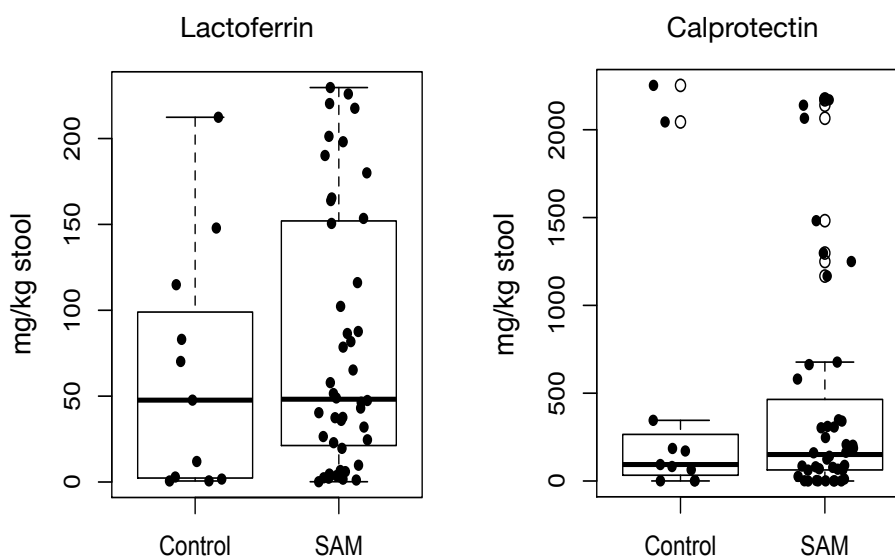


Figure 5-4. Concentration of inflammatory proteins in stool of children with SAM compared to controls as measured by ELISA.

Each point represents a single sample from a single child. The boxes represent the 25th and 75th quartiles, and the line displays the median value within each group. Points extending beyond the lines are outliers defined as values greater or less than 1.5 times the interquartile range.

5.3.4 Children with SAM have a distinct plasma metabolome

In contrast to stool results, the abundance of approximately 15% of LC-MS features detected in plasma were significantly altered by SAM, where a “feature” represents a unique m/z and retention time value (Wilcoxon test, FDR corrected $P < 0.1$, Supplemental Table 5-5). Separation between groups was observed along the 3rd and 4th component of principal component analysis (PCA) plots built from LC-MS data (Figure 5-5), demonstrating the plasma metabolome of children with SAM is distinct from non-malnourished children. The vast majority of differential features were detected exclusively by LC-MS, with only three metabolites differing significantly by GC-MS

(valine, leucine and aspartic acid). There were no metabolites that differed between children with non-edematous malnutrition (marasmus, N = 21) or edematous malnutrition (kwashiorkor, N = 26), and no significant effect of sex or age was observed (data not shown).

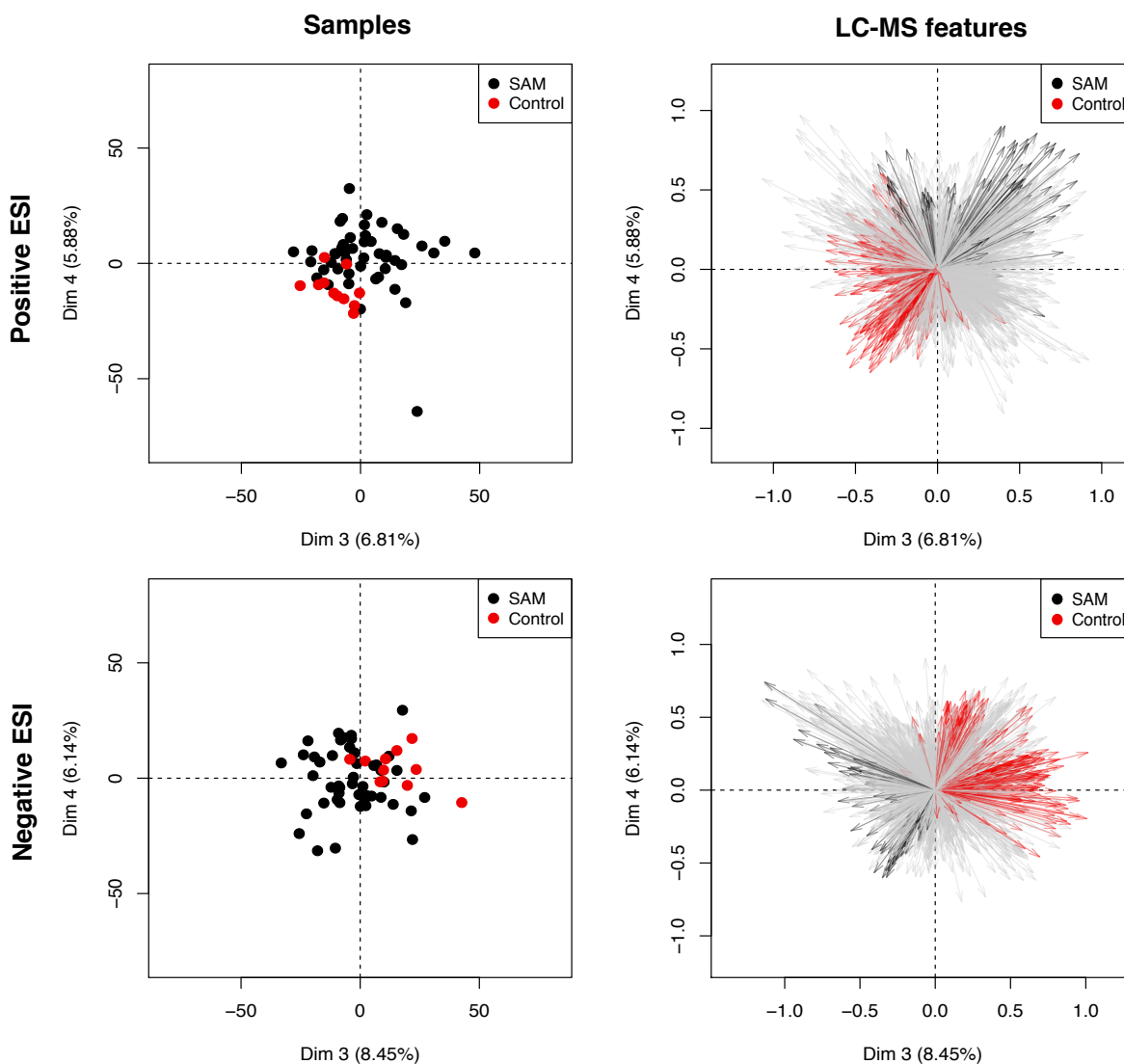


Figure 5-5. Principal component analysis (PCA) plots of plasma metabolome from children with SAM and controls as determined by LC-MS using positive (top) or negative (bottom) electrospray ionization (ESI).

Only components 3 and 4 are shown. The plots on the left display individual samples (scores). Each point represents a single sample from a single child. Positions of points display dissimilarities in the metabolome, with points furthest from one another being most dissimilar. Plots on the right display individual LC-MS features (loadings). Each ray represents a single LC-MS feature, with those significantly elevated in SAM or

Controls highlighted in black or red respectively (Wilcoxon test, FDR corrected $P < 0.1$).

5.3.5 Oxylipins, phospholipids and amino acids are depleted in the plasma of children with SAM

Significant features in LC-MS data with greater than 2-fold change between SAM and controls were selected for further identification. As expected, free amino acids and dipeptides were lower in the plasma of children with SAM including glutamine, arginine, tyrosine, leucine, valine, and the tryptophan metabolite kynurenine (Figure 5-6A). A number of ether-linked single chain phospholipids belonging to the phosphatidylcholine (PC) and phosphatidylethanolamine (PE) families were also significantly decreased (Figure 5-6B). In addition, there was a marked reduction in a number of oxylipins belonging to the eicosanoid and docosanoid family. In depth investigation of fragmentation patterns of these lipids revealed that each feature (represented as a single boxplot pair in Figure 5-6C) contained at least two different oxylipin species, differing only in the location of the hydroxyl group(s) (Figure 5-7). The precise species contributing to the differences between SAM and controls could therefore not be determined. Arachidonic acid (AA), a precursor to the eicosanoids was also significantly lower in children with SAM. While median abundance of docosahexaenoic acid (DHA, the precursor to the docosanoids), was lower in children with SAM, the difference was not significant (FDR corrected $P = 0.2$) (Figure 5-6).

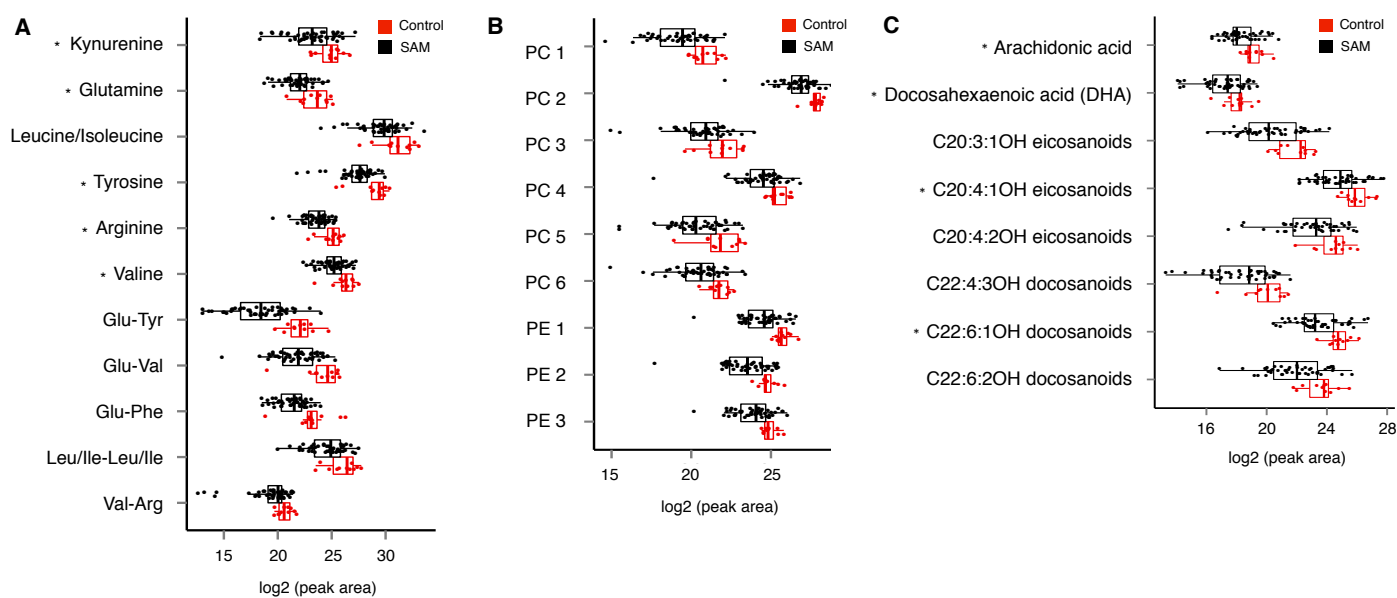


Figure 5-6. Free amino acids/dipeptides (A), phospholipids (B), and oxylipins (C) significantly decreased in children with SAM compared to controls.

> 2 fold change, Wilcoxon test, FDR corrected $P < 0.1$, with the exception of DHA which was not significant. Panels (A) and (B) were detected by positive ESI LC-MS and (C) by negative ESI LC-MS. Each point represents a single sample from a single child. The boxes represent the 25th and 75th quartiles, and the line displays the median value within each group. Points extending beyond the lines are outliers defined as values greater or less than 1.5 times the interquartile range. PC: Phosphatidylcholine, PE: Phosphatidylethanolamine. See Supplemental Table 5-5 for phospholipid identities. (*) Metabolite ID confirmed by authentic standards.

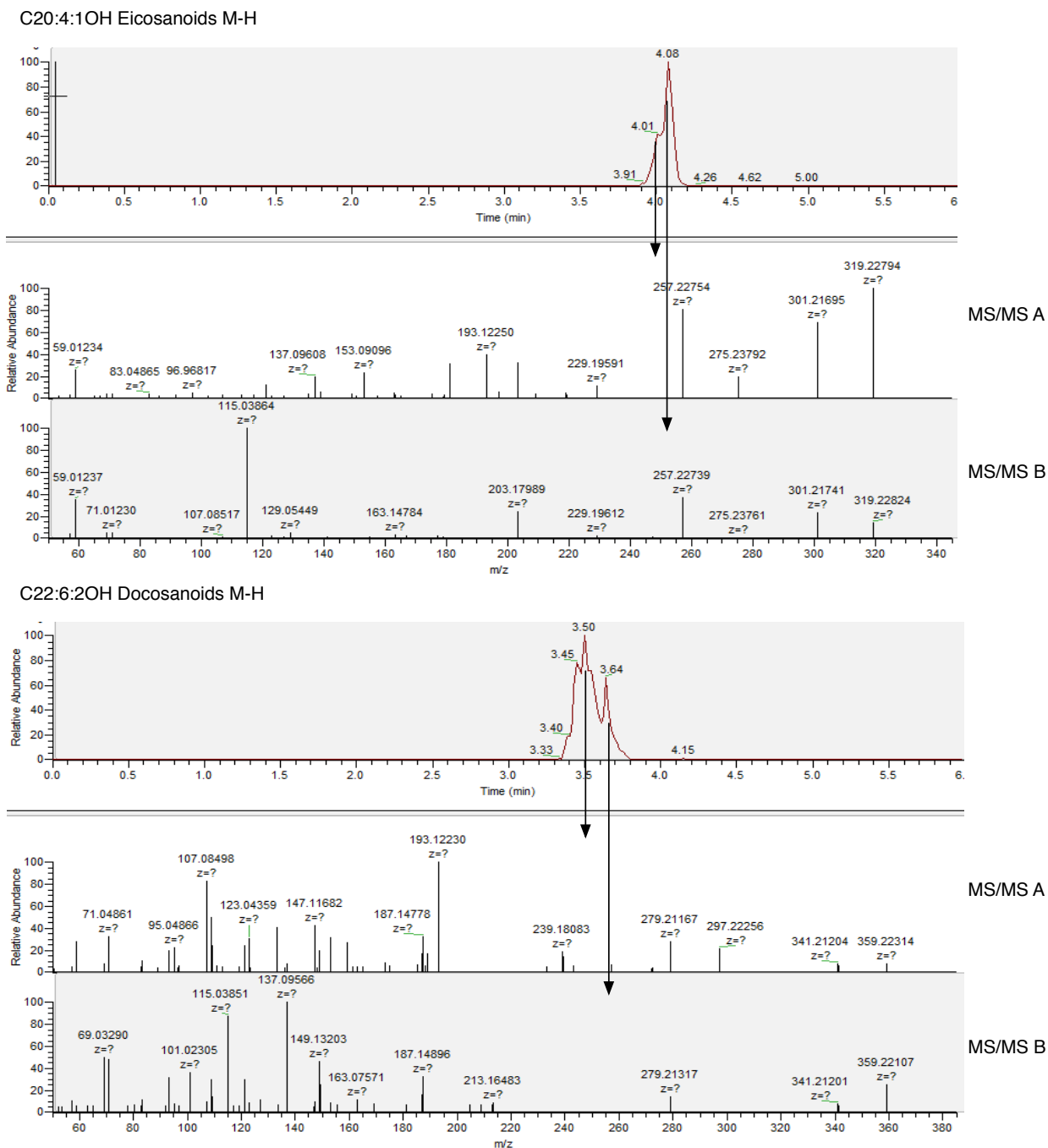


Figure 5-7. LC-MS/MS spectra examples for oxylpin features containing two or more species

5.3.6 Metabolites elevated in plasma of children with SAM

Unlike metabolites negatively associated with SAM, the classes of metabolites elevated in malnourished children were more diverse, and included sugars, peptides, lipids, short chain fatty acids and porphyrins, among others. Of interest, we detected a 2C6-disaccharide by LC-MS in both positive and negative ionization modes that was significantly elevated in children with SAM (Wilcoxon test, FDR corrected $P < 0.1$, Figure 5-9). Authentic standards of different 2C6-disaccharides confirmed that our LC-MS measurements represented the total 2C6-disaccharides in plasma. Analysis of standards by GC-MS, which is able to resolve disaccharides, and LC-MS/MS determined that sucrose was the most abundant disaccharide in plasma, with smaller amounts of lactose and maltose also present.

Two forms of truncated fibrinopeptide A (FPA) missing 2, or 3 N-terminal amino acids were also detected in significantly greater amounts in children with SAM. Specifically, des-AD (lacking N-terminal alanine and aspartate) and des-ADS FPA (lacking N-terminal alanine, aspartate and serine) were both increased approximately 6-fold (Figure 5-9). Intact FPA was outside the mass range of our initial analysis, and therefore a second analysis was conducted with a larger mass range to include the intact FPA peptide. Surprisingly, the abundance of intact FPA was not significantly different between SAM and controls (Figure 5-8), indicating that the truncated forms are not simply a degradation product of increased intact FPA, but have some other unknown origin.

