CONVERSATIONS BETWEEN SYSTEMIC IMMUNITY AND THE GUT MICROBIOME IN EARLY LIFE

by

Nelly Amenyogbe

B.Sc., The University of British Columbia, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Experimental Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2019

© Nelly Amenyogbe, 2019
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Conversations between systemic immunity and the gut microbiome in early life

submitted by  Nelly Amenyogbe          in partial fulfillment of the requirements for

the degree of  Doctor of Philosophy

in  Experimental Medicine

Examining Committee:

Dr. William Mohn, Microbiology and Immunology  
Supervisor

Dr. Scott Tebbutt  
Supervisory Committee Member

Dr. Laura Parfrey  
University Examiner

Dr. Bruce Vallance  
University Examiner

Additional Supervisory Committee Members:

Dr. Laura Sly  
Supervisory Committee Member

Dr. Soren Gantt  
Supervisory Committee Member
Abstract

Therapies targeting the microbiome hold great promise to improve the wellbeing of children. Their success will depend on understanding host-microbiome conversations. The gut microbiome differs between geographically diverse populations. Does this contribute to distinct immune phenotypes? Does the same perturbation impact these diverse populations in the same way? How does the composition of the fungal microbiome compare to bacterial? What tools are needed to listen to these conversations where it matters the most, in vulnerable populations living in low-resource settings? These questions can be ethically asked in humans.

To find answers to these questions, I assessed the relationship between systemic immune responses and the gut microbiome and compared bacterial to fungal colonization in geographically diverse child cohorts, hypothesizing that population-specific microbiomes correlated to distinct systemic innate immune phenotypes. To test this, we profiled the bacterial microbiomes of 2-year-old children from Belgium, Canada, Ecuador, and South Africa and measured their cytokine responses to innate stimuli. Certain immune differences between the cohorts correlated with the abundance of select bacterial taxa. Splenocytes of germ-free mice inoculated with human stools responded to stimulation in a manner consistent with the corresponding human donors, indicating that microbiomes can direct systemic innate immunity.

I hypothesized that immune responses of HIV-exposed uninfected (HEU) and HIV-unexposed uninfected (HUU) children within each site differed in population-specific ways and that these differences correlated to distinct microbomes. While HEU were distinguished from HUU children by immune responses and microbiome composition in population-specific ways, differences in the microbiome did not correlate with altered immune phenotypes.
I compared Bacterial vs. fungal composition over the first 5 years of life in a rural African population and found that both changed in kingdom-specific ways, suggesting they were shaped by separate selective forces. To overcome the hurdle of investigating newborns, I designed and implemented an experimental protocol in a low-resource setting that allows extraction of ‘big data’ out of the very small samples.

Taken together, these findings suggest that therapies targeting the microbiome must consider population differences, and that placing newborns in low-resource settings at the forefront of our research is not only warranted, but feasible.
Lay Summary

Our gut microbiome includes the bacterial, fungal, and viral communities living along our gastrointestinal tract. It influences our immune system, and is a crucial driver of early life health and disease. While therapies targeting the gut microbiome hold great promise to improve the wellbeing of children around the world, its relationship between host immunity must first be better understood. In this work, I demonstrated that population-specific microbiomes contribute to differences in immune responses by identifying correlations between the microbiome and immunity across diverse child populations and provided evidence such relationships can be causal using a mouse model. I also compared dynamics of less studied fungal colonization to bacterial in childhood. Lastly, I developed the experimental protocol for newborns in low resource settings, whose immune systems have thus far been understudied due to technical challenges, paving the way for a new era of host-microbiome human research.
Preface

All experimental procedures involving humans were done in accordance with ethical principles in the Declaration of Helsinki and approved by the University of British Columbia ethics board under protocol number H11-01423 for Chapters 2, 3 and 4, H11-01947 for the work involving HIV-exposed uninfected South African infants in Chapter 3. Work presented in Chapter 4 was conducted under protocol number H14-03369 for newborns from Guinea-Bissau, and H16-02409 for newborns from both the Gambia and Papua New Guinea. Animal studies conducted in Chapter 2 were approved by the University of British Columbia Ethics Board under protocol number A13-0265.

The human cohorts presented in Chapters 2 and 3 were designed by Dr. Tobias Kollmann, and blood sampling protocols optimized by Dr. Kinga Smolen and Dr. Bing Cai. Dr. Kinga Smolen collected the biospecimens. Dr. Bing Cai and Darren Sutherland conducted the assays under Dr. Kinga Smolen’s supervision. The immune data presented in Chapter 2 has been previously published independent of my involvement: Smolen KK, Ruck CE, Fortuno ES, 3rd, Ho K, Dimitriu P, Mohn WW, et al. Pattern recognition receptor-mediated cytokine response in infants across 4 continents. J Allergy Clin Immunol 2014; 133:818-26.e4.

For Chapters 2 and 3, I processed all stool specimens for microbiome analysis. Dr. Gregory Gloor provided 16S amplicon sequencing data, and Dr. Pedro Dimitriu processed the sequencing data prior to statistical analysis. The mouse experiments presented in Chapter 2 were done in collaboration with Eric Brown who reconstituted germ-free mice with human feces. I euthanized the animals at the end of the experiment and conducted the immune and microbiome analyses. Eric Brown supplied the data on urine lactulose:mannitol ratio. I chose the analytical
strategies and conducted all the analyses presented Chapters 2 and 3 and wrote the manuscripts under supervision of Dr. William Mohn and Dr. Pedro Dimitriu.

I participated in the cohort design for Chapter 4, which was done as a part of a broader collaboration involving Dr. Pinaki Panigrahi, Dr. Tobias Kollmann, and a team of scientists from the Kintampo Health Research Centre led by Dr. Seth Qwusu-Agyei. I chose the approach to sample sequencing with Dr. William Mohn, processed the breast milk specimens prior to amplicon sequencing, and facilitated multiplexed extraction of stool specimens. Multiplexed extraction and amplicon sequencing was done by Microbiome Insights. I conducted sequencing data binning under the supervision of Hilary Leung. I chose and conducted all analyses presented in Chapter 4, conducted the qPCR assay, analysed the data, and wrote the manuscript.

I participated in the study design for the Bandim Health Project (BHP; Guinea-Bissau) presented in Chapter 5, which was done as a part of a broader collaboration involving Dr. Christine Stabell Benn, Dr. Morten Bjerregaard-Andersen, and Dr. Tobias Kollmann. I designed the study protocol with input from Dr. Tobias Kollmann, Dr. Scott Tebbutt, and Dr. Leonard Foster. The Bissau study ran over a two-year period, and several BHP staff were involved in sample processing during that time. This included Guinean nurses Apala Sanca and Abdalha Cande who conducted the newborn blood draws, Gabriel Gomez who recruited the study participants and translated study materials to the local language, and Carlitos Monteiro who transported the biospecimens to the Gambia for long-term storage several times over the study period. Danish students Kristina Lindberg Larsen and Christian Golding supervised the study and processed the samples for the first year and a half of the study period, with help from Hannah Frankel and Annemette Anderson. Dr. Frederick Schaltz-Buchholzer provided supervision and study database monitoring throughout the course of the study. I designed the
sampling protocol, supervised the study from overseas and over four site visits to Guinea-Bissau. I accrued and transported all study supplies as airline luggage over the course of the site visits, including two microcentrifuges, one microscope, and several micropipettes. At the University of British Columbia, I tested the extraction protocols presented for sample quality mid-way through the study before the samples from the main set were selected, and Dr. Rym Ben Othman, and Joseph Huang processed the blood specimens presented in the chapter. I conducted all analyses and wrote the text presented.

The work presented in Chapter 5 from the Gambia and Papua New Guinea (PNG) is a part of the large international Expanded Program of Immunization Consortium (EPIC), led by Dr. Tobias Kollmann, Dr. Beate Kampmann, and Dr. Ofer Levy. The protocol I designed for Guinea Bissau was used for both Gambia and PNG, and modifications made with input from various EPIC members. A version of the sample processing flowchart presented in Chapter 5 (Figure 5.1) was prepared by Dr. Rym Ben Othman. I did not participate in sample collection. The flow cytometry data was generated by Dr. Bing Cai, and flow cytometry panel designed by Dr. Tobias Kollmann, Dr. Gemma Moncunil, and Dr. Bing Cai. I processed the data with Dr. Rym Ben Othman. I also chose the analytical strategy, conducted all analyses, and wrote the text. A version of the work presented has been previously published: Lee AH, Shannon CP, Amenyogbe N, Bennike TB, Diray-Arce J, Idoko OT, et al. Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. Nat Commun 2019; 10:1092. I was a co-first author of this paper.

Data chapters 2-4 are manuscripts in preparation. I conducted all analyses prepared all figures in this thesis except for Figure 5.1, as described above. I prepared a version of Figure 6.1 for a published review, and modified it to illustrate the final conclusions of this thesis. The
original published figure can be found here: Amenyogbe N, Kollmann TR, Ben-Othman R. Early-Life Host-Microbiome Interphase: The Key Frontier for Immune Development. Front Pediatr 2017; 5:111.
Table of Contents

Abstract ........................................................................................................................................ iii
Lay Summary .................................................................................................................................. v
Preface ........................................................................................................................................ vi
Table of Contents ........................................................................................................................ x
List of Tables .................................................................................................................................. xv
List of Figures ............................................................................................................................... xvi
List of Abbreviations .................................................................................................................... xviii
Acknowledgements ...................................................................................................................... xxiii
Dedication ....................................................................................................................................... xxv

Chapter 1: Introduction ..................................................................................................................... 1

1.1 Convering with microbes: systemic innate immunity ............................................................... 1

1.2 Innate immune responses to the microbiome .......................................................................... 2

1.2.1 Pattern recognition receptors sense the microbiome .......................................................... 2

1.2.2 Toll-like and Nod-like receptors ......................................................................................... 2

1.2.3 Cytokine responses to PRR stimulation ............................................................................. 4

1.2.4 Innate immune responses to commensals .......................................................................... 5

1.3 The gut microbiome modulates systemic immune responses ................................................. 6

1.4 Early infancy is a critical window to establish host-microbiome homeostasis ....................... 7

1.4.1 The bacterial microbiome changes across the age spectrum ............................................. 8

1.4.2 Host-microbiome interactions in infancy .......................................................................... 9

1.5 Systemic immunity and gut microbiomes across diverse populations .................................. 11
Chapter 1: The infant gut microbiome

1.5.1 Gut microbiome across geographically distinct populations .................................. 12
1.5.2 Systemic immune responses across geographically distinct populations ........... 13
1.6 The fungal microbiome in early life ........................................................................... 14
1.6.1 The fungal microbiome shapes systemic immunity ................................................ 14
1.6.2 The fungal microbiome across the age spectrum ................................................... 15
1.6.3 Interactions between bacterial and fungal microbiomes ..................................... 15
1.7 The breast milk microbiome ..................................................................................... 16
1.8 HIV-exposed uninfected infants .............................................................................. 17
1.8.1 HEU infants are more susceptible to severe infection in early life ..................... 18
1.8.2 Altered systemic innate immunity in HEU infants .............................................. 19
1.8.3 Altered gut microbiomes in HEU infants ............................................................ 20
1.9 Rationale .................................................................................................................. 21

Chapter 2: Biogeography shapes the relationship between the infant gut microbiome and innate immune responses .......................................................................................... 24

2.1 Introduction ............................................................................................................. 24
2.2 Materials and Methods .......................................................................................... 25
2.3 Results .................................................................................................................... 38
  2.3.1 Cohort characteristics ......................................................................................... 38
  2.3.2 Microbiome composition across four cohorts .................................................... 40
  2.3.3 Alpha diversity and host factors within each cohort ......................................... 43
  2.3.4 Microbiome taxonomic composition and differential abundance .................... 45
  2.3.5 Immune responses associated with each cohort .............................................. 47
  2.3.6 Correlations between TLR responsiveness and the gut microbiome .............. 50
2.3.6.1 Correlations across cohorts ................................................................. 50
2.3.6.2 Correlations within cohorts ................................................................. 52
2.3.7 Fecal transplant affects immune phenotype of germ-free mice .................. 54
2.4 Discussion ................................................................................................. 57

Chapter 3: Innate immune responses and gut microbiomes distinguish HIV-exposed from unexposed infants in a population-specific manner ............................................................................. 65

3.1 Introduction ............................................................................................... 65
3.2 Methods .................................................................................................... 66
3.3 Results ...................................................................................................... 74
  3.3.1 Cohort characteristics ........................................................................... 74
  3.3.2 Innate immune differences between HEU and HUU control children .......... 81
    3.3.2.1 Overall cytokine responses to PRR agonists ........................................ 81
    3.3.2.2 Differences in cytokine responses to PRR agonists ................................ 84
  3.3.3 Microbiome differences between HEU and HUU control children ............. 87
    3.3.3.1 Alpha and beta diversity .................................................................... 87
    3.3.3.2 OTU-level differences between HEU and HUU control children ........... 88
  3.3.4 Cytokine-microbiome relationships in HEU and HUU control children ....... 91
  3.4 Discussion .............................................................................................. 93

Chapter 4: Dramatic changes in bacterial and fungal community dynamics over the first five years of life in rural Ghana ........................................................................................................... 98

4.1 Introduction .............................................................................................. 98
4.2 Material and Methods ............................................................................... 99
4.3 Results ..................................................................................................... 106
Chapter 4: The Human Microbiome

4.3.1 16S amplicon sequencing summary .............................................................. 106
4.3.2 ITS2 amplicon sequencing summary .............................................................. 107
4.3.3 Bacterial and fungal alpha diversity across the age spectrum ....................... 109
4.3.4 Bacterial and fungal communities across the age spectrum ....................... 110
4.3.5 Maternal bacterial microbiome ................................................................. 112
4.3.6 Bacterial and fungal communities in mother’s breast milk and their newborns stool ................................................................. 114
4.3.7 Occurrence of Lactobacillus plantarum ....................................................... 116
4.4 Discussion ........................................................................................................... 116

Chapter 5: Small Sample Big Data: Towards analysis of the human immune system in the newborn period in low resource settings ................................................................. 121
5.1 Introduction ........................................................................................................ 121
5.2 Materials and Methods ................................................................................... 125
5.3 Results ............................................................................................................... 131
5.3.1 The small sample processing protocol was successfully implemented in a low-resource setting ......................................................................................... 131
5.3.2 BCGIMED-immune results ......................................................................... 133
5.3.3 Flow cytometric and plasma cytokine analysis of EPIC-Gambia newborns reveals a dynamic early-life trajectory ................................................................. 135
5.3.4 Indexing allows for identification of unified immune trajectories across independent newborn cohorts .................................................................................................................... 138
5.4 Discussion ........................................................................................................... 141

Chapter 6: Conclusion .............................................................................................. 144
6.1 Biogeography shapes the host immune-microbiome relationship ........................................ 144
6.2 Altered immune phenotypes in HEU infants is population-specific and not related to gut dysbiosis .......................................................................................................................... 145
6.3 Bacterial and fungal community dynamics are kingdom-specific ........................................ 146
6.4 Studying newborn immunity is feasible where it is most neglected .................................. 146
6.5 The way forward .................................................................................................................. 148

References .................................................................................................................................. 151

Appendix: BCGIMED-Immune sub-study materials ................................................................. 167
A.1 Standard operating procedure ............................................................................................ 167
A.2 Sample processing form: English Version .......................................................................... 179
A.3 Sample processing form: Portuguese version ................................................................... 181
A.4 Study explanation: English version .................................................................................... 183
A.5 Study explanation: Portuguese version ............................................................................... 184
List of Tables

Table 2.1 Cohort characteristics ........................................................................................................... 39
Table 2.2 Ethnic background across cohorts ........................................................................................ 40
Table 3.1 Belgian cohort characteristics ............................................................................................ 76
Table 3.2 Canadian cohort characteristics .......................................................................................... 77
Table 3.3 South African cohort characteristics .................................................................................... 78
Table 3.4 Ethnic background of Belgian HEU and HUU control children ............................................. 79
Table 3.5 Ethnic background of Canadian HEU and HUU control children ......................................... 79
Table 3.6 Ethnic background of South African HEU and HUU control children ................................. 79
Table 3.7 Initiation of maternal antiretroviral prophylaxis/therapy for HIV-positive mothers .................. 80
Table 3.8 Recruitment periods for HEU and HUU control children .................................................... 80
Table 4.1 Participant recruitment summary .......................................................................................... 107
Table 4.2 Occurrence of *Lactobacillus planarum* in newborn stools ................................................ 116
Table 5.1 Average processing times for selected BCGIMED samples from time of blood draw ......... 135
Table 5.2 Sample sizes for the Gambia-EPIC study ............................................................................ 135
Table 5.3 Sample sizes for the PNG-EPIC study ................................................................................. 138
**List of Figures**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Microbiome alpha and beta diversity across cohorts</td>
<td>42</td>
</tr>
<tr>
<td>2.2</td>
<td>Host factors associated with alpha diversity</td>
<td>44</td>
</tr>
<tr>
<td>2.3</td>
<td>Taxonomic differences across cohorts</td>
<td>46</td>
</tr>
<tr>
<td>2.4</td>
<td>Cytokine signatures across Belgian, Canadian, and Ecuadorean cohorts</td>
<td>47</td>
</tr>
<tr>
<td>2.5</td>
<td>Discriminatory cytokine features for Belgian and Canadian cohorts</td>
<td>49</td>
</tr>
<tr>
<td>2.6</td>
<td>Cytokine-microbiome correlations among Belgian, Canadian, and Ecuadorean children</td>
<td>50</td>
</tr>
<tr>
<td>2.7</td>
<td>Heatmap and select cytokine-microbiome correlations among Belgian, Canadian, and Ecuadorean children</td>
<td>51</td>
</tr>
<tr>
<td>2.8</td>
<td>Cohort-specific cytokine-microbiome relationships</td>
<td>53</td>
</tr>
<tr>
<td>2.9</td>
<td>Germ-free mouse models recapitulated immune differences between Canadian and South African children</td>
<td>55</td>
</tr>
<tr>
<td>2.10</td>
<td>Differentially abundant OTUs between mice gavaged with Canadian vs. South African feces</td>
<td>57</td>
</tr>
<tr>
<td>3.1</td>
<td>Recruitment periods for HEU and HUU control children</td>
<td>81</td>
</tr>
<tr>
<td>3.2</td>
<td>Principal components analysis discriminates cytokine responses to PRR stimulation between HEU and control infants</td>
<td>82</td>
</tr>
<tr>
<td>3.3</td>
<td>Principal component analysis highlighting cytokine data clustering</td>
<td>83</td>
</tr>
<tr>
<td>3.4</td>
<td>Principal component analysis shows sample clustering by HIV exposure for each ligand individually</td>
<td>84</td>
</tr>
</tbody>
</table>
Figure 3.5  Linear regression identified cytokine responses significantly affected by HIV exposure in a site-dependent manner ................................................................. 85
Figure 3.6  Whole blood cellular proportions of immune cells in HEU and HUU control children ........................................................................................................................................ 86
Figure 3.7  Canadian child intracellular production of IL-12p40 ........................................ 87
Figure 3.8  Microbiome alpha and beta diversity in HEU and HUU control children ........ 87
Figure 3.9  Beta diversity of HEU and HUU control children from each site separately .... 88
Figure 3.10  OTU-level differences between HEU and HUU control children ................... 90
Figure 3.11  Maximum classification error rate for discriminating HEU from HUU control children using single (sPLS-DA) or integrated (DIABLO) datasets ............................................................... 92
Figure 4.1  Amplicon sequencing coverage for bacterial and fungal communities .......... 108
Figure 4.2  Bacterial and fungal alpha diversity across the age spectrum ....................... 109
Figure 4.3  Bacterial and Fungal communities across the age spectrum ....................... 111
Figure 4.4  Postpartum maternal and newborn infant microbiome ................................ 113
Figure 4.5  Breast milk Bacterial and Fungal community composition ......................... 115
Figure 5.1  Sample processing schemes for BCGIMED and Gambia studies .................... 132
Figure 5.2  Sample collection summary for the BCGIMED study .................................. 134
Figure 5.3  Indexing cellular and soluble immune markers revealed developmental progression over the first week of life ........................................................................................................ 137
Figure 5.4  Indexing cellular and soluble markers allowed for agreement of biological trajectories from and independent validation cohort while still identifying cohort-specific differences ............................. 140
Figure 6.1  An overview of systemic immune-microbiome conversations in early life ...... 147
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>AP-1</td>
<td>Activator protein 1</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral therapy</td>
</tr>
<tr>
<td>ARV</td>
<td>Antiretroviral</td>
</tr>
<tr>
<td><em>B. fragilis</em></td>
<td><em>Bacteroides fragilis</em></td>
</tr>
<tr>
<td>BCG</td>
<td>Bacille Calmette-Guerin</td>
</tr>
<tr>
<td>BCGIMED</td>
<td>BCG-Immediate</td>
</tr>
<tr>
<td>BHP</td>
<td>Bandim Health Project</td>
</tr>
<tr>
<td>BLG</td>
<td>Belgium</td>
</tr>
<tr>
<td>Blimp1</td>
<td>B lymphocyte induced differentiation protein 1</td>
</tr>
<tr>
<td>bm</td>
<td>Breast milk</td>
</tr>
<tr>
<td><em>C. albicans</em></td>
<td><em>Candida albicans</em></td>
</tr>
<tr>
<td><em>C. perfringens</em></td>
<td><em>Clostridium perfringens</em></td>
</tr>
<tr>
<td>CAD</td>
<td>Canada</td>
</tr>
<tr>
<td>CCL</td>
<td>Chemokine C-C motif ligand</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>CD4</td>
<td>Cluster of differentiation 4</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony-forming units</td>
</tr>
<tr>
<td>CLARP</td>
<td>Caspase-like apoptosis regulatory protein</td>
</tr>
<tr>
<td>CLR</td>
<td>C-type lectin receptor</td>
</tr>
<tr>
<td>CS</td>
<td>Chi-squared test</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CXCL</td>
<td>C-X-C motif chemokine</td>
</tr>
<tr>
<td>DAT</td>
<td>Desaminotyrosine</td>
</tr>
<tr>
<td>DIABLO</td>
<td>Data Integration Analysis for Biomarker discovery using Latent variable approaches</td>
</tr>
<tr>
<td>DMM</td>
<td>Dirichlet multinomial mixtures</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DOL</td>
<td>Day of life</td>
</tr>
<tr>
<td>dsRNA</td>
<td>Double-stranded ribonucleic acid</td>
</tr>
<tr>
<td>ECD</td>
<td>Ecuador</td>
</tr>
<tr>
<td>EPIC</td>
<td>Expanded Program for Immunization Consortium</td>
</tr>
<tr>
<td>FCS</td>
<td>Flow cytometry staining</td>
</tr>
<tr>
<td>FE</td>
<td>Fisher’s exact test</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>GB</td>
<td>Guinea Bissau</td>
</tr>
<tr>
<td>GBS</td>
<td>Group B Streptococcus</td>
</tr>
<tr>
<td>GCSF</td>
<td>Granulocyte colony stimulating factor</td>
</tr>
<tr>
<td>GDM</td>
<td>Gestational diabetes mellitus</td>
</tr>
<tr>
<td>HAZ</td>
<td>Height-for-age Z-score</td>
</tr>
<tr>
<td>HEU</td>
<td>HIV-exposed uninfected</td>
</tr>
<tr>
<td>HIPC</td>
<td>Human Immune Project Consortium</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>HMO</td>
<td>Human milk oligosaccharide</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>HNSM</td>
<td>Hospital Nacional Simao Mendez</td>
</tr>
<tr>
<td>hrs</td>
<td>hours</td>
</tr>
<tr>
<td>HUU</td>
<td>HIV-unexposed uninfected</td>
</tr>
<tr>
<td>IEC</td>
<td>Intestinal epithelial cell</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IFNγ</td>
<td>Interferon gamma</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IP-10</td>
<td>Interferon-stimulated protein 10</td>
</tr>
<tr>
<td>IRAK</td>
<td>Interleukin-1 receptor-associated kinase</td>
</tr>
<tr>
<td>IRF</td>
<td>Interferon regulatory factor</td>
</tr>
<tr>
<td>ITS</td>
<td>Intergenic spacer region</td>
</tr>
<tr>
<td>KHRC</td>
<td>Kintampo Health Research Centre</td>
</tr>
<tr>
<td>KW</td>
<td>Kruskal-Wallis H-test</td>
</tr>
<tr>
<td>L. plantarum</td>
<td>Lactobacillus plantarum</td>
</tr>
<tr>
<td>LAB</td>
<td>Lactic acid bacteria</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>LRTI</td>
<td>Lower respiratory tract infection</td>
</tr>
<tr>
<td>MAMP</td>
<td>Microbe-associated molecular pattern</td>
</tr>
<tr>
<td>MBL</td>
<td>Mannose-binding lectin</td>
</tr>
<tr>
<td>MDA5</td>
<td>Melanoma differentiation-associated protein</td>
</tr>
<tr>
<td>MDC</td>
<td>Macrophage derived chemokine</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
</tbody>
</table>
min minute
MINCLE Macrophage inducible calcium-dependent lectin receptor
MIP Macrophage inflammatory protein
miR Micro RNA
mL millilitre
MOL Month of life
MyD88 Myeloid differentiation primary response 88
NEC Necrotizing enterocolitis
NF-κB Nuclear factor kappa-light-chain-enhancer of activated B cells
ng nanogram
NK Natural Killer
NKT Natural Killer T cell
NLR Nucleotide-binding oligomerization domain-like-receptor
NOD Nucleotide-binding oligomerization domain
OPV Oral polio vaccine
PAM Peptidyl-glycine alpha-amidating monooxygenase
pDC Plasmacytoid dendritic cell
PERMANOVA Permutational Analysis of Variance
pM picomolar
PNG Papua New Guinea
PRR Pattern recognition receptor
PSA Polysaccharide A
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSBI</td>
<td>Possible serious bacterial infection</td>
</tr>
<tr>
<td>QA</td>
<td>Quality assurance</td>
</tr>
<tr>
<td>QC</td>
<td>Quality control</td>
</tr>
<tr>
<td>RICK</td>
<td>RIP-like interacting CLARP kinase</td>
</tr>
<tr>
<td>RIG</td>
<td>Retinoic acid-inducible gene</td>
</tr>
<tr>
<td>RIP</td>
<td>Receptor interacting protein</td>
</tr>
<tr>
<td>RLR</td>
<td>Retinoic acid-inducible gene-like receptor</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SAF</td>
<td>South Africa</td>
</tr>
<tr>
<td>SFB</td>
<td>Segmented filamentous bacteria</td>
</tr>
<tr>
<td>sIgA</td>
<td>Secretory immunoglobulin A</td>
</tr>
<tr>
<td>ssRNA</td>
<td>Single-stranded ribonucleic acid</td>
</tr>
<tr>
<td>Th</td>
<td>T-helper</td>
</tr>
<tr>
<td>TIR</td>
<td>Toll/interleukin-1 receptor</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-like receptor</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumor necrosis factor alpha</td>
</tr>
<tr>
<td>TRIF</td>
<td>TIR-domain containing adapter-inducing interferon-β</td>
</tr>
<tr>
<td>TT</td>
<td>Two-tailed Student’s t-test</td>
</tr>
<tr>
<td>WAZ</td>
<td>Weight-for-age Z-score</td>
</tr>
<tr>
<td>WLZ</td>
<td>Weight-for-length Z-score</td>
</tr>
<tr>
<td>WT</td>
<td>Wilcoxon rank-sum test</td>
</tr>
<tr>
<td>µL</td>
<td>microlitre</td>
</tr>
</tbody>
</table>
Acknowledgements

First and foremost, I thank my supervisor Dr. William Mohn. Even before conducting my graduate work under his supervision, I first learned to appreciate the microbiome as a complex community from his microbial ecology lectures in my undergraduate years. I still have the lecture notes. The patience, support, and advice I received during my graduate years grew my appreciation for the field even more.

I extend my gratitude to my supervisory committee, Dr. Laura Sly, Dr. Scott Tebbutt, and Dr. Soren Gantt. They have taught me to be more critical of myself and careful in drawing my conclusions – likely one of the most important skills I have worked towards (and am still working on) in my scientific training.

The work presented here was in part supported by the BC Children’s Hospital Foundation, the National Institute of Allergy and Infectious Diseases, NIH, [N01 AI50023], and the Canadian Institutes of Health Research (CIHR) CMF-108029, as well as CIHR Project Grant 148781. I thank the funding agencies for their support.

I began this work here with very little knowledge of statistical tests and programs used throughout this thesis. For their guidance and patience (starting with helping me load my datasets into statistical programs) I thank Dr. Pedro Dimitriu and Casey Shannon. They made the intimidating world of statistical biology exciting and accessible.

The studies presented throughout relied on several human cohorts across eight countries and relied on the willingness of the mothers of infants whose samples made this work possible to participate in scientific research. I thank them for their time.

I acknowledge Dr. Kinga Smolen, whose dedication to her work created the highly standardized immunological datasets used for Chapters 2 and 3. Without her efforts, the
opportunity to test my hypotheses across geographical regions would not have been possible. I also thank her for the continued guidance and support in my continuation of her work.

The collaboration with the Kintampo Health Research Centre enabled the work presented in Chapter 4. Working with this group is a privilege, and I look forward to many more years of science together. Dr. Dennis Adu-Gyasi especially gave much of his time to support the work presented here.

The hospitality and support given to me and the study presented in Chapter 5 is the real reason it was a success. I sincerely thank Gabriel Gomez for providing me lessons in the local language and Lilica Sanca for her dedication to the study. I also thank Dr. Frederick Schaltz-Buchholzer, Dr. Morten Bjerregaard-Andersen, Dr. Kristina Lindberg-Larsen, and Dr. Christian Golding who provided support for the study over the years. I thank Dr. Laura Sly and the Microbiology and Immunology undergraduate program for donating the microcentrifuges used for this study. Most importantly, I thank Dr. Christine Stabell Benn and Dr. Peter Aaby for the sacrifices they have made to keep BHP alive, and all of us together.

Embarking on my own journey of motherhood during my PhD studies came with its challenges. I thank Dr. Christine Stabell Benn and Dr. Eleanor Fish for their honest advice. I thank Dr. Rym Ben Othman and Dr. Bing Cai for being exemplary role models, and by embedding the quote “Don’t stop when you’re tired. Stop when you’re done” into my life. And my little ones, Elie and Ayla, for reminding me to stay excited about the world around me.

When the going gets tough, having a tough skin helps. For that I thank my parents, Larissa and Roger. And my siblings Diana and Andrew for their wisdom and advice. And most of all, I thank my partner Tobi for changing my life’s trajectory. I look forward to many more years of discovering the world with you and our children.
Dedication

For a strong, resilient, and beautiful woman whose conversations with me shaped me most in early life; Нелли Андреевна Подболотова (English Nelly Andreyevna Podbolotova), my grandmum.
Chapter 1: Introduction

1.1 Conversing with microbes: systemic innate immunity

We are all a part of an ecosystem. Including not just the environments we can see around us, but also the one invisible to the naked eye; the micro-organisms living on and inside our bodies. This invisible world has been named the microbiome, and includes all the bacteria, fungi, viruses, and protists inhabiting our bodies. And it is not a ‘small’ world: estimates range from one thousand to up to one million bacterial species inhabiting any one person at a given time. Microbial communities assemble uniquely within individuals, and distinctly across body sites. This process begins at birth or perhaps even earlier. We benefit from the presence of commensal microbes, whose colonization serves to exclude potentially harmful microbes. Other commensals help us extract nutrients from our diet. And while invisible to us, the human body is fully aware of their presence, and is constantly conversing with our microbiomes via the immune system and other mechanisms. Our immune system has in fact been described as a sensory organ to the microbial world. And in this context, the immune system has evolved to do three things: 1. detect the presence of microbes and discriminate between different types (viruses, and bacteria for example), 2. decide whether the microbe is beneficial or harmful, and 3. based on this and on the body’s resources respond to the microbe using a diverse set of evolutionary strategies with the goal of maintaining homeostasis. The microbiome has also evolved to respond; commensal organisms educate our immune systems by delivering cues to facilitate their colonization while optimizing deleterious responses to harmful microbes in both mucosal and systemic sites. This is the host-microbiome interphase.
1.2 Innate immune responses to the microbiome

Innate immune responses to our microbiome, including commensal and pathogenic microorganisms, includes recognizing microbes by their conserved molecular patterns, producing contextually appropriate signaling molecules, and executing effector functions to either eliminate or tolerate the detected microbes.

