A proteomics-based approach to understanding female genital immunity and the relationship to the vaginal microbiome

By

Michelle Perner

A Thesis submitted to the Faculty of Graduate Studies of

University of Manitoba

In partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

Winnipeg

Abstract

Vaginal dysbiosis, characterized by *Lactobacillus* depletion and an overgrowth of facultative anaerobes, can result in the clinical condition called bacterial vaginosis (BV). BV affects 30% of women globally, affects their quality of life, and associates with an increased risk of STI's, including HIV, and reproductive complications. In some BV cases there is an increased immune response, however this is not observed in all women, indicating there may be individual factors or particular bacteria that may be responsible for negative responses. More research is needed to understand the molecular mechanisms of BV. The development of advanced tools to characterize host and microbiome information offers an opportunity to understand host-bacterial relationships important for disease pathogenesis or prevention. Here we developed a mass-spectrometry based proteomics workflow to simultaneous obtain host and functional microbial data in a large-scale analysis of cervicovaginal lavage in a human cohort. Bacterial proteome taxa profiles were used to sort women into bacterial groups, then differential host and bacterial functions were assessed. Non-*Lactobacillus* communities showed decreased epithelial barrier associated proteins and increased immune cell recruitment, particularly with the important innate leukocytes' neutrophils. Bacterial function in non-*Lactobacillus* communities showed enriched pathogenic associated pathways, such as biofilm formation and cell motility, and different carbohydrate metabolism pathways compared to *L. crispatus* dominant communities. *In vitro* experiments were used to better understand host-bacterial interactions. Experiments using vaginal epithelial cell lysate identified decreased barrier proteins in *Gardnerella*, *Mobiluncus* and *Prevotella* supernatant exposed cells compared to *L. crispatus* supernatant exposed cells. Neutrophil experiments identified increased neutrophil migration and activity upon exposure to *Mobiluncus* and not *Lactobacillus crispatus*. This study included development of a workflow

that would allow for metaproteomic analysis in large cohorts that can be applied to future studies for characterization of host bacterial interactions at mucosal surfaces. This method was then used to identify barrier disruption and neutrophil signatures in BV associated communities and created a new hypothesis of neutrophil activation as an inducer for vaginal inflammation and barrier damage observed in some cases of bacterial vaginosis, and *Mobiluncus* as one of the key bacteria to induce neutrophil migration and activation.

Acknowledgements

Thank you first and foremost to my primary supervisor Dr. Adam Burgener for the opportunity to join your lab and support to pursue my graduate degree, first as a master's student, then as a PhD student. Thank you for your mentoring in both science, networking, and good food and wine. Thank you for providing me the opportunity to travel and present my work and especially for your continuous reminders of the importance of the people behind the samples we are fortunate to work with. It has been an honour to work with you and I wish you all the best in your new position at Case Western Reserve University.

Thank you to my co-supervisor Dr. Xiao-Jian Yao for your advice, wisdom, and welcoming smile. You always had insightful questions that pushed me to be a better scientist. It has been a privilege to work with such a distinguished HIV virologist.

Thank you to my wonderful committee members Dr. Yoav Keynan, Dr. Lyle McKinnon, and Dr. Richard Sparling. Your comments and advice have pushed me to think critically about each step of my research, consider all aspects of my project, and the wider implications.

Thank you to all the members of the Burgener lab, both past and present, for their comradery and scientific help. In particular, thank you to Kenzie and Laura for their invaluable advice and help during data analysis of this dataset, without them this degree would have taken much longer. Thank you to Max and Jen for your help processing the samples. Thank you to Sarah and Sam for your statistics help and turning me into the stats person in my journal club. Thank you to Dr. Alicia Berard and Dr. Christina Farr Zuend for your assistance with my *in vitro* experiments.

A huge thank you to Dr. Garett Westmacott and Stuart McCorrister at the Mass Spectrometry and Proteomics core at the Public Health Agency of Canada for running samples on the mass spectrometers.

Thank you to the team at CAPRISA who developed, and ran the CAPRISA-004 trial, in particular Dr. Salim S Abdool Karim, Dr. Quarraisha Abdool Karim, Leila Mansoor, Anneke Grobler, Natasha Samsunder, and Jo-Ann Passmore. Thank you to the women who participated in the CAPRISA-004 clinical trial for their willingness to both test something that went into their vagina and provide samples that could be studied.

To the many graduate students I have had the pleasure of meeting throughout the years, thank you for the hours of scientific discussions, commiseration, laughs, and support during these past several years. I wish you all a successful thesis defense and career.

Thank you to the amazing administrative assistants including Angela Nelson, Jude Zieske, and Ilda Medeiros for always finding a solution to the administrative problems I faced.

Thank you to the funding agencies that provided me the ability to complete this work: Research Manitoba, Canadian Institute of Health Research, and the University of Manitoba.

Last, but not least, I would like to thank my friends and family for all their support while I completed this degree. To my wonderful fiancé Mike Lemire, words cannot express how much I appreciate and love you for your unending patience and support, but especially for the ability to always make me laugh. Thank you to my parents, Ann and Gerhard Perner, for their love and support while I have pursued my education and instilling in me a good work ethic, an eye for detail, and sense of humour that have served me well.

Dedication

This study is dedicated to all individuals who have a vagina, including ciswomen, transwomen transmen, and non-binary individuals. Too often in societies there is a taboo about vaginas, which can have numerous consequences. I hope that society continues progressing in breaking these taboos, and that this thesis provides useful information to aid in improving vaginal health.

Table of Contents

List of Tables:

List of Figures:

List of abbreviations:

A1BG: Alpha 1B-Glycoprotein

A2M: Alpha 2 macroglobulin

ABC: ATP Binding Cassette

AFM: Afamin

AGT: Angiotensin

AIDS: Acquired Immunodeficiency Syndrome

ALB: Albumin

APC: Antigen Presenting Cell

APOA1: Apolipoprotein A1

ARP-WASP: Actin Related Protein-Wiskott-Aldrich Syndrome Family Protein

ATP: Adenosine Triphosphate

AV: Aerobic Vaginitis

BAG2: B-cell Lymphoma 2 Associated Ananthogene Family Molecular Chaperone Regulator 2

BCA: Bicinchoninic Acid

BH: Benjamini-Hochberg

BLASTP: Basic Local Alignment Search Tool for Proteins

BP-Biological Process

BSA: Bovine Serum Albumin

BV: Bacterial Vaginosis

C3: Complement factor 3

CAPRISA: Centre for the AIDS Programme of Research in South Africa

CCR5: C-C Chemokine Receptor Type 5

CD: Cluster of Differentiation

CI: Confidence Interval

COC: Combined Oral Contraceptive

ComBat: Combining Batches

CoV: Coefficient of Variation

CST: Community State Type

CT: Cervicotype

CVL: Cervicovaginal Lavage

DAVID: Database for Annotation, Visualization and Integrated Discovery

DC: Dendritic Cells

DMPA: Depomedroxyprogesterone Acetate

DNA: Deoxyribonucleic Acid

DTT: Dithiothreitol

ECM: Extracellular Matrix

EIF2: Eukaryotic Initiation Factor 2

ERK: Extracellular Signal-Regulated Kinase

ES: Enrichment Score

FBS: Fetal Bovine Serum

FD: Fold Difference

FDR: False Discovery Rate

FGT: Female Genital Tract

HEPES: 4-(2-Hydroxyethyl)-1-Piperazineethanesolfonic Acid

HIV: Human Immunodeficiency Virus

HPX: Hemopexin

HSV2: Herpes Simplex Virus Type 2

GSEA: Gene Set Enrichment Analysis

IAA: Iodoacetamide

IFN: Interferon

IL: Interleukin

ILK: Integrin-linked kinase

IPA: Ingenuity Pathway Analysis

IQR: Interquartile Range

KEGG: Kyoto Encyclopedia of Genes and Genomes

KSF: Keratinocyte Serum-Free Media

LPS: Lipopolysaccharide

LXR/RXR: Liver X receptor/ Retinoid X receptor

MAIT: Mucosal Associated Invariant T cells

MAPK: Mitogen-Activated Protein Kinases

MHC: Major Histocompatibility Complex

MIP: Macrophage Inflammatory Proteins

MPO: Myeloperoxidase

MS: Mass Spectrometry

NA: Normalized Abundance

NCBI: National Center for Biotechnology Information

NES: Normalized Enrichment Score

NET-EN: Norethisterone Enanthate

NFκB: Nuclear Factor κB

NIK: NFκB Inducing Kinase

NK: Natural Killer cells

NO: Nitric Oxide

NRF-2: Nuclear Factor Erythroid 2-Related Factor 2

OR: Odds Ratio

ORM: Orosomucoid Protein

PAMP: Pathogen Associated Molecular Pattern

PI3K-AKT: Phosphatidylinositol 3-Kinase Protein AKT (also known as Protein Kinase B)

RhoA: Ras Homolog Family Member A

RhoGDI: Rho GDP (Guanine Diphosphate) Dissociation Inhibitor

ROS: Reactive Oxygen Species

RPMI: Roswell Park Memorial Institute Media

RNA: Ribonucleic Acid

rRNA: Ribosome Ribonucleic Acid

SD: Standard Deviation

SDS: Sodium Dodecyl Sulfate

SERPIN: Serine Protease Inhibitor

SPRR: Small Proline Rich proteins

STI: Sexually Transmitted Infection

TcR: T cell Receptor

TFU/ml: Total Cell Units/ ml

THOP1: Thimet Oligopeptidase 1

TLR: Toll-like Receptor

TNF: Tumour Necrosis Factor

TrEMBL: Translated European Molecular Biology Laboratory

tRNA: Transfer Ribonucleic Acid

UEB: Urea Exchange Buffer

VIRGO: Human Vaginal Non-redundant Gene Catalog

VTDB: Vitamin D binding protein

VTNC: Vitronectin

WNT: Wingless-related Integration Site

WST-1: (2-(4-iodophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium, monosodium salt)

Chapter 1. Introduction

1.1 Female Genital Tract Tissue Structure

Within the female genital tract (FGT) there is variation in how the tissue structure is organized. In the vagina the epithelial cells are organized in a multilayered squamous structure. Closest to the vaginal lumen, the cornified envelope contains keratinized, dead, flattened epithelial cells and many cross-linked proteins (*1*). These cells contain glycogen and are semi-permeable to both bacteria and virus, as well as host immune mediators, providing an opportunity for the vaginal microbiome to obtain food, and for an immune response to be mounted (*2*). Below the cornified envelope are keratinized cells that are metabolically inactive, with the majority of the proteins being keratins or their related proteins (*3*). Glycogen can also be stored in these cells, with storage induced by expression of estrogen. The supra basal cell layers underneath the keratinized cells contain metabolically active epithelial cells, which are under the regulation of estrogen produce glycogen (*2*). The basal layer is just above the lamina propria and contains the most metabolically active epithelial cells. The lamina propria is a thin layer of connective tissue cells, including stromal fibroblasts, that provide structural support (*4*). Interspersed in the epithelial layers dendritic cells (DC's) can extend pseudopods to sample antigens from the lumen and can retract their pseudopods to the lamina propria where there are immune cells to which the antigens can be presented (*5*). Tight junctions in between epithelial cells provide a semi-permeable barrier, though tight junctions are absent in the cornified envelope (*6*). There is a high turnover rate of the epithelial layers, with one layer sloughed off every 4 hours on average (*7*). As the cells slough off, they burst, releasing their glycogen into the vaginal mucus, and these cells can function as a decoy for pathogens to bind to instead of attached cells (*7*).

The organization of epithelial layers in the vagina shifts over the menstrual cycle. During high levels of estrogen, like in the proliferative phase, there are more epithelial layers, and glycogen production is increased, compared to the high progesterone secretory phase (*7*). Initiation of some hormonal contraceptives can affect the epithelia. Use of depot medroxyprogesterone acetate (DMPA), a high progesterone injectable contraceptive, decreased the number of layers and thickness of the epithelia, as well as reducing the thickness of glycogen positive epithelia (*8*), though no changes were observed for women who used combined oral contraceptive pills (COC) (*9*).

At the cervix there is a transition zone where the squamous epithelia (ectocervix) changes to a single layer of columnar cells (endocervix), with no keratinocyte or cornified envelope layer. This area is more susceptible to mechanical stress and tears in the epithelium compared to the multilayered squamous epithelium. Cervical ectopy is a condition where the ectocervix has the more susceptible single columnar epithelial layer spread beyond the endocervix, and epithelial cells are visibly red and inflamed. This increased inflammation may enhance vulnerable to STI's (*10, 11*). This has been associated with increased levels of pro-inflammatory cytokines and chemokines in adolescents (*12, 13*), including interleukin (IL)-1β, IL-8, and macrophage inflammatory protein (MIP)-1α, which have been linked to increased HIV acquisition (*14*).

1.2 Microbiome

1.2.1 The Human Microbiome

Bacteria cover every part of our body that is exposed to the environment, including the gastrointestinal and genital tracts. The collection of all bacteria, virus, fungi, protozoa, their genes, proteins and metabolites are defined as the microbiome. Each part of the body has a unique environment that facilitates the growth of specific microbial communities. The

microbiome has important roles in human health, from preventing growth or adherence of pathogens, metabolism of nutrients humans cannot process, and producing compounds that the human body requires (*15*).

The gut contains the highest concentration of bacterial cells in the human body and plays many roles in human health, including breaking down nutrients to provide energy, interacting with the brain with bacterial metabolites, and influencing fat storage in the body (*16*). Alterations in the gut microbiome in particular have been associated with a myriad of conditions including inflammatory bowel disease, psychiatric disorders like depression and anxiety, Parkinson's disease, stress, obesity, and arthritis susceptibility (*17-25*). Understanding what bacteria are beneficial and what integral functions they provide at different body sites is essential to improving human health.

1.2.2 The Vaginal Microbiome

The vaginal microbiome is different compared to most body site microbiomes, as optimal microbiomes are simple, often dominated by a single species, compared to a complex community with multiple species from many genera as is observed in the gut. An optimal vaginal microbiome generally consists of a dominant single bacterial species from the genus *Lactobacillus*, first observed by Doderlein in 1894, when he observed the gram-positive *Lactobacillus* in the vaginas of healthy women (*26*). *Lactobacillus* species that dominate the vagina most commonly are *L. iners*, *L. jensenii*, *L. crispatus*, and *L. gasseri* (*27*). Commensal bacteria are beneficial to the vaginal environment through prevention of other bacteria colonizing, and they form a barrier to the vaginal epithelia by aggregating and adhering to mucus and the epithelial cells, blocking other bacteria through steric hindrance (*28-32*). Many *Lactobacillus* species are capable of metabolizing glycogen derived products under anaerobic

conditions to make lactic acid, creating a positive feedback loop that facilitates growth of *Lactobacillus.* The low pH from lactic acid production inhibits or inactivates potential pathogens and non-*Lactobacillus* growth, indeed lactic acid production and a low pH is a hallmark of an optimal vaginal microbiome (*33-39*). Though an optimal vaginal microbiome will generally be dominated by a particular *Lactobacillus* species, low levels of other bacteria remain, which if there is a precipitating event, can overgrow the *Lactobacillus* species. Bacterial overgrowth of non-*Lactobacillus* species results in a transition to a dysbiotic microbiome, which can potentially lead to a clinical diagnosis of aerobic vaginitis (AV) or more commonly bacterial vaginosis (BV). Though these both can have a negative effect on the vaginal environment they have divergent symptoms and causes. Many women with non-*Lactobacillus* communities show no symptoms or signs.

Growth of facultative or strict anaerobes after depletion of *Lactobacillus* occurs in BV. The resulting community can contain a myriad of different bacteria including *Gardnerella, Prevotella*, *Mobiluncus*, *Atopobium,* and *Megaspheara*. Typical signs of inflammation like redness, swelling and pain are not always found in BV, however pro-inflammatory responses can occur due to BV-associated bacterial colonization, even without clinical signs (*40*). A prevalent complaint of BV is a malodourous discharge, which is due to the increased production of amine compounds. Increased pH often occurs as there is less lactic acid produced. Diagnosis of BV is through use of the Nugent score or Amsel's criteria.

Aerobic vaginitis (AV) is dysbiosis that results from overgrowth of non-*Lactobacillus* enteric associated bacteria. This includes bacteria such as *Staphylococcus*, *Streptococcus* and enteric bacteria like *Escherichia* and *Enterococcus*. Clinical symptoms or signs include inflamed reddish vaginal mucosa, abundant yellow discharge, and epithelial disruption (*41*). One notable

difference in AV is the lack of a fishy amine odour that is a signature symptom of BV. This condition is described as more inflammatory, with the most severe grade of AV described as desquamative inflammatory vaginitis. AV can be diagnosed through a vaginal smear observed under a phase contrast microscope and a scoring system. AV is less prevalent than BV, ranging from 7-12% of individuals, and often associated with postmenopausal women (*42*).

1.2.3 Prevalence of BV & Vaginal Dysbiosis

Prevalence of bacterial vaginosis varies across the world and across various demographic variables. Approximately 30% of women globally are afflicted with BV, ranging from 6% in Burkina Faso (*43, 44*) to upwards of 50% in some studies in South Africa (*45*), though a recent meta-analysis found that the rates of BV using studies from the general population to be more consistent, ranging from 23-29% in different global regions (*46*). The prevalence of BV is highest in Africa, with three study reporting a >50% prevalence of BV in South Africa (*47-49*), with a more recent study finding a 34% prevalence rate (*50*). However not all regions in Africa have a high prevalence of BV, with lower prevalence observed in western African countries (*43, 44, 51*). Lower BV prevalence has been observed in Asian countries (*45*). The lowest BV levels were found in Australia, New Zealand, and western Europe (*52-59*). Prevalence is relatively high in North America, with studies identifying a prevalence between 20-40% (*60, 61*). Much of the research used Nugent score to identify individuals with BV, which results in more BV positive women than when using Amsel's criteria. Use of molecular methods has found that many women with a negative Nugent score result can still have a dysbiotic microbiome (*62*). Importantly, many of those diagnosed display no symptoms or signs of BV.

Ethnicity has also been found to associate with variation in vaginal microbiomes. A study by Ravel et al in 2011 of 400 American women displaying no signs or symptoms of vaginal

dysbiosis found that Caucasian and Asian American women had a high proportion who were dominated by *Lactobacillus* species, mainly *L. crispatus, L. gasseri,* or *L. jensenii,* while Hispanic and African heritage women had a higher proportion of women who had non-*Lactobacillus* dominant microbiomes (*27*). Among the Hispanic and African American women who were dominated by a *Lactobacillus* species, *L. iners* was more prevalent compared to Asian and Caucasian American women (*27*). In a study of South African women, the majority of whom were black, there was a high prevalence of dysbiotic communities, and the majority of women with *Lactobacillus* dominance had *L. iners* as the most prevalent *Lactobacillus* species (*63*).

1.2.4 History of Vaginal Dysbiosis

Initial findings by Doderlein identified *Lactobacillus* in the vaginas of healthy women. Research in 1914 found that *Bacteroides* black curved rods, which we now know were *Mobiluncus,* and anaerobic cocci were associated with abnormal vaginal discharge (*64*). In 1921 Schroder categorized the vaginal flora with Gram stains into least pathogenic, intermediate and most pathogenic, with a decrease of *Lactobacillus* observed with increase in pathogenicity, similar to what we currently call the Nugent scoring system (*65*). Decades later, there was further confirmation of the lack of *Lactobacillus* and presence of *Bacteroides* associated with non-specific vaginitis. The potential role of male partners in non-specific vaginitis was first investigated in 1953, and found that non-motile, pleomorphic gram negative rods were present in both the urethral charge of men with prostatitis and discharge of women with cervicitis (*66*).

Gardner and Dukes published a paper that identified a new organism, named *Haemophilis vaginalis*, which we now refer to as *Gardnerella vaginalis* (and how it will be referenced in this thesis), that associated with non-specific vaginitis, they postulated *G. vaginalis*

as the etiologic agent of vaginitis (*67*). However, attempts to prove this using Koch's postulates faltered as *G. vaginalis* could not be isolated from all the vaginas seeded with it, and not all women developed vaginitis despite the presence of *G. vaginalis*. Importantly this paper also described clinical signs associated with non-specific vaginitis, including high pH of 5.5-6.0, an odorous discharge, absence of *Lactobacillus*, and presence of epithelial clue cells (*67*). Some of these criteria are still used to clinically diagnose BV. Further study of *G. vaginalis* identified it in the vaginas of women without vaginitis occasionally (*68*). Refinement of culture methods to isolate *G. vaginalis* proved that recovery of *G. vaginalis* in heathy women was fairly consistent, albeit at a lower abundance compared to those with symptoms (*69*). These studies also identified multiple anaerobic organisms, including *G. vaginalis*, *Bacteroides melaninogenicus*, *Bacteroides disiens*, and *Mobiluncus* species, which were present in non-specific vaginitis (*70-74*). To better represent the condition the term bacterial vaginosis (BV) was introduced to replace non-specific vaginitis. This reflected the primary complaint of odour and alluded to the diverse bacteria associated with it, whereas vaginitis implied an increased level of leukocytes in the vagina, as increased leukocytes are not always observed in BV unless there was concurrent cervicitis or vaginitis infection. Indeed, it was only in the past two decades that a delineation was made between BV and AV by Donders in 2002 (*75*). Prior to this all vaginal dysbiosis was treated as BV, but as diagnostic methods were refined the delineation between conditions was clarified.

1.2.5 Methods to Diagnose Vaginal Dysbiosis

Despite the complexity of BV, methods to diagnose it are relatively simple. The Nugent score was developed by Robert Nugent in 1991 (*76*). It uses smears of vaginal mucus on a microscope slide to assess the quantity of gram-positive rods (*Lactobacillus*), gram variable coccobacilli (*Gardnerella*) and gram-negative curved rods (*Mobiluncus*) and is more commonly

used in scientific studies to assess BV. This wet mount slide is then evaluated based on the number of *Lactobacillus, Gardnerella* and *Mobiluncus* bacteria, assigning a score of 0-4 based on the number of *Lactobacillus* (gram positive large bacilli) seen in the field of view, score of 0-4 based on the number of *G. vaginalis* (gram variable small coccobacilli), and a score of 0-2 based on the number of *Mobiluncus* (gram negative curved bacilli). Low scores of 0-3 indicate high levels of *Lactobacillus* and low levels of other bacteria, while intermediate scores of 4-6 indicate an intermediate flora with proportionally less *Lactobacillus* and more *Mobiluncus* or *Gardnerella*. A score of 7-10 is diagnosed as BV with predominately *Gardnerella* or *Mobiluncus* and low levels of *Lactobacillus*.

Amsel's criteria is used by clinicians to diagnose women with BV and was initially developed by Amsel in 1983 (*77*). For someone to be diagnosed with BV three of the four criteria must be met. The four criteria include: a pH over 4.5 which (indicates a depletion of *Lactobacillus*), abundant homogeneous white-yellow discharge, a fishy odour upon addition of 10% potassium hydroxide to the discharge (indicates the presence of quaternary amines), and presence of bacteria, commonly *Gardnerella*, adhered to epithelial cells (commonly referred to as clue cells). Clinicians can perform these tests during the patient's visit and determine if treatment is needed.

Additional diagnostic measures to categorize BV have also been explored. One of these was through measuring the reduction-oxidation potential. This works as the redox potential is reduced in women with BV, while a high reduction potential is found in a *Lactobacillus* dominant, optimal microbiome environment (*78*). Biomarkers have also been explored to differentiate BV from non-BV, including proline aminopeptidase a protein produced by non-

Lactobacillus species that was more common in BV (*79*). There is still a lack of rapid point of care testing device to quickly identify if women have non-optimal vaginal microbiomes.

Aerobic vaginitis (AV) is defined more by the high level of inflammation associated with it, with the most severe form named desquamous intraepithelial vaginitis. Similar to BV, AV can be diagnosed through a vaginal smear, though no gram stain is needed, only a phase contrast microscope is needed. The scoring system involves assessment of three categories: the bacteria present, level of epithelial disruption, and inflammation, and results in a score from 1-10 (*75*). The bacteria are assessed both for the number of gram-negative bacilli, cocci, chained bacteria, as well as the number of lactobacilli present. Epithelial disruption is measured by the proportion of parabasal cells (small, round immature epithelial cells) where more parabasal cells reflect more epithelial disruption (as these immature cells should not be sloughed off, only the mature cells should if the individual is healthy). Inflammation is assessed through counting the number of leukocytes seen, as well as the proportion of leukocytes that are "toxic" (classified as toxic based on the presence of granules in the leukocyte). A score of 7-10 reflects the most severe disease, score of 5-6 are moderate AV, a score of 3-4 is mild dysbiosis, and a score of 0-2 reflects no disease (*75*). Only individuals with scores above 5 are considered to have AV, and BV would not result in a false positive as the lack of *Lactobacillus* would give a max score of 2 (*41*).

1.2.6 Epidemiological and Clinical Variables Associated with Vaginal Dysbiosis

Though a complete understanding of the etiology of dysbiosis is not currently known, many factors are associated with vaginal dysbiosis. Multiple factors are associated with dysbiosis including menses, sexual practices, vaginal cleaning practices, and other demographic factors. During the course of the menstrual cycle the change in hormone levels alters the vaginal

environment, as glycogen is regulated by estrogen, decreased estrogen (such as in the luteal phase) which removes a potential food source for *Lactobacillus (80, 81)*. The sexual practices of an individual can associate with BV diagnosis, including new sexual partners, multiple partners or condom-less sex, though studies of women who never reported sexual activity have identified women with BV (*82*), indicating that the cause of BV is multifaceted and not solely due to sexual interactions. Cleaning practices that can disturb the general homeostasis of the vaginal flora carry risks for inducing BV, this includes douching or vaginal cleaning, use of antiseptic solutions in baths, use of vaginal deodorants or scented soaps (*83*). There are demographic factors which are associated with increased BV prevalence. The prevalence of BV is higher among individuals of Hispanic and African ethnicity (*27*). Smoking is associated with increased BV in women who have sex with men, but only inconsistently associated in women who have sex with women (*84, 85*). All of these factors: environment, genetics, and behaviour can influence the vaginal environment and predispose conditions that are more favourable to non-*Lactobacillus* bacteria growth.

1.2.7 Negative Health Consequences Associated with Vaginal Dysbiosis

The first instance of associating bacterial dysbiosis and health complications was in 1914, where it was originally postulated that anaerobic bacteria in the vagina were "part of a complex bacterial milieu" that caused postpartum endometriosis (*64*). Much later, multiple other reproductive tract infections were linked to BV, including: postpartum endometriosis (*86, 87*) post caesarian wound infection (*87*) and infection after hysterectomy (*88, 89*). Salpingitis which is inflammation of the fallopian tubes has also been suspected to be caused by BV. Other reproductive complications that have been identified include pre-term birth, low birth weight of infant, premature rupture of membranes, and chorioamnionitis (*90-93*). Reproductive

complications from dysbiosis are thought to be due to bacteria gaining entry to the upper reproductive tract and causing inflammation which is detrimental during pregnancy. This was supported by findings that identified anaerobes including *G. vaginalis* that have been recovered from amniotic fluid (*90*), even after controlling for duration of labour and rupture (*94*). Unfortunately, some treatment trials of pregnant women with BV did not reduce risk of preterm birth (*95*). Salpingitis develops when pathogenic bacteria enter the upper reproductive tract, this can be from BV but can also include traditional STI's such as *Chlamydia trachomatis* or *Neisseria gonorrhoea* (*96*). This condition is very rarely found in those who have not had sex so the presence or transmission of bacteria is likely caused by sexual transmission (*97*). Salpingitis can also result in infertility and an increased risk of ectopic pregnancy (*96, 98*). These conditions can be exacerbated by concurrent infections with STI's.

Vaginal dysbiosis is associated with increased risk of acquiring STI's, including bacterial infections (Chlamydia and Gonorrhoea) and viral infections (HIV, HPV and HSV) (*99-102*). Though the exact mechanisms whereby BV increases risk of HIV have not been fully elucidated, there are several proposed mechanisms. The main hypothesis is increased risk due to the increased genital inflammation that is sometimes observed in BV, however this is not consistently observed among women. Dysbiosis is associated with increased risk of HIV potentially due to the increased presence of target cells (T cells which express membrane proteins C-C chemokine receptor types 5 (CCR5) (HIV co-receptor) and CD4 (HIV receptor)) recruited to the vagina in the presence of non-*Lactobacillus* microbiomes (*103*). Barrier damage is another hypothesis put forth that increases women who have BV's risk of acquiring STI's, as this increases the chance that bacteria or virus are able to find a niche or cell to replicate. In BV

the depleted *Lactobacillus* bacteria that can produce hydrogen peroxide does not produce as much, increasing the risk of Gonorrhoea infection, as it is not inactivated (*99*).

1.2.8 Treatment of Vaginal Dysbiosis

Though some women will have spontaneous resolution of BV without treatment, for the women that need treatment (*104*), few treatment options exist for BV and many women can have a recurrence of BV after treatment. Treatment of male partners of women with BV have not found to be effective in reducing BV in their female partners (*105-107*), despite an overlap between bacteria identified on the partner's penis of BV afflicted women, including *Prevotella* (*108, 109*) and there is not a recommendation to treat male partners of BV diagnosed individuals (*110*). Many antibiotics have been found to be ineffective at treating BV (*111-115*). Metronidazole was initially found to be effective in treating BV when women infected with both *Trichamonas vaginalis* and BV, where metronidazole was prescribed to treat *T. vaginalis*, resulted in resolution of both conditions (*112*). Other studies found that metronidazole was effective at treating BV (*111*) (*116*), however some studies found that it was not particularly effective in treating *G. vaginalis* (*117*). Upon reduction of metronidazole radicals are produced which induce bacterial death through nucleic acid synthesis interference and destabilization. As metronidazole needs to be partially reduced in order to be effective, this means that it does not negatively affect aerobic cells (*118*). Clindamycin is an alternative antibiotic used to treat BV, which prevents bacterial protein synthesis through blocking peptide chain elongation, and results in inhibition of both aerobic and anaerobic bacterial growth (*119*). However, as there are more significant side effects and a higher cost clindamycin is not as frequently used (*118*), but it is equally effective as metronidazole (*120, 121*). New antibiotics with a structure and mechanism of action similar to metronidazole have been found to be similarly effective in treating BV (*122,*

123). The treatment prescribed for AV depends on the severity of the disease/condition. Treatment of the atrophy and epithelial damage is through estrogen replacement (*124*), while the bacterial overgrowth is treated with a topical antibiotic such as clindamycin (*75*), kanamycin ovules (*125*) or mixifloxacin (*95*). Antibiotic treatment of BV is approximately 80% effective within one month, however recurrence rates are high, with 50% of women getting BV again within a year (*126*). Alternative treatments for BV are needed due to the poor response and risk of antibiotic resistance (*127-130*).

Targeting the bacteria through use of prebiotics, probiotics, or transplant of entire microbiomes from healthy donors have been explored as alternatives alleviate BV. Prebiotics are non-digestible nutrients that promote the growth of beneficial bacteria or other microorganisms. In the vagina, these could be compounds that could encourage growth of *Lactobacillus* or inhibit the facultative anaerobes that are overabundant in women with dysbiosis, and as a beneficial side effect the compounds could interact with the host and promote wound healing (*131, 132*), or prevent attachment of potentially pathogenic organisms (*133-136*). There has been a general focus on carbohydrate polysaccharides which feed beneficial *Lactobacillus* without promoting growth of non-ideal organisms *in vitro* experiments and in a clinical trial (*135, 137, 138*). A clinical trial had women use a sucrose gel and observed a similar short-term cure rate as metronidazole (*139*). Another carbohydrate alginate oligosaccharide was found to be antibacterial and reduce vaginal pH in a clinical trial (*140*). Probiotics are a known amount and specific bacteria such as various *Lactobacillus* species, some commonly found in the vagina (*L. gasseri, L. crispatus*) (*141*). Probiotics role in improving vaginal health include production of antibacterial compounds, acid production to lower pH, compete for adherence sites against non-ideal vaginal bacteria (*142-145*). Use of probiotics were found to resolve BV in some women

but this wasn't observed in all studies (*146*), or decreased the level of vaginal cytokines IL-1β and IL-6, with comparisons against antibiotics, have shown the potential of probiotics in treating BV (*147, 148*), or beneficial in concert with antibiotic treatment (*149*). Most recently a probiotic *Lactobacillus crispatus* (Lactin-V) was found to be more effective at reducing recurrent BV than placebo and was still detectable 12 weeks after last dose (*150*). Microbiome transplants involve the use of a healthy donor's microbiome, are transplanted into the recipient to replace bacteria. The use of vaginal microbiome transplants to resolve BV was explored based on experiments that were successful with fecal microbiome transplants in treating *Clostridium difficile* gut infections (*151*). A small study focused on women with recurrent BV found long term benefits were observed in 80%, while the last individual showed partial improvement *(152)*. Alternatives to treat BV that provide beneficial bacteria or compounds to encourage growth of bacteria need more research to identify safe and effective options.

1.2.9 Bacterial Function in the Vagina

Though a lot of microbiome research focuses on bacterial community, what is less characterized is bacterial function. Bacteria can act differently in different environments based on the signals they receive. The vagina is typically a low oxygen environment with high carbon dioxide levels (*34*), though this can be altered temporarily through insertion of a diaphragm or tampon (*153, 154*), so bacteria in this environment do not require mechanisms to defend against radicals that are by-products of oxygen reduction, and use alternative molecules instead of oxygen, as a terminal electron acceptor during energy production. Some bacteria scavenge molecules such as amino acids, nucleotides, ions, vitamins, and lipids from their environment or can make them *de novo*. One of the main active transporters in bacteria are the adenosine tri-phosphate (ATP)-binding cassette (ABC) transporters, which move a variety of compounds

including sugars, peptides, ions, as well as playing a role in the efflux of drugs, contributing to anti-bacterial resistance (*155*). Phosphotransferase transporters modify molecules by addition of a phosphate upon entry of the molecule into the cell. In the case of glucose this prepares it for entry into the glycolysis pathway, with an added benefit that it keeps the non-phosphorylated glucose in a separate concentration gradient as the molecule is now modified.

In the vagina, an important nutrient is glycogen, a multi-branched polymer of glucose. Glycogen is deposited in the vagina from epithelial cells, regulated by estrogen production (*80, 81*). In most cases glycogen cannot be directly used by bacteria, instead relying on human α-amylase to depolymerize glycogen into glucose and maltose, however some non-*Lactobacillus* genera have been found to produce an α-amylase, (*156, 157*). These sugars are taken up by bacteria and can be broken down into pyruvate through the glycolysis pathway, and in low oxygen environments pyruvate is reduced into lactic acid using the enzyme lactate dehydrogenase. Lactic acid is secreted into the vaginal lumen which acidifies the vagina mucus to a low pH (3.8-4.5) that is protective against growth of pathogens or other bacteria (*158, 159*). Only a small proportion of lactic acid is produced by vaginal epithelial cells, the majority of lactic acid in the vagina is produced by bacteria (*157-160*). Though *Lactobacillus* is the main genus to produce lactic acid *Atopobium*, *Megaspheara*, and *Leptotichia* (now known as *Sneathia*) can also produce lactic acid (*161*). *Lactobacillus* are the major contributors of lactic acid, though there is variation between the different species with how much lactic acid is produced and what type, with *L. iners* only producing the L-isomer of lactic acid (*162*), and *L. crispatus*, *L. jensenii*, and *L. gasseri* produce more of the D-isomer than the L-isomer (*163*). In addition to being important for acidification, lactic acid also functions as an epigenetic regulator of human genes, through inhibition of histone deacetylase, with D-lactic acid more potent than L-lactic acid (*164*).

Inhibition of histone deacetylase facilitates gene transcription of deoxyribonucleic acid (DNA) repair enzymes and factors to aid in cell survival (*165*). High levels of lactic acid can lyse non-*Lactobacillus* bacteria (*33, 34*), as well as inactivate HIV (*35, 36*) and *Chlamydia* (*37*). In women diagnosed with BV there is a low level of lactic acid and high levels of acetate, butyrate and succinate (products of carbohydrate or amino acid metabolism) (*166, 167*). Short chain fatty acids, such as butyrate and acetate, are associated with increased levels of pro-inflammatory cytokines in the vagina, though they are anti-inflammatory in the gut (*40, 168, 169*). In the vagina, succinate has been found to reduce the capacity of neutrophils to phagocytize microbes, allowing BV associated bacteria to continue to proliferate (*170*). Bacterial production of lactic acid and short chain fatty acids can influence host responses and impact the surrounding environment.

Many bacteria are capable of producing compounds that intentionally negatively affect the surrounding bacteria or human cells. One molecule which some bacteria produce that can negatively affect surrounding cells is hydrogen peroxide, though *in vivo* it does not have much physiological relevance. Most evidence suggests that levels of hydrogen peroxide in the vagina are not high enough to cause inactivation, however it remains as a marker of beneficial *Lactobacillus* (as *L. iners* does not produce hydrogen peroxide) (*171, 172*). Some *Lactobacillus* species can produce bacteriocin-like substances, which can have toxic effects on both gram positive and negative bacteria (*173*). These compounds can control overgrowth of potentially pathogenic bacteria and help maintain *Lactobacillus* dominance in the vagina (*174*). Some compounds target host cells, such as inerolysin, produced by *L. iners*, or vaginolysin, produced by *Gardnerella*, which create a hole in the host cell membrane resulting in the cell contents leaking into the vaginal mucus. Vaginolysin activity induces p38 mitogen-activated protein

signaling, and variation in vaginolysin production by different strains of *Gardnerella* may be a partial explanation for variation in IL-8 levels in women with BV (*175*). Mucin degrading enzymes, such as glycosidase and sialidase, are produced and secreted by BV associated bacteria (*176*) which function to decrease the viscosity of the vaginal mucus, and thus its ability to sterically hinder pathogens and can result in secretion of the inflammatory cytokine IL-1β (*177*). Though these compounds are produced with specific purpose to damage cells, many bacterial products have an unintentional effect on the host and potentially increase inflammation.

Though many bacteria adhere to the body, some bacteria irreversibly bind to a surface and form biofilms, which are more difficult to treat with antibiotics (*178*). Changes in bacterial signaling aid in initiating biofilm formation; expression of flagella aids in biofilm formation or aggregation through increased bacterial motion to find biofilms and aid in attachment (*179*). During biofilm development an extracellular matrix, comprised of multiple secreted macromolecular compounds including polysaccharides and DNA, is secreted from bacteria to surround the community (*180*). Biofilms benefit bacterial communities as a biofilm is more difficult to phagocytize or be targeted by antibody, aid in nutrient collection and exchange, allow for genetic exchange (*181, 182*). In addition biofilms can attract neutrophils, which secrete antimicrobial compounds and reactive oxygen species (ROS), which also initiates collagen degradation that results in tissue damage (*183*). Biofilms in the vagina can grow directly on the vaginal epithelium and have also been found on intravaginal rings, tampons, or intrauterine devices (*179*). Though *Gardnerella* is usually the dominant bacterium in biofilms other bacteria including *Atopobium* have been identified in BV-associated biofilms (*184*). Secretion of the proteins sialidase and vaginolysin from *Gardnerella* can aid in the formation of the extracellular matrix, by providing a scaffold that other bacteria can be incorporated into (*185, 186*). The

establishment of biofilms in women with BV is thought to be one reason why there is such a high recurrence rate in women treated with antibiotics.

1.3 Immune System in the FGT

1.3.1 Innate Immunology in the Vagina

Mucosal immunology differs from systemic immunity as mucosal surfaces are not sterile environments. Therefore, a balance between responding to pathogens and not commensal organisms that pose a minimal threat is required. To prevent infection physical barriers such as the epithelium and mucous work to prevent pathogen survival or entry, defensive physiological factors work to inhibit or inactivate pathogens, and innate immune cells patrol to immediately respond to pathogenic threats. These components work together to maintain tolerance to commensal, non-threatening organisms, and neutralize and eliminate threats.

Physical barriers function to prevent entry of pathogens into the submucosa and establish a productive infection. This includes the previously discussed epithelial barrier, which prevents pathogens from gaining entry into the body, however pathogens and inflammation can result in degradation of tight junctions, compromising the barrier integrity and allowing more transport across the epithelium (*187*). Commensal bacteria also play a role by physically occupying space that, if empty, a pathogen could adhere to (*188*). Covering the epithelia in the female genital tract is a layer of mucous, which is generally at an acidic pH of <4.5 in women without dysbiosis (*189*). Mucous sterically hinders pathogen movement by slowing and trapping pathogens (*190*). The contents of the mucous represent the secretions from the various cells in the FGT, and these factors can inhibit or enhance a pathogen's infectivity.

Mucous also contains many host factors, including many with immunological properties that can serve to protect or facilitate breakdown of epithelium integrity (*191, 192*). Mucins,

negatively charged glycoproteins, capable of binding and trapping pathogens, make up a considerable amount of the protein component in mucous (*4*). Mannose binding lectin is an antimicrobial protein that recognizes and binds to carbohydrate moieties found on microbial surfaces results in opsonization or lysis of the cell (*193*), and low levels of mannose binding lectin is associated with recurrent vulvovaginal candidiasis infections (*194*). The complement system aids in pathogen removal through several mechanisms: opsonization where the pathogen is covered in a complement protein that signals to immune cells that the cell needs to be engulfed, this can be aided by antibodies that are bound to the cell, formation of the membrane attack complex which can create holes in the pathogen's membrane resulting death, and initiation of an inflammatory response (*195, 196*). Inflammatory proteins, such as proteases can exacerbate tissue damage (*197*). Many different factors can influence the mucosal proteome that can affect pathways important for infection and inflammation.

For the immune system to respond to pathogens they must first be recognized as potentially harmful non-self. This can be through detection of pathogen associated molecular patterns (PAMPs), which can be recognized by toll-like receptors (TLR), of which there are 10 in total identified in humans (*198, 199*). Some TLRs are embedded into the outer membrane of many different cell types (TLR1/2, 4, 5, 2/6, 10), to target extracellular pathogens before they have had a chance to enter cells, recognizing lipoproteins or proteins which are unique to bacteria, fungi, protozoa or viruses (*200*). Other TLRs are intracellular (TLR 3, 7/8, and 9) present on the membrane of the endoplasmic reticulum, to respond to intracellular pathogens, which identify components in the cytoplasm that would not be there unless there was a foreign pathogen (such as double or single stranded ribonucleic acid (RNA) (that is missing protein complexes that would identify it as host RNA or DNA) (*201, 202*). In the vagina lower levels of

bacterial TLR's are found compared to the upper FGT, with consistent expression of viral associated TLR's, reflecting the need to have tolerance to bacteria in the vagina where there are high levels of commensal (*203-205*). Upon detection of a PAMP, the signal cascade can result in activation of NF-κB (Nuclear factor κB), mitogen-activated protein kinases (MAPKs), and the anti-viral immune response. This results in the expression of pro-inflammatory cytokines such as IL-1α and IL-1β, IL-6, interferon (IFN)-γ, and co-stimulatory molecules that are necessary to initiate an immune response and prepare other cells to resist infection (*206, 207*). Based on what TLR initiates the signaling cascade, the resulting response is tweaked to best respond to the pathogen. Another consequence of the immune response is secretion of vasodilators that increase blood flow so that leukocytes, which circulate in the blood, can be attracted to chemokines secreted and stay in the tissue to augment the immune response.

Leukocytes involved in the innate response include a variety of cells that have multiple functions to prevent infection. Neutrophils are often the first cell recruited to a new sign of injury or infection, through the chemoattractant IL-8, which is one of the most abundant cytokines or chemokines detected in vaginal secretions (*208*). When activated, neutrophils can phagocytize pathogens, where pathogens are engulfed and digested in lysosomes, as well as radicals such as ROS or nitric oxide (NO), or enzymes contained in granules, that when released into the extracellular space that will damage cells indiscriminately (*209*). Neutrophils can also create neutrophil extracellular traps, which have been shown to trap and inhibit HIV infection (*210*). Neutrophils are important for controlling vaginal candidiasis infections, but also can be the cause of tissue damage and inflammation observed in these infections (*211-214*). Neutrophils only remain activated for a short time before apoptosis is induced, as prolonged secretion of destructive enzymes must be tightly regulated, if not excessive tissue damage and

inflammation can occur (*209, 215*). The interaction between neutrophils and the vaginal microbiome is unclear, one study identified higher levels of neutrophils in women with BV (*216*). Another study found no difference in the number of neutrophils but identified increased cell surface expression of CD16 that indicates longer neutrophil lifespan (*217*), or lower levels of neutrophil enzyme (*218*). Much like BV, neutrophils have also been implicated in spontaneous pre-term birth, indicating that there may be an interaction between the two, as in this cohort of 14,000 women elevated pH was also associated with increased risk of pre-term birth (*219*).

Natural killer (NK) cells target host cells that indicate they are infected with a pathogen. If the infected cell has low levels of the membrane protein major histocompatibility complex (MHC) type I, NK cells can induce death of the infected cell through release of perforin and granzymes. These cells are also able to release cytokines to help direct the immune response. NK cells are particularly important for the anti-viral response, as well as removal of cancerous cells (*220*). A deficiency in NK cells is associated with increased susceptibility to HSV2 infection and cervical cancer due to HPV infection (*221, 222*). There are also invariant NK T cells, which when the MHC type I-like CD1d membrane protein bind to specific glycolipid antigens, activate the cell results in cytokine production of IFN-γ, IL-2, IL-4, IL-13, and IL-17A (*223*). There are other innate T cells, such as mucosal associated invariant T (MAIT) cells and γδ-T cells. When activated, these cells secrete pro-inflammatory cytokines (such as IFN-γ and IL-17) and MAIT cells can lyse bacteria infected cells (*224-226*). The γδ-T cells can produce Tumour Necrosis Factor-α (TNF-α) which induce maturation of DC's, aiding in the activation of the adaptive immune response (*227, 228*).

Both macrophages and DC's have similar roles as both phagocytes and antigen presenting cells (APC's). These cells link in the adaptive immune cells to create a specific

response to the invading pathogen. Enzymatically digested pathogen peptides are presented by APCs on their cell surface (by MHC class II), which can be recognized by the T_{helper} cell with a T cell receptor (TcR) that is able to bind to the MHC class II with the peptide, causing activation and proliferation of the T cell, initiating the adaptive immune response that is specific to the digested pathogen peptide. A mechanism for HIV infection is through DC's that sample the antigens in the lumen and can retract pseudopods to beneath the can pull virions through the epithelium to target cells by virions associating with CD209, so an intact barrier isn't entirely protective, and another is that HIV can enter through micro abrasions (*229*). The protozoan pathogen *Trichamonas vaginalis* is capable of inhibiting activity of dendritic cells, monocytes and macrophages (*230*). As the innate immune system is attempts to eliminate the threat the adaptive immune system is linked through APC's, either to assist rapidly in the case of prior exposure, or to develop responses to particular signals from the pathogen to aid in future exposures.

1.3.2 Adaptive Immunology in the Vagina

There are two main branches of the adaptive immune response: humoral immunity with B cells producing antibodies, and cell-mediated, which involved T cells indirectly and directly targeting infected cells. Both of these cell types express cell surface proteins which are designed to bind with a single peptide that when seen, would activate the cell and initiate downstream signaling. After the T_{helper} cell is activated it directs the immune response, the response type from secreted cytokines from the initial innate inflammatory response to pathogens. This can involve activating a response that is targeted towards intracellular pathogens or extracellular pathogens and once the pathogen is eliminated, the majority of adaptive immune cells will die,

though there will be a subset that survives and circulates the body to respond should the pathogen re-infect the individual.

Matured in the thymus, T cells are characterized by their expression of CD3 and their TcR receptors. There are three main types of T cells: helper (CD4+), cytotoxic (CD8+) and regulatory (CD4+, CD25+, FoxP3+). There are several different types of T_{helper} cells, including Th1, Th2, Th17, which present antigens on their TcR to MHC class II on APC's. The primary target cell for HIV is an activated CD4+ T cell, and Th17 cells are particularly susceptible and often the initial cell where infection is established (*231*, *232*). Th17 cells are characterized by their secretion of IL-17, they are a pro-inflammatory and secrete cytokines (including IL-22) that induce antimicrobial responses from innate immune cells and B cells, and IL-22 secretion can induce fibroblast wound healing (*233*). Both Th1 and Th17 responses are important for protection from vaginal candidiasis infection, with deficiencies in Th17 signaling pathways that impaired signaling resulting in decreased neutrophil recruitment and higher fungal load (*234*). Increased recruitment of activated Th cells, in particular Th17 cells, through TLR signaling is one purported mechanism of how the vaginal microbiome influences HIV risk (*63*). Cytotoxic CD8+ T cells contain granules with toxic enzymes and can induce apoptosis in infected cells, while T_{reg} cells prevent an immune response to an antigen that not harmful. Low levels of activated immune cells, particularly T cells, have been shown to be present in women who have been exposed to HIV but not seroconverted, compared to women who did seroconvert (*235*). Though T cells have been well characterized, their interactions with the vaginal microbiome need further elucidation.

Unlike T cells, B cells mature in the bone marrow (*236*). The role of B cells is to recognize antigens that bind to their unique B cell receptor. If the antigen binds to the B cell

receptor, the B cell undergoes differentiation to modify it's the B cell receptor into an antibody, with the same target specificity. There are two main types of antibodies, those that are neutralizing and those that are non-neutralizing. Neutralizing antibodies bind to the pathogen in such a way as to incapacitate the pathogen from infecting cells, such as targeting a viral envelope protein, so it is unable to enter host cells and cause disease. Non-neutralizing antibodies do not specifically interfere with a pathogens ability to infect host cells, but can still aid the immune system by signaling that the pathogen needs to be engulfed by phagocytes or destroyed via induction of the complement system. Antibodies in the vaginal mucous are from both local B cells, as well as from circulatory transudate (*4, 237*). In the vagina, unlike other mucosal surfaces, Immunoglobulin (Ig)G is the more abundant than IgA and IgM (*238*). Some pathogens have the ability to degrade immunoglobulins, including *T. vaginalis* and *N. gonorrhea (239, 240)*.

1.4 Mass Spectrometry

1.4.1 Development of Mass Spectrometry Based Proteomic Studies

Determination of protein identification and their sequence has developed greatly in the past several decades from identification of a single protein to hundreds or thousands in a single experiment. Prior to mass spectrometry (MS), researchers used chemical methods that systematically cleaved amino acids off to identify the protein sequence. Developments in genomics lead to an increased database of protein sequences that allowed for software to have an increased ability to match measured peptide sequences to proteins in the database. This allowed for scientists to try to characterize the proteome, defined as all proteins in a specified environment or tissue. However not all proteins can be measured in a single experiment due to the high complexity of the proteome, where it is estimated that the entire proteome of human

serum is greater than 10 orders of magnitude (*241*). Focus on comparing two conditions initially used protein gels to separate out proteins, first through its isoelectric point, and second through its molecular weight. Spots of interest, such as those that were or were not in the condition of interest, could be excised, digested and analysed by MS to identify the protein. This method was limited in the number of proteins that could be assessed and the precision with which proteins could be separated. New MS technology and methods were developed to meet the desire for increased resolution and sensitivity to detect more proteins and accurately quantify. Tandem MS is frequently used for shotgun proteomic experiments where there is an unbiased attempt to characterize as many proteins as possible in a sample. First, peptides were separated along the length of a column, thus peptides can be eluted off the column separately. Second, peptides are ionized upon exiting the column and read by the MS. This gives the first MS spectrum, where intensity values relate to abundance. Increased separation by the column reduces the frequency of peptides that overlap, increasing the number of peptides that can be identified. Third, the top peptides per scan are then fragmented to give us the MS/MS spectrum. From this MS/MS spectrum the peptide's sequence can be determined and matched to proteins in the database. The next advancement was increasing the number of samples analyzed, as initial experiments used low numbers of samples, or combined samples together if there were low protein concentrations. Labelling of samples with unique chemical tags allowed for samples to be combined and analyzed as a single sample, however samples can also be analyzed individually with no labels (termed label-free) (*242, 243*). Additional concerns about the reproducibility of shotgun MS experiments have limited the number of samples that could be analyzed, as longer length runs increase technical variation of the instrument. With shotgun proteomics there is a partial

stochastic effect where if the same sample is analyzed separately, not all of the same proteins will be identified, making cross experiment comparisons more difficult.

1.4.2 Application of MS to Understand Host-Bacterial Interactions

Metaproteomics is an emerging method to study microbial systems, by studying the expressed protein complement of multiple species by high-resolution MS. Proteins provide information on the functional potential of a microbial community, as they perform the majority of biological functions within a cell. Proteome databases on microorganisms are growing and becoming more comprehensive, thus making the utility of this approach more applicable for studies of the microbiome (*244*). The origination of the idea of metaproteomics occurred in 2004 but has taken off slower compared to metagenomics or metatranscriptomics (*245*).

Analysis of bacteria using metaproteomics to study vaginal clinical samples have been limited. Using a metaproteomic approach to study clinical vaginal samples allows for detection of both host and bacterial proteins. A study by Borgdoff et al studied clinical lavage samples from Rwandan women and identified 40 *Lactobacillus* proteins, 7 proteins that were specific to *L. iners* and 11 *L. crispatus* proteins, with no focus on host proteins. Most of the *Lactobacillus* proteins were ascribed to conserved intracellular functions many involved in carbohydrate metabolism, however high levels in CVL supernatant may indicate additional specific roles when secreted into vaginal lumen (*246*). Another study identified increased transport and catabolism, folding sorting and degradation, and energy metabolism proteins were increased in *Lactobacillus* dominant women, while non-*Lactobacillus* women had increased membrane transport associated proteins (*247*). A metaproteomic study on DMPA and non-hormonal contraceptive users did not assess differential bacterial functionality but did not find differences in the microbiome community between the two groups (*248*). However more research has studied the gut

microbiome. For example, metaproteomics has been used to study the gut microbiome of infants and functional shifts that occur during development (*249-252*); the composition and functional repertoires of the healthy gut microbiome (*253*); the impact of antibiotic therapy of the microbial proteome (*254, 255*). While these studies have provided insight into the *in vivo* microbial proteome, these have been limited in scale, many did not also analyze the host proteome, and methodologies for large-scale studies of the microbiome by proteomics has been lacking.