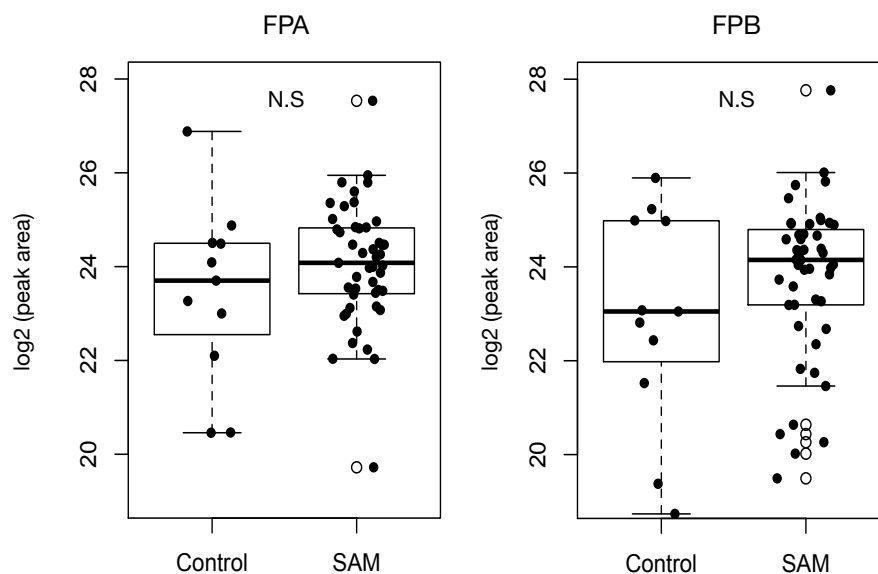


Figure 5-8. Comparison of intact fibrinopeptide A (FPA) and B (FPB) in plasma of children with SAM compared to controls measured using positive ESI LC-MS.

Each point represents a single sample from a single child. The boxes represent the 25th and 75th quartiles, and the line displays the median value within each group. Points extending beyond the lines are outliers defined as values greater or less than 1.5 times the interquartile range. $P > 0.1$ Wilcoxon test, FDR corrected.

Other metabolites of interest elevated in SAM included several acylcarnitines, the peptidehormone angiotensin I (Ang I), heme, lactate, oleoyl ethanolamide, 2,4 and 3,4-dihydroxybutyrate, an uncharacterized sphingoid base, a hydroxyvitamin D3 derivative and several other unknown compounds (Figure 5-9, Supplemental Table 5-5).

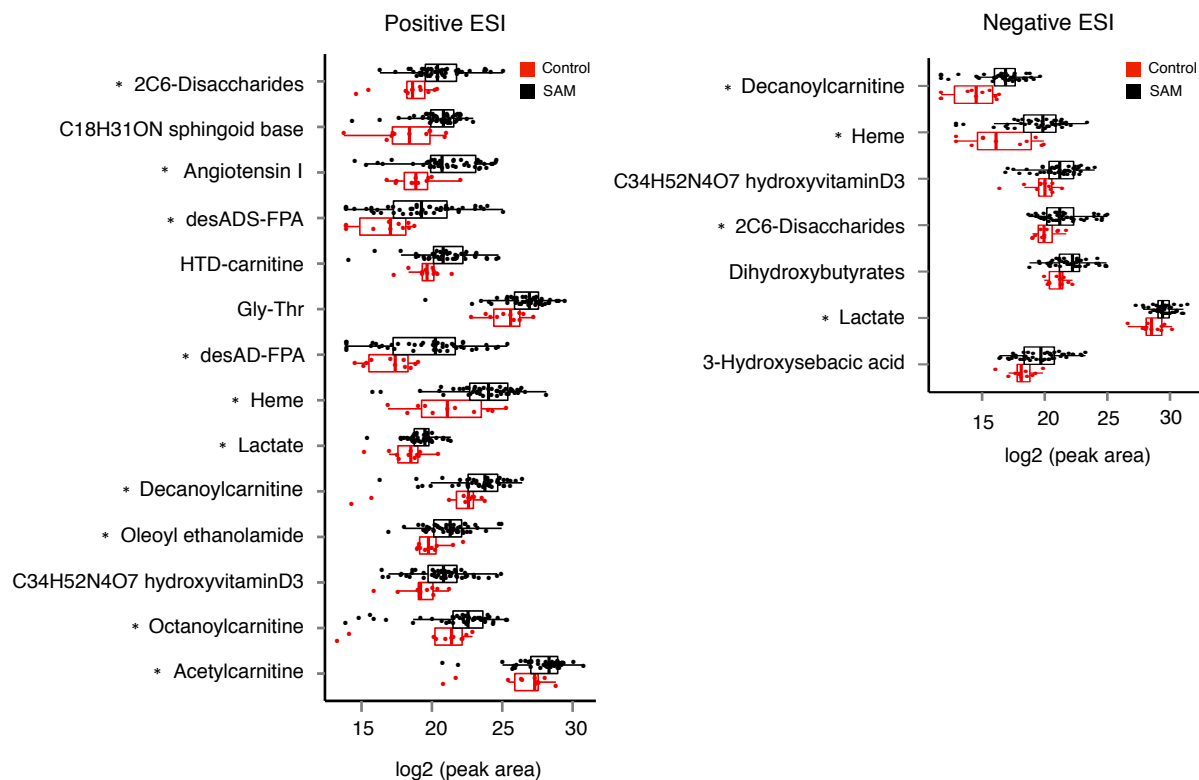


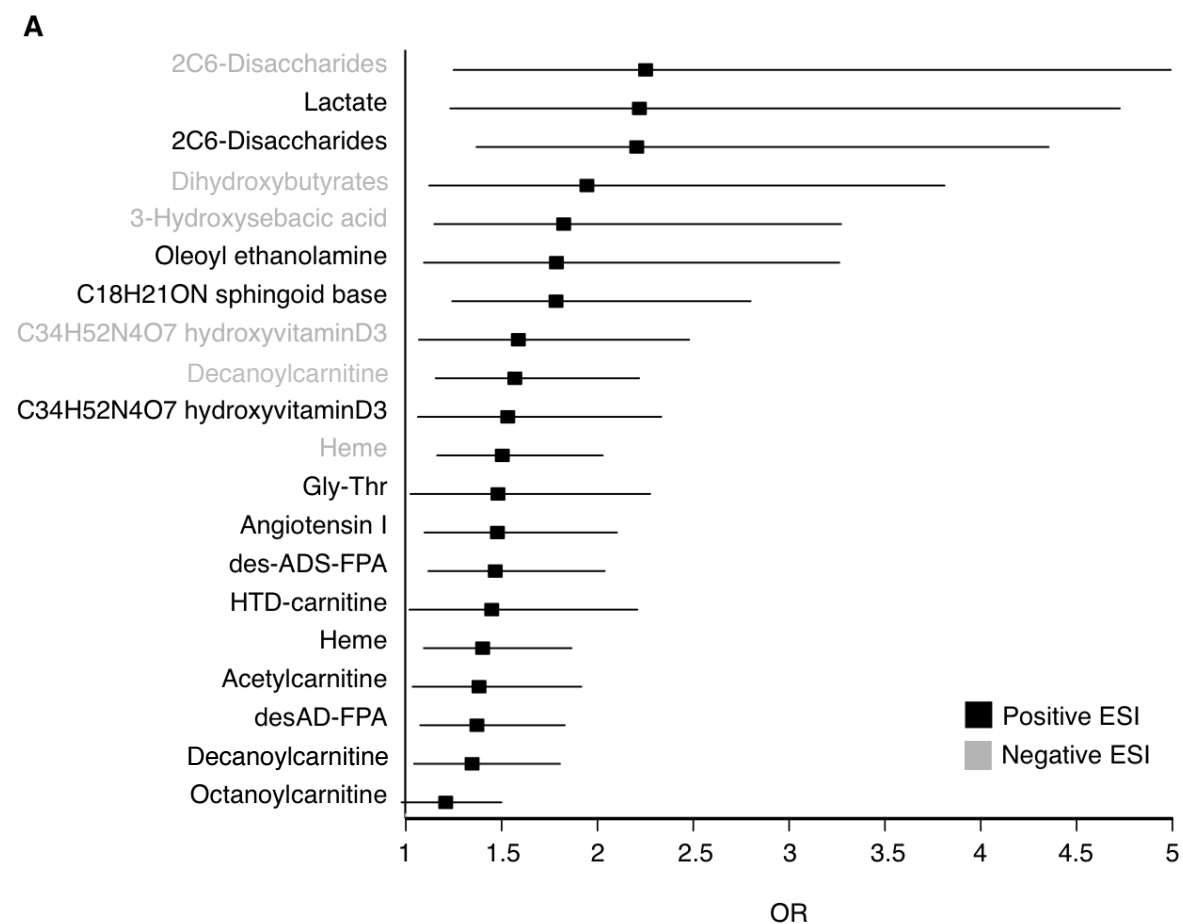
Figure 5-9. Metabolites significantly elevated at least 2-fold in the plasma of children with SAM compared to controls (Wilcoxon test, FDR corrected $P < 0.1$).

(Left) Metabolites detected using positive ESI LC-MS. (Right) Metabolites detected using negative ESI LC-MS. Each point represents a single sample from a single child. The boxes represent the 25th and 75th quartiles, and the line displays the median value within each group. Points extending beyond the lines are outliers defined as values greater or less than 1.5 times the interquartile range. FPA: Fibrinopeptide A, HTD: 3-Hydroxy-cis-5-tetradecenoyl. (*) Metabolite ID confirmed by authentic standards.

5.3.7 Assessment of biomarkers to discriminate SAM from controls

To measure the strength of the association between biomarkers and SAM, the odds ratios (OR) was calculated based on conditional logistic regressions of all metabolites elevated in children with SAM (Supplemental Table 5-6). 2C6-disaccharides and lactate were among the metabolites with the highest OR, ranging from 2.25-2.20 respectively (Figure 5-10A). ROC curves identified decanoylcarnitine, 2C6-disaccharides, an uncharacterized sphingoid base, angiotensin I, and heme as the metabolites that maximized the sensitivity and specificity for SAM, as shown by the area under the curve (AUC), which ranged

from 0.83- 0.81 for these compounds (Figure 5-10B, Supplemental Table 5-6)



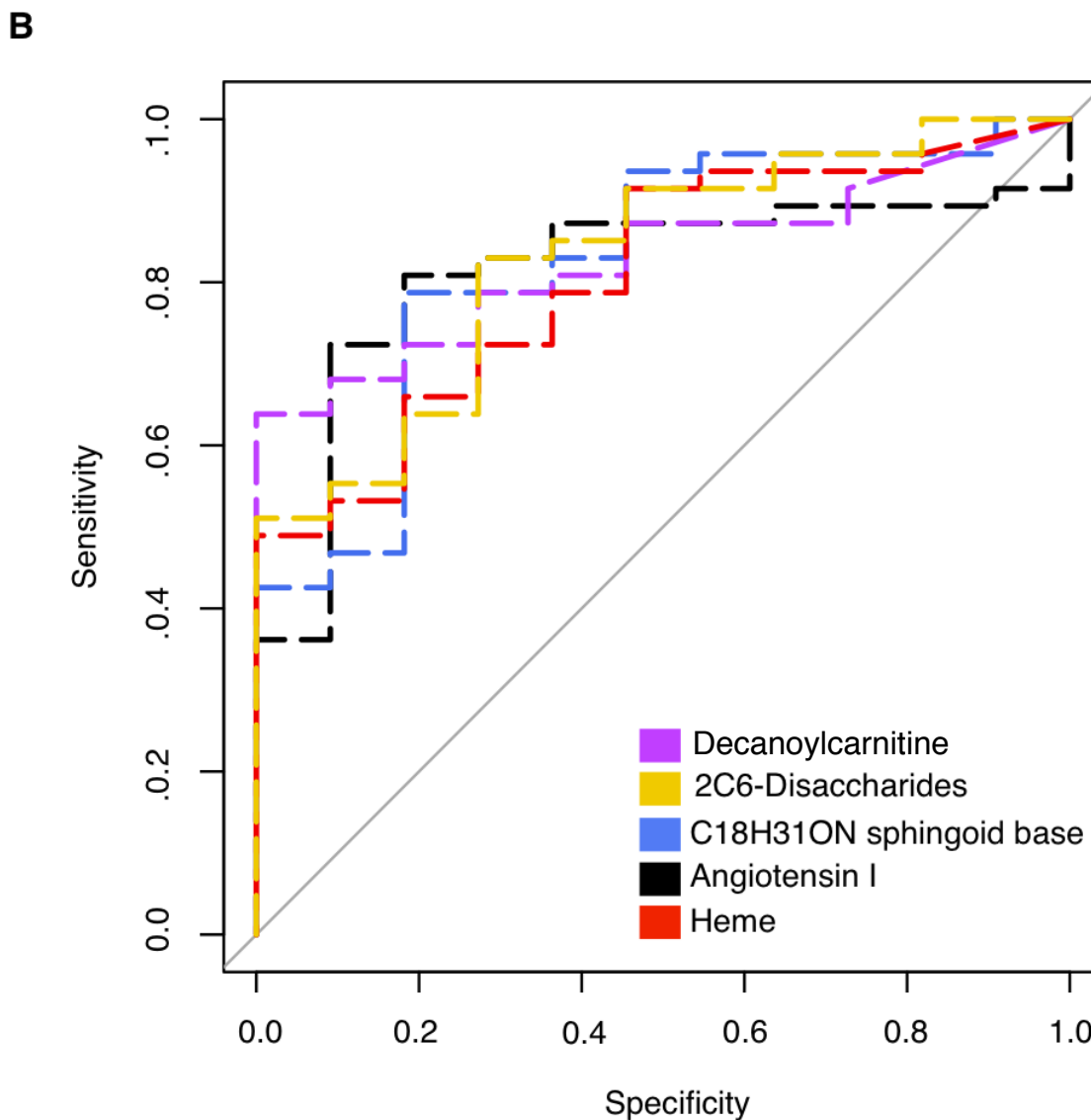


Figure 5-10. Evaluation of biomarkers to identify SAM from controls.

(A) Odds Ratios as determined by conditional logistic regressions of all validated metabolites positively associated with SAM. Bars represent 95% confidence intervals. (B) Receiver Operator Characteristics (ROC) curves. Metabolites with the highest area under the curve (AUC) are shown.

5.4 Discussion

We hereby report the first LC-MS-based untargeted metabolomic study of stool and plasma from children with SAM. In contrast to the stool microbiota and metabolome, which did not discriminate SAM from controls, approximately 15% of the plasma metabolome,

equating to 585 features, were significantly altered by malnutrition. The explanation for the lack of differences in stool are likely multifaceted, but may include the large effect of diet and inter-individual variation which cannot be accounted for with small cross-sectional studies. It is also worth noting that feces represent the net result of nutrient consumption, digestion and absorption. Malabsorption has been widely reported in SAM, and therefore it is possible that these children both consume and absorb fewer nutrients compared with non-malnourished controls, resulting in little net difference in stool nutrient composition.

The stool microbiota data mirrored the metabolome results as neither the composition or diversity were significantly altered by malnutrition, even when children were matched for age. This is in contrast to previous studies that have observed differences in the relative abundance of specific taxa and diversity for age in twin pairs and unrelated controls^{6,7}. The relatively small number of controls in the study limited our ability to model the relationship between microbial diversity and age, and therefore it is possible that a larger sample size may have revealed significant differences. Our cohort also tended to be older than the children included in the study by Subramanian et al⁷, which could partially explain the lack of differences in the microbiota. As the microbiota has been shown to stabilize by around two years of age⁷, and a large proportion (41%) of the children in our study were two years or older, it is possible that the window where microbiota stunting is most apparent was not captured by our study. Additionally, none of the plasma metabolites affected by SAM were of clear microbial origin, providing further evidence that the microbiota may not be playing a major role in this particular cohort.

Among the plasma metabolites discriminating SAM from controls, the total 2C6-disaccharides had one of the strongest positive associations, as determined by ROC and OR analyses, and were detected in both positive and negative ionization modes. Regardless of their structure and content in the diet, disaccharides are not readily absorbed, and therefore must pass through the intercellular space of the intestinal mucosa to reach systemic circulation. For this reason, sucrose, cellobiose, lactulose and other disaccharides have been used as indicators of intestinal permeability^{26,27}. Malabsorption and increased intestinal permeability are associated with SAM and EED, as measured by dual sugar

permeability tests such as the lactulose/mannitol test⁹. Despite these tests being fairly reliable, the requirement for fasting prior to administration raises ethical concerns for children with SAM, and urine must be collected over several hours for accurate results. A non-administered permeability test, for which the total disaccharides are an attractive candidate, would therefore be highly valuable.