1.2.1 Pattern recognition receptors sense the microbiome

The innate immune system has evolved sensory receptors to detect the microbiome based on conserved molecular patterns shared among related microbes. These are called pattern recognition receptors (PRRs), and include membrane-bound Toll-like receptors (TLRs)\textsuperscript{19} involved in sensing various types of bacteria, parasites, and viruses and type 1 and 2 C-type lectin receptors (CLRs)\textsuperscript{20} that recognize C-type lectins on fungi (including Dectin-1, Dectin-2, and MINCLE). Cytoplasmic PRRs that sense intracellular viruses including Rig-like receptors (RLRs); RIG-1, Melanoma differentiation-associated protein 5 (MDA5), and RLR-3\textsuperscript{21}. Nucleotide-binding oligomerization domain-containing protein (NOD) like receptors, or NLRs, sense intracellular bacteria\textsuperscript{22}. Secreted receptors include complement and mannose-binding lectin (MBL), which bind to bacteria to make them more visible to immune cells\textsuperscript{23}. Together these PRRs constitute a toolbox for our body to detect the presence of the microbes via microbe-associated molecular patterns (MAMPs).

1.2.2 Toll-like and Nod-like receptors

The TLR family of PRRs are among the best studied, and include 10 human types involved in microbe detection. These membrane-bound receptors signal as either homo- or heterodimers and
are expressed most highly by phagocytic innate immune cells, including monocytes, macrophages, and neutrophils and dendritic cells. The TLRs expressed on extracellular membranes detect the presence of MAMPs most likely encountered in the extracellular space, such as lipoproteins found on gram-positive bacteria sensed by TLR2:1 and TLR2:6 (for triacyl and diacyl lipoproteins respectively); lipopolysaccharides (LPS) on gram negative bacteria sensed by TLR4, flagellin by TLR5; and profilin-like molecules on protozoa, by TLR11. TLRs can also be expressed within endolysosomes to sense nucleic acids from bacteria and viruses; double-stranded and single-stranded ribonucleic acids (dsRNA and ssRNA) are sensed by TLR3 and TLR7:8 respectively and CpG-deoxyribonucleic acid (DNA) by TLR9. NOD1 and NOD2 recognize bacterial peptidoglycan in the cell’s cytosol. NOD1 expression is not restricted to immune cells but has been found in a wide range of tissues, and NOD2 appears more restricted to phagocytes. Both TLRs and NLRs are now known to be expressed not only by immune cells but by epithelial cells lining our body’s mucosa, and are crucial for proper barrier maintenance between microbes and the human host; this fact also emphasizes that the immune system (and in this case PRRs) are not restricted to immune cells proper, but involves all of the host.

Generally, TLR stimulation leads to initiation of pro-inflammatory responses. These can be split into myeloid differentiation primary response 88 (MyD88) protein or TIR-domain containing adapter-inducing interferon-β (TRIF) dependent signaling. Extracellular TLRs use the MyD88-dependent pathway, where upon ligation a complex called the myddosome made up of adapter proteins MyD88, IRAK4, and IRAK1 assemble at the receptor’s cytosolic TIR domain and initiate a signaling cascade that eventually ends with the activation of pro-inflammatory transcription factors NF-κB, IRF5, and AP-1. Endosomal TLRs use the TRIF-
dependent pathway, whereby TRIF binds to the receptor’s cytoplasmic domain and activates transcription factors IRF7 and IRF3 in addition to NF-κB and IRF5, leading to the production of Type 1 interferons (IFNs)\textsuperscript{34}. Uniquely, TLR4 can utilize both pathways, signaling via MyD88 when activated on the cell surface and TRIF when in the endolysosome. NOD1 and NOD2 also activate NF-κB, but through separate mechanisms. Upon activation, NOD receptors dimerize and recruit serine-threonine kinase RICK (RIP-like interacting CLARP kinase), which go on to activate NF-κB\textsuperscript{26}. Activated NOD proteins are also able to induce assembly of inflammasomes that lead to the activation of IL-1β to further promote an inflammatory response\textsuperscript{35}.

**1.2.3 Cytokine responses to PRR stimulation**

Innate immune cells mount contextually appropriate responses after PRR-mediated sensing of microbes by producing different types of signalling molecules – cytokines and chemokines. These molecules instruct the host on the response required towards any given microbe. Most of our knowledge comes from studying responses to pathogens\textsuperscript{15}, which are distinct from the types of responses elicited by commensals\textsuperscript{13}. In terms of host defence, immune responses have been characterized by type 1 immunity (defending against bacteria and viruses) and type 2 immunity (defending against worms)\textsuperscript{15}. Type 1 immunity is sometimes divided to distinguish responses to intracellular (type 1; T-helper[Th] 1) and extracellular (type 3; Th17) bacterial or viral threats\textsuperscript{36}. Type 1 and type 3 immune responses follow the activation of NF-κB and result in release of pro-inflammatory cytokines such as IFN-γ, IL-6, TNF-α, and IL-1β, along with chemokines IL-8, macrophage inflammatory protein (MIP)-1α and MIP-1β which serve to recruit macrophages and neutrophils to the site of infection\textsuperscript{37}. Physiologically, a pro-inflammatory cytokine milieu
causes blood vessels to dilate, increasing access of effector immune cells (neutrophils and macrophages in extracellular infection) to the site of infection. Once recruited, these cells are activated by the pro-inflammatory cytokines (e.g. TNF-α and IL-1β) and use several mechanisms to destroy the invading pathogens. This is perceived by the host as pain, redness, and swelling at the site of infection. For intracellular infections, interferons are often also produced, including type 1 (IFN-α and β) and type 2 (IFN-γ) alongside chemokines IP-10 that activate cell autonomous immunity or cell death. Innate pro-inflammatory responses are immediate, but are further amplified by adaptive immune responses. The interaction between innate and adaptive immunity is coordinated already at the time of PRR activation. For example, IL-12p40, IL-6, and IL-23 promote T cells differentiation into Th17 cells, which amplify responses to extracellular pathogens, while IL-12p70 promotes the differentiation into Th1 cells and cytotoxic CD8+ T cells, central in clearance of intracellular targets. These responses represent the disease strategy of resistance – a process with a high fitness cost to the host, but effective at eliminating unwanted and potentially harmful microbes. Halting the inflammatory process is critical to avoid excessive physiological damage, and IL-10 induced already alongside pro-inflammatory cytokines eventually dampens the inflammatory response by inhibiting the production of pro-inflammatory cytokines. The above short overview represents only a small portion of the cell signaling mechanisms used in host defence; I highlighted them here as they capture the broad classes of effector molecules that feature later in my work.

1.2.4 Innate immune responses to commensals

While much attention has gone to host defence, most of our lifespan (ideally) is spent sensing commensal microbes using PRRs without mounting the pro-inflammatory responses described
above. The immune system accomplishes this in partnership with commensal microbes, who have evolved ways to engage our PRRs via MAMPs that do not convey a danger signal. Much of this is done by modifying PRR signaling based on where they are sensed. The most famous example of this comes from TLR9 expression on intestinal epithelial cells (IECs), on both the luminal and basolateral surface of the intestine. However, only basolateral stimulation of TLR9 results in a pro-inflammatory response\textsuperscript{29,40}. TLR5 is only expressed basolaterally on IECs, and when stimulated results in pro-inflammatory responses\textsuperscript{27,41}. In fact, under steady state conditions, TLR signaling is necessary to maintain homeostasis in the intestine. In an unexpected finding, mice lacking MyD88 sustained much more colonic injury after experimentally induced colitis compared to wild-type mice\textsuperscript{42}, had downregulated antimicrobial peptides, and had a compromised intestinal mucus layer\textsuperscript{43}. Commensal microbes have also adapted to induce tolerogenic responses towards themselves. Double-stranded RNA from lactic acid bacteria (LAB) is sensed by dendritic cells in the gut lumen via TLR3, which in turn induces IFN-β but in the local context contributes to an anti- not pro-inflammatory environment. Potentially pathogenic bacteria such as \textit{Streptococcus pneumoniae} also contain dsRNA, but their nucleic acids did not have this effect\textsuperscript{44}. Importantly, the influence our commensals have on immune development reaches beyond the mucosa; microbial colonization influences immune development systemically.

\subsection*{1.3 The gut microbiome modulates systemic immune responses}

The intestinal microbiome is a well-recognized modulator of human health and disease throughout life, and recent work has illuminated the molecular mechanisms by which commensal microbes shape systemic immunity.
The microbiome influences how MAMPs are sensed by PRRs, and with that, helps to orchestrate immune responses to themselves and other commensals\(^4^5\). Direct effects of gut bacteria on innate immunity were demonstrated in humans; for example, lipopolysaccharides (LPS) from commensal *Bacteroides* species elicit a weaker pro-inflammatory response from innate immune cells compared to LPS from *Escherichia coli*, resulting in altered risk for autoimmune disease among populations dominated by one commensal vs. another\(^4^6\). Further, exposure to bacterial LPS in house dust decreased risk for asthma among Hutterite communities, a finding supported by studies using animal models\(^4^7\). Animal studies have also shown the necessity of microbes and their components for development of systemic immunity, with germ-free mice lacking appropriate hematopoietic cell development in the bone marrow\(^4^8\) and altered systemic responses by natural killer (NK) cells\(^4^9\), neutrophils\(^5^0, 5^1\), basophils\(^5^1\), and macrophages\(^5^2\). Innate immune sensing of bacterial components and metabolites has even been shown to be necessary for antibody responses to several vaccines in rodents\(^5^3\). Polysaccharide A (PSA), a sphingolipid specific to *B. fragilis*, was among the first bacterial products shown to induce maturation of CD4+ T cells in both the mucosa and spleens of germ-free mice\(^5^4\). Thus, commensal bacteria influence the development of our systemic innate immunity in a multitude of ways\(^4^5\).

1.4 Early infancy is a critical window to establish host-microbiome homeostasis

As summarized above, most of what we know about interactions between the gut microbiome and systemic immunity comes from adult animal models. But it is now increasingly recognized that early life represents one of the most important periods during which the conversations between the gut microbiome and systemic immunity shape the trajectory of health and disease for years, if not life, to come. This period unfortunately has not been sufficiently well
characterized in animal models, and more importantly, very rarely addressed in humans\textsuperscript{55}. This massive knowledge void is what my thesis is aiming to address.

1.4.1 **The bacterial microbiome changes across the age spectrum**

The development of microbiota throughout infancy has been the topic of numerous reviews\textsuperscript{56-59}. In one comprehensive review that I contributed to, we highlighted the major known determinants of microbiome composition in infancy to include delivery mode, diet (including breast milk versus formula and the introduction of solid foods), antibiotic use, and of course geography. Perturbations to pediatric microbiomes are associated with diseases such as necrotizing enterocolitis, allergic disease, inflammatory bowel disease, obesity, and more recently neurodevelopmental disorders\textsuperscript{60}. Here, I focus on the normal broad compositional shifts occurring in the first few years of life. These dynamic shifts have been well-studied, but mostly in resource-rich and only few resource poor settings\textsuperscript{10, 61-65}. These surveys have shown microbial diversity to increase with age and reach adult-like profiles by 2-3 years. Over the first year of life, the microbiome rapidly changes with a succession of bacterial taxa beginning with aerobes such as *Escherichia, Streptococcus* and *Staphylococcus* in the first week of life that soon are replaced by obligate anaerobes like *Bacteroides* and *Clostridia*, which then continue to feature prominently into adult life\textsuperscript{61}. Another study applied metagenomics to identify functional changes in microbiome changes with age; while infant bacteria contain folate synthesis genes, those in adults contain more genes for folate metabolism and cobalamin, vitamin B7, and B1 synthesis\textsuperscript{10}. A recent study used both 16S amplicon sequencing and metagenomics to profile gut microbiomes of 903 European infants from the first to 46\textsuperscript{th} month of life. They used Dirichlet multinomial mixtures (DMM) to identify 10 microbial clusters across this age gradient that fell
into three phases: a developmental phase during months 3-14 characterized by high abundance of *Bifidobacteria, Proteobacteria* and *Bacteroidetes*, the transitional phase during months 15-30 highlighted by unstable microbiomes across time and a decrease in *Proteobacteria*, and then a stable phase from 31 months onwards where no major taxonomic shifts were seen over time. Major transitory points to more adult-like states in microbial composition have been seen to occur around the time of weaning, but not at the introduction of solid foods, which may represent the stable phase of the microbiome described above. The introduction of a diverse array of nutrients selects for greater abundances of *Ruminococcus, Clostridium, and Bacteroides* genera, who outcompete *Bifidobacteria* and *Enterobacteriaceae* that thrive in breast-fed infants. Given the dynamic nature of microbiome community structure during this time, and known links between the microbiome and immune-mediated disease in early life, understanding the relationships between host and microbe at this time is all the more warranted.

### 1.4.2 Host-microbiome interactions in infancy

Host immunity in the first days of life likely reflects a fine-tuned and balanced match to the microbial pressures placed on the newborn around birth, all constrained further by the substantial host energetic demand of rapid growth and development. The resulting newborn adaptations in the days after birth may have evolved to both ensure safe colonization across multiple body sites while still protecting themselves from infectious threats. Only over the last decades has a body of research emerged that implicates the microbiome in newborn infections, including necrotizing enterocolitis (NEC), and both early and late-onset sepsis. Microbial community assembly in the newborn human gut has been studied extensively using molecular techniques, finding that community composition changes rapidly during this time, is largely
derived from the mother’s vagina at birth\textsuperscript{75}, and continues to share bacterial populations with their mother’s microbiome in the first days of life. For example, \textit{Bacteroides} and \textit{Clostridia} can be found in both mother and newborn in roughly equal measure\textsuperscript{64}. \textit{Bacteroides} feature prominently in both mother and baby\textsuperscript{64} likely because breast milk promotes the growth of these human milk oligosaccharide (HMO)-utilizing bacteria that may in return provide colonization resistance to potential pathogens\textsuperscript{76-78}. Meanwhile, \textit{Escherichia/Shigella}, \textit{Bifidobacterium}, \textit{Streptococcus}, and \textit{Enterococcus} occupy roughly half the space of the entire intestinal microbiome in newborns but only about 10\% in their mothers.

Mechanistic evidence that the microbiome shapes newborn immunity to facilitate colonization stems from mouse models\textsuperscript{79}. Primarily, these mechanisms seem to serve to avoid inflammatory or energetically deleterious responses of the host to commensals. Older infants and adults contain a thick mucus layer packed with anti-microbial peptides and dimerized IgA to prevent microbial translocation, but newborns do not yet have this barrier in place\textsuperscript{80}. Instead, newborns employ their own strategies to tolerate their ‘new friends’; for example, newborn intestinal epithelial cells (IECs) produce a micro-RNA molecule (miR-146a) that dampens TLR-4 mediated pro-inflammatory responses by targeting downstream signaling transducer IRAK-1\textsuperscript{81}. In fact, the entire intestinal transcriptome is highly unique in newborn mice, and expresses transcriptional regulator \textit{Blimp1} that is active only in newborns, but not adult mice. Deletion of \textit{Blimp1} results in adult-like intestinal architecture in newborn mice but is accompanied by a significant increase in early life mortality\textsuperscript{82}. In rodents, transcriptional regulation shifts post-weaning with an increase in IL-1/TLR signaling, showing that intestinal immunity is influenced by the changing microbial and nutritional landscape\textsuperscript{83}. Specific colonizers also directly influence newborn immunity in mice. Recently, clostridia-derived metabolite desaminotyrosine (DAT)
was shown to promote type 1 IFN signaling in the lung, thereby dampening pro-inflammatory responses and increasing survival post influenza challenge in newborn mice\textsuperscript{84}. Early life colonization may have lasting consequences beyond the newborn period. As I recently reviewed, microbial perturbations in the first months of life are linked to altered disease phenotypes beyond the newborn period\textsuperscript{85}. \textit{B. fragilis} for example inhibits recruitment of invariant NKT cells to the gut and lung mucosa, leaving mice less susceptible to experimentally induced colitis\textsuperscript{86}. Clostridial species (specifically, Clostridium clusters IV and XIVa) have also been shown to induce regulatory T cell accumulation in mouse colons if present during specific early life periods. Colonizing mice with these bacteria at two weeks of age protects them from colitis in adulthood and lowers their systemic IgE levels\textsuperscript{87}. On the other hand, exposure to segmented filamentous bacteria (SFB) in early life of mice is uniquely able to induce large numbers of Th17 cells\textsuperscript{88} using a mechanism dependent on adherence of SFB to the intestinal epithelium\textsuperscript{89}.

1.5 Systemic immunity and gut microbiomes across diverse populations

Given the extraordinary effects that microbes play in shaping immune responses in animal models, a compelling hypothesis that follows is that varied microbiomes across diverse populations should have population-specific effects on the human immune system. Assessing such correlation represents one of the few approaches to determine relationships at the microbiome-immune interphase in humans; this concept represents much of the rationale behind my PhD project.
### 1.5.1 Gut microbiome across geographically distinct populations

Among humans, the gut microbiome has long been recognized to differ drastically between geographically distinct populations. In a review I contributed to, I highlighted the effect of geography on gut microbiomes\(^8^5\), which I expand upon here. Early work showed striking differences between the microbiomes of West African and European infants\(^9^0, 9^1\). These populations were contrasted because of the dramatic difference in their environments and lifestyles. Infants from Burkina Faso live in a society representing the ancestral Neolithic lifestyle, with a diet mostly composed of legumes and grains. Meanwhile, Florentian infants lived a more typical western lifestyle, and with a diet including processed foods, higher in animal proteins and fats. The two populations differed predominantly by the composition of the *Bacteroidetes* phylum, whereby infants from Burkina Faso were dominated by *Prevotella* and Florentians by *Bacteroides*. Infants from Burkina Faso also had a reduced Firmicutes to *Bacteroidetes* ratio. The latter has been thought to reflect climate, whereby populations in colder climates have increased abundances of Firmicutes to facilitate extraction of fat from the diet\(^9^2\).

Since then, the conclusion that those living in industrialized nations tend to be dominated by the *Bacteroides* genus while those in more rural or non-industrialized settings dominated by *Prevotella* have been well-supported\(^9^3\). Even within the same country, urbanized lifestyles in Nigeria are associated with a microbial repertoire closer to other industrialized countries, including a decreased abundance of *Prevotella* and increase in *Bacteroides* compared to Nigerians living in a more traditional rural environment\(^9^4\), similarly to rural vs. urban dwellers in Burkina Faso\(^9^1\). Diet is of course thought to be a predominant driver of microbiome composition. Individuals dominated by either *Bacteroides* or *Prevotella* can be found within the same geographical environments\(^9^5, 9^6\) and *Prevotella* associated with a high-fibre diet.
While the *Prevotella-Bacteroides* dichotomy also applied to American versus Venezuelan or Malawian infants through the age spectrum\textsuperscript{10}, metagenomics allowed for the assessment of potential functional differences in gut microbiomes. Among infants, Malawian and Venezuelan infant microbiomes harbored more genes for in glycan and urease metabolic pathways indicative of an enhanced ability to assimilate nitrogen and glycans from breast milk. This was hypothesized to reflect lower availability of these compounds from their mother’s milk.

Despite the well-reported and dramatic differences seen in gut microbiomes across diverse populations, little is known about the consequences this has on systemic immunity. This thesis represents the first effort to do so across diverse geographic settings.

1.5.2 Systemic immune responses across geographically distinct populations

Less-studied, but still recognized, are differences in immune phenotype between distinct populations. Responses to vaccines are known to differ between populations\textsuperscript{97}. For example, in measuring *ex vivo* cytokine responses to the bacille Calmette Guerin (BCG) vaccine, Malawian infants responded with a pro-inflammatory cytokine signature highly distinct from the T-helper type 1 (Th1) cytokine signature mounted by infants from the United Kingdom\textsuperscript{98}. Cord blood antigen presenting cells from infants born in Papua New Guinea (PNG) were shown to secrete more pro-inflammatory cytokines *in vitro* at baseline than those from infants in an industrialized setting in Western Australia, but also responded less to *in vitro* stimulation\textsuperscript{99}; this was followed by a distinct trajectory over the first two years of life with pro-inflammatory responses to TLR4 agonists increasing for the PNG infants over the first two years, while anti-inflammatory IL-10 responses remained constant\textsuperscript{100} – consistent with findings from a Gambian cohort\textsuperscript{101}, but contrasting with previous observations in industrialized settings\textsuperscript{102-104}. The underlying cause for
such human population-based differences in systemic immunity in early life has never been determined; yet, in nearly all reviews of this topic it has been hypothesized that this would be due to population-based differences in the composition of the bacterial microbiome. My thesis work is the first that directly examined this.

1.6 The fungal microbiome in early life

While the bacterial microbiome has received much attention over the last decades, the fungal microbiome has been barely studied. Still, the few studies that exist demonstrate that fungi also take part in host-microbiome conversations.

1.6.1 The fungal microbiome shapes systemic immunity

The fungal microbiome (‘mycobiome’) has rarely been studied, despite emerging evidence that fungi are key modulators of health and disease\textsuperscript{105, 106}. While bacterial colonization has been shown to correct immune deficiencies in germ-free mice as described above, this is also true for fungi. Recent animal studies demonstrated that fungi effectively calibrate mucosal immune responses with an emphasis on Th17 immunity and neutrophil function\textsuperscript{105}, and colonization with Candida albicans or probiotic fungus Saccharomyces cervisiae increase murine survival of subsequent systemic viral challenge\textsuperscript{106}. Protective benefits of fungi were mediated by interactions with mannans and host TLR4, showing that TLR signaling also plays a role in host-fungal mycobiome crosstalk. And while MyD88 was shown to be important for maintenance of intestinal homeostasis during experimentally induced colitis\textsuperscript{42, 43}, mice lacking fungal PRR dectin-1 were also found to be more susceptible to experimentally induced colitis, due to inability to control fungal populations\textsuperscript{107}. These recent works highlight the importance of
commensal fungi on host health and highlight the need to also understand fungal community
dynamics across the age spectrum.

1.6.2 The fungal microbiome across the age spectrum

The human fungal microbiome in early life has thus far been surveyed in very few studies, and in
only a few select populations: Europe; Norway, Italy, Luxembourg and in the Americas; Puerto
Rico and Ecuador. Among these diverse human studies, *Candida*, *Aspergillus*, and
*Saccharomyces* feature prominently in newborns and older infants, yet with no clear age-
dependent taxonomic trajectory in the first two years of life of taxonomic composition or a difference in microbial diversity in the first two years of life aside from reduced
diversity in 10-day-old newborns compared to 3-month-olds or their mothers. This trend was
also true for a more granular assessment of the newborn fungal microbiome over the first
month. Yet, sequence-based surveys of infant mycobiomes have revealed increased abundance
of fungal taxa to be associated with the development of immune mediated allergic disease later
in childhood. Thus, assessment of the fungal microbiome in early life alongside bacterial
microbiome development appears important. My thesis work addresses this crucial knowledge
gap.

1.6.3 Interactions between bacterial and fungal microbiomes

Bacterial and fungal microbiomes interact with each other. Most of our current knowledge of
these inter-kingdom interactions across human microbiomes comes from studying potentially
pathogenic commensal fungus *Candida albicans* (*C. albicans*) with our commensal bacteria. Some of these interactions are synergistic. From *in vitro* studies it was deduced that common
fungal commensal \textit{C. albicans} forms anoxic biofilms that allow growth of common anaerobic gut commensals, \textit{Bacteroides fragilis} and \textit{Clostridium perfringens}\textsuperscript{116}. \textit{C. perfringens} specifically induced \textit{C. albicans} mini-biofilm formation, where the two species co-aggregated and \textit{C. perfringens} could grow in oxic culture conditions. Many antagonistic interactions have been described. For example, \textit{Pseudomonas aeruginosa} prevents \textit{C. albicans} filamentation\textsuperscript{117}, and kills the fungus by forming biofilms around its hyphae\textsuperscript{118}. Such relationships are also relevant \textit{in vivo}. One well-studied relationship is between lactic acid bacteria (LAB) and \textit{Candida} colonizing the vaginal epithelium\textsuperscript{119}. Adherence of LAB to the vaginal epithelium prevents colonization and growth of both bacterial and fungal pathogens, including \textit{C. albicans}\textsuperscript{120, 121}. LAB have been shown to directly influence \textit{C. albicans} filamentation\textsuperscript{122}, and modulate epithelial immune responses in responses to \textit{C. albicans}\textsuperscript{123, 124}, and in doing so, prevents fungal infection. Other gut commensals also keep the fungus in check; short-chain fatty acids such as butyrate produced by gut-residing LAB inhibit \textit{C. albicans} germination\textsuperscript{125, 126}. Still, given that fungi across a diverse array of phyla are found in the gut\textsuperscript{111, 112}, and clear beneficial uses of probiotic fungi already exist\textsuperscript{127}, a deeper understanding of interactions among commensal bacterial and fungi is clearly warranted.

1.7 The breast milk microbiome

The breast milk bacterial microbiome has also been studied using culture-based\textsuperscript{128-130} and culture-independent\textsuperscript{131, 132} approaches, which revealed its’ potential to seed the infant intestinal microbiome\textsuperscript{133}. Breast milk harbours a wide range of microbes from 100 to $10^5$ CFU per ml depending on the study \textsuperscript{77}, with \textit{Streptococcus} and \textit{Staphylococcus} being most common, but others such as \textit{Lactobacillus, Bifidobacterium, Enterococcus}, and \textit{Propinobacterium} readily
isolated from milk of healthy women\textsuperscript{77}. Many short-chain fatty-acid producing bacteria such as \textit{Veillonella}, \textit{Propionibacterium}, and \textit{Faecalibacterium} have also been isolated from breast milk\textsuperscript{77}. Breast milk also contains LAB, double-stranded RNA from which stimulates intestinal dendritic cells via TLR3 to produce IFN\textgreek{b}, which in turn promotes an anti-inflammatory environment and protects mice against colitis\textsuperscript{44,134}. However, the key mediator of the immune homeostatic function of breast milk is presumed to be its impact on bacterial gut microbiota. In mice, for example, the presence of maternal secretory IgA (sIgA) in early life moulds the composition of the gut microbiota long into adulthood, with pups born to sIgA-deficient dams harbouring more \textit{Pasteurellaceae} and \textit{Lachnospiraceae} than controls\textsuperscript{135}.

The breast milk mycobiome has barely been studied using culture-independent approaches, and only recently profiled using amplicon sequencing for the first time\textsuperscript{136}, revealing a breast milk mycobiome dominated by prominent \textit{Malassesia} genera also commonly found on skin, with compositional differences across diverse populations\textsuperscript{137}. This severely under-explored area has yet to shed light on relationships between mother and child and consequences of fungi in breast milk on infant health.

\section*{1.8 HIV-exposed uninfected infants}

Globally, over 15 million women in their reproductive years are living with HIV\textsuperscript{138}. The provision of antiretroviral (ARV) treatment during pregnancy to prevent HIV transmission to their children increased from 47\% to 76\% from 2010 to 2016. With this increase in ARV coverage, new pediatric HIV infections have declined by 47\% during that time\textsuperscript{139}. With this, HIV-exposed uninfected (HEU) infants are becoming a growing population that, despite the absence of HIV infection, remain more susceptible to other severe infections in early life\textsuperscript{140}. The
underlying cause of this altered phenotype in HEU infants has not been determined. An altered microbiome may be centrally involved, but this has never been evaluated alongside immune surveys in HEU infants. To my knowledge, my work on this topic represents the first systematic approach to this question.

1.8.1 HEU infants are more susceptible to severe infection in early life

Across multiple studies, HEU infants have been found to be more vulnerable to severe infection and mortality in infancy than HIV-unexposed uninfected (HUU) infants\(^1\)\(^{141-143}\). The reasons for this vulnerability are not understood and could include a combination of factors ranging from socioeconomic to \textit{in utero} exposure to HIV and antiretroviral medications, as well as altered microbiomes. Specifically, HEU infants suffer from more invasive bacterial disease with common bacterial etiological agents identified: \textit{Streptococcus pneumoniae}\(^{144-146}\) and neonatal Group B Streptococcus (GBS)\(^ {147, 148}\). Increased severity and incidence of GBS disease in HEU infants, despite similar GBS colonization rates to HIV-negative women\(^ {149}\) may be caused by reduced transplacental transfer of anti-GBS antibodies and reduced antibody-mediated complement deposition\(^{150}\). Increased incidence of viral infections among HEU infants have also been described for viral lower respiratory tract infection (LRTI)\(^ {151}\) and earlier Epstein-Barr virus (EBV) infection in one Kenyan\(^{152}\) and one Ugandan\(^ {153}\) cohort. Increased cytomegalovirus (CMV) transmission from mother to infant was seen prior, but not after introduction of perinatal ARV\(^{154}\). While possible mechanisms are clear for some instances, such as neonatal GBS, for other instances plausible biological mechanisms are not known and could involve alterations in innate and adaptive immune responses driven by altered microbiomes.


1.8.2 Altered systemic innate immunity in HEU infants

Few studies have evaluated the innate immune status of HEU versus HUU control infants. But of the studies done, commonalities emerge showing that HEUs have fewer circulating innate immune cells\textsuperscript{155, 156}, mount higher pro-inflammatory responses\textsuperscript{157, 158} and have lower innate antiviral responses\textsuperscript{157, 159}, compared to HUU controls.

One longitudinal study of South African HEU infants found that they exhibited altered innate immunity up to 6 weeks of life, but were indistinguishable by 1 year compared to HUU controls. Specifically, using an assay whereby whole blood was stimulated with TLR agonists and both secreted and intracellular cytokines were assessed, monocytes and dendritic cells from HEU infants produced more pro-inflammatory cytokines IL-6 and TNF-\(\alpha\) in response to PAM and LPS. Conversely, HUU controls responded more to viral ligands; monocytes produced more IL-12p40 in response to TLR3 stimulation and plasmacytoid dendritic cells (pDCs) produced more IL-6 in response to TLR7:8 stimulation\textsuperscript{157}. Proportions of circulating monocytes remained unchanged throughout the course of this study, while another study done in the Netherlands comparing HEU infants to matched HUU controls found reduced absolute numbers of monocytes in peripheral blood up to 5 months of age\textsuperscript{156}. Altered innate antiviral responses in HEU infants have also been detected in other studies. In a small Kenyan cohort, HEU infants had similar proportions of circulating natural killer (NK) cells compared to HUU control infants, but their levels of perforin expression diminished over the first year of life while levels in HUU infants increased during that time\textsuperscript{159}. Perinatal ARV exposure may account for some innate immune differences between HEU and HUU infants, especially for neutrophil counts as HEU infants exposed or unexposed to perinatal ARV had lower levels of circulating neutrophils up to in infancy\textsuperscript{160, 161}, and in one study even up to 8 years of age\textsuperscript{155}. The suppressive effect of
zidovudine on hematopoiesis was also confirmed in *in vitro* studies\textsuperscript{162}, and may be involved in the effect seen on neutrophils.

1.8.3 Altered gut microbiomes in HEU infants

Gut microbiomes of HEU compared to HUU infants have barely been studied. There are several reasons to hypothesize that gut microbiomes of HEU infants differ from HUU controls. HIV infection is associated with altered gut microbiomes, and even differs between treated and ART-naïve adults\textsuperscript{163, 164}. Human milk oligosaccharide (HMO) composition is different in HIV-positive compared to HIV-negative women\textsuperscript{165}. Previously, HIV-positive women were advised to formula feed their infants to reduce risk of HIV transmission through breast milk\textsuperscript{166} and formula feeding itself is associated with altered gut microbiomes in infants\textsuperscript{64, 65}. However, only one study on gut microbiomes of HEU infants has been performed. In a population where HEU infants were breastfed, microbiomes of Haitian HEU infants differed from HUU controls and correlated with certain HMOs in their mother’s breast milk\textsuperscript{165}. In this study, maternal stool microbiomes did not differ between HIV-positive and negative women. Infant microbiomes did, however, differ with reduced bacterial diversity, specifically among those infants whose mothers had a CD4 cell count under 350. HEU infant microbiomes were enriched for *Pseudomonaceae* and *Thermaceae*, and depleted in *Prevotellaceae*, *Desulfovibrionaceae*, and *Alcaligenaceae* compared to those of HUU controls infants. HEU and HUU infant microbiomes correlated differently with maternal HMO profiles; for example, 3’sialyllactose, found in higher abundance in breast milk of HIV-negative women was positively correlated with levels of *Enterococcaceae* and *Fusobacteriaceae* in their infants, while in HIV-negative women, both *Bifidobacteriaceae* and *Staphylococcaceae*
associated with HMOs. This study suggests that maternal HIV infection may change the way that breast milk shapes the infant gut microbiome.

1.9 Rationale

The last few decades have shed tremendous light on the effects our microbiome has on our health. Human studies have informed us about how bacterial communities assemble across diverse age and geographic spectra, how select members of the microbiome modulate mucosal and systemic immune development, and how crucial this is in the newborn window. Thus, the new challenge is to link the varied composition of gut microbiomes in humans to the consequences this has on age- and population-specific immune development, and to expand our knowledge of the microbiome outside the bacterial kingdom in the process. While mouse models will continue to be immensely useful tools dissect the effects of the microbiome on mammalian hosts, the identification of relevant taxa must first come from human studies. Such hypothesis-generating studies are challenging and require highly standardized assessment of immune phenotypes to allow for direct comparison across a diverse age spectrum. Analytical approaches used must tackle the complex, high-dimensional structure of both microbiome and immune data to resolve patterns worthy of further laboratory investigation. Moreso, while the newborn period is critical in establishing host-microbiome homeostasis, it is the least-studied due to difficulties in collecting blood specimens. This is especially true in low-resource settings, where the need to understand this interphase to prevent newborn morbidity and mortality is greatest.