1.5. CAPRISA-004 Cohort

The cohort used to study host-bacterial interactions in this thesis was the CAPRISA-004 trial was organized and ran by the Centre for the AIDS Programme of research in South Africa (CAPRISA). This trial enrolled South African women who were HIV negative, sexually active, willing to be on some form of contraceptive, abstain from pregnancy, willing to participate in a trial, not currently or in the past year in a trial that involved vaginally applied products, and planning to stay in the catchment area for the trial duration (*256*). This trial was a phase 2b randomized double-blind control trial to test the efficacy of a 1% tenofovir disoproxil fumarate in hydroxyethyl cellulose gel, or the placebo gel which contained only hydroxyethyl cellulose, and. This gel was applied vaginally up to 12 hours before a sex act, and a second dose up to 12 hours after the sex act. Enrollment began in January 2007 and the trial concluded in May 2010. Extensive questionnaires were filled out at enrolment on demographics, behaviour, and contraception, and additional questionnaires at each monthly visit covering information on behaviour, genital health, contraception, and medication use. HSV2 serology was performed at enrolment and at trial exit. Only individuals who had symptoms that could be associated with

STI's were tested, in accordance with South African national guidelines. The STI's tested for were *Chlamydia trachomatis, Neisseria gonorrhoeae, Trichomonas vaginalis, Mycoplasma genitalium, Treponema pallidum (sub species pallidum)*. Individuals were recruited from Durban or Vulindlela, both located in the South African province of Kwa-Zulu Natal, which had an HIV prevalence in pregnant women of over 35% in 2009 and 2010 (*257*). The HIV epidemic in South Africa is hyper endemic and spread beyond risk groups, such as sex workers or those who inject drugs, into the general population, with few women in the trial reporting incidence of transactional sex, forced sex, or anal sex before or during the trial (*256*). The most common reason for ineligibility for trial enrolment was an HIV positive test (26% of screened women) (*256*). The 1% TFV gel was found to be 39% effective in preventing HIV and 51% protective against HSV2 infection (*258*). In those with high gel adherence, defined as women who used the gel >80% of the time, protection was 54%. However, two studies assessed factors that may have influenced HIV prevention efficacy and identified that non-*Lactobacillus* microbiomes and high levels of pro-inflammatory mucosal cytokines reduced efficacy. In those with high gel adherence even >70% protection was observed in women who were either *Lactobacillus* dominant, or those without high levels of pro-inflammatory cytokine (*259, 260*). Non-biological factors may also contribute to HIV risk, including partner age, as it was shown in the CAPRISA-004 cohort using phylogenetic studies that young women were often infected by older male partners (>8 years older), and mid-twenties women would transmit the virus to age matched male partners (*261*). The CAPRISA-004 cohort is well described, had very detailed questionnaires, and took mucosal clinical samples that make it ideal to study host-bacterial interactions.

1.6 Study Rationale, Hypotheses, and Objectives

1.6.1 Study Rationale

Despite the high prevalence of BV and the numerous health associated risks, there is still a dearth of knowledge of the role of bacterial function and specific communities in BV, and how they contribute to the observed variable pathogenesis. Most studies have not been powered to study the different non-*Lactobacillus* communities individually. Additionally, few studies focus on bacterial function to explore the potential role this may play in pathogenesis, and fewer studies also assess host pathways. Use of MS based proteomics to characterize large numbers of samples has previously been limited due to concerns about reproducibility and instrument capabilities, both increased instrument capabilities and computational abilities, study of large cohorts is feasible. Study of large cohorts will allow for characterization of different vaginal bacterial communities individually to identify host associations and associated bacterial functional variation between communities. Metaproteomics approaches, combined with in vitro experiments to confirm and further explore the host-bacterial associations, has the ability to identify novel host-bacterial interactions, and provide new insight into vaginal dysbiosis pathogenesis molecular mechanisms.

1.6.2 Hypothesis

MS based methodologies will enable characterization of the structure and functional properties of the *in vivo* vaginal microbiome which will allow for the identification of microbiome-host interactions and/or microbial species that modulate mucosal inflammation in the female genital tract. Non-*Lactobacillus* dominant vaginal microbiomes will have increased inflammatory factors and decreased barrier associated proteins.

1.6.3 Objectives

Objective I: Development of a mass spectroscopy-based method to carry out large-scale proteome characterization of *in vivo* vaginal microbiomes and host immunity.

Objective II: To characterize the relationship between the vaginal microbial proteome and mucosal inflammation.

Objective III: To understand the relationship between the vaginal microbiome and epithelial barrier function and neutrophil inflammation using *in vitro* co-culture models.

Chapter 2. Materials & Method

2.1 CAPRISA-004 Sample Collection and Ethics

Cervicovaginal lavage (CVL) samples were taken at quarterly clinical visits by a clinician using 3mL of sterile saline, from women who were not actively menstrating, applied to the posterior vaginal fornix. Saline was recovered and then centrifuged at 800 x g for 10 minutes to remove cells and cellular debris, and supernatant was stored at -80°C until use.

This study was approved by Human Research Ethics Committees at both the University of Manitoba and the University of Kwa-Zulu Natal. Only the women who provided written infromed consent for storage and use of their specimens for future research were included in this study (NCT00441298).

2.2 Protein Quantification Assays

Protein amount in cervicovaginal lavage samples were determined by Bicinchoninic acid (BCA) micro-scale assay (Novagen, Bilerica, USA). This assay works through reduction of Cu^{+2} to Cu^{+1} by peptide bonds in alkaline conditions, and chelation of Cu^{+1} by the BCA solution produces a purple colour that can be quantitatively measured. Samples were shaken for 10 minutes, vortexed briefly, then spun down with a mini-fuge before taking 8µl and putting into a 96-well plate in duplicate, and 17µl of MS grade water were added (total well volume of 25µl). Standards ranging from 0-25µg of Bovine serum albumin (BSA) were prepared in water, then added to the 96-well plate in duplicate. The colourimetric solution was made with 1-part 4% cupric sulfate to 50-parts of the BCA solution, and 200µl was added to each well and mixed. After a 30-minute incubation at 37°C the plate was read by a H1 Synergy hybrid plate reader (BioTek Instruments Inc, Winooski USA) at 562nm with Gen5 software (BioTek Instruments Inc, v2.05). Protein concentrations of samples were determined against the standard curve.

The second protein quantitation method used the 2-D Quantkit (GE Healthcare, Chicago, USA) for cell lysate samples as the dithiothreitol (DTT, reducing agent) used in the lysis buffer interferes with the BCA assay compounds. This assay works through precipitation of all protein then addition of a copper solution which will bind to the peptide bonds, and upon the addition of the colour agent which will bind to all free copper and produce a brown/orange colour. In duplicate, 5µl of cell lysate sample and 5µl of 2% sodium dodecyl sulfate (SDS) were added to a tube. Standards were prepared, from 2-20µg with BSA and one blank with 0µg, with 5µl of 2% SDS. To all sample and standards 500µl of precipitant (renders protein insoluble) was added to each tube, vortexed, and then incubated for 5 minutes at room temperature. Then 500µl of co-precipitant (contains reagents that co-precipitate protein) was added, vortexed, and centrifuged for 5 minutes 16,100xg. Most of the supernatant was removed, avoiding the pellet, and centrifuged again, 2 minutes at 16,100xg, and all remaining supernatant was removed. To the pellet, 500µl of copper solution was added (1-part copper solution to 4-parts water), and then tubes were vortexed. Colour reagent was prepared and 1ml was added to each tube followed by inversion of the tube 15 times to mix. After a 15-minute incubation at room temperature the plate was read by a H1 Synergy hybrid plate reader at 480nm with Gen5 software. Protein concentrations of samples were determined against the standard curve.

2.3 Sample Processing for MS

After quantitation of all samples, the samples were then digested using the filter aided sample preparation method for analysis by label free MS/MS as previously described (*247, 248, 259, 262-264*). The amount of protein to digest was chosen based on the quantity that would allow for the most samples to be digested using equal amounts. For CVL samples (n=801) 25µg of protein was used, while 100µg of cell lysate (n=58) was used. A separate reference CVL

sample was created by taking 20ul from 110 CVL samples. These samples were digested in batches of 12-16, initially samples were thawed on ice and resuspended before taking out a volume that corresponded to the required amount of protein, or max volume of those that were low concentration (n=72 CVL samples), and added to 500µl of 8.0M UEB (urea exchange buffer, contains Urea (GE Healthcare, Mississauga, Canada); with 50mM HEPES buffer (4-(2-hydroxyethyl)-1-piperazineethanesolfonic acid, (Sigma Aldrich, St. Louis, USA)), for samples with a volume over 300, 800µl of UEB was used. This solution was then incubated at room temperature for 10 minutes to allow denaturation of proteins through destabilization of the tertiary protein structure. During this time the 10kDa (Dalton) nanosep columns (VWR, Radnor, USA) are prepared by spinning water and then UEB through them for 1-2 minutes at 10,000xg (all centrifugation steps are at this speed unless otherwise stated). Denatured samples are added to the columns which retain proteins larger than 10kDa, and less reliably proteins <10 kDa after a 5-minute centrifugation. After washing the columns with UEB, 100ul of 25mM of DTT (Sigma Aldrich) was added to the CVL samples (cell lysate samples have DTT in the cell lysis buffer) which will reduce cysteine disulphide bridges. A 20-minute incubation at room temperature was followed by a 3-minute centrifugation. Both cell lysate and CVL samples had 100ul of 50mM iodoacetamide (IAA) (Sigma Aldrich) added followed by a 20-minute incubation in the dark. This step alkylates the cysteine sulphide prevented the reformation of any disulphide bridges. All samples are then centrifuged for 5 minutes to remove the IAA, then washed 3 times with UEB, and then twice with 50mM HEPES buffer. Cell lysate samples were subjected to a DNA reduction step with an incubation with benzonase nuclease H (Cederlane, Burlington, Canada) to minimize the amount of DNA contamination, followed by 2 additional washes with HEPES. The final step was addition of 2µg of trypsin (Gold Mass Spectrometry

grade, Promega, Madison, USA) for overnight (16h) digestion at 37°C, which will cleave proteins at the basic amino acids (lysine and arginine).

The next day peptides were collected by addition of 50ul of 50mM HEPES buffer to the columns, shaken in fresh collection tubes for 2 minutes, then the columns were inverted in the collection tube and centrifuged to elute peptides. This step was repeated two more times, then samples were dried down through vacuum centrifugation until <5µl was left in the collection tube. Peptides were stored at -30°C before further cleaning procedures. Peptides were resuspended in 20mM ammonium·formic acid and cleaned of remaining detergents and salts by reversed-phase liquid chromatography (high pH reverse phase, Agilent 1200 series micro-flow pump, Water XBridge column) using a step-function gradient which allowed for all peptides to be collected in to one fraction. Samples were dried down via vacuum centrifugation and stored at -30°C.

To ensure all samples had an equal amount of peptide analyzed by MS, we performed peptide quantification using the LavaPep fluorescence quantitation kit (Gel Company, San Fransisco, USA) according to manufacturer's protocol. Samples were resuspended in MS grade water, shaken for 20 minutes, then 10µl of sample were diluted into 90µl of water. This dilution was shaken for 20 minutes while 8 standards were prepared ranging from 0-0.6µg/ul from a digested peptide standard. In duplicate, 50µl of samples and standard were added into a 96-well plate then the fluorometric colour reagent was prepared. To each well 100µl was added and mixed, avoiding bubble formation. After a 1 hour room temperature incubation plates were read with an excitation wavelength of 540nm, and emission wavelength of 630nm by the H1 Synergy hybrid plate reader with Gen5 software. Peptide concentrations of samples were determined against the standard curve.

2.4 Mass Spectrometry

2.4.1 MS Instrument Settings

Through optimization, it was determined that 1μg of peptide, which had been used on the previous MS instruments (*248, 263-265*) was sufficient to capture information on the Q Exactive MS, and that adding more peptide (such as 3, 5, or 10μg), though they provided a slight increase in protein identifications, was not worth the increased potential for technical issues such as clogs. As described previously (*259*), a total of 1μg of peptide from each sample was analyzed with a nano-flow Easy 1000 in-line to a Q-Exactive Plus mass spectrometer (ThermoFisher Scientific, Waltham, USA) with a nano-electrospray ion source at 2.0kV (ThermoFisher Scientific) using a 2-hour linear gradient. Samples were loaded onto an EasySpray 50cm column (C_{18}-reverse phase, 100μl inner diameter, 1.8μm particles (Thermo Fisher Scientific) using a solution of 2% acetonitrile and 0.1% formic acid. Peptides were eluted off of the column with a 2-hour linear gradient of 2-30% buffer which contained 98% acetonitrile and 0.1% formic acid, at a steady flow rate of 200nl/minute. The MS was set to scan from 300-1700 mass to charge (*m/z*) ratio (with a target resolution value of 70,000 at 200*m/z* with a target of 3,000,000 automatic gain control value (which regulates the number of ions in the MS mass anlyzer at any one time) and a maximum injection time into the mass analyzer of 80ms. With data-dependent acquisition MS method, the top 15 most abundant precursor peptide ions from each scan were chosen for further fragmentation, with the intensity selection threshold set at 100,000 ions, a charge of +2-+5, and dynamic exclusion of 10s (where identical peptide ions seen in the mass analyzer up to 10s after are excluded from fragmentation). Precursor peptide ions chosen for fragmentation were isolated, with an isolation width of 3 *m/z*, then fragmented by higher-energy collision dissociation (28% normalized collision energy), which fractured the peptide ion into individual

amino acid ions. The second round of MS scans for the fragmented ions are scanned over a dynamic *m/z* range (target resolution of 17,500 at *m/z* 200, and target of 200,000 automatic gain control value, and maximum injection time of 100 ms. The compound polysiloxane was used to calibrate the machine for reproducible, accurate mass measurement.

Peptide CVL samples were randomized and run in batches of 50. Each CAPRISA-004 batch used the CVL reference mix that was run at 10 sample intervals to monitor LC consistency, MS performance, and utilized for downstrearm normalization. A reference sample for the cell lysate experiments was created by taking 2ul from each sample after the samples were prepared for MS (to 500ng/ul concentration) and then 2ul of that was run every 10 samples. All cell lysate peptides were analyzed in a single MS batch.

2.4.2 MS Host Data Processing

The MS spectral data was processed using Progenesis QI for proteomics (v21.38.1432; Nonlinear Dynamics, Durham, USA) following a protcol previously established in the lab (*247, 248, 262-266*). Each MS batch was uploaded and processed in Progenesis separately. One reference sample was chosen by the program, and all samples were automatically aligned to the specificed sample. In case of a failed or misaligned sample, manual seeding of vectors was used to guide the subsequent re-automatic alignment. Peak picking used the programs default settings, with a maximum peptide ion charge cutoff of +7, and inclusion of peptide ions between 10 and 121 minutes. After peak picking +1 ions were removed from analysis as they are background noise. Remaining ions were normalized based on expression of that ion in a run chosen by Progenesis. Protein searches were performed with the Mascot search engine (v2.4, Matrix Science) using a database comprised of amino acid sequences from: the SwissProt Human database, National Center for Biotechnology Information (NCBI) bacterial database,

Uniprot Candida reference proteome, Uniprot Trichamonas reference proteome, specific Uniprot human viruses reference sequences (HIV, HSV2, Epstein Barr virus, Hepatitus B virus, Hepatitus C virus), and sequences from all Uniprot bacterial phages (that were available June 2015). Mascot search parameters included carbaminomethyl fixed modifications (cysteines), oxidation modification (methionine), 0.8 Da ion error tolerances, and a maximum of 1 missed trypsin cleavage allowed per peptide. Protein identifications were confirmed with Scaffold 4 (v 4.4.1, Proteome Software), with inclusion thresholds set at 95% confidence for protein identification, ≥2 unique peptides per protein, and a threshold of 80% peptide identification confidence. Normalized relative abundances of each protein within each sample were obtained from after importing data back into Progenesis QI. Each MS batch was individually processed then combined in Microsoft Excel with the consolidate function using Uniprot ID to match between MS batches. Reference samples were used to calculate coeficient of variation (CoV) to identify if particular proteins had highly variable expression, where >25% CoV value indicates a highly variable protein. If >50% of proteins identified in a MS batch had >25% CoV then this indicated a run that had too much technical variability and was rerun.

2.4.3 MS Microbiome data processing

Bacterial searches were performed as described previously (*259*). Initial bacterial searches were performed using the Mascot search engine against the TrEMBL (translated European molecular biology laboratory) database (sequences taken in August 2015) that was restricted to include only bacterial proteins. The search results for randomized batches of samples were imported into Scaffold to identify proteins that met identity confidence criteria (protein identification with a false discovery rate (FDR) of 1%, and peptide identification with an FDR of 0.1%, with ≥2 unique peptides from a protein identified). A curated database was

created using Uniprot, included the UniProtKB/SwissProt Human database first to account for homologous proteins between the bacterial genera and humans, and identified genera that accounted for ≥0.2% of the total microbial protein abundance measured in the initial search. This final list included the genera: *Lactobacillus, Gardnerella, Pseudomonas, Mobiluncus, Ruminococcus, Prevotella, Ruegeria, Bifidobacterium, Chlamydia, Megasphaera, Pedobacter, Streptococcus, Escherichia* which were observed in our initial search as well as genera commonly seen in 16S ribosomal RNA (rRNA) sequencing of vaginal samples but did not pass our 0.2% threshold (*Atopobium, Dialister, Fusobacterium, Peptoniphilus, Peptostreptococcus, Porphyromonas, Shuttleworthia, Sneathia*). To avoid homology among bacterial genera the "OR" Uniprot function was set as a database construction parameter. A curated search was then performed and re-ran through Scaffold with the same protein/ peptide identity confidence criteria. Accession number reports from Scaffold were used to identify the specificity of bacterial proteins: undistinguishable (protein mapped to multiple genera), single genera, single species, or strain specific (protein mapped to only one protein in the database that had strain information). Microbial relative abundance was calculated by summation of normalized total spectral counts for all proteins associated with each genus or species within each sample.

2.5 Bacterial Culture Methods for *in vitro* Experiments

For the VK2 epithelial cell experiments, pure type strain cultures of *Lactobacillus crispatus* (strain DSM20584), *Gardnerella vaginalis* (strain ATCC14018), *Mobiluncus mulieris* (strain DSM25311), and *Prevotella amnii* (strain DSM23384) were obtained from Leibniz Institute DSMZ (Braunschschweig, Germany). Bacteria, directly from -80°C storage, were plated and then grown in anaerobic conditions on De Man, Rogosa and Sharpe agar plates (*L. crispatus*), 5% sheep's blood agar plates (*M. mulieris* and *G. vaginalis*), or chocolate blood agar

(*P. amnii*), passaged once onto fresh agar plates, then inoculated into liquid cultures (NYCIII media for *L. crispatus* and *G. vaginalis* and 1:1 ratio of NYCIII and fastidious anaerobe broth for *M. mulieris* and *P. amnii*). After 24 hours of anaerobic incubation using Gaspak EZ anaerobe container system with indicator (Becton, Dickinson and Company) at 37°C, culture OD_{600} readings were taken to calculate the total cell units (TFU)/mL, and all bacteria were diluted to $1x10^8$TFU/mL. Bacterial supernatants were collected using a 0.22μm Millex-GV hydrophilic polyvinylidene fluoride filter (Sigma Aldrich), bacterial media used for the negative media controls were also filtered to ensure sterility and processed the same as the bacterial supernatant.

For the neutrophil experiments the type strains of *L. crispatus* and *M. mulieris* were used and prepared the same as previously stated, however in addition to incubation in liquid culture the bacteria were washed 3 times with 5ml of Roswell Park Memorial Institute (RPMI) 1640 media then diluted to $1x10^6$TFU/mL in RPMI 1640 and incubated for 24 hours in anaerobic conditions. Cultures were removed from the anaerobic environment 10 minutes prior to the start of the neutrophil migration incubation. To isolate the bacterial supernatant, the bacterial suspensions were spun at 3700xg for 10 minutes and supernatant pipetted without disturbing the bacterial pellet.

2.6 Vaginal Epithelial Cell Experiments

To assess the response of vaginal epithelial cells to bacterial supernatant we grew up vaginal epithelial cells (VK2 E6/E7) (ATCC, Old Town Manassas, USA) and exposed them to bacterial supernatants, avoiding live bacterial cells to minimize cell death which was observed if certain bacterial species were incubated with vaginal cells. The VK2 cells were grown in T75 flasks (Corning, Corning, USA) keratinocyte serum free (KSF) media (ThermoFisher, catalog #17005042) supplemented with 0.05mg/ml bovine pituitary extract, and 0.1ng/ml human

recombinant epidermal growth factor, 0.4mM calcium chloride (Sigma Aldrich), and no antibiotics. Once cells reached 50-75% confluency cells were passaged with addition of 2ml of trypsin (0.25%, ThermoFisher, catalog #25200056) and a 5-8 minute incubation at 37°C. Once cells were detached, 8ml of transition media (Dulbecco's modified eagle medium:F12 media (with 10% fetal bovine serum, ThermoFisher) was added, pipetted up and down to resuspend cells, then transferred to a 15ml conical tube and centrifuged for 10 minutes at 205xg. All transition media was removed by pipette and cells were resuspended in the KSF media and a 1:3 or 1:4 dilution was added to a fresh T75 flask, and the volume topped up to 12 ml with KSF media.

Cells for the bacterial supernatant experiments were seeded into T75s, and one 24 well plate, then let grow for 2 weeks, changing media every 2-3 days to allow for differentiation. One of two bacterial media controls in duplicate, or bacterial supernatant in triplicate from *L. crispatus*, *G. vaginalis*, *M. mulieris*, or *P. amnii* was mixed in a 1:5 ratio with KSF media and added to a T75 flask with differentiated VK2 cells for 24 hours. Cells were trypsinized for 10-15 minutes at 37°C, then the cell pellet was washed with phosphate buffered saline three times. The pellet was resuspended lysis buffer (contained DTT, HEPES, and SDS) and incubated for 10 minutes. After incubation, samples were added to Qiashredder tubes (Qiagen, Hilden, Germany, catalog #79656) then spun for 5 minutes at 10,000xg, and then samples were stored at -80°C until protein quantitation.

The 24-well plates were used to assess cell viability via WST-1 assay (Sigma Aldrich, catalog #5015944001). After a 24 hour incubation with the bacterial supernatant, or controls, 40ul of the WST-1 reagent was added to each well. For a positive control with dead cells, 8ul of Triton-X (Fisher Scientific, catalog #BP151-100) was used. After incubation for 2 hours at 37°C

with 5% CO_2 the plate was read at 450nm, and 650nm by the H1 Synergy hybrid plate reader with Gen5 software. Values measured at 650nm (background) were subtracted from the measured value at 450nm, then percent of cells actively metabolizing were calculated by division of the average value seen in the negative controls (only human cell media).

2.7 Isolation of Neutrophils from Blood

Whole blood was collected from healthy local donors <1 hour prior to use, and all donors provided informed consent (protocol approved through both the Health Canada institutional review board and University of Manitoba's Department of Medical Microbiology and Infectious Diseases research ethics board). The EasySep™ Direct Human Neutrophil Isolation Kit (StemCell, Vancouver, Canada, catalog #19666) was used for negative immunomagnetic separation of neutrophils from whole blood per the manufacturers protocol, without the use of a chelating agent, as it inhibited migration through removal of Ca^{+2} which is needed to for migration. This protocol uses a proprietary cocktail of antibodies against markers which will only bind to non-neutrophil cells. Magnetic beads were added to the blood-antibody mixture, where the magnetic beads (RapidSpheres™) will bind to the constant domain of the antibody. After a 5 minute incubation the tube is added to a magnet and incubated for 5 minutes. Without removal of the 14ml polystyrene tube from the magnet, the mixture was poured into a fresh tube and additional magnetic beads were added and incubated for 5 minutes to bind any remaining antibody. This was followed by a 5 minute magnet incubation, suspension poured into fresh tube, and after a final incubation in the magnet pour into a new tube. Isolated neutrophils were then poured out, counted by trypan blue staining, washed once with RPMI 1640, and re-suspended to $5x10^7$cells/mL in RPMI 1640 for use. Purity of neutrophils isolated from healthy donors Stemcell was stated to be 94.0 ± 3.7% by gating on CD16+ (Anti-Human CD16 Antibody

clone 3G8, Stemcell, catalog #60041), CD66b+ (Anti-Human CD66b Antibody clone G10F5, Stemcell, catalog #60086). Neutrophil purity was assessed with the in-lab neutrophil markers: CD16 (Anti-Human CD16 Antibody clone 3G8, catalog #557920, BD Biosciences, San Jose, USA), CD15 (Anti-Human CD15 Antibody clone HI98, catalog #563872, BD Biosciences) and CD49d (Anti-Human CD49d Antibody clone 9F10, catalog #559880, BD Biosciences) and a viability marker (570 (PE), catalog #564995, BD Biosciences). The purity of neutrophils based on live cells that were CD16+, CD15+, and CD49d- was 81% (See Figure 1).

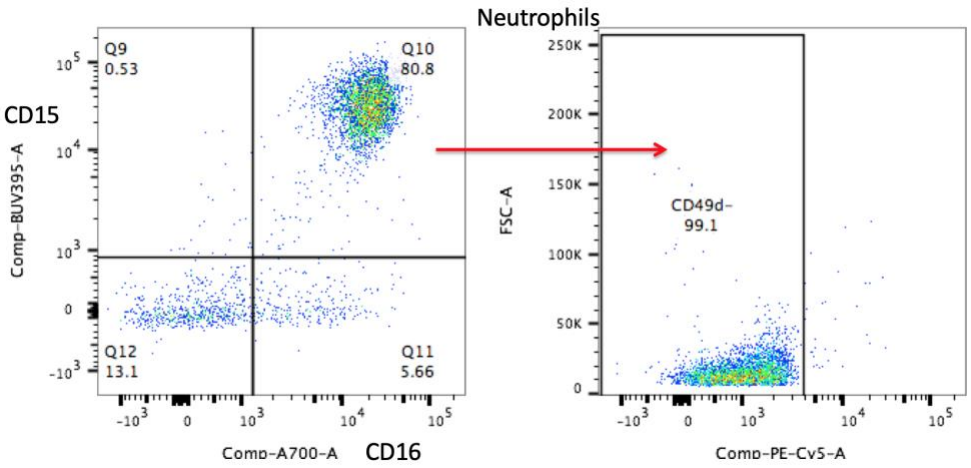

Figure 1. Assessment of the purity of isolated neutrophils through flow cytometry using EasySep™ Direct Human Neutrophil Isolation Kit.

2.8 Neutrophil Migration Experimental Protocols

The neutrophil experiments were modified from a previous experimental protocol by Kusek et al (*267*) with an experimental overview shown in Figure 2. Two days prior to the experiment, 3µm 24 well plate inserts (Millipore Sigma) were treated with 25µg/mL fibrinogen (ThermoFisher Scientific), incubated for 24 hours at 37°C, 5% CO_2, excess liquid removed, then stored at 4°C until use. Bacteria were prepared as described above and $1x10^6$ bacteria, or supernatant from $1x10^6$ bacteria, were added to each well. Fetal Bovine Serum (FBS) (10%) in anaerobic RPMI 1640 was used as a positive control for migration and activation, while

anaerobic RPMI 1640 (with no additives) was used as a negative control. The prepared plate inserts were added to each well then $1x10^6$ freshly isolated neutrophils were added into the inserts. Ten 2-fold standard dilutions of neutrophils incubated in RPMI 1640 containing 10% FBS were used to generate a standard curve. The neutrophils were incubated for 3 hours at 37°C, 5% CO_2, then the inserts were removed from the plates and the migrated neutrophils were counted using trypan blue staining. Activation status of neutrophils was inferred through measurement of relative levels of myeloperoxidase (MPO) enzyme activity, as previously described (*267*). Briefly, 50µL 0.4M citric acid was added to each well, mixed, then 100µL was transferred to a 96 well plate in duplicate. The colorimetric solution was comprised of 16.7µL 30% w/v hydrogen peroxide (Sigma Aldrich), 1.67mL 0.4M citric acid, 10mg 2,2-azino–bis-ethylbenzohiazoline-6-sulfonicacid diammonium salt (Sigma Aldrich) in sterile water. To each well 100µL colorimetric solution was added, incubated for 10 minutes in the dark at room temperature, then read immediately with an H1 Synergy hybrid plate reader at 405nm. Assays were set up in triplicate, or duplicate if neutrophil recovery was poor.

Incorporation of VK2 epithelial cells were included to understand if a synergistic effect on the neutrophil response to bacteria or bacterial supernatant would be observed. In 24-well plates VK2 cells were seeded at $1x10^5$ cells per well, and grown for 1 week to achieve 100% confluency, which was confirmed visually with microscopy. Bacteria were prepared as described above and $1x10^6$ bacteria, or supernatant from $1x10^6$TFU/ml bacteria were added to each well, followed by a 4 hour incubation. The positive control was TNF-α 100ng/ml, which induced IL-8 secretion from cells as observed previously, creating a chemokine gradient for neutrophil migration (*268*). After the 4 hour incubation isolated neutrophils were added for a 3 hour incubation as previously described. A WST-1 assay was also performed as previously

described to assess the VK2 cells metabolic capabilities after the 7 hour experiment. All conditions were compared to VK2 cells in RPMI1640 media that were not exposed to bacteria or neutrophils.

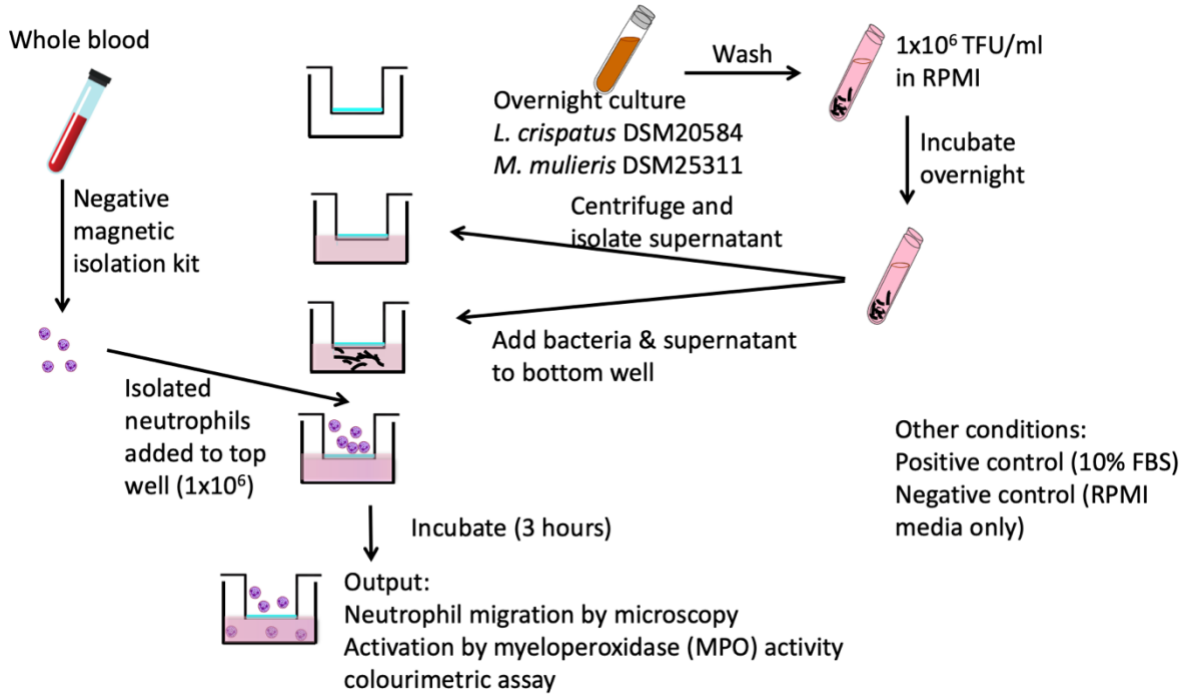

Figure 2. Neutrophil experimental overview without VK2 cells showing bacterial preparation, neutrophil isolation and transwell setup.

2.9 Statistical Analysis

Outlier samples in the host proteome dataset were identified for each MS batch by calculating the interquartile range (IQR) for each MS batch based on $\log_{10}$ protein normalized abundance, and any sample that had a median value outside the IQR was considered an outlier and re-analyzed by MS. Principal component analysis (PCA) with the R function "princomp" (built in R function, R version 3.5.2) was used to identify sample outliers based on protein expression. Principle variant component analysis (PVCA) was used to identify major contributors of biased protein expression based on technical factors was assessed using an in-

house R script created by Laura Noël-Romas, with the graphs created in PRISM (v8.3.0). Correction for MS batch was performed using the R function "ComBat" (package "sva", Leek, v3.20.0, 2013) which used a parametric Bayes framework to adjust protein expression values in a MS batch, each protein individually adjusted in each separate MS batch, to minimize effects of a known variable (*269*). All heatmaps were created with the R function "aheatmap" (package "NMF", Gaujoux and Seoighe, v0.23.0, 2010). Pearson correlations for protein expression across reference samples both within the same MS batch and across MS batches were performed with SPSS (IBM Corp. Released 2013. IBM SPSS Statistics for Macintosh, Version 22.0. Armonk, NY: IBM Corp.), and results visualized with PRISM.

To identify the proteins that had the most influence, PCA was used and the top 10 loading variables (proteins) were identified and visualized with R. To identify biological functions that were unique to proteins with 100% (proteins detected in all MS batches) or <100% sample (proteins not detected in all MS batches) coverage or pathways seen in both, the two lists were uploaded into DAVID and all biological functions (in the biological processes (BP)-DIRECT category) that were significant (BH adjusted p value≤0.05) were taken. Overlapping pathways, and pathways unique to 100% and <100% sample coverage were graphed with PRISM. Analysis of host data to identify groups of proteins that had similar expression across samples used unsupervised hierarchical clustered heatmaps.

Analysis of bacterial spectral count, number of bacterial proteins across samples were performed and graphed with PRISM. Visualization of detected MS bacterial function diversity was performed with R ((package "data.tree", Glur, v0.7.11, 2019) and (package "networkD3", Allaire, Gandrud, Russell and Yetman, v0.4, 2017)). Analysis of MS identified bacterial proteins to the human vaginal non-redundant gene catalog (VIRGO) database used all protein

accession numbers from *Lactobacillus, L. iners, L. crispatus, Gardnerella, Mobiluncus* and *Prevotella* and analyzed them through use of NCBI's BLASTP (2.10.0) against the VIRGO amino acid sequence database (*270*). Hits were considered a match if the significance level had an Evalue<$1x10^{-10}$. Thresholds were based on matching to the protein amino acid sequence matched to the initial protein searched ("identity") and the full-length proteins amino acid sequence coverage matched between the search query protein and significant match (hereafter referred to as coverage). The three different thresholds were 80% identity and 70% coverage (least strict), 90% identity and 80% coverage, and the most stringent threshold was 95% identity and 90% coverage. The bacterial proteins (identified by MS) were weighted (using the average relative abundance across all samples) to show results that were proportional for proteins that were highly abundant in the MS dataset. Visualization of circular stacked bar plots for bacterial community and function data were performed with R function "coord_polar" (package "ggplot2" Wickham, v3.3.2, 2016).

Proteomic comparisons were done using a two-tailed unpaired Students *t*-test with Benjamini-Hochberg (BH) multiple comparison correction, and significant protein lists were analyzed by DAVID and IPA to identify over and underexpressed pathways. For analysis through DAVID only significant proteins were uploaded, and the top pathways with non-redundant proteins (overlapping protein lists) were used and the enrichments BH adjusted p-value was reported. For IPA, only significant protein lists with log expression ratio were uploaded. Expression core analysis was used with default settings, with the exception of choosing experimentally observed relationships (excluding predicted relationships). Canonical pathways and functions with a calculated Z-score and significant p value were analyzed. Canonical pathways refer to well characterized signaling pathways, like acute phase response

signaling, which is an acute inflammatory response, and contain proteins that have a specific role in that pathway, such as an activator or inhibitor, transcription factor or downstream signaling molecule. Functions in IPA are more straight forward in their end biological result, so inflammatory response, and may combine different canonical pathways if they are involved in the broader function. If there was an activation Z-score of 0 was found this meant that IPA was confident that the function or pathway is neither inhibited nor activated, compared to a Z-score of N/A which means that IPA cannot confidently determine the expression of the pathway, while a positive Z-score indicates activation and a negative Z-score reflects an inhibited pathway or function. Gene set enrichment analysis (GSEA) was used to identify protein signatures in the proteomic dataset that were similar in expression to the immune cell libraries. The pre-ranked option was used to identify similar immune cell proteome signatures, using the fold change between comparison groups, and a minimum of 15 overlapping proteins.

Demographic and clinical characteristics were received from CAPRISA and each sample was matched to the corresponding patient id and visit date of sample collection. Averages and standard deviations were used to show more variation of variables within groups compared to median and IQR, along with the range seen in the cohort. Statistics were calculated in R using Pearson Chi-squared tests (with Yates continuity correction) for categorical variables, and the Wilcoxon rank sum test was used for continuous variables to as the data was not normally distributed.

To identify functional differences between bacterial communities, 3 statistical tests were used, which help account for biases within tests to avoid false positives. The three tests were: Wilcoxon rank sum test (non-parametric test for independent groups), a permutation test (run with 10,000 iterations, which regroups the data to see what the probability would be for the data

seen to be a random chance), a fisher's exact test (splits the groups into counts of above and below median values of the factor). Before analyzing the data by Wilcoxon rank sum test and permutation test, the data was changed into a centred log transformed ratio. This is used to transform compositional data into Euclidean space data where statistical tests are valid. Two different zero replacement methods were used to account for any variation that may be induced due to zero replacement.

Epithelial proteome analysis used two-way unpaired student's *t*-test after subtracting the expression observed in bacterial media samples. All comparisons were against *L. crispatus*, and significant protein lists were analyzed with DAVID and IPA as previously described. One-way paired student's *t*-test based on donor was used to assess statistical significance for neutrophil migration assays, both experiments with and without epithelial cells, after adjusting the *L. crispatus* and *M. mulieris* conditions for background migration or MPO activity observed in the media blank controls.

2.10 Contributions to this project by collaborators and lab members

The team at CAPRISA (in particular Dr. Salim S Abdool Karim, Dr. Quarraisha Abdool Karim, Dr. Leila Mansoor, Anneke Grobler, Natasha Samsunder, and Dr. Jo-Ann Passmore) developed and ran the CAPRISA-004 trial and provided the epidemiological data. Lyle McKinnon and Adam Burgener developed the initial cohort study to look at HIV risk proteomic signatures. I, Jen Butler, and Liane Arcinas performed the BCA assays for the CAPRISA-004 and VK2 cell lysate samples. Max Abou ran the offline LC to clean the peptides for all samples, while I prepared the samples for the LC and dried them down after. I performed all peptide quantitation assays and I prepared the samples for MS analysis. Stuart McCorrister at the Mass Spectrometry and Proteomics core at the Public Health Agency of Canada ran the samples on the

Q Exactive MS. I processed the host MS data and performed quality control checks, while Kenzie Birse processed the CAPRISA-004 microbial proteome dataset and helped with quality control assessment. Laura Noël-Romas batch corrected the CAPRISA-004 host proteome dataset and helped with quality control assessment. I identified the major microbial communities, sorted samples into the appropriate communities, and performed the statistical tests to identify differentially expressed proteins, and compare epidemiological variables between groups. Sarah Hoger identified the appropriate statistical tests and zero replacement methods to analyze the bacterial proteome data, and I carried out the analysis and interpretation of the functional data. Alicia Berard and I developed and performed the VK2 cell exposure to bacterial supernatant experiments. I performed the sample processing and data analysis. Christina Farr Zuend and I performed the first neutrophil migration experiment and optimized the neutrophil experiment conditions. She also suggested the protocol for neutrophil isolation which she was using for neutrophils and other cell types in the lab. Christina also ran the flow cytometer to determine the neutrophil cell purity after isolation. I performed all neutrophil experiments and data analysis, with help from Alana Lamont who grew up the bacteria for my experiments until I was trained to do so myself.

Results:

Chapter 3. MS-based proteomics methodologies to analyze mucosal biology of the FGT

3.1 Overview of MS Workflow for Large Cohorts

A total of 732 cross sectional cervicovaginal lavage (CVL) samples were received with blinded identifiers and stored at -80°C. The workflow for sample processing and the creation of both bacterial and host datasets is depicted in Figure 3. Samples were processed for MS analysis as previously described (*248, 259, 263, 264*). Samples were run on a Q Exactive MS in experimental batches of 50, resulting in a total of 18 independent experiments. Each batch included a reference mix which was injected at regular intervals after 10 samples to monitor MS technical consistency. Each batch was independently analyzed by MS over a period of 1.5 weeks, and all 18 batches were analyzed over a 12-month period. Data quality of each batch was evaluated as was individual sample quality. Quality control measures assessed for both the overall run technical consistency, and individual sample quality. Samples that did not pass quality control were reincorporated into a later MS run. The bacterial dataset was developed through an initial search of the uncurated Uniprot database tREMBL, restricted to bacterial searches. High confidence search hits determined which bacterial genera were included in the curated database (Vaginal Metaproteomic Database), which contained all bacterial genera above a 0.2% abundance threshold cut-off. Sample MS files were re-searched against the curated database, which had human included to avoid bacterial sequences that matched to human sequences. Homologous proteins were identified, and the lowest common ancestor was used for identity. Function of bacterial proteins was annotated with Ghost Koala using Kyoto Encyclopedia of Genes and Genomes (KEGG) ontology. After data processing all datasets (host, bacterial and epidemiological) were combined, then samples were grouped based on bacterial

communities for comparisons to identify dysregulated host pathways, and differential bacterial function.

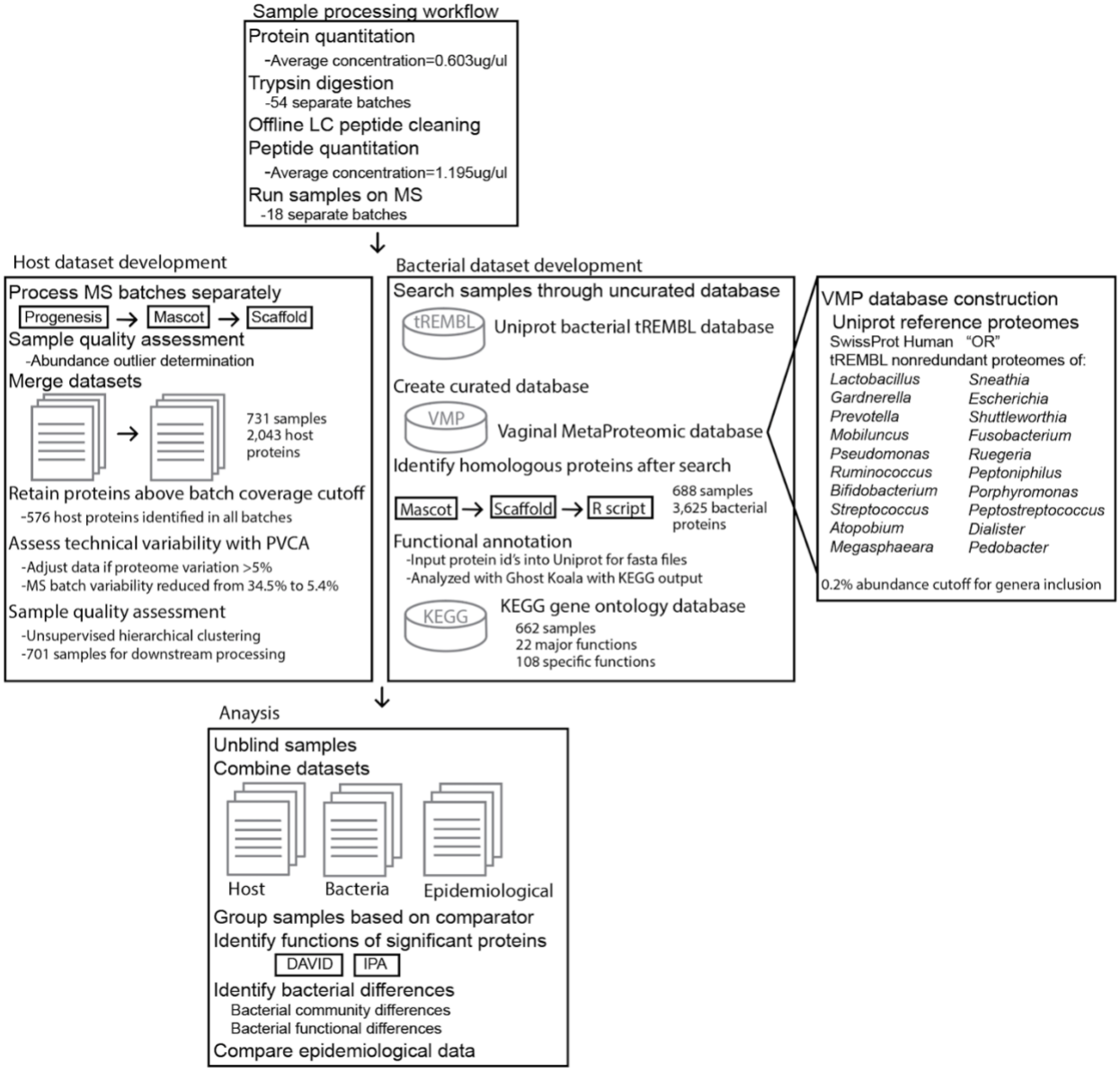

Figure 3. Workflow for sample and data processing of mucosal samples for MS analysis in multiple batches. Datasets indicated with paper, programs boxed, databases indicated with cylinders.

A total of 2,043 unique human proteins were identified, with an average of 1,219 ± 106 human proteins detected per batch MS batch. The total dynamic range of protein abundance, averaged across samples, was 7.58 $\log_{10}$ normalized abundance (NA), from 1.34 (omega

amidase, NIT2) to 8.92 $\log_{10}$ NA (small proline rich protein-3, SPRR3). Of the 2,043 proteins, 576 (28.2%) were consistently identified in 100% of MS batches. Proteins which were identified in all MS experiments were typically of higher abundance, ranging over 5 orders of magnitude (Range=3.82-8.92$\log_{10}$NA) (Figure 4A), while those that were detected less frequently were typically of lower abundance range (Range=1.30-6.90$\log_{10}$NA) (Figure 4B). Protein abundance correlations of the proteins with 100% MS batch coverage were high between reference samples ($R_{Pearson}$>0.90) showing high reproducibility across multiple experiments (Figure 4C). Only proteins that were detected in 100% of MS batches were carried forward with for analysis. Prior to downstream analysis protein abundance was adjusted using parametric Bayes statistical framework, through the program ComBat, to reduce MS batch to batch variation. Visualization of sample proteomes by principal component (PCA) analysis shows a clear reduction of clustering due to MS batch after adjustment (Figure 4D).

We utilized principal variant component analysis (PVCA) to assess if there were technical variables that biased proteome expression. The technical factors included were visible blood in CVL sample, BCA assay batch, BCA assay technician, digestion batch, digestion technician, clean up batch, peptide quantification assay batch, MS batch, and volume loaded onto MS. Pre-adjustment, MS run contributed the most (34.2%) to explained proteome variability, followed by blood contamination, with technical factors accounting for <2.0% each (Figure 4E). Post adjustment, visible blood contamination of sample contributed the most (16.3%) to explained proteome variability, with all other technical factors (BCA technician, digestion batch, peptide quant batch, MS loading volume, LC clean up batch, digestion technician, or BCA assay batch, each explaining <2% of the proteome variation (Figure 4F). The residual, which contained unexplained biological variation, as well as any unaccounted technical variation, was

increased from 43.4% pre-adjustment to 68.7% after adjustment.

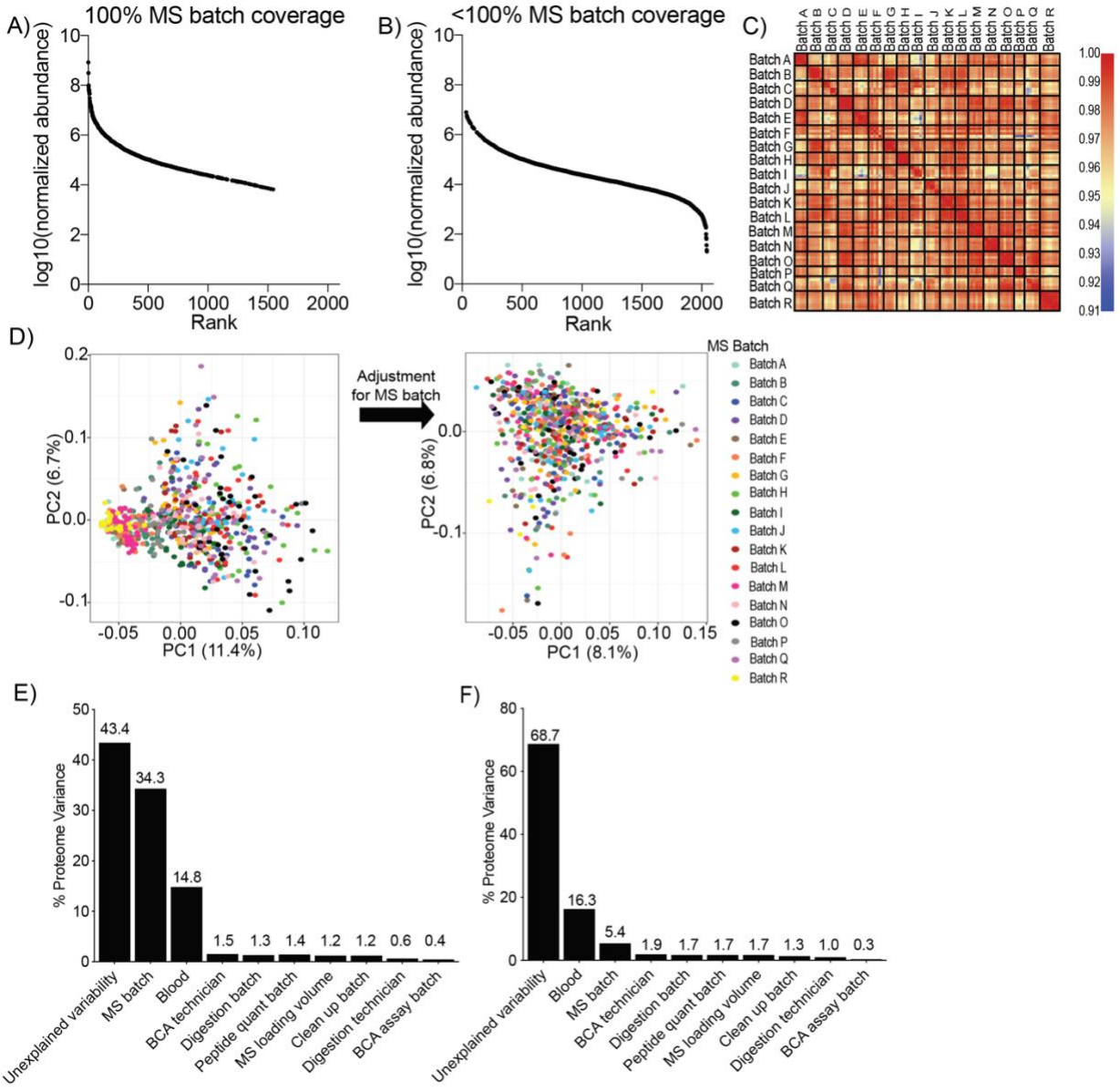

Figure 4: **Assessment of technical factors contributing to experimental variability of mass spectrometry analysis of cervicovaginal fluid. A)** Proteins ranked by averaged normalized abundance in proteins detected in 100% of MS batches. **B)** Proteins ranked by averaged normalized abundance in proteins detected in <100% of MS batches. **C)** Unadjusted Pearson correlations of 576 host proteins expressed among all reference samples from all 18 batches. **D)** PCA of samples coloured by MS batch before (left) and after (right) ComBat adjustment. Each dot represents one sample and are coloured by MS batch. **E)** Percentage of proteome variation due to technical factors before adjustment. **F)** Principal variant component analysis showing variation in the host proteome expression that is explained by technical factors after adjustment for MS batch.

3.2 Criteria to Evaluate MS Sample Quality

Quality control was an important component of ensuring the consistency between MS batches. Data quality of each batch was evaluated by identifying the percentage of host proteins in reference samples that had <25% CoV, which means consistent protein expression throughout the MS batch, and low numbers of proteins that pass this threshold indicate a poor-quality MS batch (Figure 5A). Individual sample quality was assessed by identifying samples with an overall total protein abundance outside the interquartile range within the individual experiment (Figure 5B). A total of 24 samples (3.2% total, 0-3 per batch) met these criteria, and were subsequently re-run in a final experiment after all runs had been completed. Data quality was assessed by visualization through PCA (during MS data collection) and identified 11 samples that were outliers which were incorporated into a later MS run (Figure 5C). After all data collection and adjustment unsupervised hierarchical clustering after adjustment (Figure 5D) found 29 samples respectively that were labelled as outliers, which were excluded from analysis. A total of 701 (95.7%) samples passed quality control measures and were included in downstream analysis. In total, 732 cross sectional cervicovaginal lavage (CVL) samples were analyzed across 18 separate label-free MS runs, with 31 (3.9%) samples that did not pass quality control measures which were excluded from downstream analysis.

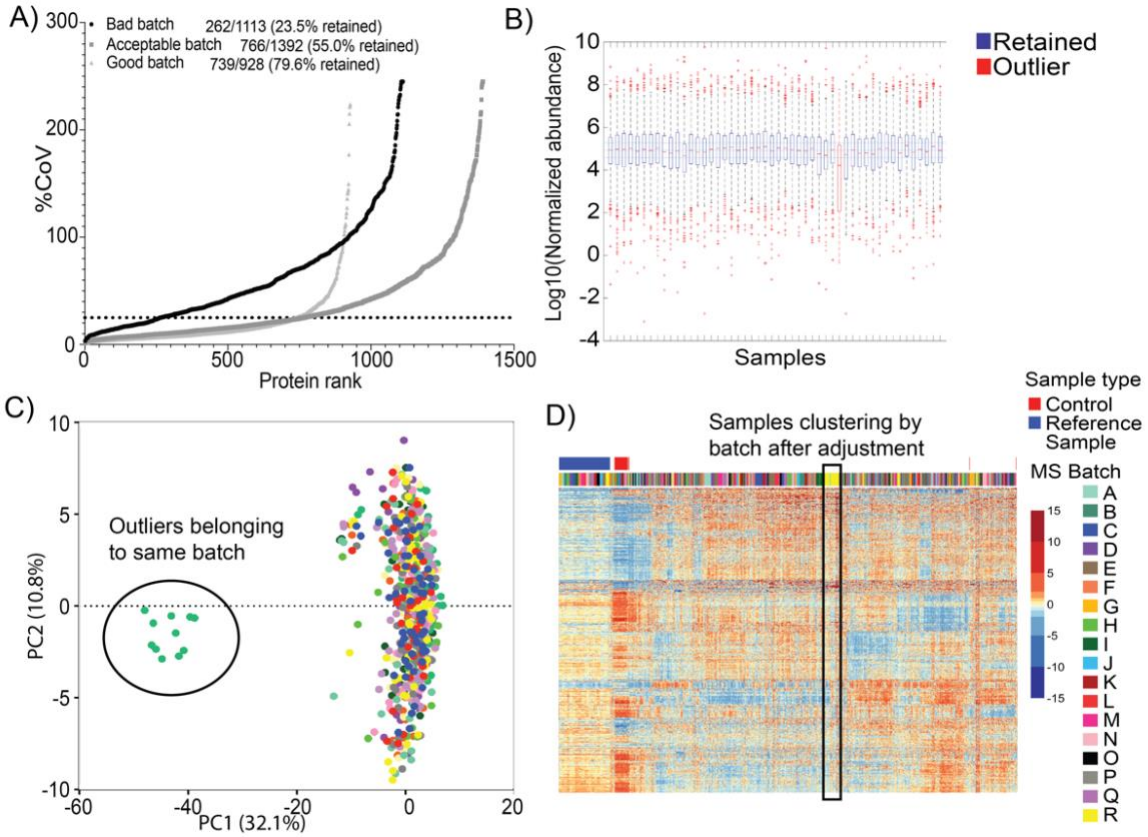

Figure 5. Measures used to assess MS batch and sample quality. **A)** Example of bad, acceptable, and good MS batch based on protein retention as determined by <25% CoV (coefficient of variation) among the reference samples in MS batches **B)** Sample outlier example based on total protein abundance being above or below the median by more than 1.5x the interquartile range, which is indicated in red **C)** PCA plot showing clustering of proteins after adjustment where a group of samples run sequentially in a particular batch clustered separately from other samples **D**) Unsupervised hierarchical clustering with the sample type indicated showing a cluster of samples from MS batch R after adjustment.