Although elevated disaccharides are suggestive of enteropathy in the SAM group, we did not observe any difference in the inflammatory markers calprotectin or lactoferrin in stool. These proteins are used as indicators of inflammation in other pathologies such as inflammatory bowel disease²⁸, but have not been evaluated as biomarkers of enteropathy in SAM. The weak correlation between lactoferrin and calprotectin (Spearman's $R = 0.33$, $P = 0.01$) limits our ability to make conclusions as to the intestinal inflammatory status of these children. Future studies evaluating the association of non-administered disaccharides with enteropathy are warranted to determine if they might also be useful in identifying EED.

The majority of metabolites discriminating SAM from controls in plasma were nutrients depleted due to malnutrition. As expected, a number of amino acids/dipeptides were reduced, indicative of protein deficiency. Oxylipins were also significantly decreased, and to the best of our knowledge, have not been examined in children with SAM previously. Importantly, each oxylipin feature contained at least two different species, and therefore the number and diversity of oxylipins altered by SAM may be vastly under-represented by our study. Oxylipins are bioactive lipids formed by oxidation of long chain polyunsaturated fatty acids (LCPUFA), with the most well studied being the AA-derived eicosanoids²⁹. These bioactive lipids perform a wide array of functions, including tissue repair, blood clotting, and regulation of the immune system. Oxylipins within the same family can have similar or opposing effects (i.e. pro- or anti-inflammatory), and thus we cannot determine the precise biological consequence of reduced oxylipins during SAM²⁹. However, as children with SAM exhibit an impaired immune response to a variety of pathogens³⁰, a lack of oxylipin mediators may be a contributing factor and warrants further investigation.

Previous studies have confirmed that children with SAM are deficient in LCPUFA^{31,32}, including the oxylipin precursors AA and DHA. As a large proportion of LCPUFA are stored as acyl-linked phospholipids, which could not be detectable by our method, a comparison of the total LCPUFA in our cohort was not possible. Nevertheless, the significant reduction in free AA, non-significant trend for lower free DHA, and significant decreases in multiple phospholipids and oxylipins are suggestive of LCPUFA deficiency in our cohort.

Despite previous evidence of LCPUFA deficiency in SAM, the levels of LCPUFA in ready-to-use therapeutic food (RUTF) are low³³. Recently, a RUTF formulation supplemented with fish oil has been developed, which translated into increased LCPUFA in recipient children³³. More studies are required to determine if the incorporation of fish oil into refeeding programs restores oxylipin levels as well and whether this leads to any clinical benefit.

Apart from dietary deficiencies, other factors may also influence LCPUFA and oxylipin levels. Children with SAM have decreased desaturase activity, resulting in decreased synthesis of AA and DHA from their precursors^{34,35}. Additionally, a large proportion of dietary fatty acids are lost to beta-oxidation, a process that is induced during starvation^{36,37}. A number of even-chain acylcarnitines, the bi-products of beta oxidation, were significantly elevated in children with SAM in our study. This suggests beta-oxidation of lipids may contribute in part to LCPUFA deficiency in SAM. Bartz et al⁴ found that even-chain acylcarnitines, including the C10, C8 and C2 species identified in our study, decreased significantly upon nutritional intervention in children with SAM from Uganda. Acylcarnitines may therefore be useful not only as biomarkers of a malnourished state, but also as measures of treatment efficacy.

The truncated FPAs were also biomarkers of interest as they are unique peptides of unknown origin and were elevated approximately 6 fold in SAM. The des-ADS form in particular has not been reported previously. FPA is a 16 amino acid peptide produced upon cleavage of the fibrinogen alpha chain by thrombin during the coagulation cascade. Thus,

it is an indicator of thrombosis, and is elevated in plasma during a number of inflammatory conditions including Crohn's disease, gastric cancer, and coronary thrombosis³⁸⁻⁴⁰. As the intact peptide was not elevated in SAM, these fragments are not simply a result of increased inflammation and thrombosis, but have another unknown origin. One hypothesis is that they may arise from increased protease activity as proposed by Zhang et al⁴¹ who reported elevated des-A FPA in gastric cancer with lymph node metastasis. Protease activity is increased during starvation to supply the body with additional amino acids⁴², and therefore it is possible that these truncated forms are a consequence of this up-regulation.

Angiotensin I (Ang I) was another peptide elevated in the SAM group. Ang I is the precursor to Ang II, a hormone directly involved in vasoconstriction and the renin-angiotensin system (RAS). Ang II was also detected by our analysis and tended to be higher in children with SAM (average 4-fold increase), but did not reach statistical significance (FDR corrected $P = 0.15$). Children with SAM have higher blood pressure compared to controls⁴³. In addition, rats fed low protein diets exhibit increased expression of multiple components of the RAS system, including renin, angiotensin I converting enzyme (ACE) and angiotensin II type 1 receptors^{44,45}. Together these results suggest increased expression of both the angiotensin hormones themselves and their receptors may contribute to vasoconstriction and increased blood pressure in SAM.

Lactate was also elevated in the plasma of malnourished children, and was one the metabolites with the highest OR for SAM. Lactate is an endproduct of cellular respiration under anaerobic conditions, and is elevated in blood due to a number of etiologies, including anemia, sepsis, trauma, and malignancy⁴⁶. In the context of malnutrition, elevated lactate may result from insufficient oxygen supply due to the anemia or increased infections associated with SAM. Interestingly, several studies have shown blood lactate to be a good predictor of mortality due to a variety of illnesses⁴⁷⁻⁴⁹, including a study of Tanzanian children with any febrile illness⁵⁰. Further research should evaluate lactate as a predictor of mortality in SAM.

Finally, we observed a significant increase in free iron (III) heme in the plasma of malnourished children. Heme is normally bound to heme proteins, but is released under conditions of oxidative stress and as a result of hemolysis. Free heme causes tissue damage, systemic inflammation^{51,52}, and exacerbates sepsis and malaria in animal models^{53,54}. In humans, malaria severity correlates with the levels of free heme, and sepsis-related mortality is associated with decreased levels of the heme binding protein hemopexin^{54,55}. These diseases frequently affect children with SAM in the developing world, and malaria in particular is significantly associated with malnutrition^{56,57}. Elevated heme during SAM may contribute to increased severity of these infections, and therefore treatments targeting free heme may be beneficial.

Our study has several limitations. Firstly, reliable information regarding other infections and pathologies was not available due to the limited local diagnostic facilities, so their possible impact on the microbiota and metabolome remains unknown. Children with malnutrition were also more likely to originate from rural settings, and therefore we cannot distinguish the role of environment from that of malnutrition. Also, although we identified clear differences in the plasma metabolome of children according to nutritional status, this was not the only source of variation. We speculate that other unaccounted for sources of variation may include time of sample collection as it relates to circadian rhythms⁵⁸ and feeding time, diet and other environmental exposures, and/or the proximity of the plasma sampling to the plasma-red blood cell interface. The samples in this study were taken at random time points, with no requirements for fasting or feeding prior to collection. This may explain why we did not identify some of the classical signs of fasting, such as ketones elevated in the SAM group. However, our pragmatic collection method emphasized metabolites robust to external factors such as time of day and feeding. This strengthens the applicability of our findings to a clinical setting, where controlling for such variables may not be feasible.

In conclusion, we have demonstrated that the plasma metabolome discriminates children with SAM from controls and identified a number of novel biomarkers of malnutrition, providing new insight into disease mechanisms and management. Future studies should

monitor these metabolites longitudinally during intervention to identify those most correlated with mortality and/or recovery. Validation of such biomarkers may enable better identification of children at highest risk of poor outcomes, the interventions they need most, and could provide a quantitative measure of treatment efficacy for a leading cause of childhood mortality worldwide.

5.5 References

- (1) Kar, B. R.; Rao, S. L.; Chandramouli, B. A. Cognitive Development in Children with Chronic Protein Energy Malnutrition. *Behav Brain Funct* **2008**, *4*, 31.
- (2) Who. WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children. *WHO Libr.* **2009**, 1–12.
- (3) Ghisolfi, J.; Charlet, P.; Ser, N.; Salvayre, R.; Thouvenot, J. P.; Duole, C. Plasma Free Amino Acids in Normal Children and in Patients with Proteinocaloric Malnutrition: Fasting and Infection. *Pediatr. Res.* **1978**, *12* (9), 912–917.
- (4) Bartz, S.; Mody, A.; Hornik, C.; Bain, J.; Muehlbauer, M.; Kiyimba, T.; Kiboneka, E.; Stevens, R.; Bartlett, J.; St Peter, J. V.; Newgard, C. B.; Freemark, M. Severe Acute Malnutrition in Childhood: Hormonal and Metabolic Status at Presentation, Response to Treatment, and Predictors of Mortality. *J. Clin. Endocrinol. Metab.* **2014**, *99* (6), 2128–2137.
- (5) Jiang, P.; Stanstrup, J.; Thymann, T.; Sangild, P. T.; Dragsted, L. O. Progressive Changes in the Plasma Metabolome during Malnutrition in Juvenile Pigs. *J. Proteome Res.* **2015**.
- (6) Smith, M. I.; Yatsunenko, T.; Manary, M. J.; Trehan, I.; Mkakosya, R.; Cheng, J.; Kau, A. L.; Rich, S. S.; Concannon, P.; Mychaleckyj, J. C.; Liu, J.; Houpt, E.; Li, J. V.; Holmes, E.; Nicholson, J.; Knights, D.; Ursell, L. K.; Knight, R.; Gordon, J. I. Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. *Science* **2013**, *339* (6119), 548–554.
- (7) Subramanian, S.; Huq, S.; Yatsunenko, T.; Haque, R.; Mahfuz, M.; Alam, M. a;

- Benezra, A.; DeStefano, J.; Meier, M. F.; Muegge, B. D.; Barratt, M. J.; VanArendonk, L. G.; Zhang, Q.; Province, M. a; Petri, W. a; Ahmed, T.; Gordon, J. I. Persistent Gut Microbiota Immaturity in Malnourished Bangladeshi Children. *Nature* **2014**, *509* (7505), 417–421.
- (8) Brown, E. M.; Wlodarska, M.; Willing, B. P.; Vonaesch, P.; Han, J.; Reynolds, L. A.; Arrieta, M.-C.; Uhrig, M.; Scholz, R.; Partida, O.; Borchers, C. H.; Sansonetti, P. J.; Finlay, B. B. Diet and Specific Microbial Exposure Trigger Features of Environmental Enteropathy in a Novel Murine Model. *Nat. Commun.* **2015**, *6*, 7806.
- (9) Denno, D. M.; VanBuskirk, K.; Nelson, Z. C.; Musser, C. a; Hay Burgess, D. C.; Tarr, P. I. Use of the Lactulose to Mannitol Ratio to Evaluate Childhood Environmental Enteric Dysfunction: A Systematic Review. *Clin. Infect. Dis.* **2014**, *59 Suppl 4* (Suppl 4), S213–S219.
- (10) Hashimoto, T.; Perlot, T.; Rehman, A.; Trichereau, J.; Ishiguro, H.; Paolino, M.; Sigl, V.; Hanada, T.; Hanada, R.; Lipinski, S.; Wild, B.; Camargo, S. M. R.; Singer, D.; Richter, A.; Kuba, K.; Fukamizu, A.; Schreiber, S.; Clevers, H.; Verrey, F.; Rosenstiel, P.; Penninger, J. M. ACE2 Links Amino Acid Malnutrition to Microbial Ecology and Intestinal Inflammation. *Nature* **2012**, *487*, 477–483.
- (11) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060–1083.
- (12) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (8), 770–781.
- (13) Styczynski, M. P.; Moxley, J. F.; Tong, L. V; Walther, J. L.; Jensen, K. L.; Stephanopoulos, G. N. Systematic Identification of Conserved Metabolites in

- GC/MS Data for Metabolomics and Biomarker Discovery. *Anal. Chem.* **2007**, *79* (3), 966–973.
- (14) Timotej, V. A Comparison of Parameters below the Limit of Detection in Geochemical Analyses by Substitution Methods. *RMZ Mater. Geoenvironment* **2011**, *58* (4), 393–404.
- (15) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536.
- (16) Patti, G. J.; Tautenhahn, R.; Siuzdak, G. Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments. *Nat. Protoc.* **2012**, *7* (3), 508–516.
- (17) Smith, C. A.; O’Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: A Metabolite Mass Spectral Database. *Ther Drug Monit* **2005**, *27* (6), 747–751.
- (18) 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.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521–D526.
- (19) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* **2014**, *42* (W1), 94–99.
- (20) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995, pp 289–300.

- (21) Gloor, G. B.; Hummelen, R.; Macklaim, J. M.; Dickson, R. J.; Fernandes, A. D.; MacPhee, R.; Reid, G. Microbiome Profiling by Illumina Sequencing of Combinatorial Sequence-Tagged PCR Products. *PLoS One* **2010**, *5* (10), e15406.
- (22) Caporaso, J. G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F. D.; Costello, E. K.; Fierer, N.; Peña, A. G.; Goodrich, J. K.; Gordon, J. I.; Huttley, G. A.; Kelley, S. T.; Knights, D.; Koenig, J. E.; Ley, R. E.; Lozupone, C. A.; McDonald, D.; Muegge, B. D.; Pirrung, M.; Reeder, J.; Sevinsky, J. R.; Turnbaugh, P. J.; Walters, W. A.; Widmann, J.; Yatsunenko, T.; Zaneveld, J.; Knight, R. QIIME Allows Analysis of High-Throughput Community Sequencing Data. *Nat. Methods* **2010**, *7* (5), 335–336.
- (23) Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Ser. B. Methodol.* **1982**, *44* (2), 139–177.
- (24) Fernandes, A. D.; Macklaim, J. M.; Linn, T. G.; Reid, G.; Gloor, G. B. ANOVA-like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One* **2013**, *8* (7), e67019.
- (25) Fernandes, A. D.; Reid, J. N.; Macklaim, J. M.; McMurrough, T. A.; Edgell, D. R.; Gloor, G. B. Unifying the Analysis of High-Throughput Sequencing Datasets: Characterizing RNA-Seq, 16S rRNA Gene Sequencing and Selective Growth Experiments by Compositional Data Analysis. *Microbiome* **2014**, *2*, 15.
- (26) Pearson, a D.; Eastham, E. J.; Laker, M. F.; Craft, a W.; Nelson, R. Intestinal Permeability in Children with Crohn’s Disease and Coeliac Disease. *Br. Med. J. (Clin. Res. Ed)*. **1982**, *285* (6334), 20–21.
- (27) Sutherland, L. R.; Verhoef, M.; Wallace, J. L.; Van Rosendaal, G.; Crutcher, R.; Meddings, J. B. A Simple, Non-Invasive Marker of Gastric Damage: Sucrose Permeability. *Lancet* **1994**, *343* (8904), 998–1000.
- (28) Smith, L. a.; Gaya, D. R. Utility of Faecal Calprotectin Analysis in Adult Inflammatory Bowel Disease. *World J. Gastroenterol.* **2012**, *18* (46), 6782–6789.

- (29) Gabbs, M.; Leng, S.; Devassy, J. G.; Monirujjaman, M.; Aukema, H. M. Advances in Our Understanding of Oxylipins Derived from Dietary PUFAs. *Adv. Nutr.* **2015**, *6* (5), 513–540.
- (30) Schaible, U. E.; Kaufmann, S. H. E. Malnutrition and Infection: Complex Mechanisms and Global Impacts. *PLoS Med.* **2007**, *4* (5), e115.
- (31) Holman, R. T.; Johnson, S. B.; Mercuri, O.; Itarte, H. J.; Rodrigo, M. A.; De Tomas, M. E. Essential Fatty Acid Deficiency in Malnourished Children. *Am. J. Clin. Nutr.* **1981**, *34* (8), 1534–1539.
- (32) Leichsenring, M.; Sütterlin, N.; Less, S.; Bäumann, K.; Anninos, A.; Becker, K. Polyunsaturated Fatty Acids in Erythrocyte and Plasma Lipids of Children with Severe Protein-Energy Malnutrition. *Acta Paediatr.* **1995**, *84* (5), 516–520.
- (33) Jones, K. D.; Ali, R.; Khasira, M. a; Odera, D.; West, A. L.; Koster, G.; Akomo, P.; Talbert, A. W.; Goss, V. M.; Ngari, M.; Thitiri, J.; Ndor, S.; Knight, M. a G.; Omollo, K.; Ndungu, A.; Mulongo, M. M.; Bahwere, P.; Fegan, G.; Warner, J. O.; Postle, A. D.; Collins, S.; Calder, P. C.; Berkley, J. A Ready-to-Use Therapeutic Food with Elevated N-3 Polyunsaturated Fatty Acid Content, with or without Fish Oil, to Treat Severe Acute Malnutrition: A Randomized Controlled Trial. *BMC Med.* **2015**, *13*, 93.
- (34) Koletzko, B.; Abiodun, P. O.; Laryea, M. D.; Bremer, H. J. Fatty Acid Composition of Plasma Lipids in Nigerian Children with Protein-Energy Malnutrition. *Eur. J. Pediatr.* **1986**, *145* (1-2), 109–115.
- (35) Smit, E. N.; Muskiet, F. A. J.; Boersma, E. R. The Possible Role of Essential Fatty Acids in the Pathophysiology of Malnutrition: A Review. *Prostaglandins Leukotrienes and Essential Fatty Acids*. 2004, pp 241–250.
- (36) Cunnane, S. C.; Anderson, M. J. The Majority of Dietary Linoleate in Growing Rats Is Beta-Oxidized or Stored in Visceral Fat. *J. Nutr.* **1997**, *127* (1), 146–152.