In this work, I present the first steps to address the knowledge gaps identified above. I conducted the first study to investigate population-specific differences in systemic cytokine responses to PRR stimulation, using a highly standardized approach, which illustrated
differences in immune phenotypes across infant populations from four continents\textsuperscript{167}. Here, I assessed the composition of their stool bacterial microbiomes across these cohorts and applied an integrated statistical approach to address my first hypothesis:

\textit{Hypothesis 1. Population-specific differences in systemic innate immune phenotypes are correlated to differences in the gut microbiome.}

Both systemic cytokine responses and gut microbiomes of HEU infants were measured in three of those populations from around the world, but had not yet been analyzed. In Chapter 3, I used a discriminatory integrated analytical approach to address my second hypothesis:

\textit{Hypothesis 2. Altered immune phenotypes in HEU infants are population-specific and correlated with compositional changes in their gut microbiomes.}

In Chapter 4, I step away from the exclusive focus on bacterial microbiomes, and apply a sequence-based approach to compare fungal to bacterial community composition across the first 5 years of life in a rural population whose microbiome had never been studied to test my third hypothesis:

\textit{Hypothesis 3. The Fungal microbiomes also assembles in an age-specific way and is partly inherited from the mother at birth.}
Lastly, in Chapter 5, I address the lack of tools available to study immune development in the newborn period at the functionally relevant fine granular complex detail to test my final hypothesis:

_Hypothesis 4. Hurdles to implement highly standardized immune phenotyping protocols in low resource settings can be overcome, making such populations more accessible to study host-microbiome interaction._

Taken together, my thesis provides a solid foundation from which to address some key questions: Are there interactions between microbiome and systemic immunity in early human life? If so, who influences whom? What are the key pieces to this likely complex puzzle to initially focus on? What tools work and what tools do we need to chart out the course to successfully reach our destination? Addressing them paves the way towards furthering our understanding of the conversations our bodies have with our commensal microbiomes. My hope is that this knowledge will enable us to optimize the host-microbiome relationship to our advantage throughout our lifespan.
Chapter 2: Biogeography shapes the relationship between the infant gut microbiome and innate immune responses

2.1 Introduction

The gut microbiome is a well-recognized modulator of host immunity throughout life, and its composition differs between geographically separated human populations. Systemic innate immune responses to microbes are largely driven by cytokine responses to PRR agonists, which have also been shown to differ between geographically distinct populations. However, the specific role of the microbiome composition and function in mediating these differing immune responses across geographical regions is largely unknown. Given the ability of host microbiome to modulate systemic innate immunity, and the known geographical differences between both gut microbiomes and innate immune phenotypes, what is missing is a mechanistic understanding of how the distinct microbiomes likely contribute to immune differences. Robust correlations from human studies are therefore needed to inform mechanistic work using animal models.

Direct comparisons of innate immune responses among distinct locations is only possible if all aspects of immune assays employed are highly controlled and harmonized among each study site. Using a standardized approach, we measured the quantity of cytokine responses to a panel of TLR agonists among 2-year-old children recruited in Brussels, (Belgium), Cape Town (South Africa), Quininide (Ecuador), and Vancouver (Canada)\textsuperscript{167}. Here, we extended this work by profiling the gut microbiomes of the same study participants via sequencing of the 16S rRNA genes (V6 region) from stools collected at the time of blood draw and present the microbiome signatures specific to each cohort\textsuperscript{167}. 
Previous studies have found associations between systemic immunity and host microbiome within single cohorts, finding that relative abundance of microbial taxa or their genes could be correlated with select cytokine responses to TLR stimulation\textsuperscript{168, 169}. Both studies relied on univariate statistics of relative abundance data to find a small subset of microbiome-immune correlations. However, given limited sample size, high dimensionality of our data, and compositional nature of microbiome data, these approaches limit quality and interconnectedness of uncovered relationships. Specifically, univariate statistical methods may often lead to spurious results as the independence assumption between predictor variables is not met. Second, by only considering one-to-one associations, univariate approaches that test each OTU individually and disregard interactions or correlations between OTUs provide a more limited insight into the microbiome\textsuperscript{170}. To overcome these hurdles, we employed a sparse Partial Least Squares (sPLS) integrative approach on centered-log-ratio (CLR) normalized OTU data to test the hypothesis that regionally distinct gut microbiomes can indeed drive development of systemic immunity.

To test the hypothesis that immune suppressive phenotype of South African infants is caused by their gut microbiomes, we employed a germ-free mouse model of human fecal transplantation. To our knowledge, this study is the first characterization of host microbiome-systemic immunity correlations among geographically diverse child populations.

\subsection{Materials and Methods}

\textbf{Recruitment of Study Participants.} The recruitment of the four cohorts were as previously published\textsuperscript{167}. Approximately two-year-old children were recruited from Vancouver, Canada; Brussels, Belgium; Quininde, Ecuador, and Cape Town, South Africa. Study participants were only included in the study if they were considered healthy based on medical history, and they
were excluded if they met one or more of the following criteria: significant chronic medical condition, primary or secondary immune deficiency, immunosuppression by disease or medication, cancer, bone marrow or organ transplantation, receipt of blood products within 3 months, bleeding disorder, major congenital malformation, genetic disorder, or born to HIV-positive mothers.

**Innate immune phenotyping.**

Given that the main purpose for innate immune phenotyping in this study was to compare immune responses across distinct locations, it was of utmost importance that every aspect of the immune phenotyping assays be controlled to attribute differences between cohorts to biology rather than technical artefacts due to differences in reagents and handling specific to each site. To that end, the innate immune phenotyping protocol was designed to standardize reagents, sample processing, and immune analyses across all study sites. These methods were published previously.\(^\text{167}\).

**Preparation of PRR stimulation plates.** Master mixes of all reagents used for PRR stimulation were prepared simultaneously and in quantities adequate for the entire study at the Vancouver, Canada study site. To prepare PRR plates used for whole blood stimulation, 1.3 mL of specific PRR ligands were placed in deep 96-well source plates (VWR Cat. No 10755-250) at 10X final desired assay concentration. The following PRR ligands were prepared: PAM3CYSK4 (PAM; stimulates TLR2; Invivogen Cat. No. tlrl-pms) at 1 mg/mL, polyinosinic-polycytidylic acid (Poly I:C, stimulates TLR3; GEHealthcare Cat. No. 27473201) at 100 mg/mL, Lipopolysaccharide (LPS, stimulates TLR4; Invivogen Cat. No. tlrl-pelps) at 10 ng/mL, resiquimod (R848;
stimulates TLR 7:8; Invivogen Cat. No. tlrl-r848) at 10 mM, Peptidoglycan (PGN, stimulates TLR2 and nucleotide-binding oligomerization domain-containing protein 1/2 [NOD1/2]; Invivogen Cat. No. tlrl-pgnsa) at 10 mg/mL, and RPMI 1640 media (RPMI; Gibco Cat. No. 61870) alone. 22 µL of each PRR ligand or media alone was dispensed into recipient 96-well round-bottom polystyrene plates (Corning, Cat. No 7007) using the Evolution P3 Precision Pipetting Platform (Perkin Elmer Cat. No. EP300011). Recipient plates were sealed with sterile aluminum plate sealers and frozen at -80°C and shipped to the other three study sites on dry ice under temperature-monitored conditions via World Courier Inc.

**Blood collection.** Blood was drawn from participants by a phlebotomist during the study visit, taking place at the hospital locations described above. Majority of samples were collected from the arm, and some from the neck. 3-5 mL blood was drawn from each participant using sterile venipuncture into sodium-heparin vacutainers (Becton Dickinson Cat. No. 8019839). Once drawn, blood samples were kept at room temperature and processed within 24 hours.

**PRR stimulation.** To eliminate processing artefacts due to various individuals conducting stimulation assays, the same person (Dr. Kinga Smolen) performed the PRR stimulation assay at each study site. Whole blood was diluted 1:1 with sterile RPMI 1640 pre-warmed to 37°C and 200 µL of diluted blood was added to each well of pre-made 96-well plates containing 22 µL PRR ligands described above. Stimulation plates were incubated at 37°C for 24 hours, and then centrifuged at 600 g for 5 minutes. After centrifugation, 100 µL supernatant was removed and frozen at -80°C for later multiplexing analysis. Samples from all sites outside of Vancouver,
Canada were shipped to the Vancouver processing laboratory on dry ice under temperature monitored conditions by World Courier Inc.

**Cytokine measurement.** Supernatants were thawed at room temperature. The concentration of cytokines in supernatants were measured using the Luminex Flex Kit system using the high-biotin manufacturer’s protocol with an overnight incubation at 4°C. The following cytokines and chemokines were measured: IFN-α2, IFN-γ, IP-10, IL-12p70, IL-12p40, IL-6, TNF-α, IL-1β, IL-8, MIP-1α, MIP-1β, and IL-10. Supernatants were diluted either 1:1, 1:20, 1:80, or 1:150 with RPMI 1640 medium. Beadlytes, biotin, and streptavidin-phycoerythrin were used at half the manufacturer’s recommended concentrations after ensuring that diluting these reagents did not affect the concentration of measured analytes. Assays were then read using the Luminex 100 Total System. Raw MFI data were analyzed using either the Bio-plex (Bio-Rad Cat. No. 171001513) or MasterPlex (MiraiBio) software. 5-parameter logistic fits were used to construct standard curves for each analyte and calculate the sample concentrations. Samples with fewer than 50 beads counted were excluded from further analysis.

**IL-23 ELISA.** As IL-23 bead analytes were not available for the Luminex assay, IL-23 supernatant concentrations were determined via Human IL-23 (p19/p40) ELISA (eBioscience Cat. No. 88-7237), whereby supernatants were diluted 1:4 in diluent supplied in the kit. The remainder of the protocol was performed according to manufacturer’s instructions. Plates were read at 450 nm with 570 nm subtraction on a SPECTRAmax Plus machine. 4-parameters logistic curves were used to generate the standard curve and determine sample concentrations.
Infant stool microbiome analysis

Stool sample collection. Parents were instructed to collect soiled diapers at home into clean containers at keep at 4°C. Samples were delivered to the laboratory within 24 hours of the bowel movement. At the laboratory, samples were aliquoted into 2-ml screw-top tubes except for samples from Vancouver, which were stored in 50-ml conical tubes. All aliquots were stored at -80°C within 24 hours of sample receipt. Stool samples from Brussels, Cape Town, and Quininde were shipped to the central study site (Vancouver, Canada) on dry ice via World Courier, Inc. with a temperature monitor to ensure all samples remained below -80°C during transport. Samples were stored at -80°C in Vancouver for no longer than 24 months prior to DNA extraction.

Stool DNA extraction. Total DNA was extracted from 180 – 220 mg stool using the QIAamp DNA stool mini kit (Qiagen Cat. No. 51504) using the manufacturer’s protocol for isolation of DNA from stool for pathogen detection. The following modifications were made: all heated incubations were at 95°C. For step two of the protocol, stool samples were homogenized by adding 1.4 ml ASL buffer to each sample and placing in the disruptor Genie (Scientific Industries Cat. No. SI-DD238) for 2-3 minutes or until stool was thoroughly homogenized.

16S amplicon sequencing. PCR amplification and sequencing closely followed previously described protocols and rationale. Amplification was carried out with between 2-5 ng template DNA, GoTaq hot start 2x colorless master mix from Promega (Cat. No. M5131) plus the following barcoded primers, targeting the V6 region of the 16S rRNA gene, at 0.8 pm/µL.
Forward primer:
5’ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnn[BC8mer]CWACGCGARGAACC TTACC3’

Reverse primer:
5’CGGTCTCGGCATTTCTGCTGAACCGCTCTTCCGATCTnnnn[BC8mer]ACRACACGAG CTGACGAC3’

Cycling conditions included 1 minute activation of the Taq at 95°C, followed by 25 cycles of 95°C, 55°C, 72°C at one minute each. Amplified samples were quantified using a QBit and the dsDNA reagent and pooled at equimolar amounts. Pooled samples were purified on a PCR cleanup column, and the eluted library was diluted 1:100 in deionized water and amplified using the following primers in order to attach Illumina adapters.

OLJ139:
5’AATGATACGGCGACCACCGAGATCTACACTCTTTTCCTACACGA3’

OLJ140:
5’CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAAC3’

The PCR was as above, except with an annealing temperature of 60°C and 10 cycles. The product was purified and 10% phiX174 DNA was added. This library was loaded onto an Illumina MiSeq instrument with a paired-end 2x100 cycle format.

16S rRNA gene sequence analysis. Paired end reads were assembled using XORRO172, and only sequences without ambiguous bases were retained for further analysis. Assembled sequences were quality-filtered and binned into OTUs with UPARSE v7.1173. Briefly, assembled reads were abundance-sorted and clustered into 97% similarity OTUs. Representative sequences
(excluding singletons) from each OTU were further chimera-filtered against the gold database (http://drive5.com/uchime/gold.fa). Reads were then mapped back to OTUs and converted into an OTU table.

Germ-free mouse model of human fecal transplantation

Mice. 3-week-old germ-free male Swiss-Webster mice were purchased from Jackson laboratories. Mice were maintained in a specific-pathogen-free environment at the University of British Columbia pharmaceutical sciences animal care facility.

Fecal transplantation. Three Canadian and three South African male stool samples were selected for fecal transplantation. Samples from male infants were selected to match the male sex of recipient mice. The samples selected belonged to subjects without outlying IFNγ and TNFα cytokine concentrations in response to LPS and R848 among other infants in their respective cohorts, as outliers were deemed to be of highest risk of assay error. No other criteria were used to select fecal samples. Each sample was used to gavage three mice. Mice receiving gavage from the same samples were co-housed. Stool samples were prepared by diluting 100 mg of stool into 10 mL of reduced PBS (PBS + 0.05% cysteine), in an anaerobic chamber. Mice were then gavaged with 100 uL of fecal homogenate every second day for 3 days, starting immediately upon arrival to the animal care facility. Mice were maintained for 21 days and then sacrificed.

Lactulose:mannitol ratio. To assess the gut barrier integrity of experimental animals, six mice per experimental group were randomly selected 24 hours prior to sacrifice. Details for this assay
were previously published\textsuperscript{174}. Briefly, all mice were gavaged with a sugar probe solution of 100 mg sucrose, 12 mg lactulose, 8 mg mannitol, and 6 mg sucralose dissolved in 0.2 mL. Mice were then immediately placed in metabolic cages with no food for 22 hours. Urine was collected during this time and prepared for high-performance liquid chromatography (HPLC) according to a published protocol\textsuperscript{175}.

**Splenocyte TLR stimulation.** TLR stimulation of mouse splenocytes used an adaptation of the protocol used for the human whole blood stimulation assay. Sterile 96-well plates (Corning Cat. No. 7007) were prepared with 22 µL of RPMI-1640 medium (Gibco Cat. No. 61870) with TLR agonists at 10-times their final concentration. The TLR agonists used were LPS (Invivogen; Cat. No. tlr1-pelps) and R848 (Invivogen Cat. No. tlr1-r848-5), both at a final concentration of 0.1 µg/mL. At the time of sacrifice, spleens were immediately removed and stored in 5 ml sterile RPMI-1640 medium. Spleens were then disrupted using the frosted glass slide method, passed through a 70-µm filter, washed with 10 ml sterile Dulbecco’s phosphate-buffered saline (dPBS; Gibco Cat. No. 14190), and centrifuged at 1400 rpm for 5 minutes at 4°C. Cell pellets were suspended in 4 ml red blood cell lysis buffer and incubated at room temperature for 4 minutes. 10 mL RPMI-1640 was then added to neutralize the RBL. Cells were centrifuged at 1400 rpm for 5 minutes at 4°C, the supernatant was discarded, and the pellet was suspended in 10 mL sterile R10 medium (RPMI-1640 supplemented with 10% fetal bovine serum). 10\textsuperscript{6} splenocytes in a total volume of 200 µL were added to each well containing either TLR agonists or RPMI media alone. Plates were incubated at 37°C for 24 hrs. Following incubation, cell cultures were centrifuged at 250 x g and 130 µL supernatant stored at -80°C until later use.
**Luminex Assay.** The eBioscience procartaplex mouse basic lumiex kit (eBioscience Cat. No. EPX010-20440-901) was used to measure cytokine concentrations in cell supernatants, including the analytes IFN-α (eBioscience Cat. No. 26027), IFN-γ (eBioscience Cat. No. 20606), IL-10 (eBioscience Cat. No. 20614), IL-23p19 (eBioscience Cat. No. 26017), IL-6 (eBioscience Cat. No. 20603), MIP-1β (eBioscience Cat. No. 26014), and TNF-α (eBioscience Cat. No. 20607). Supernatants were diluted 1:1 with sterile R10 medium prior to Luminex analysis using the manufacturer’s vacuum-manifold protocol. Data were acquired on the Luminex 100 analyzer. Raw MFI data were analyzed using the MasterPlex software. 5-parameter logistic fits were used to construct standard curves for each analyte and calculate the sample concentrations. Samples with fewer than 50 beads counted were excluded from further analysis.

**Mouse Fecal and intestinal tissue DNA extraction.** Immediately after sacrifice, the mice were sacrificed, one stool pellet and a 10 – 30 µg section of both ileum and jejunum tissue were collected from each mouse and stored in sterile 1.5 mL microcentrifuge tubes. Samples were stored at -80°C until DNA extraction. Total DNA from each sample was extracted using the Qiagen Fast DNA stool mini kit (Qiagen Cat. No. 51604) following the protocol for stool pathogen detection with the following modifications: To disrupt the stool and intestinal samples, they were placed in a 2.0-mL Eppendorf tube with one sterile 5-mm metal bead (Qiagen Cat. No. NC9257481) and 1 mL Inhibitex buffer. Samples were then homogenized using a Disruptor Genie for 2 minutes or until samples were thoroughly homogenized. The resulting suspension was heated at 95°C for 5 minutes.
16S amplicon sequencing of mouse feces and tissues and selected human samples

**PCR amplification.** To analyze the microbial communities in mouse tissues and feces, alongside re-extracted human stool used for the experiment, we used amplicon sequencing targeting the V4 region of the 16S rRNA gene. PCR amplification was done using 1-10 ng of DNA, Thermo Phusion Hot Start II DNA Polymerase (ThermoFisher Cat. No. F549S), and the following primers:

16Sf V4: GTGCCAGCMGCGCGGTAA
16Sr V4: GGACTACHVGGGTWTCTAAT

PCR cycle conditions were 98° C for 2 minutes, followed by 30 cycles of 98° C for 20 seconds, 55° C for 15 seconds, and 72° C for 30 seconds, and 72° C for 10 minutes following the cycling steps. PCR products were cleaned using the Agencourt Ampure XP beads (Beckman Coulter Cat. No. A6388) using a 0.8:1 bead-sample ratio and eluted in a final volume of 20 µL.

**Library preparation.** Library preparation was performed using a previously published SOP with the details and product information outlined below. 10 µl of the final product was used to normalize using the SepalPrep Normalization Prep Plate Kit (ThermoFisher Cat. No. A1051001) to 1-2 ng/µL and 5 µL of each normalized sample was pooled into a single library per 96-well plate. Library pools were further concentrated using the DNA Clean and Concentrator kit (Zymo Cat. No. D4013). A dilution series was performed for each of the pooled libraries for subsequent quality control steps. Each pool was analyzed using the Agilent Bioanalyzer using the High Sensitivity DS DNA assay (Agilent Cat. No. 5047-4626) to determine approximate fragment size, and to verify library integrity. Library pools with unintended amplicons were purified using the Qiagen QIAquick Gel Extraction Kit (Qiagen Cat. No. 28706). Pooled library concentrations
were determined using the KAPA Library Quantification Kit for Illumina (KAPA Cat. No KK4824). The final libraries were loaded at 8 pM, with an additional PhiX spike-in of 20%. The amplicon library was sequenced on the MiSeq using the MiSeq 500 Cycle V2 Reagent Kit (Illumina Cat. No. MS-102-2003).

**Read processing and OTU binning.** Paired-end reads were assembled using MOTHUR version 1.37.2, following the MiSeq SOP\textsuperscript{176} [accessed on Apr 2017]. OTUs were clustered at 97% identity and classified using the GreenGenes 13_8_99 database\textsuperscript{177}.

**Statistical analysis**

**Alpha diversity.** We estimated microbial diversity with abundance-dependent (Shannon) metrics, after subsampling the OTU table to account for unequal sampling depth. To test for relationships between host factors and Shannon diversity, we used linear regression. To test for a quadratic relationship between Shannon diversity and maternal age in the combined Ecuadorean and Canadian cohorts we fitted a linear model with B-spline transformed maternal age, using 5 knots and setting degrees to two (quadratic relationship) using the R package splines implemented through base R\textsuperscript{178}.

**Community composition.** We used the phyloseq R package\textsuperscript{179} to compute alpha and beta diversity. To identify subject characteristics potentially explaining multivariate community composition, we used the forward model selection algorithm of implemented in the R package vegan (function \texttt{ordiR2step})\textsuperscript{180}. This procedure minimizes Type I errors and corrects for the overestimation of the proportion of explained variance by selected variables. The significance (\(\alpha\)
of variables was evaluated stepwise and the order of variable evaluation was based on improvement in the model’s adjusted $R^2$. Model selection proceeded until the next independent variable was non-significant as determined by 1,000 permutations. Missing values in the metadata were imputed via random forest analysis using the *missForest* R package.

**OTU pre-filtering.** Following analysis of alpha and beta-diversity, data were further filtered to remove rare OTUs using two criteria. First, we removed all near zero variance OTUs using the `nearZeroVar` function in `mixOmics` with the frequency cutoff set to 90/10 and unique cutoff set to 10. Next, we removed rare OTUs by retaining only OTUs that had counts above 3 in at least 5% of samples. After filtering, 1076 OTUs were retained for subsequent analyses.

**Differential Abundance.** To identify OTUs whose normalized abundance significantly differed between cohorts, we used the likelihood ratio test implemented in the R *DESeq2* package.

**Identifying cohort-specific OTUs.** To further characterize OTUs that classified each cohort in a multivariate space, we used Partial Least Squares Discriminant Analysis (PLS-DA) of variance-stabilized counts of the differentially abundant OTUs. Differentially abundant OTUs were graphically represented by heatmaps using the *pheatmap* R package.

**Cytokine response signatures among cohorts** Because baseline, or unstimulated data cannot be incorporated into sPLS-DA, we selected only cytokines whose variance significantly differed from unstimulated control values using the Fligner-Killeen test. For baseline values, we selected only cytokines where at least 70% of subjects had unstimulated values above the detection limit,
with a median fluorescence intensity (MFI) greater than 10. The resulting cytokine-stimulus combinations were used to create a matrix of concentration values. The data were log10-normalized. Subjects with missing values for 15% or more features and cytokines with missing values for 15% or more subjects were removed. The remaining missing values were imputed using the *missForest* R package\textsuperscript{181}. Cytokine responses to PRR agonists that characterized each cohort in a multivariate space were selected using sPLS-DA.

**Preparation of OTU data for integration.** The OTU data was first total sum of squares (TSS) transformed. The centered-log-ratio (CLR) transformation was then applied to normalize OTU data. Because CLR is a log-transformation, it cannot accommodate zeroes. Thus, zero values were coded as 0.1 times the minimum relative abundance in each sample, the default offset provided within the mixOmics version 6.0 CLR function\textsuperscript{182}.

**sPLS Integration.** To uncover potential gut microbiome–host immune interactions, we examined the joint multivariate structure of gut microbiota composition and host innate immune responses via sparse partial least squares (sPLS) analysis, a method that incorporates variable selection, making it particularly suitable to high-dimensional data-sets\textsuperscript{184}. We used this approach to visualize associations between selected OTUs and extracellular cytokines produced in response to specific ligands. This analysis was conducted with the goal of identifying both interactions that were robust among cohorts and interactions specific to individual cohorts, under the hypothesis that while global pattern of microbiome-immune interactions may exist, the microbiomes and immune phenotypes specific to each cohort likely have unique interactions. To this end, separate analyses were done for all cohorts combined and for each cohort separately.
To better suit the regression framework of sPLS, we limited our analysis to OTUs present in at least 50% of the samples included in each analysis. The number of features selected per component was set to a maximum of the number of samples in the dataset, and further reduced by limiting the selection to features that significantly correlated using the Pearson correlation to at least 30% of the features in the other data type. Cytokine-OTU relationships were displayed using heatmaps generated with the *pheatmap* R package. Networks were generated using the *igraph* R package\(^\text{186}\) and exported as *graphml* objects for further editing in Cytoscape\(^\text{187}\).

### 2.3 Results

#### 2.3.1 Cohort characteristics

Cohort characteristics for the study participants included in the immune analysis can be found elsewhere\(^\text{167}\). For microbiome analysis, stool samples were collected from 17 Belgian, 32 Canadian, 41 Ecuadorean, and 8 South African children. Baseline characteristics differed between the two sites (Table 2.1). Notably among the Belgians, one of 17 study participants was female while sex was more balanced in the other cohorts. Belgians were predominantly vaginally delivered (19/21; 91%), and South Africans were exclusively vaginally delivered, meanwhile 50% and 19% of children were delivered via Caesarean section in Canadian and Ecuadorean cohorts, respectively. Other differences included younger age of recruited Canadians (mean 1.65 years, while recruitment goals were for two years of age) and differences in anthropometric measurements between cohort.
While these cohorts represent geographically distinct populations, ethnic heterogeneity varied (Table 2.2). Belgians were predominantly of African and Arab origin. Canadians were a heterogeneous cohort of white-Caucasians (63%), Latin American (9%), Chinese (9%), Filipino (3%), or mixed-race (16%). Ecuadoreans were exclusively Latin American. South Africans were mostly mixed-race (63%) and the remainder either African (Xhosa) with one white-Caucasian.

All children were born and raised at the sites of sample collection, except for a subset of the Belgian children (7%) who were born in African countries, or in one case, Germany. Immune data for these infants born outside of Belgium were not available.
Table 2.2 Ethnic background across cohorts

<table>
<thead>
<tr>
<th></th>
<th>Belgium</th>
<th>Canada</th>
<th>Ecuador</th>
<th>South Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td>4 (23.53%)</td>
<td>Asian/White-Caucasian</td>
<td>3 (9.38%)</td>
<td>African (Xhosa) 2 (25.00%)</td>
</tr>
<tr>
<td>African/Arab</td>
<td>1 (5.88%)</td>
<td>Caucasian/Filipino</td>
<td>1 (3.12%)</td>
<td>Mixed-race 5 (62.50%)</td>
</tr>
<tr>
<td>Arab</td>
<td>6 (35.29%)</td>
<td>Caucasian/South Asian</td>
<td>1 (3.12%)</td>
<td>White-Caucasian 1 (12.50%)</td>
</tr>
<tr>
<td>Black</td>
<td>3 (17.65%)</td>
<td>Chinese</td>
<td>3 (9.38%)</td>
<td></td>
</tr>
<tr>
<td>Black/Arab</td>
<td>1 (5.88%)</td>
<td>Filipino</td>
<td>1 (3.12%)</td>
<td></td>
</tr>
<tr>
<td>White-Caucasian</td>
<td>1 (5.88%)</td>
<td>Latin American</td>
<td>3 (9.38%)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (5.88%)</td>
<td>White-Caucasian</td>
<td>20 (62.50%)</td>
<td></td>
</tr>
</tbody>
</table>

2.3.2 Microbiome composition across four cohorts

The 16S amplicon libraries yielded a total of 4030 OTUs after quality filtering and binning at a 97% similarity threshold. Sequencing depth did not differ significantly among cohorts. We first sought to understand whether there was a difference in alpha diversity between the four cohorts using the Shannon diversity index. Alpha diversity of fecal microbiota did not differ among cohorts (Figure 2.1A). We conducted an exploratory non-metric multidimensional scaling analysis (NMDS) of pair-wise Bray-Curtis distances to assess whether fecal microbiome composition segregated by cohort. We found that the microbiomes differed substantially between high-resource and low-resource populations (Figure 2.1B). Of note, Belgians were distributed across both groups, with the African-born sub-group (comprised of subjects born in different African countries) clustering with South Africans and Ecuadorians (Figure 2.1C). Forward selection-based analysis of the contribution of demographic variables (sex, delivery mode, gestational age, maternal age, and anthropometric measurements) to the explanation of community composition produced a final model containing cohort (country of origin) as the sole variable of importance (adjusted-R² = 0.11, p = 0.002) and no other host factors contributed.
significantly to community composition. Because the effect of cohort could potentially mask any effect of host factors within each cohort, we performed distance-based redundancy analysis (dbRDA) and ordiR2step for these variables in each cohort separately, where applicable (e.g., as there was only one female in the Belgian cohort, sex was omitted from the Belgian analysis). We detected a significant effect of WLZ on community composition in Canadian and South African children (Canada $R^2 = 0.031$, $p = 0.01$; South Africa $R^2 = 0.17$, $p = 0.012$). No other host factors were significant for any other cohort. Thus, the host factors measured did not have major associations with microbiome composition.
Figure 2.1 Microbiome alpha and beta diversity across cohorts
A. Alpha diversity did not differ among cohorts. B. NMDS scaling reveals a separation of industrialized and non-industrialized countries along NMDS1. C. Belgian children born in African countries tend to group with the non-industrialized cluster.
2.3.3 Alpha diversity and host factors within each cohort

Within each cohort, we tested whether any host demographic factors were correlated with Shannon diversity with simple linear regression. Significant variables were then included in a multiple linear regression. We evaluated the relationship between delivery mode and diversity in Canadian and Ecuadorean cohorts with sufficient sample size. Delivery mode correlated with diversity in the Canadian cohort (Figure 2.2A). Maternal age correlated with diversity in Ecuadorean and Canadian cohorts (Figure 2.2B). Because Canadian mothers who delivered by C-section were older than those that delivered vaginally, we used multiple linear regression, where maternal age remained significant while delivery mode was no longer significantly associated with diversity. Interestingly, maternal age was positively correlated with diversity in Ecuadoreans, but negatively correlated in Canadians (Figure 2.2B). Once data from both cohorts were combined, the younger maternal age range of Ecuadoreans and older range of the Canadians revealed a significant quadratic relationship (method; p < 0.05, R2 = 0.09), with both the youngest and oldest mothers having children with lower diversity (Figure 2.2D). Gestational age was also associated with diversity in the Ecuadorean cohort, which remained significant in multiple regression (Figure 2.2D).
Figure 2.2 Host factors associated with alpha diversity
A. Gestational age correlated positively with Shannon diversity among Ecuadoreans. B. In unadjusted analysis, Canadian Caesarean-delivered children had a significantly lower microbial richness compared to Vaginally-delivered children. C. Maternal age significantly impacted microbial richness in both Ecuadorean and Canadian cohorts, but with opposite directionality. D. Because the maternal age spectrum minimally overlapped between Canadian and Ecuadorean cohorts, combining both these cohorts reveals a significant quadratic relationship between richness and maternal age with both very young and very old maternal age being associated with reduced richness (p = 0.0137, linear model fitted with basic quadratic spline).
2.3.4 Microbiome taxonomic composition and differential abundance

Taxonomic composition of microorganisms reflected commonly identified human taxa, with *Prevotellaceae, Bacteroidaceae, Lachnospiraceae, Ruminococcaceae, Vellionellaceae,* and *Enterobacteriaceae* being the top 5 abundant families (Figure 2.3A). Most individuals were dominated by either *Prevotellaceae* or *Bacteroidaceae.* We tested for associations between abundances of individual OTUs and cohort membership with the DESeq2 likelihood ratio test and found 442 OTUs to differ significantly, using an adjusted p-value cutoff of 0.05. *Prevotella* OTUs (Bacteroidetes phylum) were the most represented among those selected by DESeq2. While selecting OTUs that differed between any cohort, this analytical approach did not reveal a clear microbiome signature between cohorts. To better characterize how differentially abundant OTUs discriminated between cohorts, OTUs in PLS-DA analysis to identify the most discriminatory OTUs in a multivariate space. The top 50 OTUs selected by PLS-DA to discriminate cohorts separated subjects from high-resource origins (Canadian and European-born Belgians) from those from low-resource origins (Ecuadoreans, South Africans, and African-born Belgians) (Figure 2.3B). These OTUs did not separate South Africans from Ecuadoreans or Belgians from any other cohort. Most of the selected OTUs were enriched in the low-resource origin cluster and were almost exclusively members of the *Prevotella* genus. The smaller subset of OTUs enriched in the high-resource origin cluster included a diverse range of genera, mostly belonging to the *Firmicutes* phylum (including *Ruminococcus, Clostridia,* and unclassified *Firmicutes*). Interestingly, *Bacteroides* OTUs were not discriminatory between cohorts, as this genus was commonly found in Ecuadorean infants, some with relative abundances comparable to those in Canadians. However, *Prevotella* OTUs were very rare in Canadians, and only present at high abundances in 3 of 32 cases.
Figure 2.3 Taxonomic differences across cohorts
A. Top 10 most abundant families ordered by relative abundance of Bacteroidaceae in individual study participants. B. Top 50 taxa differentially abundant across all cohorts determined via DeSeq2 analysis (adjusted p < 0.01) and further selected by multivariate PLS-DA analysis.
2.3.5 Immune responses associated with each cohort

Cytokine responses to PRR stimulation were published previously. Most notable using univariate tests and PCA, children from South Africa profoundly under-responded to every PRR agonist except for PGN (of TLR2 and NOD1/2). However, more subtle differences existed among the other cohorts. Thus, we conducted sPLS-DA analysis to identify cytokine response signatures specific to Belgians, Canadians, and Ecuadoreans which were separate children from the three sites based in their cytokine responses (Figure 2.4A). Except for Belgians, each cohort was classified with an error rate of 10% or less with two sPLS-DA components (Figure 2.4B).