3.3 Depth and Diversity of Host Mucosal Proteome

Individual proteins that contributed to the greatest proteome variation included complement factor C3, glycoprotein A1BG, albumin ALB, heme binding protein HPX, and vitamin D binding protein (VTDB), transporters ORM1, APOA1 and AFM, anti-protease A2M, and cell adhesion factor VTNC (Figure 6A). These do not represent the most abundant proteins, as only ALB was in the top 10 most abundant, with both VTNC and AFAM average in

abundance (rank 233 and 299 respectively). These proteins were co-expressed with no Pearson correlation below 0.60, and as they can all be found in blood microparticles, these proteins may be from blood or transudate that is transported into the vaginal lumen. Many of these proteins were transporter proteins important for regulating the selectivity of the epithelial barrier and the macromolecular contents and volumes of the vaginal lumen.

We next evaluated host pathway coverage in the mucosal proteome using DAVID Bioinformatics Resources. All BH significant pathways in the BP_direct category were taken after uploading 100% batch coverage proteins and <100% MS batch coverage proteins separately to identify what pathways were consistently seen by MS and what pathways were more difficult to capture. A total of 101 pathways were significantly enriched in the 576 proteins that had 100% MS batch coverage, 101 pathways were also significantly enriched in the 1467 proteins that were not detected in all batches, with a total of 145 unique pathways. There were 52 pathways observed in both 100% and <100% MS batch coverage. The pathways were involved mainly in: epithelial barrier (cell-cell adhesion and extracellular matrix organization), protein stability (protein ubiquitination, proteolysis, protein folding), signaling pathways (MAPK cascade, Fc-epsilon receptor signaling, T cell receptor signaling, stimulatory C-type lectin receptor signaling, positive regulation of Wnt signaling, NF-κB/ NF-κB inducing kinase (NIK) signaling) (Figure 6B). The 49 pathways significantly enriched only in the 100% MS batch coverage proteome included: immune response pathways (innate immune response, complement activation, defense response to bacterium, phagocytosis, acute phase response, antibacterial humoral response), epithelial barrier pathways (epidermis development, keratinocyte differentiation/keratinization, peptide crosslinking), and metabolic pathways (oxidation-reduction process, response to ROS, protein metabolic process, glycolysis/gluconeogenesis,

pentose phosphate shunt) (Figure 6C). The 43 pathways significantly enriched in only the <100% MS batch coverage proteome included many intracellular pathways: RNA associated pathways (rRNA processing, messenger RNA processing, termination of RNA polymerase II transcription), transport (endoplasmic reticulum to Golgi vesicle mediated transport, intracellular protein transport, endosomal transport, clathrin-mediated endocytosis), antigen processing and presentation (via MHC class I and class II), metabolic pathways (fatty acid β-oxidation, cholesterol biosynthesis, glycogen biosynthesis, purine biosynthesis, xenobiotic catabolism) (Figure 6D).

Unsupervised hierarchical clustering was used to identify major clusters based on protein expression variation (Figure 6E). We identified 6 major clusters, in the first cluster top DAVID pathways were defense response (50 proteins, enrichment p value=7.1E-29) and peptidase regulator activity (24, p=5.7E-23), and IPA identified acute phase response signaling (35, p=4.5E-53) and metabolism of cellular proteins (21, p=1.7E-24) as enriched. Cluster 2 had enriched cell-cell adhesion proteins and cell adhesion binding molecule using DAVID, with IPA there was enrichment in actin cytoskeleton signaling (9, p=5.7E-23) and cellular movement (38, p=3.3E-18). In cluster 3 we observed enrichment of oxidation-reduction (34, p=4.9E-11), protein binding in cell adhesion (25, p=3.1E-15), Phosphatidylinositol 3-Kinase /Protein AKT (PI3K/AKT) signaling (11, p=2.8E-10), and necrosis (68, p=1.E-18). Cluster 4 had enrichment in small molecule metabolic process (40, p=2.9E-7), peptidase activity (25, p=4.0E-8), protein ubiquitination pathway (13, p=3.2E-9), and degranulation of neutrophils (31, p=3.3E-24). In cluster 5 cell-cell adhesion (30, p=2.0E-10), protein binding in cell adhesion (22, p=2.8E-16), glucocorticoid receptor signaling (20, p=2.1E-17), and keratinization (27, p=3.9E-33). Cluster 6 contained proteins in tissue development (24, p=6.3E-4), protein binding in cell adhesion (15,

p=2.4E-8), B-cell Lymphoma 2 Associated Ananthogene Family Molecular Chaperone Regulator 2 (BAG2) signaling pathway (4, p=2.6E-5), and degranulation of cells (19, p=1.1E-11). We observed that in cluster 4 samples with high expression of neutrophil degranulation generally had lower levels of keratinization or cell-cell adhesion proteins (cluster 5).

3.4 Depth and Diversity of Bacterial Proteome

Mass spectrometry identified a total of 3,625 bacterial proteins in 688 (98.1%) cross sectional samples, and 13 samples (1.9%) had no bacterial proteins detected. Total bacterial spectral counts ranged from 0.7-1,425 with an average of 153 ± 180 spectral counts per sample, and an average of 43 ± 48 (range= 1-521) bacterial proteins detected per sample. The average bacterial to human protein signal ratio was 0.0313 ± 0.0418 (Figure 7A). Bacterial spectral count and protein count were highly correlated ($R_{Spearman}$=0.86, p<1E-15), as shown in Figure 7B. The majority of bacterial proteins identified were species level specific (1688 proteins, 48.4%) or strain specific (998, 24.9%, where peptides matched to a single protein in the database which had strain level information), while 586 (16.4%) belonged to a single genus, and only 357 (9.8%) could not be assigned to one genus and were classified as 'undistinguishable', however, apart from *Lactobacillus,* all proteins were binned to the genera level. The phylogenetic diversity included 8 genera from the phylum Firmicutes, 3 genera from Bacteroidetes, 3 genera from Proteobacteria, 4 genera from Actinobacteria, and 2 genera from Fusobacteria. Of the 20 bacterial genera identified, *Lactobacillus* had the most bacterial proteins detected (858, 23.7%), followed by *Gardnerella* (812, 22.4%), *Prevotella* (619, 17.1%), *Mobiluncus* (287, 7.9%), and *Pseudomonas* (239, 6.6%) (Figure 7C).

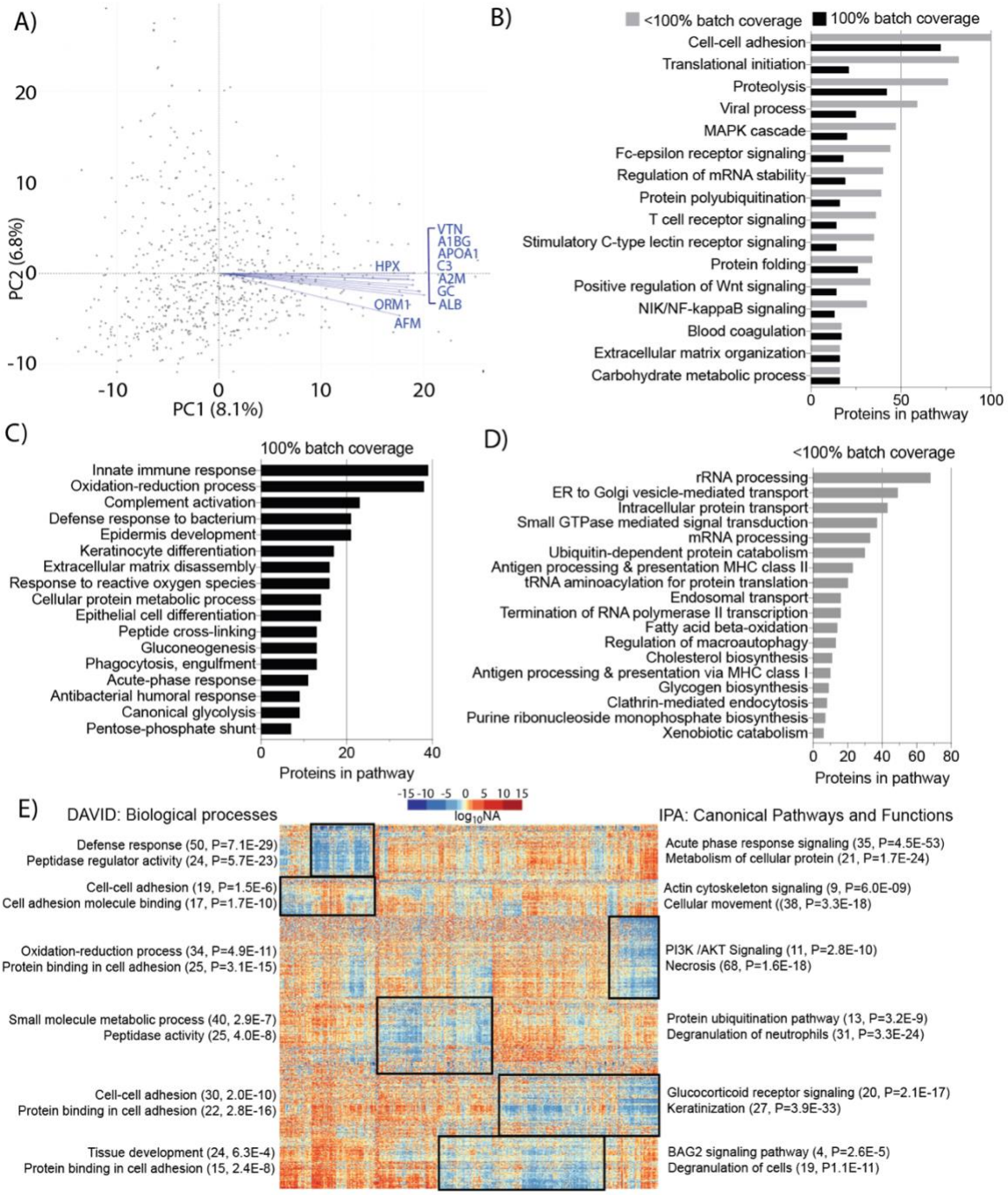

Figure 6. Biological pathways observed in the host proteome. A) PCA plot showing the sample distribution, top 10 most influential proteins indicated. Bar graphs showing the number of proteins in significantly enriched biological pathways in both 100% and <100% MS batch coverage **(B)**, only in 100% MS batch coverage proteome **(C)**, or only <100% MS batch coverage proteome (D). **E)** Heatmap of 100% MS batch coverage proteins across 701 samples. The top IPA (right) or DAVID (left) functions observed in proteins within highlighted boxes. Brackets indicating number of proteins identified in that pathway and P value which represents the enrichment of that pathway compared to a random dataset.

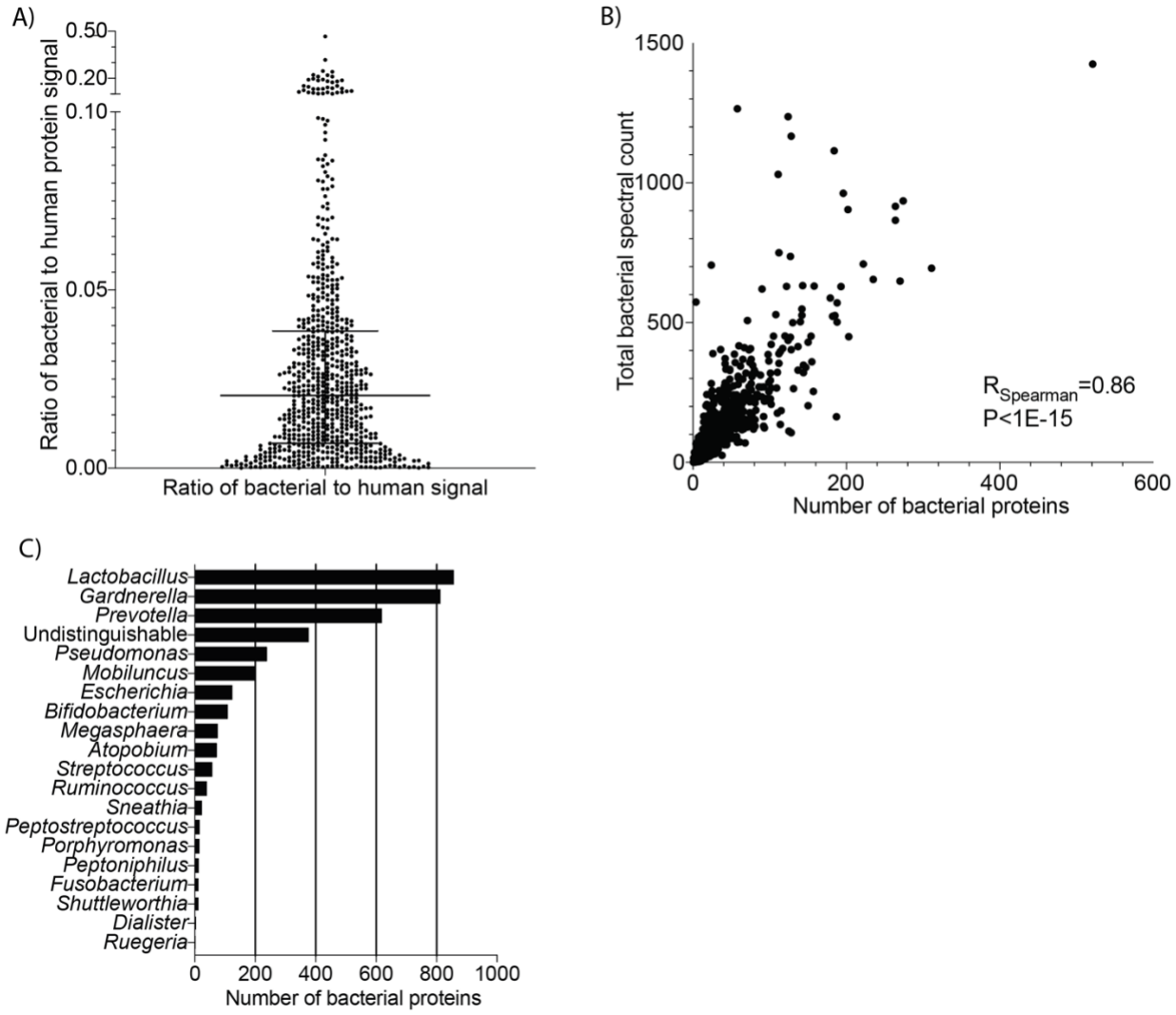

Figure 7. Overview of bacterial metaproteome data observed in CAPRISA-004. A) Plot showing the ratio of bacterial to human protein signal across cross sectional samples. **B)** Plot showing Spearman correlation between number of bacterial proteins and the total bacterial spectral count across cross sectional samples. **C)** Number of proteins identified stratified by genera.

A comparison of the metaproteomic data was made using the bacterial proteins identified in our dataset and assess how specific they were in a separate database. The published VIRGO database contains curated non-redundant metagenomic sequences from vaginal clinical samples and bacterial isolates, though some sequences in this database are classified as unannotated (sequence could not be assigned to a specific genera or species as it was seen in multiple species

or genera or had not yet been associated with any genera or species) (*270*). All proteins detected by MS from the most abundant genera (*Lactobacillus*, *Gardnerella*, *Prevotella*, and *Mobiluncus*) had their protein identifier (which links to the protein sequence) analyzed through the basic local alignment search tool for proteins (BLASTP) with the VIRGO database as the comparator database. Significant hits (Evalue (match p value) <1E-10) were assessed as a match to the protein identifier searched if they surpassed thresholds based on agreement of amino acid sequence between the input protein sequence and the database matches ("identity"), and total length of the protein that had matched amino acid sequence ("coverage"). The three thresholds were: 80% identity and 70% coverage, 90% identity and 80% coverage, and 95% identity and 90% coverage.

Genera specificity was assessed in four of the most abundant bacterial genera. At the strictest threshold 94.5% of Lactobacillus proteins (n=854) were specific to *Lactobacillus* (Figure 8A). Of the 809 *Gardnerella* proteins detected at the strictest threshold 85.7% were specific with the majority (85.7%) matching only to *Gardnerella* at the strictest threshold (Figure 8B). *Prevotella* (n=616 proteins) had the lowest percentage of specific proteins at the least strict threshold, 68.8% of proteins matched to other genera, however at the strictest threshold the non-specific protein percentage was reduced to 8.9% (Figure 8C). *Mobiluncus* was the most specific, at the strictest threshold 100% of the proteins only matched to *Mobiluncus* (Figure 8D).

The species *L. crispatus* and *L. iners* were the two most abundant *Lactobacillus* species and both have important implications for vaginal health (*63*, *271*, *272*). Thus, we also compared the species-specific proteins in *L. crispatus* and *L. iners* to the VIRGO dataset. Of the 311 proteins specific to *L. crispatus*, 77.6% at the strictest threshold were specific to *L. crispatus* (Figure 8E) though all sequences matched to *Lactobacillus* and/or unannotated sequences. *L. iners* proteins

(n=207 proteins) were more specific, with 98.4% not matching to other species at the strictest threshold (Figure 8F).

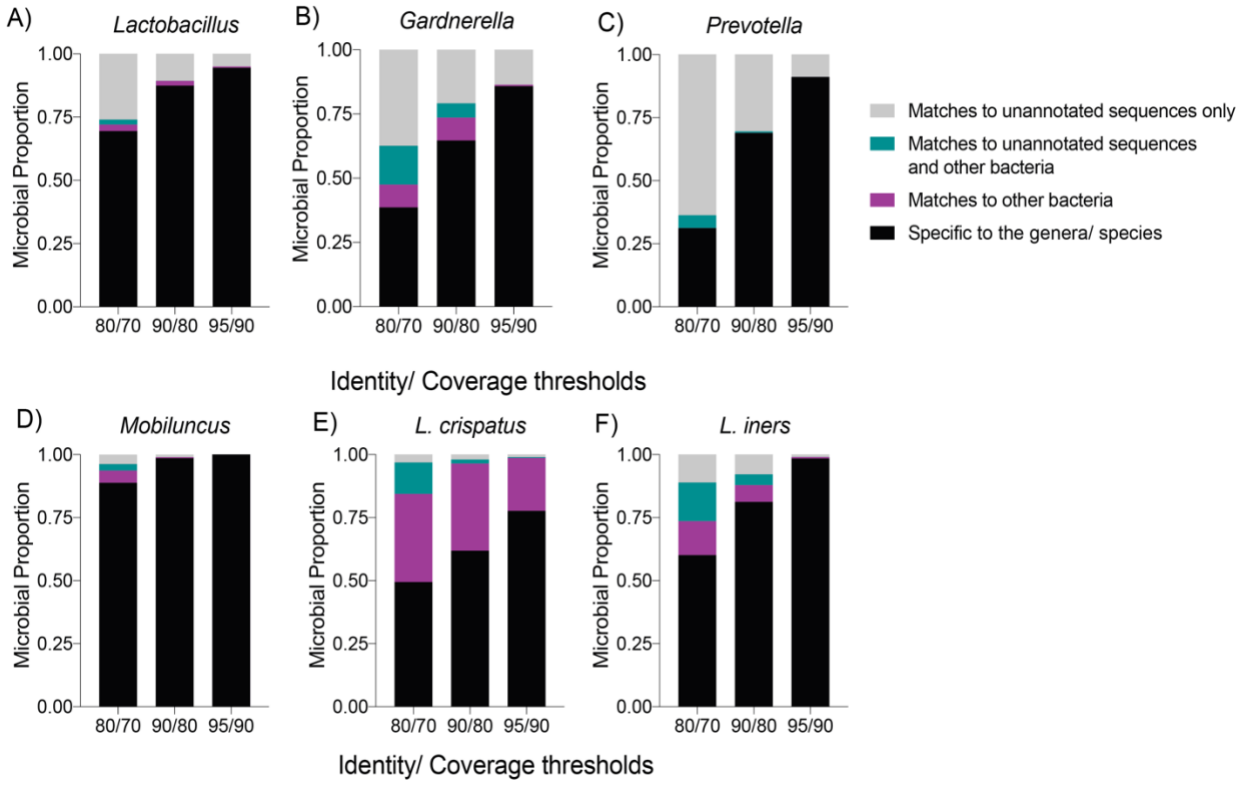

Figure 8. Specificity of proteins identified by MS compared to the vaginal metaproteomic database VIRGO using protein BLAST. Unannotated sequences were sequences not linked to a specific genera or species in the VIRGO database. Identity similarity and coverage overlap percentage of protein similarity were used at 3 separate thresholds: 80% identity and 70% coverage, 90% identity and 80% coverage, 95% identity and 90% coverage. **A)** Results of all *Lactobacillus* proteins (n=854) showed that at the strictest threshold. **B)** Results of all *Gardnerella* proteins (n=809). **C)** Results of all *Prevotella* proteins (n=616). **D)** Results of all *Mobiluncus* proteins (n=195). **E)** Results of all *L. crispatus* proteins (n=311). **F)** Results of all *L. iners* proteins (n=207).

We also wanted to assess if the specific peptides identified by MS were unique enough to specify a bacterial protein to the species and strain level. To do this, 10 bacterial proteins were randomly chosen for analysis by BLASTP to find bacterial proteins that closely match to the protein. For each protein, all of the associated peptides seen across all MS batches, the sequences were recorded, and these peptide sequences were used to assess if a unique amino acid

sequence of the protein was covered with these detected peptides compared to the closest (>80% amino acid sequence agreement) protein matches. We found that 6/10 proteins had ≥1 peptide(s) detected that covered unique portions of the protein compared to the closest alignment matches (Table 1). We also assessed the capability of MS to resolve proteins to strain level specificity as different bacterial strains or subtypes, such as *G. vaginalis*, can vary in virulence factor expression (*273*) and 27.5% of the proteins had strain level information, based on our database searches and confidence thresholds. When abundant proteins from each *G. vaginalis* subtype (A-E) were analyzed for unique peptides that differentiated them from close amino acid sequence matches, only 1 protein (subtype C) had unique peptides, though 2/5 proteins had non-specific peptides, but were the only protein that had all identified peptides (other matches were truncated) (Table 2).

Table 1. Randomly chosen bacterial proteins to determine appropriateness of species specificity of protein based on peptides detected by MS.

Protein	Bacteria (strain)	Protein length	# peptides identified	Coverage of protein (%)	Alignment identity, source of closest bacterial match	≥1 unique segment covered by detected peptides
Gram-positive signal peptide protein, YSIRK family	*L. iners* (LactinV 01V1-a)	157	3	36/157 (22.9%)	75.1%, *L. iners*	Yes
Uncharacterized protein	*Fusobacterium hwasookii* (ChDC F300)	2381	2	19/ 2381 (0.8%)	88.5%, *F. periodonticum*	Yes
Glyceraldehyde-3-phosphate dehydrogenase	*L. gallinarum*	338	9	104/338 (30.8%)	99.4%, *L. helveticus*	Yes
Glutamate dehydrogenase	*Prevotella sp.* (S7-1-8)	445	4	45/445 (10.1%)	95.5%, *P. colorans*	Yes
1,4-alpha-glucan branching enzyme	*Gardnerella vaginalis* (5-1)	1888	21	227/1888 (12.1%)	97.7%, *G. vaginalis*	Yes
D-lactate dehydrogenase	*L. crispatus* (DSM 20584)	337	12	121/337 (35.9%)	100% *L. crispatus*	No
MalE-type ABC sugar transport system periplasmic component	*Gardnerella vaginalis* (1500E)	448	11	126/448 (28.1%)	99.1% *Bifidobacteriaceae bacterium*	No
Formate C-acetyltransferase	*Megasphaera genomosp. type_1* (str. 28L)	762	20	198/762 (26.0%)	95.1% *M. genomosp. type_1*	Yes
Glyceraldehyde-3-phosphate dehydrogenase	*Prevotella bivia* (DNF00320)	342	8	85/342 (24.9%)	99.4%, *P. bivia*	No
Sugar ABC transporter substrate-binding protein	*Streptococcus iniae* (IUSA1)	444	3	22/444 (5.0%)	100%, *S. iniae*	No

Table 2. *G. vaginalis* proteins used to assess capability of MS to resolve to strain level based on peptides detected.

Subtype	Protein	Bacteria	Protein length	Number of peptides identified	Coverage of protein (%)	Alignment identity and source of closest bacterial match	≥1 unique segment covered by detected peptides
A	1,4-alpha-glucan branching enzyme	*G. vaginalis* 5-1	1,888	63	759/1888 (40.2%)	97.7%, *G. vaginalis*	No
B	M protein repeat protein	*G. vaginalis* JCP8017A	2,405	49	573/ 2405 (23.8%)	91.0% *G. vaginalis* or 96.4% *G. vaginalis JCP7659*	No
C	Uncharacterized protein	*G. vaginalis* 55152	548	14	161/458 (35.2%)	96.7% *G. vaginalis* 1400E	Yes
D	Putative sugar-binding secreted protein	*G. vaginalis* 6119V5	409	16	179/409 (43.8%)	100% *Bifidobacteriaceae bacterium* WP012	No
E	ABC transporter solute-binding protein	*G. vaginalis* JCP8481B	413	14	143/413 (34.6%)	99.8% *G. vaginalis*	No

Functional information from bacterial proteins was obtained using KEGG ontology through annotation by Ghost Koala. There were 2,309 (63.7%) bacterial proteins that mapped to one or more functions, and 676 (98.3%) of women who had bacterial proteins detected had at least one protein with functional annotations. In total we identified 22 KO level functions, after removal of non-bacterial associated pathways (such as cancer, or cardiovascular health). Within the B level functions, there were 108 KO level pathways (Figure 9). Carbohydrate metabolism

had the most KO levels, though among the two amino acid metabolism B levels there were 21 unique KO levels. Third was metabolism of cofactors and vitamins, followed by signal transduction (8), lipid metabolism (8), energy metabolism (7) and biosynthesis of other secondary metabolites (7). Both environmental adaptation and transcription both had a single KO pathway, while transport and catabolism, cell motility, and nucleotide metabolism had two KO level pathways each.

Through stratification of the number of proteins per B level pathway we observed that carbohydrate metabolism had the most proteins (n=1059) annotated, though unannotated proteins were more numerous (n=1319) (Figure 10A). Translation (n=486), energy metabolism (n=464), nucleotide metabolism (n=276), and amino acid metabolism rounded out the top five pathways with the most proteins. The five pathways that had the fewest proteins annotated were transport and catabolism (n=11), glycan biosynthesis and metabolism (n=17), metabolism of terpenoids and polyketides (n=20), environmental adaptation (25) and replication and repair (28). When looking at the frequency of detected pathways among the women's bacterial communities, carbohydrate metabolism and energy metabolism were the most ubiquitous, with 93.6% and 90.3% of communities respectively having that function (Figure 10B). Signal transduction (88.7%), folding sorting and degradation (83.6%), and membrane transport (76.3%) were the next most frequently detected pathways. Intermittently detected pathways in samples included translation (71.4%), amino acid metabolism (70.5%), nucleotide metabolism (68.9%), cell growth and death (40.8%), cellular community of prokaryotes (39.5%), metabolism of co-factors and vitamins (29.8%), transcription (21.2%), metabolism of terpenoids and polyketides (19.9%), metabolism of other amino acids (14.5%), xenobiotics biodegradation and metabolism (13.8%), and lipid metabolism (13.2%). Infrequently detected pathways in samples included transport and

catabolism (1.7%), biosynthesis of secondary metabolites (4.2%), glycan biosynthesis (4.5%), environmental adaptation (9.9%), replication and repair (10.5%), and cell motility (10.5%).

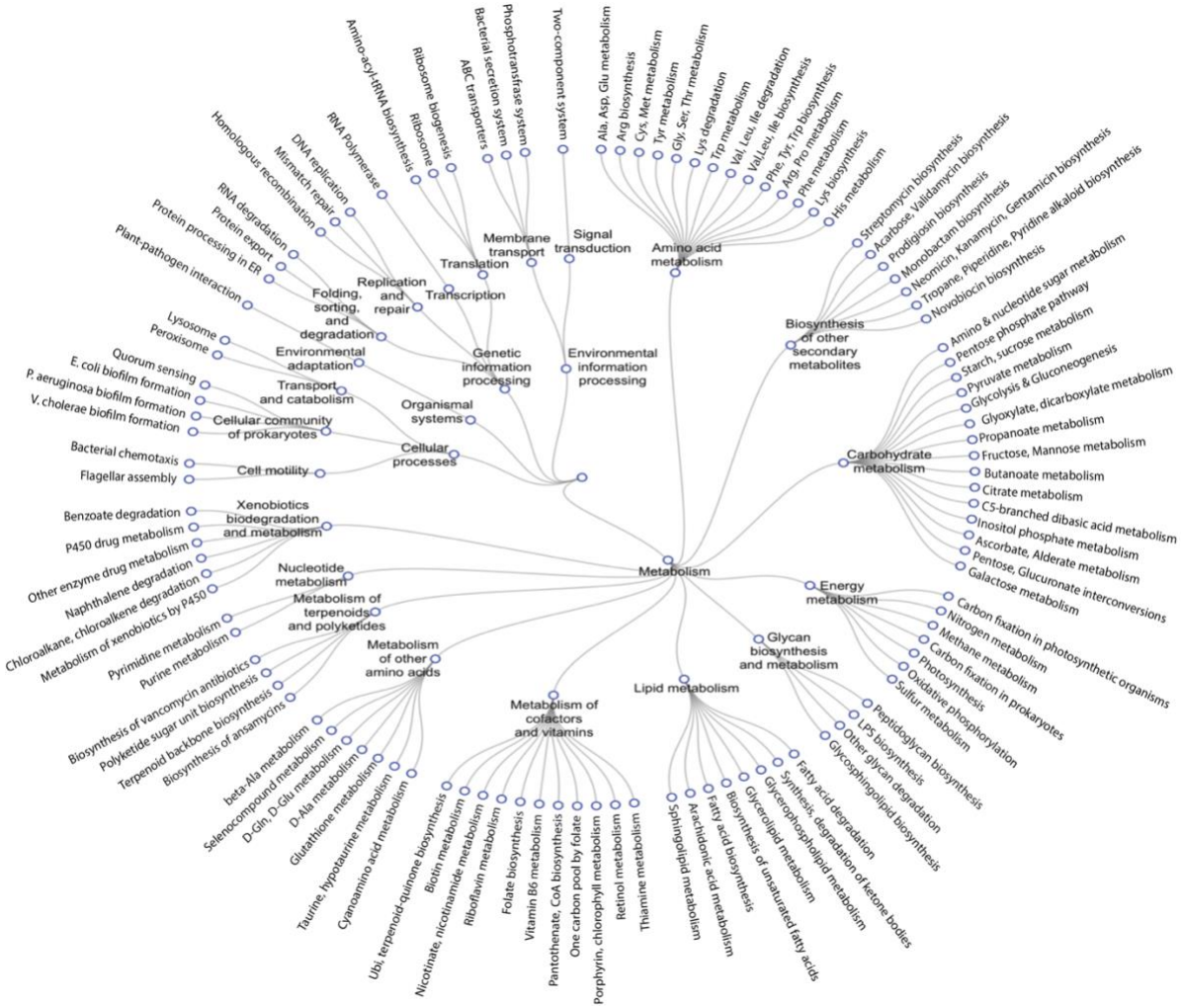

Figure 9. Diversity of bacterial KO level functions identified grouped by their B level category. Nodes in the centre show broader categories while the outer edges represent the ko level pathways that have more specified functions.

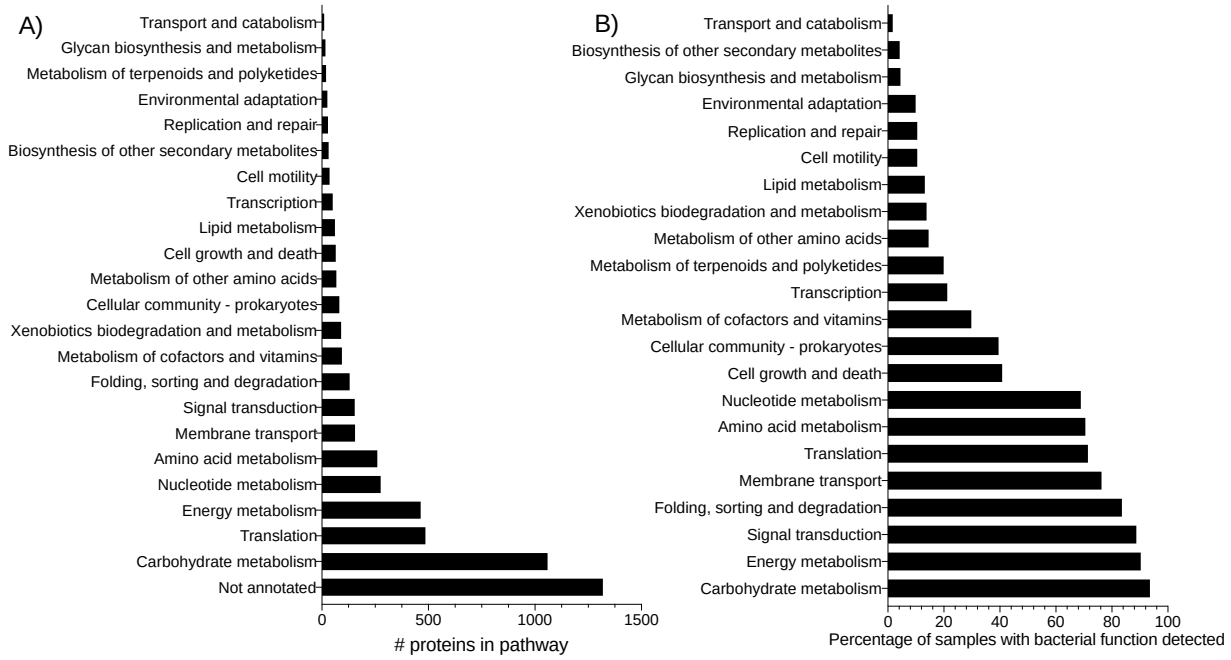

Figure 10. Bacterial functional pathways by abundance and frequency of detection. A) Number of proteins assigned to B level pathways, with a protein counted multiple times if it was annotated to multiple pathways. **B)** Percentage of samples with indicated B level functions detected.

We stratified *Lactobacillus* into *L. crispatus*, *L. iners*, and *Lactobacillus spp. (*non-*crispatus,* non-*iners*, still species specific (grouped together due to low prevalence of other *Lactobacillus* species*)* and *Lactobacillus* (not species specific) to assess how similar or dissimilar they may be acting, as shown in Figure 11A. Carbohydrate metabolism and energy metabolism were the most prevalent functions. It was observed that *L. iners* and *Lactobacillus spp.* had very similar functional profiles. While *L. crispatus* was somewhat similar to *L. iners* and *Lactobacillus spp*. functional profiles, the proportion of signal transduction proteins were highest in *L. crispatus* (12.6%) compared to *L. iners* (7.1%) or *Lactobacillus spp*. (8.4%), and non-specific *Lactobacillus* had a high proportion of signal transduction (12.4%) and translation (9.5%).

Among non-*Lactobacillus* bacteria, membrane transport contributed to a high proportion of the proteome in *Gardnerella* (16.6%), *Streptococcus* (17.3%), and *Sneathia* (16.2%), and these bacteria also had low levels of energy metabolism (<11%) relative to other genera (Figure 11A). *Mobiluncus* proteome consisted largely of functions involved in signal transduction (28.3%), carbohydrate metabolism (28.1%), and cell motility (10.4%). The proteome of *Prevotella* was mostly involved in energy production (24.9%) and carbohydrate metabolism (47.0%). *Pseudomonas* had the highest nucleotide metabolism proportion (9.8%).

As carbohydrate metabolism was the most abundant bacterial function identified, we further analyzed at the KO pathway level. Carbohydrate metabolism was most prominent in the genera *Prevotella*, *Gardnerella*, and *Lactobacillus* (Figure 11B). The most abundant carbohydrate metabolic pathways were glycolysis and pyruvate metabolism. All *Lactobacillus* had ≤5 proteins involved in the citrate cycle, which may reflect the preference for homolactic fermentation which produces lactic acid that vaginal lactobacilli are associated with (*159*, *274*). Conversely, proteins in the citrate cycle were most prominent in *Prevotella* (33 proteins, 12.7%), and *Pseudomonas* (16, 23.9%), and *Escherichia* (7, 15.6%) while *Gardnerella* did not have enzymes from the citrate cycle proteins detected. *Gardnerella* had the largest number of proteins in starch and sucrose metabolism (31, 18.3%), followed by *L. iners* (11, 17.2%), and *L. crispatus* (10, 15.2%), which may provide a possible explanation as to why these three are often dominant species in the vagina, they can catabolize complex carbohydrates, where other bacteria are more reliant on them to provide utilizable nutrients. Only *Gardnerella*, *Prevotella*, and *Megasphaera* had >10 proteins involved in butanoate metabolism which may indicate that these genera are the primary producers of butanoate, a common metabolite associated with vaginal dysbiosis (*274*).

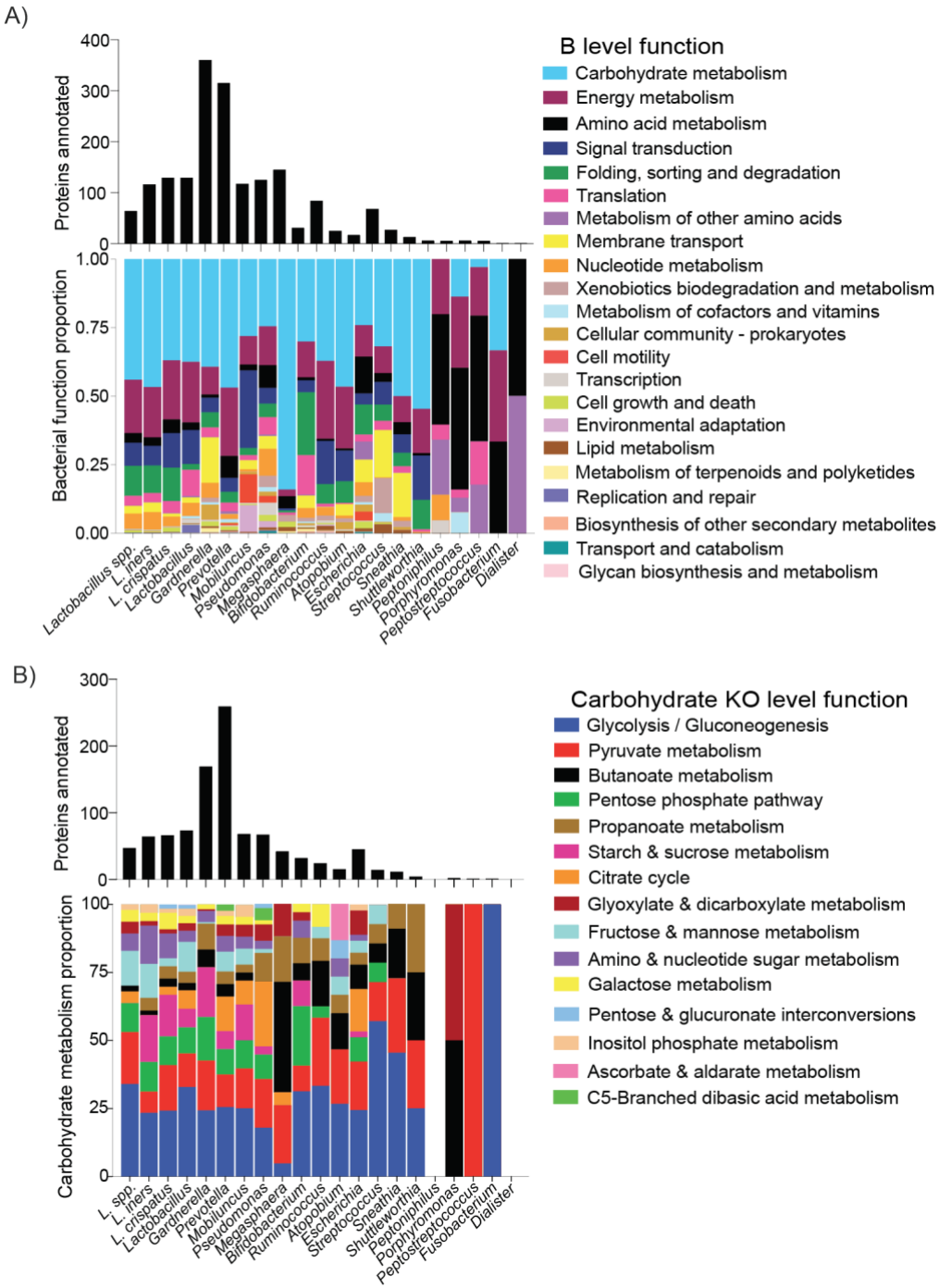

Figure 11. Overall functional pathway coverage across bacterial taxa in cervicovaginal lavage fluid by MS. A) Number of bacterial proteins identified by genera above proportional plots showing distribution of B-level KEGG ontology pathways within major bacterial taxa identified by mass spectrometry analysis, with *Lactobacillus* split into specific groups. Only proteins which were assigned a function are shown. **B)** Distribution and relative abundance of bacterial proteins involved in carbohydrate metabolism pathways at the KO level, separated by bacterial taxa, with total number of proteins annotated to a carbohydrate B level KO function above.

Chapter 4. A Metaproteomic Analysis of the Relationship Between the Vaginal Microbiome and Host Mucosal Proteome

4.1 Cohort Description

This cohort consisted of 889 women who were involved in the CAPRISA-004 There were 701 cross sectional samples from women of the total 889 women who participated in the CAPRISA-004 clinical trial, who consented to storage of samples and passed quality control measures. Overall, there was a 9.0% incidence of HIV, 9.7% incidence of HSV2, with a baseline HSV2 prevalence of 51.4%. The clinical and demographic characteristics of participants are shown in Table 3. Though the trial accepted women from ages 18-40, the majority were younger women with 66.0% under the age of 25 at enrolment, and of those 20.4% were under the age of 20. The trial had an equal balance between individuals assigned to the 1% tenofovir gel or the placebo gel, and the 49.9% of women assigned to the tenofovir arm in our subset reflects this. The majority (73.2%) of women were recruited from the rural area Vulindlela, and the minority were from the city of Durban. Most of the women had previously given birth to one or more children, and the majority of women had their first sexual act between the ages of 15-19, while a small subset (7.1%) had sex before the age of 15. Most women reported 2-5 sexual partners in their lifetime, while 35.4% only reported a single partner. Condom use was not assessed for every sex act during the trial, at baseline only 28.5% of women reported always using condoms, 16.8% reported never using condoms, while 18.4% reported using condoms most times and 36.2% reported occasionally using condoms.

Many behavioural and clinical questions were asked at each sample collection visit to obtain more accurate information on variables that can change. At the clinical visit when samples were collected the majority of women were using injectable contraceptives DMPA (65.8%) or norethisterone enanthate (NET-EN) (17.7%), while a smaller minority used COC

pills (14.1%), and the remaining 17 (2.4%) women used intrauterine devices, underwent surgery (3 tubal ligation, 2 had hysterectomies), with one woman who reported using male condoms only, and one woman did not list a contraceptive method. Only 11.6% of women reported any genital symptoms (including discharge, pain, bleeding, etc.), and 7.0% of women had at least some cervical ectopy, as diagnosed by a clinician. Testing for STI's was not performed on all women, only if there were indications, of the women tested, only 6.7% were positive at that clinical visit, though over the course of the trial 19.7% had at least one positive STI test result. Vaginal cleaning data was collected at each visit, with 4.9% of women reporting vaginal cleaning around or after sex acts (not including use of tenofovir or placebo gel), with 18.0% reporting doing it at some point during the trial. The vast majority of women reported using fingers only, of the 1103 reported acts of vaginal cleaning during the trial, 99.5% were using fingers only. As it has been reported that partners ≥5 years older are a risk for HIV transmission in South Africa (*261*), we assessed age difference between participants who reported having at least one sexual act in the month prior, 25.4% reported that their partner was at least 5 years older. Another risk factor is the HIV status of the partner, however only 3.6% of women reported knowing that their partner was HIV positive. Concurrent partnerships also increase risk of STI transmission and 11.0% of women were aware that their partner had other partners, while 6.6% of women reported having more than 6 sexual partners during their lifetime. There was an average of 5.7±5.3 sex acts per month reported by women, though 16.1% reported no sexual acts in the month preceding sample collection, and the majority (55.3%) reported sex in the week prior to sample collection. Median adherence was calculated for each individual woman by the CAPRISA team, and 60.6% had over 50% adherence to the proper gel use during the trial.

Table 3. Demographic and clinical characteristics of the 701 women analyzed by MS.

Variable	Average±SD; Range	n	%
Infections		n=701	
HIV seroconverters		63	9.0
HSV2 seroprevalence at baseline ▲		360	51.4
HSV2 incidence during trial ▲		68	9.7
Baseline Demographics			
Age at enrolment	23.9 ± 5.2; 18-40		
≤19		143	20.4
20-24		320	45.6
≥25		238	34.0
TFV arm		350	49.9
Rural		513	73.2
Live births	1.2 ± 0.9; 0-8		
None		147	21.0
1		378	53.9
≥2		176	25.1
Age at sexual debut	17.4 ± 2.1; 12-26		
Early (≤14)		50	7.1
Average (15-19)		554	79.0
Late (≥20)		97	13.8
Total lifetime sex partners	3.2 ± 9.5; 1-202		
1		248	35.4
2-5		407	58.1
≥6		46	6.6
Condom use			
Always		200	28.5
Most times		129	18.4
Occasionally		254	36.2
Never		118	16.8
Demographics at visit of sample collection			
Contraception use ¤			
DMPA		461	65.8
NET-EN		124	17.7
COC		99	14.1
Tubal ligation		13	1.9
Any genital symptoms*		81	11.6
Any level of cervical ectopy ¶		49	7.0
Positive test for STI infection at visit §		47	6.7
Positive STI test during trial §		138	
Vaginal cleaning reported at visit		34	
Vaginal cleaning ever during trial		126	
Antibiotic use in past 30 days		55	
Age difference of oldest partner and woman †	3.2 ± 3.9; (-35)-(+19)		
≥5 years older		178	25.4
Aware partner had positive HIV test		25	3.6
Aware partner has other partners		77	11.0
Sex acts in past 30 days	5.7 ± 5.3; 0-42		
Sex in the week before visit		388	55.3
>50% median adherence during trial		425	60.6

▲ 1 individual had no information on HSV2 status, 6 others had indeterminate results at either the enrolment or trial end serology test
¤ 1 woman had no contraception information, 3 used other forms not listed here
* 4 women had no information on genital symptoms reported
¶ 25 women did not have information on cervical ectopy at visit
§ Not all women were tested for STI's, 494 had no STI test at visit, 312 were not tested for STI's during trial
† Calculated as partner age - age of woman at visit. 113 women did not report sex that month so had no corresponding partner age, 2 women did not know age of partner

4.2 Major Bacterial Communities Observed in CAPRISA-004

Among the 688 women with bacterial proteins identified, there were 8 unique bacterial communities identified, which categorized 655 (95.2%) of the women with bacterial profiles based on proportional bacterial abundance. The communities were: >50% *L. crispatus*, >50% *Lactobacillus* (non-*L. crispatus*, non-*L. iners*, noted as *L. spp*), >50% *L. iners*, codominant *Lactobacillus* and *G. vaginalis*, *G. vaginalis* dominant, >25% *Mobiluncus*, >25% *Prevotella*, >40% *Pseudomonas*, with the proportional average for each community shown in Figure 12A, and the individual communities in Figure 12B. Functional profiles were available for the vast majority (639 women, 97.6%) of samples in the major bacterial community groups, with the proportional functional average profile for each community shown in Figure 13A, and the individual communities in Figure 13B.

There were 77 *L. crispatus* dominant women (11.2%), The average proportion of the top three major bacteria in the community were: 72.1% *L. crispatus*, 10.4% *Lactobacillus* (non-species specific), and 6.0% *L. iners*. The total bacterial spectral count median was 110.6 (IQR=41.3-163.9), with a median bacterial protein count of 45 (IQR=19-77). The Shannon's H diversity index median was 0.84 (IQR=0.71-1.00). There were 75 women (97.4%) who had bacterial functional information, a median of 9 functions were detected per sample (IQR=5.25-

11). The top 3 functions on average were: carbohydrate metabolism (37.9%), energy metabolism (23.6%), and signal transduction (13.5%).

A small proportion of women (21 women, 3.1%) were dominated by other *Lactobacillus* species or their microbiome contained multiple *Lactobacillus* species. The four most abundant bacteria were all lactobacilli: 26.6% *L. crispatus*, 23.4% *L. iners*, 21.4% non-specific *Lactobacillus*, 10.0% *L. jensenii*, with a total bacterial spectral count median of 75.4 (IQR=15.4-115.9). There was a median of 20 bacterial proteins per sample (IQR=5.5-40), with the median Shannon's H index of 0.98 (IQR=0.77-1.07). There were 19 women (90.5%) who had bacterial functional information, with a median of 7 functions detected (IQR=4-9). The top 3 functions on average were: carbohydrate metabolism (43.1%), energy metabolism (25.3%), and signal transduction (12.8%).

The most common *Lactobacillus* dominant community was *L. iners* dominant (286 women, 41.6%), with an average of 81.4% protein abundance belonging to *L. iners*, followed by *non-specific Lactobacillus* (9.7%), and then *L. crispatus* (3.7%). The total bacterial spectral count median was 123.0 (IQR=58.3-194.9), with a median of 26 proteins per sample (IQR=17-41). The median Shannon's H index was 0.48 (IQR=0.31-0.60). There were 283 women (99.0%) who had bacterial functional information, with a median of 8 functions detected in each sample (IQR=7-10). The top 3 B level functions on average were: carbohydrate metabolism (45.7%), energy metabolism (19.1%), and folding sorting and degradation (10.1%).

There was an intermediate community identified that had both *Lactobacillus* and *Gardnerella* codominance (30 women, 4.4%). The most prevalent genera observed were 37.5% *L. iners*, 29.4% *Gardnerella*, 7.2% of the proteins were undistinguishable at the genera level, and non-specific *Lactobacillus* (6.6%). The total bacterial spectral count median among the samples

was 80.6 (IQR=20.9-219.3) and the median number of bacterial proteins detected per sample was 28 (IQR=9.5-74.75). The Shannon's H diversity index median was 1.15 (IQR=0.56-1.37). There were 29 women (96.7%) who had bacterial functional information, and a median of 9 functions detected per woman (IQR=6.5-13). The top 3 B level functions on average were: carbohydrate metabolism (45.2%), energy metabolism (15.3%), and signal transduction (12.8%).

Amongst the non-*Lactobacillus* communities, *Gardnerella* dominance was the most prevalent (172 women, 25.0%) where the average proportion of *Gardnerella* was 68.8%, with undistinguishable proteins as the next most prevalent (13.5%), *Lactobacillus* (6.0%). The total bacterial spectral count median was 104.8 (IQR=38.0-241.5), with a median of 40 bacterial proteins detected per sample (IQR=17-67), with a median Shannon's H diversity index of 0.69 (IQR=0.40-0.92). There were 171 women (99.4%) who had bacterial functional information, with a median of 10 functions detected per woman (IQR=7-14). The top 3 B level functions on average were: carbohydrate metabolism (39.2%), membrane transport (20.0%), and energy metabolism (11.2%).

A community codominated by *Prevotella* was the second most prevalent non-*Lactobacillus* community (n=29, 4.2%). The *Prevotella* genus was the most abundant (41.6%), followed by *Gardnerella* at 28.0%, and *Lactobacillus* (9.3%). The total bacterial spectral count median among samples was 79.3 (IQR=22.4-279.6) with a median of 41 bacterial proteins detected per sample (11.5-102). The median Shannon's H diversity index was 1.06 (IQR=0.85-1.28). There were 28 women (96.6%) who had bacterial functional information, where a median of 11.5 functional pathways were identified per person (IQR=5-16.5). The top 3 B level functions on average were: carbohydrate metabolism (41.8%), energy metabolism (17.7%), and membrane transport (10.4%).

There was a community codominated by *Mobiluncus* (n=26, 3.8%). The most abundant genus was *Mobiluncus* at 43.3%, followed by *Gardnerella* (25.8%), and *Prevotella* 9.5%. The median total bacterial spectral count was 138.5 (IQR=36.4-451.6), with a median of 43 bacterial proteins detected per sample (IQR=16-137). The Shannon's H diversity index median across samples was 1.32 (IQR=1.01-1.55). All women in this group had bacterial functional information, and there was a median of 12 functions per person detected (IQR=7.75-18). The top 3 B level functions on average were: carbohydrate metabolism (31.1%), signal transduction (19.7%), and energy metabolism (12.0%).

A community with a high prevalence of *Pseudomonas* was observed in CAPRISA-004, though it was infrequent (15 women, 2.2%). The dominant genus was *Pseudomonas* (79.9%), followed by *Gardnerella* 6.7%, then *Lactobacillus* 6.4%. The total bacterial spectral count median was 4.5 (IQR=2.6-705.6). Few proteins were often identified in these samples (median= 2, IQR=1-43). The Shannon's H diversity index median was 0.19 (IQR=0.00-0.69). This group had the highest percentage of samples without functional information, with only 8 women (53.3%) having corresponding functional information, with a median of 10.5 functions identified per woman (IQR=6.5-18.25). The top 3 B level functions on average were: carbohydrate metabolism (25.3%), cellular community of prokaryotes (14.8%), and energy metabolism (12.8%).