- (37) Nagao, M.; Parimoo, B.; Tanaka, K. Developmental, Nutritional, and Hormonal Regulation of Tissue-Specific Expression of the Genes Encoding Various Acyl-CoA Dehydrogenases and Alpha-Subunit of Electron Transfer Flavoprotein in Rat. *J. Biol. Chem.* **1993**, *268* (32), 24114–24124.
- (38) Edwards, R. L.; Levine, J. B.; Green, R.; Duffy, M.; Mathews, E.; Brande, W.; Rickles, F. R. Activation of Blood Coagulation in Crohn's Disease. Increased Plasma Fibrinopeptide A Levels and Enhanced Generation of Monocyte Tissue Factor Activity. *Gastroenterology* **1987**, *92* (2), 329–337.
- (39) Ebert, M. P. a; Niemeyer, D.; Deininger, S. O.; Wex, T.; Knippig, C.; Hoffmann, J.; Sauer, J.; Albrecht, W.; Malfertheiner, P.; Röcken, C. Identification and Confirmation of Increased Fibrinopeptide a Serum Protein Levels in Gastric Cancer Sera by Magnet Bead Assisted MALDI-TOF Mass Spectrometry. *J. Proteome Res.* **2006**, *5* (9), 2152–2158.
- (40) Eisenberg, P. R.; Sherman, L. a.; Schectman, K.; Perez, J.; Sobel, B. E.; Jaffe, a. S. Fibrinopeptide A: A Marker of Acute Coronary Thrombosis. *Circulation* **1985**, *71* (5), 912–918.
- (41) Zhang, M. H.; Xu, X. H.; Wang, Y.; Linq, Q. X.; Bi, Y. T.; Miao, X. J.; Ye, C. F.; Gao, S. X.; Gong, C. Y.; Xiang, H.; Dong, M. S. A Prognostic Biomarker for Gastric Cancer With Lymph Node Metastases. *Anat. Rec. Integr. Anat. Evol. Biol.* **2013**, *296* (4), 590–594.
- (42) Medina, R.; Wing, S. S.; Haas, A.; Goldberg, A. L. Activation of the Ubiquitin-ATP-Dependent Proteolytic System in Skeletal Muscle during Fasting and Denervation Atrophy. *Biomed. Biochim. Acta* **1991**, *50* (4-6), 347–356.
- (43) Sesso, R.; Barreto, G. P.; Neves, J.; Sawaya, A. L. Malnutrition Is Associated with Increased Blood Pressure in Childhood. *Nephron - Clin. Pract.* **2004**, *97* (2), C61–C66.
- (44) Martinez-Maldonado, M.; Benabe, J. E.; Wilcox, J. N.; Wang, S.; Luo, C. Renal

- Renin, Angiotensinogen, and ANG I-Converting-Enzyme Gene Expression: Influence of Dietary Protein. *Am. J. Physiol.* **1993**, *264* (6 Pt 2), F981–F988.
- (45) Gomide, J. M.; Demenezes, R. C.; Fernandes, L. G.; Silva, F. C.; Cardoso, L. M.; Miranda, P. H.; Dasilvajr, L. G.; Lima, M. P.; Pesquero, J. L.; Foureaux, G.; Ferreira, A. J.; Chiancajr, D. Increased Activity of the Renin-Angiotensin and Sympathetic Nervous Systems Is Required for the Regulation of the Blood Pressure in Rats Fed a Low Protein Diet. *Exp. Physiol.* **2012**, 1–25.
- (46) Reddy, A. J.; Lam, S. W.; Bauer, S. R.; Guzman, J. A. Lactic Acidosis: Clinical Implications and Management Strategies. *Cleve. Clin. J. Med.* **2015**, *82* (9), 615–624.
- (47) Shapiro, N. I.; Howell, M. D.; Talmor, D.; Nathanson, L. A.; Lisbon, A.; Wolfe, R. E.; Weiss, J. W. Serum Lactate as a Predictor of Mortality in Emergency Department Patients with Infection. *Ann. Emerg. Med.* **2005**, *45* (5), 524–528.
- (48) Mikkelsen, M. E.; Miltiades, A. N.; Gaieski, D. F.; Goyal, M.; Fuchs, B. D.; Shah, C. V.; Bellamy, S. L.; Christie, J. D. Serum Lactate Is Associated with Mortality in Severe Sepsis Independent of Organ Failure and Shock. *Crit. Care Med.* **2009**, *37* (5), 1670–1677.
- (49) Krishna, S.; Waller, D. W.; ter Kuile, F.; Kwiatkowski, D.; Crawley, J.; Craddock, C. F.; Nosten, F.; Chapman, D.; Brewster, D.; Holloway, P. A. Lactic Acidosis and Hypoglycaemia in Children with Severe Malaria: Pathophysiological and Prognostic Significance. *Trans. R. Soc. Trop. Med. Hyg.* *88* (1), 67–73.
- (50) Mtove, G.; Nadjm, B.; Hendriksen, I. C. E.; Amos, B.; Muro, F.; Todd, J.; Reyburn, H. Point-of-Care Measurement of Blood Lactate in Children Admitted with Febrile Illness to an African District Hospital. *Clin. Infect. Dis.* **2011**, *53* (6), 548–554.
- (51) Balla, G.; Vercellotti, G. M.; Muller-Eberhard, U.; Eaton, J.; Jacob, H. S. Exposure of Endothelial Cells to Free Heme Potentiates Damage Mediated by Granulocytes and Toxic Oxygen Species. *Lab. Invest.* **1991**, *64* (5), 648–655.

- (52) Wagener, F. A.; Eggert, A.; Boerman, O. C.; Oyen, W. J.; Verhofstad, A.; Abraham, N. G.; Adema, G.; van Kooyk, Y.; de Witte, T.; Figdor, C. G. Heme Is a Potent Inducer of Inflammation in Mice and Is Counteracted by Heme Oxygenase. *Blood* **2001**, *98* (6), 1802–1811.
- (53) Seixas, E.; Gozzelino, R.; Chora, A.; Ferreira, A.; Silva, G.; Larsen, R.; Rebelo, S.; Penido, C.; Smith, N. R.; Coutinho, A.; Soares, M. P. Heme Oxygenase-1 Affords Protection against Noncerebral Forms of Severe Malaria. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (37), 15837–15842.
- (54) Larsen, R.; Gozzelino, R.; Jeney, V.; Tokaji, L.; Bozza, F. a; Japiassú, A. M.; Bonaparte, D.; Cavalcante, M. M.; Chora, A.; Ferreira, A.; Marguti, I.; Cardoso, S.; Sepúlveda, N.; Smith, A.; Soares, M. P. A Central Role for Free Heme in the Pathogenesis of Severe Sepsis. *Sci. Transl. Med.* **2010**, *2* (51), 51ra71.
- (55) Dalko, E.; Das, B.; Herbert, F.; Fesel, C.; Pathak, S.; Tripathy, R.; Cazenave, P.-A.; Ravindran, B.; Sharma, S.; Pied, S. Multifaceted Role of Heme during Severe Plasmodium Falciparum Infections in India. *Infect. Immun.* **2015**, *83* (10), 3793–3799.
- (56) Berkley, J. A.; Bejon, P.; Mwangi, T.; Gwer, S.; Maitland, K.; Williams, T. N.; Mohammed, S.; Osier, F.; Kinyanjui, S.; Fegan, G.; Lowe, B. S.; English, M.; Peshu, N.; Marsh, K.; Newton, C. R. HIV Infection, Malnutrition, and Invasive Bacterial Infection among Children with Severe Malaria. *Clin Infect Dis* **2009**, *49* (3), 336–343.
- (57) Khogali, M.; Zachariah, R.; Keiluhu, A.; Van den Brande, K.; Tayler-Smith, K.; Ayada, L.; Jima, D.; Hinderaker, S. G.; Harries, A. D. Detection of Malaria in Relation to Fever and Grade of Malnutrition among Malnourished Children in Ethiopia. *Public Heal. Action* **2011**, *1* (1), 16–18.
- (58) Dallmann, R.; Viola, A. U.; Tarokh, L.; Cajochen, C.; Brown, S. A. The Human Circadian Metabolome. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (7), 2625–2629.

Chapter 6

6 Post-acquisition artifact filtering for LC-MS based human metabolomic studies

McMillan, A, Renaud, J, Gloor, GB, Reid, G, and MW, Sumarah

This chapter is based on a manuscript currently under revisions for publication as a preliminary communication in The Journal of Cheminformatics.

Supplemental Tables 6-(1-3) are available for download as additional files (.xlsx)

6.1 Findings

Untargeted metabolomics has a wide array of applications, from biomarker discovery to elucidating disease mechanisms, and characterizing the function of microbial communities. Of all available platforms, liquid chromatography-high resolution mass spectrometry (LC-MS) is capable of detecting the widest range of metabolites. The resulting data from a single untargeted LC-MS experiment contains thousands of “features”, where each represents a single mass-to-charge ratio (m/z). Unlike other ‘omics’ fields, annotation of the complete metabolome is not realistic, and therefore efforts to identify features are focused on those selected via robust statistical approaches. A consequence of selective annotation is that the proportion of features originating from true metabolites versus background contamination or ionization source-generated artifacts remains unknown.

Our recent work on the plasma metabolome of children with severe acute malnutrition prompted us to address this issue. In this dataset, statistical analysis identified approximately 300 features (positive and negative mode combined) which met our pre-defined P value and fold change cut-offs (Wilcoxon test, FDR corrected $P < 0.1$, > 2 fold change, see Supplemental Table 6-1). However, upon further investigation, we noted that a large proportion of significant features were not endogenous metabolites, but rather salt clusters composed of potassium and/or sodium, with chloride and/or formate anions. Although most of these clusters eluted early in the void volume, their retention times overlapped with a number of metabolites of interest, indicating retention time alone is not a suitable filter to remove these source-generated artifacts (Figure 6-1). As the masses of

these clusters are not consistent across datasets, they cannot be removed based on m/z alone (data not shown).

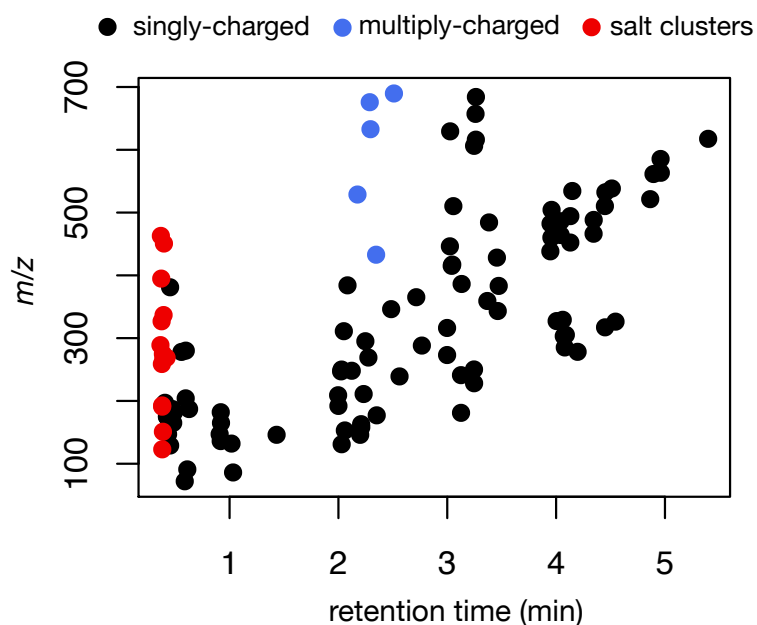


Figure 6-1. Retention time and m/z of plasma metabolites differing significantly between children with severe acute malnutrition and controls

Wilcoxon test, FDR corrected $P < 0.1$, > 2 fold change. Salt clusters are shown in red, with validated metabolites colored according to their charge state. Data was acquired by positive ESI.

Electrospray ionization in particular is a ‘soft’ ionization technique and is known to generate non-covalent complexes, including salt clusters, which can occur irrespective of the extraction method, solvent, column or platform used¹⁻⁴. Salt clusters are derived from compounds present in the LC buffer and/or compounds present in the sample itself therefore, with the most commonly observed clusters consisting of combinations of small cations such as Na^+ and/or NH_4^+ with chlorides and/or small organic acids, such as formate (HCOO^-) and acetate (H_3CCOO^-)¹. These salt clusters have m/z values with characteristically high mass defects, which for this work we simply define as the decimal numbers after the nominal mass. This is the result of the relatively high ratio of elements such as chlorine (34.9688527 Da), sodium (22.9897697 Da), potassium (38.963708 Da), and oxygen (15.9949146 Da), compared to hydrogen (1.007825 Da) and nitrogen (14.003074 Da).

When the mass defect of all endogenous or food-derived metabolites in the human metabolome database⁵ are plotted by m/z (Figure 6-2, Supplemental Table 6-2), only 0.39 % of endogenous compounds fall within the mass defect space occupied by salt clusters, confirming the rarity of human metabolites with such high mass defects.

Given the ubiquity of salt clusters in LC-MS data¹⁻⁴, and their predictable mass defect, we developed a method to identify and remove salt cluster artifacts from untargeted LC-MS data using mass defect filtering. This comprised performing a linear regression of compounds with the highest mass defect in the human metabolome database (Figure 6-2), then modelling C_nH_{n+2} alkanes, which represent the theoretical maxima mass defect for compounds containing only carbon, hydrogen, oxygen, and/or nitrogen. Both methods yielded the same linear equation ($y=0.00112x + 0.01953$). We then applied this equation to experimental datasets and removed features with mass defects greater than our model equation. Given there are a small number of endogenous metabolites in the human metabolome database with high mass defects, such as iodine containing thyroid hormones and triphosphates, we also incorporated an “inclusion list” into our model. Features in experimental datasets with the same m/z as compounds in this inclusion list (within a pre-set error range) will be retained; ensuring known endogenous compounds are not removed with salt clusters.

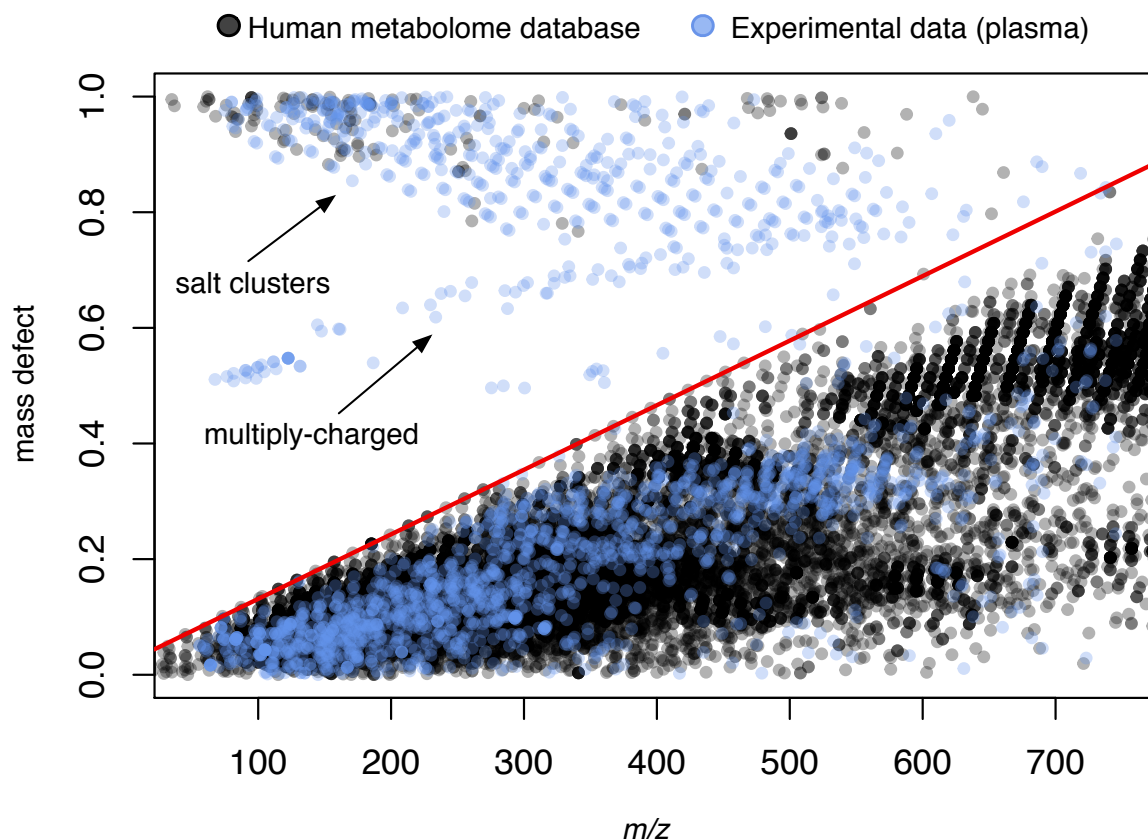


Figure 6-2. Mass defect as a function of m/z for all endogenous and food-derived compounds between 50-750 m/z in the human metabolome database (hmdb, $n=17957$) compared to features detected experimentally in plasma ($n=2227$).