![Figure 2.4 Cytokine signatures across Belgian, Canadian, and Ecuadorean cohorts](image)

A. sPLS-DA ordination of Cohort-defining cytokines separates Canada along x-variate 1 and Belgium along x-variate 2. B. Maximum error rates overall and per class shows that BLG, CAD, and ECD are classified with an error rate of 18%, 10%, and 0% respectively, with a minimum overall error rate of 6.8%.

Canadians were classified by their lower responses to PAM (TLR2) stimulation compared to Ecuadoreans and Belgians (Figure 2.5A), while Belgians had lower responses to endosomal PRR stimulants, Poly I:C (TLR3) and R848 (TLR7:8) with select cytokines (Figure 2.5C). Cytokine-specific trends were also identified. Canadians were further characterized by significantly lower
IL-6, IP-10, and IL-12p40 production in response to select ligands and higher IL-12p40 responses to PGN and LPS compared to the other cohorts (Figure 2.5A). Belgians mounted a higher IL-10 response to every PRR agonist (Figure 2.5B). Ecuadoreans were classified uniquely by not having lower or higher responses to any PRR stimulation, and were thus accurately identified by their exclusion from each other cohort. All cytokines selected by sPLS-DA were also significantly different between cohorts in previous univariate analysis$^{167}$. 
Figure 2.5 Discriminatory cytokine features for Belgian and Canadian cohorts
A-B. Discriminatory features in component 1 with lower responses (C) and higher responses (D) in Canadians. C-D. Discriminatory features in component 2 with lower responses (E) and higher responses (F) in Belgians.
2.3.6 Correlations between TLR responsiveness and the gut microbiome

2.3.6.1 Correlations across cohorts

Initial integration of OTU and cytokine data using all available subjects yielded a very poor correlation structure between the two datasets (not shown). Because there were no significant changes between South African and Ecuadorean infant microbiomes, yet South African immune profiles were highly distinct from all the cohorts, we hypothesized that the extreme immune phenotype of this cohort resulted in poor agreement among all the cohorts. When the South African subjects were removed from the analysis, co-varying OTUs and cytokines were selected along the first component (Figure 2.6).

![Figure 2.6 Cytokine-microbiome correlations among Belgian, Canadian, and Ecuadorean children](image)

A. Correlation circle plot showing OTUs and cytokines selected by sPLS with significant network correlations. B. Network showing correlations between selected OTUs and cytokines. Colours denote Pearson correlation strength and nodes are sized according to number of connections.
The selected features from both datasets were dominated by negative correlations between *Bacteroides* OTUs and cytokine responses to PAM (TLR2) stimulation, and positive correlations between *Prevotella* OTUs and those responses (Figure 2.7A). These associations remained significant across all three cohorts (Figure 2.7B) in the Ecuadorean cohort alone (Figure 2.7C). IL-6, IL-8, and IP-10 responses to PGN (TLR2, NOD1/2) and MIP-1α and MIP-1β responses to endosomal TLR stimulation followed the same pattern. IL-23 responses to both PGN and LPS were selected for their distinct relationships to the selected OTUs, correlating negatively with *Prevotella* but not with *Bacteroides*.

![Figure 2.7](image.png)

**Figure 2.7 Heatmap and select cytokine-microbiome correlations among Belgian, Canadian, and Ecuadorean children**

A. Correlation circle plot showing OTUs and cytokines selected by sPLS with significant network correlations. B-C. Heatmap (C) and network (D) showing correlations between selected OTUs and cytokines. Colours denote Pearson correlation strength and nodes are sized according to number of connections. D-E. Selected features include negative correlations between *Bacteroides* OTU_1 and cytokine responses to PAM3Cys stimulation (IL-10 and IP-10) and
positive correlations between Prevotella OTU_2215 and those responses in Belgium, Canada, and Ecuador (D) and Ecuador only (E).

2.3.6.2 Correlations within cohorts

For the Belgian cohort, the first two selected sPLS components yielded networks of OTU-cytokine correlations. The first component again selected predominantly cytokine responses to PAM stimulation and to R848 (TLR7:8) stimulation, all negatively correlated to *Firmicutes*, including *Lachnospiraceae* and *Oscillospira* (Figure 2.8A). The second component selected mostly *Firmicutes*, including *Clostridia* and *Ruminococcus*, having positive correlations with responses to LPS (TLR4) stimulation and negative correlations with IL-8 and MIP-1β in response to multiple ligands.

The Canadian cohort also contained network structures along the first two sPLS components. The first component was dominated by pro-inflammatory and Th17 responses (including IL-1β, TNF-α, IL-23 and IL-12p40) to LPS and Th1-supporting responses (including IFN-γ, IL-12p40, and IL-12p70) to R848 (Figure 2.8B). All these responses all were negatively correlated with several *Lachnospiraceae* OTUs and positively correlated with *Bacteroides* OTUs. The 2nd component was again dominated by cytokine responses to PAM having a positive correlated with several *Firmicutes*, again including *Lachnospiraceae*

Correlations of *Prevotella, Bacteroides* OTUs with cytokine responses to PAM were also evident within the Ecuadorean cohort (Figure 2.8C), where those OTUs were additionally correlated with cytokine responses to PGN. In addition to the above OTUs, several belonging to the *Lachnospiraceae* family had negative correlations to cytokine responses to PAM and PGN. Features selected along the 2nd component were largely driven by one individual and yielded no significant correlations among the selected features. However, component 3 selected a smaller
subset of OTUs, predominantly Firmicutes and Proteobacteria, having negative correlations with MIP-1α and MIP-1β responses to several ligands.

While feature selection for the South African cohort was limited to 7 per component, due to low sample size, 3 sPLS components selected features with correlation structure (Figure 2.8D). Unlike for all the other cohorts, no cytokine responses to PAM were selected. Component 1 was

---

**Figure 2.8 Cohort-specific cytokine-microbiome relationships**
Heatmaps illustrating Pearson correlations between sPLS selected OTUs and cytokines in Belgium (A), Canada (B), Ecuador (C), and South Africa (D). Colours denote Pearson correlation strength. Non-significant correlations are shown in white.
instead dominated by cytokine responses to PGN correlating negatively to a group of *Firmicutes*, mostly unclassified past the genus level. Component 2 was dominated by baseline levels of IFN-α and IL-10. Baseline levels of IL-8 correlated also positively with a subset of *Firmicutes*, while IL-8 in responses to all other stimuli correlated negatively with the same subset. Component 3 selected TNF-α and IL-10 responses to LPS having negative correlations to *Firmicutes*, including *Lachnospiraceae*, as well as MIP-1α and MIP-1β responses to R848 having positive correlations with the same OTUs.

### 2.3.7 Fecal transplant affects immune phenotype of germ-free mice

We tested whether the host microbiota could modulate systemic immunity using a germ-free mouse model. South African children had the most divergent immune phenotype and contained the most distinct microbiota-immune relationships, but the integrated correlation analysis did not reveal any correlations between the host microbiota and the immune suppression in these children. This motivated the comparison of South African versus Canadian microbiomes (chosen for both distinct microbiome composition and sample availability) for their potential effects on splenocyte responses to TLR stimulation in germ-free mice (Figure 2.9A). Three weeks after the final gavage, a subset of mice were tested for gut barrier integrity. The following day, all animals were sacrificed, splenocytes harvested, and separately stimulated with LPS (TLR4) or R848 (TLR7:8), chosen because South Africans demonstrated the most profoundly low responses to these stimulants. In mice inoculated with South African feces versus those inoculated with Canadian feces, IFN-γ and IL-10 responses to LPS and the IL-10 response to R848 were significantly lower (Figure 2.9B-C). And, the IL-6 response to R848 showed a similar trend (*p* = 0.06). Responses to pro-inflammatory cytokines TNF-α and MIP-1β were
similar between the groups, while IL-23 and IFN-α2 were not produced in this assay. Overall, mice inoculated with South African feces mounted lower cytokine responses than those inoculated with Canadian feces.

**Figure 2.9** Germ-free mouse models recapitulated immune differences between Canadian and South African children
A. Schematic of experimental design. B. Splenocyte cytokine responses to LPS (TLR4 stimulation). C. Responses to R848 (TLR7:8) stimulation. *(p < 0.05, Wilcoxon test; unadjusted). D. Single NMDS showing similarity of microbiomes in mouse feces, ileum, jejunum and human feces. E. Individual NMDS showing similarity of microbiomes in mouse feces, ileum, and jejunum. F. Lactulose-mannitol ratios of SAF and CAD mice. **(p < 0.01, student’s t-test).

Mouse microbiomes at the end of the experiment were assessed in feces, ileum, and jejunum samples using 16S amplicon sequencing. Human samples used to inoculate the mice
were re-extracted and sequenced in parallel. The microbiota engrafted in the mice was distinct in the feces versus the ileum and jejunum (Figure 2.9D). The microbiomes of the mice were distinct from their respective source feces. Still, within each tissue and the feces, the microbiomes were distinct between mice with the different fecal sources (Figure 2.9E). Further, the small intestinal barrier integrity was significantly lower between mice inoculated with South African feces versus those inoculated with Canadian feces (Figure 2.9F).

To identify differentially-abundant OTUs between mice inoculated with Canadian versus South African feces, we built a DESeq2 likelihood ratio model adjusting for the donor, thus ensuring that selected OTUs were not enriched in only one cage. A subset of the differentially abundant OTUs in the original human samples were identified in the mice (Figure 2.10A-C), and some overlapped between the intestinal regions and feces. Of the OTUs differentially abundant in both the human donors and the mice, six consistently distinguished the treatment groups in the ileum, jejunum and feces. Among these, four OTUs belonging to the Alistipes, Odoribacter, and Prevotella genera and one to the Rikenellaceae family were enriched in mice inoculated with South African feces and nearly absent in those inoculated with Canadian feces (Figure 2.10D). Furthermore, these OTUs were detected in only South African human donor stools. Clostridium OTUs were more abundant in the mice inoculated with Canadian feces, but were also detected in the other treatment (Figure 2.10E). The remainder of differentially abundant OTUs were not detected in the human stools.
Figure 2.10 Differentially abundant OTUs between mice gavaged with Canadian vs. South African feces
A-C. Differentially abundant OTUs between CAD versus SAF inoculated mice in the jejunum (A), ileum (B), and feces (C). Four OTUs enriched in SAF inoculated mice were also present in only in South Africans (D). Two OTUs enriched in CAD inoculated mice were also present in humans (E).

2.4 Discussion

The goal of this study was to identify host microbiome and systemic immune responses both within and across diverse populations. Using multi-omic integration, we identified relationships between cytokine responses to TLR2 and relative abundance of *Prevotella* and *Bacteroides* taxa in stool. Further, responses to TLR2 featured in cohort-specific integrations,
except for South African children. However, fecal gavage of stools from South Africans resulted in an altered immune phenotype in gavaged mice consistent with changes seen in the humans. Together, these findings suggest that universal host-microbiome relationships can exist even across diverse populations, and germ-free mouse models of human fecal transplantation can be used to dissect these further.

We found that infant microbiomes differed primarily between low- and high-resource regions of birth and were not distinct among all four cohorts. South African microbiomes did not differ from Ecuadorean ones, despite the environmental differences between the two countries. Ecuadoreans represented a more homogeneous ethnic group sampled within one rural village, meanwhile South Africans were a racially diverse group sampled within resource-poor areas of Cape Town, an urban region in South Africa. The similarity of microbiomes of African-born Belgians to those of Ecuadorian and South Africans further suggests that resource availability, and therefore diet of the region of birth is an important driver of gut microbiome composition. This composition may be maintained even when the infants move to a different region at a very early age. Microbiomes in low-resource regions were associated with increased abundance of *Prevotella*, while those in high-resource regions were enriched in a smaller repertoire of *Firmicutes*. Microbiomes dominated by *Bacteroides* were more common in, but not exclusive to, high-resource regions. This general trend is consistent with previous microbiome surveys of diverse populations. American microbiomes were distinct from rural Malawian and Venezuelan microbiomes, while the latter two were not easily distinguished, despite their geographical separation. Florentian infant microbiomes were distinct from rural microbiomes in Burkina Faso. The relationship between gut microbiomes and resource availability are likely complex. Settings considered low-resource, such as Burkina Faso, are also accompanied by
more traditional lifestyles and diets more like rural regions in resource-rich regions. Increased resources are often accompanied by urbanization, which has on its own been associated with distinct microbiomes; microbiomes of urban-dwellers in Nigeria and Burkina-Faso both have microbiomes resembling those living in industrialized urban centres, and distinct from rural counterparts from the same countries\textsuperscript{91,94}. In each of these instances, microbiomes associated with low-resource environments were associated with increased abundance of \textit{Prevotella}, while industrialized or urban ones, with \textit{Bacteroides}. However, we found that \textit{Bacteroides} was well-represented within the Ecuadorean infants, even in instances where \textit{Prevotella} abundance was high and even though all infants were born and lived in the same rural village. In contrast, while \textit{Bacteroides} dominated almost every Canadian subject, \textit{Prevotella} OTUs were rarely identified in this high-resource region cohort. This suggests that while the low-resource environments may promote \textit{Prevotella} dominance, \textit{Bacteroides} species are still able to colonize in those environments. In contrast, high-resource environments may be highly unfavorable to \textit{Prevotella} colonization of the gut?

We found a quadratic relationship between maternal age and microbial diversity in the Canadian and Ecuadorean cohorts. Lack of relationship between maternal age within the Belgian and South African cohorts may reflect the lower sample size in these cohorts, although it is possible that it reflects a region-specific effect of maternal age on child microbiomes. While the effect of maternal age on child microbiomes has not been previously evaluated, both extremes of maternal age have been associated with increased risk for adverse birth outcomes\textsuperscript{188}, stunting at two years of age, and altered glucose metabolism in adulthood\textsuperscript{189}. To what extent the microbiome is involved in such outcomes associated with maternal age warrants more attention.
While we did not identify microbial signatures unique to each cohort, we were able to accurately classify each cohort based on their immune phenotypes. For instance, although South African and Ecuadorean microbiomes could not be distinguished, their immune phenotypes were very distinct, with South Africans under-responding to almost all PRR agonists. Also, while the Belgian cohort included microbiomes similar to both low- and high-resource region microbiomes, infants of the Belgian cohort could be identified by their increased IL-10 responsiveness and decreased responsiveness to endosomal TLR stimulation independent of their microbiome. Thus, there are likely additional environmental or genetic determinants of systemic innate immune phenotypes that are not directly influenced by the taxonomic composition of gut microbiomes.

Using a multivariate method, we could identify immune-microbiome correlation networks shared among Belgian, Canadian, and Ecuadorean children alongside distinct networks within each of the four cohorts. Within the global network, cytokine responses to TLR2 were predominant—correlating positively with Prevotella OTUs and negatively with Bacteroides OTUs. Both genera have been associated with modulating mucosal immune responses through TLR2. Prevotella species induce a more robust pro-inflammatory cytokine response from human dendritic cells in a TLR2-dependent manner\textsuperscript{190, 191}. These effects on dendritic cells link Prevotella-rich gut dysbiosis in humans to rheumatoid arthritis\textsuperscript{192, 193} and periodontal disease\textsuperscript{194}, both of which are associated with increased frequencies of Th17 cells induced by the pro-inflammatory DC subsets. Conversely, Bacteroides fragilis sphingolipid Polysaccharide A (PSA) also signals through TLR2, instead stimulating dendritic cells to produce IL-10, thereby contributing to an anti-inflammatory environment, systemically\textsuperscript{195} and in the mucosa\textsuperscript{196, 197}. 
Given that we recruited only healthy infants, and that *Prevotella* colonization is very common in rural populations, where metabolic and inflammatory diseases are less common than in urban settings\(^{198}\), this opens the possibility that enhanced TLR2 responsiveness in these settings is advantageous, and the associations between *Prevotella* and inflammatory disease manifest when *Prevotella*-dominated individuals are living in an urban environment. It is also possible that there are important interactions between *Bacteroides* or *Prevotella* and other taxa that co-occur with them, particularly Firmicutes that were also selected in the correlation network. Additionally, both viruses and fungi could have important interactions, but they were not measured in this study.

The above TLR2 responses were associated with distinct OTUs in Belgian, Canadian, and Ecuadorean networks. The Ecuadorean network selected both *Bacteroides* and *Prevotella* OTUs, alongside Firmicutes mostly belonging to the *Lachnospiraceae* family, while Canadian and Belgian networks selected only Firmicutes OTUs which correlated negatively with cytokine responses in Ecuadorians and Belgians, but positively, in Canadians. The lack of substantial overlap between selected OTUs combined with overlapping selected responses suggests that the TLR2 response pathway is more closely associated with the host microbiome than other PRRs, and that microbes that modulate the TLR2 pathway, aside from *Bacteroides* and *Prevotella*, may be region-specific.

The absence of a global integrative network when South African children were included in the analysis was likely due to their extremely divergent immune profiles, which were not reflected in their microbiomes. While cytokine responses to TLR2 were selected in all integrations of the other three cohorts, combined and separately, these responses were not
selected in the South African cohort, suggesting that the hypo-responsiveness observed in these children is caused by unique factors yet to be identified.

The fecal transplants in germ-free mice suggest that South African microbiomes induced lower splenocyte cytokine responses to TLR4 and TLR7:8 compared to Canadian microbiomes. That Th1- and Th17- supporting cytokines were affected, while inflammatory cytokines were not, is inconsistent with our published human data\textsuperscript{167} which shows that both types of responses were reduced in South African children.

The mouse microbiomes clustered according to fecal donor but were distinct from those of the corresponding donors at the end of the experiment. Thus, the original fecal inoculum was not maintained after transplant, but it strongly influenced the trajectory of the developing microbiome. Based on these data it is impossible to conclude whether microbial components in the inoculum, metabolic activities of engrafted microbes, or both modulated the immune phenotype. However, we identified four OTUs that were present in all mice inoculated with South African feces, were absent in mice inoculated with Canadian feces, and were detectable in only South African donor feces. These OTUs belonged to the \textit{Alistipes}, \textit{Odoribacter}, and \textit{Prevotella} genera and one to the \textit{Rikenellaceae} family. \textit{Aistipes} and \textit{Rikenellaceae} have been associated with innate immune deficiency in mice. \textit{Rikenellaceae} are enriched in both MyD88 and NOD2 knockout mice\textsuperscript{199,200}. Both taxa were enriched in male versus female mice, where the male mice had decreased type 1 interferon responses\textsuperscript{201}. Importantly, these sex-specific immune responses are also present in germ-free mice. In humans, \textit{Alistipes} has been associated with healthy gut microbiomes, compared to those of subjects with inflammatory bowel diseases\textsuperscript{202} and allergic disease\textsuperscript{203}. \textit{Prevotella}, as discussed above, is associated with increased inflammation in mice and inflammatory disease in mice and humans. \textit{Odoribacter}, while a common gut
commensal, has not been prominently associated with human health outcomes. Taken together, it is possible that enrichment of *Alistipes* and *Rikenellaceae* are a result, rather than cause, of immune suppression in this mouse model, whereas neither *Prevotella* nor *Odoribacter* can be strongly implicated. It is possible that immune suppression was driven by a component of the fecal microbiome not detected using the 16S amplicon sequencing approach.

There are several limitations to this study. This study was designed as a pilot. Thus, the study was limited in sample number and also suffered unequal sample numbers among the four cohorts. With that, the integration results were more heavily influenced by the Ecuadorean and Canadian cohorts than the Belgian or South African cohorts. This limited the number of OTUs retained for integrated analysis to roughly 5% of those identified among the four cohorts. We also did not analyze the metagenome, which has been shown to associate with systemic immunity in previous work \(^{169}\). While the cohorts are referred to by their countries of recruitment, the enrolled study participants are not representative of the overall populations in those countries. This is reflected by the ethnic background of recruited study participants. Thus, these cohorts are geographically separated, but belong to sub-populations not necessarily represented by the resource availability of their home countries. The germ-free mouse experiments were not conducted in a germ-free facility, which may have contributed to the divergence of engrafted mouse microbiomes from the original inocula.

In conclusion, this was the first attempt to investigate the relationships among geographically distinct immune phenotypes and the gut microbiomes. We established a predominant association between systemic cytokine responses to TLR2 stimulation and colonization with *Bacteroides* and *Prevotella* populations. We also demonstrate that, even in absence of statistical correlations, it is possible to demonstrate causal relationships between the
host microbiome and systemic immunity, using germ-free mouse models of human fecal transplantation. Further investigation of the biogeography of microbiome-immune interactions will continue to expand our understanding of health and disease in the increasingly mobile global human population.
Chapter 3: Innate immune responses and gut microbiomes distinguish HIV-exposed from unexposed infants in a population-specific manner

3.1 Introduction

HIV transmission from mother to infant has been drastically reduced due to increased coverage of antiretroviral therapy during pregnancy. However, HIV-exposed uninfected (HEU) infants are not unaffected as they still suffer more severe infectious disease in early infancy than HIV-unexposed uninfected (HUU) controls. Altered immunity in HEU infants has been identified as a potential cause\textsuperscript{155, 156}. However, each of these studies involved only single country populations. It thus remains unclear if alterations of the immune system following HIV exposure in early life are similar around the globe. This is a crucial aspect to delineate, as immune alterations common to all HEU infant populations would suggest a common mechanism (such as HIV or ARV medication itself) and thus targeting this pathway may be of benefit to all (e.g. starting ARV treatment before conception\textsuperscript{158}). If the immune differences between HEU and HUU infants were not globally present but population specific, the likely culprits likely wouldn’t be prenatal HIV or ARV exposure (i.e. factors common to all sites) and avenues to improve outcomes for HEU infants would have to focus on population-specific factors.

Most immune studies to date contrasting HEU to controls have been done in African populations living either in Africa or Europe\textsuperscript{155, 157-159} and the one survey in Haiti\textsuperscript{165}. Yet, both microbiomes\textsuperscript{10, 91} and immune responses\textsuperscript{167, 204} have been shown to differ from one population to another among healthy infants. Thus, to put the effect of HIV exposure on immune responses and microbiomes, and importantly, the interactions between the two, in context, they must be
evaluated simultaneously but in diverse populations to determine to what extent differences between HEU and HUU control infants can apply from one study to the next.

To address this, we recruited three cohorts of HEU and HUU control infants from contrasting regions and measured both cytokine production in response to pattern recognition receptor (PRR) agonists and in parallel profiled their stool microbiomes with 16S amplicon sequencing. Using a stringently standardized immune assay designed to eliminate the effect of site, we previously identified site-specific immune signatures among HUU controls from these cohorts\textsuperscript{167, 204}. In the present study, we contrasted site-specific alterations in immune responses and stool microbiomes of HEU vs. HUU control infants. Furthermore, using multi-omic integration of immune and microbiome data, we assessed whether a relationship between these compartments that related to HIV exposure existed.

Given the powerful influence of the bacterial microbiome on systemic immunity, and known differences in the microbiome around the world, we hypothesized that the impact of HIV exposure on immunity could be driven by population-specific influence of the gut microbiome.

3.2 Methods

Study Participants

HEU and HUU control infants were recruited between January 2011 and March 2012 in all study sites.

**HIV unexposed uninfected (HUU) control participants.** The recruitment of the control subjects was previously published, and were the same children whose immune phenotypes and microbiomes were presented in Chapter 2\textsuperscript{167}. Infants of approximately two years of age were
recruited from Vancouver, Canada; Brussels, Belgium; and Cape Town, South Africa. Control subjects were only included in the study if the infant was considered healthy based on medical history. They were excluded if they met one or more of the following criteria: significant chronic medical condition, immune deficiency, immunosuppression by disease or medication, cancer, bone marrow or organ transplantation, receipt of blood products within 3 months, bleeding disorder, major congenital malformation, genetic disorder, or born to HIV-positive mothers.

**HIV-exposed uninfected participants.** HEU subjects were recruited at the same centers as HUU controls from each respective cohort, except for Canadian HEU infants. Canadian HEU infants were recruited in the Oak Tree Clinic at the British Columbia women’s hospital. This clinic specifically cares for HIV infected women and follows infants born to HIV positive mothers over the first two years of life. For all sites, the same exclusion criteria for used for HUU controls applied to HEU infants, aside from the criterion to be born to HIV-negative mothers. Inclusion criteria were perinatal HIV exposure in addition to criteria used for HUU control participants.

**Innate immune phenotyping**

**Blood draw and supernatant cytokine measurement.** Blood draw and supernatant cytokine analysis following PRR stimulation was done simultaneously with samples from HUU control subjects presented in Chapter 2. In brief, stimulation plates containing 10x the final desired concentration of PRR ligands were prepared from a single lot of reagents simultaneously at the Vancouver laboratory and shipped to Belgian and South African study sites. PRR agonists used
included PAM3CYSK4 (PAM; stimulates TLR2), polyinosinic-polycytidylic acid (Poly I:C, stimulates TLR3), Lipopolysaccharide (LPS, stimulates TLR4), Peptidoglycan (PGN, stimulates TLR2 and nucleotide-binding oligomerization domain-containing protein 1/2 [NOD1/2]), resiquimod (R848; stimulates TLR7:8) and RPMI 1640 medium alone. The following serum cytokine concentrations were measured at the Vancouver laboratory using the Luminex multiplex assay (Luminex, Upstate/Millipore “Flex Kit” system): IFN-α2, IFN-γ, CXCL10, IL-12p70, IL-12p40, IL-6, TNF-α, IL-1β, CXCL8, CCL3, CCL4, and IL-10. IL-23 was measured via human IL-23 ELISA.

**Preparation of PRR stimulation plates for flow cytometry.** Stimulation plates for flow cytometry were prepared using the same protocol as preparation of PRR stimulation plates for supernatant cytokine analysis in Chapter 2 with the following modification: All stimulation wells contained brefeldin A at 10 µg/mL (Sigma-Aldrich Cat. No. B6542). Pre-made stimulation plates for flow cytometry were shipped to each respective study site alongside stimulation plates for supernatant cytokine analysis on dry ice under temperature-monitored conditions by World Courier Inc.

**Whole blood PRR stimulation for flow cytometry.** PRR stimulation of whole blood for flow cytometry was done alongside the stimulation assay for cytokine measurement by Kinga Smolen at each site. 200 µL whole blood diluted 1:1 in RPMI 1640 medium was added to each well. Plates were incubated at 37°C at 5% CO₂ for 6 hours, then treated with 2 mM EDTA for 10 min at 37°C. Cells were then suspended in 1.4 mL BD FACS Lysing solution (BD Cat. No. 349202) and stored at -80°C in 2.0 mL screw-top tubes (VWR Cat. No. 16466-060). Samples from the
Belgian and South African sites were shipped to the Vancouver laboratory for further analysis on dry ice under temperature monitored conditions by World Courier Inc.

**Intracellular cytokine staining and flow cytometry.** At the Vancouver laboratory, flow cytometry samples were thawed at 37°C and immediately spun at 600 g for 5 min. Supernatant was discarded and pellets were resuspended in 200 µL 1x BD FACS permeabilizing solution 2 (BD Cat. No. 340973) and incubated at room temperature for 10 min. Cells were washed twice by spinning at 600 g for 5 min, discarding supernatant, and resuspending in 100 µL PBS containing 0.5% Bovine Serum Albumin (BSA) and 0.1% sodium azide (PBSAN). Cells were stained with antibodies against cell-surface markers CD11c (APC; clone S-HCL-3, CD Cat. No. 340544), CD123 (PE-Cy7; clone 6H6, eBioscience Cat. No. 25-1239), CD3 (PE-CF594; clone UCHT1; BD Cat. No. 562280), γδTCR (FITC; clone B1.1, eBioscience Cat. No. 11-9959), HLA-DR (eFlour605; clone LN3; eBioscience Cat. No. 93-9956), and CD14 (V500; clone M5E2, BD Cat. No. 561391) and intracellular cytokines TNF-α (Alexa 700; clone Mab11, BD Cat. No. 557996), IFN-α (PE; clone 7N4-1, BD Cat. No. 560097), IL-6 (Percp-eFlour710; clone MQZ13A5, eBioscience Cat. No. 46-7069), IFN-γ (BV711, clone 4S.B3, BD Cat. No. 564793), and IL-12p40 (eFlour450; clone LN3, eBioscience Cat No. 48-7129). Cells were stained in 100 µL PBSAN for 45 minutes at room temperature in the dark. Cells were then washed twice and resuspended in a final volume of 250 µL PBSAN for flow cytometric analysis on an LSR II Flow Cytometer (BD Biosciences). Flow cytometry samples were randomized for HEU exposure and site. Prior to each acquisition, tracking beads (BD Biosciences) were used to calibrate the machine and compensation beads (BD Biosciences Cat. No 552843) were used to standardize
voltage settings. Positive controls for each acquisition were used by staining an aliquot of pre-
frozen stock of one adult whole-blood sample stimulated with R848. A total of 500,000
uncompensated events were acquired per sample. Compensation was conducted in FlowJo
version 9 (TreeStar). Viability was assessed via forward and side-scatter appearance only. After
compensation, gates for each cell population of interest were set on the fluorescence-minus-one
principle. Proportions of the following cell types were identified: monocytes (HLA-DR+, 
CD14+), cDCs (HLA-DR+, CD14-, CD11c+, CD123-), pDCs (HLA-DR+, CD14-, CD11c-, 
CD123+), αβ T cells (CD3+, γδTCR-), γδ T cells (CD3+, γδTCR+), B cells (HLA-DR+, CD14-, 
CD11c-, CD123-), and granulocytes (HLA-DR-, CD14+). Intracellular production of cytokines
IL-6, IL-12p40, IFN-α, IFN-γ, and TNF-α were also measured via median fluorescence intensity
(MFI) within monocytes, pDCs, and cDCs.

**Infant fecal microbiome amplicon sequencing.** All samples for 16S amplicon sequencing were
done simultaneously with samples from HUU control samples presented in Chapter 2. In Brief,
total DNA was extracted from each stool specimen using the Qiagen QIAamp Stool DNA mini-
kit, and microbiome composition measured via amplicon sequencing targeting the V6 region of
the 16S rRNA gene.

**Statistical Analysis**

**Baseline Demographics.** Baseline demographics between HEU and HUU children were
compared in each cohort separately. For continuous variables, the data among HEU or HUU
cohorts displayed skewedness (outside the range of -1 to 1) or kurtosis (outside the range of -3 to
3) for some cohorts, but not others. The student’s t-test was used to compare normal continuous
variables and data presented as mean with standard deviation. The non-parametric Wilcoxon rank-sum test for all non-normal continuous variables and the data presented as medians and interquartile ranges. For categorical variables, the Fisher’s exact test was applied where there were fewer than 10 individuals in each sub-grouping, and the Chi-Squared test for instances were each group contained sample sizes of 10 or more. Unadjusted p-values under 0.05 were considered significant.

**Principal components analysis of cytokine data.** Prior to PCA analysis, cytokine values were log10-transformed and scaled. The PCA was constructed with each unique data point constituting one individual for one stimulus. Variance explained by site and HIV exposure were determined in cytokine data subset by stimulus, and variance explained by HIV exposure within each site was determined using the data subset by site within each stimulus. The variance explained by stimulus was determined using the Adonis test on Euclidean distance in the R package vegan\textsuperscript{180}.