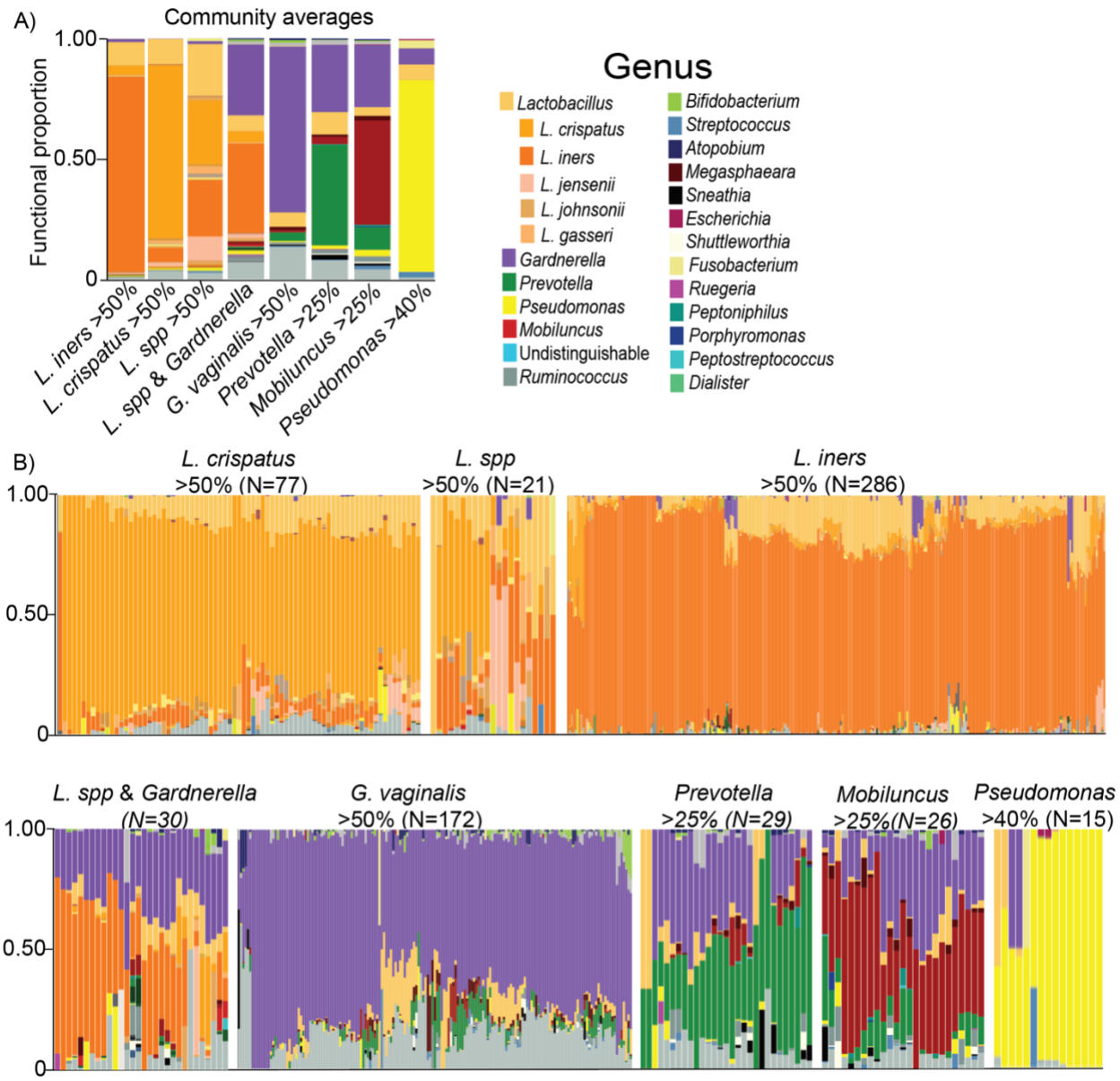

Figure 11. Bacterial communities' microbial profiles separated by community. A) Averages of the bacterial community proportion and legend. Only within the *Lactobacillus* groups is *Lactobacillus* genera separated into species, with the top 5 most abundant species specified in the legend **B)** Community profiles of samples by group.

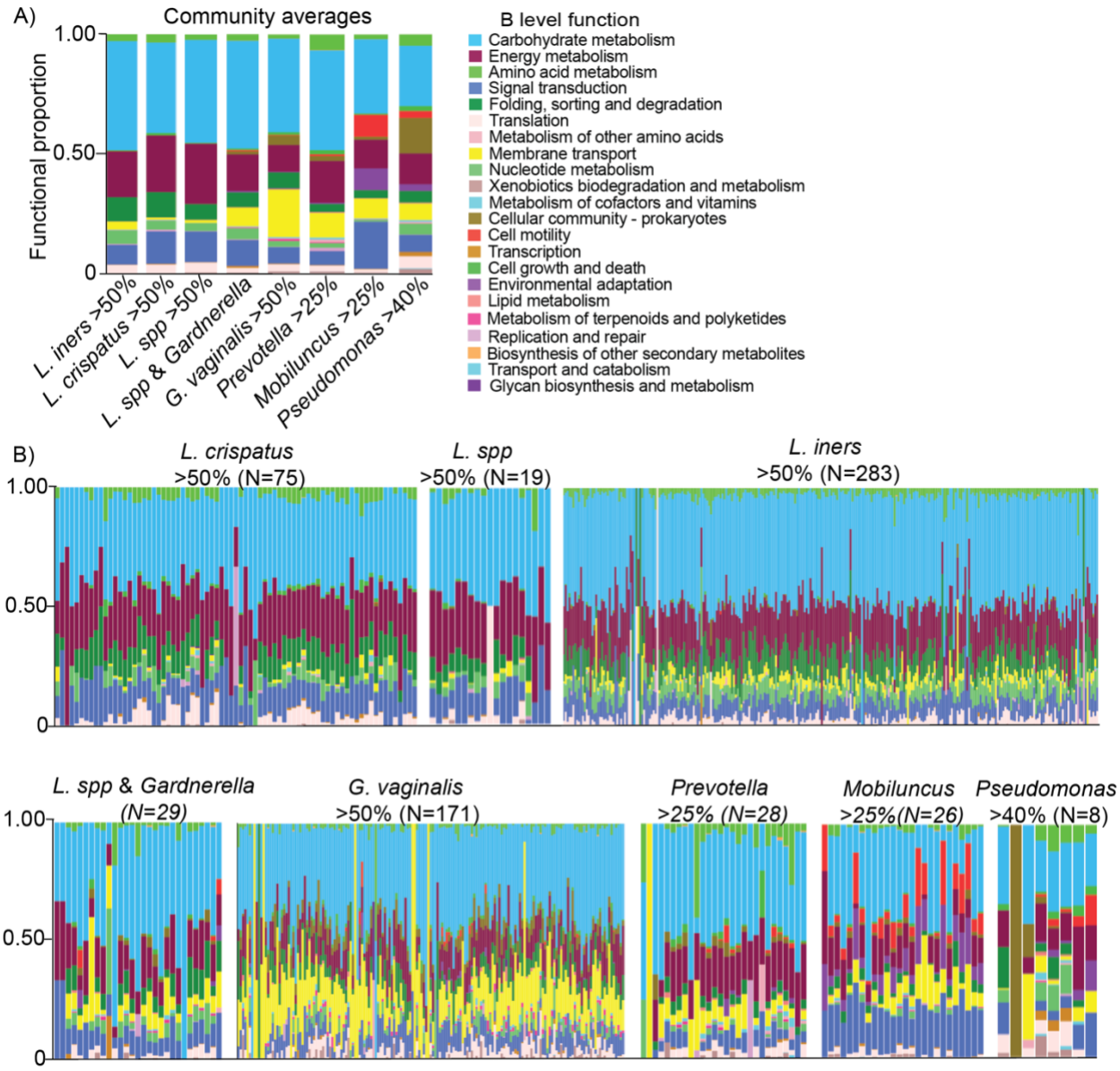

Figure 13. Bacterial functional profiles separated by community. A) Averages of the bacterial functional proportion at the B level by community and legend. **B)** Averaged functional profiles of samples by community group, with number of samples with information on function listed at the B level (as not all bacterial proteins were annotated to a function).

4.3 Host Proteome Differences of Major Bacterial Communities

In the vagina, a microbiome dominated by *Lactobacillus* is considered optimal for maintenance of genital health. However not all Lactobacillus species are equal, with *L. crispatus* thought to be one of ideal vaginal *Lactobacillus* species. Conversely, the presence of anaerobic

bacteria has been associated with vaginal symptoms, increased acquisition of sexually transmitted infections, and adverse reproductive health outcomes (*47, 91, 93, 101, 275-280*). However, little is known about the host immunological effects in the female genital tract in the presence of these bacteria. To identify differences in host proteome expression associated with the different vaginal communities, each community was compared to the *L. crispatus* dominant community.

Mild host differences were observed between women with *L. spp* and *L. crispatus* communities. There were no clinical or demographic differences between the two groups (Table 4). There were 66 host proteins (11.5%) that were significantly different ($p<0.05$), as shown in the volcano plot in Figure 14A. Only ribosome-binding protein 1 (RRBP1) was significant after multiple comparison corrections (BH). A heatmap of the proteins that were significant showed 2 clusters, though this did not perfectly sort the two different communities (Figure 14B). Among the 47 proteins that were downregulated in women with *L. spp* community there was an enrichment in cytoskeleton organization proteins (17 proteins, p.adj (BH) =2.1E-5), while the 17 slightly upregulated proteins were in pathways involved in nucleoside phosphate catabolic process (3 proteins p=0.0024, p.adj=0.69) and proteolysis (3 proteins p=0.019, p.adj=0.78). Analysis of IPA canonical pathways, which represent very specific signaling pathways and can be part of major biological functions such as inflammation, only nuclear factor erythroid 2-related factor 2 (NRF-2)-mediated oxidative stress response (involved in inflammation) was activated in *L. spp* communities (Z-score=2.00, 4 proteins, p=3.80E-4), while Liver-X-Receptor/Retinoid-X-Receptor (LXR/RXR) activation (also involved in inflammation) (Z-score=0, 4 proteins, p=0.00204) and ILK (integrin linked kinase) signaling (involved in cell-adhesion) (Z-score=0, 4 proteins, p=0.00200) showed no activation or inhibition. Activated

major functions identified by IPA in *L. spp* communities were cytolysis (Z-score=1.71, 9 proteins, p=1.2E-5), catabolism of proteins (Z-score=1.24, 20 proteins, p=0.00373), cell death of epithelial cells (Z-score=0.96, 8 proteins, p=9.3E-4) and apoptosis (Z-score=0.32, 26 proteins, p=4.5E-5) (Figure 14C). Inhibited IPA major functions organization of cytoskeleton (Z-score=-1.14, 15 proteins, p=0.00122), cell death of T lymphocytes (Z-score=-0.85, 5 proteins, p=0.00695), cell movement (Z-score=-0.35, 22 proteins, p=3.4E-4), and invasion of cells (Z-score=-0.30, 11 proteins, p=0.00525) (Figure 14D). Comparison to the immune cell proteome libraries using Gene Set Enrichment Analysis (GSEA) identified no protein signatures that matched the protein signature of women with *L. spp* communities that were significant ($p<0.05$).

Table 4. Demographic and clinical characteristics of women with communities of *L. crispatus* and *L. spp*.

Variable	>50% *L. crispatus* women	>50% *Lactobacillus* women (*L. spp*)	P value
	n=77	n=21	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	24.7±5.5; 18-40	0.584
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.7±0.4; 0.9-2.3	0.400
HIV seroconverters (n, %)	4 (5.2)	2 (9.5)	0.829
HSV2			
Baseline seroprevalence (n, %)	32 (41.6)	9 (42.9)	0.887
HSV2 incidence during trial (n, %)	4/45 (8.9)	2/12 (16.7)	0.0629
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	9 (42.9)	0.699
Rural (n, %)	51 (66.2)	18 (85.7)	0.143
Given Birth (n, %)	60 (77.9)	15 (71.4)	0.740
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.3±1.9; 12-20	0.652
Total lifetime sex partners (Mean ± SD; range)	3.8±11.4; 1-100	3.1±4.3; 1-20	0.618
Condom use (n, %)			
Always	24 (31.2)	5 (23.8)	0.836
Most times	16 (20.8)	4 (19.0)	
Occasionally	25 (32.5)	9 (42.9)	
Never	12 (15.6)	3 (14.3)	
Demographics at visit of sample collection			
Contraception use (n, %)			
DMPA	55 (71.4)	14 (66.7)	0.910
NET-EN	13 (16.9)	4 (19.0)	
COC	9 (11.7)	3 (14.3)	
Vaginal discharge (n, %)	7 (9.1)	3 (14.3)	0.772
Presence of STI infection (n, %) §	2/20 (10.0)	2/8 (25.0)	0.669
Any STI during trial (n, %) §	11/43 (25.6)	4/15 (26.7)	1.000
Any cervical ectopy (n, %) ¶	4 (5.2)	2 (9.5)	0.763
Antibiotic use in past 30 days (n, %)	6 (7.8)	1 (4.8)	1.000
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	7.0±6.1; 0-20	0.500
Age difference of oldest partner (Mean ± SD;	3.7±3.1; (-3)-12	3.3±3.7; (-2)-15	0.471
Median adherence during trial >50% (n, %)	46 (59.7)	13 (61.9)	1.00
Vaginal cleaning reported at visit (n, %)	8 (10.4)	0 (0.0)	0.275
Vaginal cleaning ever during trial (n, %)	19 (24.7)	2 (9.5)	0.230

§ 70 women were not tested for STI's at the visit, 40 women were not tested for STI's during the trial

¶ 4 women had no information at the visit of cervical ectopy

† 16 women did not report sexual acts in the previous 30 days so had no partner age reported

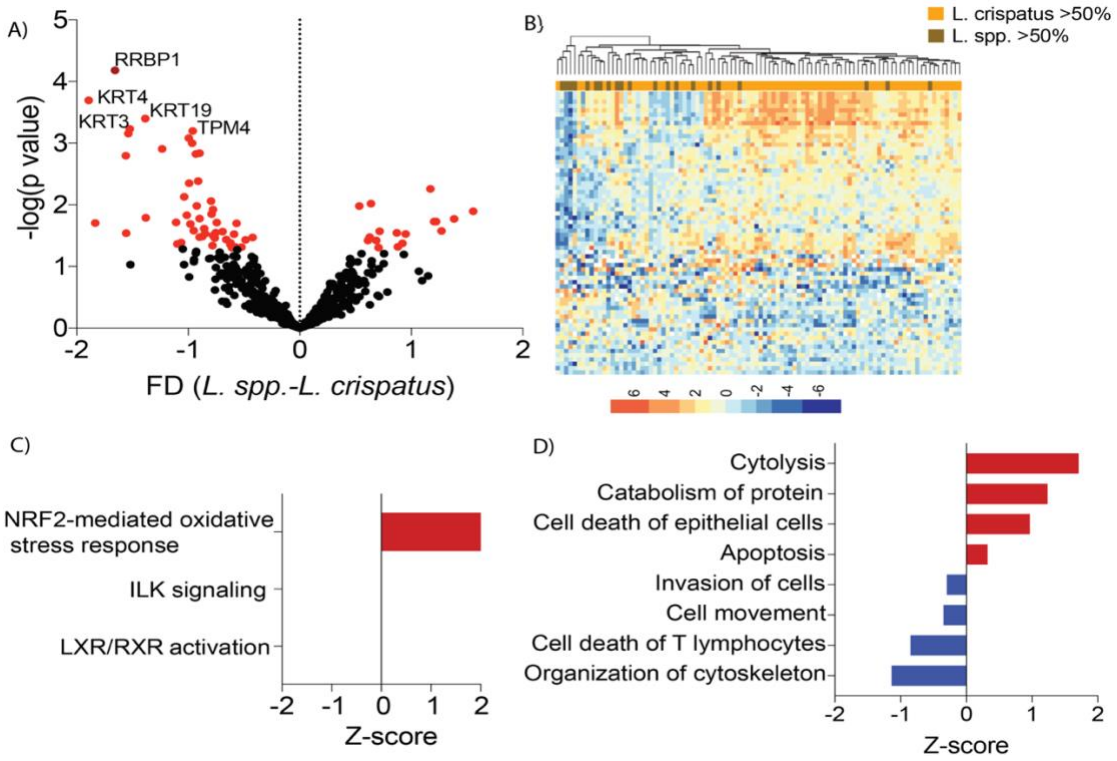

Figure 14. Overview of host differences between women with *L. spp.* and *L. crispatus* communities. A) Volcano plot showing the proteins significantly different ($p<0.05$) in red, with the five most significant proteins indicated. **B)** Unsupervised hierarchical clustered heatmap of the significant proteins different between *L. spp* and *L. crispatus* women. **C)** Significant IPA canonical pathways that had a numerical Z-score assigned to reflect the activation or inhibition in *L. spp* women. **D)** Significant IPA host functions showing activation (red) or inhibition (blue) observed in *L. spp* community samples.

There has been much debate about how beneficial *L. iners* is in the vagina. We identified 82 host proteins (14.2%) that were significantly different after BH corrections ($p.adj<0.05$) between women with *L. iners* and *L. crispatus* communities, as shown in the volcano plot in figure 15A. There were no demographic or clinical differences observed between *L. iners* and *L. crispatus* women, apart from an increased HSV2 incidence during the trial (24.5% vs 8.9%, $p=0.0423$) as shown in Table 5. A heatmap of the proteins that were significant showed some separation between those with *L. iners* and *L. crispatus* though the clustering did not sort the two groups well (Figure 15B). The 42 proteins that were higher in women with *L. iners* community had pathways enriched in defense response to bacteria (8 proteins, $p.adj=0.0034$), extracellular

matrix organization (8 proteins, p.adj=0.0074), and innate immune response (11 proteins, p.adj=0.0081). The 40 downregulated proteins were in in cytoskeleton organization (15 proteins, p.adj=1.2E-5), and regulation of humoral immune response (3 proteins, p.adj=0.42, p=0.0060). The only activated IPA canonical pathway was the inflammatory acute phase response signaling (Z-score=2.00, 4 proteins, p=0.00407) (Figure 15C). The two inhibited canonical pathways in *L. iners* women were ILK signaling (Z-score=-1.00, 5 proteins, p=6.31E-4) and LXR/RXR activation (Z-score=-1.00, 4 proteins, p=9.77E-4). Activated functions identified by IPA in *L. iners* women included pathways associated with increased innate immune responses including: inflammatory response (Z-score=2.12, 13 proteins, p=4.30E-4), migration of cells (Z-score=1.92, 35 proteins, p=1.09E-9), and activation of cells (Z-score=1.77, 22 proteins, p=1.32E-8), in particular activation of leukocytes (Z-score=1.09, 15 proteins, p=2.96E-6), as shown in Figure 15D. *L. iners* women showed decreased adaptive immune response involvement, with inhibited cell proliferation of T lymphocytes (Z-score=-1.10, 11 proteins, p=3.50E-4) and response of APCs (Z-score=-0.56, 8 proteins, p=6.98E-6). The other overall function observed that was inhibited in *L. iners* women was functions associated with an increased adaptive response including inhibition of apoptosis of leukocytes (Z-score=-0.93, 10 proteins, p=1.11E-4) in particular apoptosis of macrophages (Z-score=-2.00, 4 proteins, p=9.40E-4), however phagocytosis of cells was also significantly inhibited (Z-score=-0.69, 13 proteins, p=3.22E-9). Two immune cell signatures matched the host proteome signature of *L. iners* women, the up signature of CD8 T cells compared to NK cells (NES (Normalized enrichment score) =1.69, 22 proteins, p=0.00984) (Figure 15E), and the down signature of CD4 T cells compared to neutrophils (where the neutrophil up signature matched the *L. iners* protein signature) (NES=1.65, 52 proteins, p=0.00781 (Figure 15F).

Table 5. Demographic and clinical characteristics of women with *L. crispatus* and *L. iners* communities.

Variable	>50% *L. crispatus* women	>50% *L. iners* women	P value
	n=77	n=286	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	24.3±5.0; 18-41	0.203
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.6±0.5; 0.3-2.4	0.741
HIV seroconverters (n, %)	4 (5.2)	24 (8.4)	0.489
HSV2 ▲			
Baseline seroprevalence (n, %)	32 (41.6)	145 (50.7)	0.177
HSV2 incidence during trial (n, %)	4/45 (8.9)	34/139 (24.5)	0.0423
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	142 (49.7)	0.978
Rural (n, %)	51 (66.2)	203 (71.0)	0.505
Given Birth (n, %)	60 (77.9)	229 (80.1)	0.798
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.5±2.1; 12-26	0.325
Total lifetime sex partners (Mean ± SD;	3.8±11.4; 1-100	3.4±12.1; 1-202	0.858
Condom use (n, %)			
Always	24 (31.2)	90 (31.5)	0.733
Most times	16 (20.8)	46 (16.1)	
Occasionally	25 (32.5)	108 (37.8)	
Never	12 (15.6)	42 (14.7)	
Demographics at visit of sample			
Contraception use (n, %) ¤			
DMPA	55 (71.4)	192 (67.1)	0.872
NET-EN	13 (16.9)	53 (18.5)	
COC	9 (11.7)	36 (12.6)	
Vaginal discharge (n, %)	7 (9.1)	20 (7.0)	0.705
Presence of STI infection (n, %) §	2/20 (10.0)	19/81 (23.5)	0.308
Any STI during trial (n, %) §	11/43 (25.6)	56/147 (38.1)	0.184
Any cervical ectopy (n, %) ¶	4 (5.2)	19 (6.6)	0.816
Antibiotic use in past 30 days (n, %)	6 (7.8)	27 (9.4)	0.823
Sex acts in past 30 days (Mean ± SD;	6.2±6.3; 0-39	5.8±5.4; 0-42	0.846
Age difference of oldest partner (Mean ±	3.7±3.1; (-3)-12	3.4±3.0; (-12)-12	0.524
Median adherence during trial >50% (n,	46 (59.7)	172 (60.1)	1
Vaginal cleaning reported at visit (n, %)	8 (10.4)	13 (4.5)	0.094
Vaginal cleaning ever during trial (n, %)	19 (24.7)	54 (18.9)	0.334

▲ 2 women had undetermined HSV2 status

¤ 4 woman were using another form of birth control, 1 woman had no information at visit

§ 258 women were not tested for STI's at the visit, 173 women were not tested for STI's during the trial

¶ 14 women had no information at the visit of cervical ectopy

† 61 women did not report sexual act in the previous 30 days so had no partner age reported

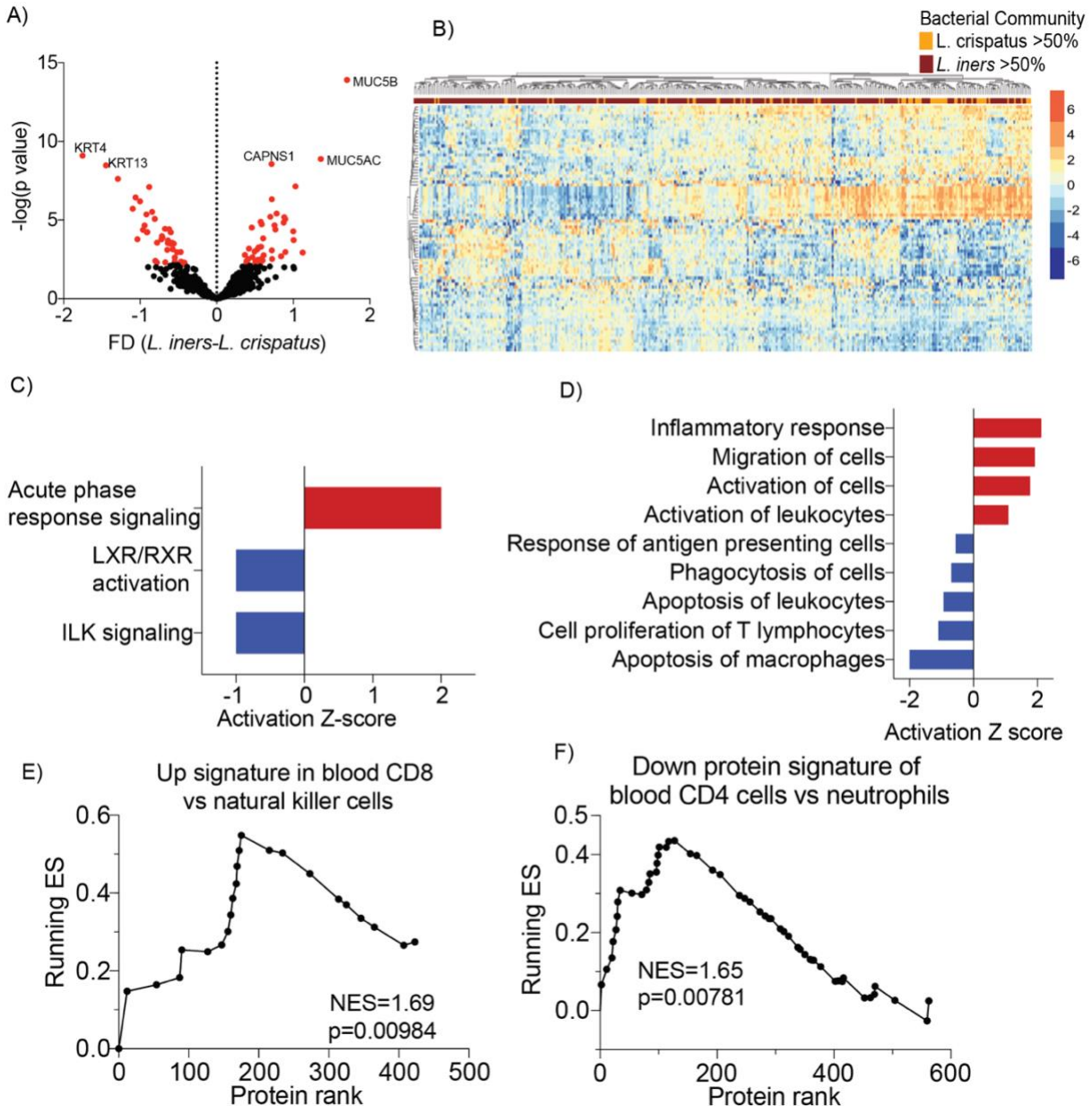

Figure 14. Overview of host mucosal proteome differences between women with *L. iners* and *L. crispatus* vaginal communities. A) Volcano plot showing the proteins significantly different ($p<0.05$) in red, with the five most significant proteins indicated. **B)** Unsupervised hierarchical clustered heatmap of the significant proteins different between *L. iners* and *L. crispatus* women. **C)** Significant IPA canonical pathways that were activated (red) or inhibited (blue) in *L. iners* women. **D)** Significant IPA host functions that were activated (red) or inhibited (blue) in *L. iners* women. **E)** Overlapping protein signature of *L. iners* with the up signature in blood CD8 T cells vs NK cells. **F)** Overlapping protein signature of *L. iners* with the down signature in blood CD4 T cells vs neutrophils (*L. iners* protein signature matches neutrophil up signature vs CD4 T cells).

The subset of women with codominance of *Lactobacillus* and *Gardnerella* may represent women who are transitioning to or from a dysbiotic state. There were 183 host proteins (31.8%) that were significantly different (BH corrected $p<0.05$) between women with codominant *Lactobacillus* and *Gardnerella*, and *L. crispatus* communities, as shown in the volcano plot in figure 16A. There were no significant differences in demographic or clinical characteristics between the two communities (Table 6). A heatmap of the proteins that were significant showed 2 clusters (Figure 16B) with the cluster that was upregulated in women with *Lactobacillus* and *Gardnerella* communities had pathways enriched in innate immune response (20 proteins, p.adj=8.2E-7) and small molecule metabolic process (36 proteins, p.adj=1.0E-10), in particular the amine metabolic process (11 proteins, p.adj=5.5E-8) and detoxification (9 proteins, p.adj=1.1E-6). The proteins that were decreased in women with *Lactobacillus* and *Gardnerella* communities were in epithelial barrier pathways including cell-cell adhesion (32 proteins, p.adj=2.3E-11), tissue development (30 proteins, p.adj=4.2E-6), and keratinization (10 proteins, p.adj=1.6E-9). IPA canonical pathways that were activated in *Lactobacillus* and *Gardnerella* communities were acute phase response signaling (Z-score=1.89, 7 proteins, p=5.25E-4), gluconeogenesis I (Z-score=1.00, 4 proteins, p=4.68E-5), apoptosis signaling (Z-score=0.45, 5 proteins, p=0.00107) and ILK signaling (Z-score=0.38, 7 proteins, p=7.4E-4) (Figure 16C). The LXR/RXR pathway was not activated or inhibited (Z-score=0, 4 proteins, p=0.0148), while eukaryotic initiation factor 2 (EIF2) signaling was inhibited (Z-score=-2.00, 6 proteins, p=0.00832). Activated functions in *Lactobacillus* and *Gardnerella* women identified by IPA showed increased immune activation (chemotaxis of granulocytes (Z-score=2.57, 9 proteins, p=4.27E-7), accumulation of neutrophils (Z-score=1.45, 5 proteins, p=6.96E-4), antimicrobial response(Z-score=1.34, 12 proteins, p=5.92E-4), and degranulation of cells (Z-score=1.19, 42

proteins, p=1.04E-22), and leukocyte migration (Z-score=0.66, 24 proteins, p=6.67E-4)), and increased cell movement of keratinocytes (Z-score=2.57, 9 proteins, p=4.27E-7) Figure 16D. The inhibited functions in samples with codominant *Lactobacillus* and *Gardnerella* communities were apoptosis (Z-score=-1.07, 73 proteins, p=3.47E-12) and cytotoxicity (Z-score=-1.78, 9 proteins, p=5.81E-4). There were three significant protein signatures that matched the protein signature of women with *Lactobacillus* and *Gardnerella* codominant communities: the up signature in CD8 T cells vs NK cells (NES=1.66, 22 proteins, p=0.00525, Figure 16E) and the down signature of monocytes vs neutrophils (NES=1.54, 45 proteins, p=0.00120 Figure 16F), with a similar weaker down signature observed between B cells and neutrophils (NES=1.32, 68 proteins, p=0.0360, data not shown).

Table 6. Demographic and clinical characteristics of women with *L. crispatus* and *Lactobacillus* and *Gardnerella* communities.

Variable	>50% *L. crispatus* community women	*Lactobacillus* & *Gardnerella* women	P value
	n=77	n=30	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	24.9±4.4; 19-36	0.865
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.6±0.5; 0.5-2.3	0.481
HIV seroconverters (n, %)	4 (5.2)	3 (10.0)	0.64
HSV2			
Baseline seroprevalence (n, %)	32 (41.6)	14 (46.7)	0.793
HSV2 incidence during trial (n, %)	4/45 (8.9)	2/16 (12.5)	0.943
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	16 (53.3)	0.973
Rural (n, %)	51 (66.2)	21 (70.0)	0.886
Given Birth (n, %)	60 (77.9)	22 (73.3)	0.803
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.7±2.3; 12-22	0.961
Total lifetime sex partners (Mean ± SD; range)	3.8±11.4; 1-100	2.6±2.1; 1-10	0.906
Condom use (n, %)			
Always	24 (31.2)	8 (26.7)	0.906
Most times	16 (20.8)	8 (26.7)	
Occasionally	25 (32.5)	10 (33.3)	
Never	12 (15.6)	4 (13.3)	
Demographics at visit of sample collection			
Contraception use (n, %) ¤			
DMPA	55 (71.4)	18 (60.0)	0.279
NET-EN	13 (16.9)	4 (13.3)	
COC	9 (11.7)	7 (23.3)	
Vaginal discharge (n, %)	7 (9.1)	2 (6.7)	0.986
Presence of STI infection (n, %) §	2/20 (10.0)	1/10 (10.0)	1.000
Any STI during trial (n, %) §	11/43 (25.6)	3/20 (15.0)	0.539
Any cervical ectopy (n, %) ¶	4 (5.2)	2 (6.7)	1.000
Antibiotic use in past 30 days (n, %)	6 (7.8)	4 (13.3)	0.607
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	6.3±5.3; 0-20	0.694
Age difference of oldest partner (Mean ± SD;	3.7±3.1; (-3)-12	3.4±3.3; (-3)-10	0.706
Median adherence during trial >50% (n, %)	46 (59.7)	17 (56.7)	0.943
Vaginal cleaning reported at visit (n, %)	8 (10.4)	5 (16.7)	0.573
Vaginal cleaning ever during trial (n, %)	19 (24.7)	8 (26.7)	1

¤ 1 woman was using another form of birth control

§ 77 women were not tested for STI's at the visit, 44 women were not tested for STI's during the trial

¶ 2 women had no information at the visit of cervical ectopy

† 19 women did not report sexual act in the previous 30 days so had no partner age reported

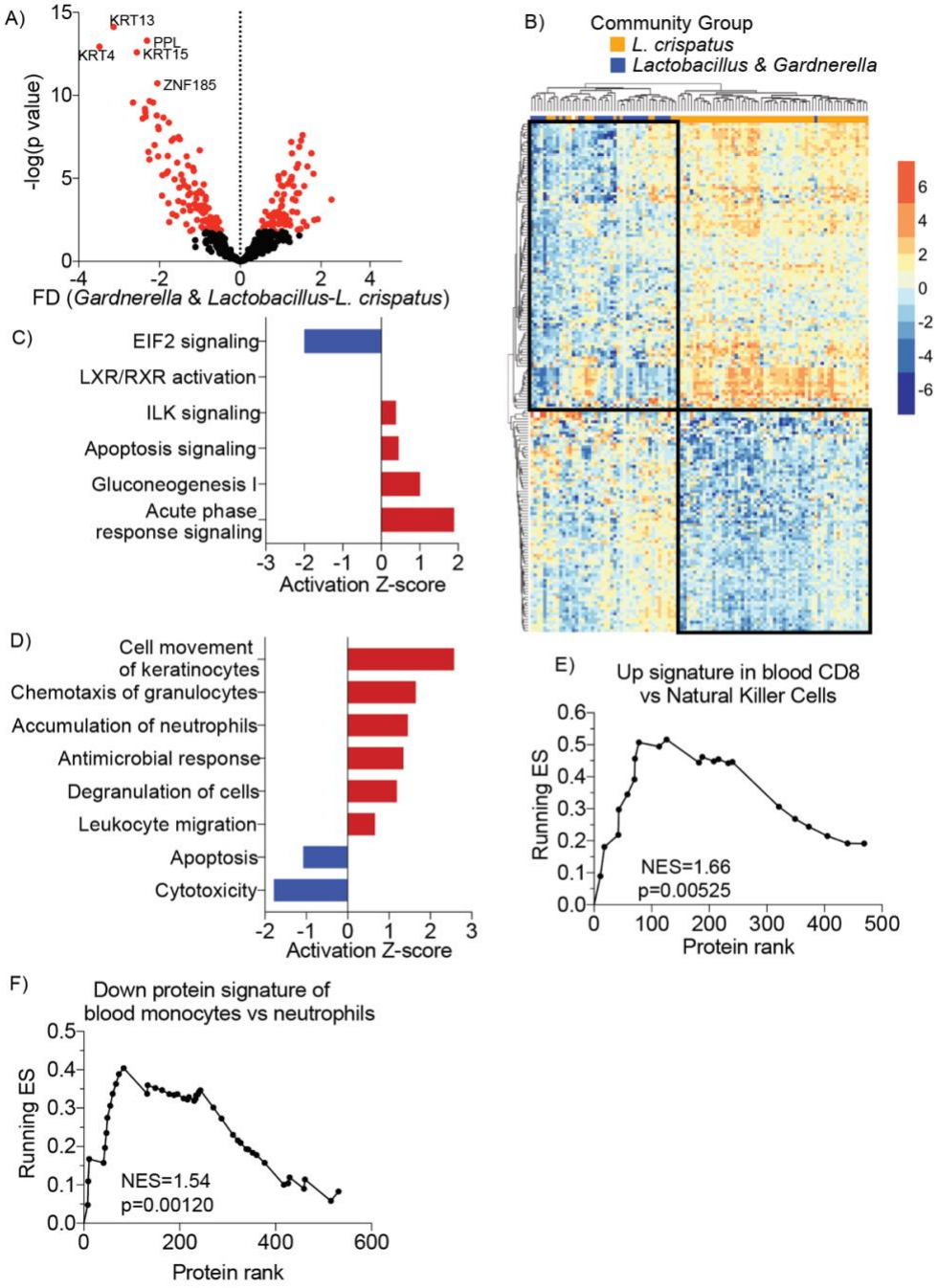

Figure 16. Overview of host differences between women with codominance of *Lactobacillus/Gardnerella* and *L. crispatus* communities. A) Volcano plot showing the proteins significantly different (BH adjusted p<0.05) in red, with the five top proteins indicated. **B)** Heatmap of the significant proteins different between codominant *Lactobacillus/Gardnerella* and *L. crispatus* women. **C)** IPA canonical pathways with the Z-score indicating activated (red) or inhibited (blue) in *Lactobacillus/Gardnerella* women. **D)** IPA functions that were activated (red) or inhibited (blue) in codominant women. Overlapping protein signatures of *Lactobacillus/Gardnerella* communities with **E)** the up protein signature in blood CD8 T cells compared to NK cells or **F)** the down protein signature in blood monocytes vs neutrophils.

Gardnerella is one of the most common non-*Lactobacillus* bacteria identified in the vagina and can be found in both asymptomatic women and those diagnosed with BV. There were 346 host proteins (60.1%) that were differentially expressed between women with *Gardnerella* and *L. crispatus* communities after multiple comparison correction (BH), as shown in the volcano plot in Figure 17A. The only difference among clinical and demographic characteristics was decreased reports of vaginal cleaning at the visit of sample collection in *Gardnerella* dominant women (2.3% vs 10.4%, p=0.0153 (Table 7). A heatmap of the proteins that were significant showed 3 clusters (Figure 17B). The protein cluster upregulated in the majority of women with a *Gardnerella* dominant community had pathways enriched in metabolism pathways pyridine compound metabolic process (16 proteins, p.adj=2.6E-10) and detoxification (12 proteins, p.adj=9.0E-8), though innate immune response proteins were also enriched (25 proteins, p.adj=6.1E-6). A small proportion of women dominated by *Gardnerella* had increased expression of immune related pathways such as humoral immune response (19 proteins, p.adj=3.4E-20), and response to wounding (11 proteins, p.adj=5.0E-5). There was consistent downregulation of proteins seen in both clusters of *Gardnerella* dominant women, these proteins were associated with the epithelial barrier including cell-cell adhesion (46 proteins, p.adj=4.3E-15), tissue development (41 proteins, p.adj=1.2E-6), and keratinization (11 proteins, p.adj=5.2E-9), with proteins in translation initiation (11 proteins, p.adj=4.5E-4) and intracellular transport (31 proteins, p.adj=0.0025) decreased as well. IPA canonical pathways that were activated in *Gardnerella* dominant women could be divided into three main categories: Epithelial barrier associated pathways (integrin signaling (Z-score=2.71, 12 proteins, p=9.33E-9) and remodeling of epithelial adherens junctions (Z-score=0.82, 10 proteins, p=6.61E-8)), immune response pathways (acute phase response signaling (Z-score=1.73, 21 proteins,

p=3.16E-13), complement system (Z-score=1.34, 9 proteins, p=2.88E-9), leukocyte extravasation signaling (Z-score=1.13, 8 proteins, p=0.00977) and production of NO and ROS in macrophages (Z-score=0.38, 8 proteins, p=0.00759)) and metabolic pathways (glycogen degradation II (Z-score=2.00, 4 proteins, p=2.19E-5), pentose phosphate pathway (Z-score=2.00, 4 proteins, p=9.55E-6), Gluconeogenesis I (Z-score=1.63, 6 proteins, p=1.91E-6), and NRF2-mediated oxidative stress response (Z-score=1.34, 9 proteins, p=0.00224)) (Figure 17C). Only the translation canonical pathway EIF2 signaling (Z-score=-2.45, 11 proteins, p=0.00) was inhibited in *Gardnerella* women.

Activated functions in women with *Gardnerella* dominance identified by IPA showed increased neutrophil response (cell movement of neutrophils (Z-score=3.67, 25 proteins, p=1.34E-8), adhesion of neutrophils (Z-score=2.46, 12 proteins, p=7.95E-8), accumulation of neutrophils (Z-score=2.40, 10 proteins, p=9.72E-7), and formation of neutrophil extracellular traps (Z-score=1.74, 6 proteins, p=1.71E-7)), as well as general immune response (cell movement of phagocytes (Z-score=2.97, 40 proteins, p=3.03E-10), activation of cells (Z-score=2.67, 60 proteins, p=4.44E-12), inflammatory response (Z-score=2.51, 45 proteins, p=2.65E-8), accumulation of blood cells (Z-score=2.44, 22 proteins, p=9.57E-8), and degranulation of phagocytes (Z-score=2.09, 70 proteins, p=6.12E-40) (Figure 17D). The inhibited functions in samples with *Gardnerella* communities were quantity of ROS (Z-score=-2.36, 19 proteins, p=1.66E-8), necrosis (Z-score=-1.47, 143 proteins, p=1.82E-23), and cytotoxicity (Z-score=-0.69, 17 proteins, p=3.00E-6).

Only the down protein signature of monocytes vs neutrophils (NES=1.60, 45 proteins, p=0.006) matched the protein signature of women with *Gardnerella* communities (Figure 17E).

Table 7. Demographic and clinical characteristics of women with *L. crispatus* and *Gardnerella* communities.

Variable	>50% *L. crispatus* community women	>50% *Gardnerella* community women	P value
	n=77	n=172	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	24.8±5.7; 18-40	0.338
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.6±0.5; 0.2-2.4	0.172
HIV seroconverters (n, %)	4 (5.2)	9 (5.2)	1
HSV2▲			
Baseline seroprevalence (n, %)	32 (41.6)	93 (54.1)	0.0684
HSV2 incidence during trial (n, %)	4/45 (8.9)	16/76 (21.1)	0.137
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	85 (49.4)	0.966
Rural (n, %)	51 (66.2)	128 (74.4)	0.24
Given Birth (n, %)	60 (77.9)	141 (82.0)	0.565
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.2±2.0; 12-24	0.103
Total lifetime sex partners (Mean ± SD;	3.8±11.4; 1-100	2.9±7.2; 1-94	0.501
Condom use (n, %)			
Always	24 (31.2)	46 (26.7)	0.561
Most times	16 (20.8)	27 (15.7)	
Occasionally	25 (32.5)	66 (38.4)	
Never	12 (15.6)	33 (19.2)	
Demographics at visit of sample collection			
Contraception use (n, %) ¤			
DMPA	55 (71.4)	102 (59.3)	0.352
NET-EN	13 (16.9)	29 (16.9)	
COC	9 (11.7)	30 (17.4)	
Vaginal discharge (n, %) ★	7 (9.1)	9 (5.2)	0.399
Presence of STI infection (n, %) §	2/20 (10.0)	16/56 (28.6)	0.171
Any STI during trial (n, %) §	11/43 (25.6)	37/101 (36.6)	0.274
Any cervical ectopy (n, %) ¶	4 (5.2)	16 (9.3)	0.398
Antibiotic use in past 30 days (n, %)	6 (7.8)	10 (5.8)	0.758
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	5.3±4.8; 0-28	0.481
Age difference of oldest partner (Mean ± SD; range) †	3.7±3.1; (-3)-12	3.4±3.6; (-7)-19	0.343
Median adherence during trial >50% (n, %)	46 (59.7)	106 (61.6)	0.887
Vaginal cleaning reported at visit (n, %)	8 (10.4)	4 (2.3)	0.0153
Vaginal cleaning ever during trial (n, %)	19 (24.7)	28 (16.3)	0.165

▲ 3 women had undetermined HSV2 status

¤ 11 women were using another form of birth control

★ 2 women had no information on vaginal discharge at visit

§ 173 women were not tested for STI's at the visit, 105 women were not tested during the trial

¶ 6 women had no information at the visit of cervical ectopy

† 44 women did not report sexual act in the previous 30 days so had no partner age reported

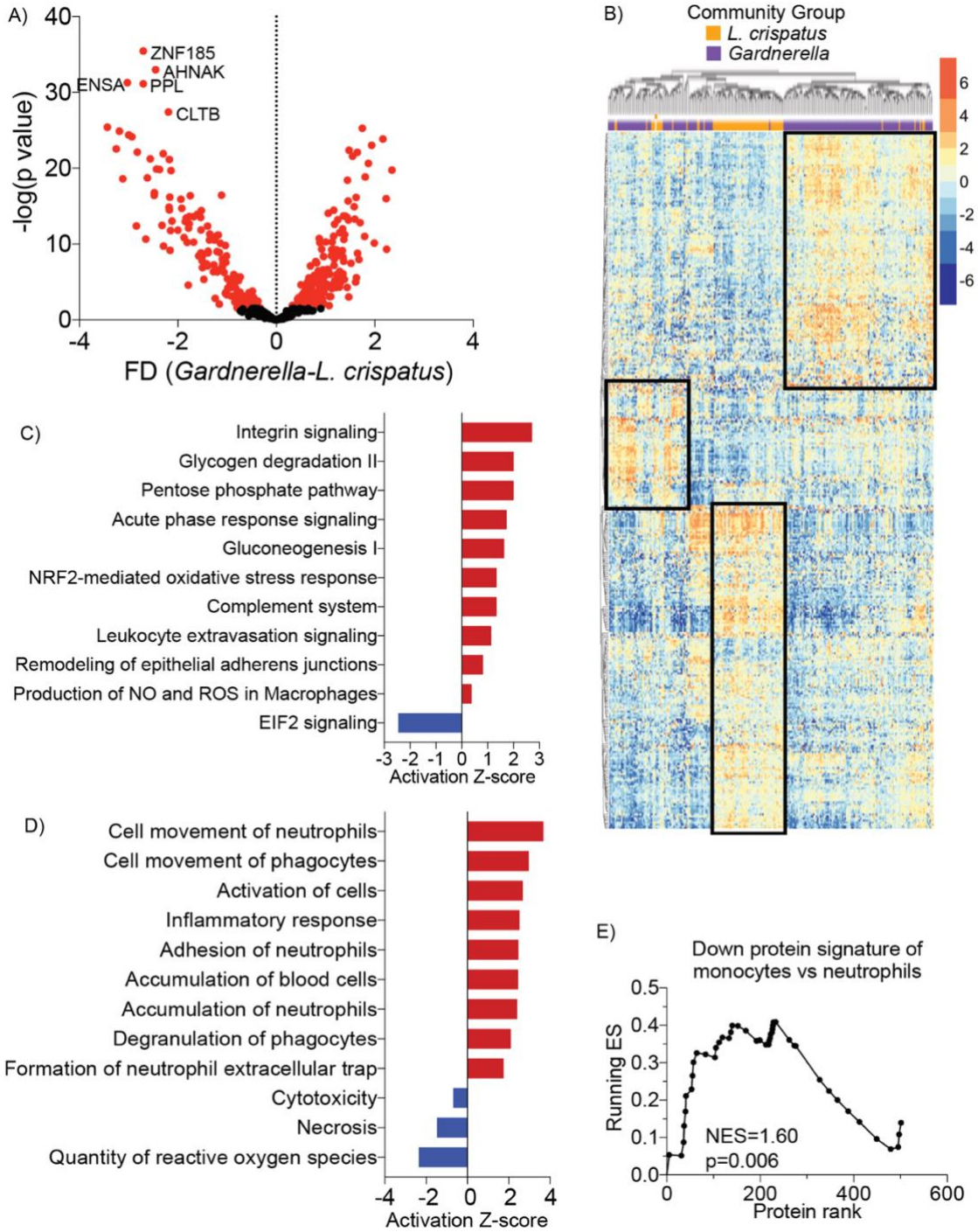

Figure 17. Overview of host differences between women with *Gardnerella* and *L. crispatus* communities. A) Volcano plot of the proteins significantly different (BH adjusted p<0.05) in red, with the top five indicated. **B)** Heatmap of the significant proteins different between *Gardnerella* and *L. crispatus* women, with three major protein clusters boxed. **C)** IPA canonical pathways activated (red) or inhibited (blue) in *Gardnerella* women. **D)** IPA functions activated (red) or inhibited (blue) in *Gardnerella* women. **E)** Overlap of *Gardnerella* host proteome signature with the down protein signature in blood monocytes vs neutrophils.

Prevotella has often been associated with BV and has been associated with increased risk of HIV infection (*103*). There were 335 host proteins (58.2%) that were significantly different (BH p.adj<0.05) between women with *Prevotella* and *L. crispatus* communities, as shown in the volcano plot in figure 18A. Women who were codominated by *Prevotella* were more likely to be younger (22.4±3.7 vs 25.4±5.8 years, p=0.00931) and in the trial for less time (1.4±0.5 vs 1.6±0.5 years, p=0.0112), but otherwise there were no significant differences in the clinical or demographic characteristics (Table 8). A heatmap of the proteins that were significant showed 2 clusters (Figure 18B). The proteins in the cluster that was upregulated in women with *Prevotella* had pathways enriched in proteolysis (52 proteins p.adj=3.2E-11) and pyridine containing compound metabolic process (16 proteins, p.adj=1.4E-9) and innate immune response (30 proteins, p.adj=2.2E-7). Downregulated proteins were in epithelial barrier pathways such as cell-cell adhesion (51 proteins, p.adj=5.1E-18), tissue development (46 proteins, p.adj=3.7E-8), and keratinization (10 proteins, p.adj=1.3E-7), and metabolism pathways like translation initiation (10 proteins, p.adj=0.0043) and protein metabolic process (75 proteins, p.adj=0.0049). The majority of IPA canonical pathways that were activated with a Z-score<1.00 were involved in pathways such as epithelial barrier associated pathways (integrin signaling (Z-score=3.21 14 proteins, p=2.57E-6) and remodeling of epithelial adherens junctions (Z-score=1.90, 15 proteins, p=3.16E-14)), inflammatory pathways (LXR/RXR activation (Z-score=2.24, 20 proteins, p=6.31E-16), Fcγ receptor-mediated phagocytosis in macrophages and monocytes (Z-score=1.89, 7 proteins, p=4.07E-4), production of NO and ROS in macrophages (Z-score=1.13, 7 proteins, p=0.0186)), metabolic pathways (glycogen degradation II (Z-score=2.00, 4 proteins, p=1.86E-5), pentose phosphate pathway (Z-score=2.00, 4 proteins, p=8.13E-6), gluconeogenesis I (Z-score=1.89, 7 proteins, p=5.89E-8), and NRF2-mediated oxidative stress response (Z-

score=1.63, 8 proteins, p=0.00603)) and a signal transduction pathway, Rac signaling (Z-score=2.45, 6 proteins, p=0.00550) (Figure 18C). The two significant inhibited canonical pathways were EIF2 signaling (Z-score=-2.24, 10 proteins, p=0.00151) and RhoGDI (Rho GDP dissociation inhibitor) signaling (Z-score=-1.41, 8 proteins, p=0.00447). Activated functions identified by IPA in women codominated by *Prevotella* included activation of the immune response (activation of leukocytes (Z-score=3.33, 39 proteins, p=2.08E-8), activation of phagocytes (Z-score=2.68, 30 proteins, p=9.06E-11), migration of phagocytes (Z-score=2.41, 25 proteins, p=4.58E-10), degranulation of phagocytes (Z-score=2.32, 67 proteins, p=2.78E-38), accumulation of leukocytes ((Z-score=2.15, 21 proteins, p=5.41E-8), and inflammatory response (Z-score=2.04, 41 proteins, p=4.07E-7)), particularly neutrophils (accumulation of neutrophils (Z-score=2.77, 12 proteins, p=6.01E-9), adhesion of neutrophils (Z-score=2.65, 12 proteins, p=4.68E-8), cell movement of neutrophils (Z-score=2.63, 27 proteins, p=2.39E-10), binding of neutrophils (Z-score=2.34, 13 proteins, p=3.97E-8), and activation of neutrophils (Z-score=2.26, 12 proteins, p=3.17E-8) (Figure 18D). Inhibited functions in samples with *Prevotella* communities were involved in response to radicals (quantity of ROS (Z-score=-2.83, 18 proteins, p=4.57E-8) and metabolism of hydrogen peroxide (Z-score=-1.50, 13 proteins, p=1.89E-7)), phagocytosis of cells (Z-score=-0.59, 22 proteins, p=1.81E-7), aggregation of blood cells (Z-score=-0.62, 21 proteins, p=2.45E-9), apoptosis (Z-score=-0.69, 133 proteins, p=1.15E-20), and microcytosis (Z-score=-1.95, 8 proteins, p=4.33E-7). The one signature which matched the protein signature of women with *Prevotella* communities was the down protein signature of monocytes vs neutrophils (NES=1.83, 45 proteins, p<0.0001, Figure 18E) while the up signature of CD8 T cells vs natural killer cells was trending (NES=1.45, 22 proteins, p=0.055, Figure 18F).

Table 8. Demographic and clinical characteristics of women with *L. crispatus* and *Prevotella*.