The red line depicts the theoretical mass defect maxima for compounds containing only C, H, N and/or O ($y=0.00112x + 0.01953$). Features above this line which are present only in the experimental dataset are likely to be salt clusters or multiply charged as indicated by the arrows. Hmdb masses are shown as theoretical $[M+H]$. Experimental data was acquired by positive ESI, and contains a variety of adducts including but not limited to $[M+H]$, $[M+Na]$, $[M+K]$ and $[M+NH_4]$, as well as isotopic peaks.

To test the ability of our method to remove artifacts while retaining validated metabolites, we applied this filter to plasma data from the metabolomics study of severe acute malnutrition mentioned previously (Supplemental Table 6-1). Importantly, the metabolites of interest contained multiply-charged peptides, which were not modelled by the human metabolome dataset (Figure 6-2). Some of these peptides occupied the same mass defect space as the salt clusters, and therefore would be removed from the analysis by our original, 'mass defect only' method (Figure 6-3A,B). However, using a C18 column, the salt clusters

elute in, or shortly after the void volume, while peptides are retained. We therefore incorporated retention time into the model as a third variable to further isolate the salt clusters. Incorporation of retention time removed all salt clusters while retaining all identified metabolites of interest, confirming the validity of the method (Figure 6-3C).

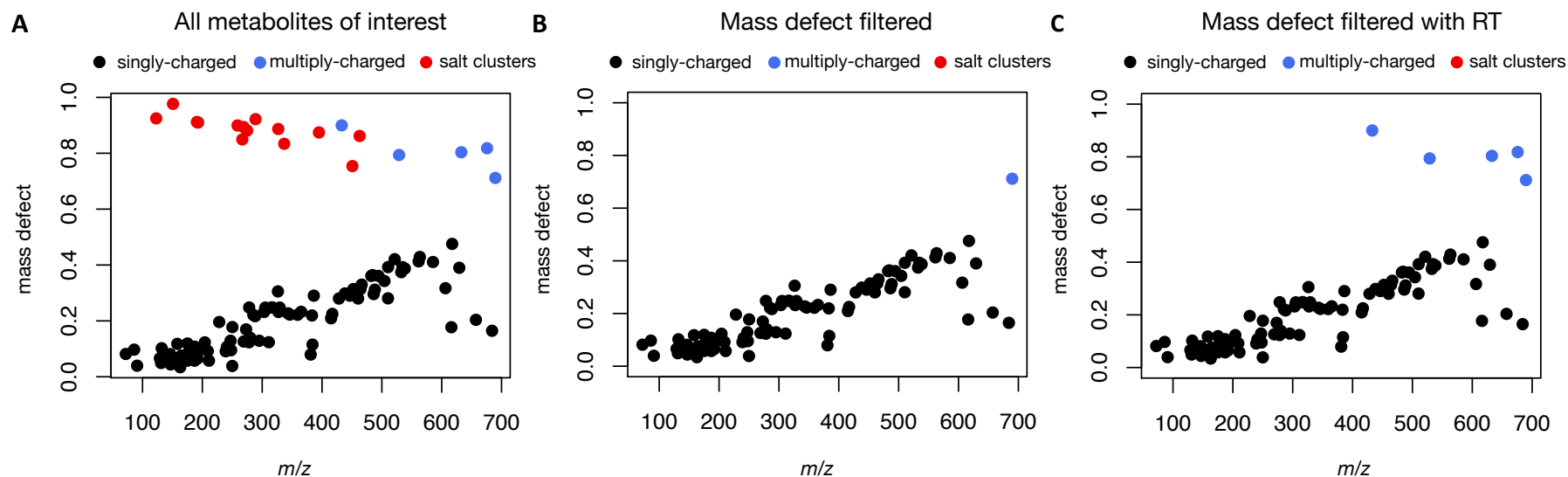


Figure 6-3. Significant metabolites (Wilcoxon test, FDR corrected $P < 0.1$, > 2 fold change) remaining in the malnutrition dataset after A. no filter, B. mass defect filter only, C, mass defect filter with retention time.

Each point represents a single metabolite, colored according to their charge states, with salt clusters shown in red. Data was acquired by positive ESI. Singly-charged features encompass a variety of adducts including but not limited to $[M+H]$, $[M+Na]$, $[M+K]$ and $[M+NH_4]$.

We next applied the mass defect filter to all features detected in plasma (2227 in positive mode, 1742 in negative mode) to determine the percentage of features in the complete dataset with mass defects corresponding to salt clusters. This analysis revealed a large percentage (15.94% in positive mode, 28.82% in negative mode) of total features were likely salt clusters (Table 6-1).

Table 6-1. Percent data reduction after mass defect filtering alone or in combination with retention time and hmdb inclusion list. All features detected in plasma in the malnutrition dataset are shown in both positive and negative ionization mode. RT: retention time, md: mass defect.

ionization mode	filter	features remaining	% features removed
positive	none	2227	0.00
	md	1730	22.32
	md+RT	1853	16.79
	md+RT+inclusion	1872	15.94
negative	none	1742	0.00
	md	1107	36.45
	md+RT	1225	29.68
	md+RT+inclusion	1240	28.82

To determine if the proportion of salt clusters could be reduced instrumentally, and if they occurred in other biological matrices, we ran a series of tests comparing the effect of sweep gas and column type (reverse phase or HILIC) on salt cluster formation in a set of three plasma, urine and stool samples. The sweep gas did not significantly reduce the salt cluster proportion (Supplemental Table 6-3), although fewer features were detected overall, indicating lower sensitivity with this method. Surprisingly, the use of HILIC columns consistently increased the proportion of salt clusters (Figure 6-4, Supplemental Table 6-3), perhaps due to less ion suppression at later retention times where these salt clusters elute.

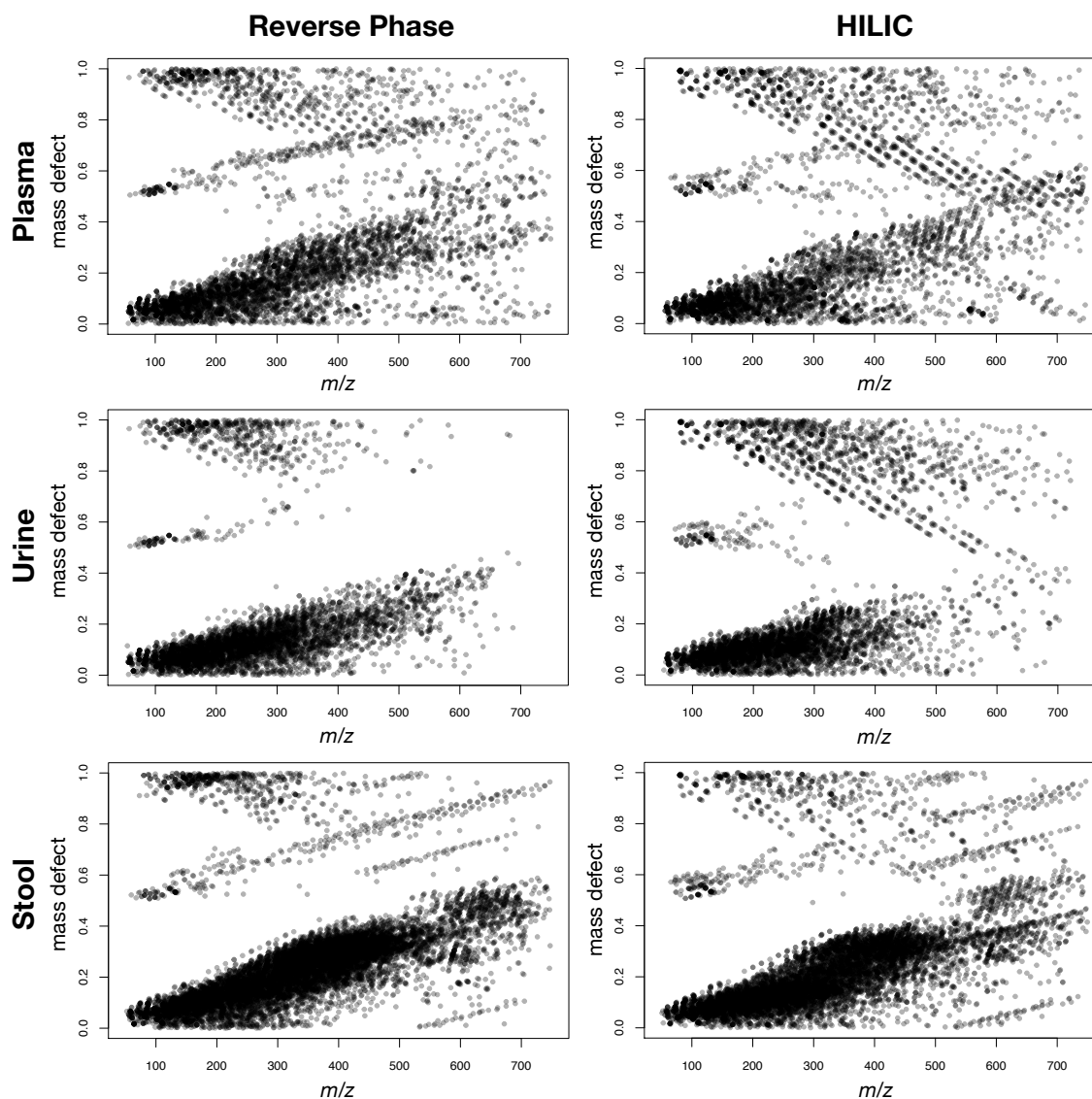


Figure 6-4. Effect of column type on salt cluster formation in plasma, urine and stool.

The left and right panels display mass defect as a function of m/z for the same set of samples run on reverse phase or HILIC columns respectively. Data was acquired by positive ESI.

Although salt clusters were identified in all sample types, they consistently occurred at a lower proportion in stool and urine compared to plasma. The use of K_2EDTA tubes for blood collection in our study may be responsible for this observation. Barri *et al.*⁶

demonstrated that plasma collected with EDTA tubes had a significantly higher number of potassium clusters compared to heparin tubes, while citrate tubes resulted in more sodium clusters. These results indicate that salt cluster removal is applicable to stool, urine and plasma, but is particularly important for plasma collected with tubes containing salt-based anticoagulants.

It is worth noting that our method was designed for studies pertaining to human physiology, and therefore synthetic compounds were not included in the analysis. Synthetic compounds such as drugs and pesticides, which are more likely to contain halogens^{7,8}, can occupy the salt cluster mass defect space. Investigators concerned with these types of molecules may therefore wish to incorporate these masses in the inclusion list or refrain from applying this correction to their datasets.

The advantage of removing artifacts prior to annotation and statistical analyses is three-fold. Firstly, removing a large number of unknown features will allow for more complete annotation of the metabolome, and decreased reporting of false positives. Secondly, feature reduction may change the relationship between samples as determined by multivariate modelling methods such as principal component analysis. Most importantly, removing hundreds of features will affect the distribution of P values generated from univariate analyses. This has important implications for multiple testing corrections, such as the false discovery rate (FDR) and Bonferroni adjustment, which rely on this distribution (FDR), or on the total number of features compared (Bonferroni) for P value adjustment^{9,10}.

In conclusion, we propose a method to filter out salt cluster artifacts in untargeted LC-MS data using mass defect and retention time. This filter can be easily applied to processed data using a set of R functions, and requires only a list of detected m/z and retention time values. The code for these analyses as well as example datasets are freely available at (https://github.com/amcmil/mz_defect_filter).

6.2 Materials and Methods

6.2.1 Metabolite extraction

For stool, approximately 250 mg of wet sample was lyophilized overnight. After drying, 40 mg was weighed into microcentrifuge tubes and extracted with 8:2 methanol:H₂O to a final concentration of 40 mg/mL. Samples were then vortexed for 30 sec, followed by centrifugation for 15 min at 10 000 rpm. Supernatant was then transferred to LC-MS vials for with micro-inserts for analysis.

Metabolites were extracted from plasma according to the methods of Dunn *et al.*, 2011¹¹. Briefly, plasma samples were thawed on ice for 30 min. Once thawed, 805 μ L of 8:2 methanol:H₂O was added to 230 μ L of plasma to make a 4.5 fold dilution. Samples were vortexed for 15 sec and centrifuged at 15 000 rpm for 15 min to pellet precipitated proteins. 370 μ L of supernatant was then transferred to separate vials and dried down for LC-MS using a speedvac with no heat. Samples were then reconstituted with 90 μ L ddH₂O and transferred to LC-MS vials with micro-inserts for analysis.

For urine, 200 μ L of sample was extracted with 800 μ L 1:9 Acetonitrile:H₂O as per the methods of Warth *et al.* 2012¹². After centrifugation, 500 μ L of supernatant was transferred to LC-MS vials with micro-inserts for analysis.

6.2.2 LC-MS analyses

Samples were analyzed using an Agilent 1290 Infinity HPLC coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, USA) with a HESI (heated electrospray ionization) source. For reverse phase HPLC, 2 μ L of each sample was injected into a ZORBAX Eclipse plus C18 2.1 x 50mm x 1.8 micron column. Mobile phase (A) consisted of 0.1% formic acid in water and mobile phase (B) consisted of 0.1% formic acid in acetonitrile. The initial composition of 0% (B) was held constant for 30 s and increased to 100% over 3.0 min. Mobile phase B was held at 100% for 2 minutes and returned to 0% over 30s for a total run time of 6 min. For normal phase HPLC, 2 μ L of each sample was injected into a ZORBAX RRHD HILIC plus 2.1 x 50mm x 1.8 micron column. Mobile phase (A) consisted of 0.1% formic acid in water and mobile phase (B) consisted of 0.1% formic acid in acetonitrile. The initial composition of 95% (B) was held constant for 30 s and decreased to 5% over 3.0 min. Mobile phase B was held at 5% for 1 minute and returned to 95% over 30s and held for 1 minute for a total run time of 6 min.

Full MS scanning between the ranges of m/z 50-750 was performed at 140 000 resolution. The HESI source was operated under the following conditions: nitrogen flow of 30 and 8 arbitrary units for the sheath and auxiliary gas respectively, probe temperature and capillary temperature of 450 °C and 250 °C respectively and spray voltage of 3.9 kV and 3.5 kV in positive and negative mode respectively. The automatic gain control (AGC) target and maximum injection time were 1e6 and 500 ms respectively. For experiments testing the affect of sweep gas on cluster formation, sweep gas was set to 2 arbitrary units. Blanks of pure methanol were run between every sample to limit carryover. After data acquisition Thermo .RAW files were converted to .MZML format and centroided using ProteoWizard¹³. Files were then imported into R using the XCMS package¹⁴ for chromatogram alignment and deconvolution. Features were detected with the “xcmsSet” function using the “centWave” method and a ppm tolerance of 1. Prefilter was set to 3-5000, noise 1E5, and signal to noise threshold was set to 5. Due to a lower overall noise and signal in negative mode, noise was set to 1E3 for this mode. Retention time correction was conducted using the “obiwarp” method, grouping included features present in at least one samples, allowable retention time deviation was 5 seconds, and m/z width set to 0.015. Areas of features below the signal to noise threshold in the data were integrated using the “fillPeaks” function with default settings.