**Univariate analysis of cytokine response data.** The effect of HIV exposure on cytokine responses to PRR ligands was tested within each regional cohort separately. Cytokine data were log-10 transformed prior to univariate analysis. For cytokine production in response to ligand, we used multiple linear regression and adjusted for levels produced in the unstimulated sample. For unstimulated cytokine production, we selected only cytokines where at least 70% of subjects had unstimulated values above the detection limit, with a median fluorescence intensity (MFI) greater than 10. Both baseline and stimulation response analysis were adjusted for recruitment date for the Canadian cohort only, where HEU children were recruited at a significantly different
point in time compared to HUU children. Recruitment day was calculated using the number of days from the recruitment of the first to the last participant in the study. P-values were adjusted for false discovery across all ligand-cytokine combinations within each cohort using the Benjamini-Hochberg method. For all significant findings, multiple linear regression was used and included factors that were not balanced within that cohort: child sex and delivery mode in the Belgian cohort, and WLZ scores, HAZ scores, and maternal age for the Canadian cohort. Due to sample size, each demographic model was tested in a separate model. Within the Belgian cohort only, multiple linear regression was used to test whether any significant cytokines differed between HEU children whose mothers initiated ARV therapy preconception or during pregnancy.

**Analysis of flow cytometry data.** Proportions of cells types in HEU and HUU children, within each regional cohort, were compared using the Wilcoxon test, and p-values adjusted within each cohort using the Benjamini-Hochberg method. Differences in the percent of cDCs, pDCs, and monocytes producing IL-6, IL-12p40, TNF-α, and IFN-α, and IFN-γ were tested using linear regression, adjusted for percent producers in the unstimulated controls. Both baseline and stimulation response analyses were adjusted for recruitment date for the Canadian cohort only, as described above for cytokine response analysis. P-values were adjusted for all ligand-cytokine combinations for each cell type separately using the Benjamini-Hochberg method.

**Microbiome Analysis**

**Alpha diversity.** We estimated microbial diversity with abundance-dependent (Shannon) metrics, after subsampling the OTU table to account for unequal sampling depth. To test for relationships between host factors and Shannon diversity, we used the Wilcoxon test.
**Beta diversity.** We used the *phyloseq* R package\(^{179}\) to compute beta diversity using the Bray-Curtis index. Sample clustering was visualized using non-metric-multidimensional scaling (NMDS). To test whether HIV exposure explained community composition among all infants or among those within each cohort, we used the Adonis test from R package *vegan*\(^{180}\).

**Differential abundance.** Following analysis of alpha and beta diversity, we stratified all further analyses by site. Only OTUs present in at least 6% of subjects were retained within each site. To identify differentially abundant OTUs between HEU and control infants, we used the Wald test implemented in the R package *DESeq2*\(^{183}\). OTUs were considered significant if with a Benjamini-Hochberg adjusted p-value under 0.05.

**Integrated analysis of cytokine and microbiome data**

**sPLS-DA analysis of microbiome and Luminex data.** For both microbiome and Luminex datasets, we tested the ability of significantly different features to classify HEU from control infants within each cohort using sparse Partial Least Squares Discriminant Analysis\(^{184}\), implemented in the R *mixOmics* package. Cytokine data were log10-transformed. OTU data were normalized using the variance stabilizing transformation implemented in the *DESeq2* R package. Models were first tuned to select the number of features required to achieve the minimum classification error rate, using Mfold validation with at least 10 subjects per fold and 100 iterations. The final models were run on two components, using the numbers of features selected by the tuning process. Classification error was assessed using the maximum error rate.
**Cytokine and microbiome data integration.** To test whether OTU-cytokine correlations distinguished HEU from control subjects within each cohort, we used sparse Generalized Canonical Correlation Discriminant Analysis, or mixDiablo\textsuperscript{205} implemented in the R *mixOmics* package. Cytokine data were log10-transformed. OTU count data were first transformed into relative abundance, then normalized using the CLR transformation with an offset of 0.1 times the minimum value for each individual (the default offset in the *mixOmics* CLR function). All data were scaled. We used the DIABLO full model, where selected features across data types are constrained to the discriminant factor (HIV exposure) and to the features in the other data block – i.e. to maximize correlation across OTU and microbiome data. We first tuned the model to determine the maximum number of OTUs and cytokines to achieve the lowest classification error rate across two components, with a minimum of 5 features per dataset and a maximum equal to the number of samples. The final models used the selected number of features, and error assessed using the maximum error rate.

### 3.3 Results

#### 3.3.1 Cohort characteristics

HEU and HUU control infants differed demographically in a cohort-specific manner. Belgian HUU children differed from HEU infants in sex and delivery mode where, 95% of controls versus 52% of HEUs were male, and 90% of controls versus 52% of HEUs were delivered vaginally (Table 3.1.1). Canadian HUU differed from HEU children in age, WLZ, LAZ, and maternal age, where HEU children were younger, had higher WLZ, but lower LAZ scores than controls. Canadian HEU children were also born to younger mothers than controls (Table 3.1.2). South African HEU and HUU controls differed for WAZ and HAZ scores (Table 3.1.3). For all
three cohorts, ethnic background differed between HEUs versus controls (Table 3.2). Belgian HEU children were predominantly African, while HUU controls were either mixed race (mostly Arab and African) or Arab. Canadian HUU controls were mostly Caucasian while HEUs were represented by diverse ethnic groups and were predominantly mixed-race. South African HUU controls were mostly coloured (African-Caucasian) while the HEU children were mostly Africans from the Xhosa tribe. Maternal ARV initiation differed among the three sites (Table 3.3). The highest proportion of Belgian mothers started ARV prior to pregnancy. Most Canadian mothers began ARV in the second trimester, while most South African mothers began ARV in the third trimester. While HEU and HUU controls were recruited at different time periods between sites, recruitment time within site differed between HEU and HUU control infants in Canada only (Table 3.4, Figure 3.1).
Table 3.1 Belgian cohort characteristics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>21</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex: n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1 (4.8)</td>
<td>10 (47.6)</td>
<td>0.004</td>
<td>FE</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>20 (95.2)</td>
<td>11 (52.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode: n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caesarean</td>
<td>2 (9.5)</td>
<td>10 (47.6)</td>
<td>0.015</td>
<td>FE</td>
<td></td>
</tr>
<tr>
<td>Vaginal</td>
<td>19 (90.5)</td>
<td>11 (52.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in months: mean (sd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23.12 (6.54)</td>
<td>23.67 (4.11)</td>
<td>0.747</td>
<td>TT</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational age: median [IQR]</td>
<td>39.00 [38.00, 40.00]</td>
<td>38.00 [37.00, 39.00]</td>
<td>0.099</td>
<td>WT</td>
<td></td>
</tr>
<tr>
<td>Birthweight (g): mean (sd)</td>
<td>3119.17 (724.59)</td>
<td>2908.10 (637.66)</td>
<td>0.339</td>
<td>TT</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAZ: mean (sd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.60 (1.21)</td>
<td>0.74 (1.01)</td>
<td>0.68</td>
<td>TT</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLZ: mean (sd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.06 (1.33)</td>
<td>0.03 (1.07)</td>
<td>0.799</td>
<td>TT</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAZ: mean (sd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.36 [0.59, 1.85]</td>
<td>1.88 [0.68, 2.49]</td>
<td>0.382</td>
<td>WT</td>
<td></td>
</tr>
<tr>
<td>Maternal age: mean (sd)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.67 (6.48)</td>
<td>29.90 (5.01)</td>
<td>0.33</td>
<td>TT</td>
<td></td>
</tr>
</tbody>
</table>

WAZ = weight for age Z-score, WLZ = weight for length Z-score, HAZ = height for age Z-score
FE = Fisher’s exact test, TT = two-tailed student’s t-test, WT = Wilcoxon rank-sum test
Table 3.2 Canadian cohort characteristics

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex: n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16 (48.5)</td>
<td>7 (43.8)</td>
<td>0.995</td>
<td>CS</td>
</tr>
<tr>
<td>Male</td>
<td>17 (51.5)</td>
<td>9 (56.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delivery mode: n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caesarean</td>
<td>16 (48.5)</td>
<td>4 (25.0)</td>
<td>0.208</td>
<td>CS</td>
</tr>
<tr>
<td>Vaginal</td>
<td>17 (51.5)</td>
<td>12 (75.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in months: median [IQR]</td>
<td>18.87 [18.38, 20.74]</td>
<td>18.29 [18.25, 19.34]</td>
<td>0.041</td>
<td>WT</td>
</tr>
<tr>
<td>Gestational age: median [IQR]</td>
<td>39.00 [38.00, 40.00]</td>
<td>38.50 [37.83, 39.35]</td>
<td>0.11</td>
<td>WT</td>
</tr>
<tr>
<td>Birthweight (g): mean (sd)</td>
<td>3291.15 (415.96)</td>
<td>3016.38 (595.01)</td>
<td>0.067</td>
<td>TT</td>
</tr>
<tr>
<td>WAZ: mean (sd)</td>
<td>0.36 (1.01)</td>
<td>0.73 (1.16)</td>
<td>0.271</td>
<td>TT</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WLZ: median [IQR]</td>
<td>0.33 [-0.33, 0.80]</td>
<td>1.27 [0.75, 1.56]</td>
<td>0.004</td>
<td>WT</td>
</tr>
<tr>
<td>HAZ: mean (sd)</td>
<td>0.30 (0.99)</td>
<td>-0.58 (0.87)</td>
<td>0.005</td>
<td>TT</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal age: mean (sd)</td>
<td>34.97 (4.24)</td>
<td>29.00 (4.98)</td>
<td>&lt;0.001</td>
<td>TT</td>
</tr>
</tbody>
</table>

WAZ = weight for age Z-score, WLZ = weight for length Z-score, HAZ = height for age Z-score
FE = Fisher’s exact test, TT = two-tailed student’s t-test, WT = Wilcoxon rank-sum test
Table 3.3 South African cohort characteristics

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
<th>test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex: n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>10 (50.0)</td>
<td>11 (68.8)</td>
<td>0.32</td>
<td>FE</td>
</tr>
<tr>
<td>Male</td>
<td>10 (50.0)</td>
<td>5 (31.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Delivery mode: n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caesarean</td>
<td>0 (0.0)</td>
<td>1 (6.2)</td>
<td></td>
<td>0.444</td>
<td>FE</td>
</tr>
<tr>
<td>Vaginal</td>
<td>20 (100.0)</td>
<td>15 (93.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age in months: mean (sd)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>24.29 (0.62)</td>
<td>24.37 (0.63)</td>
<td>0.718</td>
<td>TT</td>
</tr>
<tr>
<td><strong>Gestational age: median [IQR]</strong></td>
<td>Missing</td>
<td>38.00 [37.00, 40.00]</td>
<td>39.00 [38.00, 40.00]</td>
<td>0.467</td>
<td>WT</td>
</tr>
<tr>
<td><strong>Birthweight (g): mean (sd)</strong></td>
<td></td>
<td>3005.85 (371.02)</td>
<td>2939.38 (455.51)</td>
<td>0.632</td>
<td>TT</td>
</tr>
<tr>
<td><strong>WAZ: mean (sd)</strong></td>
<td></td>
<td>-0.53 (1.08)</td>
<td>0.32 (1.24)</td>
<td>0.035</td>
<td>TT</td>
</tr>
<tr>
<td><strong>WLZ: median [IQR]</strong></td>
<td></td>
<td>0.00 (0.95)</td>
<td>0.42 (1.39)</td>
<td>0.288</td>
<td>TT</td>
</tr>
<tr>
<td><strong>HAZ: mean (sd)</strong></td>
<td></td>
<td>-1.02 (1.28)</td>
<td>-0.06 (1.22)</td>
<td>0.028</td>
<td>TT</td>
</tr>
<tr>
<td><strong>Maternal age: mean (sd)</strong></td>
<td></td>
<td>25.73 (5.76)</td>
<td>27.42 (5.05)</td>
<td>0.434</td>
<td>TT</td>
</tr>
</tbody>
</table>

WAZ = weight for age Z-score, WLZ = weight for length Z-score, HAZ = height for age Z-score
FE = Fisher’s exact test, TT = two-tailed student’s t-test, WT = Wilcoxon rank-sum test
### Table 3.4 Ethnic background of Belgian HEU and HUU control children

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td>2 (10.5)</td>
<td>17 (81.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>African/Arab</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Arab</td>
<td>7 (36.8)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Belgian/Guinean</td>
<td>0 (0.0 )</td>
<td>1 (4.8)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>3 (15.8)</td>
<td>1 (4.8)</td>
<td></td>
</tr>
<tr>
<td>Black/Arab</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Caucasian/Arab</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Metis/Caucasian/Arab</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>North African</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>1 (5.3 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0.0 )</td>
<td>1 (4.8)</td>
<td></td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>1 (5.3 )</td>
<td>1 (4.8)</td>
<td></td>
</tr>
</tbody>
</table>

Statistical analysis done using the Chi-squared test.

### Table 3.5 Ethnic background of Canadian HEU and HUU control children

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aboriginal</td>
<td>0 (0.0 )</td>
<td>1 (6.2)</td>
<td>0.001</td>
</tr>
<tr>
<td>Asian</td>
<td>0 (0.0 )</td>
<td>1 (6.2)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0 (0.0 )</td>
<td>2 (12.5)</td>
<td></td>
</tr>
<tr>
<td>Caucasian-mixed</td>
<td>5 (15.2)</td>
<td>9 (56.2)</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>3 (9.1 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Filipino</td>
<td>1 (3.0 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Latin American</td>
<td>3 (9.1 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>South Asian</td>
<td>0 (0.0 )</td>
<td>1 (6.2)</td>
<td></td>
</tr>
<tr>
<td>White/Caucasian</td>
<td>21 (63.6)</td>
<td>2 (12.5)</td>
<td></td>
</tr>
</tbody>
</table>

Statistical analysis done using the Chi-squared test.

### Table 3.6 Ethnic background of South African HEU and HUU control children

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>HEU</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>African (Xhosa)</td>
<td>4 (20.0)</td>
<td>11 (68.8)</td>
<td>0.004</td>
</tr>
<tr>
<td>African (Zimbabwean)</td>
<td>0 (0.0 )</td>
<td>1 (6.2)</td>
<td></td>
</tr>
<tr>
<td>Mixed-race</td>
<td>15 (75.0)</td>
<td>4 (25.0)</td>
<td></td>
</tr>
<tr>
<td>White-Caucasian</td>
<td>1 (5.0 )</td>
<td>0 (0.0)</td>
<td></td>
</tr>
</tbody>
</table>

Statistical analysis done using the Chi-squared test.
Table 3.7 Initiation of maternal antiretroviral prophylaxis/therapy for HIV-positive mothers

<table>
<thead>
<tr>
<th></th>
<th>Belgium</th>
<th>Canada</th>
<th>South Africa</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No PMTCT</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>3 (21.4)</td>
<td>0.001</td>
</tr>
<tr>
<td>Preconception</td>
<td>7 (38.9)</td>
<td>3 (25.0)</td>
<td>1 (7.1)</td>
<td></td>
</tr>
<tr>
<td>First Trimester*</td>
<td>3 (16.7)</td>
<td>2 (16.7)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Second Trimester*</td>
<td>5 (27.8)</td>
<td>7 (58.3)</td>
<td>2 (14.3)</td>
<td></td>
</tr>
<tr>
<td>Third Trimester*</td>
<td>3 (16.7)</td>
<td>0 (0.0)</td>
<td>8 (57.1)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

*First trimester = 0 to under 13 weeks, second trimester = 13 to under 26 weeks, third trimester = 26 weeks or more. Statistical analysis done using the Chi-squared test.

Table 3.8 Recruitment periods for HEU and HUU control children

<table>
<thead>
<tr>
<th></th>
<th>Recruitment Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>07-Sep-2011 to 24-Oct-2011</td>
</tr>
<tr>
<td>HEU</td>
<td>07-Sep-2011 to 19-Oct-2011</td>
</tr>
<tr>
<td>Canada</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>01-Nov-2011 to 31-Jan-2012</td>
</tr>
<tr>
<td>HEU</td>
<td>27-Jan-2011 to 05-Mar-2012</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>04-May-2011 to 31-May-2011</td>
</tr>
<tr>
<td>HEU</td>
<td>04-May-2011 to 30-May-2011</td>
</tr>
</tbody>
</table>
3.3.2 Innate immune differences between HEU and HUU control children

3.3.2.1 Overall cytokine responses to PRR agonists

Principal component analysis of cytokine responses to PRR ligands revealed that stimulus had the dominant effect on sample clustering, as previously observed\(^{167}\) (Figure 3.2; PERMANOVA \(R^2 = 0.49, p = 0.001\)). For each stimulus, country of origin contributed significantly to variance (Figure 3.2 for LPS, Figure 3.3 for other PRR ligands). This masked the effect of HIV exposure within each site. When determining the effect of site HIV exposure within each stimulus, cohort of origin contributed significantly to clustering for every ligand, while the effect of HIV exposure contributed minimal variance in comparison (Figure 3.2C). Within each site, the effect of HIV exposure was significant for all ligands among Canadians, for select ligands among Belgians, and not at all among South Africans (Figure 3.2D; Figure 3.4E-F for site-specific PCA).
Figure 3.2  Principal components analysis discriminates cytokine responses to PRR stimulation between HEU and control infants

Each PCA data point represents one unique individual for one stimulus for A and B, and one individual for E-G. A. Type of stimulus has a significant effect on cytokine responses (PERMANOVA R² = 0.49; p = 0.001). B. PCA highlighting LPS showing within-stimulus cohort of origin dominates sample clustering. C. For each stimulus, site significantly explains variability of responses to every stimulus, while HIV exposure minimally explains variability, significantly for only unstim, PGN, and p: t:C. D. For each site, HIV exposure significantly explains variability of responses to every stimulus among Canadian infants, responses to only PGN and LPS among Belgian infants, and to no stimuli among South African infants. E-G. PCA analysis showing HIV exposure affects cytokine responses to LPS in Belgian and Canadian but not South African infants.
Figure 3.3 Principal component analysis highlighting cytokine data clustering for PGN (A), PAM (B), plC (C), and R848 (D) stimulated samples.
Figure 3.4 Principal component analysis shows sample clustering by HIV exposure for each ligand individually for Belgium (A), Canada (B), and South Africa (C). R2 and p-values are plotted in Figure 3.1D.

3.3.2.2 Differences in cytokine responses to PRR agonists

We tested the effect of HIV exposure on cytokine responses for each stimulus-cytokine combination separately for each regional cohort using simple linear regression. Consistent with PCA analysis, HIV exposure was associated with responses to LPS and PGN in the Belgian cohort, with greater IL-6 and IL-8 production by HEU children (Figure 3.5A). These findings remained significant after adjusting for sex. Because ARV initiation during pregnancy has been reported to account for differences between HEU and HUU control infants, we tested whether
timing of ARV initiation affected IL-6 and IL-8 responses in the Belgian cohort. No effect was found.

Within the Canadian cohort, HIV exposure was associated with higher levels of IFN-α2 and MDC and IL-23 both at baseline (unstimulated) and in response to pI:C (Figure 3.5B). Canadian HUU had higher levels than HEU children of IL-10 in response to LPS, PGN, and R848, IL-12p40 in response to LPS and PGN, IL-1β in response to LPS only, and IL-12p70 in response to R848 only (Figure 3.5C). These findings remained significant when adjusting for unbalanced demographic factors. Within the South African cohort, HIV exposure was not associated with responses to any ligands.

![Figure 3.5](image)

**Figure 3.5** Linear regression identified cytokine responses significantly affected by HIV exposure in a site-dependent manner
A. Among Belgians, significantly different cytokines include only higher responses by HEU compared to HUU control children. B-C. Among Canadians, significantly affected cytokines include higher responses to pI:C (B) and lower responses to LPS, PGN, R848, and baseline (C) by HEU compared to HUU control children.
At all three sites, HEU and HUU control children had similar proportions of monocytes, granulocytes, dendritic cells, and T cells. For Canadians, there were higher proportions of γδ T cells and lower proportions of B cells in HEU versus HUU control children (Figure 3.6).

![Whole blood cellular proportions of immune cells in HEU and HUU control children](image)

**Figure 3.6 Whole blood cellular proportions of immune cells in HEU and HUU control children**

Comparisons were made for HEU vs HUU control children in each site separately for cellular proportions of A. gamma-delta T cells, B. T cells, C. B cells, D. conventional dendritic cells, E. plasmacytoid dendritic cells, F. Granulocytes and G. monocytes

* p < 0.05, ** p < 0.01, Wilcoxon test, adjusted using Benjamini-Hochberg method

Canadian intracellular cytokine responses supported the trends seen for IL-12p40, as HEU children had fewer cDCs and producing IL-12p40 in response to PAM, LPS and R848 and fewer monocytes producing IL-12p40 in response to LPS and R848 (Figure 3.7). Few other differences in intracellular cytokine responses were found in Canadians; HEU child cDCs produced less IFN-γ in response to PGN (q = 0.048), and fewer monocytes produced IL-6 in response to LPS (q = 0.008). No other differences between HEU and HUU control children in any other site were found.
Figure 3.7 Canadian child intracellular production of IL-12p40
A. cDCs from Canadian HUU control children produced more IL-12p40 in response to PAM, LPS, and R848 compared to HEU children. B. Monocytes from Canadian HUU control children produced more IL-12p40 in response to LPS and R848.
* p < 0.05, ** p <= 0.01, *** p <= 0.001 linear regression, adjusted for unstimulated values

3.3.3 Microbiome differences between HEU and HUU control children

3.3.3.1 Alpha and beta diversity

Country of origin had the greatest effect on beta diversity (Figure 3.8A) and there were no differences in alpha diversity using the Shannon index between HEU and HUU control children at any site (Figure 3.8B). HIV exposure had minimal effects (Figure 3.9).

Figure 3.8 Microbiome alpha and beta diversity in HEU and HUU control children
A. Microbiome beta diversity is mainly driven by site in NMDS ordination of all samples. B. Alpha diversity computed with the Shannon index was not affected by HIV exposure for any site.
Figure 3.9 Beta diversity of HEU and HUU control children from each site separately
Belgian (A), Canadian (B), and South African (C) HEU and control children with inset
PERMANOVA R2 and p values for effect of HIV exposure on community composition illustrate
minimal effects of HIV exposure within each site.

3.3.3.2 OTU-level differences between HEU and HUU control children

We identified differentially abundant OTUs between HEU and HUU control children within
each site (Figure 3.10A-C), with 17, 63, and 3 OTUs discriminating between HEU and HUU
control children from Belgium, Canada, and South Africa, respectively. Belgian HEU children
were characterized by increased abundances 12 OTUs, including of several Firmicutes
(Eubacterium, Ruminococcus, and Blautia) and Bacteroidetes Rikenellaceae and Odoribacter.
HUU control children had a higher abundance of only 5 OTUs, representing the Proteobacteria
and Firmicutes phyla. OTUs belonging to the Bacteroides genus had both greater and lower
abundances in HEU children (Figure 3.10A).

Canadian HEU children had increased abundances of 22 OTUs. OTUs classified at the
highest level as Firmicutes (phylum), Clostridiales (order), and Bacteroides (genus) had both
greater and lower abundances in HEU children. OTUs always more abundant in HEU children
represented 10 classifications, including Bacteroidetes Parabacteroides (genus) and
Bacteroidales (order), and Firmicutes Blautia, Dorea, Eubacterium, Oscillospira, and
Streptococcus (genera). Multiple Akkermansia OTUs were more abundant in HEU children. Conversely, HUU control children where characterised by 12 classifications always less abundant in HEU children. These included mostly Firmicutes Veillonella, Ruminococcus, Lachnospira, Roseburia, Faecalibacterium, and Clostridium – all known short chain fatty acid producers. Others included Gammaproteobacteria, Wolbachia, Bifidobacterium, Prevotella, and Rikenellaceae belonging to the Proteobacteria, Bacteroides, and Actinobacterium phyla (Figure 3.10B).

Among the South African HEU and HUU control children, only 3 OTUs were differentially abundant. One Rikenellaceae OTU was more abundant in HEUs and one Succinivibrio OTU was more abundant in control children (Figure 3.10C). Differentially abundant OTUs were ranked for their discriminatory capacity using sPLS-DA analysis, and the top 4 OTUs discriminating between HEU and HUU control children are shown in Figure 3.10 D-G.
Figure 3.10 OTU-level differences between HEU and HUU control children
A-C. Differentially abundant OTUs in HEU vs control infants in Belgians (A), Canadians (B), and South Africans (C). OTUs with fold-changes above zero are enriched in HEUs. D-F. Differentially-abundant OTUs with highest discriminatory classes for control and HEU children classified above the Phylum level.
3.3.4 Cytokine-microbiome relationships in HEU and HUU control children

To test whether any differences in the microbiomes of HEU children correlated with their aberrant cytokine responses to PRR stimulation, we employed DIABLO; a sparse Generalized Canonical Correlation Discriminant Analysis designed to identify features that both co-vary across OMIC datatypes and discriminate between classes of interest\(^\text{205}\) – here HEU and HUU control children. Because achieving a minimal classification error rate is a primary goal for this analysis, we compared classification error rates using each dataset alone using sPLS-DA and DIABLO models, both tuned to select the number of features resulting in the lowest error rate for Luminex cytokine and Microbiome data. Cytokine data alone distinguished between Canadian HEU and HUU control children with an overall error rate of 11%, compared to 30% and 55% for Belgians and South Africans, respectively. Microbiome data alone distinguished HEU from HUU control children with 14%, 20%, and 11% error for Belgian, Canadian, and South African children (Figure 3.11; sPLS-DA models).

Except for cytokine data for the Canadian cohort, DIABLO selected features with inferior classification accuracy to sPLS-DA models. With only 5 OTUs (the minimum allowed) selected for each DIABLO model, they discriminated HEU from HUU control children with error rates of 50%, 58%, and 69% for Belgian, Canadian, and South African cohorts respectively. Thus, no OTUs were identified that both correlated and discriminated between HEU and HUU control children for any site.
Figure 3.11 Maximum classification error rate for discriminating HEU from HUU control children using single (sPLS-DA) or integrated (DIABLO) datasets
For all three sites (Belgium, A; Canada, B; South Africa, C) sPLS-DA analysis using each dataset separately distinguished HEU from HUU control children with error rates either equal, or superior to integrated DIABLO analysis, except for cytokine data for the Canadian cohort.
3.4 Discussion

The goal of this study was to evaluate, for the first time, immune as well as microbiome differences between HIV-exposed and unexposed infants across geographically diverse settings. Our findings suggest that site of origin is the primary driver of innate immune phenotype. However, an effect of HIV exposure on immune response was still evident within two of three sites. Notably, the immune differences between HEU and HUU control children were also cohort-specific; immune differences in one population and could not simply be extrapolated to another.

Belgian HEU children were distinguished from HUU control infants by higher IL-6 and IL-8 production in response to LPS and PGN, and higher IL-8 production in response to PAM. We determined that child sex did not drive these associations but could not determine to what extent they were driven by ethnicity, as the HEU children in this cohort were almost exclusively of African descent, while the control group was predominantly either Arab or mixed race. Increased inflammatory responses in HEU infants were described previously\textsuperscript{157}, including HEU children recruited at the exact same centre used in the present study\textsuperscript{158}. However, this is the first study to report differences at two years of life; previous studies found the more differential phenotypes at birth that sometimes persisted up to 1 year of life\textsuperscript{157, 158}.

Immune differences in HEU children have not been previously measured in any North American populations; this study reports that Canadian HEU children in this cohort differed from HUU control children by their lower production of IL-10 and IL-12 in response to LPS, PGN, and R848. Thus, while responses to similar ligands were affected by HIV exposure in both the Belgian and Canadian cohorts, the specific cytokines affected were completely distinct. Canadian HEU and HUU control children differed from one another demographically in several
aspects – child age, maternal age, WLZ, and HAZ scores, and ethnic background. Thus, several host factors independent of HIV infection may have contributed to the distinct immune responses. After adjusting for maternal age, WLZ scores, or HAZ scores, these differences remained significant, and thus were unlikely to contribute to observed immune phenotypes.

The lack of immune differences between South African HEU and HUU control children is in keeping with previous findings from this cohort, where different bulk and cell-specific cytokine responses to LPS and PAM were evident up to 6 months of life, but not by 1 year\textsuperscript{157}. Notably, the South African children overall were most distinct from the other cohorts due to lower responses to almost all PRR agonists\textsuperscript{167}. Thus, it is possible that environment influences immune responses to a degree that any effects of HIV exposure are no longer evident, or we did not have the power to detect differences with a smaller effect size.

To our knowledge, this is the first study to measure stool microbiome composition of HEU children at two years of age. Only one other study has reported stool microbiomes of HEU infants, but at 3 months\textsuperscript{165}. Gut microbiomes in early infancy, prior to addition of solid foods, are known to be highly distinct from those in later infancy or toddlerhood when solid foods are introduced and after weaning\textsuperscript{10,64}. Thus, it is not surprising that taxonomic differences in the 3-month-old infant study did not at all apply to our findings in 2-year-old children. Breast versus formula feeding may have contributed to microbiome differences between HEU and HUU control children in the present study. All HIV-positive mothers in the present study were advised to exclusively formula-feed their infants from birth to prevent HIV transmission. All the HUU control children in the present study were breastfed. Duration of breastfeeding was not recorded. Thus, exclusive formula feeding may have contributed to distinct microbial composition in HEU children.
Abundances of short-chain fatty acid (SCFA)-producing bacteria were reduced in Canadian HEU children. We found that gut microbiomes of Canadian HEU children included lower relative abundances of SCFA-producing taxa that have been isolated from breast milk and are associated with breastfeeding – *Bifidobacterium* and *Veillonella*\(^{128,206}\). As well, other prominent SCFA producers normally present in adult-like microbiomes, *Ruminococcus*, *Lachnospira*, *Clostridia*, *Faecalibacterium*, and *Roseburia*\(^{207}\), were also less abundant in Canadian HEU children. Diets rich in non-digestible fibers heavily increase the relative abundance of SCFA-producing bacteria\(^{208}\). However, as dietary composition was not analyzed in this study, we cannot determine whether this underlies the observed difference. Given the known roles SCFAs have in maintaining intestinal and immunological health, possible effects of reduced SCFA-producing bacteria in Canadian HEU children deserve further attention. However, such differences were not apparent in any of our other cohorts. In Belgian HEU infants, OTUs representing SCFA-producing *Ruminococcus* and *Eubacterium* were increased compared to HUU control children. Several classes in the Belgian cohort contained OTUs that either decreased or increased in Belgian HEU children, making the functional impact of the microbial differences seen in this cohort difficult to infer. Only 3 OTUs were differentially abundant between South African HEU and HUU control children, suggesting that this compartment was not affected by HIV exposure in this cohort.

*sPLS-DA* and block *sPLS-DA* analyses both demonstrated that while both cytokine and OTU datasets had discriminatory power for Belgian and Canadian cohorts, no discriminatory multivariate relationships between the two datasets (immune and microbiome) could be identified using this powerful data integration approach. This suggests that both compartments were affected, but not by the same factors associated with HIV exposure. For instance, while
formula feeding may have contributed to microbial differences seen between Canadian HEU and HUU control infants, another aspect of HIV exposure, partly accounted for by differences in maternal age, were behind the observed immune differences.

There are several limitations to this study. Our study only evaluated immune and microbial differences in a small number of HEU and HUU control children. Thus, we did not capture enough severe infectious events to assess the relationships between biological findings and clinical outcome. HIV and ARV exposure are both associated with adverse birth outcomes including low birth weight and pre-term birth\(^{209, 210}\), which are both in turn associated with increased risk for infection in early life. Our HEU child cohorts were almost all term and normal birth weight, and thus a low-risk population where increased severity of infection compared to HUU controls may be less apparent. Given the diverse make-up of both HEU and HUU control children in the three populations, we were unable to test whether additional host factors, such as ARV regimen or ethnic background, contributed to any differences seen between HEU and controls. The use of the word ‘control’ for HUU children is somewhat of a misnomer, in that there were many demographic differences between them and HEU children from each site, and were only matched for age and broad geographic region. To appropriately assess both the effect of HIV exposure and socioeconomic factors associated with HIV infection, either appropriately matched controls or adequate and diverse sample sizes are needed to adjust for possible confounders. Our small sample sizes made it impossible to test whether any immune or microbiome differences contributed to clinical phenotypes in these infants, as increased severity of disease has been noted earlier in infancy, and less-so at the two year time point. Our approach to measure the immune phenotype relies on whole blood – a physiological compartment distinct
from the mucosal immune system, where host-microbe interactions (both commensal and pathogenic) take place.

In conclusion, while differences between HEU and HUU control children in immune status as well as microbiome composition were readily detectable at 2 years of age, these differences were not present across all the populations included in our study. This indicates that there was no universal signature of HIV exposure impacting the immune or microbiome compartment. Furthermore, the microbiome differences between HEU and control children within a given population did not correlate with their respective differences in immune status; thus, our study did not find any evidence indicating that the microbiome was a mechanistic driver for any immune difference even within a site. Together, this strongly suggests that factors common to all sites in our study, specifically biological exposure to HIV and ARV were unlikely the sole culprit for immune or microbial differences at two years of life.
Chapter 4: Dramatic changes in bacterial and fungal community dynamics over the first five years of life in rural Ghana

4.1 Introduction

Bacterial microbiomes are increasingly recognized as important in health and disease starting at birth\textsuperscript{10, 61-65}. Fungi also colonize our intestines early in life and emerging evidence demonstrates that they also may influence health and disease. Bacterial and fungal communities are known to interact intensely, and in this interaction, along with the host selection, determine the composition of the combined microbiome.