Variable	>50% *L. crispatus* community women	>25% *Prevotella* community women	P value
	n=77	n=29	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	22.4±3.7; 18-33	0.00931
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.4±0.5; 0.3-2.2	0.0112
HIV seroconverters (n, %)	4 (5.2)	5 (17.2)	0.111
HSV2▲			
Baseline seroprevalence (n, %)	32 (41.6)	20 (69.0)	0.0129
HSV2 incidence during trial (n, %)	4/45 (8.9)	1/8 (12.5)	0.738
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	17 (58.6)	0.607
Rural (n, %)	51 (66.2)	25 (86.2)	0.0729
Given Birth (n, %)	60 (77.9)	23 (79.3)	1
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.3±1.8; 14-21	0.461
Total lifetime sex partners (Mean ± SD;	3.8±11.4; 1-100	2.3±1.3; 1-5	0.84
Condom use (n, %)			
Always	24 (31.2)	5 (17.2)	0.429
Most times	16 (20.8)	9 (31.0)	
Occasionally	25 (32.5)	9 (31.0)	
Never	12 (15.6)	6 (20.7)	
Demographics at visit of sample collection			
Contraception use (n, %)			
DMPA	55 (71.4)	21 (72.4)	0.903
NET-EN	13 (16.9)	4 (13.8)	
COC	9 (11.7)	4 (13.8)	
Vaginal discharge (n, %)	7 (9.1)	3 (10.3)	1.000
Presence of STI infection (n, %) §	2/20 (10.0)	6/10 (60.0)	0.146
Any STI during trial (n, %) §	11/43 (25.6)	7/16 (43.8)	0.303
Any cervical ectopy (n, %) ¶	4 (5.2)	4 (13.8)	0.273
Antibiotic use in past 30 days (n, %)	6 (7.8)	1 (3.4)	0.716
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	5.3±3.5; 0-16	0.963
Age difference of oldest partner (Mean ± SD; range) †	3.7±3.1; (-3)-12	3.0±3.1; (-4)-10	0.354
Median adherence during trial >50% (n, %)	46 (59.7)	16 (55.2)	0.838
Vaginal cleaning reported at visit (n, %)	8 (10.4)	0 (0.0)	0.164
Vaginal cleaning ever during trial (n, %)	19 (24.7)	3 (10.3)	0.176

▲ 1 woman had an undetermined HSV2 status

§ 76 women were not tested for STI's at the visit, 47 women were not tested during the trial

¶ 3 women had no information at the visit of cervical ectopy

† 15 women did not report sexual act in the previous 30 days so had no partner age reported

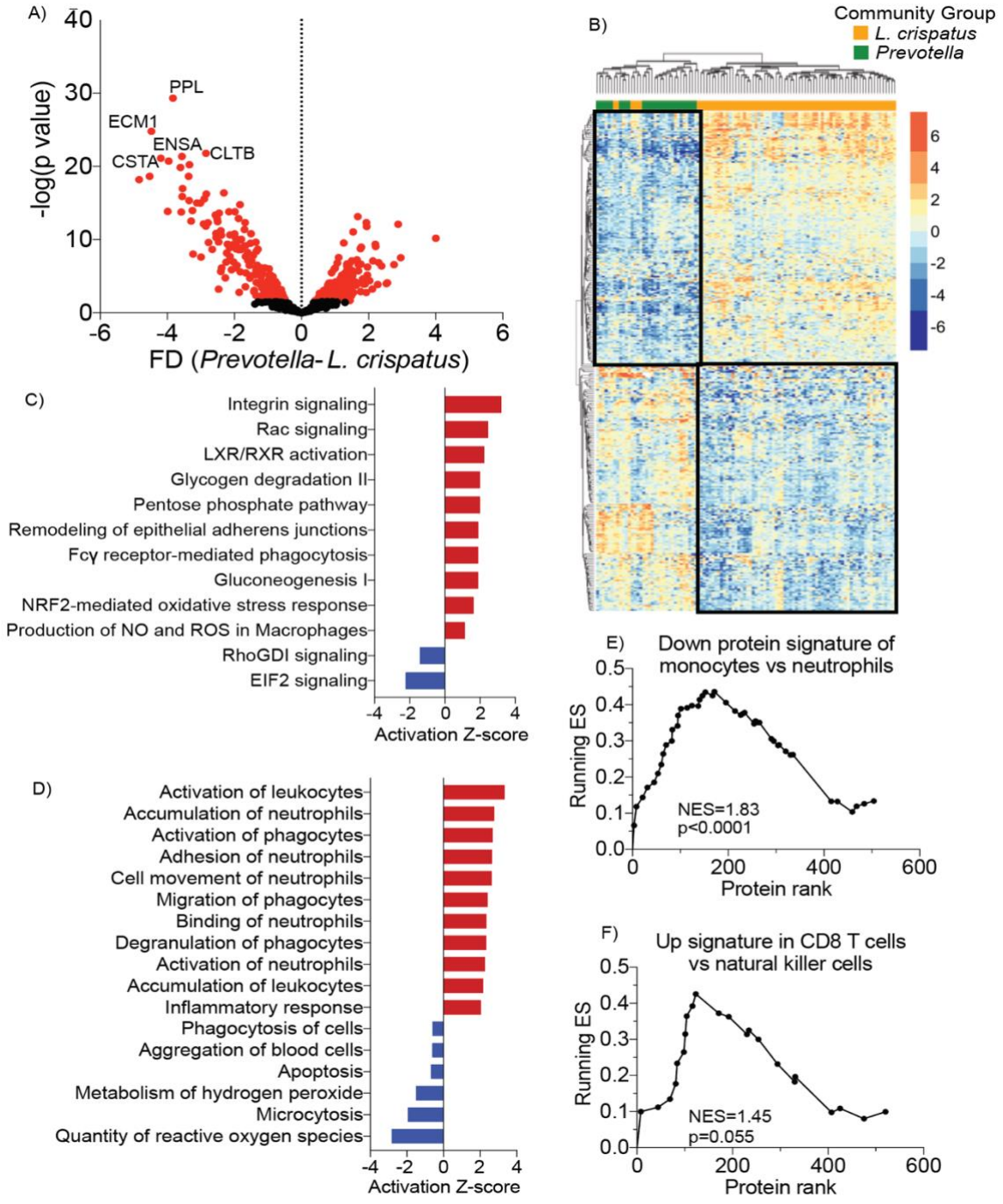

Figure 18. Overview of host differences between women with >25% *Prevotella* and *L. crispatus* communities. A) Volcano plot showing the proteins significantly different (BH corrected p<0.05) in red, with the top five indicated. **B)** Heatmap of the significant proteins different between *Prevotella* and *L. crispatus* women. **C)** Significant IPA canonical pathways activated (red) or inhibited (blue) in >25% Prevotella women. **D)** Significant IPA functions that were activated (red) or inhibited (blue). Overlapping protein signature of women with *Prevotella* with **E)** the down signature in blood monocytes vs neutrophils and **F) the** up signature in blood CD8 T cells vs NK cells.

In this study 26 women had *Mobiluncus* as the dominant or codominant (>25%) organism in the vagina, which has previously been associated with increased mucosal cytokine production (*63*) and pre-term birth (*281*). Baseline clinical, behavioral, and sexually transmitted infection information was not different between these two groups, except women with codominant *Mobiluncus* were slightly younger (23.0 vs 25.4, p=0.0419) (Table 9). A total of 264 host proteins (45.8%) were differentially abundant (BH corrected p<0.05), with 115 overabundant and 149 under-abundant protein in women with >25% *Mobiluncus* communities (Figure 19A). The cluster that contained proteins expressed at low levels in women with *Mobiluncus* codominant communities consisted of proteins involved in epithelial associated pathways such as cell-cell adhesion (40 proteins, p=8.8E-12), keratinocyte differentiation (10, p.adj=4.9E-8), and membrane organization (28 proteins, p.adj=6.2E-6). Cluster two contained proteins upregulated in women with *Mobiluncus* codominant communities which were broadly involved in immune response (defense response 29, p=5.9E-7), innate immune response (23, p=3.2E-8)), and increased cellular catabolism (30, p=6.2E-7) (Figure 19B). The activated IPA pathways were involved in epithelial barrier associated pathways (integrin signaling (Z-score=2.83 8 proteins, p=0.00309) and remodeling of epithelial adherens junctions (Z-score=1.63, 10 proteins, p=4.90E-9), and actin cytoskeleton signaling (Z-score=0.30, 11 proteins, p=4.17E-5)), inflammatory pathways (acute phase response signaling (Z-score=1.41, 13 proteins, p=1.45E-7), blood coagulation (Z-score=0.45, 5 proteins, p=4.47E-5), leukocyte extravasation signaling (Z-score=0.38, 7 proteins, p=0.00741), and LXR/RXR activation (Z-score=0.28, 13 proteins, p=1.26E-9)), metabolic pathways (pentose phosphate pathway (Z-score=2.00, 4 proteins, p=3.24E-6), gluconeogenesis I (Z-score=1.00, 4 proteins, p=2.00E-4), and NRF2-mediated oxidative stress response (Z-score=0.82, 9 proteins, p=3.24E-4)), and a signal transduction

pathway signaling by Rho family GTPases (Z-score=0.71, 10 proteins, p=5.01E-4) (Figure 19C). The only significant inhibited canonical pathway was EIF2 signaling (Z-score=-2.45, 9 proteins, p=0.00110). Activated functions identified by IPA in women codominated by *Mobiluncus* included activation of the immune response (accumulation of granulocytes (Z-score=1.97, 8 proteins, p=1.22E-4), activation of myeloid cells (Z-score=1.92, 18 proteins, p=8.36E-6), activation of phagocytes (Z-score=1.63, 19 proteins, p=7.70E-6), accumulation of myeloid cells (Z-score=1.61, 11 proteins, p=7.39E-5), activation of granulocytes (Z-score=1.61, 9 proteins, p=2.31E-5), migration of phagocytes (Z-score=1.50, 14 proteins, p=1.56E-4), and inflammatory response (Z-score=1.21, 30 proteins, p=6.62E-5)), with neutrophil associated pathways prominent (cell movement of neutrophils (Z-score=2.13, 16 proteins, p=4.59E-5), activation of neutrophils (Z-score=2.04, 8 proteins, p=2.48E-5), and accumulation of neutrophils (Z-score=1.70, 7 proteins, p=7.48E-5)), binding of neutrophils (Z-score=2.34, 13 proteins, p=3.97E-8), and activation of neutrophils (Z-score=2.26, 12 proteins, p=3.17E-8), as well as migration of keratinocytes (Z-score=2.20, 7 proteins, p=1.13E-4), and cell cycle progression (Z-score=1.80, 36 proteins, p=8.59E-5) (Figure 19D). Inhibited functions in samples with *Mobiluncus* communities were quantity of ROS (Z-score=-2.04, 14 proteins, p=1.88E-6), and indications of tissue damage (wound (Z-score=-1.98, 15 proteins, p=1.85E-6)), aggregation of blood cells (Z-score=-1.60, 13 proteins, p=4.07E-5), and growth of connective tissue (Z-score=-0.99, 29 proteins, p=2.86E-6)). Both significant protein signatures of *Mobiluncus* codominant women that matched identified significant overlap with neutrophils, one compared to monocytes (NES=1.73, 45 proteins, p=0.00424, Figure 19E), the other compared to B cells (NES=1.29, 68 proteins, p=0.0042, Figure 19E).

Table 9. Demographic and clinical characteristics of women with *L. crispatus* and *Mobiluncus*.

Variable	>50% *L. crispatus* women	>25% *Mobiluncus* women	P value
	n=77	n=26	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	23.0±4.5; 18-34	0.0435
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.6±0.5; 0.4-2.3	0.563
HIV seroconverters (n, %)	4 (5.2)	3 (11.5)	0.509
HSV2			
Baseline seroprevalence (n, %)	32 (41.6)	13 (50.0)	0.602
HSV2 incidence during trial (n, %)	4/45 (8.9)	3/13 (23.1)	0.368
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	12 (46.2)	0.865
Rural (n, %)	51 (66.2)	17 (65.4)	1
Given Birth (n, %)	60 (77.9)	21 (80.8)	0.976
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	17.8±1.8; 15-22	0.8509
Total lifetime sex partners (Mean ± SD; range)	3.8±11.4; 1-100	2.2±1.4; 1-6	0.485
Condom use (n, %)			
Always	24 (31.2)	8 (30.8)	0.996
Most times	16 (20.8)	6 (23.0)	
Occasionally	25 (32.5)	8 (30.8)	
Never	12 (15.6)	4 (15.4)	
Demographics at visit of sample collection			
Contraception use (n, %) ¤			
DMPA	55 (71.4)	13 (50.0)	0.179
NET-EN	13 (16.9)	8 (30.8)	
COC	9 (11.7)	4 (15.4)	
Vaginal discharge (n, %) ★	7 (9.1)	2 (7.7)	1.000
Presence of STI infection (n, %) §	2/20 (10.0)	0/5 (0.0)	1.000
Any STI during trial (n, %) §	11/43 (25.6)	6/11 (54.5)	0.138
Any cervical ectopy (n, %) ¶	4 (5.2)	1 (3.8)	1.000
Antibiotic use in past 30 days (n, %)	6 (7.8)	2 (7.7)	1.000
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	4.5±3.5; 0-10	0.403
Age difference of oldest partner (Mean ± SD; range) †	3.7±3.1; (-3)-12	3.6±3.4; 0-14	0.547
Median adherence during trial >50% (n, %)	46 (59.7)	16 (61.5)	1
Vaginal cleaning reported at visit (n, %)	8 (10.4)	0 (0.0)	0.198
Vaginal cleaning ever during trial (n, %)	19 (24.7)	4 (15.4)	0.477

¤ 1 woman was using another form of birth control

★ 1 women had no information on vaginal discharge at visit

§ 76 women were not tested for STI's at the visit, 47 women were not tested during the trial

¶ 4 women had no information at the visit of cervical ectopy

† 15 women did not report sexual act in the previous 30 days so had no partner age

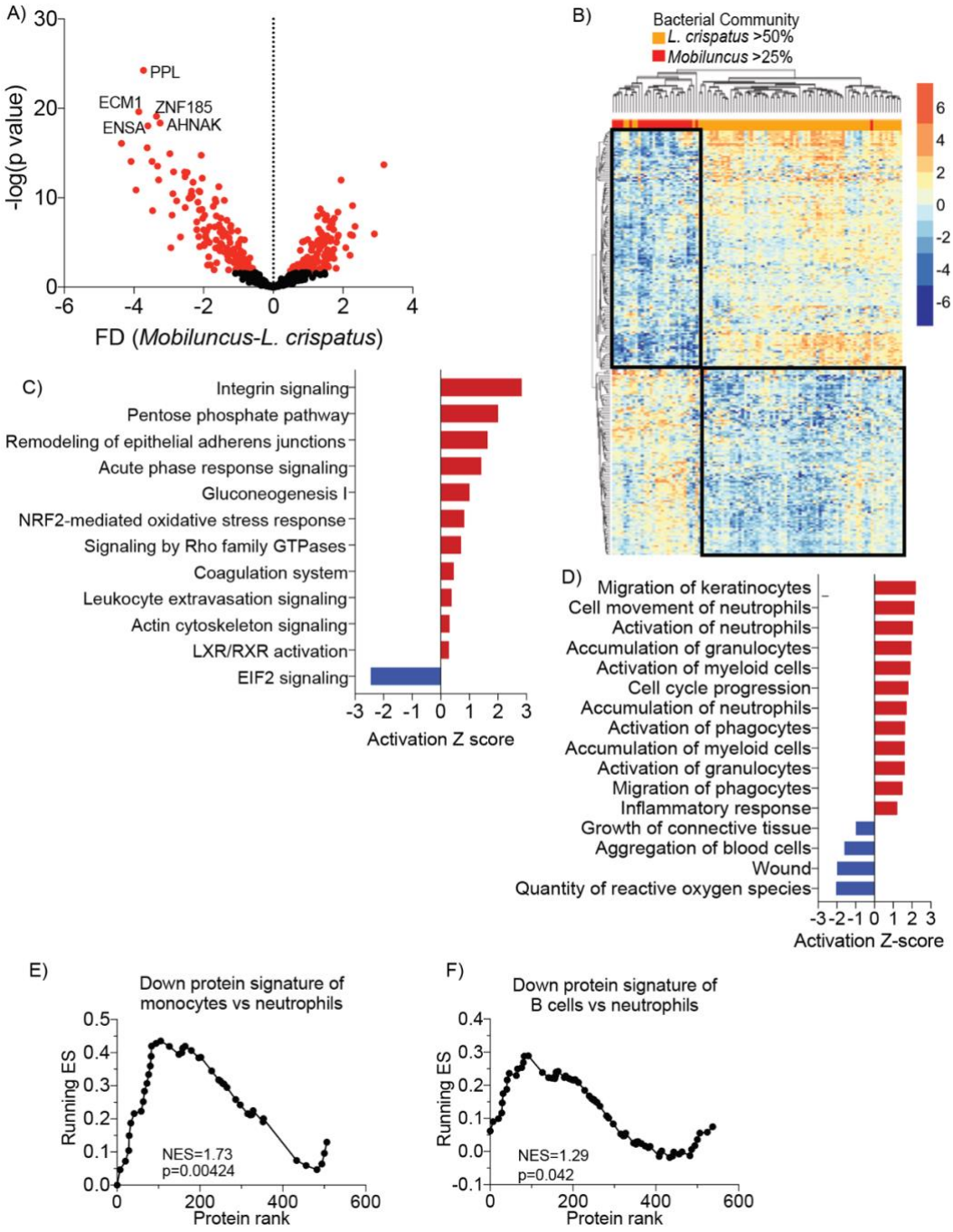

Figure 19. Overview of host differences between women with >25% *Mobiluncus* and *L. crispatus* communities. A) Volcano plot showing the proteins significantly different (BH corrected p<0.05) in red, with the top five indicated. **B)** Heatmap of the significant proteins different between *Mobiluncus* and *L. crispatus* women. **C)** IPA canonical pathways activated (red) or inhibited (blue) in women codominated by *Mobiluncus*. **D)** Significant IPA functions activated (red) or inhibited (blue) in women with *Mobiluncus*. Overlapping protein signature of women with *Mobiluncus* with **E)** the down signature in blood monocytes vs neutrophils and **F)** the down signature in blood B cells vs neutrophils.

Pseudomonas is not a commonly identified vaginal bacteria, potentially due to exclusion by individuals who believe it is an environmental contaminant, however studies have identified it in the vagina (*282, 283*). There were 219 host proteins (38.0%) that were significantly different after BH correction between women with *Pseudomonas* and *L. crispatus* communities, as shown in the volcano plot in figure 20A. *Pseudomonas* community women had a higher baseline prevalence of HSV2 (73.3% vs 41.6%, p=0.0238) but otherwise there were no differences in clinical or demographic variables (Table 10). A heatmap of the proteins that were significant showed 2 clusters (Figure 20B). The smaller cluster of 41 upregulated proteins in women with *Pseudomonas* community had pathways enriched in proteolysis (22 proteins, p.adj=1.4E-8), and blood coagulation (10 proteins, p.adj=6.2E-6). The larger cluster (178 proteins) of downregulated proteins were in pathways involved in the epithelial barrier including cell-cell adhesion (49 proteins, p.adj=1.5E-15), keratinization (11 proteins, p.adj=9.5E-9), and tissue development (46 proteins, p.adj=1.3E-7) as well as immune associated pathways such as humoral immune response (12 proteins, p.adj=0.0026) and protease inhibition (13 proteins, p.adj=5.4E-5). Contrary to other bacterial communities, the majority of IPA canonical pathways were inhibited in *Pseudomonas* community women, with only the LXR/RXR activation pathway (Z-score=1.41, 8 proteins, p=1.91E-5) activated (Figure 20C). Two pathways showed no activation or inhibition: leukocyte extravasation (Z-score=0, 5 proteins, p=0.0380) and ILK signaling (Z-score=0, 8 proteins, p=4.37E-4). Inhibited canonical pathways included blood associated pathways (coagulation system (Z-score=-0.45, 5 proteins, p=1.78E-5), and intrinsic prothrombin activation (Z-score=-0.82, 6 proteins, p=2.51E-6)), cytoskeleton associated pathways (actin cytoskeleton signaling (Z-score=-0.38, 7 proteins, p=0.00457), and RhoA signaling (Z-score=-1.00, 4 proteins, p=0.0288)), protein assembly and folding pathways (EIF2

signaling (Z-score=-2.00, 6 proteins, p=0.0191), and BAG2 signaling (Z-score=-1.00, 4 proteins, p=6.92E-4)), as well as the inflammatory associated pathway acute phase response signaling (Z-score=-0.33, 15 proteins, p=1.55E-10). Activated functions identified by IPA included immune associated pathways (recruitment of myeloid cells (Z-score=2.10, 12 proteins, p=2.11E-4), accumulation of granulocytes (Z-score=1.85, 10 proteins, p=4.22E-7), recruitment of leukocytes (Z-score=1.74, 14 proteins, p=9.06E-5), accumulation of myeloid cells (Z-score=1.62, 15 proteins, p=5.61E-9), and activation of leukocytes (Z-score=1.58, 25 proteins, p=1.01E-5)) with an emphasis in neutrophil pathways (cell movement of neutrophils (Z-score=2.04, 20 proteins, p=6.67E-9), infiltration by neutrophils (Z-score=1.60, 13 proteins, p=3.31E-7), and accumulation of neutrophils (Z-score=1.59, 9 proteins, p=1.72E-7)) as well as signs of tissue damage (accumulation of blood cells (Z-score=1.62, 18 proteins, p=2.65E-8), degradation of DNA (Z-score=2.40, 10 proteins, p=7.09E-5), and damage of epithelial tissue (Z-score=2.03, 9 proteins, p=7.99E-6) (Figure 20D). Inhibited functions in samples with *Pseudomonas* communities were cell viability (Z-score=-2.02, 42 proteins, p=9.47E-6), endocytosis (Z-score=-1.97, 25 proteins, p=1.80E-7), and growth of connective tissue (Z-score=-1.78, 23 proteins, p=5.57E-5). There was only one protein signature that matched the protein signature of women with *Pseudomonas* communities: the down protein signature of B cells compared to neutrophils (NES=1.21, 68 proteins, p=0.0588, Figure 20E).

Table 10. Demographic and clinical characteristics of women with *L. crispatus* and *Pseudomonas* communities.

Variable	>50% *L. crispatus* women	>40% *Prevotella* women	P value
	n=77	n=15	
Age at visit (Mean ± SD; range)	25.4±5.8; 18-40	22.7±4.7; 18-35	0.0634
Years in trial (Mean ± SD; range)	1.6±0.5; 0.5-2.4	1.4±0.6; 0.4-2.1	0.444
HIV seroconverters (n, %)	4 (5.2)	3 (20.0)	0.148
HSV2 ▲			
Baseline seroprevalence (n, %)	32 (41.6)	11 (73.3)	0.0238
HSV2 incidence during trial (n, %)	4/45 (8.9)	0/3 (0.0)	0.59
Baseline Demographics			
TFV arm (n, %)	39 (50.6)	7 (46.7)	1
Rural (n, %)	51 (66.2)	13 (86.7)	0.205
Given Birth (n, %)	60 (77.9)	13 (86.7)	0.677
Age at sexual debut (Mean ± SD; range)	17.7±2.6; 12-25	16.5±1.6; 14-18	0.0885
Total lifetime sex partners (Mean ± SD; range)	3.8±11.4; 1-100	2.3±1.4; 1-5	0.844
Condom use (n, %)			
Always	24 (31.2)	3 (20.0)	0.106
Most times	16 (20.8)	4 (26.7)	
Occasionally	25 (32.5)	2 (13.3)	
Never	12 (15.6)	6 (40.0)	
Demographics at visit of sample collection			
Contraception use (n, %) ¤			
DMPA	55 (71.4)	13 (86.7)	0.191
NET-EN	13 (16.9)	0 (0.0)	
COC	9 (11.7)	1 (6.7)	
Vaginal discharge (n, %) ★	7 (9.1)	1 (6.7)	1.000
Presence of STI infection (n, %) §	2/20 (10.0)	0/3 (0.0)	1.000
Any STI during trial (n, %) §	11/43 (25.6)	2/8 (25.0)	1.000
Any cervical ectopy (n, %) ¶	4 (5.2)	1 (6.7)	1.000
Antibiotic use in past 30 days (n, %)	6 (7.8)	1 (6.7)	1.000
Sex acts in past 30 days (Mean ± SD; range)	6.2±6.3; 0-39	4.1±4.9; 0-20	0.193
Age difference of oldest partner (Mean ± SD; range) †	3.7±3.1; (-3)-12	2.2±1.5; 0-5	0.0792
Median adherence during trial >50% (n, %)	46 (59.7)	10 (66.7)	0.831
Vaginal cleaning reported at visit (n, %)	8 (10.4)	0 (0.0)	0.42
Vaginal cleaning ever during trial (n, %)	19 (24.7)	1 (6.7)	0.228

▲ 1 woman had an undetermined HSV2 status

¤ 1 woman was using another form of birth control

§ 69 women were not tested for STI's at the visit, 41 women were not tested during the trial

¶ 2 women had no information at the visit of cervical ectopy

† 16 women did not report sexual act in the previous 30 days so had no partner age reported

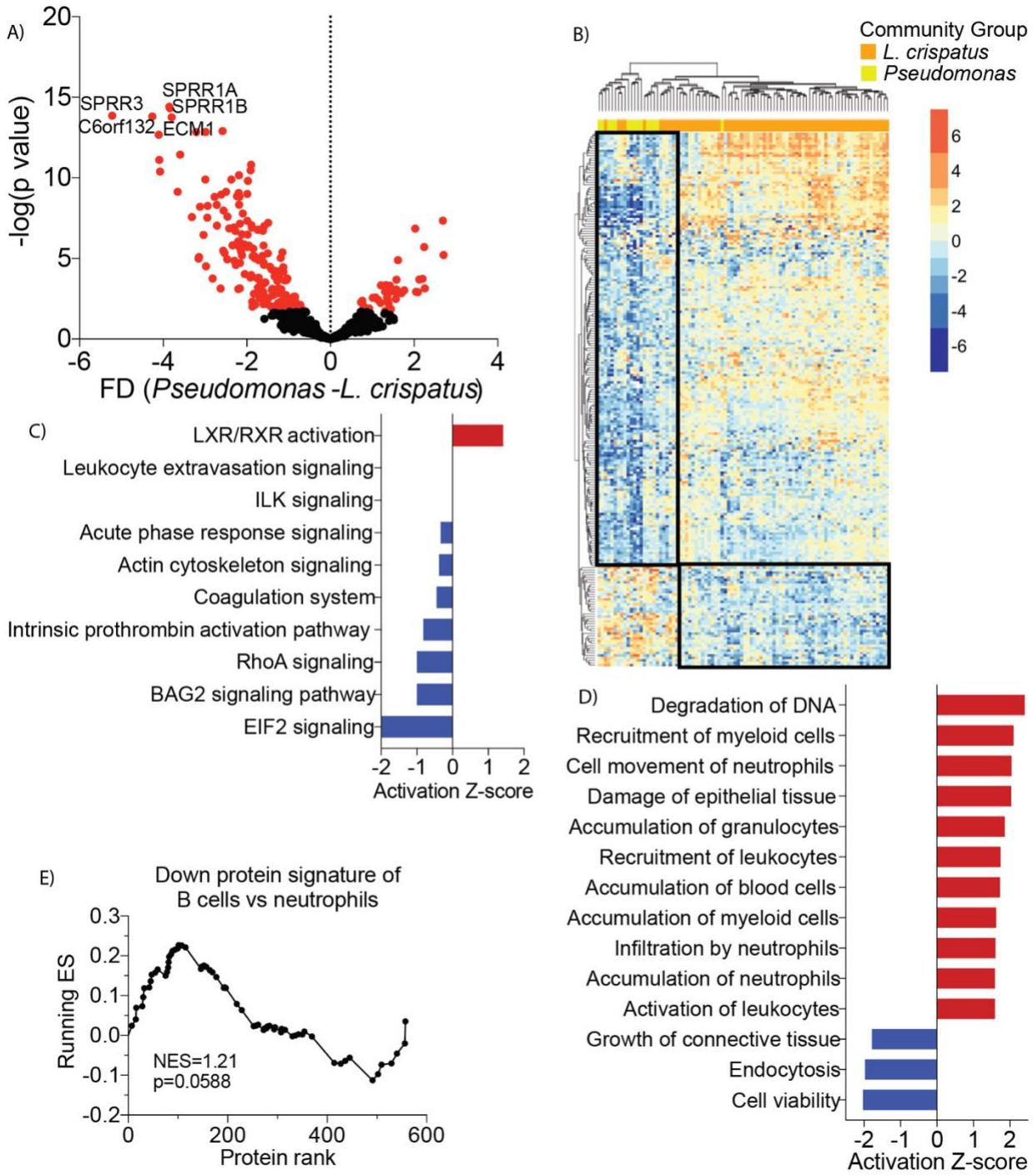

Figure 20. Overview of host differences between women with *Pseudomonas* and *L. crispatus* communities. A) Volcano plot showing the proteins significantly different (BH adjusted p<0.05) in red, with the top five indicated. **B)** Clustered heatmap of the significant proteins different between *Pseudomonas* and *L. crispatus* women. **C)** Significant IPA canonical pathways activated (red) or inhibited (blue) in *Pseudomonas* women. **D)** Significant IPA functions that were activated (red) or inhibited (blue) in *Pseudomonas* women. **E)** Overlapping protein signature with the down protein signature in blood B cells vs neutrophils.

4.4 Overlapping Dysregulated Proteins and Pathways

We wanted to assess if there were proteins that were dysregulated in multiple bacterial communities compared to the expression in *L. crispatus* communities. Of the 576 proteins in 100% of MS batches, 32 (5.6%) proteins were dysregulated in 6 of the 7 non-*L. crispatus* groups, with many of the proteins associated with epithelial integrity, such as keratins (13), periplakin, small proline rich protein 3, and serine protease inhibitor kazal type 5. Interestingly 123 (21.4%) proteins were not dysregulated in any bacterial community. These proteins were involved in vesicle mediated transport (n=31 proteins), and the inflammatory response (n=18 proteins). The inflammatory proteins that were not altered were involved in platelet degranulation (APOA1, ORM1, ORM2, SERPINA1, SERPINA3, SERPING1), and there were multiple anti-proteases (AGT, SERPINA1, SERPINA3, SERPING1). Some proteins were not differentially expressed in the different bacterial communities, but at the pathway level similarities could be observed.

The similarly dysregulated pathways largely belonged to pathways relating to the epithelial barrier and inflammation and immune cell associated pathways. Figure 21A shows the dysregulated pathways involving bleeding associated pathways (bleeding, aggregation of blood platelets, and coagulation of blood), connective tissue pathways (proliferation and death of connective tissue cells) and fibroblast pathways (fibrinogenesis and proliferation of fibroblasts), and epithelial pathways (proliferation, adhesion, growth, migration, and death). Particularly interesting was that *L. iners* had the highest activation of the coagulation of blood pathway, and that damage to epithelial tissue was observed mostly in *Pseudomonas* and *Prevotella* communities, while *Mobiluncus*, *Gardnerella* & *Lactobacillus*, and *Gardnerella* communities activated death and necrosis of epithelial cells. The *L. iners* community did not show decreased epithelial growth or increased cell death of epithelial associated pathways, with mostly blood

associated increased signatures. Signatures that may indicate a wound healing response, such as decreased formation of skin and increased differentiation or movement of epithelial cells and keratinocytes were observed in all non-*Lactobacillus* dominant communities. Overall, there was a consistent inhibition of epithelial, and underlying tissue, pathways that indicate a compromised or weakened epithelial barrier in non-*Lactobacillus* communities.

Inflammation related dysregulated pathways that focused on immune cells belonged to neutrophils, phagocytes, leukocytes, and myeloid cells (Figure 21B). Activation of the inflammatory response was observed in *L. iners*, *Prevotella*, and *Gardnerella* communities, while there was an inhibition of immune response of cells signature in *Mobiluncus* and *Pseudomonas* communities. Activation in the pathway infection of cells was seen in all but the *Pseudomonas* community, while an antimicrobial response signature was only observed in *Gardnerella* & *Lactobacillus*, *Prevotell*a, and *Gardnerella* communities. Activation of movement and accumulation of neutrophils signatures were consistently observed in non-*L. crispatus* communities, though the *Gardnerella* & *Lactobacillus* community did not show increased movement of neutrophils, and the *L. iners* community only showed minor accumulation of neutrophils. The activation of neutrophils pathway was increased in *Mobiluncus*, *Prevotell*a, and *Gardnerella* communities, with a slight increase seen in *L. iners* community. Pathways associated with neutrophil function, such as degranulation of neutrophils were activated in *Prevotella* and *Gardnerella* communities, while there was only slight activation in reactive oxygen species synthesis pathway in *Mobiluncus*, *Prevotella*, and *Gardnerella* communities. Though phagocyte activation, movement and accumulation pathways were observed in all but the *Gardnerella* & *Lactobacillus* community, it was most activated in *Prevotella* and *Gardnerella* communities, though inhibition of phagocytosis signatures,

particularly in *Mobiluncus* and *Pseudomonas* communities was observed. Interestingly only two communities, *Gardnerella* & *Lactobacillus* and *Mobiluncus*, showed minimal leukocyte response signatures (only moderate leukocyte migration) with no increased recruitment, activation or accumulation, unlike the other communities. Movement of monocytes was decreased slightly in *Mobiluncus* communities, and no recruitment of myeloid cells was observed, though there was increased signatures of accumulation and activation of myeloid cells. The strongest immune activation signatures were in *Prevotella* and *Gardnerella* communities, *Mobiluncus* and *Pseudomonas* communities had strong response signatures but not by all immune cell subsets.

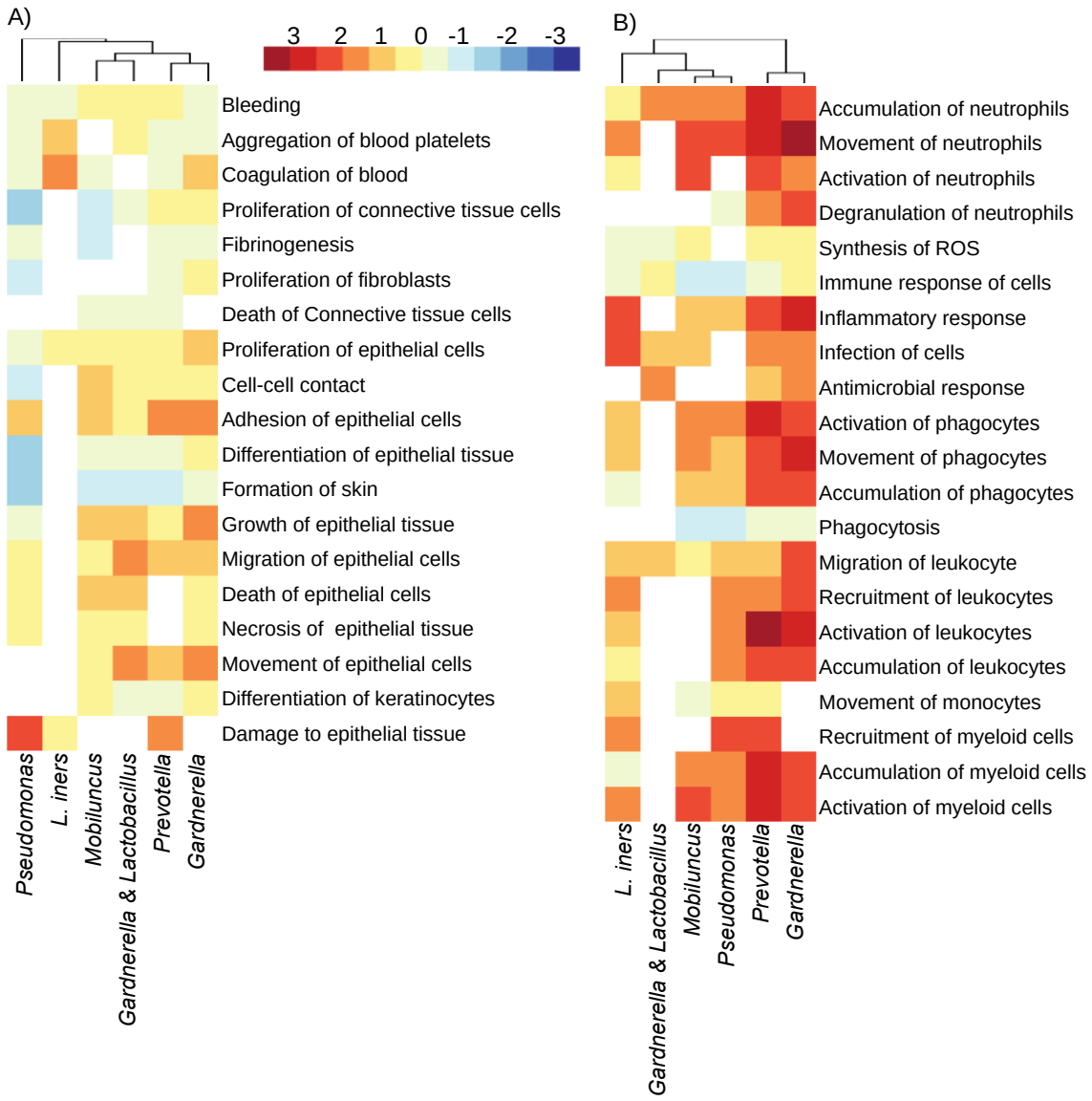

Figure 21. Commonly dysregulated pathways in non-*L. crispatus* communities. Pathways (from IPA) included were activated or inhibited in ≥3 communities using proteins significantly enriched (Benjamini-Hochberg corrected, $p<0.05$), with >50% *L. spp* community removed (only 1 protein passed multiple comparison correction so no list could be entered to IPA for pathway analysis). **A)** Z-score activation or inhibition of blood and epithelial associated pathways dysregulated in non-*L. crispatus* communities. **B)** Z-score activation or inhibition of inflammation and immune cell associated pathways dysregulated in non-*L. crispatus* communities.

4.5 Variation in Bacterial Function Among Major Communities

Though the composition of the vaginal microbiome is important, what the bacterial community is doing is also an important facet to consider. Changes in protein production can be initiated by bacteria in a myriad of circumstances such as the presence of a new carbon source, potential threat, or crowding of bacteria. Thus, we sought to identify if there were functional differences observed between the bacterial communities, as compared to the *L. crispatus* dominant community.

There were no significant bacterial functional differences between communities with *L. crispatus* and *L. spp* at the B level with no function passing any of the 5 tests. When analyzed at the KO level, the most significant hits were fatty acid degradation, (passed two tests, largest p value=0.639), chloroalkane and chloroalkene degradation (passed two tests, largest p value=0.639), naphthalene degradation (passed two tests, largest p value=0.639), and degradation of aromatic compounds (passed two tests, largest p value=0.639). Overall, these results indicate *L. crispatus* and non-*L. iners Lactobacillus* bacterial communities have similar functions at the community level.

There were 5 significant bacterial functional differences ($p<0.05$ in all five tests) between communities with *L. crispatus* and *L. iners* at the B level. Both membrane transport (odds ratio (OR)=2.34 (95% confidence interval (CI)=1.28-4.44) largest p value=0.0118, Figure 22A) and nucleotide metabolism (OR=2.01 (CI=1.17-3.47) largest p value=0.0274, Figure 22B) were higher in *L. iners* dominant women (more likely to have values above median). Three pathways were less likely to be above median levels in *L. iners* dominant women: transcription (OR=0.16 (CI=0.085-0.30) largest p value=4.35E-8, Figure 22C), cell growth and death (OR=0.23 (CI=0.13-0.39) largest p value=2.76E-7, Figure 22D), and signal transduction (OR=0.30,

(CI=0.16-0.55) largest p value= 0.00709, Figure 22E). At the KO level there were 14 pathways that passed all five statistical tests (Table 11). The majority were decreased in *L. iners* communities: Galactose metabolism, citrate cycle, two-component system, RNA polymerase, aminoacyl tRNA biosynthesis, glyoxylate and dicarboxylate metabolism, nitrogen metabolism, arginine metabolism, glycine, serine, and threonine metabolism, pyrimidine metabolism and alanine, aspartate and glutamate metabolism. Amino sugar and nucleotide sugar metabolism, purine metabolism, and ABC transporters were increased in *L. iners* communities. Overall, these results suggest that there is an increased uptake of nutrients in *L. iners* communities and lower levels of metabolism of most nutrients with a few exceptions in nucleotide metabolism, where *L. crispatus* communities showed higher levels of metabolism, and in pathways that indicated higher rates of cell growth.

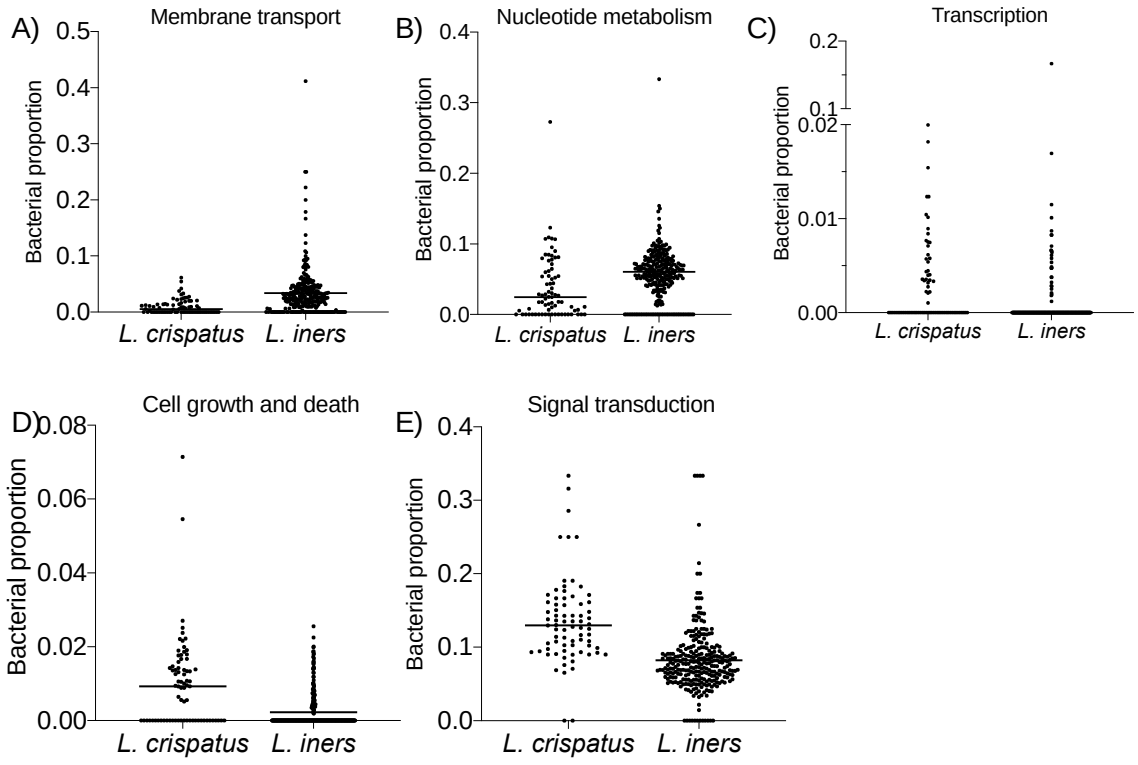

Figure 22. Box plots of significantly different B level pathways between *L. iners* and *L. crispatus* communities. A) Membrane transport **B)** nucleotide metabolism **C)** transcription **D)** cell growth and death **E) s**ignal transduction.

Table 11. KO level pathways that were significantly different between *L. crispatus* and *L. iners* communities.

KO pathway	B level pathway	Fishers p.adj	OR (*L. iners*/ *L. crispatus*), 95% CI
Galactose metabolism	Carbohydrate metabolism	3.48E-14	0.05, 0.02-0.11
Citrate cycle	Carbohydrate metabolism	6.40E-14	0.03, 0.01-0.09
Two component system	Signal transduction	4.84E-08	0.20, 0.11-0.36
RNA polymerase	Transcription	4.84E-08	0.16, 0.09-0.30
Aminoacyl tRNA biosynthesis	Translation	1.01E-07	0.20, 0.11-0.35
Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism	1.55E-07	0.22, 0.12-0.38
Nitrogen metabolism	Energy metabolism	2.94E-07	0.22, 0.13-0.39
Arginine biosynthesis	Amino acid metabolism	4.22E-07	0.23, 0.13-0.40
Glycine serine and threonine metabolism	Amino acid metabolism	1.21E-06	0.22, 0.12-0.40
Pyrimidine metabolism	Nucleotide metabolism	3.95E-06	0.24, 0.14-0.43
Alanine aspartate and glutamate metabolism	Amino acid metabolism	5.93E-05	0.31, 0.18-0.53
Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism	3.86E-05	3.47, 1.97-6.27
Purine metabolism	Nucleotide metabolism	0.00425	2.51, 1.45-4.35
ABC transporters	Membrane transport	0.0345	2.24, 1.21-4.32

There were significant functional differences identified between codominant *Lactobacillus* and *Gardnerella* and *L. crispatus* communities. Four B level functions were significantly different, three were more prevalent in codominant *Lactobacillus* and *Gardnerella* communities: metabolism of terpenoids and polyketides (OR=48.4 (CI=6.44-2164.73) largest p value=1.42E-5, Figure 23A), cellular community of prokaryotes (OR=8.52 (CI=2.96-26.31) largest p value=1.77E-4, Figure 23B), lipid metabolism (OR=9.14 (CI=1.51-98.38) largest p value=0.0403, Figure 23C). Folding, sorting and degradation (OR=0.26 (CI=0.09-0.68) largest p value=0.0195, Figure 23D) was the only pathway that was decreased in codominant *Lactobacillus* and *Gardnerella* communities compared to *L. crispatus* communities. At the more

specific KO level there were 17 pathways that were differentially expressed (Table 12). The nine pathways that were decreased in *Lactobacillus* and *Gardnerella* codominant communities were: amino acid metabolism pathways (arginine biosynthesis, nitrogen metabolism, alanine, aspartate and glutamate metabolism, and glycine, serine and threonine metabolism), the transport pathway two component system, aminoacyl tRNA biosynthesis, and carbohydrate metabolism pathways (glyoxylate and dicarboxylate metabolism, and methane metabolism). The eight pathways that were higher were: ABC transporters, quorum sensing, butanoate metabolism, and biofilm formation (in *E. coli*), biosynthesis of ansamycins, fatty acid degradation, and chloroalkane and chloroalkene degradation. Overall, these results suggest in *Gardnerella* and *Lactobacillus* communities there was increased membrane transport and less emphasis of the bacteria metabolizing or producing its own metabolites and macromolecules. There were increases in pathways associated with negative health effects in the vagina, like butanoate metabolism and biofilm formation, and the increased quorum signaling could be have been signaling to begin producing biofilm related proteins.

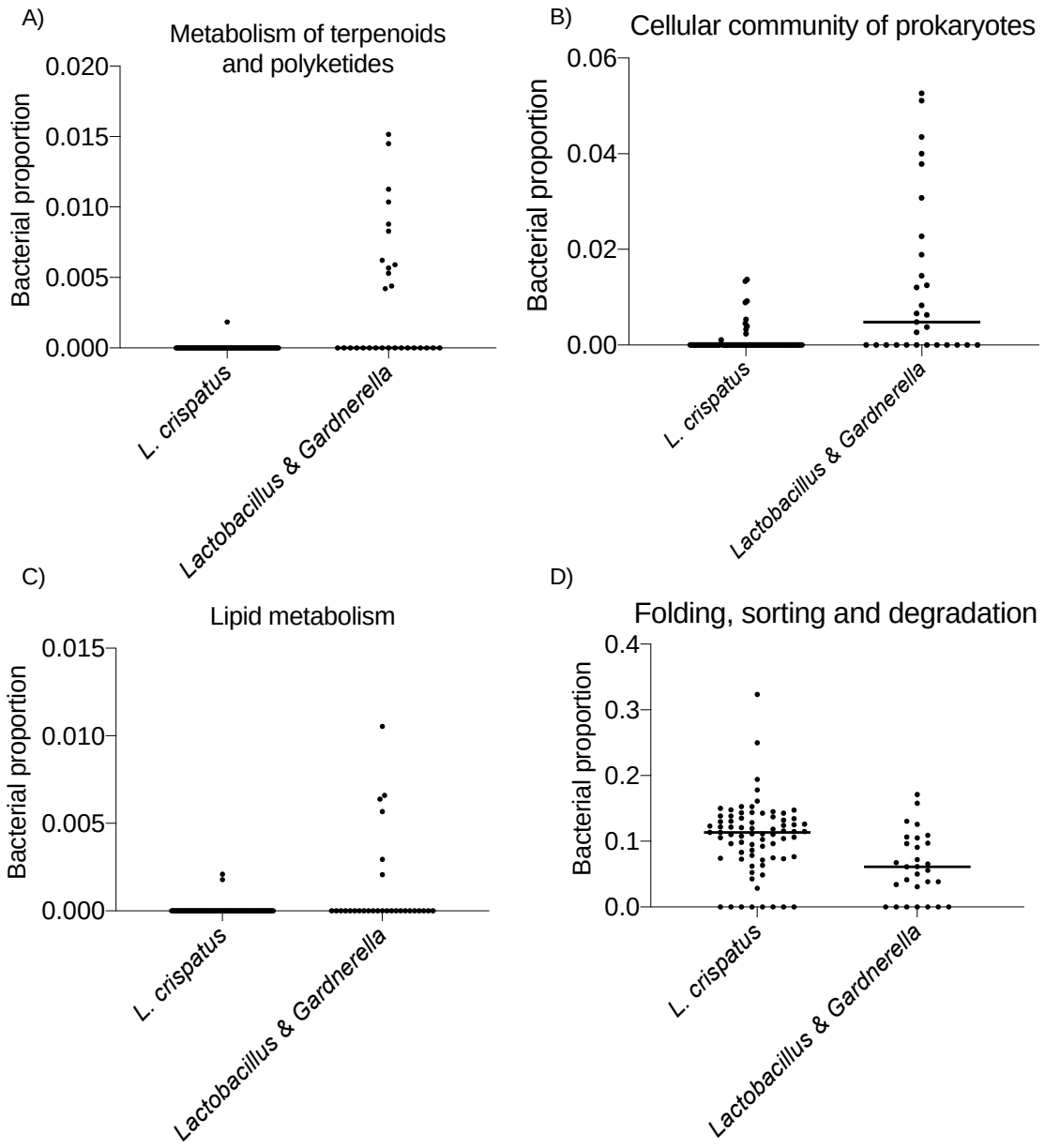

Figure 23. Box plots of significantly different B level pathways between *Lactobacillus* and *Gardnerella* codominant communities and *L. crispatus* communities. A) metabolism of terpenoids and polyketides **B)** cellular community of prokaryotes **C)** folding, sorting and degradation **D)** lipid metabolism.

Table 12. KO level pathways that were significantly different between *L. crispatus* and *Lactobacillus* and *Gardnerella* codominant communities.

KO pathway	B level pathway	Fishers p.adj	OR (*Gardnerella* & *Lactobacillus*/ *L. crispatus)*, 95% CI
Arginine biosynthesis	Amino acid metabolism	7.03E-04	0.11, 0.03-0.365
Nitrogen metabolism	Energy metabolism	7.03E-04	0.11, 0.03-0.365
Alanine aspartate and glutamate metabolism	Amino acid metabolism	0.00155	0.14, 0.04-0.44
Two component system	Signal transduction	0.00262	0.18, 0.06-0.49
Aminoacyl tRNA biosynthesis	Translation	0.0102	0.19, 0.05-0.57
Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism	0.0123	0.22, 0.07-0.61
Inositol phosphate metabolism	Carbohydrate metabolism	0.0303	0.23, 0.06-0.70
Glycine serine and threonine metabolism	Amino acid metabolism	0.0433	0.24, 0.07-0.74
Methane metabolism	Energy metabolism	0.0474	0.31, 0.11-0.79
ABC transporters	Membrane transport	0.0479	3.48, 1.31-9.47
Quorum sensing	Cellular community of prokaryotes	0.00246	6.46, 2.18-20.34
Butanoate metabolism	Carbohydrate metabolism	7.03E-04	11.79, 3.12-56.22
Biofilm formation in *E. coli*	Cellular community of prokaryotes	0.00201	11.96, 2.75-73.98
Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides	5.90E-06	∞, 10.32-∞
Fatty acid degradation	Lipid metabolism	0.0126	∞, 2.57-∞
Chloroalkane and chloroalkene degradation	Xenobiotics biodegradation and metabolism	0.0345	∞, 1.80-∞

There were several differences observed in the functional profiles of women with *Gardnerella* communities compared to those dominated by *L. crispatus*. One that was decreased in *Gardnerella* dominant women was proportion of proteins that could not be assigned a function (OR=0.15, CI=0.03-0.51, p.adj=1.05E-4). At the B level, there were six pathways decreased in women with *Gardnerella* communities compared to *L. crispatus*: glycan biosynthesis, signal transduction, energy metabolism, amino acid metabolism, transcription, and nucleotide

metabolism (Table 13). The pathways increased in *Gardnerella* communities were: metabolism of cofactors and vitamins, xenobiotics biodegradation and metabolism, lipid metabolism, membrane transport, cellular community of prokaryotes, and metabolism of terpenoids and polyketides. Many pathways were significant at the KO level were involved in pathways that were significantly different at the B level: amino acid metabolism, carbohydrate metabolism, energy metabolism, membrane transport, cellular community of prokaryotes. In particular KO pathways butanoate metabolism, biofilm formation (in *E. coli*), ABC transporters, and drug metabolism were KO levels increased in *Gardnerella* dominant women (Table 14). These results suggest that compared to *L. crispatus* communities, *Gardnerella* communities have shifted away from metabolizing macromolecules, instead focusing on transport, and increased pathogenesis associated pathways (butanoate metabolism, flagellar assembly, biofilm formation, and drug metabolism). This data also indicates different carbon source and metabolism preference with *Gardnerella* communities having higher levels of starch and sucrose metabolism and pentose phosphate pathway, while *L. crispatus* had high levels of fructose and mannose metabolism and glycolysis.

Table 13. B level pathways that were significantly different between *L. crispatus* and *Gardnerella* dominant communities.

B level	Fishers p.adj	OR (*Gardnerella*/ *L. crispatus*), 95% CI
Glycan biosynthesis and metabolism	0.00187	0.15, 0.03-0.51
Signal transduction	7.08E-05	0.28, 0.15-0.53
Energy metabolism	7.08E-05	0.30, 0.16-0.54
Amino acid metabolism	0.000178	0.32, 0.17-0.57
Transcription	0.0185	0.47, 0.26-0.87
Nucleotide metabolism	0.0284	0.51, 0.29-0.92
Metabolism of cofactors and vitamins	0.00372	2.68, 1.41-5.33
Xenobiotics biodegradation and metabolism	0.00429	3.56, 1.49-9.85
Lipid metabolism	3.73E-05	11.75, 2.91-103.14
Membrane transport	3.38E-17	14.03, 7.09-28.98
Cellular community of prokaryotes	5.63E-21	23.07, 10.55-55.31
Metabolism of terpenoids and polyketides	1.28E-17	89.30, 14.84-3589.72

Table 14. KO level pathways that were significantly different between *L. crispatus* and *Gardnerella* dominant communities.

KO pathways	B level pathway	Fishers p.adj	OR (*Gardnerella*/ *L. crispatus*), 95% CI
Glycine serine and threonine metabolism	Amino acid metabolism	5.92-07	0.18, 0.09-0.35
Arginine biosynthesis	Amino acid metabolism	2.46E-06	0.23, 0.13-0.43
Alanine aspartate and glutamate metabolism	Amino acid metabolism	9.24E-05	0.30, 0.16-0.53
Tyrosine metabolism	Amino acid metabolism	0.000971	5.72, 1.96-22.88
Cysteine and methionine metabolism	Amino acid metabolism	0.00129	0.37, 0.20-0.66
Fructose and mannose metabolism	Carbohydrate metabolism	8.36E-16	0.07, 0.04-0.14
Butanoate metabolism	Carbohydrate metabolism	8.79E-15	23.10, 8.10-91.17
Inositol phosphate metabolism	Carbohydrate metabolism	3.56E-11	0.09, 0.04-0.19
Galactose metabolism	Carbohydrate metabolism	8.78E-07	0.17, 0.08-0.34
Glycolysis/ Gluconeogenesis	Carbohydrate metabolism	1.14E-05	0.25, 0.14-0.47
Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism	5.28E-05	0.29, 0.16-0.52
Starch and sucrose metabolism	Carbohydrate metabolism	0.000287	3.11, 1.73-5.68
Pentose phosphate pathway	Carbohydrate metabolism	0.000304	3.11, 1.73-5.68
Amino sugar and nucleotide sugar metabolism	Carbohydrate metabolism	0.00111	2.85, 1.56-5.34
Flagellar assembly	Cell motility	0.00583	11.16, 1.74-468.16
Quorum sensing	Cellular community of prokaryotes	7.85E-20	22.83, 10.25-56.70
Biofilm formation in *E. coli*	Cellular community of prokaryotes	3.75E-15	29.78, 9.22-153.57
Nitrogen metabolism	Energy metabolism	4.45E-07	0.20, 0.11-0.38
Carbon fixation in photosynthetic organisms	Energy metabolism	4.67E-05	0.28, 0.15-0.52
Carbon fixation pathways in prokaryotes	Energy metabolism	0.000794	4.59, 1.84-13.80
Methane metabolism	Energy metabolism	0.00113	0.36, 0.19-0.65
Fatty acid degradation	Lipid metabolism	7.51E-07	∞, 5.83-∞
ABC transporters	Membrane transport	6.44E-20	20.23, 9.88-43.49

KO pathways	B level pathway	Fishers p.adj	OR (*Gardnerella*/ *L. crispatus*), 95% CI
Phosphotransferase system	Membrane transport	1.39E-07	0.12, 0.05-0.27
Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides	6.15E-19	∞, 23.04-∞
Thiamine metabolism	Metabolism of cofactors and vitamins	0.000101	4.50, 2.05-10.96
Retinol metabolism	Metabolism of cofactors and vitamins	0.000132	∞, 3.51-∞
D Alanine metabolism	Metabolism of other amino acids	2.18E-05	0.03, 0.00-0.23
Two component system	Signal transduction	8.44E-07	0.22, 0.12-0.40
Aminoacyl tRNA biosynthesis	Translation	8.78E-07	0.21, 0.11-0.39
Chloroalkane and chloroalkene degradation	Xenobiotics biodegradation and metabolism	7.51E-07	∞, 5.83-∞
Drug metabolism - cytochrome P450	Xenobiotics biodegradation and metabolism	0.000132	∞, 3.51-∞

Multiple differences were observed in the functional profiles of women with *Prevotella* codominant communities compared to those with *L. crispatus* communities. *Prevotella* codominant women had increased abundance of proteins that could not be assigned a function (OR=5.32, CI= 1.97-15.20, p.adj=0.00126). At the B level, there were 10 pathways that were different between *Prevotella* and *L. crispatus* communities. The pathways increased in *Prevotella* communities were mainly metabolism pathways (metabolism of cofactors and vitamins, xenobiotics biodegradation and metabolism, metabolism of terpenoids and polyketides, and lipid metabolism), membrane transport, cellular community of prokaryotes, environmental adaptation, cell motility, and (Table 15). Only signal transduction and folding, sorting and degradation pathways were decreased in *Prevotella* codominant communities compared to *L. crispatus* communities. At the KO level there were 17 differentially abundant KO level pathways between *Prevotella* codominant and *L. crispatus* communities (Table 16). The KO

pathways that were increased in *Prevotella* communities included glutathione metabolism, ABC transporters, citrate cycle, quorum sensing, biofilm formation (in *E. coli*), butanoate metabolism, and flagellar assembly. Lower levels of RNA degradation, phosphotransferase system, aminoacyl tRNA biosynthesis, and two component system signaling were found in *Prevotella* codominant communities (Table 16). These data indicate that *Prevotella* communities focused on metabolism of cofactors and vitamins, in particular vitamin B6, compared to *L. crispatus* communities, and had higher levels of ABC transporter proteins, though there were fewer differences in metabolism of carbohydrates and amino acids compared to other non-*Lactobacillus* communities, but *Prevotella* did have increased lipid metabolism and metabolism of ansamycins (compound with antimicrobial activity). As seen in other communities, *Prevotella* communities were more likely to have expression of proteins involved in pathogenesis related pathways like flagellar assembly, biofilm formation, butanoate metabolism.