6.3 References

- (1) Zhou, S.; Hamburger, M. Formation of Sodium Cluster Ions in Electrospray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10* (7), 797–800.
- (2) Zhang, D.; Cooks, R. G. Doubly Charged Cluster Ions $[(\text{NaCl})_m(\text{Na})_2]^{2+}$: Magic Numbers, Dissociation, and Structure. *Int. J. Mass Spectrom.* **2000**, *195-196*, 667–684.
- (3) Hao, C.; March, R. E.; Croley, T. R.; Smith, J. C.; Rafferty, S. P. Electrospray Ionization Tandem Mass Spectrometric Study of Salt Cluster Ions. Part 1--

- Investigations of Alkali Metal Chloride and Sodium Salt Cluster Ions. *J. Mass Spectrom.* **2001**, *36* (1), 79–96.
- (4) Konermann, L.; McAllister, R. G.; Metwally, H. Molecular Dynamics Simulations of the Electrospray Process: Formation of NaCl Clusters via the Charged Residue Mechanism. *J. Phys. Chem. B* **2014**, *118* (41), 12025–12033.
- (5) 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.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Macinnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35* (Database issue), D521–D526.
- (6) Barri, T.; Dragsted, L. O. UPLC-ESI-QTOF/MS and Multivariate Data Analysis for Blood Plasma and Serum Metabolomics: Effect of Experimental Artefacts and Anticoagulant. *Anal. Chim. Acta* **2013**, *768*, 118–128.
- (7) Hernandez, M. Z.; Cavalcanti, S. M. T.; Moreira, D. R. M.; de Azevedo Junior, W. F.; Leite, A. C. L. Halogen Atoms in the Modern Medicinal Chemistry: Hints for the Drug Design. *Curr. Drug Targets* **2010**, *11* (3), 303–314.
- (8) Jeschke, P. The Unique Role of Halogen Substituents in the Design of Modern Agrochemicals. *Pest Manag. Sci.* **2010**, *66* (1), 10–27.
- (9) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995, pp 289–300.
- (10) Bonferroni, Carlo, E. Teoria Statistica Delle Classi E Calcolo Delle Probabilità. *Pubbl. del R Ist. Super. di Sci. Econ. e Commer. di Firenze* **1936**, *8*, 3–62.

- (11) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry. *Nat. Protoc.* **2011**, *6* (7), 1060–1083.
- (12) Warth, B.; Sulyok, M.; Fruhmann, P.; Mikula, H.; Berthiller, F.; Schuhmacher, R.; Hametner, C.; Abia, W. A.; Adam, G.; Fröhlich, J.; Krska, R. Development and Validation of a Rapid Multi-Biomarker Liquid Chromatography/tandem Mass Spectrometry Method to Assess Human Exposure to Mycotoxins. *Rapid Commun. Mass Spectrom.* **2012**, *26* (13), 1533–1540.
- (13) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536.
- (14) Patti, G. J.; Tautenhahn, R.; Siuzdak, G. Meta-Analysis of Untargeted Metabolomic Data from Multiple Profiling Experiments. *Nat. Protoc.* **2012**, *7* (3), 508–516.

Chapter 7

7 General discussion

The composition of the microbiome has only recently become accessible, resulting from advancements in NGS techniques. Due to the infancy of the field, as well as the cost of metagenomics/transcriptomics, the majority of microbiome studies to date have been limited to “who is there”. Metagenomics has become increasingly more affordable and common, but validation of predicted products is still extremely rare, and conclusions are usually drawn from metagenomics alone¹⁻³. Our work on BV in Chapters 2 & 3 has discovered that many metabolites discriminating BV from health are of unknown origin (i.e. 2HV, GHB, 2-hydroxyisocaproate, 2-hydroxyglutarate), with the genes responsible for their production either being poorly characterized or completely unknown. Indeed, none of the above mentioned metabolites were predicted based on meta-transcriptome analysis⁴. Our work therefore serves as a cautionary tale for those tempted to draw sweeping conclusions from metagenomic data alone. As demonstrated by the case of the amines in BV, a single metabolite can have a large impact on phenotype^{5,6}. Conversely, it is often overlooked that genes and transcripts identified by NGS experiments are merely predictions based on sequence similarity to genes whose functions have been proven experimentally. In the absence of classical microbiology methods, metabolomics offers some confirmation of these predictions. The findings within Chapters 2 & 3 highlight this fact, and advocate that a combination of techniques is absolutely necessary to obtain a complete picture of the function of microbial communities.

7.1 The vaginal metabolome in BV

Despite being the most common vaginal condition, the diagnostics for BV have remained stagnant for the past 25 years⁷. The current methods are time consuming and unreliable, prompting the search for new diagnostic biomarkers by our group. Upon commencement of my PhD studies there were no metabolomic studies of the vaginal environment. Though we were not the first to publish work on the vaginal metabolome⁸⁻¹⁰, the Rwandan study in Chapter 2 represents the largest study to date, and the only one to include pregnant women. We were also the first to identify GHB and 2HV as biomarkers

of BV¹¹. These biomarkers have since been confirmed in two independent cohorts by other groups¹², as well as in our own study of Canadian women in Chapter 3. Whether these metabolites play any functional role in BV is yet to be determined, however we have shown that they do not have any effect on growth of vaginal bacteria, nor do they effect cytokine production by vaginal epithelial cell lines stimulated with TNF-alpha (data not shown).

Replication is an important aspect of these studies. Validation of biomarkers is encouraged in the metabolomics field and is often required by high impact journals¹³⁻¹⁶. This is due in part to the large number of variables analyzed in a single experiment, increasing the likelihood of false positives. The same could be said for microbiome data, however validation of findings within the same scientific article is extremely rare in this field. Given the high degree of irreproducibility among microbiome studies^{17,18}, replication would be of enormous benefit, especially when the findings are controversial.

The studies of the vaginal metabolome Chapters 2 & 3 have also allowed direct comparison of the three most common metabolomics platforms; NMR, GC-MS and LC-MS. Although there was some degree of overlap between them, each method identified some degree of unique information. There were many metabolites which could be detected by all three methods, such was the case with 2HV, cadaverine, succinate, and many amino acids. In other cases, only one or two of the methods was able to detect a given metabolite. For example, trimethylamine was detected by NMR and LC-MS, but not GC-MS. Due to time constraints, the metabolites identified exclusively by LC-MS were not investigated for the Rwandan study, however this method would have uncovered the largest number of unique metabolites, as detection of large lipids, peptides, and poryphrins is possible^{19,20}.

An additional point of interest is the surprising amount of reproducibility in the metabolome between ethnicities and geographical locations. This undoubtedly arises from the highly conserved nature of the vaginal microbiota. Although the prevalence of certain taxa has been shown to differ between ethnicities, with African Americans having

a higher prevalence on BV associated bacteria^{21,22}, the core vaginal microbiota is remarkably conserved worldwide. Indeed, in every instance, *L. iners* and *L. crispatus* dominate the microbiota of healthy women^{21,23–26}. The reason for the dominance of these two species is still not well understood, but has led to the theory of universality necessary for human reproduction²⁷. The findings in Chapter 4 suggest *L. iners*' unique ability to bind fibronectin may be one of the many factors involved^{28,29}. Interestingly, we found that women of African descent from Rwanda had a high prevalence of *Lactobacillus* species. Although recruitment was biased to include an equal number of women with and without BV, we were still not able to recruit enough women with BV in the time provided. This raises an interesting question as to whether the increased rates of BV in African American women are of genetic or environmental origin. Addressing this question will require further studies directly comparing African American with African women in an unbiased fashion, but could provide novel insight into the factors which shape the vaginal microbiota.

7.2 Integration of metabolite and NGS data

By marrying the metabolome and microbiome, we have shown in Chapters 2 & 3 that these data can identify biologically relevant species-product relationships. A surprising finding of this thesis was that correlations between metabolites and taxa and/or transcripts are relatively weak. This likely stems from multiple factors, including the high degree of co-occurrence between BV associated bacteria, as well as the fact that absolute abundance of bacteria and/or transcript cannot be measured by NGS approaches. Although we confirmed production of GHB by *G. vaginalis in vitro* and identified the putative enzyme in the transcriptome, the Spearman's correlation between GHB and *G. vaginalis* was only 0.56 in the Rwandan study and 0.74 in the Canadian study. In the case of succinate and *L. crispatus*, the relationship is even more convoluted by the fact that many organisms other than *L. crispatus* produce succinate. Despite these obstacles, we have shown it is possible to identify meaningful correlations armed with the knowledge of the system and data at hand.

This work has also highlighted the enormous amount of spurious correlations between these data. Conclusions based on these correlations should therefore be made with extreme caution. Ideally one would require multiple lines of evidence in addition to *in vitro* validation for confirmation of these species-product relationships. Due to the demanding nutritional and growth requirements of many vaginal isolates, the latter may not always be possible. This is the case for many environments, including the gut. The ability to culture the unculturable is thus an underappreciated yet essential piece to understanding the microbiome and its role in health and disease³⁰.

7.3 The microbiome and metabolome in malnutrition

The malnutrition study in Chapter 5 created an opportunity to apply the methods developed in the earlier chapters to a more complex system. Contrary to previous studies^{31,32}, we could not identify any significant differences in the microbiome of children with SAM compared to non-malnourished controls. While this could be a result of the modest sample size of our study, the lack of metabolites in plasma of clear microbial origin differentiating children with SAM from controls argues against this. We also explored correlations between the stool microbiome (16S) and stool metabolome in this dataset, and found these to be relatively weak (data not shown), consistent with what was found in the vaginal environment. This may be due in part to functional overlap between taxa, but may also be related to the compositional nature of NGS data as was discussed previously.

Finally, this chapter highlighted the usefulness of an untargeted approach. Many of the plasma metabolites differentiating children with SAM from controls have not been linked to the condition previously (heme, oxylipins, truncated FPA and others), allowing generation of new hypotheses regarding the pathology of SAM in humans.

7.4 Improving upon metabolomics methods

A major aspect of this thesis was developing workflows for analysis of metabolomic data. This involved evaluation of various software available for metabolite integration, deconvolution and identification, as well as methods of visualizing these data. Although

many more software exist from the ones employed in this thesis, we found these (AMDIS³³, Spectconnect³⁴, Xcaliber, XCMS³⁵, R³⁶) to be most appropriate for our specific needs. In terms of metabolite annotation, we opted for a low throughput approach, with each metabolite being manually curated. For LC-MS, annotation is by far the most time consuming portion of analysis, creating a push for more automated methods. While there are software available which claim accurate annotation based on *m/z* and/or MS/MS (Mummuchog³⁷, CSI-FingerID³⁸, CFM-ID³⁹, LipidBlast⁴⁰), we found none of these to be as accurate as manual annotation. Indeed, in most cases the software either could not identify any plausible match, or returned false positives. Thus, although significant progress has been made in the area of metabolite annotation, there is much to be done before automated annotation is a reality. Despite the fact that automated annotation is not reliable in most cases, many studies continue to publish metabolomics data without the use of standards or even MS/MS data. This is concerning, especially given that many of these studies are multidisciplinary, with many of reader's having expertise outside the chemistry field and therefore unable to critically analyze the work.

As noted in Chapter 6, there is also an enormous amount of noise in LC-MS data, with the identity of most features in a given experiment remaining unknown. Separating the metabolites of biological origin from the noise is a key first step in making annotation of the complete metabolome feasible. We demonstrated in Chapter 6 that by utilizing intrinsic properties of salt clusters (mass defect), these artefacts can be easily identified and removed. Data reduction methods such as this will decrease reporting of false positives, impact multivariate modelling results as well as statistical analysis.

7.5 Future directions

There are many aims which could not be addressed during the time span of my PhD. An obvious, and I believe feasible goal is to develop a rapid diagnostic test for BV based on GHB, 2HV and/or their tyrosine ratios. GHB is a controlled substance⁴¹, which has driven the development of screening tests for specimens such as urine and hair, making this metabolite the most logical choice for BV diagnostics^{42,43}. A rapid diagnostic test for GHB in urine does exist⁴², however the limits of detection are not likely within the

range of vaginal fluid. We estimate GHB concentrations to be in the range of 2µg/mL in women with BV, with currently available tests reporting limits of detection of 10µg/mL. Additional method work would therefore be required to produce a test sensitive enough to detect BV.

Development of tests specific for an organisms and/or their metabolic by-products such as this may lead to more personalized treatments in the future. In the example of GHB and *G. vaginalis*, women with this organism may require an alternative treatment to someone dominated by BVAB1 for example. *G. vaginalis* is thought to be one of the more pathogenic organisms in BV due in part to sialidase production which breaks down the mucus layer of the vaginal epithelium, potentially contributing to the increased susceptibility to HIV^{44,45}. Sialidase levels also predict preterm birth in a dose dependent manner in women with BV, and *G. vaginalis* supernatants directly stimulate HIV-1 expression *in vitro*, providing further evidence for the role of this organism in adverse outcomes associated with BV⁴⁶. Despite this knowledge, many clades of *G. vaginalis* are intrinsically resistant to the front-line treatment for BV, metronidazole⁴⁷. It is therefore not surprising that the BV recurrence rates are high, and that alternative treatments are needed. Probiotics have been shown to be effective in increasing cure rates when combined with antibiotic treatment^{48,49}, perhaps due in part to disruption of *G. vaginalis* biofilms by these organisms^{50,51}. Alternative therapies such as this have yet to make their way into common practice, but undoubtedly will in the future given the increasing awareness of the benefits of probiotics⁵².

Longitudinal sampling of the microbiome and metabolome of women with BV would also be highly valuable. If metabolic markers can be detected in advance of symptoms, treatment before onset may be possible. Further studies are required to determine if early treatment translates into reduced risk of co-morbidities associated with BV. At the very least, the unpleasant symptoms associated with the condition might be prevented, which would benefit those with recurrent BV most. Longitudinal analysis is also needed to determine how the metabolome varies with menstrual cycle. Other groups have demonstrated that the proportion of *G. vaginalis* increases significantly during menses⁵³.

This observation suggests GHB levels may also increase during this time, creating the potential for false positives if diagnostics based on GHB are implemented.

In terms of future directions for the malnutrition study in Chapter 5, replication is a key next step. Once metabolic changes due to SAM have been validated in an independent cohort, a number of avenues are worth following. The theory of non-administered disaccharides being indicative of enteropathy is an intriguing finding, and should be validated in animal models and humans by comparison to histology scores and dual sugar permeability tests respectively. Confirmation of these associations would be highly valuable, allowing children with enteropathy to be rapidly identified. Without such a method of identifying children with enteropathy, we cannot determine how this contributes to malnutrition, and evaluate treatments effective in reducing it.

The reduction in oxylipins in children with SAM is also worth investigating to determine if reductions contribute to an impaired immune response. If oxylipin reduction is shown to result mainly from limited availability of long-chain polyunsaturated fatty acids (PUFAs) precursors, this would provide evidence for modification of the current nutritional interventions include higher levels of PUFAs.

Metabolomics is a rapidly evolving field, and its application to microbiome studies holds much promise. This thesis has demonstrated how these data can be integrated to find useful information, whether it be in the form of disease biomarkers or improved understanding of microbial metabolism. It is my hope that some of these findings will one day be translated into clinical practice to improve the diagnosis, management and/or treatment of disease as this is the ultimate goal of medical research.