Despite the importance of both the bacterial and fungal microbiome in influencing the developmental trajectory towards health or disease, to our knowledge these two have never been analyzed in the same samples and over the likely most important period of development, namely in early life. Equally surprising, despite the well-established fact that the maternal stool and breast milk microbiome shapes the newborn microbiome, no study so far has captured this interaction between mother and infant across the bacterial and fungal kingdoms simultaneously. We thus conducted a cross-sectional bacterial and fungal microbiome survey encompassing the first 5 years of life, alongside stool and breast milk microbiomes of mothers to newborn infants to contrast bacterial and fungal community composition during this time. Study participants were recruited from the newborn period (Day of life; DOL 0-5, 13-17, and 26-35), infancy (DOL 83-115, 165-200 and Month of life; MOL 11-13), and later childhood (MOL 22-26, 33-39, and 57-63). Further, we tested for the presence of a potential probiotic, \textit{Lactobacillus plantarum}, with the potential to reduce newborn infections\textsuperscript{211} using a qPCR-based approach.
4.2 Material and Methods

Participant recruitment. Study participants were recruited in the Brong Afahoe region central in Ghana at the Kintampo Health Research Centre (KHRC; Kintampo, Brong Ahafo Region, Ghana). A well-trained cadre of community workers of the KHRC already conducts weekly home visits as part of their Community Based Surveillance Volunteers (CBSVs), where they identify pregnant women in the community, and then make home visits during pregnancy and three in the first week of life of the neonate to address essential maternal and neonatal care practices, and to assess and refer sick babies. Eligibility criteria for this study were for participants to be living in the study area, to be healthy with no known congenital defects, to be under 63 months or be mothers of children under 63 months, and to have consented to their participation in the study. An exclusion criterion was refusal to participate in the study by means of not providing informed consent. Participants were selected prospectively using the KHRC demographic surveillance database. Mothers of newborns recruited into the study were recruited during pregnancy. KHRC field workers approached the identified subjects and requested informed consent for mothers and their children’s participation. For those consenting to participate, field workers supplied mothers with sterile containers and scoops to collect stool samples from soiled diapers. Samples were transported to the KHRC within 30 minutes. Breast milk was hand-expressed into sterile screw-top containers and transported to the KHRC alongside stool samples. At the clinical laboratory, stool and breast milk samples were split into multiple aliquots in 1.7 mL Eppendorf tubes and stored at -80°C until further analysis. Samples were transported to the University of British Columbia on dry ice with temperature monitoring via World Courier Inc.
**Stool DNA extraction for amplicon sequencing.** DNA was extracted using the MagAttract PowerSoil DNA KF kit (Qiagen Cat. No. 27000-4-KF) using the manufacturer’s protocol for the KingFisher Flex platform. Approximately 200 mg bulk stool was loaded into each 96-well plate with sterile wooden picks.

**Breast milk DNA extraction for amplicon sequencing.** Total DNA was extracted from 0.7 – 1.0 µL breast milk using the Qiagen DNeasy PowerSoil DNA extraction kit (Qiagen Cat. 12830-50) with the following modifications: frozen breast milk samples were thawed on ice and transferred to 2.0 mL screw-top tubes (VWR Cat. 211-0440) and spun for 10 min at 4°C at 20,000 g to pellet all particles. The lipid portion of the milk remained on top of the aqueous phase after centrifugation. This layer was retained, and the aqueous phase between the pellet and lipid layer carefully removed with a 200 µL micropipette. The contents of one Qiagen 0.1 mm glass bead tube (Qiagen 13118-50) was then added to each sample with 500 µL Bead Solution and 200 µL phenol:chloroform:IAA pH 7-8 (Ambion Cat. AM 9730, pH adjusted with included Tris buffer). 60 µL of C1 was then added, and samples were homogenized in a Fasprep bead mill for 30s at 5.5 m/s for two cycles with a 5 min wait in between. The remainder of the protocol was carried out according to manufacturer’s instructions and samples eluted in 60 µL elution buffer.
16S amplicon sequencing

16S amplicon sequencing was done using the same V4-16S sequencing pipeline applied to mouse intestinal and fecal samples presented in Chapter 2.

PCR amplification. To analyze the microbial communities in mouse tissues and feces, alongside re-extracted human stool used for the experiment, we used amplicon sequencing targeting the V4 region of the 16S rRNA gene. PCR amplification was done using 1-10 ng of DNA, Thermo Phusion Hot Start II DNA Polymerase (ThermoFisher Cat. No. F549S), and the following primers:

16Sf V4
GTGCCAGCMGCCGCGGTAA

16Sr V4
GGACTACHVGGGTWTCTAAT

PCR cycle conditions were 98°C for 2 minutes, followed by 30 cycles of 98°C for 20 seconds, 55°C for 15 seconds, and 72°C for 30 seconds, and 72°C for 10 minutes following the cycling steps. PCR products were cleaned using the Agencourt Ampure XP beads (Beckman Coulter Cat. No. A6388) using a 0.8:1 bead-sample ratio and eluted in a final volume of 20 μL.

Library Preparation. Library preparation was done using a previously published SOP176 with the details and product information outlined below. 10 μL of the final PCR product was used to normalize using the SepalPrep Normalization Prep Plate Kit (ThermoFisher Cat. A1051001) to 1-2 ng/μl and 5 μl of each normalized sample was pooled into a single library per 96-well plate. Library pools were further concentrated using the DNA Clean and Concentrator kit (Zymo Cat.
D4013). A dilution series was performed for each of the pooled libraries for subsequent quality control steps. Each pool was analyzed using the Agilent Bioanalyzer using the High Sensitivity DS DNA assay (Agilent Cat. 5047-4626) to determine approximate fragment size, and to verify library integrity. Library pools with unintended amplicons were purified using the Qiagen QIAquick Gel Extraction Kit (Qiagen Cat. 28706). Pooled library concentrations were determined using the KAPA Library Quantification Kit for Illumina (KAPA Cat. KK4824). The final libraries were loaded at 8 pM, with an additional PhiX spike-in of 20%. The amplicon library was sequenced on the MiSeq using the MiSeq 500 Cycle V2 Reagent Kit (Illumina Cat. MS-102-2003)

Read processing and OTU binning. Paired-end reads were assembled using MOTHUR version 1.37.2, following the MiSeq SOP\textsuperscript{213} (https://www.mothur.org/wiki/MiSeq_SOP; accessed Aug 2018). OTUs were clustered at 97% identity and classified using the GreenGenes 13_8_99 database\textsuperscript{177}.

ITS2 amplicon sequencing. Processing for ITS2 amplicon sequencing was identical to that for 16S amplicon sequencing, except with the following differences: Primers and PCR conditions were as previously described\textsuperscript{214}. Paired-end reads were trimmed at the non-overlapping ends, and high-quality reads classified using UNITE (v. 7.1)\textsuperscript{215}.

16S and ITS2 sequencing analysis

Quality control (QC). The potential for contamination was addressed by co-sequencing DNA amplified from specimens and from four each of template-free controls and extraction kit
reagents processed the same way as the specimens (two for breast milk specimens). Two positive controls, consisting of cloned SUP05 DNA, were also included (number of copies = 2*10^6). Breast milk samples were extracted and PCR-amplified separately from the stool samples but sequenced on the same MiSeq run. Given the different distribution of contaminating OTUs from breast milk blanks compared to stool, and that PCR amplification for breast milk samples was done on separate plates from the stools, QC was done for the breast milk samples separately.

Contaminating OTUs can come from two main sources: those from PCR and extraction reagents, and carry-over from neighboring samples during the extraction and amplification process. Potential reagent contaminants for removal from the dataset were identified as those present in over 50% of blank samples whose count geometric mean plus one standard deviation was greater than that for the samples. Using this approach, OTUs with high counts in many samples (mainly including taxa commonly known to be highly abundant in these sample types) were not flagged for removal, despite their presence in blanks. For OTUs whose mean counts were high in the samples and were also found in over 50% of blanks, the mean count plus standard deviation from the blanks was subtracted from each of the samples. Samples with total counts under 1000 were also removed. Once filtered, data from breast milk and stool were recombined for further analysis.

**Statistical Analysis**

**Alpha diversity.** We used the *phyloseq* R package\textsuperscript{179} to compute alpha and beta diversity. We estimated microbial diversity with abundance-dependent (Shannon) and independent (Observed richness) metrics, after subsampling the OTU table to account for unequal sampling depth. To test for relationships between age bin, we used the Wilcoxon test comparing each age bin to the
Mothers 26-35 day-old infants, as this was considered the adult microbiome group. P-values were adjusted using the Bonferroni correction and adjusted p-values under 0.05 considered significant.

**Community composition.** Beta diversity was computed using the Bray-Curtis index. Sample clustering was visualized using non-metric-multidimensional scaling (NMDS). To test whether child age (all age bins except maternal samples) explained community composition, we used the Adonis test from R package vegan\textsuperscript{180} using age in days calculated between the time interval between participant’s date of birth and visit date. To test if community composition between 5-year-olds (age bin MOL 57-63) and younger children differed, we computed the Bray-Curtis distance of individuals in each bin to their distance to all individuals in the MOL 57-63 bin. The Wilcoxon rank-sum test was then used to compare distances between each bin to the median distance among the 5-year-olds to others in the same bin. P-values were adjusted using the Bonferroni correction and adjusted p-values under 0.05 considered significant.

**Differential Abundance.** To identify differentially abundant OTUs between mothers of 0-5 day old vs. 26-35 day-old infants, we used the Wald test implemented in the R package DESeq\textsuperscript{2183}. OTUs were considered significant if with a Benjamini-Hochberg adjusted p-value under 0.05. To further visualize select bacterial genera differentially abundant between maternal groups and their abundance across the age spectrum, we aggregated OTU counts by genus-level assignment and confirmed that the genera differed between groups using the Wilcoxon rank-sum test.
**Shared OTUs between mother and infant.** For shared OTU analysis, OTUs were considered present if their relative abundance was over 0.01%. For stool and breast milk separately, we calculated the number of shared OTUs between newborns and their mothers and newborns and unrelated mothers. Because the number of unrelated pairs was much greater than related pairs, we randomly sub-sampled the unrelated pairs, selecting the same number as related pairs in each analysis. The median number of shared OTUs between related and unrelated pairs and 0-5 and 25-36 days post-partem was then compared using the Wilcoxon rank-sum test.

**Detection of Lactobacillus plantarum DNA in newborn stools**

**On-site DNA extraction for qPCR.** Total DNA was extracted from 200-300 mg of stool using the Qiagen QIAamp Fast DNA Stool mini kit (Qiagen Cat. 51604) at the KHRC with the following modifications: 15-15 mg 0.1 mm Zirconia beads (Qiagen 13118-50) were added to the Inhibitex/stool mixture following the 10-minute incubation (step 3 of the protocol) and vortexed for 2-3 minutes. The bead/stool suspension was then centrifuged at 13,000 RPM to pellet stool particles and the remainder of the protocol carried out as per manufacturer’s instructions.

**qPCR for Lactobacillus plantarum.** *L. plantarum* DNA presence in stool samples was assessed for stools of newborns under 36 days of age using DNA extracted on-site. qPCR was done in the research laboratory at the University of British Columbia. The primer pair used for detection targeted the 16S-23S intergenic spacer region specific to *L. plantarum*, as previously published$^{216, 217}$:

**Lpn-1:** TGG ATC ACC TCC TTT CTA AGG AAT

**Lpn-2:** TGT TCT CGG TTT CAT TAT GAA AAA ATA
Primers were reconstituted in sterile dH2O (Invitrogen Cat. 10977-015) to a working stock of 10 µM. Standard curves were constructed with *L. plantarum* genomic DNA (ATCC Cat. BAA-793D-5) from $10^7$ to $10^4$ copies in duplicate. Each 10 µL PCR reaction contained 5 µL SYBR PCR master mix (BioRad Cat. 172-5124), 0.3 µL Lpn-1 and Lpn-2 primers, 0.4 µL sterile dH2O and 1 µL sample, ranging from 1 to 25 ng. PCR cycle conditions were as follows: 50°C for 2 min, 95°C for 3 min, 40 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec using the Applied Biosystems StepOnePlus PCR system. Samples were considered positive for *L. plantarum* genomic DNA if their Ct values were above 37.

### 4.3 Results

#### 4.3.1 16S amplicon sequencing summary

We collected stool samples from 14 to 16 infants or children per age bin across the first 5 years of life, alongside stool and breast milk from mothers of 0-5 or 26-36 day-old newborns (Table 4.1). After quality filtering, we retained a total of 2.9 million reads, representing 4492 bacterial OTUs. These OTUs represented 435 unique taxa classified at the genus level or above. Five samples contained fewer than 1000 reads and were excluded from downstream analysis. The remainder contained between 2.6 and 31.2 thousand reads per sample (Figure 4.1A).
Table 4.1 Participant recruitment summary

<table>
<thead>
<tr>
<th>Sex: n (%)</th>
<th>n</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOL 0-5</td>
<td>15</td>
<td>9 (60.0)</td>
<td>6 (40.0)</td>
</tr>
<tr>
<td>DOL 13-17</td>
<td>16</td>
<td>8 (50.0)</td>
<td>8 (50.0)</td>
</tr>
<tr>
<td>DOL 26-35</td>
<td>15</td>
<td>4 (26.7)</td>
<td>11 (73.3)</td>
</tr>
<tr>
<td>DOL 165-200</td>
<td>16</td>
<td>5 (31.2)</td>
<td>11 (68.8)</td>
</tr>
<tr>
<td>DOL 83-115</td>
<td>14</td>
<td>7 (50.0)</td>
<td>7 (50.0)</td>
</tr>
<tr>
<td>MOL 11-13</td>
<td>13</td>
<td>8 (61.5)</td>
<td>5 (38.5)</td>
</tr>
<tr>
<td>MOL 22-26</td>
<td>17</td>
<td>9 (52.9)</td>
<td>8 (47.1)</td>
</tr>
<tr>
<td>MOL 33-39</td>
<td>16</td>
<td>7 (43.8)</td>
<td>9 (56.2)</td>
</tr>
<tr>
<td>MOL 57-63</td>
<td>15</td>
<td>8 (53.3)</td>
<td>7 (46.7)</td>
</tr>
<tr>
<td>Mothers_DOL 0-5</td>
<td>15</td>
<td>15</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Mothers_DOL 26-35</td>
<td>15</td>
<td>15</td>
<td>0 (0.0)</td>
</tr>
</tbody>
</table>

DOL, Day of life; MOL, Month of life

4.3.2 ITS2 amplicon sequencing summary

After quality filtering, we retained a total of 2.5 million reads representing 1772 fungal OTUs, representing 375 taxa classified at the genus level or above. Fifty-nine samples contained fewer than 1000 reads and were excluded from downstream analysis. The remainder contained between 1.2 and 96.5 thousand reads per sample (Figure 4.1B). Samples with no detectable fungi over our count threshold were found in both infant and adult stool samples alike. The proportion of samples in each age bin below that threshold was highly variable (Figure 4.1C).
Figure 4.1 Amplicon sequencing coverage for bacterial and fungal communities
A. Amplicon sequencing allowed for bacterial profiling of nearly all samples at a sequencing depth over 10,000 for most samples.  B. For fungal profiling, several samples did not produce sufficient ITS2 sequencing reads for downstream analysis.  C. Samples with no detectable fungi spanned the age spectrum. Abbreviations: bm, breast milk; DOL, day of life; MOL, month of life
4.3.3  Bacterial and fungal alpha diversity across the age spectrum

To measure alpha diversity, we used both observed richness, the number of distinct OTUs present, and the Shannon Diversity index, which reflects both richness and evenness. We compared diversity of each age bin to that of the one month postpartum mothers, representing an adult microbiome. Alpha diversity of bacterial communities increased with age. Observed richness was lower than that of adults until month of life (MOL) 33-39, and Shannon diversity was lower until MOL 22-26 (Figure 4.2A). To our surprise, alpha diversity of fungal communities did not follow this or any other clear pattern. Gut fungal microbiome diversity of only a few age bins differed from adult diversity, and there was no clear trend for diversity versus age (Figure 4.2B). Overall, greater numbers of bacterial taxa than fungal taxa were found per sample than fungal, where most samples from after 2 years of age contained less than 60 fungal OTUs but between 100 and 300 bacterial OTUs.

Figure 4.2  Bacterial and fungal alpha diversity across the age spectrum (A) and fungal communities (B). Coding indicates values significantly different from those of adult communities (Mothers of 26-35 day–old newborns): **** p < 0.0001, *** p < 0.001, ** p < 0.01, * p < 0.05; Wilcoxon Rank Sum test adjusted using the Bonferroni correction.
4.3.4 Bacterial and fungal communities across the age spectrum

Non-metric multidimensional scaling (NMDS) based on Bray-Curtis distances revealed distinct community structures among bacterial and fungal communities across the age spectrum and, expectedly, in relation to breast milk communities (Figure 4.3A-B). While child age contributed to 18% of the variance for bacterial communities (PERMANOVA p = 0.001), it accounted for only 2% of variance in fungal communities (PERMANOVA = 0.015). We did not identify one NMDS axis that drove an age gradient for either community. Thus, to further assess community composition across childhood, we measured the distance between age bins by assessing the Bray-Curtis distance between all samples within an age bin to the oldest child age bin, MOL 57-63. For bacterial communities, all microbiomes from all age bins differed from this 5-year time point except MOL 33-36. Meanwhile, the distance between 5-year-olds and all other ages did not differ for fungal communities (Figure 4.3C). Taxonomic composition of the top 25 most abundant Genera in each dataset clearly revealed Bacterial taxonomic shifts over the age spectrum. In keeping with the literature on bacteria\textsuperscript{63,66}, newborns were dominated by \textit{Escherichia coli} in the first week of life, which persisted up to DOL 165-200. \textit{Bifidobacteria} and \textit{Bacteroides} then dominated up to MOL 11-13. Subsequently, community diversity increased and taxa associated with adult microbiomes dominated, including \textit{Prevotella} and \textit{Faecalibacterium} (Figure 4.3D). Conversely for fungi, \textit{Candida} (primarily \textit{C. albicans} and \textit{C. tropicalis}) dominated throughout the age spectrum. \textit{Malassezia} featured more prominently in the first 3 months of life, and \textit{Aspergillus}, after the first year (Figure 4.3E).
Figure 4.3 Bacterial and Fungal communities across the age spectrum
A. Bacterial community composition follows a trajectory across the age spectrum. B. Fungal communities do not follow a defined gradient. C. Bray-Curtis distance differs between 5-year-olds and younger infants up to MOL 22-26 for bacterial communities, but for fungal communities only 0-5 day-old newborns are significantly dissimilar from 5-year-old children. D-E. Taxonomic composition of the 25 most abundant genera in bacterial (D) and fungal (E) communities across the age spectrum. **** p < 0.0001, *** p < 0.001, ** p < 0.01, * p < 0.05, Wilcoxon Rank Sum test compared to 57-63 month-olds, adjusted using the Bonferroni correction.
4.3.5 Maternal bacterial microbiome

Unexpectedly, our study revealed dramatic differences in stool microbiomes of mothers in the first postpartum week versus those of mothers 1 month postpartum (Figure 4.3A). Furthermore, clear shifts in their taxonomic profiles were also evident (Figure 4.3D). We further explored this difference by first measuring the Bray-Curtis distance between mothers’ microbiomes at both time points to their newborn’s (DOL 0-5) microbiome, finding that the one week postpartum mothers’ bacterial microbiomes were significantly more similar to newborn bacterial microbiomes in the first week of life (Figure 4.4A). Using DESeq2, we identified 68 differentially abundant OTUs between the two mother groups, representing 34 bacterial taxa at the genus level or above. These included differences in bacterial taxa that contributed to increased similarity to newborns; a decrease in the relative abundance of *Prevotella*, which is absent in newborns and only dominated during the 2nd year of life, and an increase in *Escherichia* (*E. coli*), which was more highly abundant in newborns than older infants or children (Figure 4.4C-D). However, other taxa of greater relative abundance were found that were nearly absent in newborns such as *Faecalibacterium* and *Blautia* (Figure 4.4E-F). Given this finding, we also tested whether newborns share OTUs with their mother’s stool more so than with unrelated mothers’ stool, finding that this was significant for DOL 0-5 but not DOL 26-35 pairs (Figure 4.4G). OTUs most commonly shared by mother-infant pairs include members of *Escherichia*, *Enterococcus*, and *Streptococcus*, all of which are of higher relative abundance in the newborns than in the mothers. A parallel analysis for fungal data did not identify any OTUs shared between mother’s stools and their newborns (not shown).
Figure 4.4 Postpartum maternal and newborn infant microbiome
A. Average of pairwise similarity of maternal microbiome versus corresponding infant microbiome on the same sample collection date. B. Differentially-abundant bacterial OTUs identified by DESeq2 between maternal microbiome one week and one month postpartum. Negative fold-difference, indicates greater abundance at one week. C-F. Bacterial taxa that
differ by post-partem time include shifts that bring the mothers closer to the newborn microbiome (C; reduction in *Prevotella* and D; increase in *Escherichia*) but also include increases in taxa not prevalent in newborn microbiomes (E; *Faecalibacterium*, F; *Blautia*). G-H. Number of bacterial OTUs shared between mother-infant pairs is significantly greater than those shared between unrelated pairs for DOL 0-5 infants only (G). OTUs most commonly shared between mother and newborn stools include *Escherichia* and *Enterococci*, and *Streptococci*. Statistics: G. **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, Wilcoxon Rank Sum test, adjusted using the Bonferroni correction.

### 4.3.6 Bacterial and fungal communities in mother’s breast milk and their newborns stool

Breast milk is a well-recognized modulator of the infant gut microbiome, serving both as a source of colonizing bacteria and bacterial growth substrates. This has been less well described for fungal communities. Thus, we sought to compare the community composition and OTUs shared between both microbiomes. Breast milk microbiomes were dominated by skin taxa of both bacteria (*Streptococcus*, *Staphylococcus*, and *Corynebacterium*; Figure 4.5A) and fungi (*Malassezia*, including *Malassezia restricta*, *globosa*, and *furfur*; Figure 4.5B). For bacterial communities, mother-infant pairs shared significantly more OTUs than unrelated pairs at both postpartum time points (Figure 4.5C). While fewer pairs were available for fungal data, we found a significant difference between related and unrelated pair sharing for the 26-35 day time point only (Figure 4.5D). The most shared bacterial OTUs belonged to common skin commensal genera, *Streptococcus* and *Staphylococcus*, which were more abundant in the breast milk than in the newborn stool. Shared OTUs, also included gut commensals *Bifidobacterium* and *Bacteroides*, both of which were of higher abundance in the newborns stool than in their mother’s milk (Figure 4.5E). Shared fungal OTUs were much more limited, and did not have a clear pattern of differing abundance between mother and child (Figure 4.5F).
Figure 4.5 Breast milk Bacterial and Fungal community composition

A-B. Top 25 most abundant bacterial (A) and fungal (B) genera across all samples reveal that both communities are dominated by common skin taxa, but also contain a diverse repertoire of microbes at lower abundances.

C-D. OTUs shared between mother-infant pairs are significantly
greater than the number of OTUs shared between a random sample of unrelated mothers and infants for bacteria (C). But for fungi this was the case only at 26-35 days postpartum (D). E-F. Most common shared bacterial (E) and fungal (F) OTUs between mother-infant pairs. Statistics. C-D: **** p < 0.0001, *** p < 0.001, ** p < 0.01, * p < 0.05, Wilcoxon Rank Sum test, adjusted using the Bonferroni correction.

4.3.7 Occurrence of Lactobacillus plantarum

Recent findings indicate that a synbiotic administration of *Lactobacillus plantarum* (*Lp*) in the first week of life significantly reduce incidence of newborn infection\(^{211}\). There is great interest in expanding the use of this probiotic to other regions with high burdens of newborn infection, but there is also reluctance to introduce a bacterial species not normally found in newborns. Thus, we sought specifically to determine the prevalence of this species in our population prior to considering a trial. Using unfiltered data only (i.e, data retaining OTUs with fewer than 3 counts across the dataset), we found that only one sample in the first month of life contained an OTU classified as *Lp*. Given the low expected relative abundance of *Lp*, we also used a qPCR assay specific for *Lp* using DNA extracted from separate stool aliquots at the study site, and found that 12 newborn stools (29\%) were positive for *Lp* (Table 4.2).

<table>
<thead>
<tr>
<th>Table 4.2 Occurrence of <em>Lactobacillus planarum</em> in newborn stools</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
</tr>
<tr>
<td>Positive for Lp: n (%)</td>
</tr>
</tbody>
</table>

4.4 Discussion

Here, we report for the first time the combined bacterial and fungal community dynamics during early childhood. In conducting a direct comparison between bacterial and fungal communities,
we demonstrated how the two follow very distinct developmental trajectories, and were shared differently between mothers and their newborns. We also found a previously unreported and highly surprising compositional difference in maternal stool microbiomes by postpartum period.

The absence of fungi in a large subset of stool samples is consistent with previous findings, where fungi have been undetectable in infant and adult samples alike. However, the samples with undetectable fungi contained readily amplifiable bacterial genes, even at the youngest age groups, thus making the absence of fungi in these samples more likely a reflection of their biological absence or extremely low abundance rather than sample processing artefacts. Interestingly, most breast milk samples were positive for fungi, especially 26-35 days postpartum, while we did not detect any fungi in half of newborn stools at that age.

Increasing bacterial and unchanging fungal richness and diversity during infancy have both been reported previously, but never together in the same study. Bacterial richness and diversity in our study remained stable over the first year and dramatically increased after the first year of life. In contrast, diversity in other studies on rural, low-resource settings increased steadily from birth to two years or 3 months to 5 years in American and European cohorts, respectively. However, numbers of observed OTUs also increased steadily from birth to 3 years among Amerindian and Malawian infants. Thus, the different trend in our study may reflect lifestyles and environmental exposures unique to this study cohort in Ghana.

The bacterial succession, with early dominance by *E. coli*, *Streptococcus*, and *Staphylococcus* in the first weeks, followed by *Bifidobacterium* and *Bacteroides* until the first year, and subsequent high abundance of common adult taxa such as *Prevotella* and *Faecalibacterium* were also consistent with previous studies of newborn and adult microbiomes. The Bacteroidetes phylum in older children or adults is typically dominated by
Bacteroides or Prevotella but not both\textsuperscript{10, 90, 93}. As our study was done in a population dominated by Prevotella in adulthood, we also saw that this genus only colonized infants between the first and second year of life, where beforehand Bacteroides still dominated. The ability of Bacteroides species to utilize HMOs\textsuperscript{218} as well as diet-derived polysaccharides\textsuperscript{219} as energy sources, is thought to contribute to this taxon’s unique ability to persist in both infant and adult guts. However, in this study Bacteroides was eventually outcompeted by Prevotella associated with a fiber-rich diet in adults\textsuperscript{95}.

Fungal microbiome dominance by Candida sp. has been found in some\textsuperscript{108, 112} but not other\textsuperscript{110, 111} studies, where fungal taxa such as Aspergillus and Penicillium feature more prominently. We also detected an early increased abundance of Malassezia that was not seen in other newborn studies\textsuperscript{108, 109}. Given the high relative abundance of this taxon in breast milk in our study and another\textsuperscript{136}, and evidence of its transmission from mother to newborn\textsuperscript{220}, absence of this taxon in other studies is puzzling but may reflect differences in maternal breast milk mycobiomes or sample processing.

We were surprised to find substantial differences between the bacterial microbiomes from mothers of 0-5 and 26-35 day old newborns. These dramatic differences were not due to cross-contamination between maternal and infant samples, as mothers in both groups were sampled using the same protocol and time points making it highly unlikely that cross-contamination occurred among the mothers of 0-5 but not 26-35 day-old newborns. Given the cross-sectional study design, we cannot confirm that the microbiome has changed in these women over this time. However, the 3-month recruitment window and the increased similarity of the earlier postpartum mothers’ microbiomes to newborn microbiomes specifically suggest that the mother’s microbiome changed between childbirth and the 4\textsuperscript{th} post-partem week. Studies of
the stool microbiome during pregnancy have yielded mixed results, with some studies showing no change in community composition during pregnancy\textsuperscript{221, 222} and others showing substantial shifts\textsuperscript{223, 224}, including an increase in Proteobacteria that we also observed in mothers 0-5 days post-partum. However, our finding of a reduction of \textit{Prevotella} relative abundance is entirely novel, and was one of the most differentially abundant taxa in our study. Also, we observed a greater abundance of \textit{Faecalibacterium} and \textit{Blautia} 0-5 days post-partum. \textit{Faecalibacterium} is associated with decreased inflammation and protection from inflammatory bowel diseases\textsuperscript{225}. \textit{Blautia} was increased in healthy controls compared to type 1 diabetes patients\textsuperscript{226} and in pregnant women diagnosed with gestational diabetes mellitus (GDM) that then followed prescribed dietary intervention compared to those diagnosed with GDM with no change in diet\textsuperscript{224}. Together, our findings suggest that differences in maternal microbiome around the time of birth reflect both increases in taxa common to newborns (\textit{Escherichia}) and decreases in taxa absent in newborns (\textit{Prevotella}) but also increases in abundance of taxa with potential health benefits to the mother (such as \textit{Faecalobacteria} and \textit{Blautia}). Accordingly, the most shared OTU between mother’s stool and newborns was classified as \textit{E. coli}. Also, while we did not identify significantly different fungal taxa by postpartum time, we also did not identify commonly shared fungal OTUs between pairs, suggesting that the bacterial microbiome specifically may undergo changes to accommodate newborn colonization while fungal communities do not.

Common skin taxa dominated breast milk, including bacterial (\textit{Streptococcus}, \textit{Staphylococcus}) and fungal (\textit{Malasezzia}) communities, as previously shown\textsuperscript{132, 136, 227, 228}, and the bacterial taxa commonly shared between mother-infant pairs include not only skin taxa, but also other gut commensals \textit{Bifidobacterium}, \textit{Bacteroides}, \textit{Lactobacillus} and \textit{Rothia} have also been found\textsuperscript{227}. For \textit{Bacteroides}, sharing between mother and infant was seen only 6 weeks but not 1 or
3 weeks postpartum, but this was seen already in the first week of life in our study. Overall, our study adds strong and novel data indicating that the degree of sharing between mother and infant microbiomes is much greater for the bacterial microbiome than fungal.

Our study is limited by its cross-sectional design, leaving it impossible to conclude whether differences in maternal microbiomes represent changes over time. Our sample size only allowed for an analysis focussed on effect of age, but not other host factors that could also have shaped this community. And given the pilot nature of this cohort, data on child and mother health, and more importantly, mode of delivery, were not available. As is the case with any sequence-based approaches, we cannot distinguish microbial colonizers from DNA transiently passing through the infant’s intestine. In the case of bacterial genera *Streptococcus* and *Staphylococcus*, and the fungal genus *Malassezia*, their dominance in breast milk necessitates the presence of microbial DNA in the newborn’s stool – but whether these taxa reside and are active in the gut would require further culture-based approaches.
Chapter 5: Small Sample Big Data: Towards analysis of the human immune system in the newborn period in low resource settings

5.1 Introduction

While the microbiome shapes host immune responses throughout our entire lifespan, the host-microbiome interaction in the early neonatal period likely has the most impact on host immunity. This is all the more important, as globally 2.8 million neonates die in the first month of life, with three-quarters of these deaths occurring in the first week. In low-income countries, where most of these deaths occur, mortality for children ages 1-59 months has declined by an average 4.6% per year since the year 2000 – an improvement from only 2.4% per year in the decade preceding. However, progress on reducing neonatal mortality has lagged, with a decrease of 2.5% per year since the year 2000 – in incremental change from 1.7% per year in the preceding decade. As a result, today neonatal deaths make up nearly half of all childhood deaths worldwide. Twenty percent of neonatal deaths are attributed to possible serious bacterial infection (pSBI) – a broad term for all severe infections, including sepsis, meningitis, and lower respiratory tract infections (LRTI). More importantly, while pathogenic microbes such as bacteria can trigger a cascade that ultimately leads to mortality, it is increasingly recognized that it is the newborn’s own immune response that causes much of the damage. The vast majority of these newborn ‘infectious’ deaths occur in low resource settings. Despite this, immune development in this vulnerable population remains vastly understudied.

