Norwegian University
of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of Chemistry, Biotechnology, and Food Science

Associations in infant gut microbiota's taxonomic and inferred metabolic composition with immune cells at 12 months

Cecilie Fredheim
MSc Biotechnology

Acknowledgements

This thesis was performed at the Faculty of Chemistry, Biotechnology, and Food Sciences, at the Norwegian University of Life Sciences, under the supervision of Professor Knut Rudi (Main) and PhD student Morten Nilsen (Co).

First, I would like to thank my main supervisor, Knut Rudi, for allowing me to take part in the PreventADALL project. He has been an outstanding mentor with bright ideas and solutions to any problem, big or small. His enthusiasm and positive mindset lay the path for a dedicated, active, and inclusive research environment in the Microbial Diversity (MiDiv) lab group. I also owe a massive thanks to my co-supervisor, Morten Nilsen, who has been an enormous support in the research project. His devotion to the subject has inspired and taught me so much.

I would also like to thank the PreventADALL study members for all their help in making this project possible. Hilde Aaneland for organizing the fecal sample collection and Axel Olin, Petter Brodin with members of the BRODIN LAB in Stockholm for analyzing the immune cells

Furthermore, I am very grateful to the astonishing laboratory engineers Inga Leena Angell and Ida Ormaasen who have guided me through all the bumps in the road of this process with their academic and technical expertise. A huge thanks to my fellow master students in the MiDiv lab group, Unni Lise Albertsdottir Jonsmoen, Mari Raudstein, Regina Sørensen, and Fredrik Johansen that have created a great working environment and made the time in the lab enjoyable and memorable.

Lastly, I would like to give my biggest gratitude to my family, roommates, friends, and colleagues for their encouragement and support throughout this project and my academic degree.

Ås, 2020

Cecilie Fredheim

Sammendrag

Den menneskelige mage-tarmkanalen rommer et sammensatt og mangfoldig økologisk felleskap av kommensal tarmflora tett tilknyttet verten. Fra dyrestudier er det bevist at tarmfloraens komponenter og metabolitter er essensielle for korrekt modulering og modning av vertens immunsystem, særlig i spedbarnsalder, da perturbasjon av tarmflora tidlig i livet kan forstyrre utviklingen av immunforsvaret og senere medføre immunrelaterte sykdommer. Fremskritt i metagenomiske analyser tillater nå en mer dyptgående undersøkelse av sammenhengene også mellom menneskelig immunitet og tarmflora. Dette prosjektet tok dermed i sikte på å bestemme tarmfloraens relative taksonomiske og utledete metabolske sammensetning det første leveåret for å undersøke mulige assosiasjoner til den relative immuncellesammensetningen ved 12 måneders alder.

Studien inkluderte deler av biologiske prøver samlet i Prevent Atopic Dermatitis and ALLergies kohortstudien. Tarmfloraens taksonomiske og utledete metabolske sammensetning ble bestemt fra langsgående fekale prøver ved henholdsvis Redusert Metagenomisk Sekvensering med Kraken2 HumGut- og Virtual Metabolic Human-databasen. De fekale prøvene fordeler seg over følgende tidspunkt: 60 prøver fra spedbarnets første avføring (mekonium) og deres gravide mødre, 59 prøver fra 3 og 6 måneder og 180 prøver fra 12 måneder. Korrelasjonsanalyser med korreksjon for multippel testing ble deretter utført på disse tarmflora sammensetningene til den tilgjengelige immuncellesammensetning for 67 av barna ved 12 måneders alder.

Resultatene viste negative korrelasjoner mellom relativ forekomst av multiple slekter, arter og utledede metabolitter i prøver fra mødre, mekonium og 12 måneder, til relativ forekomst av klassiske, ikke-klassiske og proinflammatoriske monocytter ved 12 måneder. Et annet funn var den positive korrelasjonen mellom den relative forekomsten av fire arter ved 3 og 6 måneder til den relative forekomsten av naive $CD8^+$ T-celler ved 12 måneder. Den relative forekomsten av to slekter ved 3 måneder var positivt korrelert til den relative forekomsten av T-hukommelsesceller ved 12 måneder, mens den relative forekomsten av utledet smørsyre var negativt korrelert til den relative forekomsten av $CD56^{bright}$ NK-celler, begge ved 12 måneder. Oppsummert indikerer funnene at den relative forekomsten av flere mikrobielle taksonomiske og utledete metabolske egenskaper i tarmen er assosiert med den relative forekomsten av flere immunceller. Av immuncellene viser monocytter den sterkeste assosiasjonen til tarmflora. Imidlertid trengs det ytterligere forskning for å utforske de biologiske implikasjonene til de identifiserte korrelasjonene.