7.6 References

- (1) Forslund, K.; Hildebrand, F.; Nielsen, T.; Falony, G.; Le Chatelier, E.; Sunagawa, S.; Prifti, E.; Vieira-Silva, S.; Gudmundsdottir, V.; Krogh Pedersen, H.; Arumugam, M.; Kristiansen, K.; Yvonne Voigt, A.; Vestergaard, H.; Hercog, R.; Igor Costea,

- P.; Roat Kultima, J.; Li, J.; Jørgensen, T.; Levenez, F.; Dore, J.; MetaHIT consortium; Bjørn Nielsen, H.; Brunak, S.; Raes, J.; Hansen, T.; Wang, J.; Dusko Ehrlich, S.; Bork, P.; Pedersen, O. Disentangling Type 2 Diabetes and Metformin Treatment Signatures in the Human Gut Microbiota. *Nature* **2015**, *528* (7581), 262–266.
- (2) Qin, N.; Yang, F.; Li, A.; Prifti, E.; Chen, Y.; Shao, L.; Guo, J.; Le Chatelier, E.; Yao, J.; Wu, L.; Zhou, J.; Ni, S.; Liu, L.; Pons, N.; Batto, J. M.; Kennedy, S. P.; Leonard, P.; Yuan, C.; Ding, W.; Chen, Y.; Hu, X.; Zheng, B.; Qian, G.; Xu, W.; Ehrlich, S. D.; Zheng, S.; Li, L. Alterations of the Human Gut Microbiome in Liver Cirrhosis. *Nature* **2014**, *513* (7516), 59–64.
- (3) 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. S. S.; Li, D.; Cao, J.; Wang, B.; Liang, H.; Zheng, H.; Xie, Y.; Tap, J.; Lepage, P.; Bertalan, M.; Batto, J.-M.; Hansen, T.; Le Paslier, D.; Linneberg, A.; Nielsen, H. B.; Pelletier, E.; Renault, P.; Sicheritz-Ponten, T.; Turner, K.; Zhu, H.; Yu, C.; Li, S. S. S.; Jian, M.; Zhou, Y.; Li, Y.; Zhang, X.; Li, S. S. S.; Qin, N.; Yang, H.; Wang, J. J.; Brunak, S.; Doré, J.; Guarner, F.; Kristiansen, K.; Pedersen, O.; Parkhill, J.; Weissenbach, J.; MetaHIT Consortium; Bork, P.; Ehrlich, S. D.; Wang, J. J.; Ruiqiang Li¹; 3, M. A. * J. R.; Burgdorf⁴, K. S.; Manichanh⁵, C.; Nielsen⁴, T.; Pons⁶, N.; Levenez⁶, F.; Yamada², T.; Mende², D. R.; 7, J. T. Y. X. J. X. J. L.; Li¹, S. S. S.; 8, J. C. D. L.; Wang¹, B.; Liang¹, H.; Zheng¹, H.; 7, J. T. Y. X. J. X. J. L.; Lepage⁶, P.; Bertalan⁹, M.; Batto⁶, J.-M.; Hansen⁴, T.; Le, D.; Allan Linneberg¹¹, H. B. N. P.; Eric Pelletier¹⁰, P. R.; Sicheritz-Ponten⁹, T.; Keith Turner¹², H. Z.; Yu¹, C.; Li¹, S. S. S.; Jian¹, M.; Zhou¹, Y.; Li¹, Y.; Zhang¹, X.; Li¹, S. S. S.; Qin¹, N.; Yang¹, H.; Wang¹, J.; Brunak⁹, S.; Dore⁶, J.; Guarner⁵, F.; MetaHIT Consortium {, P. B. K. K. O. P. 14 J. P. J. W.; Wang, S. D. E. & J. A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing: Commentary. *Nature* **2010**, *11* (1), 28.
- (4) Macklaim, J. M.; Fernandes, A. D.; Di Bella, J. M.; Hammond, J.-A.; Reid, G.;

- Gloor, G. B. Comparative Meta-RNA-Seq of the Vaginal Microbiota and Differential Expression by *Lactobacillus iners* in Health and Dysbiosis. *Microbiome* **2013**, *1* (1), 12.
- (5) Wolrath, H.; Forsum, U.; Larsson, P. G.; Borén, H. Analysis of Bacterial Vaginosis-Related Amines in Vaginal Fluid by Gas Chromatography and Mass Spectrometry. *J. Clin. Microbiol.* **2001**, *39* (11), 4026–4031.
- (6) Sobel, J. D.; Karpas, Z.; Lorber, A. Diagnosing Vaginal Infections through Measurement of Biogenic Amines by Ion Mobility Spectrometry. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2012**, *163* (1), 81–84.
- (7) Nugent, R. P.; Krohn, M. A.; Hillier, S. L. Reliability of Diagnosing Bacterial Vaginosis Is Improved by a Standardized Method of Gram Stain Interpretation. *J. Clin. Microbiol.* **1991**, *29* (2), 297–301.
- (8) Gajer, P.; Brotman, R. M.; Bai, G.; Sakamoto, J.; Schütte, U. M. E.; Zhong, X.; Koenig, S. S. K.; Fu, L.; Ma, Z. S.; Zhou, X.; Abdo, Z.; Forney, L. J.; Ravel, J. Temporal Dynamics of the Human Vaginal Microbiota. *Sci. Transl. Med.* **2012**, *4* (132), 132ra52.
- (9) Yeoman, C. J.; Thomas, S. M.; Miller, M. E. B.; Ulanov, A. V.; Torralba, M.; Lucas, S.; Gillis, M.; Cregger, M.; Gomez, A.; Ho, M.; Leigh, S. R.; Stumpf, R.; Creedon, D. J.; Smith, M. A.; Weisbaum, J. S.; Nelson, K. E.; Wilson, B. A.; White, B. A. A Multi-Omic Systems-Based Approach Reveals Metabolic Markers of Bacterial Vaginosis and Insight into the Disease. *PLoS One* **2013**, *8* (2), e56111.
- (10) Laghi, L.; Picone, G.; Cruciani, F.; Brigidi, P.; Calanni, F.; Donders, G.; Capozzi, F.; Vitali, B. Rifaximin Modulates the Vaginal Microbiome and Metabolome in Women Affected by Bacterial Vaginosis. *Antimicrob. Agents Chemother.* **2014**, *58* (6), 3411–3420.
- (11) McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J.; Gloor, G. B.; Reid, G. A Multi-platform Metabolomics Approach Identifies Novel

Biomarkers Associated with Bacterial Diversity in the Human Vagina
<http://arxiv.org/abs/1504.02816> (accessed Jun 8, 2016).

- (12) Srinivasan, S.; Morgan, M. T.; Fiedler, T. L.; Djukovic, D.; Hoffman, N. G.; Raftery, D.; Marrazzo, J. M.; Fredricks, D. N. Metabolic Signatures of Bacterial Vaginosis. *MBio* **2015**, *6* (2).
- (13) Wang, T. J.; Larson, M. G.; Vasan, R. S.; Cheng, S.; Rhee, E. P.; McCabe, E.; Lewis, G. D.; Fox, C. S.; Jacques, P. F.; Fernandez, C.; O'Donnell, C. J.; Carr, S. A.; Mootha, V. K.; Florez, J. C.; Souza, A.; Melander, O.; Clish, C. B.; Gerszten, R. E. Metabolite Profiles and the Risk of Developing Diabetes. *Nat. Med.* **2011**, *17* (4), 448–453.
- (14) 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.; Lysis, A. J.; Hazen, S. L. Gut Flora Metabolism of Phosphatidylcholine Promotes Cardiovascular Disease. *Nature* **2011**, *472* (7341), 57–63.
- (15) Elliott, P.; Pasma, J. M.; Chan, Q.; Garcia-Perez, I.; Wijeyesekera, A.; Bictash, M.; D. Ebbels, T. M.; Ueshima, H.; Zhao, L.; van Horn, L.; Daviglus, M.; Stamler, J.; Holmes, E.; Nicholson, J. K. Urinary Metabolic Signatures of Human Adiposity. *Sci. Transl. Med.* **2015**, *7* (285), 285ra62–ra285ra62.
- (16) Mapstone, M.; Cheema, A. K.; Fiandaca, M. S.; Zhong, X.; Mhyre, T. R.; MacArthur, L. H.; Hall, W. J.; Fisher, S. G.; Peterson, D. R.; Haley, J. M.; Nazar, M. D.; Rich, S. A.; Berlau, D. J.; Peltz, C. B.; Tan, M. T.; Kawas, C. H.; Federoff, H. J. Plasma Phospholipids Identify Antecedent Memory Impairment in Older Adults. *Nat. Med.* **2014**, *20* (4), 415–418.
- (17) Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's All Relative: Analyzing Microbiome Data as Compositions. *Ann. Epidemiol.* **2016**, 1–8.
- (18) Sinha, R.; Abnet, C. C.; White, O.; Knight, R.; Huttenhower, C. The Microbiome

- Quality Control Project: Baseline Study Design and Future Directions. *Genome Biol.* **2015**, *16*, 276.
- (19) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* **2015**, *3* (March), 23.
- (20) Wishart, D. S. Computational Approaches to Metabolomics. *Methods Mol. Biol.* **2010**, *593*, 283–313.
- (21) Ravel, J.; Gajer, P.; Abdo, Z.; Schneider, G. M.; Koenig, S. S. K.; McCulle, S. L.; Karlebach, S.; Gorle, R.; Russell, J.; Tacket, C. O.; Brotman, R. M.; Davis, C. C.; Ault, K.; Peralta, L.; Forney, L. J. Vaginal Microbiome of Reproductive-Age Women. *Proc. Natl. Acad. Sci.* **2010**, *108* (Supplement_1), 4680–4687.
- (22) Zhou, X.; Brown, C. J.; Abdo, Z.; Davis, C. C.; Hansmann, M. A.; Joyce, P.; Foster, J. A.; Forney, L. J. Differences in the Composition of Vaginal Microbial Communities Found in Healthy Caucasian and Black Women. *ISME J.* **2007**, *112* (10), 121–133.
- (23) Hummelen, R.; Fernandes, A. D.; Macklaim, J. M.; Dickson, R. J.; Chantalucha, J.; Gloor, G. B.; Reid, G. Deep Sequencing of the Vaginal Microbiota of Women with HIV. *PLoS One* **2010**, *5* (8).
- (24) Fredricks, D. N.; Fiedler, T. L.; Marrazzo, J. M. Molecular Identification of Bacteria Associated with Bacterial Vaginosis. *N. Engl. J. Med.* **2005**, *353* (18), 1899–1911.
- (25) McMillan, A.; Rulisa, S.; Sumarah, M.; Macklaim, J. M.; Renaud, J.; Bisanz, J. E.; Gloor, G. B.; Reid, G. A Multi-Platform Metabolomics Approach Identifies Highly Specific Biomarkers of Bacterial Diversity in the Vagina of Pregnant and Non-Pregnant Women. *Sci. Rep.* **2015**, *5*, 14174.
- (26) Zhou, X.; Hansmann, M. A.; Davis, C. C.; Suzuki, H.; Brown, C. J.; Schütte, U.; Pierson, J. D.; Forney, L. J. The Vaginal Bacterial Communities of Japanese Women Resemble Those of Women in Other Racial Groups. *FEMS Immunol. Med.*

Microbiol. **2010**, *58* (2), 169–181.

- (27) Reid, G. Cervicovaginal Microbiomes-Threats and Possibilities. *Trends in Endocrinology and Metabolism*. 2016.
- (28) Macklaim, J. M.; Gloor, G. B.; Anukam, K. C.; Cribby, S.; Reid, G. At the Crossroads of Vaginal Health and Disease, the Genome Sequence of *Lactobacillus iners* AB-1. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108 Suppl* , 4688–4695.
- (29) McMillan, A.; Macklaim, J. M.; Burton, J. P.; Reid, G. Adhesion of *Lactobacillus iners* AB-1 to Human Fibronectin: A Key Mediator for Persistence in the Vagina? *Reprod. Sci.* **2012**, *20* (7), 791–796.
- (30) Walker, A. W.; Duncan, S. H.; Louis, P.; Flint, H. J. Phylogeny, Culturing, and Metagenomics of the Human Gut Microbiota. *Trends in Microbiology*. 2014, pp 267–274.
- (31) Smith, M. I.; Yatsunenko, T.; Manary, M. J.; Trehan, I.; Mkakosya, R.; Cheng, J.; Kau, A. L.; Rich, S. S.; Concannon, P.; Mychaleckyj, J. C.; Liu, J.; Houghton, E.; Li, J. V.; Holmes, E.; Nicholson, J.; Knights, D.; Ursell, L. K.; Knight, R.; Gordon, J. I. Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor. *Science* **2013**, *339* (6119), 548–554.
- (32) Subramanian, S.; Huq, S.; Yatsunenko, T.; Haque, R.; Mahfuz, M.; Alam, M. a; Benezra, A.; DeStefano, J.; Meier, M. F.; Muegge, B. D.; Barratt, M. J.; VanArendonk, L. G.; Zhang, Q.; Province, M. a; Petri, W. a; Ahmed, T.; Gordon, J. I. Persistent Gut Microbiota Immaturity in Malnourished Bangladeshi Children. *Nature* **2014**, *509* (7505), 417–421.
- (33) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (8), 770–781.
- (34) Styczynski, M. P.; Moxley, J. F.; Tong, L. V.; Walther, J. L.; Jensen, K. L.;

- Stephanopoulos, G. N. Systematic Identification of Conserved Metabolites in GC/MS Data for Metabolomics and Biomarker Discovery. *Anal. Chem.* **2007**, *79* (3), 966–973.
- (35) Smith, C. A.; Want, E. J.; O’Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779–787.
- (36) R Development Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* **2015**, *1*, 409.
- (37) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B. Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput. Biol.* **2013**, *9* (7).
- (38) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12580–12585.
- (39) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* **2014**, *42* (W1), 94–99.
- (40) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. LipidBlast in Silico Tandem Mass Spectrometry Database for Lipid Identification. *Nat. Methods* **2013**, *10* (8), 755–758.
- (41) Busardò, F. P.; Jones, A. W. GHB Pharmacology and Toxicology: Acute Intoxication, Concentrations in Blood and Urine in Forensic Cases and Treatment of the Withdrawal Syndrome. *Curr. Neuropharmacol.* **2015**, *13* (1), 47–70.
- (42) DrugCheck GHB Test <http://drugcheck.com/drugcheck-ghb-test> (accessed Jun 9, 2016).
- (43) Kintz, P.; Cirimele, V.; Jamey, C.; Ludes, B. Testing for GHB in Hair by

- GC/MS/MS after a Single Exposure. Application to Document Sexual Assault. *J. Forensic Sci.* **2003**, *48* (1), 195–200.
- (44) Lewis, W. G.; Robinson, L. S.; Gilbert, N. M.; Perry, J. C.; Lewis, A. L. Degradation, Foraging, and Depletion of Mucus Sialoglycans by the Vagina-Adapted Actinobacterium *Gardnerella vaginalis*. *J. Biol. Chem.* **2013**, *288* (17), 12067–12079.
- (45) Petrova, M. I.; van den Broek, M.; Balzarini, J.; Vanderleyden, J.; Lebeer, S. Vaginal Microbiota and Its Role in HIV Transmission and Infection. *FEMS Microbiol. Rev.* **2013**, *37* (5), 762–792.
- (46) Hashemi, F. B.; Ghassemi, M.; Roebuck, K. A.; Spear, G. T. Activation of Human Immunodeficiency Virus Type 1 Expression by *Gardnerella vaginalis*. *J. Infect. Dis.* **1999**, *179* (4), 924–930.
- (47) Schuyler, J. A.; Mordechai, E.; Adelson, M. E.; Sobel, J. D.; Gygyax, S. E.; Hilbert, D. W. Identification of Intrinsically Metronidazole-Resistant Clades of *Gardnerella vaginalis*. *Diagn. Microbiol. Infect. Dis.* **2016**, *84* (1), 1–3.
- (48) Martinez, R. C. R.; Franceschini, S. A.; Patta, M. C.; Quintana, S. M.; Gomes, B. C.; De Martinis, E. C. P.; Reid, G. Improved Cure of Bacterial Vaginosis with Single Dose of Tinidazole (2 G), *Lactobacillus rhamnosus* GR-1, and *Lactobacillus reuteri* RC-14: A Randomized, Double-Blind, Placebo-Controlled Trial. *Can. J. Microbiol.* **2009**, *55* (2), 133–138.
- (49) Macklaim, J. M.; Clemente, J. C.; Knight, R.; Gloor, G. B.; Reid, G. Changes in Vaginal Microbiota Following Antimicrobial and Probiotic Therapy. *Microb. Ecol. Health Dis.* **2015**, *26*, 27799.
- (50) Saunders, S.; Bocking, A.; Challis, J.; Reid, G. Effect of *Lactobacillus* Challenge on *Gardnerella vaginalis* Biofilms. *Colloids Surf. B. Biointerfaces* **2007**, *55* (2), 138–142.

- (51) McMillan, A.; Dell, M.; Zellar, M. P.; Cribby, S.; Martz, S.; Hong, E.; Fu, J.; Abbas, A.; Dang, T.; Miller, W.; Reid, G. Disruption of Urogenital Biofilms by Lactobacilli. *Colloids Surfaces B Biointerfaces* **2011**, *86* (1), 58–64.
- (52) Reid, G. Probiotic and Prebiotic Applications for Vaginal Health. *J. AOAC Int.* *95* (1), 31–34.
- (53) Srinivasan, S.; Liu, C.; Mitchell, C. M.; Fiedler, T. L.; Thomas, K. K.; Agnew, K. J.; Marrazzo, J. M.; Fredricks, D. N. Temporal Variability of Human Vaginal Bacteria and Relationship with Bacterial Vaginosis. *PLoS One* **2010**, *5* (4).

Appendix A: Content License from Scientific Reports

(available at <http://www.nature.com/srep/journal-policies/editorial-policies#license-agreement>)

SCIENTIFIC REPORTS



License agreement and author copyright

Scientific Reports does not require authors to assign copyright of their published original research papers to the journal. Articles are published under a [CC BY license](#) (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor ([read full legal code](#)).

Alternative Creative Commons licenses are available on request: authors should contact the editorial office (scientificreports@nature.com) on acceptance of an article to discuss these options.

Visit our open research site for more information about [Creative Commons licensing](#).

Appendix B: Content License from Reproductive Sciences

(Available at <https://us.sagepub.com/en-us/nam/journals-permissions>)



DISCIPLINES PRODUCTS RESOURCES ABOUT

Search: keyword, title, author, ISBN



Cart 0

Authors Reusing Their Own Work

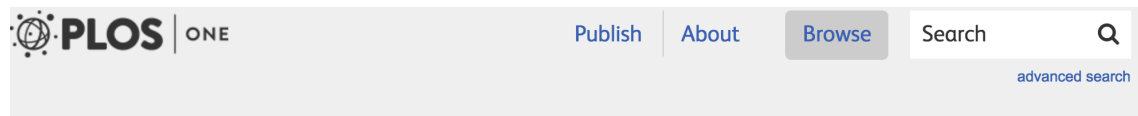
SAGE Journal authors are able to use their article in certain circumstances without any further permission. The chart above includes common requests and an explanation of which 'version' of the article can be used in each circumstance.