Most current efforts to reduce the global burden of early-life infection do not target the neonatal period. Vaccines, our most powerful tool to combat disease in infancy, are mostly
administered outside the neonatal window or do not target the pathogens responsible for neonatal infectious morbidity and mortality. Development of appropriate interventions require in-depth knowledge of the immune responses unique to newborns. Furthermore, little is known about how the microbiota that the newborns is exposed to in the first few hours to days of life impact on immune development during this critical period. The limited amount of sample available for research purposes no doubt contributed to this knowledge gap.

Modulating the human microbiome in early life holds promise. In fact, probiotic administration is already used to reduce incidence of NEC in premature human infants, and most recently pSBI including late-onset sepsis and LRTI in term newborns. However, there is a long way to go, and to advance this field, the human newborn immune system must be more accessible to scientific investigation, and the assays employed feasible for execution in low-resource settings, where newborns face the biggest burden of disease.

To this end, a collaborative opportunity arose to join forces with an epidemiological study site in Guinea-Bissau (GB), a country ranked fifth in the world for newborn mortality rate. The Bandim Health Project (BHP; www.bandim.org) has been conducting epidemiological research on childhood mortality in GB since 1978, and currently maintains demographic surveillance on over 100,000 inhabitants in the capital Bissau. BHP is also closely liaised with the only tertiary care centre in the country, Hospital Nacional Simão Mendez (HNSM), where BHP staff are responsible for providing BCG and OPV vaccines to newborns. Following on previous observations that newborn vaccination with the bacille Calmette-Guerin (BCG) vaccine reduces all-cause mortality in the first week of life, the BHP initiated the BCG-Immediate study (BCGiMED; clinical trials.gov registration NCT01989026) at the newborn nursery at HNSM. Here, following consent from the mothers, newborns admitted to nursery care were
randomized to BCG and OPV vaccines at birth vs. at hospital discharge, which is the regular policy. Newborns were followed for in-hospital survival, and the over several home visits over the first year of life. However, what was lacking in this study was any effort to delineate the mechanistic insight into both differences between vaccinated and unvaccinated infants, and immunological development in the newborns, alongside perturbations predisposing newborns to pSBI and subsequent death. Working closely with BHP, we designed an immunological sub-study (the BCGIMED-Immune sub-study) with the goals of i) identifying predictors of newborn mortality and ii) changes induced by the BCG/OPV vaccines associated with survival in newborn blood specimens using OMIC technologies. OMIC approaches are not without their challenges; they require highly standardized sample processing and diligent adherence to protocol to avoid the “garbage-in, garbage-out” trap whereby poorly processed specimens provide data shaped more by artifact than biology. Sampling protocols implemented in such low resource settings with limited laboratory infrastructure must be extremely simple, require minimal equipment, and include stringent quality control steps all while accommodating safe and feasible blood volumes that could be taken from newborns. I successfully developed a sample processing protocol that required a minimal blood volume of only 0.4 mL, could be successfully implemented in this low-resource setting, yet still yielded a wealth of information from flow cytometry, proteomic, metabolomics, and transcriptomic profiling (See Appendix A for study materials; A.1 Study SOP, A.2 sample processing forms, and A.3 Participant information, all supplied in English and Portuguese).

I went to Guinea-Bissau several times to train the local staff and maintained supervision of study progress even from a distance over the coming years. My protocol was subsequently modified for 1 mL blood, and implemented at two low-resource study centres (The Gambia,
Medical Research Council and Papua New Guinea, Institute of Medical Research) as a part of the larger collaborative EPIC-Human Immune Project Consortium (EPIC-HIPC), an international team of scientists with the goal of shedding light on immune ontogeny in low-resource settings to better understand newborn responses to vaccination. The pilot data generated by the EPIC-HIPC study, using the protocol I originally designed for BHP, allowed for the first time an in-depth look at the changing immune landscape measured over the first week of human life by flow cytometry, whole blood transcriptome, and plasma proteome as well as metabolome profiling.

I specifically designed the Guinea-Bissau protocol to fit the specific research site where it was to be implemented and at the same time removed as many processing artefacts as possible as they could mask biological signatures in future OMIC analysis. The protocol was successful both because of its simplicity but also deliberate design rationale for each step adapted to the local set up. First, because sterile venipuncture was not feasible in Guinea-Bissau, RNA samples were preserved immediately after blood draw to (i) avoid changes to the blood transcriptome due to possible contamination; and (ii) whole blood transcriptomes are known to be heavily altered already within 4 hours of collection. Optimal plasma processing for proteomic analysis should also avoid delays between collection and processing, as blood cells can break down extracellular proteins and release intracellular contents into the plasma. Cold storage alters innate immune cell composition for flow cytometry staining; this led us to maintain samples at ambient room temperature at all times post blood collection. Taking all these into account, blood was spun immediately (after 200 µL was removed for RNA processing) to separate plasma from the cellular fractions. We supplied a microcentrifuge to be stored at the maternity ward for this purpose. Because multiple plasma aliquots were required for different downstream purposes (i.e,
proteomic and metabolomics analyses), plasma was stored in two homogenous aliquots of 25 µL; they were generated by removing 50 µL plasma, mixing the sample by micropipetting, and then transferring 25 µL to a second microcentrifuge tube for storage. Flow cytometry processing of the cellular fraction that remained after plasma removal needed to accommodate several logistical challenges. The National Research Laboratory in Bissau has designated one workroom for BHP projects. While it is equipped with a workspace and two biosafety cabinets, there is no -80°C storage available. The only -80°C freezer available in the country is maintained by BHP personnel in a storage room located at the living compound. Thus, any samples processed in the lab would still have to be transported prior to cold storage, adding to processing time and increasing cellular degradation prior to storage. A workspace was instead set up next to the -80°C freezer on the BHP compound allowing specimens to be frozen immediately after processing, including a second microcentrifuge supplied for this study.

Here, I present the protocol design that enabled both these studies, together with evidence of the successful implementation in GB, and the immune ontogeny revealed by both flow cytometric and plasma cytokine measurements of EPIC-HIPC newborns over the first week of life based on my analysis of those data. This platform will continue to serve as an invaluable tool to study immunity and host-microbe relationships in the newborn period.

5.2 Materials and Methods

The BCGIMED study

Samples collected from Guinea-Bissau were auxiliary to the BCGIMED study (clinical trial registration number NCT01989026), where newborns admitted to the neonatal nursery at the tertiary care center Hospital National Simao Mendes were randomized to receive BCG-Danish
strain 1331 (Statens Serum Institute) and oral polio vaccine (OPV) vaccines either at birth, or at hospital discharge as per regular practice for newborns admitted to nursery care. Inclusion criteria were 1 minute Apgar scores $\geq 2$ and admission weight $\geq 1,250$ grams. Exclusion criteria were having gross malformations, being moribund, or prior vaccination. Consent was requested immediately following enrolment in the main BCGIMED trial. Newborns whose mothers consented to their participation were bled 24 hours after randomization, when half of newborns had received the vaccines and half were unvaccinated.

**BCGIMED Peripheral blood processing.** At the newborn nursery, capillary blood (400 µL) was collected via heel prick into lithium-heparin microtainers (BD Cat. 365965). Aliquots of 200 µL were immediately placed into a 2.0 mL cryovial (Corning Cat. 430488) pre-loaded with 1.3 mL RNA Later™ (Invitrogen, AM7020). The remaining blood sample was transferred to a 0.5 mL microcentrifuge tube (Diamed Cat. SPE055-N) and spun at 500 x g for 10 minutes. 50 µL plasma was then removed and placed in a 2.0 mL microfuge tube (VWR Cat. 16466-060), mixed by pipetting up and down 10x, and 25 µL transferred to a second 2.0 mL microfuge tube. The blood specimen was then transported to the BHP staff living compound, where a -80°C freezer is located a storage room, and is closely monitored by BHP personnel with generator support upon the inevitable power outages. As cells need to be frozen shortly after processing, the remainder of sample processing was done here. 0.5 mL tubes with blood were short-spun for 5 seconds to remove blood from the lid. Whole blood was resuspended and transferred to a 2.0 mL microfuge tube pre-loaded with 1.4 mL 1x BD FACSLyse prepared by diluting 10X BD FACSLyse (BD Cat. 394202) in deionized water (Invitrogen Cat. 10977-015) and stored at 4°C. Samples were
incubated at room temperature for 10 minutes, and centrifuged at 2500 RPM for 5 minutes. Supernatant was discarded, and the cell pellet resuspended in 200 µL 1xFACSLyse. All processed samples were then immediately stored at -80°C. Samples were subsequently ground-transported to the Medical Research Council (MRC, Gambia) on dry ice by study personnel via multiple trips crossing 2 borders over the course of the study period, and then flown to the University of British Columbia under temperature controlled and monitored conditions (World Courier; New Hyde Park, NY, USA) over three shipments.

**BCGIMED Whole blood RNA extraction.** Total RNA was extracted from each sample using the Human RiboPure RNA purification kit (Ambion™ ThermoFisher; AM1928) following the manufacturer’s protocol. Quantification and quality assessment of total RNA was performed using an Agilent 2100 Bioanalyzer (Santa Clara, CA, USA).

**The EPIC study**

This study is registered under Clinical Trials Number NCT03246230. In the Gambia and PNG (EPIC consortium study sites), following informed consent, mothers were screened for HIV-I and –II, and Hepatitis B with positivity for either virus representing an exclusion criterion. Inclusion criteria were a healthy appearing infant as determined by physical examination, born by vaginal delivery at gestational age of ≥ 36 weeks, 5 - minute Apgar scores ≥ 8, and a birth weight of ≥ 2.5 kilograms. Our non-vaccinated control group had samples obtained before the day of life 7 administration of BCG-Russia, OPV and hepatitis B vaccine (all from Serum Institute of India) as per local standard of care. The treatment group of newborns received BCG, OPV and hepatitis B vaccine at birth. Peripheral blood samples were obtained from all infants on the day of birth and
one additional sample was then collected at either day 1, 3 or 7 post-vaccination in order to reduce venipunctures to a maximum of two within the first week of life. This allowed contrasting of vaccinated vs. unvaccinated newborns over the first week of life.242

**EPIC Peripheral Blood Processing.** Venous blood was collected directly into 4 mL pre-heparinized collection tubes (75 USP, VWR Cat. 367871) by “drip-blood” collection method until 2 mL was collected. Aliquots (~ 200 µl) were immediately placed in RNALater™ (Invitrogen, AM7020) with the remaining blood kept in the collection tubes at RT until further processing within 4 hours. Whole blood was centrifuged on site at 500 g for 10 minutes at RT, and plasma removed. The amount of plasma removed from the whole blood after centrifugation was subsequently replaced with RPMI. For later assessment of cellular composition by Flow Cytometry, EDTA (0.2 mM final concentration; Invitrogen Cat. 15575-038) was added to the whole blood / RPMI mixture to ensure adherent cells were not lost. 225 µL blood sample aliquots were stained for viability with 2 µL of fixable viability dye (eBioscience, Cat 65-0865-14) for 30 minutes at 4°C in the dark, then red blood cells were lysed and cells fixed in 350 µL Stable-Lyse V2 buffer (Smart tube, cat. STBLYSE2-250) for 15 minutes at room temperature, then 1 mL of Stable-Store V2 (Smart tube, cat. STBLSTORE2-250) was added and incubated for 15 minutes at room temperature prior to storage at -80°C. Samples were subsequently shipped to UBC under temperature controlled and monitored conditions (World Courier; New Hyde Park, NY, USA).

**EPIC Flow Cytometry.** At the immunophenotyping laboratory flow cytometry samples were thawed, washed in staining buffer (PBSAN; 0.5% BSA, 0.1% sodium azide) and stained on ice in PBSAN with a cocktail of anchor markers to determine the frequency and the absolute counts
of immune cells in peripheral blood. Cell surface antibodies included CD45 (V450; clone HI30; BD Cat. No. 560367), CD3 (PE-CF-594; clone UCHT1; BD Cat. No. 562280), CD56 (BV650; clone HCD56; Biolegend Cat. No. 318343), CD16 (FITC; clone 3G8; Pharmingeen Cat. No. 560996), γδ TCR (PE; clone B1.1; eBiosciences Cat. No. 12-9959-42), HLADR (eFluor605; clone LN3; eBiosciences Cat. No. 83-9956), CD11c (APC; clone Bu15; BD Cat. No. 340544), CD123 (PE-Cy7, clone 6H6, eBiosciences Cat. No. 25-1239-42), CD19 (BV711; clone HiB19; Biolegend Cat. No. 302245), CD11b (BV786; clone ICRF44; Biolegend Cat. No. 301346), and CD123 (PE-Cy7; clone 6H6; eBiosciences Cat. No. 25-1239-42). These markers were used to identify T cells (CD45+/CD3+/CD56−/CD16−), γδ T cells (CD45+/CD3+/γδ TCR+), B cells (CD45+/CD3−/HLADR+/CD11c−/CD123−/CD19+), classical monocytes (CD45+/HLADR+/CD14+/CD16−), non-classical monocytes (CD45+/HLADR+/CD14+/CD16+), NK T cells (CD45+/CD3+/CD56+/CD16+), myeloid dendritic cells (CD45+/CD3−/HLADR+/CD11c+), plasmacytoid dendritic cells (CD45+/CD3−/HLADR+/CD123+/CD11c+), CD56-high NK cells (CD45+/CD3−/HLADR−/CD16−/CD56hi), CD56-dim NK cells (CD45+/CD3−/HLADR−/CD16−/CD56dim), mature neutrophils (CD45+/Boolean gate excluding all other cell types/CD16+/CD11b+), immature neutrophils (CD45+/Boolean gate excluding all other cell types/CD16−/CD11b+/−), eosinophils (CD45hi−+/CD16−), and basophils (CD45+/CD3−/CD16−/CD56−/CD123+). Flow cytometric analyses were acquired on a custom-built LSRII flow cytometer (BD Biosciences) and compensation set according to the methods outlined in Chapter 3, with the following modifications. TruCount beads (BD Cat. No. 340335) were added to each sample immediately after they were thawed. 240,000 beads were added per sample. To calculate cells per microliter blood, the cell counts in each sample were divided by the bead count in each sample, then multiplied by 240,000 times.
the original blood volume of 225µL. All further data analyses were done with the cells per µL blood values.

**EPIC Luminex Cytokine quantification.** Plasma (25 µl) was used to measure cytokine concentrations using a custom-designed multi-analyte Cytokine Human Magnetic Panel bead array, (Invitrogen/Life Technologies, Carlsbad; CA) consisting of CCL2, CCL3, CCL5, CXCL8, CXCL10, GM-CSF, IFN-α2, IL-10, IL-12p40, IL-12p70, IL-1β, IL-6, and TNFα. Results were obtained with a Flexmap 3D system with Luminex xPONENT software version 4.2 (both from Luminex Corp.; Austin, TX, USA). Cytokine concentrations were determined using Milliplex Analyst software (version 3.5.5.0, Millipore).

**EPIC Cytokine and flow cytometry statistical analysis.** Raw values were normalized with a 1 + Log2-transformation. To index each subject to their pre-vaccination cellular levels, a WithinVariation matrix was computed using the WithinVariation function in R package mixOmics version 6.1.2. The Kruskal-Wallis H-test (kruskal.test in base R), followed by the Dunn post-test was used to determine differentially regulated features within each data type, using the WithinVariation values for each feature. P-values were adjusted for each data type separately using the Benjamini-Hochberg method (p.adjust function, base R). Features were considered statistically different over the days of life if their adjusted p-values were below 0.05 for the Kruskal-Wallis test. All analyses were performed in R version 3.3.2 (2016-10-31).
5.3 Results

5.3.1 The small sample processing protocol was successfully implemented in a low-resource setting

I designed the blood processing protocol for this study to use minimal equipment in a low-resource setting. Because venous blood draws in newborns require technical skills not available in this region of GB, the initial protocol was also designed to accommodate a 0.4 mL blood volume that was feasible to obtain via heel-prick capillary blood draw (Figure 5.1A). This protocol was slightly adapted for implementation in low-resource populations where laboratory infrastructure and trained phlebotomists were available (e.g., The Gambia), but still utilized only 1 mL venous blood safely collected from newborns multiple times over the first week of life (Figure 5.1B).
Figure 5.1 Sample processing schemes for BCGIMED and Gambia studies
Sample processing workflows for BCGIMED (A) and EPIC (B). Steps done at the site of blood draw are boxed in green, while steps done either at the provisional laboratory (BCGIMED; storage closet) or standard BSL 2 laboratory (EPIC) in blue.
5.3.2 BCGIMED-immune results

The BCGIMED-Immune sub-study ran from April 9th, 2015 to August 28th, 2017. Study participants were followed for one year as a part of the main BCGIMED trial. Over the course of the study period, 1332 biospecimens were collected according to my protocol (Figure 5.2A). Of these, a subset of 350 samples have now been selected after the completion of the BCGIMED study for immune phenotyping, representing newborns that died, or were given a clinical diagnosis of pSBI within two months of blood draw, alongside a random sample of control infants that did not die nor receive a pSBI diagnosis during this time. While the biological analyses of these data are ongoing, the quality of the RNA extracted from whole blood demonstrated that processing, storage, and shipping conditions preserved optimal sample quality (Figure 5.2B) and provided an RNA yield suitable for downstream OMIC applications (Figure 5.2C). Storage time did not affect RNA quality or quantity. By conducting sample processing immediately at time of blood draw instead of transferring samples to the National Research Laboratory, and the flow cytometry specimens in our make-shift laboratory, it was possible for RNA to be preserved in a mean 1.23 minutes, plasma separated a mean 3.66 minutes, flow cytometry samples processed a mean 1.73 hours, and all samples frozen a mean 1.75 hours after blood draw across all 350 selected biospecimens. This level of standardization is difficult to achieve even where well-equipped and highly-resourced facilities are available.
Figure 5.2 Sample collection summary for the BCGIMED study
A. Flow chart summarizing the recruitment and sample collection metrics for the study. B. RNA integrity of the selected subset were not affected by study period, and were all above 7.5 (mean 9.26, range 7.8 to 9.9) C. RNA yield in micrograms ranged from 1.1 to 60.8 ug (mean 6.5 ug).
Table 5.1 Average processing times for selected BCGIMED samples from time of blood draw

<table>
<thead>
<tr>
<th></th>
<th>Time from blood draw: mean (SD)</th>
<th>n samples</th>
<th>n missing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA preservation (min)</td>
<td>1.28 (2.30)</td>
<td>350</td>
<td>0</td>
</tr>
<tr>
<td>Plasma separation (min)</td>
<td>3.66 (3.75)</td>
<td>327</td>
<td>0</td>
</tr>
<tr>
<td>FACS processing (hrs)</td>
<td>1.73 (0.57)</td>
<td>326</td>
<td>0</td>
</tr>
<tr>
<td>-80°C Freezing storage (hrs)</td>
<td>1.75 (0.62)</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>

5.3.3 Flow cytometric and plasma cytokine analysis of EPIC-Gambia newborns reveals a dynamic early-life trajectory

The Gambian study recruited 30 newborns sampled on the first day of life, and then were randomized to a second sample on day of life 1, 3, or 7 (Table 5.2).

Table 5.2 Sample sizes for the Gambia-EPIC study

<table>
<thead>
<tr>
<th></th>
<th>Flow Cytometry</th>
<th>Plasma cytokines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples post QC/QA</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>DOL 1 vs DOL 0</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>DOL 3 vs DOL 0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>DOL vs DOL 0</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

DOL = day of life

Using multi-OMIC integration revealed a stable immune trajectory in the first week of life marked by changes in interferon signaling, complement, and neutrophil pathways\textsuperscript{242}. While this study analyzed a wide range of data as summarized above, I present here in more detail only the
portion of the work that I contributed to (flow cytometry and cytokine multiplex analysis). As for the other OMIC data types, analysis of our pre-defined targeted cell populations and plasma cytokines in the EPIC-Gambia study revealed substantial between-subject variability in peripheral blood samples obtained over the first week of life, where day of life contributed less than 10% variance in both FCS and cytokine data (Figure 5.3A-B). However, consistent within-subject changes over the first week of life amongst the entire cohort of 30 Gambian newborns emerged when samples were indexed as displayed by Principal component analysis (PCA). After indexing, day of life contributed 17% and 21% to variance across FCS and Cytokine data respectively. Univariate analysis identified eight discriminating cellular features over the first week of life: basophils, plasmacytoid dendritic cells (pDC), natural killer cells, and neutrophils decreased; in contrast, myeloid DCs increased after Day of Life 0 (DOL0), while many other cell types remained stable (Figure 5.3E). We also found that plasma concentrations of C-X-C motif chemokine 10 (CXCL10), IL-17A, MDC, and IFN $\gamma$ increased, while IL-10, Chemokine C-C motif ligand (CCL) 5, granulocyte colony stimulating factor 2 (GCSF), and IL-6 decreased with age over the first week of life; many other soluble immune markers remained unchanged.
Figure 5.3 Indexing cellular and soluble immune markers revealed developmental progression over the first week of life

A,B. PCA analysis was used to plot cellular composition (A) and plasma cytokines/chemokine concentration (B) for each sample; this highlighted the substantial variability between participants and lack of defined clustering by DOL due to higher influence of individual variance over ontogeny. C,D. Accounting for repeat measures from the same individual across different sampling days compared to DOL0 (indexing to DOL0) revealed sample clustering by day of life between samples. E. Normalized cell counts showing developmental trajectories for cell populations that significantly changed. G. Normalized plasma cytokine/chemokine concentrations showing developmental trajectories for cell populations that significantly changed over the first week of life. **** p <= 0.0001, *** p <= 0.001, ** p <= 0.01, * p <= 0.05, ns p > 0.05, Kruskal-Wallis test, Benjamini-Hochberg adjusted p-values. FCS; Flow cytometry staining.
5.3.4 Indexing allows for identification of unified immune trajectories across independent newborn cohorts

To validate findings in the Gambian cohort, a confirmatory cohort was recruited in Papua New Guinea using the same study design and sample processing procedure.

Table 5.3 Sample sizes for the PNG-EPIC study

<table>
<thead>
<tr>
<th>PNG</th>
<th>Flow Cytometry</th>
<th>Plasma cytokines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of samples post QC/QA</td>
<td>59</td>
<td>40</td>
</tr>
<tr>
<td>DOL 1 vs DOL 0</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>DOL 3 vs DOL 0</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>DOL vs DOL 0</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

The highly standardized nature my protocol allows comparison of biological findings across diverse populations and independent studies. In our case, the PNG cohort served as a validation cohort of the Gambia cohort, as the same processing and sampling protocol was implemented across both sites. Prior to indexing, cohort of origin had a significant effect on cellular and plasma cytokine composition, contributing 72% and 26% of variance to the respective datasets (Figure 5.4A-B). After indexing to DOL0, cohort no longer had a measurable effect on either dataset, while day of life again contributed to 24% and 20% of the variance in FCS and Cytokine datasets respectively (Figure 5.4C-D). However, while newborns from PNG followed similar trajectories to Gambian newborns, differences could still be detected in the magnitude of their changes. This was only found for two cell types, pDCs and CD56-dim NK cells, where PNG newborns had higher cell counts on DOL0 compared to the later days than Gambians (Figure
Newborns from each cohort also differed for various cytokines; IL-17A, IFN-γ, FGF-2, Eotaxin, GCSF, and IL-4. In each case, PNG newborns had higher cytokine levels on DOL0 than Gambian infants (Figure 5.4F). GCSF was the most dramatically affected cytokine, whereby PNG newborn samples contained relatively higher concentrations on DOL0 but significantly lower by DOL7.
Figure 5.4  Indexing cellular and soluble markers allowed for agreement of biological trajectories from an independent validation cohort while still identifying cohort-specific differences

A-B. PCA analysis demonstrated that cohort contributed significantly to sample clustering while day of life trajectories contributed minimal variance. C-D. After indexing, samples from the
PNG validation cohort co-clustered with the Gambian cohort, with day of life contributing significantly to sample clustering while effect of cohort was no longer evident. E-F. Within immune cells (E) or plasma cytokines (F) significantly changing over the day of life with consistent trajectories between cohorts, differences were still identified between Gambian and PNG infants. **** p <= 0.0001, *** p <= 0.001, ** p <= 0.01, * p <= 0.05, ns p > 0.05, Wilcoxon Rank-Sum test, Bonferroni adjusted p-values. FCS; Flow cytometry staining

5.4 Discussion

The first week of life is characterized by heightened susceptibility to infections and is increasingly recognized as a major determinant of overall health for the entire human lifespan. Knowledge of the molecular drivers involved in these processes in newborns (defined as those $\leq 28$ days of life) is fragmentary. Systems biology approaches, employing high-dimensional molecular and cellular measurements, along with unbiased analytic approaches, have increased our understanding of basal and altered molecular states in adults and recently in newborns and infants after the first week of life, but such approaches have not been applied systematically to characterize molecular ontogeny over the most critical period, i.e. the first week of life. This is likely due to the analytical challenges posed by the limited amount of biosample that can be obtained and the many rapid physiological changes around birth.

To overcome these hurdles, I developed a robust experimental and analytical approach feasible with $\leq 1$ mL of newborn blood that was tested under field conditions in multiple low resource settings. The results from analysis of several hundred of biological specimen from 3 different cohort studies implementing my protocol proved that sample quality was sufficient to provide robust data for downstream systems biology analysis with starting blood volumes of even $< 500$ microliters. Beyond the technical feasibility, my protocol already provided deep biological insight, as the data produced using my protocol represent the most comprehensive systems biology study yet performed during the first week of human life. Specifically, despite
substantial between-subject variation, normalizing (indexing) all samples from a given newborn enabled identification of consistent and robust changes over the first week of life across the entire cohort. Furthermore, data integration using independent strategies not only validated signatures across methodologically- and biologically-distinct datasets, but also provided novel findings. The major observations derived from a cohort from West Africa (The Gambia) were validated for an Australasian (Papua New Guinea) cohort. Specifically, the results from these pioneering cohort studies highlight that, contrary to the relatively steady-state biology observed in healthy adults\textsuperscript{249, 250}, the first week of human life is highly dynamic. Nevertheless, despite the substantial variability between participants and these dramatic changes, ontogeny followed a robust trajectory common to newborns from very different areas of the world.

Still, this protocol represents only a starting point. Newer technologies already built on the Bissau protocol for the benefit of the EPIC study. For instance, the Bissau protocol utilized FACSLyse for preservation of cells. The EPIC Study used the Smart Tube system, where sample quality of preserved cells is maintained at four degrees for several hours (unpublished; in-house data). This system was not yet available at the commencement of the Bissau study. Direct comparison for FACSLyse and Smart Tube systems on whole blood in our laboratory demonstrated that cellular compositions slightly differed between the two; so while Smart Tube would have been a better choice, it was not possible to change the protocol mid-way through the study.

Implementing protocols requires field experience, strong partnerships with local groups, and an understanding of the populations invited to participate in the study. Blood sampling is not routinely done for newborns in most low resource settings, and cultural attitudes towards the procedure are skeptic. For example, we decided against multiple blood draws from the same
newborn in Bissau for the concern expressed to us by the local ethics committee about acceptability by the parents; meanwhile in the Gambia where MRC presence has made this practice more familiar, multiple samples were taken from each study participant over the first week of life.

Perhaps most importantly of all, conducting research in low-resource settings while coming from a high-resource one comes with many ethical considerations that, at this point of my training, I am only starting to understand. This issue has been addressed by other biomedical scientists with extensive field experience\textsuperscript{251}. This study in Bissau was only made possible by the long-standing relationship BHP has built with Guinea Bissau since 1978. BHP also employs around 170 locals to be a part of the research on child health that BHP is conducting, and has trained a cadre of Guineans to the PhD level. As such, my study employed two Guinean nurses to conduct blood draws and relied on several Guinean BHP employees to regularly collect and enter data. An auxiliary study to this main protocol, where peripheral white blood cell stains were prepared from 267 newborn samples was also used as a research project for the outstanding staff from Guinea Bissau responsible for sample processing over the last year of the study. This was presented at the remote Beyond Science conference in 2017, and won a top-presenter award\textsuperscript{252}. As a result of her exceptional skill and supported by my long-term long-distance tutelage, she now is pursuing her PhD in Portugal. This is only a first step towards training highly motivated and capable individuals to come to the forefront of biomedical research designed to improve health outcomes in their homeland, but it is a step in the right direction that I am very proud to have contributed to.
Chapter 6: Conclusion

The central goal of this work was to shed more light on the conversations between our bodies and their microbial component, and especially early in life where these conversations have profound effects on human health. By using a highly-standardized approach to analyze systemic immune responses from for geographically distinct cohorts alongside their stool microbiomes, comparing fungal to bacterial communities during a time of rapid development, and taking to the field to develop a platform to study immune responses in the most vulnerable newborn period in low resource settings, I made several contributions to the field that allow us to not only identify the areas most pertinent for further study, but also the approach. Specifically, the answers provided to several key questions have led to four main conclusions.

6.1 Biogeography shapes the host immune-microbiome relationship

Are there interactions between the host microbiome and systemic immunity in early life? If so, who influences whom? The work presented in Chapter 2 addressed these questions by testing the hypothesis that geographically distinct gut microbiomes correlated to differences in immune responses across four child cohorts. We were able to support that hypothesis by establishing a predominant association between systemic cytokine responses to TLR2 stimulation and colonization with the bacterial genera, Bacteroides and Prevotella. We also demonstrate that, even in absence of statistical correlations, causal relationships between the host microbiome and systemic immunity can still exist and can be dissected using germ-free mouse models of human fecal transplantation. Further investigation of the biogeography of microbiome-immune
interactions will continue to expand our understanding of health and disease in the increasingly mobile global human population.

6.2 Altered immune phenotypes in HEU infants is population-specific and not related to gut dysbiosis

After establishing that the gut microbiome and systemic immunity correlated with each other across diverse populations, and that region-specific microbiomes could shape region-specific immune responses, a new question arises: do the same perturbances to microbiome and immune development hold the same consequence within diverse populations, or are also population-specific? This question was addressed in Chapter 3 where we tested the hypothesis that immune alterations in HEU compared to HUU control infants were correlated with differences in their gut microbiomes, and that these differences would be consistent across distinct populations. We found that while differences between HEU and HUU control children in immune status as well as microbiome composition were readily detectable at 2 years of age, these differences did not correlate with each other. Further, found that perturbations related to maternal HIV infection to both immune phenotype and gut microbiomes were region-specific. Aberrant systemic immune phenotypes can be caused by various other host and environmental factors, and the microbiome is not always the culprit. This is also important as it suggests that there was no universal signature of HIV exposure impacting the immune or microbiome compartments, and factors common to all sites in our study, specifically biological exposure to HIV and ARV may not have been responsible for immune or microbial differences at two years of life. With that, extrapolating from one human population to another is not an ideal approach to understand host perturbations in health and disease.
6.3 Bacterial and fungal community dynamics are kingdom-specific

While the focus of this thesis (and of the human microbiome field at large) has thus far been the bacterial microbiome, fungi are known to colonize the gastrointestinal tract and modulate immune responses as well. None of the questions above were answered with respect to fungi. While evaluating the correlations between fungi and systemic immunity is only a future direction at this point, another question arises: how does the composition of the fungal microbiome directly compare to the bacterial microbiome, already known to follow a dynamic developmental trajectory throughout early life? This was addressed in Chapter 4, where I tested the hypothesis that fungal microbiomes in the gut followed age-dependent developmental trajectories distinct from bacterial microbiomes. I found that that the two communities behave very differently across early life and thus are likely driven by separate selective forces. Further, while mothers are a major source of bacteria for newborns, this was less apparent for fungi in these data. And surprisingly, mother’s bacterial but not fungal stool microbiomes differed drastically between one week and one month post-partum in a way that may relate to optimizing the inoculation of their newborns. Taken together, our study clearly demonstrates that commensal bacteria and fungi colonize our guts in highly kingdom-specific ways, and that the dynamics of the mother-newborn interaction are specific to a degree that suggests purpose. This in turn indicates that targeted interventions, e.g. probiotics administered to the mother, may have significant potential to influence the developmental trajectory of the newborn microbiome.

6.4 Studying newborn immunity is feasible where it is most neglected

Newborns are an understudied population, mostly due to technical hurdles, and especially in low resource setting where their health is most threatened. In Chapter 5, I challenged the notion that
systems biology could not be applied in a highly-standardized way to newborns in low-resource settings and instead set out to answer the last questions of my thesis: What tools work and what tools do we need to chart out the course to successfully study the most vulnerable? I hypothesized that stringently implemented protocols custom designed to fit low resource settings could successfully yield biologically informative information. And my hypothesis was supported. My protocol proved feasible and will in the next phase of these studies be applied to analyze the interaction of microbiome and immune ontogeny. Together, this will help fill the knowledge gap of this important interaction in early life and hopefully facilitate discovery of effective interventions to finally reduce newborn infectious mortality in low resource settings.