Table 15. B level pathways that were significantly different between *L. crispatus* and *Prevotella* codominant communities.

B level pathway	Fishers p.adj	OR (*Prevotella*/ *L. crispatus*), 95% CI
Signal transduction	0.0120	0.27, 0.10-0.72
Folding sorting and degradation	0.0270	0.32, 0.12-0.82
Metabolism of cofactors and vitamins	0.00556	4.27, 1.59-11.88
Membrane transport	0.00323	4.57, 1.70-12.73
Xenobiotics biodegradation and metabolism	0.00323	5.98, 1.82-21.06
Cellular community of prokaryotes	2.37E-05	10.62, 3.62-33.73
Lipid metabolism	0.000176	19.01, 3.63-192.34
Environmental adaptation	0.000563	27.88, 3.43-1291.15
Cell motility	6.13E-05	38.32, 4.94-1739.74
Metabolism of terpenoids and polyketides	2.37E-05	44.36, 5.81-2000.10

Table 16. KO level pathways that were significantly different between *L. crispatus* and *Prevotella* codominant communities.

KO pathway	B level pathway	Fishers p.adj	OR (*Prevotella*/ *L. crispatus*), 95% CI
Aminoacyl tRNA biosynthesis	Translation	0.0168	0.24, 0.07-0.71
Two component system	Signal transduction	0.0287	0.30, 0.11-0.80
RNA degradation	Folding sorting and degradation	0.0453	0.32, 0.12-0.82
Phosphotransferase system	Membrane transport	0.000331	0, 0-0.27
Glutathione metabolism	Metabolism of other amino acids	0.00530	5.98, 1.82-21.06
ABC transporters	Membrane transport	0.000497	6.56, 2.39-19.14
Citrate cycle	Carbohydrate metabolism	0.000515	6.96, 2.39-23.53
Quorum sensing	Cellular community of prokaryotes	1.57E-05	11.95, 3.99-39.11
Carbon fixation pathways in prokaryotes	Energy metabolism	1.57E-05	14.03, 4.28-52.70
Biofilm formation in *E. coli*	Cellular community of prokaryotes	3.40E-05	19.25, 4.59-117.81
Butanoate metabolism	Carbohydrate metabolism	3.56E-06	21.46, 5.78-102.74
Plant pathogen interaction	Environmental adaptation	0.000834	27.88, 3.43-1291.15
Flagellar assembly	Cell motility	9.91E-05	38.32, 4.94-1739.74
Vitamin B6 metabolism	Metabolism of cofactors and vitamins	9.91E-05	38.32, 4.94-1739.74
Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides	5.26E-06	∞, 9.34-∞
Fatty acid degradation	Lipid metabolism	0.0001604	∞, 5.57-∞

At the B level, there were four pathways that were different between *Mobiluncus* and *L. crispatus* communities. All of the pathways that were significant (membrane transport, cellular community of prokaryotes, lipid metabolism, and cell motility) were all increased in *Mobiluncus* codominant women (Table 17). At the KO level there were no pathways that were significantly lower in *Mobiluncus* communities, while two carbohydrate metabolism pathways were increased (citrate cycle and), energy metabolism associated pathway carbon fixation pathways in prokaryotes, ABC transporters, quorum sensing, two component system signaling, and

pathogenesis associated pathways (flagellar assembly, butanoate metabolism, and biofilm formation) (Table 18). Codominant *Mobiluncus* communities showed a preference for the citrate cycle for carbohydrate metabolism, and increased pathogenesis associated pathways (flagellar assembly, biofilm formation, and butanoate metabolism), but did not show a difference in amino acid metabolism unlike what was observed with some other communities.

Table 17. B level pathways that were significantly different between *L. crispatus* and *Mobiluncus* codominant communities.

B level pathway	Fishers p.adj	OR (*Mobiluncus*/ *L. crispatus*), 95% CI
Membrane transport	0.00017514	7.20, 2.50-22.65
Cellular community of prokaryotes	2.38E-05	10.37, 3.40-34.26
Lipid metabolism	5.02E-05	22.42, 4.22-228.99
Cell motility	2.25E-20	1170.62, 94.38-∞

Table 18. KO level pathways that were significantly different between *L. crispatus* and *Mobiluncus* codominant communities.

KO level pathway	B level pathway	Fishers p.adj	OR (*Mobiluncus*/ *L. crispatus*), 95% CI
Citrate cycle	Carbohydrate metabolism	0.000492	7.61, 2.44-28.83
ABC transporters	Membrane transport	0.000211	7.74, 2.67-24.52
Quorum sensing	Cellular community of prokaryotes	6.81E-05	9.98, 3.22-33.40
Two component system	Signal transduction	0.00406	14.05, 2.06-605.39
Carbon fixation pathways in prokaryotes	Energy metabolism	2.03E-07	21.25, 6.17-85.23
Biofilm formation in *E. coli*	Cellular community of prokaryotes	4.25E-06	23.49, 5.48-146.43
Butanoate metabolism	Carbohydrate metabolism	5.74E-07	23.63, 6.16-116.73
Flagellar assembly	Cell motility	6.67E-20	1170.62, 94.38-∞

Though there were few individuals with a *Pseudomonas* community several differences were still observed. At the B level, only cell motility was increased (OR=43.38, CI=4.19-2237.21, p.adj=0.00205) between *Pseudomonas* and *L. crispatus* communities. More pathways

were different at the KO level, with lower levels of pathways involved in amino acid metabolism (cysteine and methionine metabolism and alanine, aspartate and glutamate metabolism) and energy metabolism (methane metabolism, nitrogen metabolism, and carbon fixation in photosynthetic organisms) in *Pseudomonas* communities (Table 19). Pathways that were higher in *Pseudomonas* communities included carbon fixation in prokaryotes, flagellar assembly, C5-branched dibasic acid metabolism, and fatty acid deregulation. These results indicate that *Pseudomonas* communities focused on different carbohydrate metabolism pathways (carbon fixation in prokaryotes and c5 branched dibasic acid metabolism) and increased fatty acid metabolism. Though *Pseudomonas* communities did have increased pathogenic associated pathways butanoate metabolism and flagellar assembly, it did not have increased biofilm formation which was seen in all other non-*Lactobacillus* communities.

Table 19. KO level pathways that were significantly different between *L. crispatus* and *Pseudomonas* codominant communities.

KO level pathway	B level pathway	Fishers p.adj	OR (*Pseudomonas*/ *L. crispatus*), 95% CI
Cysteine and methionine metabolism	Amino acid metabolism	0.00249	0.00, 0.00-0.30
Fructose and mannose metabolism	Carbohydrate metabolism	0.00215	0.04, 0.00-0.33
Alanine aspartate and glutamate metabolism	Amino acid metabolism	0.00770	0.06, 0.00-0.45
Methane metabolism	Energy metabolism	0.00527	0.09, 0.00-0.43
Carbon fixation in photosynthetic organisms	Energy metabolism	0.0121	0.10, 0.01-0.48
Nitrogen metabolism	Energy metabolism	0.0253	0.13, 0.01-0.67
Carbon fixation pathways in prokaryotes	Energy metabolism	0.00371	13.10, 2.82-67.25
Flagellar assembly	Cell motility	0.00249	43.38, 4.19-2237.21
C5 Branched dibasic acid metabolism	Carbohydrate metabolism	0.000466	∞, 9.90-∞
Fatty acid degradation	Lipid metabolism	0.0125	∞, 2.70-∞

Chapter 5. An Evaluation of the Impact of Vaginal Bacteria on Epithelial and Neutrophil Cell Function

5.1 Vaginal Epithelial Proteome Changes After Exposure to Bacterial Products

We wanted to further understand the impact of bacteria, or their secreted products, on vaginal epithelial cells. Whole live bacteria were not used as initial attempts resulted in high levels of cell death when using *G. vaginalis*, while use of supernatant did not induce cell death. Vaginal cell (VK2) lysates were collected and analyzed by MS from VK2 cells exposed to: VK2 cell media controls, bacterial media controls, or one of four bacterial supernatants in bacterial media (*L. crispatus*, *G. vaginalis*, *M. mulieris*, *P. amnii*). A total of 2,343 human proteins were retained for analysis, and after quality checks using total sample protein abundance and unsupervised hierarchical clustering no outliers were identified.

There was a total of 257 proteins (11.0% of the detected proteins) that were significantly different between *G. vaginalis* supernatant exposed cells and *L. crispatus* supernatant exposed cells at a p value<0.05, with no proteins significant at the 5% FDR level (Figure 24A). When the significant protein list was analyzed through DAVID the two most significant biological processes enriched in the protein list were cell-cell adhesion (21 proteins, BH p.adj=3.80E-6) and keratinocyte differentiation (7 proteins, BH p.adj=0.043), where both of these pathway's proteins were lower in VK2 cells exposed *to G. vaginalis* bacterial products (Figure 24B and C). Analysis by IPA identified top inhibited functions associated with reduced leukocyte movement (Z score=-2.29, 27 proteins, p=3.85E-4) and proliferation (Z-score=-2.13, 25 proteins, p=0.0013) and decreased molecule transport (Z-score=-2.09, 44 proteins, p=0.00103) as shown in Figure 24D. Activated IPA functions showed increased cell death, both through apoptotic pathways (Z-score=2.23, 93 proteins, p=1.97E-12) and necrotic pathways (Z-score=2.27, 101 proteins,

p=1.02E-16), increased inflammatory ROS pathways including production of ROS (Z-score=2.04, 16 proteins, p=1.83E-4), and quantity of ROS (Z-score=1.58, 9 proteins, p=0.00166), and increased translation (Z-score=1.47, 13 proteins, p=1.61E-4). The non-redundant (removal of subsequent pathways that identify the same proteins) canonical pathways in IPA that were activated or inhibited significantly are shown in Figure 24E. Only regulation of cellular mechanisms of calpain protease was activated (Z-score=-1.00, 4 proteins, p=0.00457), while the following pathways were inhibited: NRF2-mediated oxidative stress response (Z-score=-2.00, 8 proteins, p=0.00214), actin cytoskeleton signaling (Z-score=-1.41, 8 proteins, p=0.00575), extracellular signal-regulated kinases/ mitogen-activated protein kinases (ERK/MAPK) signaling (Z-score=-1.00, 6 proteins, p=0.00316), and ephrin receptor signaling (Z-score=-0.378, 7 proteins, p=0.0049). Use of GSEA to identify what immune cell signatures identified only the down signature in CD8 T cells compared to monocytes, which had a modest signature overlap (NES=-1.31, p=0.0331, 198 proteins, Figure 24F). The match to the down signature reflects that the *G. vaginalis* exposed cells have a more similar protein signature of monocytes.

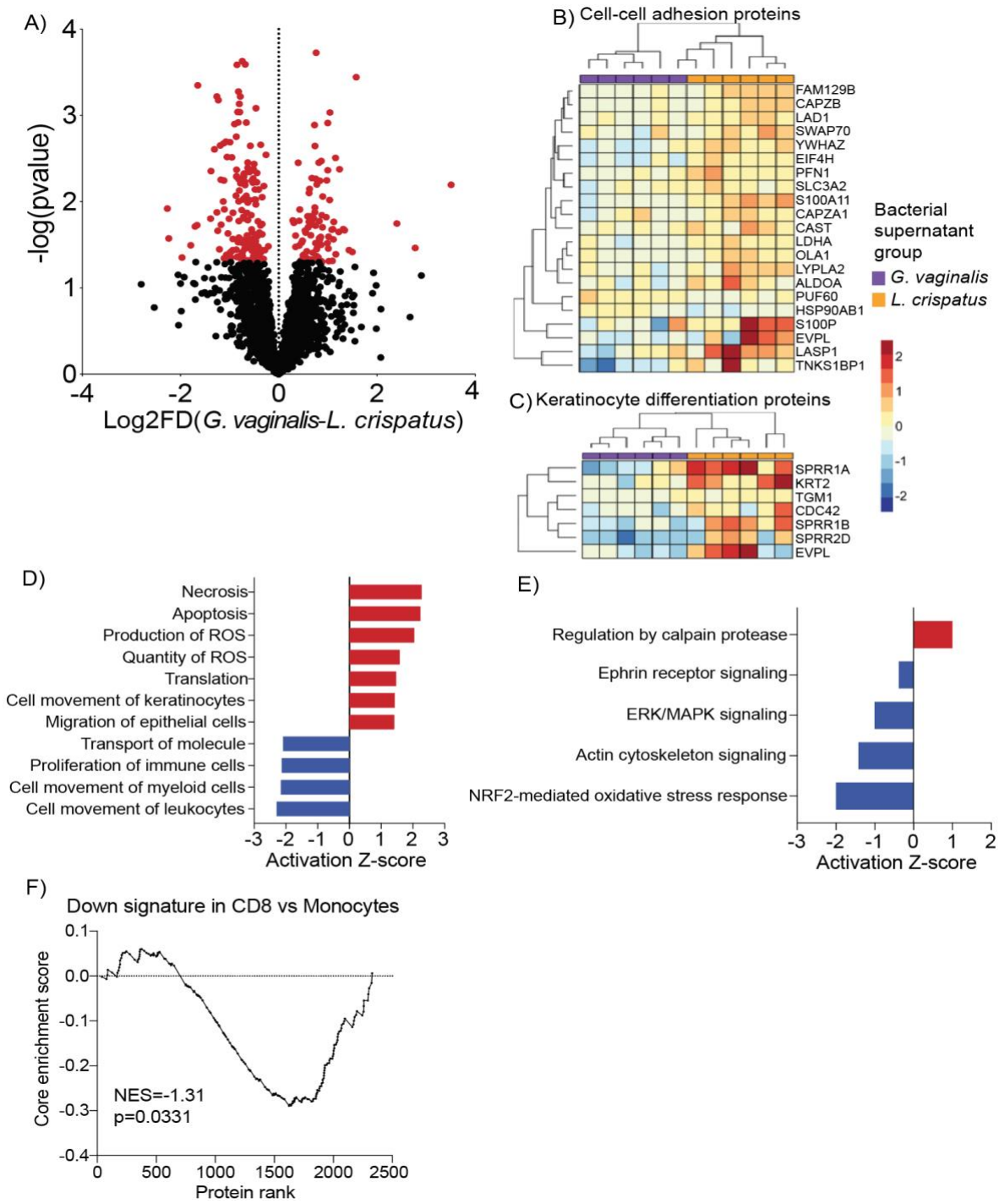

Figure 24. Proteome differences observed between VK2 cells exposed to *G. vaginalis* compared to *L. crispatus*. A) Volcano plot showing proteins significantly different proteins (P<0.05) in red. Heatmaps showing differentially expressed proteins with the legend and scale shown to the side of **B)** the cell-cell-adhesion pathway and **C)** the keratinocyte differentiation pathway. **D)** Functions in IPA significantly activated (red) or inhibited (blue) in *G. vaginalis* exposed cells compared to *L. crispatus* E) Canonical pathways that were activated (red) or inhibited (blue) in *G. vaginalis* exposed cells compared to *L. crispatus*. **F)** Down signature of CD8 T cells compared to monocytes matched to *G. vaginalis* proteome signature.

There was a total of 261 proteins (11.1% of the detected proteins) that were significantly differentially expressed between *M. mulieris* exposed cells and *L. crispatus* exposed cells at a p value<0.05, with 6 proteins significant at the 5% FDR level (Figure 25A). Using unsupervised hierarchical clustering of the top DAVID pathways only leukocyte migration (7 proteins, p.adj=0.56, p unadjusted=0.00950) was able to perfectly sort *M. mulieris* from *L. crispatus* supernatant exposed VK2 cells (Figure 25B). Similar to *G. vaginalis* exposed VK2 cells, cell-cell adhesion was the most significantly enriched DAVID biological process (23 proteins, p.adj=1.40E-7) (Figure 25C). Analysis with IPA identified that the top inhibited functions associated with autophagy of cells (Z score=-1.30, 14 proteins, p=1.82E-4) infection of epithelial cell lines (Z-score=-1.02, 14 proteins, p=4.46E-5) accumulation of blood cells (Z-score=-0.39, 13 proteins, p=4.03E-4), and binding of leukocytes (Z-score=-0.38, 16 proteins, p=3.94E-4) as shown in Figure 25D. Activated IPA functions showed increased killing of cells (Z-score=3.26, 11 proteins, p=4.25E-4), with cell death of epithelial lines (Z-score=2.15, 15 proteins, p=4.83E-5) having a much stronger signal than generic cell death (Z-score=0.39, 118 proteins, p=1.81E-15). Immune associated pathways that were activated included synthesis of ROS (Z-score=2.91, 23 proteins, p=3.41E-6) and accumulation of phagocytes (Z-score=0.56, 9 proteins, p=2.46E-4). Among the canonical pathways in IPA the 5 activated or inhibited pathways are shown in Figure 25E. Both RhoA (Ras homolog family member A) signaling (Z-score=-2.24, 5 proteins, p=0.0162) and sirtuin signaling (Z-score=-0.71, 9 proteins, p=0.00832) were inhibited in *M. mulieris* supernatant exposed VK2 cells. The three activated canonical pathways were: leukocyte extravasation signaling (Z-score=0.38, 7 proteins, p=0.0135), Rac (GTPase signaling protein) signaling (Z-score=0.45, 5 proteins, p=0.00575), and neuroprotective role of THOP1 (Thimet oligopeptidase) in Alzheimer's disease (Z-score=1.63, 6 proteins, p=0.00302). No immune cell

signatures from our in-house immune cell library matched to the *M. mulieris* supernatant exposed epithelial cell lysate proteome.

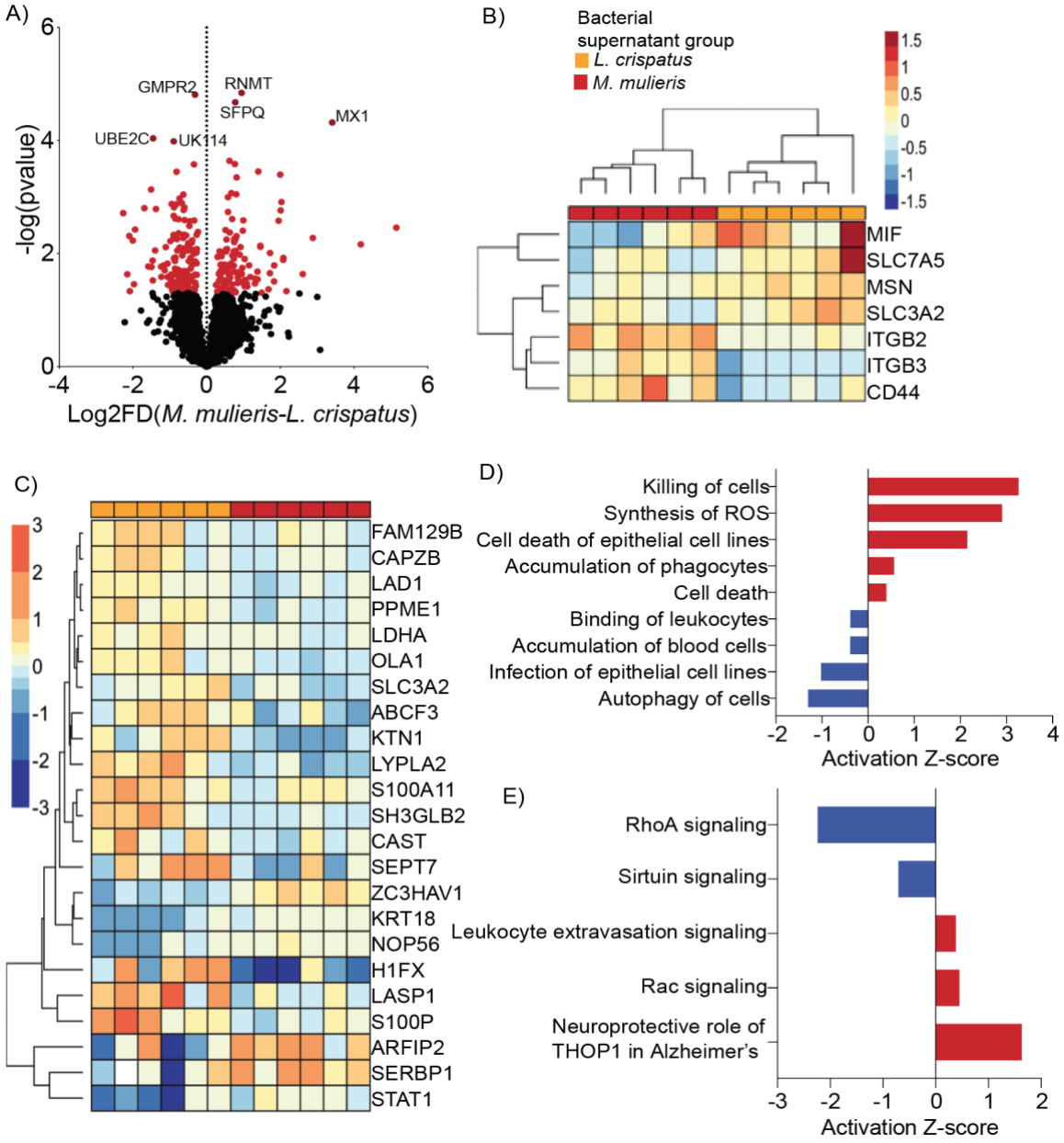

Figure 25. Proteome differences observed between VK2 cells exposed to supernatant from *M. mulieris* compared to *L. crispatus*. A) Volcano plot showing proteins significantly different (P<0.05) in red with the 6 proteins significant at 5% FDR in maroon and labelled. **B)** Heatmap showing differentially expressed proteins in the leukocyte migration pathway with legend and scale shown to the side. **C)** Heatmap with unclustered samples showing differentially expressed proteins in the cell-cell adhesion pathway with the scale shown to the side. **D)** Functions in IPA significantly activated (red) or inhibited (blue) in *M. mulieris* compared to *L. crispatus* exposed VK2 cells. **E)** Canonical pathways that were activated (red) or inhibited (blue) in *M. mulieris* exposed cells compared to *L. crispatus* exposed VK2 cells.

There was a total of 441 proteins (18.8% of the detected proteins) significantly different between *P. amnii* exposed cells and *L. crispatus* exposed cells at a p value<0.05, with 79 proteins significant at the 5% FDR level (Figure 26A). Two of the top biological processes that were enriched in the significant protein list included cell-cell adhesion (45 proteins, p.adj=0.7.40E-20, Figure 26B) and translation initiation (15 proteins, p.adj=0.0072, Figure 26C) where the protein expression showed clear separation between *P. amnii* exposed cells and *L. crispatus* exposed cells. The activated and inhibited cellular functions in *P. amnii* exposed cells identified in IPA are shown in Figure 26D. Similar to both *G. vaginalis* and *M. mulieris* exposed cells, ROS pathways were activated, including quantity (Z score=2.20, 14 proteins, p=2.58E-4) and production of ROS (Z score=2.07, 21 proteins, p=7.82E-4). Similar to *G. vaginalis*, but not *M. mulieris* exposed VK2 cells, increased cell death (Z score=3.01, 182 proteins, p=2.05E-17), cell movement of keratinocytes (Z score=1.78, 9 proteins, p=6.00E-4) and decreased transport of molecules (Z score=-2.90, 78 proteins, p=4.89E-6). Some overlap with *M. mulieris* activated pathways included increased aggregation of blood cells (Z score=-1.00, 17 proteins, p=-1.47E-4), and cell death of epithelial cell lines (Z score=1.42, 17 proteins, p=1.86E-4), though increased growth (Z score=1.01, 35 proteins, p=6.69E-4) and adhesion of epithelial cell lines (Z score=1.20, 12 proteins, p=1.38E-5) were also increased in *P. amnii*. Also observed in *P. amnii* supernatant exposed cells were inhibited synthesis of nucleotides (Z score=-1.48, 22 proteins, p=2.15E-4) and leukocyte migration (Z score=-1.26, 43 proteins, p=8.36E-4), with activation of ubiquitination pathways (Z score=1.66, 22 proteins, p=4.29E-5). Among the canonical pathways in IPA there were 11 activated or inhibited pathways in *P. amnii* supernatant exposed cells after removing pathways with no calculated z-score, pathways with redundant protein lists, and non-significant pathways, which are shown in Figure 26E. Similar to *M. mulieris* exposed cells,

RhoA signaling (Z-score=-1.27, 10 proteins, p=0.00220) and sirtuin signaling (Z-score=-1.90, 13 proteins, p=0.00676) were inhibited, with increased leukocyte extravasation signaling activated (Z-score=0.33, 10 proteins, p=0.0112). Similar to *G. vaginalis* supernatant exposed cells, the most inhibited pathway was NRF2-mediated oxidative stress response (Z-score=0.38, 7 proteins, p=0.0135). Pathways that were inhibited that were unique to *P. amnii* exposed cells were: EIF2 signaling (Z-score=-1.13, 16 proteins, p=1.20E-5), Protein kinase A signaling (Z-score=-1.07, 17 proteins, p=0.00355), and production of NO and ROS in macrophages (Z-score=-0.71, 8 proteins, p=0.0447). Pathways that were activated in *P. amnii* supernatant exposed cells, and not observed in *M. mulieris* or *G. vaginalis* supernatant exposed cells, were regulation of actin based motility by Rho (Z-score=0.45, 6 proteins, p=0.0100), actin nucleation by Actin related protein-Wiskott-Aldrich syndrome family protein (ARP-WASP) complex (Z-score=1.00, 5 proteins, p=0.00550), integrin signaling (Z-score=1.00, 10 proteins, p=0.0145), and Hippo signaling (a kinase involved in angiogenesis) (Z-score=2.45, 8 proteins, p=3.90E-4) which is involved in inhibition of cell proliferation and activation of apoptotic pathways. Similar to *M. mulieris* exposed cells, no immune cell signatures from our in-house immune cell library matched to the ranked protein list that was significant ($p<0.05$).

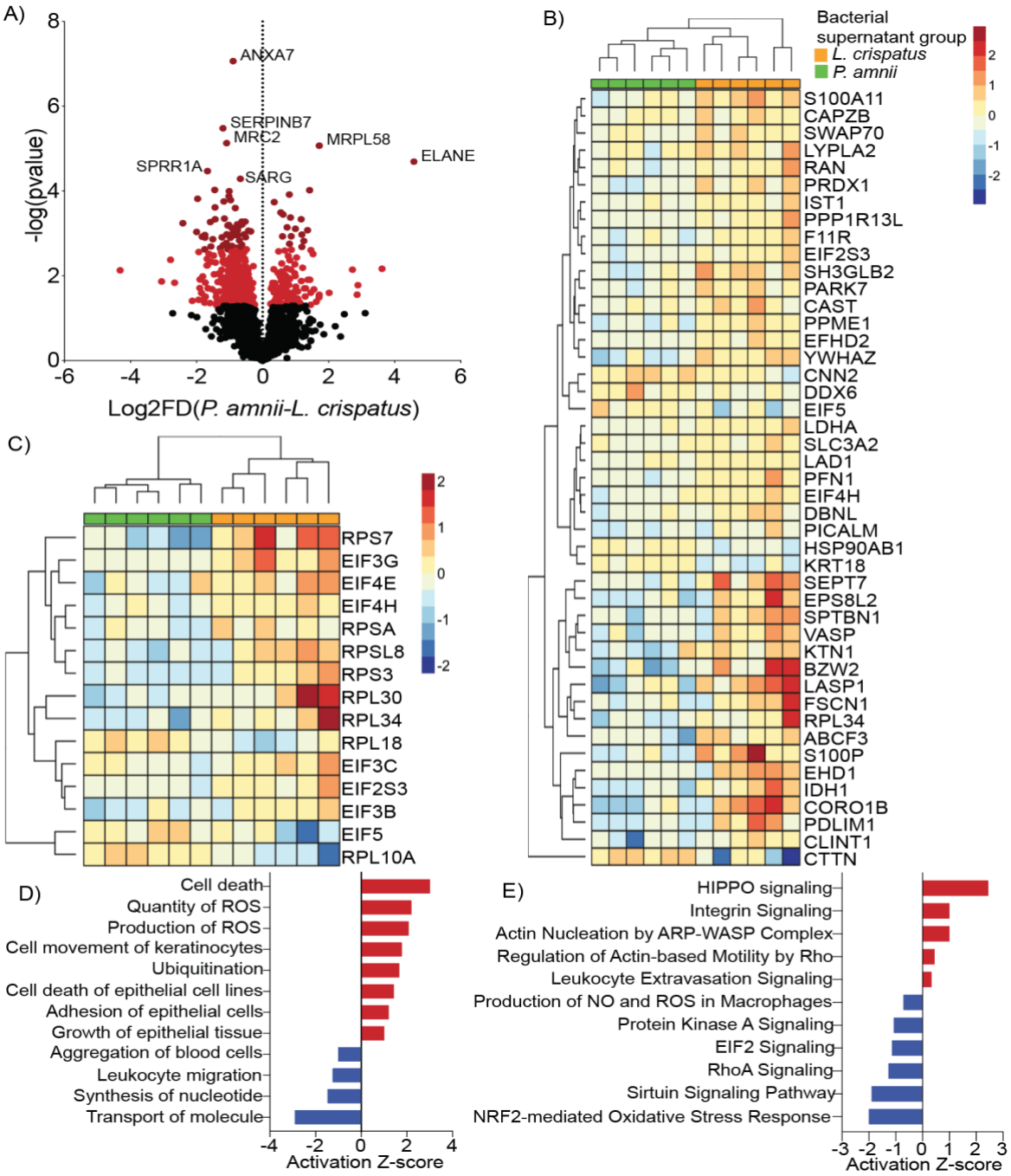

Figure 26. Proteome differences observed between VK2 cells exposed to supernatant from *P. amnii* compared to *L. crispatus*. A) Volcano plot showing proteins significantly different (P<0.05) in red with the 76 proteins significant at 5% FDR in maroon and the top 7 labelled **B)** Heatmap showing differentially expressed proteins in the cell-cell adhesion pathway with the legend and scale shown to the side. **C)** Heatmap showing differentially expressed proteins in the translation initiation pathway with the scale shown to the side. **D)** Functions in IPA significantly increased (in red) or decreased (in blue) in *P. amnii* exposed VK2 cells compared to *L. crispatus* exposed VK2 cells. **E)** Canonical pathways that were activated (in red) or inhibited (in blue) in *P. amnii* exposed cells compared to *L. crispatus* exposed VK2 cells.

5.2 Neutrophil Activation and Migration Towards Vaginal Bacteria or Secreted Bacterial Products

In the previous sections we observed increased neutrophil signatures association with BV associated bacterial communities were chosen for further exploration. To characterize the response of neutrophils to bacteria, *L. crispatus* and *M. mulieris* were chosen. The hypothesis was *L. crispatus*, or its secreted products, would not induce neutrophil migration or activation, while *M. mulieris*, or its secreted products, would induce both. We used transwell inserts that neutrophils were added to and, if attracted to the contents in the bottom well, would move through the 3μm pores into the bottom well. If the neutrophils were activated, they would secrete their granules, release the enzyme myeloperoxidase (MPO) which produces of ROS. For a positive control 10% fetal bovine serum was used because serum activates neutrophils, however the bacterial media also contained horse serum (NYCIII and fastidious anaerobe broth), so it could not be used during the neutrophil experiment (Figure 27). Bacteria were incubated in RPMI before the experiment to avoid background activation due to media.

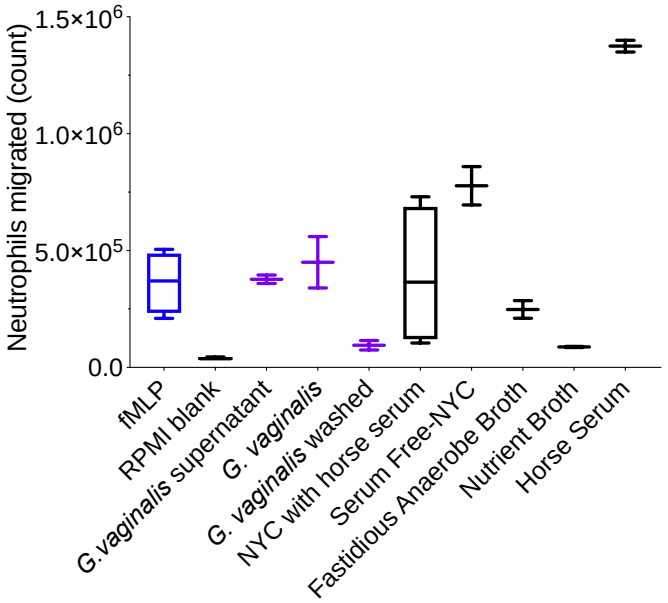

Figure 27. Variation in background neutrophil migration by media type or processing.

In total, six experiments from five separate donors were performed for the *L. crispatus* and *M. mulieris* (bacteria and supernatant in RPMI) experiments, with the two results from the same donor grouped together. Comparison of the positive and negative controls across donors showed a low background for neutrophil migration and relative MPO activity in the negative RPMI media control and a high migration and activation in the positive control (Figure 28A & B). There was a significant increase in migration towards *M. mulieris* compared to *L. crispatus* (average increase of 18,200 neutrophils, p=0.0306, one tailed paired *t*-test, Figure 28C) and increased relative MPO activity (average increase of 4.35%, p=0.0274, one tailed paired *t*-test), Figure 28D.

The secreted products in the supernatant from *L. crispatus* or *M. mulieris* was also assessed to understand if these could induce activation or migration, though in separate experiments from the previous experiment where live bacterial cells were used, in five separate experiments with five separate donors. As seen in the previous experiment a low background was observed for the negative media control for both neutrophil migration and activation (Figure 28E-F). After adjustment for migration observed in the negative media control, there was a significant increase in migrated neutrophils in the wells with *M. mulieris* secreted products compared to the *L. crispatus* supernatant conditions (average increase 12,700, p=0.00751, one tailed paired t-test), and no difference between relative MPO activity (average increase 1.00%, p=0.186, one tailed paired t-test) (Figure 28G-H).

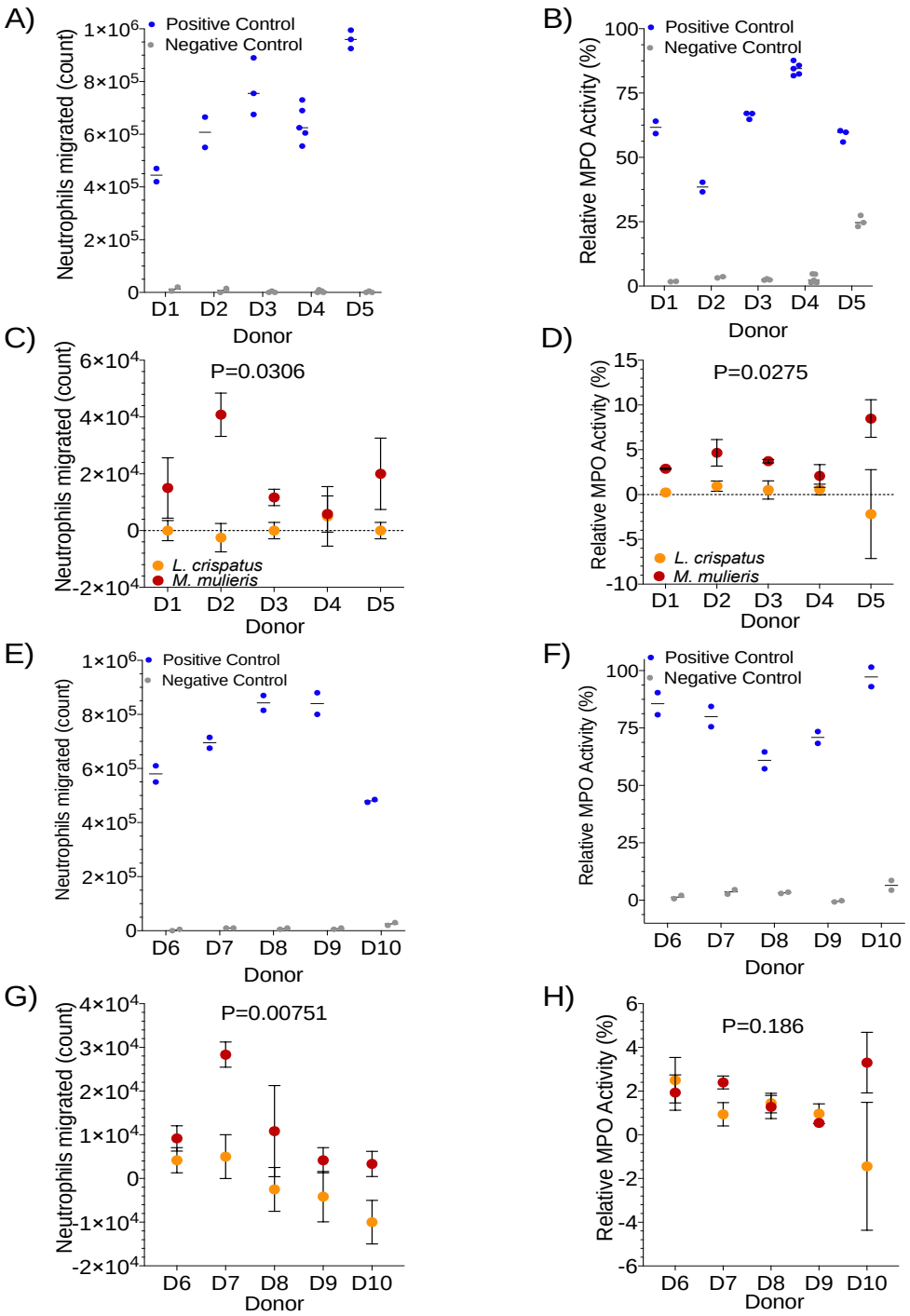

Figure 28. Results of neutrophil migration and activation experiments with bacteria or bacterial secreted products. Comparison of positive and negative controls across donors for bacteria in supernatant experiments for **A)** neutrophil migration and **B)** relative MPO activity. Donor paired data (after adjustment) of bacteria in supernatant experiments for **C)** neutrophil migration and **D)** relative MPO activity. Comparison of positive and negative controls across donors for bacterial supernatant experiments for **E)** neutrophil migration and **F)** relative MPO activity. Donor paired data (after adjustment) of bacterial supernatant experiments for **G)** neutrophil migration and **H)** relative MPO activity after adjustment.

5.3 Neutrophil Activation and Migration Towards Epithelial Cells Exposed to Vaginal Bacteria or Bacterial Products

After assessing the response of vaginal epithelial cells to bacterial secreted products and the response of neutrophils to bacteria or their secreted products, it was further investigated if incorporation of vaginal epithelial cells into the neutrophil model would modify the strength of the neutrophil response. I hypothesized that the response of VK2 cells to *M. mulieris* or *M. mulieris'* secreted products would augment the neutrophil response, while exposure to *L. crispatus* or its products still would not induce migration or activation.

Following the experimental set up from Kusek et al (*267*), the underside of the 3μm trans well insert was seeded with VK2 cells, however the media did not have sufficient surface tension and the liquid rapidly filtered through the pores. This meant that this experimental approach could not be followed exactly, so 24 well plates were seeded with VK2 cells (no inserts added until neutrophil migration). The cells were grown for seven days to ensure confluency before neutrophil experiments were performed. Before the experiments could be performed the appropriate length incubation period of the epithelial cells and bacteria, or bacterial secreted products, needed to be determined so that the epithelial cell viability was not greatly affected during the experiment by bacteria that potentially have deleterious effects.

Incubation times of 1, 2, and 4 hours were tested with each bacterial condition, negative control of RPMI media only, and the positive control 100ng/ml TNF-α which has been previously shown to induce IL-8 secretion in *in vitro* experiments with vaginal cells (*268*). Compared to VK2 cells exposed to just RPMI media, incubation with either *L. crispatus* or *M. mulieris*, their products, or TNF-α did not affect cell viability as measured by WST-1 assay after 1, 2 or 4 hours (Figure 29A-C). For all neutrophil experiments a 4 hour incubation was used to

allow for the maximal cellular response against the bacteria. An additional cell viability WST-1 assay was performed after the first set of neutrophil epithelial cell experiments and found that compared to VK2 cells in RPMI with no neutrophils, neutrophil exposure to VK2 cells in RPMI reduced viability significantly (100% vs. 78%, p=0.0335) as shown in Figure 29D. This reduced viability was trending in the TNF-α positive control exposed VK2 cells with neutrophils added (100% vs. 77%, p=0.0646), but reduced viability was not observed in cells that had bacteria or secreted bacterial products. In total, five experiments from five separate donors were performed. Comparison of each donors positive and negative control revealed a higher background for neutrophil migration (Figure 29E) compared to neutrophil experiments with no epithelial cells, however a threefold increase in migration comparing the TNF-α positive control to the negative RMPI media control was still observed. Neutrophil activation was also assessed, through use of the proxy marker MPO activity in the supernatant as used in the previous neutrophil experiments that did not incorporate VK2 cells. In the same 5 VK2 neutrophil experiments used for the neutrophil migration the relative MPO activity was compared. Surprisingly, positive and negative control revealed no difference with a much higher background for neutrophil activation (Figure 29F) compared to neutrophil experiments with no epithelial cells.

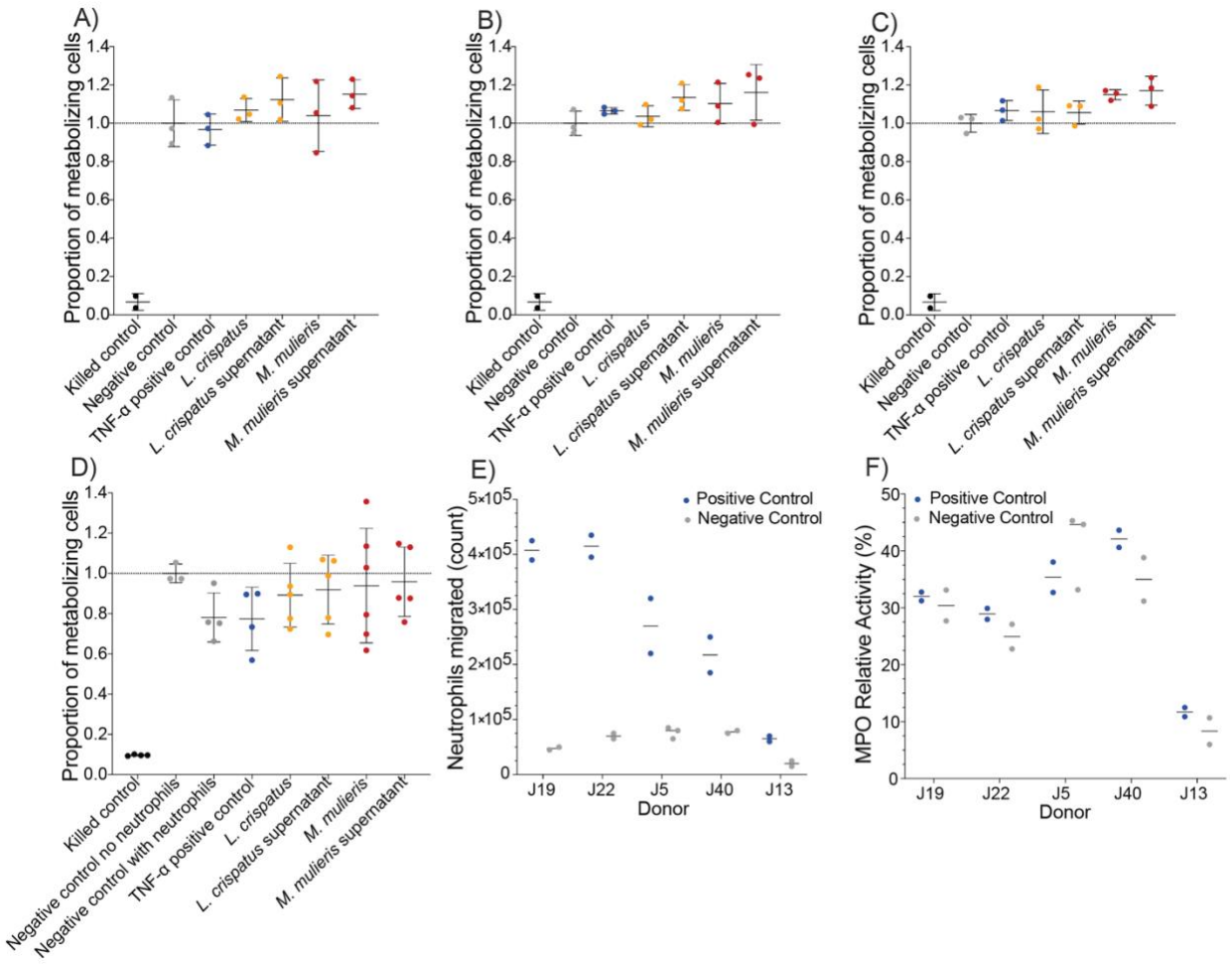

Figure 29. Optimization of experiments incorporating vaginal epithelial cells into neutrophil experiment. Results of WST-1 assay of incubation of VK2 cells and conditions used in neutrophil experiments for **A)** 1 hour incubation **B)** 2 hour incubation **C)** 4 hour incubation **D)** 4 hour incubation plus 3 hour incubation for neutrophil migration experiment. **E)** Neutrophil migration observed in positive control TNF-α and negative media control stratified by donor. **F)** Neutrophil activation, inferred through relative MPO activity, observed in positive control TNF-α and negative media control stratified by donor.

Incorporation of vaginal epithelial cells into the neutrophil migration experiments showed an increased response of the neutrophils. After adjustment for migration observed in the negative media control, there was no significant difference between migrated neutrophil in the *L. crispatus* and the *L. crispatus* supernatant conditions (p=0.459, with no significant difference between migration in the *M. mulieris* or its supernatant condition (p=0.631). When comparing

between the bacteria, there was a significant increase in migration towards *M. mulieris* bacteria in supernatant compared to *L. crispatus* bacteria in supernatant (average increase of 91,300 neutrophils, p= 0.0116, one tailed paired *t*-test) and compared to *L. crispatus* supernatant only (average increase of 84,800 neutrophils, p=0.0132, one tailed paired *t*-test). When neutrophils were exposed to *M. mulieris* supernatant secreted products, there was no significant increase compared to *L. crispatus* bacteria in supernatant (average increase of 118,800, p=0.0661, one tailed paired *t*-test) or *L. crispatus* supernatant (average increase of 125,300 neutrophils, p=0.0704, one tailed paired *t*-test) due to a large standard deviation. After adjustment for migration observed in the negative media control, there was no significant increase between migrated neutrophil in the *L. crispatus* and the *L. crispatus* supernatant conditions (p=0.459, and no difference between migration in the *M. mulieris* or its supernatant condition (p=0.0687). After adjustment for relative MPO activity observed in the negative media control, there was no significant increase between *L. crispatus* and *L. crispatus* supernatant conditions (p=0.624, and a trend towards increased relative MPO activity in the *M. mulieris* compared to its supernatant condition (p=0.0687). When comparing between bacteria, there was a significant increase in migration towards *M. mulieris* compared to *L. crispatus* (average increase 10.12%, p= 0.00634, one tailed paired *t*-test) and when compared to *L. crispatus* supernatant (average increase 9.36%, p= 0.0252, one tailed paired *t*-test). Comparison of *M. mulieris* supernatant showed no significant difference in relative MPO activity comparing both *L. crispatus* bacteria and supernatant (average increase 3.96%, p= 0.0905, one tailed paired *t*-test) or *L. crispatus* supernatant (average increase 3.19%, p= 0.199, one tailed paired *t*-test).

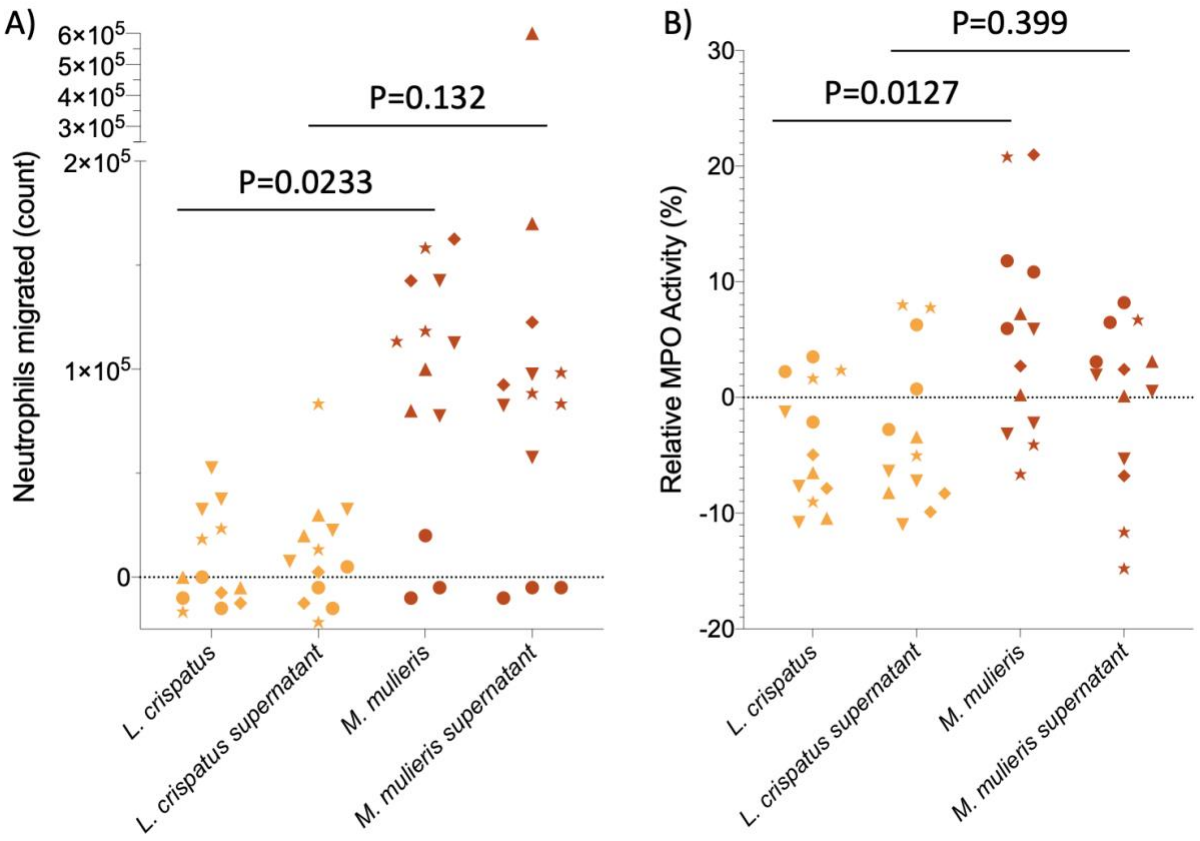

Figure 29. Vaginal epithelial cells response to *L. crispatus*, *M. mulieris*, or their products and how neutrophils migrate or release MPO containing granules. **A)** Neutrophil migration observed in bacteria or bacterial secreted product conditions, each donor indicated with a different symbol. **B)** Relative MPO activity observed in bacteria or bacterial secreted product conditions, each donor is indicated with a different symbol, with the symbols matched across graphs.

Chapter 6. Discussion

6.1 General Discussion

Vaginal dysbiosis is highly prevalent and can have major consequences for reproductive health and susceptibilities to STI's including HIV, however many people with vaginal dysbiosis have no signs or symptoms. Molecular mechanisms of vaginal dysbiosis pathogenesis are not well characterized and though treatments are available for vaginal dysbiosis, patients often have recurrence. Understanding of how vaginal dysbiosis occurs develops and results in pathogenesis is needed to be able to develop more effective treatments and improve reproductive health. The works of this thesis included method development utilizing MS to detect both host and bacterial proteins in CVL samples from a large clinical cohort to further our understanding of the impact of vaginal dysbiosis on host inflammation and epithelial damage proteome signatures. With this technique we were able to characterize not only bacterial communities and their functional profiles but also host proteins and pathways that were consistently identified by MS. We used the bacterial communities to group women to understand differential host inflammation and epithelial associated signatures and how they varied in the different communities, and how the bacterial functional profiles differed. Signals of disrupted epithelial barriers and increased leukocyte migration and activity, in particular neutrophils, were observed in non-*Lactobacillus* communities. We observed increased protein levels of pathogenesis related pathways in non-*Lactobacillus* communities, and differential carbohydrate metabolism. Use of *in vitro* experiments to explore the response of vaginal epithelial cells to bacterial supernatant identified lowered cell adhesion in BV-associated bacteria. Neutrophil experiments found increased neutrophil migration and MPO activity, indicative of neutrophil activation, in response to *M. mulieris* and not *L. crispatus*, and when exposed to the secreted products from these bacteria, *M.*

mulieris supernatant induced migration but not activation of neutrophils. Incorporation of vaginal epithelial cells showed a synergistic effect in the numbers of migrated neutrophils but not in the level of relative MPO activity in the presence of *M. mulieris*, however supernatant of *M. mulieris* did not significantly increase neutrophil migration or MPO activity.