- Version 1 – original submission to the journal (before peer review)
- Version 2 – original submission to the journal with your revisions after peer review, often the version accepted by the editor
- Version 3 – copy-edited and typeset proofs and the final published version

I WANT TO:	CLEARED	REQUIRES PERMISSIONS
supply my article to my students or use the article for teaching purposes	Version 3	
supply my article to a research colleague at an academic institution	Version 3	
supply my article to a commercial organization for republication, distribution or a web posting		X
upload my article to my institution repository or department website	Version 2	
upload my article to a repository NOT affiliated with my institution, 12 months after publication	Version 2	

Appendix C: Content License from PLoS ONE

(Available at <http://journals.plos.org/plosone/s/content-license>)



Using PLOS Content
Figures, Tables, and Images
Data
Submitting Copyrighted or
Proprietary Content

Content License

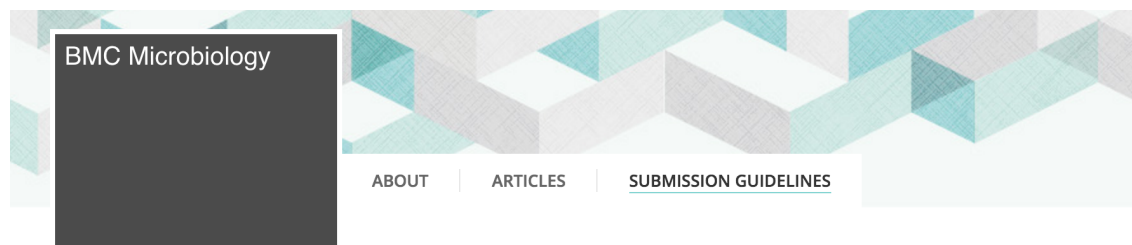
The following policy applies to all of PLOS journals, unless otherwise noted.

PLOS applies the [Creative Commons Attribution \(CC BY\) license](#) to works we publish. This license was developed to facilitate open access – namely, free immediate access to, and unrestricted reuse of, original works of all types.

Under this license, authors agree to make articles legally available for reuse, without permission or fees, for virtually any purpose. Anyone may copy, distribute or reuse these articles, as long as the author and original source are properly cited.

Appendix D: Content License from BMC Microbiology

(Available at <https://bmcmicrobiol.biomedcentral.com/submission-guidelines/copyright>)



[Aims and scope](#)

[Fees and funding](#)

[Language editing services](#)

[Copyright](#)

[Prepare your manuscript](#)

[Prepare supporting information](#)

[Conditions of publication](#)

Copyright

- Copyright on any open access article in a journal published by BioMed Central is retained by the author(s).
- Authors grant BioMed Central a [license](#) to publish the article and identify itself as the original publisher.
- Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified.
- The [Creative Commons Attribution License 4.0](#) formalizes these and other terms and conditions of publishing articles.

Appendix E: Content License from Frontiers in Bioengineering and Biotechnology

(Available at <http://journal.frontiersin.org/journal/bioengineering-and-biotechnology#about>)



Copyright Statement

Under the **Frontiers Conditions for Website Use** and the **Frontiers General Conditions for Authors**, authors of articles published in Frontiers journals retain copyright on their articles, except for any third-party images and other materials added by Frontiers, which are subject to copyright of their respective owners. Authors are therefore free to disseminate and re-publish their articles, subject to any requirements of third-party copyright owners and subject to the original publication being fully cited. Visitors may also download and forward articles subject to the citation requirements and subject to any fees Frontiers may charge for downloading licenses. The ability to copy, download, forward or otherwise distribute any materials is always subject to any copyright notices displayed. Copyright notices must be displayed prominently and may not be obliterated, deleted or hidden, totally or partially.

[Field Chief Editors](#)

[Mission Statement](#)

Appendix F: Ethics Approval for studies

UWO Ethical approval for study in chapter 2



Use of Human Participants - Ethics Approval Notice

Research Ethics

Principal Investigator:Dr. Gregor Reid
File Number:102886
Review Level:Full Board
Approved Local Adult Participants:0
Approved Local Minor Participants:0
Protocol Title:Pilot study investigating bacteria (microbiota), their by-products (metabolome) and environmental toxins in relation to reproductive health in Rwandan women.
Department & Institution:Schulich School of Medicine and Dentistry/Microbiology & Immunology,Western University
Sponsor:UWO

Ethics Approval Date:December 11, 2012
Ethics Expiry Date:September 30, 2013

Documents Reviewed & Approved & Documents Received for Information:

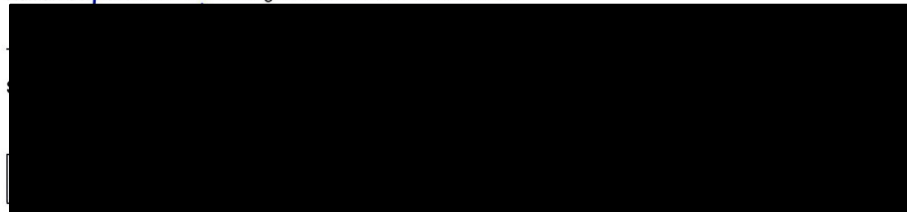
Document Name	Comments	Version Date
Western University Protocol		
Letter of Information & Consent	Sample Collection Group Only	
Letter of Information & Consent	Probiotic Trial Group	

This is to notify you that the University of Western Ontario Health Sciences Research Ethics Board (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced study on the approval date noted above. The membership of this HSREB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request form.

Member of the HSREB that are named as investigators in research studies, or declare a conflict of interest, do not participate in discussions related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.



This is an official document. Please retain the original in your files.

University of Kigali Teaching Hospital (CHUK) approval for study in chapter 2



CENTRE HOSPITALIER UNIVERSTAIRE
UNIVERSITY TEAHING HOSPITAL

Centre Hospitalier Universitaire de Kigali

Ethics Committee / Comité d'éthique

July 13th 2012

Ref.: EC/CHUK/035/12

Review Approval Notice

Dear Amy McMillan,

Your research project: "Pilot Study Investigating Bacteria (Microbiota), their by-products (Metabolome) and Environmental Toxins in Relation to Reproductive Health in Rwandan Women"

During the meeting of the Ethics Committee of Kigali University Teaching Hospital (KUTH) that was held on 13/07/2012 to evaluate your protocol of the above mentioned research project, we are pleased to inform you that the Ethics Committee/CHUK has approved your protocol. You are required to present the results of your study to KUTH Ethics Committee before publication.

PS: Please note that the present approval is valid for 12 months.

Yours sincerely,



Dr George NTAKIYIRUTA.
The Vice President, Ethics Committee,
Kigali University Teaching Hospital

<<University teaching hospital of Kigali Ethics committee operates according to standard operating procedures (Sops) which are updated on an annual basis and in compliance with GCP and Ethics guidelines and regulations>>

B.P. :655 Kigali- RWANDA Tél. Fax : 00 (250) 576638E-mail : chuk.hospital@chukigali.org

UWO Ethical approval for study in chapter 3



Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Gregor Reid
File Number:100751
Review Level:Delegated
Approved Local Adult Participants:100
Approved Local Minor Participants:0
Protocol Title:Investigating small molecules produced by naturally occurring bacteria of the vaginal tract - 18203E
Department & Institution:Schulich School of Medicine and Dentistry/Microbiology & Immunology,Western University
Sponsor:
Ethics Approval Date:June 15, 2012 **Expiry Date:**June 30, 2016
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Revised Western University Protocol	Revised study methods, participant recruitment, eligibility of participants and administrative changes.	
Other	Instrument - Participant Questionnaire	
Other	Instrument - Vaginal Examination - Clinician Questionnaire	
Revised Letter of Information & Consent		2012/04/26

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

This is an official document. Please retain the original in your files.

University of Ibadan Ethical approval for study in chapter 5



INSTITUTE FOR ADVANCED MEDICAL RESEARCH AND TRAINING (IMRAT)
COLLEGE OF MEDICINE, UNIVERSITY OF IBADAN, IBADAN, NIGERIA.
 E-Mail - imratcomui@yahoo.com




UI/UCH EC Registration Number: NHREC/05/01/2008a

NOTICE OF FULL APPROVAL AFTER FULL COMMITTEE REVIEW

Re: Non-Invasive Biomarkers of Enteropathy in Infants and Children with Severe Acute Malnutrition; a Pilot Study

UI/UCH Ethics Committee assigned number: UI/EC/11/0067

Name of Principal Investigator: **Prof. O. O. Akinyinka**

Address of Principal Investigator: 

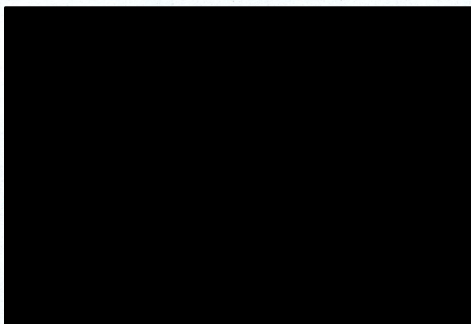
Date of receipt of valid application: 25/03/2011

Date of meeting when final determination on ethical approval was made: N/A

This is to inform you that the research described in the submitted protocol, the consent forms, and other participant information materials have been reviewed and *given full approval by the UI/UCH Ethics Committee.*

This approval dates from 03/06/2011 to 02/06/2012. If there is delay in starting the research, please inform the UI/UCH Ethics Committee so that the dates of approval can be adjusted accordingly. Note that no participant accrual or activity related to this research may be conducted outside of these dates. *All informed consent forms used in this study must carry the UI/UCH EC assigned number and duration of UI/UCH EC approval of the study.* It is expected that you submit your annual report as well as an annual request for the project renewal to the UI/UCH EC early in order to obtain renewal of your approval to avoid disruption of your research.

The National Code for Health Research Ethics requires you to comply with all institutional guidelines, rules and regulations and with the tenets of the Code including ensuring that all adverse events are reported promptly to the UI/UCH EC. No changes are permitted in the research without prior approval by the UI/UCH EC except in circumstances outlined in the Code. The UI/UCH EC reserves the right to conduct compliance visit to your research site without previous notification.



Research Units: ■ Genetics & Bioethics ■ Malaria ■ Environmental Sciences ■ Epidemiology Research & Service
 ■ Behavioural & Social Sciences ■ Pharmaceutical Sciences ■ Cancer Research & Services ■ HIV/AIDS

Curriculum Vitae

Name: Amy McMillan

Post-secondary University of Western Ontario

Education and London, Ontario, Canada

Degrees: 2007-2011 H.B.Sc.

Honours and Frederick Banting and Charles Best in Canada

Awards: Graduate Scholarship Doctoral Award 2014-2017 (National Award,
\$105 000 over 3 years)

Metabolomics Society Student Travel Award

12th Annual International Conference of the Metabolomics Society,
Dublin, Ireland, 2016 (\$500).

Ontario Graduate Scholarship

2014 (Ontario Government Award, \$15 000, Award Declined)

Ontario Graduate Scholarship

2013 (Ontario Government Award, \$15 000)

CIDA Students for Development Grant

2012 (Funded by the Canadian International Development Agency,
\$6 500)

Ontario Graduate Scholarship

2012 (Ontario Government Award, \$15 000)

eBiosciences Award of Excellence in Graduate Poster
 Nov 4, 2011 (Infection and Immunity Research Forum
 (IIRF), \$100)

**Related Work
 Experience**

Teaching Assistant
 The University of Western Ontario
 Biology of Infection and Immunity (Micro 2500A)
 Sept 2013-Dec 2013

Teaching Assistant
 The University of Western Ontario
 First year General Biology laboratory
 Sept 2011-Dec 2014

Teaching Assistant
 The University of Western Ontario
 Biological and Social Determinants of Infectious Disease
 Jan 2016-April 2016

Research Assistant
 Lawson Health Research Institute
 Jan 2010-Sept 2011

Publications:

1. **McMillan A**, Rulisa S, Sumarah M, Macklaim JM, Renaud J, Bisanz JE, Gloor GB, and G Reid. 2015. A multi-platform metabolomics approach identifies highly specific biomarkers of bacterial diversity in the vagina of pregnant and non-pregnant women. *Sci Rep* 5:14174. doi: 10.1038/srep14174

Contribution: Participated in study design, supervised patient recruitment and sample collection, performed sample processing, developed methods for metabolite profiling by GC-MS and LC-MS, analyzed and interpreted microbiome and metabolome data, performed *in vitro* experiments and wrote the manuscript. 90% contribution.

2. Urbaniak C, **McMillan A**, Angelini M, Gloor GB, Sumarah M, Burton JP. and G Reid. 2014. Effect of chemotherapy on the microbiota and metabolome of human milk, a case report. *Microbiome* 2: 24- 2618-2-24.

Contribution: Developed methods for metabolite profiling of breast milk by GC-MS. Performed metabolomic analysis and contributed to manuscript generation. 30% contribution.

3. Bisanz JE, Seney S, **McMillan A**, Vongsa R, Koenig D, Wong L, Dvoracek B, Gloor GB, Sumarah M, Ford B, Herman D, Burton JP, and G Reid. 2014. A systems biology approach investigating the effect of probiotics on the vaginal microbiome and host responses in a double blind placebo-controlled clinical trial of post-menopausal women. PLoS ONE 9(8): e104511.

Contribution: Performed metabolomic analysis of vaginal samples and analyzed the resulting data. Performed DNA extraction and PCR of vaginal samples for 16S rRNA gene sequencing and participated in data analysis. 15% contribution.

4. Reid G, Nduti N, Sybesma W, Kort R, Kollmann TR, Adam R, Brown EM, Einerhand A, El-Nezami H, Gloor GB, Kavere II, Lindahi J, Manges A, Mamo W, Martin R, **McMillan A**, Obiero J, Ochieng PA, Onyango A, Rulisa S, Salminen E, Salminen S, Sije A, Swann JR, van Treuren W, Waweru D, and SJ Kemp. 2014. Harnessing microbiome and probiotic research in sub-Saharan Africa: recommendations from an African workshop. *Microbiome* 2, 12-2618-2-12.

Contribution: Participated in working groups and generation of recommendations presented in the manuscript. 5% contribution.

5. **McMillan A**, Macklaim MM, Burton JP, and G Reid. 2012. Adhesion of *Lactobacillus iners* AB-1 to human fibronectin: a key mediator for persistence in the vagina? *Reprod Sci* 20 (7): 791-796.

Contribution: Performed *in vitro* experiments, analyzed the resulting data and co-wrote the manuscript. 90% contribution.

6. Hummelen R, Macklaim MM, Bisanz JE, Hammond J, **McMillan A**, Vongsa B, Koenig D, Gloor GB, and G Reid. 2011. Vaginal microbiome diversity and epithelial cell changes in post-menopausal women with dryness and atrophy. *PLoS ONE* 6(11):e26602.

Contribution: Extracted DNA from vaginal samples, performing 16S rRNA gene PCR, and standardizing PCR product concentrations across all samples by quantification of DNA electrophoresis band intensity. 10% contribution.

7. **McMillan A**, Dell M, Zellar MP, Cribby S, Martz S, Hong E, Fu J, Abbas A, Dang T, Miller W, and G Reid. 2011. Disruption of urogenital biofilms by lactobacilli. *Colloids Surf B Biointerfaces* 86(1): 58- 64.

Contribution: Responsible for compiling data from multiple studies and co-writing the manuscript in full. 50% contribution.

Publications in progress:

8. **McMillan A**, Orimadegun A, Sumarah M, Renaud J, Muc M, Gloor GB, Akinyinka OO, Reid G, and S Allen. 2016. Metabolic derangements identified through untargeted metabolomics in a cross-sectional study of Nigerian children with severe acute malnutrition. *J Nutr*. Under Review.

Contribution: Developed methods for metabolite profiling of stool and plasma by GC-MS and LC-MS, processed samples and analyzed the resulting metabolomic data, performed stool ELISAs for quantification of inflammatory proteins, performed DNA extractions, PCR, and data analysis microbiome profiling of stool, and co-wrote the manuscript. 70% contribution.

9. **McMillan A**, Renaud J, Reid G, and Sumarah MW. Post-acquisition artifact filtering for LC-MS based human metabolomic studies. *J. Cheminform.* Under Revisions..

Contribution: Developed method to identify and remove salt cluster artifacts in LC-MS based metabolomics data using mass defect filtering. Conducted all experiments, developed R code, and co-wrote the manuscript. 90% contribution.

Patents:

Application 62/146,094, Use of gamma-hydroxybutyrate, 2-hydroxyisovalerate and their tyrosine ratios for the diagnosis of Bacterial Vaginosis.