Figure 6.1 An overview of systemic immune-microbiome conversations in early life
Systemic immunity follows a developmental trajectory across the early-life age spectrum. Composition of the bacterial, but less-so not fungal microbiomes also follows a trajectory in early life. Throughout this process, there is cross-talk between immune and microbial compartments (depicted in light blue between immune and microbial development). Early infancy represents a heightened period of vulnerability to perturbed immune and microbial
homeostasis, deviations from which result in lifelong disease states. Two host factors tested for their influence on the host-microbiome crosstalk are depicted in orange circles; maternal HIV infection is associated with immune development and bacterial microbiome composition separately, and geography interacts with crosstalk between the two.

6.5 The way forward

The studies presented here were highly informative, despite only pilot sample sizes. Insights gained and lessons learned pave the way for new approaches to study host-microbiome conversations. By turning limitations into lessons, future work will benefit even more. With the falling costs of DNA sequencing, highly-powered microbiome surveys are now within grasp and will allow for the incorporation of vast host metadata to study host-microbiome relationships. Statistical approaches applied to immune and microbiome research separately have already provided much insight into how host demographic factors shape both compartments. We must now move to highly standardized studies across diverse cohorts to gain insight into population-specific host-microbiome interactions, and study perturbations linked to human disease in this way. Also, given the known functional redundancy in gut microbiomes, understanding the links between functional microbial genes and their expression to immune phenotypes will provide much more insight than taxonomic composition alone. To that end, my thesis work generated the opportunity for me to now look for links between gut metagenomes and systemic immune cytokine responses in a Canadian cohort of infants, in the context of allergic disease. As a co-applicant on the grant enabling this work, I am excited to apply the skills I learned here to this very important health issue. Specifically, I will analyze cytokine responses to PRR agonists and stool metagenomic data from 120 Canadian children recruited to the Canadian Healthy Longitudinal Infant (CHILD) study at birth, 1 year, and 5 years of life. These children were either diagnosed with asthma at 5 years of age, had a positive skin prick test
to a range of environmental or food allergens from 3 to 5 years of age, or were not diagnosed with allergy or asthma at any time. Using the analytical tools applied in Chapters 2 and 3, I will test the hypothesis that children who develop asthma or allergy later in life have aberrant innate immune responses to PRRs early in life, and that these differences correlate to differences in their gut microbiomes. This will directly test the hypotheses I generated in my thesis.

One long-term goal of these surveys is to optimize host microbiomes to prevent diseases associated with deviations from a beneficial commensal community. With the recent exciting findings that probiotic *L. plantarum* can reduce newborn pSBI\textsuperscript{211}, the next steps will be to repeat this in another randomized clinical trial. And in fact the work presented in Chapter 4 was used to secure funding to conduct a feasibility pilot of *L. plantarum* administration to newborns in Ghana for this purpose. I have the privilege to be involved in this endeavor, and have used the knowledge gained from this work to shape the upcoming trial, including longitudinal sampling of mothers before birth across body sites and of postnatally for mothers and children. Specifically, mothers that consent to themselves and their newborns taking part in the study will be randomized to either administer a synbiotic preparation of *L. plantarum* and fructo-oligosaccharide (FOS) or a placebo (maltodextrin alone) to their babies daily for the first week of life. We will collect stool and blood specimens from babies throughout the first 6 months of their lives. By working closely with local scientists, we will capture the necessary lifestyle factors needed to better understand microbial community dynamics in their population. The primary outcome for this pilot study are safety, or intolerance of either synbiotic or placebo administration presenting as diarrhea, vomiting, or abdominal distension. The secondary outcome is gut colonization by the probiotic strain, and the third outcome safety and feasibility of longitudinal blood sampling and implementation of associated OMIC protocols at this study.
site. While we will only generate pilot sample sizes, we will still apply the blood processing protocol used in Chapter 5 to test whether administration of *L. plantarum* and FOS is associated to changes in systemic immune signatures.

The immediate future for the Guinea-Bissau study presented in Chapter 5 is to undertake the analysis of immune phenotyping data from the selected subset of infants – the data is being generated as this thesis is being written. Specifically, the samples selected belonged to infants that either died in the first week of life following blood draw, were given a diagnosis of sepsis in the first month of life, or controls that neither died in the first two months of life nor received a diagnosis of sepsis during that time. Each group contains newborns that were either vaccinated with BG and OPV prior to blood draw or not. The goals of these analyses are to determine whether flow cytometric, proteomic, metabolomic, or transcriptomic datasets either separately or integrated with each other could predict either sepsis or mortality in the first week of life. Further, given the phenomenon of BCG-vaccination of newborns in Guinea-Bissau to be associated with decreased mortality in the first month of life, we will test the hypothesis that biological signatures preceding a diagnosis of sepsis are different among BCG and OPV vaccinated compared to unvaccinated infants. Also, this close and fruitful collaboration has paved the way for other studies to lend biological insight to the exemplary epidemiological work being conducted at the Bandim Health Project for years to come, and with the recognized importance of the microbiome, this will also become a central focus in future study designs. Specifically, this study design is being incorporated into another randomized clinical trial testing whether pre-conceptual maternal vaccination with the measles vaccine increases the non-specific benefits of measles vaccination in the mother’s infants.


Duerr CU, Hornef MW. The mammalian intestinal epithelium as integral player in the establishment and maintenance of host-microbial homeostasis. Semin Immunol 2012; 24:25-35.


251. Baumann LC.


Appendix A : BCGIMED-Immune sub-study materials

A.1 Standard operating procedure
A. **Study Materials**

1. **Reagents:**
   A. RNALater solution (Ambion Cat. No. AM7021)
   B. 10x BD FACSLysing solution (BD Cat. No. 394202)
   C. UltraPure distilled water (Invitrogen Cat. No. 10977-015)

2. **Sample tubes and labeling:**
   A. BD microtainer collection tubes (BD Cat. No. 365965)
   B. Corning 2.0 ml cryovials (Corning Cat. No. 430488)
   C. Diamed 0.5 ml microtubes (Diamed Cat. No. SPE055-N)
   D. VWR 2.0 ml microfuge tubes with cap (VWR Cat. No. 16466-060)
   E. BD 50 ml Falcon tubes (BD Cat. No. 352070)
   F. Sample tube labels (supplied by Vancouver, Kollmann Lab)
   G. Plastic bags (various, for sample transport)

3. **Blood collection materials**
   A. BD quickheel lancets (BD Cat. No. 368101)
   B. Gauze
   C. Alcohol prep pads
   D. Gauze tape

4. **Equipment & PPE**
   A. Accu-chek Aviva glucose meter and strips
   B. Accu-chek Aviva glucose strips (Roche 0645988)
   B. p200*, p1000 micropipettes
   C. p200 sterile pipette filter tips (VWR Cat. No. 82003-196)
   D. p1000 pipette tips
   C. Microcentrifuges: Eppendorf 5415C, BD Microfuge 16
   D. 110/120V to 220/240V voltage transformers, 400W or higher
   E. Biohazard bags, sharps bins

*The p200 pipettes supplied for this study are Rainin LTS pipettes. Tips used with this pipette MUST be LTS tips, otherwise they will not fit.
B. Preparing Reagents

1. RNALater

Preparation:
• RNALater solution is used as-is.
• For this study, we purchased 500 ml bottles of RNALater.
• Prepare 50 ml working aliquots of RNALater from the main stock for ease of use and avoidance of contamination by pouring solution from the 500 ml bottle into a 50 ml falcon tube.

Storage:
• Store both stock solutions and working aliquots of RNALater at room temperature.
• Store in a cool, dark place where the temperature is not likely to reach over 30°C.
• Do not refrigerate.

2. BDFACSLyse

Preparation Materials:
• FACSLyse 10x buffer
• Ultra-pure distilled water (dH$_2$O)
• p1000 pipette with tips
• 50 ml Falcon tube

Procedure:
1. Label a 50 ml Falcon tube “1 x FACSLyse”, with your initials and date
2. Pipet 5 x 1 ml (total of 5 ml) of 10x FACSLyse into 50 ml falcon tube
3. Pour ultra-pure distilled water into the 50 ml tube to the 50 ml line

NOTE: to prepare smaller volumes, dilute the 10x buffer 1/10 with distilled water:
vol. 10x FACSLyse = desired 1x FACSLyse volume x 0.1
vol. dH$_2$O = desired final volume – vol. 10x FACSLyse

e.g. for 5 ml:
Vol. 10x FACSLyse = 5 ml x 0.1 = 0.5 ml
Vol. dH$_2$O = 5 ml - 0.5 ml = 4.5 ml
…so add 0.5 ml 10x FACSLyse to 4.5 ml dH$_2$O.

Storage:
• Store both 10x and 1x FACSLyse at 4°C
• 10x FACSLyse CAN be stored between 4°C and 25°C, but preferably keep in the fridge! It can get much warmer than 25°C…
• Store dH$_2$O at room temperature
C. Preparing specimen bags and sample tubes

It is most efficient to prepare about 20 sample tube sets at a time. This will last for 1-2 weeks.

Materials (per specimen bag)

Tubes:
A. 1 x BD microtainer
B. 1 x Corning 2.0 ml cryovial, orange lid (For RNALater)
C. 1 x 0.5 ml microfuge tube
D. 3 x 2.0 ml VWR screw-top tubes, clear lids

Reagents:
E. RNALater (1.3 ml per sample)
F. 1x FACSLyse (1.9 ml per sample)

Other supplies:
G. Sample sticker labels (6):
   ____ WB, ____ RNA, ____ PLAS1, ____ PLAS2, ____ FACS, Subject ID____

H. Plastic specimen bag
I. p1000 pipette and tips
J. VWR Ethanol-proof marker

Procedure (per tube set)
1. Add 2 x 0.65 ml (total 1.3 ml) RNALater solution to tube B (orange cryovial)
2. Add 2 x 0.85 ml (total 1.7 ml) 1x FACSLyse to one of tube D (clear screw-top tube)
3. Place all sample tubes EXCEPT FACSLyse tube into one plastic bag with the sample sticker label.
   **These tube sets can be kept at room temperature, but not above ~25 degrees**
   **Best to keep them at Casa de Peter and take what you need for the day to HNSM**
4. Store all the 1x FACSLyse tubes at 4°C, Casa de Peter lab.

Preparing tube sets for sample collection:
1. Write the BE number for the subject on each of the 6 stickers
2. Place the labels on the tubes:
   ____ WB: BD Microtainer (Put the sticker over the “microtainer” label
   ____ RNA: Orange-lid cryovial
   ____ PLAS1: empty clear 2.0 ml tube
   ____ PLAS2: empty clear 2.0 ml tube
   ____ FACS: Clear 2.0 ml tube filled with FACSLyse (label this tube once you get to lab Casa de Peter. Leave the label in the specimen
   ** the 0.5 ml tube does not get a sticker. Simply write the BE number on the side of the tube
D. Sample processing protocols

1. Heel prick blood draw

Materials:

A. Quickheel lancet
B. Alcohol prep pad
C. Gauze
D. Gauze tape
E. Paper tray
F. Pre-labelled Microtainer
G. Sharps bin
H. Accu-Chek glucose meter and strips

Before you begin
Prepare all blood collection and processing materials on collection tray like pictured

Protocol:

A. Collecting the blood sample
1. Clean the bottom of the foot with an alcohol wipe. Dry skin with gauze
2. Press lancet firmly against side of heel (face logo side up).
3. Press the trigger to make the cut
4. Dispose of the used lancet into a sharps container
5. Squeeze the foot to make the first drop of blood
6. WIPE THE FIRST DROP AWAY with gauze before proceeding with sample collection
7. Squeeze the foot to form the next drop of blood
8. Scoop the blood drops once they run just under the cut
   **DO NOT scrape the tube over the cut! It will hurt 😞**
   **Pause for 1-2 seconds before squeezing the next drop**
9. Repeat drop-by-drop until the microtainer is filled to the 2nd line (400 mcl)
   **BELOW the line is NOT OK!**
   **ABOVE the line is NOT OK!**
10. Once the tube is filled, dispose the flo-top collector, close the lid,
    **IMMEDIATELY invert tube 10x to mix**
11. RECORD TIME OF COLLECTION ON THE SPF
2. Measuring blood glucose

Materials:
Accu-chek glucose meter
Glucose strips

Protocol:

While the blood sample is being collected, the 2nd person should set up the glucose meter. If there is only one person working, set up the glucose meter after the microtainer has been collected.

1. Turn on the glucose meter. The screen displays a flashing glucose strip
2. Place a strip into the meter, chip-side up
   **The screen will now show a flashing hourglass.
   **When the meter is ready, it will show the glucose strip (not flashing) and a blood drop (flashing)
3. Touch the strip to the drop of blood until the display shows an hourglass
4. Read blood glucose

Bandaging the cut:

- After the sample is collected and the glucose measurement has been taken, hold a piece of gauze over the cut until there is no more bleeding.
- Fold a piece of gauze over the cut and secure it with gauze tape.
NOTE: Processing according to blood volumes

If the blood draw is difficult and you do not get a full sample, process according the the flowchart below.

**Blood sample processing: Priorities**

- **Microtainer volume <200 mcl**
  - Entire sample in RNALater

- **Microtainer volume At least 200, below 400**
  - RNA: 200 mcl whole blood
  - Pipette all remaining blood into 0.5 ml tube
  - <100 mcl
    - Plasma: None
  - >100 mcl
    - Plasma: 1 x 30 mcl
    - FACS: 100 mcl, or rest of sample

- **Microtainer volume =400 mcl**
  - RNA: 200 mcl whole blood
  - Pipette all remaining blood into 0.5 ml tube
  - Plasma: 2 x 30 mcl

- **Microtainer volume >400 mcl**
  - You may notice some blood clotting. If yes, then do not process. If none, process according to 400 mcl.
3. RNA Sample (Orange tube, ___ RNA)

Process this sample WITHIN 15 MINUTES of blood draw

Materials:
A. Microtainer blood sample
B. Pre-labelled ___ RNA tube
C. p200 pipette with tips

Protocol:

1. Set the pipette to 200 mcl and open the RNA tube
2. Pipette blood up and down 10x to mix
3. Draw up 200 mcl blood and place in RNALater. Pipette 10x to mix (If less than 200 mcl, do not process).

4. RECORD TIME ON SAMPLE PROCESSING FORM

5. Pipet rest of blood from microtainer into 0.5 ml tube
6. Record sample volume:
   **If volume of blood transferred = 200 mcl, check “0.4 ml” sample volume. If it is less than 200 mcl, check “between 0.2 and 0.4 ml”.
   **If it more than 200 mcl, check “>0.4 ml”**
4. Plasma Samples (Clear tubes, ____ PLAS1, ____ PLAS2)

Blood must be centrifuged immediately after RNA sample is processed.

Materials:

A. Blood sample (now in 0.5 ml tube)
B. Pre-labelled ___ PLAS1, ____ PLAS2 tubes
C. p200 pipette with tips
D. Microcentrifuge

Procedure:

1. Place the remaining whole blood in the centrifuge:
2. Spin the sample at 500 x g for 10 minutes
   RECORD TIME OF SPIN ON SAMPLE PROCESSING FORM
3. Set the pipette to 25 mcl
4. Remove 1x 25 mcl plasma for low volume or 2x 35 mcl (=7 total) for full-volume samples.
   Place all plasma in tube “PLAS1”.
5. For full volume samples, once all plasma has been removed, pipette the plasma up and down 10x
6. With a NEW TIP, draw up 35 mcl of mixed plasma and place in tube “PLAS2”. Each tube now has 30 mcl (~5 mcl is lost during pipetting).
   RECORD TIME OF ALIQUOTTING ON SAMPLE PROCESSING FORM
5. FACS Samples (Clear tube with FACSLyse liquid)

Process this sample Casa de Peter lab WITHIN 3 HOURS of blood draw. Blood sample must be placed in freezer immediately after completing this protocol.

Materials:

A. Blood sample (in 0.5 ml tube)  
B. ____ FACS tube (pre-filled with FACSLyse)  
C. FACSLyse buffer (200 mcl per sample)  
D. Bleach  
E. Glass waste container  
F. Paper towel

Protocol:

1. Short-spin the blood tubes for ~5 sec  
2. Remove a FACS tube from the fridge and label with ____ FACS sticker  
3. Set the micropipette to 100 mcl. Pipet blood up and down 10 x to re-mix the sample  
4. Pipet 100 mcl mixed blood (or remainder of sample if less than 100 mcl) into ____ FACS tube  
5. **RECORD TIME IN FACS ON SAMPLE PROCESSING FORM**  
6. Pipet up and down 10x to thoroughly mix blood with FACLyse  
7. Leave FACS tube at room temperature for 10 minutes  
8. Invert tube 10x 5 and 9 minutes after incubation  
9. Centrifuge samples at 2500 RPM for 5 minutes  
10. Discard supernatant by pouring into waste bottle (glass boddle with 10-20 ml bleach)  
11. Tap tube onto paper towel to get rid of most of the liquid  
12. Resuspend cells in 200 mcl fresh FACSLyse buffer:  
   *Wash over cell pellet 5x  
   *Wash around the rim of the tube 5x  
   *Wash the bttom of the tube 5x  
13. **RECORD TIME OF RESUSPENSION ON SAMPLE PROCESSING FORM**  
14. IMMEDIATELY store all sample tubes at -80°C:  
   ____ RNA, ____ PLAS1, ____ PLAS2, ____ FACS  
15. **RECORD TIME OF FREEZING ON SAMPLE PROCESSING FORM**
BCGIMED sample processing during power outages at HNSM
Nelly Amenyogbe, 27 January 2016-01-27

General notes:
- If the power is off, delay the day’s sample collection. Allow ~2 hrs, or until 10 am. If power is still off, then collect the day’s samples under SOP A.

- If power outages become more frequent, do not delay in removing a sample from the centrifuge once the spin is done. This will prevent samples being stuck inside should the power go off.

A. Power outage before sample is centrifuged

This is for any samples that have been collected, but cannot be placed in centrifuge.

SOP:
1. Aliquot 200 mcl into RNALater as soon as blood is drawn, as per routine SOP
2. Place remainder of blood into 0.5 ml tube for centrifugation
3. Leave aliquot of blood upright, in test-tube holder. The plasma will separate from the blood on it’s own. Take care not to disturb the tube such that the plasma does not mix with the blood fraction.
4. Spin vial when power becomes available and note the time on the sample processing form.

4.b. If power does not become available at HNSM before blood is to be transported to the compound, (within ~2.5 hrs of blood draw),
- Ensure sample has been sitting for at least 30 minutes
- Remove two aliquots of plasma as per regular SOP
- Proceed with the remainder of the sample to the compound FACS processing

B. Power outage when sample is in centrifuge

This applies to any samples stuck in Richard, the HNSM centrifuge, when the power goes out during a spin.

SOP:
1. Wait until the power turns back on. Spin the sample again for another 10 minutes and record time of 2nd spin on the sample processing form
2. Proceed to compound for processing of FACS sample

If power remains off for ~2 hrs: If able, transport Richard to the compound. Do not bring the transformer – the one being used with Howard (centrifuge at compound) will work.
- Plug in centrifuge
- Visually observe the sample and note whether the plasma is still separated from the blood. Record observation on sample processing form. **DO NOT delay this observation – it is necessary to record the degree of sample disturbance during transport.
- Spin sample for additional 10 min
- Proceed with the rest of SOP as usual

**If centrifuge cannot be transported to the compound:**
Especially if other samples have been collected: Bring these to Compound for remainder of processing. The sample stuck in the centrifuge will have to be discarded the following work day.

**C. Power outage in compound**

If power is out at the compound, do not place any blood into FACSLyse. Leave it at room temperature and proceed with sample processing when power becomes available.

Remainder of samples (RNA, plasma) should still be placed in the -80 freezer.
A.2 Sample processing form: English Version

**BCG Immunogical Sub-study:**
Sample collection and processing forms

Regno:_______

Was the infant consented into the BCG-Immune sub-study:  Sim __ Não __

Inclusion date (dd/mm/aaaa):_____/_____/_____ Sample collection date:_____ /_____/_____

**General information:**

**Infant indentification:** Please ensure this information corresponds to the inclusion form from the BCGIMED study. If there is discrepancy please do not proceed with sample collectoin.

Mother’s name:_________________ age:______

Father’s name:_________________

Is the infant a twin: Yes__ No__ If yes: Twin number:____

Infant sex: M___ F___ Date of birth (dd/mm/yyyy):_____/_____/_______ Birthweight:____ g

Location of the infant when the sample was taken? Crib___ With the mother___ Incubador___

Is the infant being breastfed? Yes___ No___ Unknown___

If yes, time of last breastfeeding:____:____, Unknown___

Has the infant received antibiotics? Yes__ No__ Unknown__ If yes:

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Antibiotics administered</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amoxicillin</td>
<td>Metronidazole</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Antibiotics administered</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amoxicillin</td>
<td>Metronidazole</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Antibiotics administered</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amoxicillin</td>
<td>Metronidazole</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Has the infant received other medications? Yes__ No__ Unknown__ If yes:

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Medication</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Multivitamin</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Medication</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Multivitamin</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date (dd/mm/yyyy)</th>
<th>Medication</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Multivitamin</td>
</tr>
<tr>
<td>Other:___________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# BCG Immunology Sub-study:
Sample collection and processing form

Regno: __________

**Lot Numbers:**

<table>
<thead>
<tr>
<th>Microtainer:</th>
<th>Glucose Strips:</th>
</tr>
</thead>
<tbody>
<tr>
<td>__ Lot 1 (410038N, Exp. 2015-09)</td>
<td>__ Lot 1 (493903, Exp. 2015-31)</td>
</tr>
<tr>
<td>__ Lot 2 (415761N, Exp. 2015-11)</td>
<td>__ Lot 2 (494014, Exp. 2016-04-30)</td>
</tr>
<tr>
<td>__ Lot 3 (424118N, Exp. 2016-02)</td>
<td>__ Lot 3 (494126, Exp. 2016-06-30)</td>
</tr>
<tr>
<td>__ Lot 4 (425211N, Exp. 2016-02)</td>
<td><strong>Sample collection:</strong></td>
</tr>
</tbody>
</table>

Sample collected by: _____________________  Sample processed by: _____________________

Sample was collected?: Yes__  No__  Time of collection (hh:mm): ___:___

If no, reason:  Mother was discharged ___ Infant transferred to Pedriatria ___ Infant died ___

Glucose: ________mmol/L

Blood volume: <200 mcl___  >400 mcl___

Coagulated: Yes__  No__  Hemolysis: Yes__  No__

**Processing times:**

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Time of completion (hh:mm)</th>
<th>Sample was processed:</th>
<th>Blood volume processed (in mcl)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA preservation</td>
<td><strong>:</strong></td>
<td>Y_ N__</td>
<td>Vol:___</td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>Spin: <strong>:</strong></td>
<td>Y_ N__</td>
<td>Vol. PLAS1: ___ mcl</td>
<td>Vol. PLAS2: ___ mcl</td>
</tr>
<tr>
<td></td>
<td>Aliquot: <strong>:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cellular preservation</td>
<td>FACS: <strong>:</strong></td>
<td>Y_ N__</td>
<td>Blood volume: &lt;100mcl_</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wash: <strong>:</strong></td>
<td></td>
<td>=100 mcl_</td>
<td></td>
</tr>
</tbody>
</table>

Time of sample freezing (hh:mm): ___:___

**Additional notes (use back of form if necessary)**
A.3 Sample processing form: Portuguese version

BCG Subestudo de Imunologia Imediata:
Ficha de coleta e processamento de amostras

Regno: _______

Número (adesivo): _______
Número (escrito): _______

A criança faz parte do estudo de BCGIMED: Sim __ Não __
Data de inclusão (dd/mm/aaaa): __/__/____
Data de coleta de amostra: __/__/____

Informação geral:

Identificação da criança: Tudo deve corresponder com a ficha de inclusão do BCG-Imediato.
Se não corresponde, por favor não tire a amostra do sangue.
Nome da mãe: ____________________________ Idade:______
Nome da pape: ____________________________
A criança é gêmeo: Sim __ Não __ Se Sim: No. Gemlar:______
Sexo da criança: M __ F ___ Data de nascimento (dd/mm/aaaa): __/__/____ Peso ao nascer: ___ g

Localização da criança no momento da amostragem? Berço ___ Com a mãe ___ Incubador ___
A criança foi amamentado já? Sim ___ Não ___
A hora da última vez que a criança foi amamentado: __:__:__, NS___

A criança chegou a receber um antibiótico? Sim __ Não__ NS__ Se sim:

<table>
<thead>
<tr>
<th>Data (dd/mm/aaaa)</th>
<th>Antibiótico administrado</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amoxicilina</td>
<td>Metronidazole</td>
</tr>
<tr>
<td></td>
<td>Amoxicilina</td>
<td>Metronidazole</td>
</tr>
<tr>
<td></td>
<td>Amoxicilina</td>
<td>Metronidazole</td>
</tr>
<tr>
<td></td>
<td>Amoxicilina</td>
<td>Metronidazole</td>
</tr>
</tbody>
</table>

A criança chegou a receber qualquer outro medicação? Sim __ Não__ NS__ Se Sim:

<table>
<thead>
<tr>
<th>Data (dd/mm/aaaa)</th>
<th>Medicação</th>
<th>Dose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Polivitamina</td>
</tr>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Polivitamina</td>
</tr>
<tr>
<td></td>
<td>Paracetamol</td>
<td>Polivitamina</td>
</tr>
</tbody>
</table>
BCG Subestudo de Imunologia Imediata:
Ficha de coleta e processamento de amostras

Número (adesivo):  
Número (escrito):  

Regno: _______

Lot Numbers:

**Microtainer:**  
___ Lot 1 (410038N, Exp. 2015-09)  
___ Lot 2 (415761N, Exp. 2015-11)  
___ Lot 3 (424118N, Exp. 2016-02)  
___ Lot 4 (425211N, Exp. 2016-02)  

**Tiras de glicose:**  
___ Lot 1 (493903, Exp, 2015-03-31)  
___ Lot 2 (494014, Exp. 2016-04-30)  
___ Lot 3 (494126, Exp. 2016-06-30)  

Coleta:

Sangue coletado por: _____________________  
Amostra processada por: _____________________

Coleta de sangue:  
Sim  
Não

Hora da coleta: ___:___

Se não, porque: Mãe foi para casa ___ infantil em Pediatria ___ infantil morreu ___

Glicose: ______ mmol/L

Quantidade da amostra: <200 mcl ___  >400 mcl ___

Coágulo: Sim  
Não

Hemólise: Sim  
Não

**Tempo de procedimento:**

<table>
<thead>
<tr>
<th>Procedimento</th>
<th>Hora de realização (hh:mm)</th>
<th>Amostra processado:</th>
<th>Volume de processão (em mikrolitros, mcl)</th>
<th>Notações</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservação de RNA</td>
<td><em><strong>:</strong></em></td>
<td>S__ N__</td>
<td>Vol:___</td>
<td></td>
</tr>
<tr>
<td>Plasma</td>
<td>Spin: <em><strong>:</strong></em></td>
<td>S__ N__</td>
<td>Vol. PLAS1: ___ mcl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aliquot: <em><strong>:</strong></em></td>
<td></td>
<td>Vol. PLAS2: ___ mcl</td>
<td></td>
</tr>
<tr>
<td>Preservação das células</td>
<td>FACS: <em><strong>:</strong></em></td>
<td>S__ N__</td>
<td>Volume de sangue: ___ mcl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lavado: <em><strong>:</strong></em></td>
<td></td>
<td>&lt;100 mcl ___</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>=100 mcl ___</td>
<td></td>
</tr>
</tbody>
</table>

Hora de congelamento da amostra (hh:mm): ___:___

**Para notas adicionais**
A.4 Study explanation: English version

**Informed Consent: Study information**
Immunological Substudy within the BCG-Immediate study

Within the BCG immediate study, in which mothers that have recently given birth choose an envelope, which decides whether their child is vaccinated when we talk to the mother or at the time when the mother and child is feeling good and is ready to go home, we from the BHP are also doing another type of work. We would like to know better how BCG vaccine works and the effects it has on the child’s ability to combat diseases. To do this, we wish to study the child’s blood.

For these reasons we have started this study where we ask mothers that have given birth to infants that are included in the BCG Immediate study if they would like to help us. We ask for permission to take a blood sample from the child’s heel. The quantity to be taken is equal to the contents of this small tube (0.4 mL), like that of a big peanut.

Some of the blood will be used to measure the glucose level in the blood of the child so we know if it is good. The result of the glucose test will be given to you right after we have taken the blood sample. If the infant has a low glucose level, we will provide treatment.

The remaining blood will be analyzed “fora de Bissau”, where they have the big machines needed for this. Some of the sample will be taken to the Gambia, and some to Canada. The result will only be available after the study has ended. Hence, it will not help your baby, but it can hopefully be used to help babies in the future.

**How are blood samples from newborn infants taken:**
- First, we will clean the feet of the child well with alcohol and after that we draw a small amount of blood from the child’s heel.
- To do this, we have to prick the child’s heel with a small piece of equipment.
- We will help to get the blood out by a gentle touch to the child’s foot. This can take a couple of minutes. The child might feel some pain and it might start crying.

For a newborn child this is an absolutely secure and normal procedure. It will be performed by a trained nurse who is skilled at taking such samples. When the nurse sees that no blood is coming out of the child’s foot anymore, he will put a bandage on the foot.

If you want, you can take this piece of paper and read it carefully with your husband and other family members so you can consider if you want to take part of the study or not. Participation in this study is voluntary and if you don’t wish to, we from the BHP don’t have a problem with that. You will not have a bad relationship with us or the BHP because of that. Also, you can take part in the BCG Immediate study without taking part in the blood sample study.

Do you have anything you wish to ask us? Now, you’ll have some time to think about it. Tomorrow we will come and talk to you again about the study.
A.5 Study explanation: Portuguese version

Konsentimentu Informadu
Subestudu immunologiku do estudu BCG-Immediatu

Dentro do estudo BCG Immediatu, nunde ku padidas na bin kudji um envelope ku ta dici di si se fidju na toma BCG na hora ku no ta papia ku mame ou na hora ku padida ku meninu sta diritu i na bai pa kasa, fora di es anos di PSB no ta fasi utru tarbadju. No ta sibi di kuma barcina di mon, BCG, i bon, i ta tadja manga di doensas. Gossi, no misti sibi mas mindjor kuma ku i ta tissi es efeitos, kuma ku i ta stimula sistema immunologiku na korpu pa i pudi tadja utru doensas. Pa no fasi es, no misti studa sangue di meninu. Tambi, no misti midu nivel di glukose na sangue di meninu pa no sibi si i sta diritu. Resultadu di testi di glukose i na dadu pa bo dipus ku no kaba fasi analise. Si meninu tene nivel di glukose bas, no na bin djudal na fasi si tratamentu.

Es ku mandu no ranka es studu nunde ku no ta punta padidas inkludus na studu di BCG-Immediatu si i pudi djudano. No ta pidi lisensa pa no tira um amostra di sangue na kalkanhada di meninu. Kuantidadi ku ta tirado i ta djusa ku ke ku na fika na es frasku pikininu (0.4 mL), suma um mankara garandi.

No misti fasi analisis di sangue avansadu ku ta usa makinas garandi tchiu. Suma no ka ta tene tudu es makinas li na Bissau, no na bin tene nesissidade pa manda parte di kada amostra pa Gambia pa fasi analisasi la. Utral analisis avansadu tchiu dimas i pudi fashidu so na Canada, nunde i tene mindjoris makinas ku djintis treinadus diritu pa djubi sinai di korpu na sangue.

Kuma ku sangue di meninu ku kumsa padidu ta tirado:
- Primeiri no ta limpa pe di meninu diritu ku alkoal i dipus i ta tirado um kuantidadi pekininu di sangue na kalkanhada di mininu.
- Pa es fashidu, no dibidi fidi kalkanhada di meninu ku un pekininu materiul di tira sangue.
- No na bin djuda meninu pa sangue pudi bim sa ku un toki lebi na si pe. Es i pudi toma alguns minutus. Meninu pudi na bin sinti dur un bokadu i meninu pudi bin tchora.

Pa meninu ku kumsa padidu es i um procedimentu absolutamente seguru i normal. I na sedu fashidu pa un emfermeiro ku sibi diritu kuma ku i ta fashidu. Hora ku emfermeiro na odja kuma pe di meninu ka ta sangra mas, i ta pui un adisivu na pe di meninu.

Si bu misti, bu pudi toma es papel pa bu lei diritu ku bu omi i utrus familias pa bu pensa si bu misti fasi parti di es studu ou nao. Partisipason na es studu i si algum misti i si bu ka misti, anos di PSB no ka tene problema. I ka na fasiu tene mau relason ku anos ou ku PSB. Tambi, bo pudi fasi parti di studu di BCG-immediatu sin fasi parti di estudu ku amostra di sangue.

Bu tene kualker kusa ku bu misti puntano? Dipus di es, bu tene um bokadinhu di tempo pa bu pensa. Papia amanha no na bin papia mas ku abo.