6.1.1 Large Scale MS Metaproteomic Analysis

Mass spectrometry is a powerful technique used to provide in-depth characterization of the proteome, but several considerations need to be made for large-scale studies that require experiments to occur over long periods of time. Due to the stochastic nature of shotgun mass spectrometry, where some peptides are identified in one sample and not the next (*284*), reproducibility of proteome coverage in each sample is a major consideration in large experiments. While this study showed variability in protein identification between samples, there was a consistently detected proteome that was identified across all experiments, which represented the vast majority of host protein abundance, providing good candidates for biomarker discovery. This provides large scale proteomics studies with a major statistical advantage, with increased power to observe differences, identify and correct for differences in clinical variables, as well as compare subsets of data. This is in contrast to most shotgun MS studies of clinical samples, of which many have been limited to pooling samples or low numbers. Additionally, this approach does not limit the ability to examine within experiment proteomes, allowing for more in-depth analysis of protein biomarkers or pathways of interest.

Concerns about contributions of sample preparation steps on downstream proteome variation, including technician who prepared the samples, liquid chromatography run, digestion batch, and other experimental steps were rather minimal after incorporation of a computational approach for batch-correction (*269*). This allowed for 701 samples to be analyzed as a single

experiment, eliminating the need for multiple within-batch comparisons to maintain statistical power to identify biological relationships important for mucosal inflammation.

6.1.2 Depth and Diversity of Host Mucosal Proteome

There are multiple techniques that can be used to gain insight on the host response in different conditions. In comparison to cytokine arrays, unbiased shotgun MS requires no *a priori* knowledge and is able to capture proteins that represent more information on host biological pathways, instead of only focusing on very specific immune response. Immune cell phenotyping by flow cytometry can identify what cells are present and what markers are expressed, though this does not inform on the epithelial barrier, while immunohistochemistry or metatranscriptomic analysis of a biopsy can provide information on both the epithelial barrier and immune cells, however both of these samples are more invasive than lavage.

Cervicovaginal lavage MS analysis contained proteins involved in immune, metabolic, and barrier pathways. The proteins that had 100% sample coverage contained proteins involved in adaptive and innate immunity, inflammation, epithelial barrier, and metabolism pathways, and proteins in these pathways were not enriched in the <100% sample coverage proteins. Pathways enriched in the low sample coverage proteome and not the 100% sample coverage proteome included intracellular pathways including transport and metabolism, however it is understandable that these were more inconsistently detected as intracellular proteins, unless they are secreted for an additional functional role, are going to be lower abundance in extracellular fluid. As there was coverage of pathways often studied for research on infection, transmission and pathogenesis, use of proteins with 100% sample coverage only can still be capable of identifying important associations. Additionally, these proteins are ideal biomarker candidates if they are increased or

decreased in a condition, as they are easily detectable by MS. Use of multiple proteins to identify a disease signature can be more accurate than a single protein alone.

6.1.3 Comparison of Metaproteomics to Other Bacterial Analysis Methods

A metaproteomic experimental approach allows for both compositional and functional profiling of the vaginal microbiome. This provides some advantages over targeted approaches to study the microbiome, including 16S rRNA sequencing, which only provides community compositional data. However, 16S rRNA sequencing is able to achieve a better resolution level as it can read millions of sequences, identifying low abundance taxa more confidently than MS, due to the vast dynamic range of the proteome where it is difficult to capture bacteria that are not active or in low abundance due to low protein output. Conversely, MS can provide some species level information that may not be accessible by 16S rRNA sequencing (*285*, *286*). Some species level resolution was able to be obtained with metaproteomics, but in most cases strain or subtype level confidence was not achieved based on identified peptides. Species level specificity is very important, in particular for Lactobacillus species can vary in their ability to produce lactic acid and/or inflammation responses in vaginal epithelial cells (*158, 159, 272, 287*). Fortunately, the majority of detected *L. iners* proteins were species specific, agreeing with previous literature that *L. iners* is a unique *Lactobacillus* species, as it has evolved to survive in the FGT, while *L. crispatus*, like other *Lactobacillus* species, can be found in other body sites, such as the gastrointestinal tract (*288, 289*). Use of shotgun metagenomics, like 16S rRNA, provides greater depth of coverage, provided that host DNA sequences can be removed as this can be sequenced instead of bacterial sequences. Metagenomics also can provide strain level resolution and identify differences in inferred functional potential, though these genes identified may or may

not be expressed (*290*). While metagenomic and metatranscriptomic analyses provides pathway information on microbial communities, there is a poor correlation with proteins expressed (*291-293*), indicating that direct examination of proteins provides unique information on microbiome phenotype. Metaproteomics and metabolomics are excellent techniques to identify the phenotype of the bacteria, important as bacteria can change what they are doing based on environmental signals, and this will not be captured by other methods. Metaproteomics has an advantage over metabolomics in that it is often able to assign proteins to specific taxa, providing more insight into the role of specific bacteria in the community, as metabolomics cannot determine what bacteria produced which metabolites.

6.1.4 Bacterial Function in the Vagina

The benefit of metaproteomic analysis was that it allowed for the functional annotation of bacterial pathways, which provides an important opportunity to understand how changes in bacterial function influence mucosal immunity (*246, 247, 251, 252, 259, 294-296*). In comparison to bacteria grown in isolation on plates, bacteria in the human body have a greater number of considerations to survive, they may have to compete with other bacteria for nutrients, there may be synergy where bacteria share metabolites and products, or they may have to contend with the host that may try to remove the bacteria from the body. Specific bacteria can also play a key role as producers of a compound that other bacteria cannot, such as bacteria that can produce α-amylase to breakdown vaginal glycogen into compounds that the bacteria can metabolize, avoiding reliance on the host (*156, 157*).

All the abundant genera had detected proteins in pathways essential for life, such as carbohydrate metabolism, energy metabolism, amino acid metabolism, and translation, though

the number of proteins annotated to these functions varied between genera. Folding, sorting and degradation was a pathway that was well represented in *Lactobacillus* species and *Bifidobacterium*, both which are generally considered healthy. In low pH environments proteins are more susceptible to degradation and so correct protein folding is essential for proper bacterial function, and *Lactobacillus* species have been found to modify surface proteins in in response to an acidic environment which aid in survival (*297*). RNA molecules can be secreted by bacteria and used for cell-cell communication between bacteria, however if this is recognized by the host, an inflammatory response can be initiated, so high levels of RNA degradation could prevent unintended export of RNA molecules that could activate the host immune system, though some bacteria secrete RNA molecules that can inhibit immune responses (*298-301*). *L. crispatus* and *L. spp* communities had the lowest levels of membrane transport which imply that either they rely more on de novo synthesis than macromolecule uptake or an indicator that less pathogenic bacteria do not adhere to epithelial cells that can slough off into the vaginal lumen. Studies comparing the genome of *L. crispatus* and *L. iners* have shown that L. crispatus is able to synthesize all molecules required, whereas *L. iners* is deficient in several pathways, such as *de novo* amino acid biosynthesis, and thus is dependent on the host, and needs additional transporters (*289*). Some membrane transport was observed in *L. iners* and *Lactobacillus* and *Gardnerella* communities, with the highest levels seen in *Gardnerella*, *Prevotella*, *Mobiluncus*, and *Pseudomonas* communities, which were also the communities with proteins involved in biofilm formation, which would help bacteria adhere to epithelial cells. *Prevotella* was interesting as there were very few proteins involved in membrane transport, though this could indicate that these cells were not in the CVL supernatant, or that it relies on *de novo* synthesis more than uptake of preformed macromolecules, and that there were other bacteria contributing

to the membrane transport proteins detected in this community. Higher rates of membrane transport indicate that BV associated bacterial communities are more dependent on the host, which may be due to the excess nutrients accessible from increased epithelial cell death observed. Only *Mobiluncus*, *Prevotella*, *Pseudomonas*, and *Escherichia* had detected proteins in the cell motility pathway, which is comprised mainly of flagellar proteins. This agrees with known literature as of the included genera as only these have been found to produce flagella (in our dataset), with the exception of *Prevotella* which is considered non-motile, however the only protein detected was MotB, which is part of the basal body of the flagella so this likely represents non-functional flagella in *Prevotella*. An interesting observation was that *Megasphaera* had the majority of its proteome annotated to carbohydrate metabolism, with butanoate metabolism as the predominant carbohydrate metabolism pathway, indicating that this may be a key species in butanoate production, and *Megasphaera* isolates from the gut, were found to have the ability to produce butanoate (*302*), but has not previously been associated with butanoate production in vaginal dysbiosis. There was often decreased amino acid metabolism in non-*Lactobacillus* communities and agrees with previous research that found amino acids phenylalanine, tyrosine, glutamate, and (glycyl)proline were negatively correlated with BV associated symptoms amine odour, elevated pH, and clue cells (*303*). A difference between BV associated bacteria and *Lactobacillus* communities seems to be how nitrogen, or amine, metabolism is utilized. More health associated communities, usually *Lactobacillus*, produce amino acids, while BV associated communities produce tertiary amines (which could partially be responsible for the observed increased pH, and are associated with malodour). A significant portion of bacterial proteins detected in this study were not annotated to a function, indicating substantial biological information remains to be uncovered in the microbial proteome, and these

proteins could have important roles in both bacterial function and influencing levels of host inflammation.

6.1.5 Major Bacterial Communities Observed in CAPRISA-004

Several major communities were identified by metaproteomics. The types of communities observed fit in with what was seen in other studies of African, or African ancestry, populations with *L. crispatus* the second most frequent *Lactobacillus* species in the vagina after *L. iners*, and a high prevalence of non-*Lactobacillus* communities compared to Caucasian and Asian populations (*27*, *63*). These studies categorized vaginal microbiomes into groups termed community state types (CSTI-V) (*27*) or cervicotypes (CT1-4) (*63*). For our *L. crispatus* communities they match closely to CST1 or CT1, both of which are characterized by *L. crispatus* dominant microbiomes. The percentage of women with these *L. crispatus* communities varied, from 45% in Caucasian American, 22% in Black American women, and a low of 14% in Hispanic American women, compared to 11% in our study of South African women, which was very similar to another South African cohort that found 10% of women had *L. crispatus* communities, while higher proportions of women with *L. iners* communities were seen in CAPRISA-004 (42%) compared to Caucasian Americans (27%), Black Americans (31%) or South African women (32%) (*27*, *103*). A small subset of women in our study had *Lactobacillus* such as *L. gasseri* and *L. jensenii*, which would reflect CSTIII or CSTV respectively, and these are the two infrequent CST groups, which is also observed in our dataset. Among the non-*Lactobacillus* dominant, a large proportion of women in our study (25%) were dominated by *Gardnerella*, which would be assigned to the diverse microbiome CSTIV, which includes all non-*Lactobacillus* dominant communities, while CT3 specifies *Gardnerella* dominant

microbiome, which is a more accurate reflection of our data, and 29% of South African women belonged to this group. Fewer women in our study had the stereotypical BV polymicrobial community (8%), which would be assigned into the diverse CSTIV group, and still showed a lower rate of CSTIV microbiomes compared to Black Americans (40%), and compared to the other South African study where it would be classified as CT4 (30%), and this community was linked with increased risk of HIV acquisition in a South African cohort (*27, 103*). Neither the CST or CT grouping included the *Pseudomonas* community, as they did not detect it in their samples (*27, 63, 103*). *Pseudomonas* in the vagina has been infrequently observed, but some studies have reported it, in association with BV associated bacteria (*283, 304*) or in low proportions (*282*). Another study identified *Pseudomonas* in women with vaginal infections, indicating that *Pseudomonas* may be an important aerobic pathogen in women with vaginal infections such as AV, which is less common that BV (*305*). African women, or women of African descent, are more likely to have microbiomes that are associated with less health benefits, including *L. iners* and diverse communities. Genetic factors may play a role in influencing the vaginal microbiome, as has been suggested for the gut microbiome (*306-309*), indeed a study of Korean twins found associations between IL-5 genetic variation and abundance of *Prevotella* species (*310*). Other studies have found that environment is more influential to determining microbiome composition (*311*), and there are many environmental differences between countries including diet, urban or rural living, and climate.

6.1.6 Metaproteome Differences Across Major Bacterial Communities

Only mild differences were observed between women with *L. spp* and *L. crispatus* communities. Very few host changes were observed between *L. spp*, with the main finding a

slight decrease in barrier proteins, and a slight increase in oxidative stress response, though these differences were not significant after Benjamini-Hochberg multiple comparisons correction. No significant differences were identified in bacterial function. Our findings of minimal differences agrees with previous data that though *L. crispatus* is the *Lactobacillus* most associated with vaginal health, *L. gasseri* and *L. jensenii,* are also beneficial *Lactobacillus* species in the vagina, producing D and L lactic acid (*163*), and are not associated with BV, unlike *L. iners* which has frequently been isolated and detected in dysbiotic vaginal microbiomes (*312-321*).

The most commonly identified bacterial community in our study was *L. iners* dominant. Major overall host protein differences observed include increase in mucins, and a slight increase in inflammation associated pathways. The mild inflammation observations have been identified previous studies where *L. iners* is more detrimental than *L. crispatus*, inducing TLR activation and showing increased immune activation and inflammation (*315, 322-324*). An interesting observation that was not observed in other communities was decreased apoptosis of immune cells, including macrophages and T cells, which may indicate that these cells are not undergoing apoptosis because there is still a pathogenic threat that has not been controlled. The increase in mucin expression is contrary to results from Doerflinger *et al*, where there was not an increase in mucin expression from a 3D vaginal epithelial cell model, while we saw increased expression of mucins, though a different study found increased mucin expression in BV associated microbiomes (*322, 325*). This may indicate that this is in response to increased immune activation signaling a need for increased pathogen trapping by mucins, or to compensate for degradation by pathogens which produce carbohydrate degrading enzymes. Though we did not detect the inerolysin protein we did detect a gametolysin, a matrix metalloprotease that is involved in cell division but also is capable of degrading the extracellular matrix (*326, 327*). Our

observation of increased membrane transport associated protein abundance and decreases in amino acid metabolism kolevel pathways supports data that *L. iners* has a reduced genomic capacity to produce or metabolize amino acids and relies on uptake of nutrients from the vaginal lumen compared to *L. crispatus*, and it may use its ability to lyse host cells to gain the nutrients from the environment (*289*). This ability to rely on the host may counter the decreased signal transduction which implies *L. iners* may be less able to sense and respond to changes in the vaginal lumen, as it is more resilient than *L. crispatus*, capable of survival in higher pH, non-ideal growth conditions.

Lactobacillus and *Gardnerella* communities may represent transitory communities as women switch to or from a *Lactobacillus* dominant community to non-dominant *Lactobacillus* community. Transitions in the vaginal microbiome can occur in less than a week, or take much longer, though a study that looked at the persistence of a community very similar to the one we identified, of mixed *L. iners* and *Gardnerella* in a study by Ravel et al that involved daily sampling found this community persisted 3-5 days (*328, 329*). Interestingly this community had proteomic profiles more similar to other non-*Lactobacillus* communities than *Lactobacillus* dominant communities, with decreased host translation (EIF2 signaling), which was only observed in non-*Lactobacillus* dominant communities, and increased immune response, more than was observed in the *L. iners* community. In particular the signatures of neutrophil recruitment and accumulation could have important consequences in providing an environment more conducive to *Gardnerella* growth. Both butyrate metabolism and biofilm formation were increased in the *Lactobacillus* and *Gardnerella* codominant community, which may also help shift the microbiome to one of *Gardnerella* dominance, or more polymicrobial. The decreased amino acid metabolism may reflect the increased scavenging of nutrients in this codominant

community. Use of GSEA identified increased CD8+ T cell and neutrophil proteomic signatures in this community, which agrees with the increased degranulation of cells pathway observed, and may relate to the increased gluconeogenesis, as CD8+ T cells when activated greatly increase their glucose metabolism (*330*). In this community the increased signatures of inflammation and barrier disruption are indicators that the host is likely trying to remove the *Gardnerella*, using phagocytes and neutrophils, as well as potentially increasing the rate of epithelial cells sloughing off in an attempt to dislodge the *Gardnerella* as it binds tighter to the epithelial cells, potentially trying to initiate biofilm formation. Increased sloughing and apoptosis of epithelial cells has been observed in BV women (*331, 332*). This increased sloughing off of epithelial cells may be augmented by neutrophils through degradation of the extracellular matrix.

In our study the majority of women with non-*Lactobacillus* communities were dominated by *Gardnerella*. There were two separate immune responses to *Gardnerella* with respect to the immune system, with the majority of women increase metabolism and the innate immune response, while there was a subset that showed increased humoral immunity. This may reflect different responses to *Gardnerella* or potentially that women with increased humoral immunity have been colonized by *Gardnerella* for a longer period of time. A previous study that assessed inflammation associated pathways or factors found increased IFN-γ, but not increased IL-8, compared to *L. crispatus* dominant women, whereas we observed increased neutrophil proteomic signatures (*63*). Interestingly in this same cohort, an increased risk of HIV acquisition in *Gardnerella* dominant women was observed however in our study there was not an increased risk of HIV seroconversion (*103*). Epithelial barrier associated pathways were consistently lower in *Gardnerella* dominant women, which has been previously reported in studies that assessed women with BV (*247, 333, 334*). Vaginolysin, a cytotoxic pore forming bacterial

protein produced by *Gardnerella*, contributes to epithelial cell damage and may contribute to the compromised epithelial barrier and increased immune activation observed in our study (*335-339*), and MS was able to detect vaginolysin, as well as several other cytolysins (putative perfringolysin O, pneumolysin, cytolysin, thiol activated cytolysin). One dramatic difference in bacterial function was increased membrane transport, specifically ABC transporters, which implies increased uptake of nutrients and more metabolically active bacteria. Within carbohydrate metabolism pathways, most were downregulated, however butyrate metabolism, starch and sucrose metabolism, and pentose phosphate pathway were higher in *Gardnerella* communities, indicating it is utilizing different pathways than *L. crispatus*. *Gardnerella* is potentially able to utilize glycogen directly with its α-1,4 glucan phosphorylase, which we were able to detect with proteomics, instead of relying on host to provide enzymes potentially in an attempt to overgrow the commensal *Lactobacillus*. Increased biofilm formation and quorum sensing were observed which agrees with previous findings of biofilm formation in the vagina (*179*, *184*). Further analysis on how bacterial functional change can impact host immunity are needed to understand *Gardnerella* pathogenesis. Much like what is potentially occurring in *Lactobacillus* and *Gardnerella* codominant communities, the host appears to be attempting to remove *Gardnerella*, in potentially a dose-dependent response to the amount of *Gardnerella*. Increased attempts by *Gardnerella* dominant communities could initiate the increased epithelial sloughing, which by our pathway analysis may be categorized into the various epithelial barrier pathways as an attempt to remove the *Gardnerella* from the vaginal lumen. It would be interesting to evaluate individuals from the two different clusters based on host protein expression to see if there is variable bacterial function between individuals who had increased humoral response or those with increased innate responses.

Microbiomes with codominance of *Prevotella* is a frequently seen in women with vaginal dysbiosis. Similar to *Gardnerella* dominant communities there were strong signatures of barrier disruption and increased immune activation. In Anahtar *et al*, they observed high levels of pro-inflammatory cytokines in women with *Prevotella*, as well as in response to presence of *Prevotella* in epithelial in vitro experiments (*63*). Use of a 3D vaginal epithelial model found that *Prevotella* colonization induced IL-1β and MIP-3α (*322*), which binds to CCR6, and in a separate study was associated with Th17 cell facilitated inflammation in a chronic oral infections (*340*), which are very susceptible to HIV infection (*231, 232*). This may be one reason why this community has been associated with increased HIV infection (*103*), though not in our study. Increased flagellin assembly, though not *Prevotella* derived, was seen in *Prevotella* codominant communities. Flagellin is a TLR5 agonist that initiates a signaling cascade to activate the innate immune response. There was also increased biofilm formation, indicating *Prevotella* may be involved in vaginal biofilms, or can help provide an environment or metabolites that are conducive to biofilm formation. Butyrate metabolism was increased in *Prevotella* communities, which is a common metabolite associated with BV (*303, 333, 341-343*), and can induce inflammatory responses in the vagina (*344-346*). Only *Prevotella* codominant communities showed higher levels of vitamin B6 metabolism, which is interesting as vitamin B6 has several roles that could aid in pathogenesis, as well as being required for survival. Vitamin B6 is capable of quenching ROS, providing a defensive mechanism for these communities against host defenses (*347*). It is a required co-factor for enzymes that are used to break down glycogen, so bacteria in the community can utilize it for a carbon source. More research is needed to understand how *Prevotella* function can influence the vaginal environment and immune responses. *Prevotella* dominant communities seemed to show a similar pathogenesis as

Gardnerella dominant women, with increased biofilm formation and butyrate production, and high levels of inflammatory, immune cell recruitment, and metabolism signatures, with low levels of barrier associated pathways.

Mobiluncus mulieris is a BV-associated bacterium that has recently been associated with increased risk of spontaneous pre-term birth (*281*). *Mobiluncus* codominance associated with decreased barrier proteins and increased host metabolism of sugars via gluconeogenesis and pentose phosphate pathway, and barrier disruption was been observed with an in vitro experiment using *Mobiluncus* supernatant (*348*). We observed signatures of phagocytosis inhibition in *Mobiluncus* codominant women, and an older study also identified decreased indicators of phagocytosis upon exposure to *Mobiluncus curtsii* (*349*). We again observed an association between increased neutrophil-associated inflammation pathways and non-*Lactobacillus* microbial communities. Increased abundance of neutrophil chemokine IL-8 in response to *Mobiluncus* has been observed several times in *in vitro* epithelial experiments (*63, 348, 350*), and correlated highly with pro-inflammatory cytokines in CVL in a subset of another South African cohort (*63*). *Mobiluncus* derived flagellin proteins correlated with infiltration of neutrophils, and has also been found to promote neutrophil survival, providing a mechanism of why *Mobiluncus* codominant communities may induce barrier disruption and immune activation (*351*). *Mobiluncus* codominant communities may have severe effects on the vaginal barrier as there were signatures of connective tissue and fibroblast proliferation decrease (and increased cell death), which are below the lamina propria, that may indicate exposure of these cells to the vaginal lumen. Despite recruitment and activation signatures of many immune cells, the phagocytosis signatures, and cellular immune response were decreased, indicating the potential of immune evasion by *Mobiluncus* that merits further exploration.

Pseudomonas vaginal communities have not been well characterized as they have been rarely studied, especially in association with how the host may be affected. Similar to other non-*Lactobacillus* communities we observed increased inflammatory associated pathways, however what was most striking was the signatures of barrier disruption, which were stronger and more numerous than other non-*Lactobacillus* communities. This included decreased proliferation of connective tissues cells such as fibroblasts, which could be indicative of severe barrier disruption that exposes the underlying cells, which aligns more with desquamative aerobic vaginitis where there is severe epithelial damage that is observable using microscopy. There is a dearth of knowledge on how *Pseudomonas* functions in the vagina may affect the host, however studies of chronic wounds have found that *Pseudomonas aeruginosa* impairs epithelial repair and form biofilms (*352-354*), but we observed no increase in biofilm formation in *Pseudomonas* communities. Both amino acid metabolism and energy metabolism KO level pathways were decreased in *Pseudomonas* women, apart from carbon fixation in prokaryotes, which may indicate *Pseudomonas* scavenges amino acids and does not utilize the same carbon source as *L. crispatus*. Unlike other non-*Lactobacillus* communities, there was no difference between *L. crispatus* and *Pseudomonas* in butyrate metabolism, which further supports that *Pseudomonas* may not be a BV associated bacteria, and instead an AV pathogen. Future analysis of how bacterial function is altered within communities could provide important insight on why there is variation in host responses, with some women who have inflammation, immune cell recruitment, and epithelial disruption, and other women who do not. Comparison of differential bacterial function in women diagnosed with BV compared to AV would be helpful in identifying the functions associated with each condition, as well as the potential differential role in inflammation and epithelial damage that bacterial function has in each.

6.1.7 Epithelial Proteome Differences Upon Exposure to Vaginal Bacterial Products

All three conditions, supernatant from *G. vaginalis*, *M. mulieris*, and *P. amnii*, showed differential protein expression compared to vaginal epithelial cells exposed to *L. crispatus* supernatant. The most ubiquitous finding was decreased cell-cell adhesion proteins in response to non-*L. crispatus* supernatant, which aligns with our *in vivo* cohort data and previous studies that BV associated bacteria can impair the epithelial barrier (*247, 333, 334*). Only *P. amnii* exposed cells replicated our *in vivo* data with respect to decreased host translation. Research on the pathogenic *Legionella pneumophila* has found that inhibition of host translation elongation blocks the activation of the host's unfolded protein pathway and limits the innate immune response (*355*), which implies *P. amnii* is attempting to circumvent the immune response. All bacteria inhibited the RhoA signaling pathway in IPA, involved in host cell cycle progression, adhesion and epithelial wound repair, matching results from an experiment with bacterial supernatant, which found that *G. vaginalis* impaired epithelial wound repair in vitro with HeLa cells in a scratch assay, however *P. amnii* and *M. mulieris* were not assessed (*247*). All three bacteria had increased signatures of production of ROS, which was not a consistent finding in the *in vivo* data, which is surprising since this is a method employed by both macrophages and neutrophils to destroy pathogens. Sirtuin signaling was decreased in both *P. amnii* and *M. mulieris*, and this pathway is involved in inflammation and cellular response to stress, and inhibitors of sirtuins, which are histone deacetylases, has been explored as treatment for cancer and immunodeficiency viruses (*356*). As this was epithelial cell lysate, bacterial proteins could not be assessed to incorporate how bacterial function could impact host responses to the bacteria. Future studies that focus on identifying the bacterial secreted proteins and metabolites to better

understand what specifically the vaginal cells are reacting to would help identify bacterial functions that may be the perpetrators of the negative health consequences observed in BV.

6.1.8 Response of Neutrophils Upon Exposure to *M. mulieris* or its Secreted Products

Our *in vitro* co-culture data showed that *M. mulieris* can activate neutrophils as well as increase the chemotaxis of neutrophils. Neutrophil involvement has been poorly characterized in the vaginal mucosal microenvironment and the vaginal microbiome, despite the neutrophil chemokine IL-8 being one of the most abundant chemokines or cytokines in the vagina (*208, 357*), primarily due to the challenges in studying these immune cells from mucosal specimens, as they are not viable after freezing. However, recruitment and accumulation of neutrophils has been found to be important for control of *Candida* vaginal infections through mechanisms such as phagocytosis, neutrophil extracellular traps, reactive oxygen species, and destructive enzymes (*211-214, 358-362*). Previous studies have found increased IL-8 secretion in *in vitro* vaginal epithelial cells exposed to *Mobiluncus curtsii (350)* or other BV-associated bacteria (*345, 363*), although not all BV associated bacteria induce IL-8 secretion (*364*). Cohort studies have found variable results, which suggests not all BV associated bacteria may associate with neutrophil recruitment (*365-369*). Our results suggest that *Mobiluncus*-associated neutrophil activity is associated with increased mucosal inflammation and epithelial barrier disruption. While neutrophils are critical in protection from infections, there is evidence that aberrant neutrophil responses can be harmful. Indeed, models of inflammatory bowel disease (IBD) suggest that neutrophils in the GI tract during intestinal inflammation may contribute to disease (*370, 371*), and it is clear that neutrophils can induce tissue damage (*372-376*). Indeed, the unregulated release of proteases and other factors into the extracellular space can paradoxically damage host tissues, by degrading structural proteins of mucosal surfaces (*372, 373*). High levels of IL-1, IL-

6, and particularly IL-8 were found to be good indicators of compound toxicity when assessing anti-HIV compounds such as nonoxyl-9, which was discontinued as a prophylactic as it was found to increase risk of HIV acquisition (*377, 378*), high IL-8 levels were also part of the HIV risk cytokine and chemokine signature identified in CAPRISA-004 (*379*). As such, neutrophil recruitment and activation may be an underlying feature contributing to reproductive health outcomes, including pre-term birth, and HIV acquisition risk. Further studies on the role of this bacterium, and others, in vaginal immunity is warranted to better understand these potential contributing mechanisms. Future work on studying the neutrophil response to other bacteria would further clarify what bacteria induce neutrophil migration and how the neutrophils respond, in particular proteins or metabolites that increase or decrease the level of migration or neutrophil activation.

6.2 Study Limitations

This study analyzed mucosal samples of 701 women using MS metaproteomics methods. Ion suppression of bacterial peptides by more abundant host peptides reduces the ability of MS to detect and identify bacterial peptides, resulting in low or no coverage in some samples (*380*), however MS data-independent acquisition would likely improve detection as this technique does not choose which peptides to sequence based on abundance like data-dependent acquisition (*381*). There are limitations of metaproteomics in comparison to other methods. One is that protein identification is dependent upon the availability and accuracy of database sequences that match the species and strains potentially in samples (*382, 383*). However, inclusion of too many sequences increase the false discovery rate, reducing the confidence in protein identification. Better databases that are specific to vagina could increase identifications/ confidence of protein

identification. Ideally the best proteomic database would be formed from metagenomes found in the same samples. However, errors or misidentifications/ or low levels of confidence in the database (that can be incurred by using non-curated lists, such as TREMBL) are able to provide species level specificity more reliably than 16S rRNA sequencing, however many proteins were not strain specific, and were only identified as such due to database set up, and a more comprehensive database for identifications. On a subset of our samples 16S rRNA sequencing was performed by collaborators and there was good agreement between the two methods in classification of individuals into *Lactobacillus* dominant or non-*Lactobacillus* dominant, and the more abundant genera showed significant positive correlations, including *Lactobacillus Gardnerella*, *Prevotella* and *Mobiluncus* (*259*). There can also be issues with database creation if there are bacteria with many species that end up being a disproportionate contributor, as this may skew the protein results by having a high false discovery rate. Another issue is that, compared to other sample types, there can be low recovery of bacterial proteins from mucosal samples. Protein post translational modifications and metabolites can play an important role in signaling, activation of pathways, location and these were not assessed in this thesis.

Analysis of the different bacterial communities in the CAPRISA-004 cohort could have been augmented by comparing all communities against each other to identify differentially regulated pathways between the non-*Lactobacillus* communities that could have aided in determining unique protein or pathway associations in these communities. This would have provided better insight into the potential differential pathogenesis mechanisms. Use of mathematical models to identify the minimum proteins that differentiate communities could have helped us understand what is unique about each community. Adjustment for various epidemiological variables known to influence the vaginal microbiome may have also helped

tease out interesting associations, though fortunately there were very few differences between *L. crispatus* and other communities. Our study also relies on the *L. crispatus* community being an accurate representation of the ideal vaginal microbiome, with accompanying healthy host environment, and this may not be the case, as there is variation among individuals. This was a cross-sectional study that can only identify associations and cannot provide temporal information about whether the bacterial community or function changed and then influenced the host to change, or if the host changed and provided a different environment that allowed non-*Lactobacillus* bacteria to grow.

There were also several limitations of the *in vitro* experiments. In the epithelial cell experiments bacterial supernatants were used as initial tests found that some bacteria, such as *Gardnerella* induced cell death. Bacteria in specific laboratory culture conditions may act a certain way and may not be reflective of protein expression *in vivo*. Mixed bacterial cultures were not used that would have been more reflective of dysbiotic microbiomes, but this could be assessed in future studies.

Neutrophil experiments had to grow bacteria in RPMI, of which the contents were amino acids, glucose, inorganic salts, and B vitamins, which may not be the bacteria's preferred food and not represent the nutrients available in the vaginal environment. Changes in initial pH of the RPMI over time may make it more difficult for the *L. crispatus* to be viable after plating the next day. Fortunately, both *L. crispatus* and *M. mulieris* would be similar pink-yellow colours, which indicate a slightly acidic environment when using the pH indicator phenol red. Future studies should note solution pH before and after bacterial incubation. Inclusion of flagellin as a neutrophil chemoattractant would have been very informative in the experiments to see if its presence induced both activation and recruitment, only one, or neither, and compare that to the

whole bacteria and supernatant experiments to see if it may explain most or only some of the observed neutrophil differences. Unfortunately, we were unable to recapitulate the experimental set up for the VK2 neutrophil experiments used in Kusek *et al* due to weak surface tension that prevented seeding of VK2 cells on underside of transwell inserts (*267*). This would have reflected a more *in vivo* situation where neutrophils would have had to move through the epithelial barrier to reach the bottom well where the chemoattractants were located, simulating the vaginal barrier and lumen. Future work on studying the neutrophil response and additional bacteria would further clarify what bacteria induce neutrophil migration and how the neutrophils respond, identifying if particular proteins or metabolites increase the level of migration or neutrophil activation.

6.3 Contributions and Importance

In the course of my studies, I have developed a method to characterize human and bacterial proteomes of hundreds of mucosal clinical samples, which was previously limited to smaller experiment sizes of <100 reported in the literature, and generally focused on either the host or bacterial proteins, not both simultaneously. This method can be applied to other sample types and studies to gain knowledge into host-bacterial interactions important for disease related pathogenesis. This method resulted in a dataset that was able to identify several major bacterial communities and identify the host proteome differences. A major finding is that there was a large proportion, nearly 40%, of women who had non-*Lactobacillus* dominant microbiomes. These women showed higher levels of inflammation and barrier damage proteomic signatures that could potentially mean they are at increased risk of STI's, including HIV. One of the major findings was a consistent decrease in barrier associated proteins in non-*Lactobacillus*

communities, particularly in *Pseudomonas* communities, a bacterium infrequently references with respect to vaginal colonization. The other more novel host finding was increased neutrophil signatures in non-*Lactobacillus* communities, which could be perpetrators of inflammation. This study identified increased levels of proteins in various pathogenesis associated pathways in non-*Lactobacillus* communities, though there was variation in which pathways. There were differences in bacterial function in each community, highlighting the importance of studying bacterial communities separately. This variation in bacterial function both across communities, and within communities could have important implications for inflammation and HIV prevention efficacy, and merits further exploration. Use of *in vitro* experiments were able to further confirm proteomic changes in cell adhesion in epithelial cells exposed to BV associated bacteria. Neutrophil experiments identified *Mobiluncus* as a recruiter and activator of neutrophils and may be an important initiator in bacterial induced vaginal inflammation and epithelial barrier disruption. These data were also used to identify an important observation between the efficacy of the pre-exposure prophylactic anti-HIV vaginally applied gel used in this cohort (1% tenofovir gel) in women who did not have a *Lactobacillus* dominant microbiome, highlighting the importance of considering the vaginal microbiome in prevention technology development and testing (*259*). Additionally, this dataset was able to identify increased HIV seroconversion in DMPA users, only in those who were *Lactobacillus* dominant, significant differences were not observed in women who were non-Lactobacillus dominant or in those using other forms of contraceptives (*384*). This work identifies potential important interactions between hormonal contraceptives, the vaginal microbiome, and how this may impact HIV susceptibility.

This work is important because vaginal dysbiosis is still not well understood, and treatment options are still limited and often ineffective. Mechanisms of damage and the role

different bacterial communities may play has previously been poorly understood. Further study of bacterial pathways different in communities associated with STI susceptibility or reproductive conditions could help identify a pathway important for the pathogenesis and symptoms seen in some cases of dysbiosis. Longitudinal cohort studies and animal models will be essential for linking cause and effect between specific bacteria, pathways, and health associated outcomes.

6.4 Conclusion

This thesis describes a workflow and utility of mass spectrometry based metaproteomics to study mucosal specimens in a large cohort setting, and describe the scope of the detected host and bacterial proteome. This comprehensive host and bacteria proteome data identified relationships between microbial composition and function with mucosal inflammation and neutrophil activation *in vivo*, which was validated using multiple *in vitro* experiments. We were able to show novel relationships between anaerobic bacteria associated with poor reproductive health outcomes and neutrophil activation. Application of large-scale metaproteomic studies to clinical mucosal samples is a useful tool to understand host-microbial interactions in disease and identify novel interactions.

References:

1. H. H. Bragulla, D. G. Homberger, Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. *J Anat* **214**, 516-559 (2009).
2. D. J. Anderson, J. Marathe, J. Pudney, The structure of the human vaginal stratum corneum and its role in immune defense. *Am J Reprod Immunol* **71**, 618-623 (2014).
3. A. Baroni *et al.*, Structure and function of the epidermis related to barrier properties. *Clin Dermatol* **30**, 257-262 (2012).
4. C. R. Wira, M. Rodriguez-Garcia, M. V. Patel, The role of sex hormones in immune protection of the female reproductive tract. *Nat Rev Immunol* **15**, 217-230 (2015).
5. X. Zhao *et al.*, Vaginal submucosal dendritic cells, but not Langerhans cells, induce protective Th1 responses to herpes simplex virus-2. *J Exp Med* **197**, 153-162 (2003).
6. C. D. Blaskewicz, J. Pudney, D. J. Anderson, Structure and function of intercellular junctions in human cervical and vaginal mucosal epithelia. *Biol Reprod* **85**, 97-104 (2011).
7. D. L. Patton *et al.*, Epithelial cell layer thickness and immune cell populations in the normal human vagina at different stages of the menstrual cycle. *Am J Obstet Gynecol* **183**, 967-973 (2000).
8. L. Miller *et al.*, Depomedroxyprogesterone-induced hypoestrogenism and changes in vaginal flora and epithelium. *Obstet Gynecol* **96**, 431-439 (2000).
9. D. A. Eschenbach *et al.*, Effects of oral contraceptive pill use on vaginal flora and vaginal epithelium. *Contraception* **62**, 107-112 (2000).
10. L. Y. Hwang *et al.*, Active squamous metaplasia of the cervical epithelium is associated with subsequent acquisition of human papillomavirus 16 infection among healthy young women. *The Journal of infectious diseases* **206**, 504-511 (2012).
11. G. B. Moss *et al.*, Association of cervical ectopy with heterosexual transmission of human immunodeficiency virus: results of a study of couples in Nairobi, Kenya. *J Infect Dis* **164**, 588-591 (1991).
12. L. Y. Hwang, M. E. Scott, Y. Ma, A. B. Moscicki, Higher levels of cervicovaginal inflammatory and regulatory cytokines and chemokines in healthy young women with immature cervical epithelium. *Journal of reproductive immunology* **88**, 66-71 (2011).
13. R. P. Madan *et al.*, Altered biomarkers of mucosal immunity and reduced vaginal Lactobacillus concentrations in sexually active female adolescents. *PloS one* **7**, e40415 (2012).
14. L. Masson *et al.*, Genital inflammation and the risk of HIV acquisition in women. *Clin Infect Dis*, (2015).
15. V. B. Young, The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* **356**, j831 (2017).
16. P. Vernocchi, F. Del Chierico, L. Putignani, Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. *Front Microbiol* **7**, 1144 (2016).
17. D. Atkins, G. T. Furuta, Mucosal immunology, eosinophilic esophagitis, and other intestinal inflammatory diseases. *J Allergy Clin Immunol* **125**, S255-261 (2010).

18. D. N. Frank *et al.*, Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* **104**, 13780-13785 (2007).
19. C. Juste *et al.*, Bacterial protein signals are associated with Crohn's disease. *Gut* **63**, 1566-1577 (2014).
20. J. R. Kelly *et al.*, Transferring the blues: Depression-associated gut microbiota induces neurobehavioural changes in the rat. *J Psychiatr Res* **82**, 109-118 (2016).
21. F. Scheperjans *et al.*, Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov Disord* **30**, 350-358 (2015).
22. J. D. Galley *et al.*, Exposure to a social stressor disrupts the community structure of the colonic mucosa-associated microbiota. *BMC Microbiol* **14**, 189 (2014).
23. M. Ferrer *et al.*, Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environ Microbiol* **15**, 211-226 (2013).
24. R. E. Ley, P. J. Turnbaugh, S. Klein, J. I. Gordon, Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022-1023 (2006).
25. J. U. Scher *et al.*, Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife* **2**, e01202 (2013).
26. A. Doderlein, Die scheidensekret untersuchugen. *Zentralb Gynakol* **18**, 10-14 (1894).
27. J. Ravel *et al.*, Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4680-4687 (2011).
28. S. Boris, J. E. Suarez, F. Vazquez, C. Barbes, Adherence of human vaginal lactobacilli to vaginal epithelial cells and interaction with uropathogens. *Infect Immun* **66**, 1985-1989 (1998).
29. J. Osset, R. M. Bartolome, E. Garcia, A. Andreu, Assessment of the capacity of Lactobacillus to inhibit the growth of uropathogens and block their adhesion to vaginal epithelial cells. *J Infect Dis* **183**, 485-491 (2001).
30. N. Phukan, T. Parsamand, A. E. Brooks, T. N. Nguyen, A. Simoes-Barbosa, The adherence of Trichomonas vaginalis to host ectocervical cells is influenced by lactobacilli. *Sex Transm Infect* **89**, 455-459 (2013).
31. F. Atassi, D. Brassart, P. Grob, F. Graf, A. L. Servin, Vaginal Lactobacillus isolates inhibit uropathogenic Escherichia coli. *FEMS Microbiol Lett* **257**, 132-138 (2006).
32. F. Atassi, D. Brassart, P. Grob, F. Graf, A. L. Servin, Lactobacillus strains isolated from the vaginal microbiota of healthy women inhibit Prevotella bivia and Gardnerella vaginalis in coculture and cell culture. *FEMS Immunol Med Microbiol* **48**, 424-432 (2006).
33. D. E. O'Hanlon, T. R. Moench, R. A. Cone, In vaginal fluid, bacteria associated with bacterial vaginosis can be suppressed with lactic acid but not hydrogen peroxide. *BMC Infect Dis* **11**, 200 (2011).
34. D. E. O'Hanlon, T. R. Moench, R. A. Cone, Vaginal pH and microbicidal lactic acid when lactobacilli dominate the microbiota. *PLoS One* **8**, e80074 (2013).
35. M. Aldunate *et al.*, Vaginal concentrations of lactic acid potently inactivate HIV. *J Antimicrob Chemother* **68**, 2015-2025 (2013).
36. D. Tyssen *et al.*, Anti-HIV-1 Activity of Lactic Acid in Human Cervicovaginal Fluid. *mSphere* **3**, (2018).

37. Z. Gong, Y. Luna, P. Yu, H. Fan, Lactobacilli inactivate Chlamydia trachomatis through lactic acid but not H2O2. *PLoS One* **9**, e107758 (2014).
38. C. E. Isaacs, W. Xu, Theaflavin-3,3'-digallate and lactic acid combinations reduce herpes simplex virus infectivity. *Antimicrob Agents Chemother* **57**, 3806-3814 (2013).
39. M. A. Graver, J. J. Wade, The role of acidification in the inhibition of Neisseria gonorrhoeae by vaginal lactobacilli during anaerobic growth. *Ann Clin Microbiol Antimicrob* **10**, 8 (2011).
40. C. Mitchell, J. Marrazzo, Bacterial vaginosis and the cervicovaginal immune response. *Am J Reprod Immunol* **71**, 555-563 (2014).
41. M. J. Mason, A. J. Winter, How to diagnose and treat aerobic and desquamative inflammatory vaginitis. *Sex Transm Infect* **93**, 8-10 (2017).
42. G. G. G. Donders, G. Bellen, S. Grinceviciene, K. Ruban, P. Vieira-Baptista, Aerobic vaginitis: no longer a stranger. *Res Microbiol* **168**, 845-858 (2017).
43. F. Kirakoya-Samadoulougou *et al.*, Bacterial vaginosis among pregnant women in Burkina Faso. *Sex Transm Dis* **35**, 985-989 (2008).
44. F. Kirakoya-Samadoulougou *et al.*, Epidemiology of herpes simplex virus type 2 infection in rural and urban Burkina Faso. *Sex Transm Dis* **38**, 117-123 (2011).
45. C. Kenyon, R. Colebunders, T. Crucitti, The global epidemiology of bacterial vaginosis: a systematic review. *Am J Obstet Gynecol* **209**, 505-523 (2013).
46. K. Peebles, J. Velloza, J. E. Balkus, R. S. McClelland, R. V. Barnabas, High Global Burden and Costs of Bacterial Vaginosis: A Systematic Review and Meta-Analysis. *Sex Transm Dis* **46**, 304-311 (2019).
47. L. Myer *et al.*, Bacterial vaginosis and susceptibility to HIV infection in South African women: a nested case-control study. *J Infect Dis* **192**, 1372-1380 (2005).
48. J. A. Frohlich, Q. Abdool Karim, M. M. Mashego, A. W. Sturm, S. S. Abdool Karim, Opportunities for treating sexually transmitted infections and reducing HIV risk in rural South Africa. *J Adv Nurs* **60**, 377-383 (2007).
49. L. Govender, A. A. Hoosen, J. Moodley, P. Moodley, A. W. Sturm, Bacterial vaginosis and associated infections in pregnancy. *Int J Gynaecol Obstet* **55**, 23-28 (1996).
50. J. A. Dols *et al.*, Microarray-based identification of clinically relevant vaginal bacteria in relation to bacterial vaginosis. *Am J Obstet Gynecol* **204**, 305 e301-307 (2011).
51. K. C. Anukam, E. O. Osazuwa, I. Ahonkhai, G. Reid, Lactobacillus vaginal microbiota of women attending a reproductive health care service in Benin city, Nigeria. *Sex Transm Dis* **33**, 59-62 (2006).
52. K. A. Fethers *et al.*, Early sexual experiences and risk factors for bacterial vaginosis. *J Infect Dis* **200**, 1662-1670 (2009).
53. K. H. Lim *et al.*, Is there a correlation between bacterial vaginosis and preterm labour in women in the Otago region of New Zealand? *Aust N Z J Obstet Gynaecol* **50**, 226-229 (2010).
54. L. Cristiano, S. Rampello, C. Noris, V. Valota, Bacterial vaginosis: prevalence in an Italian population of asymptomatic pregnant women and diagnostic aspects. *Eur J Epidemiol* **12**, 383-390 (1996).
55. E. Gratacos *et al.*, Prevalence of bacterial vaginosis and correlation of clinical to Gram stain diagnostic criteria in low risk pregnant women. *Eur J Epidemiol* **15**, 913-916 (1999).

56. D. Guerreiro, M. A. Gigante, L. C. Teles, Sexually transmitted diseases and reproductive tract infections among contraceptive users. *Int J Gynaecol Obstet* **63 Suppl 1**, S167-173 (1998).
57. B. Martinez de Tejada, O. Coll, M. de Flores, S. L. Hillier, D. V. Landers, [Prevalence of bacterial vaginosis in an obstetric population of Barcelona]. *Med Clin (Barc)* **110**, 201-204 (1998).
58. U. Nilsson, D. Hellberg, M. Shoubnikova, S. Nilsson, P. A. Mardh, Sexual behavior risk factors associated with bacterial vaginosis and Chlamydia trachomatis infection. *Sex Transm Dis* **24**, 241-246 (1997).
59. J. E. Tolosa *et al.*, The International Infections in Pregnancy (IIP) study: variations in the prevalence of bacterial vaginosis and distribution of morphotypes in vaginal smears among pregnant women. *Am J Obstet Gynecol* **195**, 1198-1204 (2006).
60. H. N. Simhan, L. M. Bodnar, M. A. Krohn, Paternal race and bacterial vaginosis during the first trimester of pregnancy. *Am J Obstet Gynecol* **198**, 196 e191-194 (2008).
61. R. A. Royce *et al.*, Race/ethnicity, vaginal flora patterns, and pH during pregnancy. *Sex Transm Dis* **26**, 96-102 (1999).
62. L. R. McKinnon *et al.*, The Evolving Facets of Bacterial Vaginosis: Implications for HIV Transmission. *AIDS Res Hum Retroviruses* **35**, 219-228 (2019).
63. M. N. Anahtar *et al.*, Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* **42**, 965-976 (2015).
64. A. H. Curtis, On the etiology and bacteriology of leucorrhoea. *Surg Gynecol Obstet* **18**, 299-306 (1914).
65. R. Schroder, Zur pathogenese und Klinik des vaginalem fluors. *Zentralb Gynakol* **38**, 1350-1361 (1921).
66. S. Leopold, Heretofore undescribed organism isolated from the genitourinary system. *U S Armed Forces Med J* **4**, 263-266 (1953).
67. H. L. Gardner, C. D. Dukes, Haemophilus vaginalis vaginitis: a newly defined specific infection previously classified non-specific vaginitis. *Am J Obstet Gynecol* **69**, 962-976 (1955).
68. R. F. Smith, H. A. Rodgers, P. A. Hines, R. M. Ray, Comparisons between direct microscopic and cultural methods for recognition of Corynebacterium vaginale in women with vaginitis. *J Clin Microbiol* **5**, 268-272 (1977).
69. P. A. Totten, R. Amsel, J. Hale, P. Piot, K. K. Holmes, Selective differential human blood bilayer media for isolation of Gardnerella (Haemophilus) vaginalis. *J Clin Microbiol* **15**, 141-147 (1982).
70. J. Martius *et al.*, Relationships of vaginal Lactobacillus species, cervical Chlamydia trachomatis, and bacterial vaginosis to preterm birth. *Obstet Gynecol* **71**, 89-95 (1988).
71. A. L. Blackwell, A. R. Fox, I. Phillips, D. Barlow, Anaerobic vaginosis (non-specific vaginitis): clinical, microbiological, and therapeutic findings. *Lancet* **2**, 1379-1382 (1983).
72. C. A. Spiegel, R. Amsel, D. Eschenbach, F. Schoenknecht, K. K. Holmes, Anaerobic bacteria in nonspecific vaginitis. *N Engl J Med* **303**, 601-607 (1980).
73. J. L. Thomason, P. C. Schreckenberger, W. N. Spellacy, L. J. Riff, L. J. LeBeau, Clinical and microbiological characterization of patients with nonspecific vaginosis associated with motile, curved anaerobic rods. *J Infect Dis* **149**, 801-809 (1984).

74. P. Piot, E. Van Dyck, P. Godts, J. Vanderheyden, The vaginal microbial flora in non-specific vaginitis. *Eur J Clin Microbiol* **1**, 301-306 (1982).
75. G. G. Donders *et al.*, Definition of a type of abnormal vaginal flora that is distinct from bacterial vaginosis: aerobic vaginitis. *BJOG* **109**, 34-43 (2002).
76. R. P. Nugent, M. A. Krohn, S. L. Hillier, Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. *J Clin Microbiol* **29**, 297-301 (1991).
77. R. Amsel *et al.*, Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med* **74**, 14-22 (1983).
78. K. K. Holmes, K. C. Chen, C. M. Lipinski, D. A. Eschenbach, Vaginal redox potential in bacterial vaginosis (nonspecific vaginitis). *J Infect Dis* **152**, 379-382 (1985).
79. J. L. Thomason *et al.*, Proline aminopeptidase activity as a rapid diagnostic test to confirm bacterial vaginosis. *Obstet Gynecol* **71**, 607-611 (1988).
80. W. B. Ayre, The glycogen-estrogen relationship in the vaginal tract. *J Clin Endocrinol Metab* **11**, 103-110 (1951).
81. G. I. Gorodeski, U. Hopfer, C. C. Liu, E. Margles, Estrogen acidifies vaginal pH by up-regulation of proton secretion via the apical membrane of vaginal-ectocervical epithelial cells. *Endocrinology* **146**, 816-824 (2005).
82. E. H. Koumans *et al.*, The prevalence of bacterial vaginosis in the United States, 2001-2004; associations with symptoms, sexual behaviors, and reproductive health. *Sex Transm Dis* **34**, 864-869 (2007).
83. J. R. Schwebke, C. M. Richey, H. L. Weiss, Correlation of behaviors with microbiological changes in vaginal flora. *J Infect Dis* **180**, 1632-1636 (1999).
84. D. Hellberg, S. Nilsson, P. A. Mardh, Bacterial vaginosis and smoking. *Int J STD AIDS* **11**, 603-606 (2000).
85. D. S. Forcey *et al.*, Factors Associated with Bacterial Vaginosis among Women Who Have Sex with Women: A Systematic Review. *PLoS One* **10**, e0141905 (2015).
86. K. Rosene, D. A. Eschenbach, L. S. Tompkins, G. E. Kenny, H. Watkins, Polymicrobial early postpartum endometritis with facultative and anaerobic bacteria, genital mycoplasmas, and Chlamydia trachomatis: treatment with piperacillin or cefoxitin. *J Infect Dis* **153**, 1028-1037 (1986).
87. D. H. Watts, M. A. Krohn, S. L. Hillier, D. A. Eschenbach, Bacterial vaginosis as a risk factor for post-cesarean endometritis. *Obstet Gynecol* **75**, 52-58 (1990).
88. D. E. Soper, R. C. Bump, W. G. Hurt, Bacterial vaginosis and trichomoniasis vaginitis are risk factors for cuff cellulitis after abdominal hysterectomy. *Am J Obstet Gynecol* **163**, 1016-1021; discussion 1021-1013 (1990).
89. P. G. Larsson, J. J. Platz-Christensen, U. Forsum, C. Pahlson, Clue cells in predicting infections after abdominal hysterectomy. *Obstet Gynecol* **77**, 450-452 (1991).
90. M. G. Gravett, D. Hummel, D. A. Eschenbach, K. K. Holmes, Preterm labor associated with subclinical amniotic fluid infection and with bacterial vaginosis. *Obstet Gynecol* **67**, 229-237 (1986).
91. H. Leitich, H. Kiss, Asymptomatic bacterial vaginosis and intermediate flora as risk factors for adverse pregnancy outcome. *Best Pract Res Clin Obstet Gynaecol* **21**, 375-390 (2007).

92. J. A. Svare, H. Schmidt, B. B. Hansen, G. Lose, Bacterial vaginosis in a cohort of Danish pregnant women: prevalence and relationship with preterm delivery, low birthweight and perinatal infections. *BJOG* **113**, 1419-1425 (2006).
93. B. M. Mercer *et al.*, The Preterm Prediction Study: prediction of preterm premature rupture of membranes through clinical findings and ancillary testing. The National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. *Am J Obstet Gynecol* **183**, 738-745 (2000).
94. S. L. Hillier *et al.*, A case-control study of chorioamnionic infection and histologic chorioamnionitis in prematurity. *N Engl J Med* **319**, 972-978 (1988).
95. Z. L. Wang *et al.*, Diagnosis and microecological characteristics of aerobic vaginitis in outpatients based on preformed enzymes. *Taiwan J Obstet Gynecol* **55**, 40-44 (2016).
96. M. C. Brown-Harrison, J. C. Christenson, A. M. Harrison, M. E. Matlak, Group A streptococcal salpingitis in a prepubertal girl. *Clin Pediatr (Phila)* **34**, 556-558 (1995).
97. I. S. Arda, M. Ergeneli, M. Coskun, A. Hicsonmez, Tubo-ovarian abscess in a sexually inactive adolescent patient. *Eur J Pediatr Surg* **14**, 70-72 (2004).
98. S. D. Algren, J. L. Strickland, Beta hemolytic streptococcus group f causing pelvic inflammatory disease in a 14-year-old girl. *J Pediatr Adolesc Gynecol* **18**, 117-119 (2005).
99. M. F. Gallo *et al.*, Bacterial vaginosis, gonorrhea, and chlamydial infection among women attending a sexually transmitted disease clinic: a longitudinal analysis of possible causal links. *Ann Epidemiol* **22**, 213-220 (2012).
100. E. Gillet *et al.*, Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect Dis* **11**, 10 (2011).
101. J. Atashili, C. Poole, P. M. Ndumbe, A. A. Adimora, J. S. Smith, Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. *AIDS* **22**, 1493-1501 (2008).
102. T. L. Cherpes, L. A. Meyn, M. A. Krohn, J. G. Lurie, S. L. Hillier, Association between acquisition of herpes simplex virus type 2 in women and bacterial vaginosis. *Clin Infect Dis* **37**, 319-325 (2003).
103. C. Gosmann *et al.*, Lactobacillus-Deficient Cervicovaginal Bacterial Communities Are Associated with Increased HIV Acquisition in Young South African Women. *Immunity* **46**, 29-37 (2017).
104. M. A. Klebanoff *et al.*, Time course of the regression of asymptomatic bacterial vaginosis in pregnancy with and without treatment. *Am J Obstet Gynecol* **190**, 363-370 (2004).
105. M. B. Mengel *et al.*, The effectiveness of single-dose metronidazole therapy for patients and their partners with bacterial vaginosis. *J Fam Pract* **28**, 163-171 (1989).
106. S. Mehta, Patterns of systemic uptake of 18-FDG with positron emission tomography/computed tomography (PET/CT) studies in patients with presumed ocular tuberculosis. *Ocul Immunol Inflamm* **20**, 434-437 (2012).
107. E. Bukusi *et al.*, Topical penile microbicide use by men to prevent recurrent bacterial vaginosis in sex partners: a randomized clinical trial. *Sex Transm Dis* **38**, 483-489 (2011).
108. M. Zozaya *et al.*, Bacterial communities in penile skin, male urethra, and vaginas of heterosexual couples with and without bacterial vaginosis. *Microbiome* **4**, 16 (2016).
109. C. M. Liu *et al.*, Penile Anaerobic Dysbiosis as a Risk Factor for HIV Infection. *mBio* **8**, (2017).

110. K. A. Workowski, G. A. Bolan, C. Centers for Disease, Prevention, Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm Rep* **64**, 1-137 (2015).
111. J. Symonds, A. K. Biswas, Amoxycillin, augmentin and metronidazole in bacterial vaginosis associated with Gardnerella vaginalis. *Genitourin Med* **62**, 136 (1986).
112. T. A. Pheifer, P. S. Forsyth, M. A. Durfee, H. M. Pollock, K. K. Holmes, Nonspecific vaginitis: role of Haemophilus vaginalis and treatment with metronidazole. *N Engl J Med* **298**, 1429-1434 (1978).
113. M. A. Durfee, P. S. Forsyth, J. A. Hale, K. K. Holmes, Ineffectiveness of erythromycin for treatment of Haemophilus vaginalis-associated vaginitis: possible relationship to acidity of vaginal secretions. *Antimicrob Agents Chemother* **16**, 635-637 (1979).
114. P. Piot, E. Van Dyck, P. Godts, J. Vanderheyden, A placebo-controlled, double-blind comparison of tinidazole and triple sulfonamide cream for the treatment of nonspecific vaginitis. *Am J Obstet Gynecol* **147**, 85-89 (1983).
115. B. Andersch, L. Forssman, K. Lincoln, P. Torstensson, Treatment of bacterial vaginosis with an acid cream: a comparison between the effect of lactate-gel and metronidazole. *Gynecol Obstet Invest* **21**, 19-25 (1986).
116. R. F. Smith, W. E. Dunkelberg, Jr., Inhibition of Corynebacterium vaginale by metronidazole. *Sex Transm Dis* **4**, 20-21 (1977).
117. E. D. Ralph, Y. E. Amatnieks, Relative susceptibilities of Gardnerella vaginalis (Haemophilus vaginalis), Neisseria gonorrhoeae, and Bacteroides fragilis to Metronidazole and its two major metabolites. *Sex Transm Dis* **7**, 157-160 (1980).
118. R. Sobel, J. D. Sobel, Metronidazole for the treatment of vaginal infections. *Expert Opin Pharmacother* **16**, 1109-1115 (2015).
119. J. Spizek, T. Rezanka, Lincomycin, clindamycin and their applications. *Appl Microbiol Biotechnol* **64**, 455-464 (2004).
120. S. Hillier, M. A. Krohn, D. H. Watts, P. Wolner-Hanssen, D. Eschenbach, Microbiologic efficacy of intravaginal clindamycin cream for the treatment of bacterial vaginosis. *Obstet Gynecol* **76**, 407-413 (1990).
121. W. L. Greaves, J. Chungafung, B. Morris, A. Haile, J. L. Townsend, Clindamycin versus metronidazole in the treatment of bacterial vaginosis. *Obstet Gynecol* **72**, 799-802 (1988).
122. M. A. B. Petrina, L. A. Cosentino, L. K. Rabe, S. L. Hillier, Susceptibility of bacterial vaginosis (BV)-associated bacteria to secnidazole compared to metronidazole, tinidazole and clindamycin. *Anaerobe* **47**, 115-119 (2017).
123. S. M. Elghazaly *et al.*, Efficacy and safety of single dose of oral secnidazole 2 g in treatment of bacterial vaginosis: A systematic review and meta-analysis. *Eur J Obstet Gynecol Reprod Biol* **238**, 125-131 (2019).
124. G. G. Donders, K. Ruban, G. Bellen, Selecting anti-microbial treatment of aerobic vaginitis. *Curr Infect Dis Rep* **17**, 477 (2015).
125. G. Tempera, P. M. Furneri, Management of aerobic vaginitis. *Gynecol Obstet Invest* **70**, 244-249 (2010).
126. C. S. Bradshaw *et al.*, High recurrence rates of bacterial vaginosis over the course of 12 months after oral metronidazole therapy and factors associated with recurrence. *J Infect Dis* **193**, 1478-1486 (2006).

127. Y. Shafquat *et al.*, Antimicrobial susceptibility against metronidazole and carbapenem in clinical anaerobic isolates from Pakistan. *Antimicrob Resist Infect Control* **8**, 99 (2019).
128. J. D. Sobel *et al.*, Suppressive antibacterial therapy with 0.75% metronidazole vaginal gel to prevent recurrent bacterial vaginosis. *Am J Obstet Gynecol* **194**, 1283-1289 (2006).
129. M. N. Austin, R. H. Beigi, L. A. Meyn, S. L. Hillier, Microbiologic response to treatment of bacterial vaginosis with topical clindamycin or metronidazole. *J Clin Microbiol* **43**, 4492-4497 (2005).
130. O. O. Oduyebo, R. I. Anorlu, F. T. Ogunsola, The effects of antimicrobial therapy on bacterial vaginosis in non-pregnant women. *Cochrane Database Syst Rev*, CD006055 (2009).
131. F. H. Al-Ghazzewi, Elamir A., Tester R., Elzagze A., Effect of depolymerised konjac glucomannan on wound healing. *Bioactive Carbohydrates and Dietary Fibre* **5**, 125-128 (2015).
132. M. Shahbuddin *et al.*, High molecular weight plant heteropolysaccharides stimulate fibroblasts but inhibit keratinocytes. *Carbohydr Res* **375**, 90-99 (2013).
133. R. F. Tester, F. H. Al-Ghazzewi, Beneficial health characteristics of native and hydrolysed konjac (Amorphophallus konjac) glucomannan. *J Sci Food Agric* **96**, 3283-3291 (2016).
134. T. R. Sutherland A., Al-Ghazzewi F., McCulloch E., Connolly M., Glucomannan hydrolysate (GMH) inhibition of Candida albicans growth in the presence of Lactobacillus and Lactococcus species. *Microbial Health in Health and Disease* **20**, 127-134 (2008).
135. I. Coste, P. Judlin, J. P. Lepargneur, S. Bou-Antoun, Safety and efficacy of an intravaginal prebiotic gel in the prevention of recurrent bacterial vaginosis: a randomized double-blind study. *Obstet Gynecol Int* **2012**, 147867 (2012).
136. I. M. Poynten *et al.*, The safety of candidate vaginal microbicides since nonoxynol-9: a systematic review of published studies. *AIDS* **23**, 1245-1254 (2009).
137. V. Rousseau, J. P. Lepargneur, C. Roques, M. Remaud-Simeon, F. Paul, Prebiotic effects of oligosaccharides on selected vaginal lactobacilli and pathogenic microorganisms. *Anaerobe* **11**, 145-153 (2005).
138. S. L. Collins *et al.*, Promising Prebiotic Candidate Established by Evaluation of Lactitol, Lactulose, Raffinose, and Oligofructose for Maintenance of a Lactobacillus-Dominated Vaginal Microbiota. *Appl Environ Microbiol* **84**, (2018).
139. Z. M. Zeng *et al.*, Directed shift of vaginal flora after topical application of sucrose gel in a phase III clinical trial: a novel treatment for bacterial vaginosis. *Chin Med J (Engl)* **123**, 2051-2057 (2010).
140. H. L. Hou W., Li M., Chen J., Chen Y., Effectiveness evaluation of alginate oligosaccharides antibacterial gel for bacterial vaginosis. *Life Science Journal* **11**, 528-531 (2014).
141. S. Borges, J. Silva, P. Teixeira, The role of lactobacilli and probiotics in maintaining vaginal health. *Arch Gynecol Obstet* **289**, 479-489 (2014).
142. S. J. Klebanoff, S. L. Hillier, D. A. Eschenbach, A. M. Waltersdorph, Control of the microbial flora of the vagina by H2O2-generating lactobacilli. *J Infect Dis* **164**, 94-100 (1991).

143. O. Gillor, L. M. Nigro, M. A. Riley, Genetically engineered bacteriocins and their potential as the next generation of antimicrobials. *Curr Pharm Des* **11**, 1067-1075 (2005).
144. C. Charlier, M. Cretenet, S. Even, Y. Le Loir, Interactions between Staphylococcus aureus and lactic acid bacteria: an old story with new perspectives. *Int J Food Microbiol* **131**, 30-39 (2009).
145. S. Coudeyras, G. Jugie, M. Vermerie, C. Forestier, Adhesion of human probiotic Lactobacillus rhamnosus to cervical and vaginal cells and interaction with vaginosis-associated pathogens. *Infect Dis Obstet Gynecol* **2008**, 549640 (2008).
146. A. Homayouni *et al.*, Effects of probiotics on the recurrence of bacterial vaginosis: a review. *J Low Genit Tract Dis* **18**, 79-86 (2014).
147. Z. Ling *et al.*, The restoration of the vaginal microbiota after treatment for bacterial vaginosis with metronidazole or probiotics. *Microb Ecol* **65**, 773-780 (2013).
148. K. C. Anukam *et al.*, Clinical study comparing probiotic Lactobacillus GR-1 and RC-14 with metronidazole vaginal gel to treat symptomatic bacterial vaginosis. *Microbes Infect* **8**, 2772-2776 (2006).
149. C. Li *et al.*, Probiotics for the treatment of women with bacterial vaginosis: A systematic review and meta-analysis of randomized clinical trials. *Eur J Pharmacol* **864**, 172660 (2019).
150. C. R. Cohen *et al.*, Randomized Trial of Lactin-V to Prevent Recurrence of Bacterial Vaginosis. *N Engl J Med* **382**, 1906-1915 (2020).
151. T. Malikowski, S. Khanna, D. S. Pardi, Fecal microbiota transplantation for gastrointestinal disorders. *Curr Opin Gastroenterol* **33**, 8-13 (2017).
152. A. Lev-Sagie *et al.*, Vaginal microbiome transplantation in women with intractable bacterial vaginosis. *Nat Med* **25**, 1500-1504 (2019).
153. G. Wagner, R. J. Levin, L. Bohr, Diaphragm insertion increases human vaginal oxygen tension. *Am J Obstet Gynecol* **158**, 1040-1043 (1988).
154. G. Wagner, L. Bohr, P. Wagner, L. N. Petersen, Tampon-induced changes in vaginal oxygen and carbon dioxide tensions. *Am J Obstet Gynecol* **148**, 147-150 (1984).
155. W. Li *et al.*, Overcoming ABC transporter-mediated multidrug resistance: Molecular mechanisms and novel therapeutic drug strategies. *Drug Resist Updat* **27**, 14-29 (2016).
156. G. T. Spear *et al.*, Human alpha-amylase present in lower-genital-tract mucosal fluid processes glycogen to support vaginal colonization by Lactobacillus. *J Infect Dis* **210**, 1019-1028 (2014).
157. D. Nasioudis *et al.*, alpha-Amylase in Vaginal Fluid: Association With Conditions Favorable to Dominance of Lactobacillus. *Reprod Sci* **22**, 1393-1398 (2015).
158. E. R. Boskey, K. M. Telsch, K. J. Whaley, T. R. Moench, R. A. Cone, Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. *Infect Immun* **67**, 5170-5175 (1999).
159. E. R. Boskey, R. A. Cone, K. J. Whaley, T. R. Moench, Origins of vaginal acidity: high D/L lactate ratio is consistent with bacteria being the primary source. *Hum Reprod* **16**, 1809-1813 (2001).
160. J. B. Ewaschuk, J. M. Naylor, G. A. Zello, D-lactate in human and ruminant metabolism. *J Nutr* **135**, 1619-1625 (2005).

161. X. Zhou *et al.*, Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology* **150**, 2565-2573 (2004).
162. H. Mendes-Soares, H. Suzuki, R. J. Hickey, L. J. Forney, Comparative functional genomics of Lactobacillus spp. reveals possible mechanisms for specialization of vaginal lactobacilli to their environment. *J Bacteriol* **196**, 1458-1470 (2014).
163. S. S. Witkin, I. M. Linhares, Why do lactobacilli dominate the human vaginal microbiota? *BJOG* **124**, 606-611 (2017).
164. T. Latham *et al.*, Lactate, a product of glycolytic metabolism, inhibits histone deacetylase activity and promotes changes in gene expression. *Nucleic Acids Res* **40**, 4794-4803 (2012).
165. W. Wagner, W. M. Ciszewski, K. D. Kania, L- and D-lactate enhance DNA repair and modulate the resistance of cervical carcinoma cells to anticancer drugs via histone deacetylase inhibition and hydroxycarboxylic acid receptor 1 activation. *Cell Commun Signal* **13**, 36 (2015).
166. R. Stanek, R. E. Gain, D. D. Glover, B. Larsen, High performance ion exclusion chromatographic characterization of the vaginal organic acids in women with bacterial vaginosis. *Biomed Chromatogr* **6**, 231-235 (1992).
167. A. N. Chaudry *et al.*, Analysis of vaginal acetic acid in patients undergoing treatment for bacterial vaginosis. *J Clin Microbiol* **42**, 5170-5175 (2004).
168. J. R. Schwebke, H. L. Weiss, Interrelationships of bacterial vaginosis and cervical inflammation. *Sex Transm Dis* **29**, 59-64 (2002).
169. N. Singh *et al.*, Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128-139 (2014).
170. A. W. Sturm, Chemotaxis inhibition by Gardnerella vaginalis and succinate producing vaginal anaerobes: composition of vaginal discharge associated with G vaginalis. *Genitourin Med* **65**, 109-112 (1989).
171. D. E. O'Hanlon, B. R. Lanier, T. R. Moench, R. A. Cone, Cervicovaginal fluid and semen block the microbicidal activity of hydrogen peroxide produced by vaginal lactobacilli. *BMC Infect Dis* **10**, 120 (2010).
172. D. F. Muench *et al.*, Hydrogen peroxide-producing lactobacilli inhibit gonococci in vitro but not during experimental genital tract infection. *J Infect Dis* **199**, 1369-1378 (2009).
173. J. A. McGroarty, G. Reid, Detection of a Lactobacillus substance that inhibits Escherichia coli. *Can J Microbiol* **34**, 974-978 (1988).
174. S. Boris, C. Barbes, Role played by lactobacilli in controlling the population of vaginal pathogens. *Microbes Infect* **2**, 543-546 (2000).
175. S. E. Gelber, J. L. Aguilar, K. L. Lewis, A. J. Ratner, Functional and phylogenetic characterization of Vaginolysin, the human-specific cytolysin from Gardnerella vaginalis. *J Bacteriol* **190**, 3896-3903 (2008).
176. S. S. Olmsted, L. A. Meyn, L. C. Rohan, S. L. Hillier, Glycosidase and proteinase activity of anaerobic gram-negative bacteria isolated from women with bacterial vaginosis. *Sex Transm Dis* **30**, 257-261 (2003).

177. S. Cauci, J. F. Culhane, M. Di Santolo, K. McCollum, Among pregnant women with bacterial vaginosis, the hydrolytic enzymes sialidase and prolidase are positively associated with interleukin-1beta. *Am J Obstet Gynecol* **198**, 132 e131-137 (2008).
178. N. Cerca *et al.*, Comparative assessment of antibiotic susceptibility of coagulase-negative staphylococci in biofilm versus planktonic culture as assessed by bacterial enumeration or rapid XTT colorimetry. *J Antimicrob Chemother* **56**, 331-336 (2005).
179. L. Hardy, N. Cerca, V. Jespers, M. Vaneechoutte, T. Crucitti, Bacterial biofilms in the vagina. *Res Microbiol* **168**, 865-874 (2017).
180. H. C. Flemming, T. R. Neu, D. J. Wozniak, The EPS matrix: the "house of biofilm cells". *J Bacteriol* **189**, 7945-7947 (2007).
181. G. M. Teitzel, M. R. Parsek, Heavy metal resistance of biofilm and planktonic Pseudomonas aeruginosa. *Appl Environ Microbiol* **69**, 2313-2320 (2003).
182. M. C. Walters, 3rd, F. Roe, A. Bugnicourt, M. J. Franklin, P. S. Stewart, Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of Pseudomonas aeruginosa biofilms to ciprofloxacin and tobramycin. *Antimicrob Agents Chemother* **47**, 317-323 (2003).
183. J. Hirschfeld, Dynamic interactions of neutrophils and biofilms. *J Oral Microbiol* **6**, 26102 (2014).
184. L. Hardy *et al.*, A fruitful alliance: the synergy between Atopobium vaginae and Gardnerella vaginalis in bacterial vaginosis-associated biofilm. *Sex Transm Infect* **92**, 487-491 (2016).
185. B. W. Catlin, Gardnerella vaginalis: characteristics, clinical considerations, and controversies. *Clin Microbiol Rev* **5**, 213-237 (1992).
186. M. Pleckaityte, M. Janulaitiene, R. Lasickiene, A. Zvirbliene, Genetic and biochemical diversity of Gardnerella vaginalis strains isolated from women with bacterial vaginosis. *FEMS Immunol Med Microbiol* **65**, 69-77 (2012).
187. A. Nazli *et al.*, Exposure to HIV-1 directly impairs mucosal epithelial barrier integrity allowing microbial translocation. *PLoS Pathog* **6**, e1000852 (2010).
188. R. C. Chan, A. W. Bruce, G. Reid, Adherence of cervical, vaginal and distal urethral normal microbial flora to human uroepithelial cells and the inhibition of adherence of gram-negative uropathogens by competitive exclusion. *J Urol* **131**, 596-601 (1984).
189. E. Amabebe, D. O. C. Anumba, The Vaginal Microenvironment: The Physiologic Role of Lactobacilli. *Front Med (Lausanne)* **5**, 181 (2018).
190. S. S. Olmsted *et al.*, Diffusion of macromolecules and virus-like particles in human cervical mucus. *Biophys J* **81**, 1930-1937 (2001).
191. P. Levinson *et al.*, Levels of innate immune factors in genital fluids: association of alpha defensins and LL-37 with genital infections and increased HIV acquisition. *AIDS* **23**, 309-317 (2009).
192. A. Burgener *et al.*, Comprehensive proteomic study identifies serpin and cystatin antiproteases as novel correlates of HIV-1 resistance in the cervicovaginal mucosa of female sex workers. *J Proteome Res* **10**, 5139-5149 (2011).
193. N. J. Klein, Mannose-binding lectin: do we need it? *Mol Immunol* **42**, 919-924 (2005).
194. O. Babula, G. Lazdane, J. Kroica, W. J. Ledger, S. S. Witkin, Relation between recurrent vulvovaginal candidiasis, vaginal concentrations of mannose-binding lectin, and a

mannose-binding lectin gene polymorphism in Latvian women. *Clin Infect Dis* **37**, 733-737 (2003).
195. M. M. Markiewski, B. Nilsson, K. N. Ekdahl, T. E. Mollnes, J. D. Lambris, Complement and coagulation: strangers or partners in crime? *Trends Immunol* **28**, 184-192 (2007).
196. A. P. Sjoberg, L. A. Trouw, A. M. Blom, Complement activation and inhibition: a delicate balance. *Trends Immunol* **30**, 83-90 (2009).
197. B. Fingleton, Matrix metalloproteinases as regulators of inflammatory processes. *Biochim Biophys Acta Mol Cell Res* **1864**, 2036-2042 (2017).
198. S. Akira, S. Uematsu, O. Takeuchi, Pathogen recognition and innate immunity. *Cell* **124**, 783-801 (2006).
199. R. Medzhitov, Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819-826 (2007).
200. M. Yamamoto, K. Takeda, Current views of toll-like receptor signaling pathways. *Gastroenterol Res Pract* **2010**, 240365 (2010).
201. E. Latz *et al.*, TLR9 signals after translocating from the ER to CpG DNA in the lysosome. *Nat Immunol* **5**, 190-198 (2004).
202. T. Nishiya, E. Kajita, S. Miwa, A. L. Defranco, TLR3 and TLR7 are targeted to the same intracellular compartments by distinct regulatory elements. *J Biol Chem* **280**, 37107-37117 (2005).
203. P. A. Pioli *et al.*, Differential expression of Toll-like receptors 2 and 4 in tissues of the human female reproductive tract. *Infect Immun* **72**, 5799-5806 (2004).
204. K. M. Hart *et al.*, Functional expression of pattern recognition receptors in tissues of the human female reproductive tract. *J Reprod Immunol* **80**, 33-40 (2009).
205. M. Ghosh *et al.*, Pathogen recognition in the human female reproductive tract: expression of intracellular cytosolic sensors NOD1, NOD2, RIG-1, and MDA5 and response to HIV-1 and Neisseria gonorrhea. *Am J Reprod Immunol* **69**, 41-51 (2013).
206. M. Yamamoto, K. Takeda, S. Akira, TIR domain-containing adaptors define the specificity of TLR signaling. *Mol Immunol* **40**, 861-868 (2004).
207. T. Kawai, S. Akira, Innate immune recognition of viral infection. *Nat Immunol* **7**, 131-137 (2006).
208. J. Deese *et al.*, Injectable Progestin-Only Contraception is Associated With Increased Levels of Pro-Inflammatory Cytokines in the Female Genital Tract. *Am J Reprod Immunol* **74**, 357-367 (2015).
209. T. N. Mayadas, X. Cullere, C. A. Lowell, The multifaceted functions of neutrophils. *Annu Rev Pathol* **9**, 181-218 (2014).
210. F. D. Barr, C. Ochsenbauer, C. R. Wira, M. Rodriguez-Garcia, Neutrophil extracellular traps prevent HIV infection in the female genital tract. *Mucosal Immunol* **11**, 1420-1428 (2018).
211. J. P. Richardson *et al.*, Candidalysin Drives Epithelial Signaling, Neutrophil Recruitment, and Immunopathology at the Vaginal Mucosa. *Infect Immun* **86**, (2018).
212. A. R. Huppler *et al.*, Role of neutrophils in IL-17-dependent immunity to mucosal candidiasis. *J Immunol* **192**, 1745-1752 (2014).
213. F. Essig, K. Hunniger, S. Dietrich, M. T. Figge, O. Kurzai, Human neutrophils dump Candida glabrata after intracellular killing. *Fungal Genet Biol* **84**, 37-40 (2015).

214. S. Altmeier *et al.*, IL-1 Coordinates the Neutrophil Response to C. albicans in the Oral Mucosa. *PLoS Pathog* **12**, e1005882 (2016).
215. E. Mortaz, S. D. Alipoor, I. M. Adcock, S. Mumby, L. Koenderman, Update on Neutrophil Function in Severe Inflammation. *Front Immunol* **9**, 2171 (2018).
216. V. Krishnamurthy, S. Satish, M. G. Vimalambike, Cannonballs in Pap Smears: Double Whammy of Bacterial Vaginosis and Associated Infections. *Acta Cytol* **60**, 53-57 (2016).
217. J. Beghini *et al.*, Altered CD16 expression on vaginal neutrophils from women with vaginitis. *Eur J Obstet Gynecol Reprod Biol* **167**, 96-99 (2013).
218. J. Beghini, P. C. Giraldo, I. M. Linhares, W. J. Ledger, S. S. Witkin, Neutrophil Gelatinase-Associated Lipocalin Concentration in Vaginal Fluid: Relation to Bacterial Vaginosis and Vulvovaginal Candidiasis. *Reprod Sci* **22**, 964-968 (2015).
219. H. N. Simhan, S. N. Caritis, M. A. Krohn, S. L. Hillier, Elevated vaginal pH and neutrophils are associated strongly with early spontaneous preterm birth. *Am J Obstet Gynecol* **189**, 1150-1154 (2003).
220. S. Paul, G. Lal, The Molecular Mechanism of Natural Killer Cells Function and Its Importance in Cancer Immunotherapy. *Front Immunol* **8**, 1124 (2017).
221. M. A. Spinner *et al.*, GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* **123**, 809-821 (2014).
222. J. S. Orange, Human natural killer cell deficiencies and susceptibility to infection. *Microbes Infect* **4**, 1545-1558 (2002).
223. Y. Kinjo *et al.*, Functions of CD1d-Restricted Invariant Natural Killer T Cells in Antimicrobial Immunity and Potential Applications for Infection Control. *Front Immunol* **9**, 1266 (2018).
224. M. Lawand, J. Dechanet-Merville, M. C. Dieu-Nosjean, Key Features of Gamma-Delta T-Cell Subsets in Human Diseases and Their Immunotherapeutic Implications. *Front Immunol* **8**, 761 (2017).
225. R. J. Napier, E. J. Adams, M. C. Gold, D. M. Lewinsohn, The Role of Mucosal Associated Invariant T Cells in Antimicrobial Immunity. *Front Immunol* **6**, 344 (2015).
226. L. J. Howson, M. Salio, V. Cerundolo, MR1-Restricted Mucosal-Associated Invariant T Cells and Their Activation during Infectious Diseases. *Front Immunol* **6**, 303 (2015).
227. D. S. Leslie *et al.*, CD1-mediated gamma/delta T cell maturation of dendritic cells. *J Exp Med* **196**, 1575-1584 (2002).
228. J. Ismaili, V. Olislagers, R. Poupot, J. J. Fournie, M. Goldman, Human gamma delta T cells induce dendritic cell maturation. *Clin Immunol* **103**, 296-302 (2002).
229. T. B. Geijtenbeek, Y. van Kooyk, DC-SIGN: a novel HIV receptor on DCs that mediates HIV-1 transmission. *Curr Top Microbiol Immunol* **276**, 31-54 (2003).
230. R. N. Fichorova, Impact of T. vaginalis infection on innate immune responses and reproductive outcome. *J Reprod Immunol* **83**, 185-189 (2009).
231. D. J. Stieh *et al.*, Th17 Cells Are Preferentially Infected Very Early after Vaginal Transmission of SIV in Macaques. *Cell Host Microbe* **19**, 529-540 (2016).
232. Y. Alvarez *et al.*, Preferential HIV infection of CCR6+ Th17 cells is associated with higher levels of virus receptor expression and lack of CCR5 ligands. *J Virol* **87**, 10843-10854 (2013).

233. H. M. McGee *et al.*, IL-22 promotes fibroblast-mediated wound repair in the skin. *J Invest Dermatol* **133**, 1321-1329 (2013).
234. S. L. Gaffen, N. Hernandez-Santos, A. C. Peterson, IL-17 signaling in host defense against Candida albicans. *Immunol Res* **50**, 181-187 (2011).
235. C. M. Card, T. B. Ball, K. R. Fowke, Immune quiescence: a model of protection against HIV infection. *Retrovirology* **10**, 141 (2013).
236. M. D. Cooper, The early history of B cells. *Nat Rev Immunol* **15**, 191-197 (2015).
237. J. Mestecky, M. W. Russell, Induction of mucosal immune responses in the human genital tract. *FEMS Immunol Med Microbiol* **27**, 351-355 (2000).
238. J. Z. Zhou, S. S. Way, K. Chen, Immunology of the Uterine and Vaginal Mucosae. *Trends Immunol* **39**, 302-314 (2018).
239. M. R. Mendoza-Lopez *et al.*, CP30, a cysteine proteinase involved in Trichomonas vaginalis cytoadherence. *Infect Immun* **68**, 4907-4912 (2000).
240. R. Halter, J. Pohlner, T. F. Meyer, IgA protease of Neisseria gonorrhoeae: isolation and characterization of the gene and its extracellular product. *EMBO J* **3**, 1595-1601 (1984).
241. N. L. Anderson, N. G. Anderson, The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845-867 (2002).
242. W. M. Old *et al.*, Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* **4**, 1487-1502 (2005).
243. X. Wang *et al.*, Feature extraction in the analysis of proteomic mass spectra. *Proteomics* **6**, 2095-2100 (2006).
244. K. Rooijers *et al.*, An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* **12**, 6 (2011).
245. M. Kleiner *et al.*, Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* **8**, 1558 (2017).
246. H. Borgdorff *et al.*, Unique Insights in the Cervicovaginal Lactobacillus iners and L. crispatus Proteomes and Their Associations with Microbiota Dysbiosis. *PLoS One* **11**, e0150767 (2016).
247. A. S. Zevin *et al.*, Microbiome Composition and Function Drives Wound-Healing Impairment in the Female Genital Tract. *PLoS Pathog* **12**, e1005889 (2016).
248. K. D. Birse *et al.*, Genital Injury Signatures and Microbiome Alterations Associated With Depot Medroxyprogesterone Acetate Usage and Intravaginal Drying Practices. *J Infect Dis* **215**, 590-598 (2017).
249. E. S. Klaassens, W. M. de Vos, E. E. Vaughan, Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl Environ Microbiol* **73**, 1388-1392 (2007).
250. B. Brooks *et al.*, Strain-resolved microbial community proteomics reveals simultaneous aerobic and anaerobic function during gastrointestinal tract colonization of a preterm infant. *Front Microbiol* **6**, 654 (2015).
251. R. D. Zwittink *et al.*, Metaproteomics reveals functional differences in intestinal microbiota development of preterm infants. *Mol Cell Proteomics* **16**, 1610-1620 (2017).
252. T. Cerdo *et al.*, Gut microbial functional maturation and succession during human early life. *Environ Microbiol*, (2018).

253. P. Y. Lee, S. F. Chin, H. M. Neoh, R. Jamal, Metaproteomic analysis of human gut microbiota: where are we heading? *J Biomed Sci* **24**, 36 (2017).
254. M. Ferrer, V. A. Martins dos Santos, S. J. Ott, A. Moya, Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut Microbes* **5**, 64-70 (2014).
255. A. E. Perez-Cobas *et al.*, Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* **62**, 1591-1601 (2013).
256. Q. A. Karim *et al.*, Recruitment of high risk women for HIV prevention trials: baseline HIV prevalence and sexual behavior in the CAPRISA 004 tenofovir gel trial. *Trials* **12**, 67 (2011).
257. A. Nel *et al.*, HIV incidence remains high in KwaZulu-Natal, South Africa: evidence from three districts. *PLoS One* **7**, e35278 (2012).
258. Q. Abdool Karim *et al.*, Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* **329**, 1168-1174 (2010).
259. N. R. Klatt *et al.*, Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *Science* **356**, 938-945 (2017).
260. L. R. McKinnon *et al.*, Genital inflammation undermines the effectiveness of tenofovir gel in preventing HIV acquisition in women. *Nat Med* **24**, 491-496 (2018).
261. T. de Oliveira *et al.*, Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV* **4**, e41-e50 (2017).
262. K. M. Birse *et al.*, Unbiased proteomics analysis demonstrates significant variability in mucosal immune factor expression depending on the site and method of collection. *PLoS One* **8**, e79505 (2013).
263. K. Birse *et al.*, Molecular signatures of immune activation and epithelial barrier remodeling are enhanced during the luteal phase of the menstrual cycle: implications for HIV susceptibility. *J Virol*, (2015).
264. K. B. Arnold *et al.*, Increased levels of inflammatory cytokines in the female reproductive tract are associated with altered expression of proteases, mucosal barrier proteins, and an influx of HIV-susceptible target cells. *Mucosal Immunol* **9**, 194-205 (2016).
265. L. M. Romas *et al.*, A comparative proteomic analysis of the soluble immune factor environment of rectal and oral mucosa. *PLoS One* **9**, e100820 (2014).
266. K. D. Birse *et al.*, Non-Cationic Proteins Are Associated with HIV Neutralizing Activity in Genital Secretions of Female Sex Workers. *PLoS One* **10**, e0130404 (2015).
267. M. E. Kusek, M. A. Pazos, W. Pirzai, B. P. Hurley, In vitro coculture assay to assess pathogen induced neutrophil trans-epithelial migration. *J Vis Exp*, e50823 (2014).
268. J. L. Wilmer, M. I. Luster, Chemical induction of interleukin-8, a proinflammatory chemokine, in human epidermal keratinocyte cultures and its relation to cytogenetic toxicity. *Cell Biol Toxicol* **11**, 37-50 (1995).
269. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
270. B. Ma *et al.*, A comprehensive non-redundant gene catalog reveals extensive within-community intraspecies diversity in the human vagina. *Nat Commun* **11**, 940 (2020).
271. R. R. Spurbeck, C. G. Arvidson, Lactobacilli at the front line of defense against vaginally acquired infections. *Future Microbiol* **6**, 567-582 (2011).

272. M. I. Petrova, E. Lievens, S. Malik, N. Imholz, S. Lebeer, Lactobacillus species as biomarkers and agents that can promote various aspects of vaginal health. *Front Physiol* **6**, 81 (2015).
273. J. J. Schellenberg *et al.*, Gardnerella vaginalis Subgroups Defined by cpn60 Sequencing and Sialidase Activity in Isolates from Canada, Belgium and Kenya. *PLoS One* **11**, e0146510 (2016).
274. M. Aldunate *et al.*, Antimicrobial and immune modulatory effects of lactic acid and short chain fatty acids produced by vaginal microbiota associated with eubiosis and bacterial vaginosis. *Front Physiol* **6**, 164 (2015).
275. T. E. Taha *et al.*, Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. *AIDS* **12**, 1699-1706 (1998).
276. P. Mirmonsef, L. Krass, A. Landay, G. T. Spear, The role of bacterial vaginosis and trichomonas in HIV transmission across the female genital tract. *Curr HIV Res* **10**, 202-210 (2012).
277. M. C. Meltzer, R. A. Desmond, J. R. Schwebke, Association of Mobiluncus curtisii with recurrence of bacterial vaginosis. *Sex Transm Dis* **35**, 611-613 (2008).
278. J. A. McGregor, J. I. French, Bacterial vaginosis and preterm birth. *N Engl J Med* **334**, 1337-1338; author reply 1338-1339 (1996).
279. L. Manns-James, Bacterial vaginosis and preterm birth. *J Midwifery Womens Health* **56**, 575-583 (2011).
280. A. Buve, V. Jespers, T. Crucitti, R. N. Fichorova, The vaginal microbiota and susceptibility to HIV. *AIDS* **28**, 2333-2344 (2014).
281. M. A. Elovitz *et al.*, Cervicovaginal microbiota and local immune response modulate the risk of spontaneous preterm delivery. *Nat Commun* **10**, 1305 (2019).
282. C. Zhang *et al.*, The direct and indirect association of cervical microbiota with the risk of cervical intraepithelial neoplasia. *Cancer Med* **7**, 2172-2179 (2018).
283. P. Dobbler *et al.*, The vaginal microbial communities of healthy expectant Brazilian mothers and its correlation with the newborn's gut colonization. *World J Microbiol Biotechnol* **35**, 159 (2019).
284. D. L. Tabb *et al.*, Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* **9**, 761-776 (2010).
285. R. H. Deurenberg *et al.*, Reprint of "Application of next generation sequencing in clinical microbiology and infection prevention". *J Biotechnol* **250**, 2-10 (2017).
286. P. D. Schloss, The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6**, e1000844 (2010).
287. S. S. Witkin *et al.*, Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *MBio* **4**, (2013).
288. T. Ojala *et al.*, Genome sequence of Lactobacillus crispatus ST1. *J Bacteriol* **192**, 3547-3548 (2010).
289. M. T. France, H. Mendes-Soares, L. J. Forney, Genomic Comparisons of Lactobacillus crispatus and Lactobacillus iners Reveal Potential Ecological Drivers of Community Composition in the Vagina. *Appl Environ Microbiol* **82**, 7063-7073 (2016).

290. D. S. A. Goltsman *et al.*, Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome. *Genome Res* **28**, 1467-1480 (2018).
291. J. Bathke, A. Konzer, B. Remes, M. McIntosh, G. Klug, Comparative analyses of the variation of the transcriptome and proteome of Rhodobacter sphaeroides throughout growth. *BMC Genomics* **20**, 358 (2019).
292. T. Teng *et al.*, Comparative transcriptomic and proteomic analyses reveal upregulated expression of virulence and iron transport factors of Aeromonas hydrophila under iron limitation. *BMC Microbiol* **18**, 52 (2018).
293. J. A. Freiberg *et al.*, Global Analysis and Comparison of the Transcriptomes and Proteomes of Group A Streptococcus Biofilms. *mSystems* **1**, (2016).
294. N. Grassl *et al.*, Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Med* **8**, 44 (2016).
295. F. Fouhy *et al.*, A pilot study demonstrating the altered gut microbiota functionality in stable adults with Cystic Fibrosis. *Sci Rep* **7**, 6685 (2017).
296. C. A. Kolmeder *et al.*, Faecal Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and Limited Effects of a Probiotic Intervention in Adults. *PLoS One* **11**, e0153294 (2016).
297. G. L. Lorca, G. Font de Valdez, A. Ljungh, Characterization of the protein-synthesis dependent adaptive acid tolerance response in Lactobacillus acidophilus. *J Mol Microbiol Biotechnol* **4**, 525-532 (2002).
298. A. Ghosal, Importance of secreted bacterial RNA in bacterial-host interactions in the gut. *Microb Pathog* **104**, 161-163 (2017).
299. D. Simonov, S. Swift, C. Blenkiron, A. R. Phillips, Bacterial RNA as a signal to eukaryotic cells as part of the infection process. *Discoveries (Craiova)* **4**, e70 (2016).
300. P. C. Campos, M. T. Gomes, E. S. Guimaraes, G. Guimaraes, S. C. Oliveira, TLR7 and TLR3 Sense Brucella abortus RNA to Induce Proinflammatory Cytokine Production but They Are Dispensable for Host Control of Infection. *Front Immunol* **8**, 28 (2017).
301. M. A. Milillo *et al.*, B. abortus RNA is the component involved in the down-modulation of MHC-I expression on human monocytes via TLR8 and the EGFR pathway. *PLoS Pathog* **13**, e1006527 (2017).
302. S. A. Shetty, N. P. Marathe, V. Lanjekar, D. Ranade, Y. S. Shouche, Comparative genome analysis of Megasphaera sp. reveals niche specialization and its potential role in the human gut. *PLoS One* **8**, e79353 (2013).
303. S. Srinivasan *et al.*, Metabolic signatures of bacterial vaginosis. *MBio* **6**, (2015).
304. R. W. Hyman *et al.*, Microbes on the human vaginal epithelium. *Proc Natl Acad Sci U S A* **102**, 7952-7957 (2005).
305. W. Mulu, M. Yimer, Y. Zenebe, B. Abera, Common causes of vaginal infections and antibiotic susceptibility of aerobic bacterial isolates in women of reproductive age attending at Felegehiwot Referral Hospital, Ethiopia: a cross sectional study. *BMC Womens Health* **15**, 42 (2015).
306. A. K. Benson *et al.*, Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**, 18933-18938 (2010).

307. M. J. Bonder *et al.*, The effect of host genetics on the gut microbiome. *Nat Genet* **48**, 1407-1412 (2016).
308. J. Wang *et al.*, Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat Genet* **48**, 1396-1406 (2016).
309. J. K. Goodrich *et al.*, Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe* **19**, 731-743 (2016).
310. J. Si, H. J. You, J. Yu, J. Sung, G. Ko, Prevotella as a Hub for Vaginal Microbiota under the Influence of Host Genetics and Their Association with Obesity. *Cell Host Microbe* **21**, 97-105 (2017).
311. D. Rothschild *et al.*, Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210-215 (2018).
312. J. P. Burton, G. Reid, Evaluation of the bacterial vaginal flora of 20 postmenopausal women by direct (Nugent score) and molecular (polymerase chain reaction and denaturing gradient gel electrophoresis) techniques. *J Infect Dis* **186**, 1770-1780 (2002).
313. D. N. Fredricks, T. L. Fiedler, J. M. Marrazzo, Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* **353**, 1899-1911 (2005).
314. R. Verhelst *et al.*, 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* **5**, 61 (2005).
315. R. Tamrakar *et al.*, Association between Lactobacillus species and bacterial vaginosis-related bacteria, and bacterial vaginosis scores in pregnant Japanese women. *BMC Infect Dis* **7**, 128 (2007).
316. M. J. Ferris, J. Norori, M. Zozaya-Hinchliffe, D. H. Martin, Cultivation-independent analysis of changes in bacterial vaginosis flora following metronidazole treatment. *J Clin Microbiol* **45**, 1016-1018 (2007).
317. E. De Backer *et al.*, Quantitative determination by real-time PCR of four vaginal Lactobacillus species, Gardnerella vaginalis and Atopobium vaginae indicates an inverse relationship between L. gasseri and L. iners. *BMC Microbiol* **7**, 115 (2007).
318. N. A. El Aila *et al.*, Identification and genotyping of bacteria from paired vaginal and rectal samples from pregnant women indicates similarity between vaginal and rectal microflora. *BMC Infect Dis* **9**, 167 (2009).
319. R. Hummelen *et al.*, Deep sequencing of the vaginal microbiota of women with HIV. *PLoS One* **5**, e12078 (2010).
320. G. L. Santiago *et al.*, Longitudinal study of the dynamics of vaginal microflora during two consecutive menstrual cycles. *PLoS One* **6**, e28180 (2011).
321. S. Srinivasan *et al.*, Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS One* **7**, e37818 (2012).
322. S. Y. Doerflinger, A. L. Throop, M. M. Herbst-Kralovetz, Bacteria in the vaginal microbiome alter the innate immune response and barrier properties of the human vaginal epithelia in a species-specific manner. *J Infect Dis* **209**, 1989-1999 (2014).
323. L. Petricevic *et al.*, Characterisation of the vaginal Lactobacillus microbiota associated with preterm delivery. *Sci Rep* **4**, 5136 (2014).

324. J. Wertz, N. Isaacs-Cosgrove, C. Holzman, T. L. Marsh, Temporal Shifts in Microbial Communities in Nonpregnant African-American Women with and without Bacterial Vaginosis. *Interdiscip Perspect Infect Dis* **2008**, 181253 (2008).
325. B. J. Moncla, C. A. Chappell, B. M. Debo, L. A. Meyn, The Effects of Hormones and Vaginal Microflora on the Glycome of the Female Genital Tract: Cervical-Vaginal Fluid. *PLoS One* **11**, e0158687 (2016).
326. T. Kinoshita, H. Fukuzawa, T. Shimada, T. Saito, Y. Matsuda, Primary structure and expression of a gamete lytic enzyme in Chlamydomonas reinhardtii: similarity of functional domains to matrix metalloproteases. *Proc Natl Acad Sci U S A* **89**, 4693-4697 (1992).
327. J. Abe *et al.*, The transcriptional program of synchronous gametogenesis in Chlamydomonas reinhardtii. *Curr Genet* **46**, 304-315 (2004).
328. J. Ravel *et al.*, Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome* **1**, 29 (2013).
329. J. P. Brooks *et al.*, Changes in vaginal community state types reflect major shifts in the microbiome. *Microb Ecol Health Dis* **28**, 1303265 (2017).
330. M. Swamy *et al.*, Glucose and glutamine fuel protein O-GlcNAcylation to control T cell self-renewal and malignancy. *Nat Immunol* **17**, 712-720 (2016).
331. D. E. O'Hanlon, P. Gajer, R. M. Brotman, J. Ravel, Asymptomatic Bacterial Vaginosis Is Associated With Depletion of Mature Superficial Cells Shed From the Vaginal Epithelium. *Front Cell Infect Microbiol* **10**, 106 (2020).
332. E. Roselletti *et al.*, Apoptosis of vaginal epithelial cells in clinical samples from women with diagnosed bacterial vaginosis. *Sci Rep* **10**, 1978 (2020).
333. C. J. Yeoman *et al.*, A multi-omic systems-based approach reveals metabolic markers of bacterial vaginosis and insight into the disease. *PLoS One* **8**, e56111 (2013).
334. H. Borgdorff *et al.*, Cervicovaginal microbiome dysbiosis is associated with proteome changes related to alterations of the cervicovaginal mucosal barrier. *Mucosal Immunol* **9**, 621-633 (2016).
335. J. L. Patterson, A. Stull-Lane, P. H. Girerd, K. K. Jefferson, Analysis of adherence, biofilm formation and cytotoxicity suggests a greater virulence potential of Gardnerella vaginalis relative to other bacterial-vaginosis-associated anaerobes. *Microbiology* **156**, 392-399 (2010).
336. F. C. Los, T. M. Randis, R. V. Aroian, A. J. Ratner, Role of pore-forming toxins in bacterial infectious diseases. *Microbiol Mol Biol Rev* **77**, 173-207 (2013).
337. L. Gurcel, L. Abrami, S. Girardin, J. Tschopp, F. G. van der Goot, Caspase-1 activation of lipid metabolic pathways in response to bacterial pore-forming toxins promotes cell survival. *Cell* **126**, 1135-1145 (2006).
338. S. K. Cassidy, M. X. O'Riordan, More than a pore: the cellular response to cholesterol-dependent cytolysins. *Toxins (Basel)* **5**, 618-636 (2013).
339. T. M. Randis *et al.*, Vaginolysin drives epithelial ultrastructural responses to Gardnerella vaginalis. *Infect Immun* **81**, 4544-4550 (2013).
340. S. G. de Aquino *et al.*, Periodontal pathogens directly promote autoimmune experimental arthritis by inducing a TLR2- and IL-1-driven Th17 response. *J Immunol* **192**, 4103-4111 (2014).

341. A. Aroutcheva *et al.*, Defense factors of vaginal lactobacilli. *Am J Obstet Gynecol* **185**, 375-379 (2001).
342. A. A. Aroutcheva, J. A. Simoes, S. Faro, Antimicrobial protein produced by vaginal Lactobacillus acidophilus that inhibits Gardnerella vaginalis. *Infect Dis Obstet Gynecol* **9**, 33-39 (2001).
343. E. V. Valore, C. H. Park, S. L. Igreti, T. Ganz, Antimicrobial components of vaginal fluid. *Am J Obstet Gynecol* **187**, 561-568 (2002).
344. P. Mirmonsef *et al.*, Short-chain fatty acids induce pro-inflammatory cytokine production alone and in combination with toll-like receptor ligands. *Am J Reprod Immunol* **67**, 391-400 (2012).
345. E. K. Libby, K. E. Pascal, E. Mordechai, M. E. Adelson, J. P. Trama, Atopobium vaginae triggers an innate immune response in an in vitro model of bacterial vaginosis. *Microbes Infect* **10**, 439-446 (2008).
346. W. A. Rose, 2nd *et al.*, Commensal bacteria modulate innate immune responses of vaginal epithelial cell multilayer cultures. *PLoS One* **7**, e32728 (2012).
347. M. Havaux *et al.*, Vitamin B6 deficient plants display increased sensitivity to high light and photo-oxidative stress. *BMC Plant Biol* **9**, 130 (2009).
348. C. M. Dude, A. Saylany, A. Brown, M. Elovitz, L. Anton, Microbial supernatants from Mobiluncus mulieris, a bacteria strongly associated with spontaneous preterm birth, disrupts the cervical epithelial barrier through inflammatory and miRNA mediated mechanisms. *Anaerobe* **61**, 102127 (2020).
349. H. Moi, H. Fredlund, E. Tornqvist, D. Danielsson, Mobiluncus species in bacterial vaginosis: aspects of pathogenesis. *APMIS* **99**, 1049-1054 (1991).
350. C. R. Eade *et al.*, Identification and characterization of bacterial vaginosis-associated pathogens using a comprehensive cervical-vaginal epithelial coculture assay. *PLoS One* **7**, e50106 (2012).
351. G. V. Salamone *et al.*, Flagellin delays spontaneous human neutrophil apoptosis. *Lab Invest* **90**, 1049-1059 (2010).
352. M. M. Mihai *et al.*, Identification and phenotypic characterization of the most frequent bacterial etiologies in chronic skin ulcers. *Rom J Morphol Embryol* **55**, 1401-1408 (2014).
353. S. Roy *et al.*, Mixed-species biofilm compromises wound healing by disrupting epidermal barrier function. *J Pathol* **233**, 331-343 (2014).
354. K. S. Brandenburg *et al.*, Inhibition of Pseudomonas aeruginosa biofilm formation on wound dressings. *Wound Repair Regen* **23**, 842-854 (2015).
355. A. D. Hempstead, R. R. Isberg, Inhibition of host cell translation elongation by Legionella pneumophila blocks the host cell unfolded protein response. *Proc Natl Acad Sci U S A* **112**, E6790-6797 (2015).
356. J. M. Villalba, F. J. Alcain, Sirtuin activators and inhibitors. *Biofactors* **38**, 349-359 (2012).
357. K. G. Michel, R. P. Huijbregts, J. L. Gleason, H. E. Richter, Z. Hel, Effect of hormonal contraception on the function of plasmacytoid dendritic cells and distribution of immune cell populations in the female reproductive tract. *J Acquir Immune Defic Syndr* **68**, 511-518 (2015).
358. G. B. Lazenby, D. E. Soper, F. S. Nolte, Correlation of leukorrhea and Trichomonas vaginalis infection. *J Clin Microbiol* **51**, 2323-2327 (2013).

359. O. D. Jarrett *et al.*, T. vaginalis Infection Is Associated with Increased IL-8 and TNFr1 Levels but with the Absence of CD38 and HLADR Activation in the Cervix of ESN. *PLoS One* **10**, e0130146 (2015).
360. N. Branzk *et al.*, Neutrophils sense microbe size and selectively release neutrophil extracellular traps in response to large pathogens. *Nat Immunol* **15**, 1017-1025 (2014).
361. N. Whibley *et al.*, Delinking CARD9 and IL-17: CARD9 Protects against Candida tropicalis Infection through a TNF-alpha-Dependent, IL-17-Independent Mechanism. *J Immunol* **195**, 3781-3792 (2015).
362. K. A. Glass, S. J. Longley, J. M. Bliss, S. K. Shaw, Protection of Candida parapsilosis from neutrophil killing through internalization by human endothelial cells. *Virulence* **6**, 504-514 (2015).
363. R. N. Fichorova *et al.*, The villain team-up or how Trichomonas vaginalis and bacterial vaginosis alter innate immunity in concert. *Sex Transm Infect* **89**, 460-466 (2013).
364. C. Marconi, G. G. Donders, C. M. Parada, P. C. Giraldo, M. G. da Silva, Do Atopobium vaginae, Megasphaera sp. and Leptotrichia sp. change the local innate immune response and sialidase activity in bacterial vaginosis? *Sex Transm Infect* **89**, 167-173 (2013).
365. T. L. Cherpes *et al.*, Hormonal contraceptive use modulates the local inflammatory response to bacterial vaginosis. *Sex Transm Infect* **84**, 57-61 (2008).
366. P. Losikoff *et al.*, Genital tract interleukin-8 but not interleukin-1beta or interleukin-6 concentration is associated with bacterial vaginosis and its clearance in HIV-infected and HIV-uninfected women. *Infect Dis Obstet Gynecol* **2007**, 92307 (2007).
367. C. Mitchell, J. Balkus, K. Agnew, R. Lawler, J. Hitti, Changes in the vaginal microenvironment with metronidazole treatment for bacterial vaginosis in early pregnancy. *J Womens Health (Larchmt)* **18**, 1817-1824 (2009).
368. A. Rebbapragada *et al.*, Bacterial vaginosis in HIV-infected women induces reversible alterations in the cervical immune environment. *J Acquir Immune Defic Syndr* **49**, 520-522 (2008).
369. M. H. Yudin, D. V. Landers, L. Meyn, S. L. Hillier, Clinical and cervical cytokine response to treatment with oral or vaginal metronidazole for bacterial vaginosis during pregnancy: a randomized trial. *Obstet Gynecol* **102**, 527-534 (2003).
370. M. Natsui *et al.*, Selective depletion of neutrophils by a monoclonal antibody, RP-3, suppresses dextran sulphate sodium-induced colitis in rats. *Journal of gastroenterology and hepatology* **12**, 801-808 (1997).
371. M. G. Buell, M. C. Berin, Neutrophil-independence of the initiation of colonic injury. Comparison of results from three models of experimental colitis in the rat. *Digestive diseases and sciences* **39**, 2575-2588 (1994).
372. U. K. Gursoy, E. Kononen, N. Luukkonen, V. J. Uitto, Human neutrophil defensins and their effect on epithelial cells. *Journal of periodontology* **84**, 126-133 (2013).
373. E. Kolaczkowska, P. Kubes, Neutrophil recruitment and function in health and inflammation. *Nat Rev Immunol* **13**, 159-175 (2013).
374. D. A. Weber *et al.*, Neutrophil-derived JAML inhibits repair of intestinal epithelial injury during acute inflammation. *Mucosal Immunol* **7**, 1221-1232 (2014).
375. M. Somsouk *et al.*, Gut epithelial barrier and systemic inflammation during chronic HIV infection. *AIDS* **29**, 43-51 (2015).

376. B. Drescher, F. Bai, Neutrophil in viral infections, friend or foe? *Virus research* **171**, 1-7 (2013).
377. R. N. Fichorova *et al.*, Interleukin (IL)-1, IL-6, and IL-8 predict mucosal toxicity of vaginal microbicidal contraceptives. *Biol Reprod* **71**, 761-769 (2004).
378. L. Van Damme *et al.*, Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-1 transmission in female sex workers: a randomised controlled trial. *Lancet* **360**, 971-977 (2002).
379. L. Masson *et al.*, Genital Inflammation and the Risk of HIV Acquisition in Women. *Clin Infect Dis* **61**, 260-269 (2015).
380. W. Xiong, R. J. Giannone, M. J. Morowitz, J. F. Banfield, R. L. Hettich, Development of an enhanced metaproteomic approach for deepening the microbiome characterization of the human infant gut. *J Proteome Res* **14**, 133-141 (2015).
381. A. Bilbao *et al.*, Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **15**, 964-980 (2015).
382. C. A. Kolmeder *et al.*, Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS One* **7**, e29913 (2012).
383. B. A. Petriz, O. L. Franco, Metaproteomics as a Complementary Approach to Gut Microbiota in Health and Disease. *Front Chem* **5**, 4 (2017).
384. L. Noel-Romas *et al.*, Vaginal microbiome-hormonal contraceptive interactions associate with the mucosal proteome and HIV acquisition. *PLoS Pathog* **16**, e1009097 (2020).