Abstract

The human gastrointestinal tract harbors a complex and diverse ecological community of commensal gut microbes closely connected to the host. Animal-based studies have found that gut microbial components and metabolites are essential for correct modulation and maturation of the host's immune system, especially during infancy. An early-life perturbation in the gut microbial community can disrupt the immune system development and lead to later immune-related diseases. Recent advancement in metagenomic analysis enables now a more in-depth exploration of human immunity and gut microbiota connections. Thus, did this project aim to determine the gut microbiota's relative taxonomic and inferred metabolic composition in the first year of life to investigate possible associations to the relative immune cell composition at 12 months of age.

The study included a subset of biological samples from the Prevent Atopic Dermatitis and ALLergies cohort study. The gut microbiota's taxonomic and inferred metabolic composition was determined from longitudinal fecal samples by Reduced Metagenome Sequencing with the Kraken2 HumGut and the Virtual Metabolic Human database, respectively. The following time points distribute the fecal samples: 60 samples from the infants' first feces (meconium) and corresponding pregnant mother, 59 samples from 3 and 6 months, and 180 samples from 12 months. Correlation analysis with correction for multiple testing was conducted on these gut microbial compositions to the available immune cell composition for 67 of the children at 12 months of age.

The analysis identified negative correlations of the relative abundance of multiple gut microbial species, genera, and inferred metabolites in mother, meconium, and 12 months samples to the relative abundance of classical, nonclassical, and proinflammatory monocytes at 12 months. Another finding was the positive correlation in the relative abundance of four species at 3 and 6 months to the relative abundance of naïve $CD8^+$ T cells at 12 months. The relative abundance of two genera at 3 months were positively correlated to the relative abundance of memory T cells at 12 months, while the inferred relative abundance of butyric acid at 12 months was negatively correlated to the relative abundance of $CD56^{bright}$ NK cells at 12 months. These findings indicate that the relative abundance of several gut microbial taxonomic and inferred metabolic characteristics is associated with the relative abundance of several immune cells. Of the immune cells, monocytes show the strongest connection to the gut microbiota. However, further research is necessary to explore the biological implications of the identified correlations.

Abbreviations

GI tract	Gastrointestinal tract
NCDs	Noncommunicable Diseases
SCFAs	Short Chain Fatty Acids
HMOs	Human Milk Oligosaccharides
GALT	Gut Associated Lymphoid Tissue
Tregs	T regulatory cells
APCs	Antigen Presenting Cells
IgA	Immunoglobulin A
GF animals	Germ Free Animals
rRNA	Ribosomal Ribonucleic Acid
RMS	Reduced Metagenome Sequencing
PCR	Polymerase Chain Reaction
AFLP	Amplification Fragment Length Polymorphism
NGS	Next Generation Sequencing
dNTP	deoxynucleotide Triphosphate
TGS	Third Generation Sequencing
BLAST	Basic Local Alignment Search Tool
LCA	Least Common Ancestor
VMH database	Virtual Metabolic Human database
CyTOF™	Cytometry Time Of Flight
PreventADALL	Preventing Atopic Dermatitis and ALLergies in children
qPCR	quantitative Polymerase Chain Reaction
NSC	Norwegian Sequencing Centre
bp	Base pair
NK cells	Natural killer cells

Table of Contents

1. Introduction

1.1 The human gut microbiota

1.1.1 The human gastrointestinal tract

The human gastrointestinal (GI) tract is a continuous channel through the body composed of a series of hollow organs. It starts with the oral cavity, following the pharynx, esophagus, stomach, small intestine, large intestine, rectum, and anal canal, ending with the anus. The tongue, salivary glands, liver, pancreas, and gallbladder serve as accessory organs. This complex organ system serves essential functions that are crucial for maintaining human health. Firstly, it is responsible for digesting and absorbing food and drinks. The nutrients and water are taken up in the blood and lymph vessels via transcellular transport through the epithelial lining and then transported to the different body sites for energy and macromolecule extraction (Liao, Zhao, & Gregersen, 2009). Lastly, it is responsible for expelling undigested food and waste products that the body does not need through feces (Liao et al., 2009).

Crypts, villi, and microvilli compose the intestinal wall in the GI tract, which enlarges the luminal surface (Walton, Freddo, Wang, & Gumucio, 2016). The large surface area leads to more efficient absorption. However, it also makes the human body more vulnerable to potential dangers in the luminal content. Thus, the mucosal membranes serve as a physical, chemical, and immunological barrier to protect the body against dangerous luminal content (Okumura & Takeda, 2018). The most crucial physical barrier is the lining of the mucosal membrane by several kinds of specialized and polarized epithelial cells bound together by tight junctions in a monolayer. Additionally, epithelial goblet cells secrete glycosylated proteins called mucins forming mucus that reside outside the epithelial layer in the lumen (Okumura & Takeda, 2018). Chemical barriers include antimicrobial peptides secreted by epithelial paneth cells. Immunological barriers consist of intraepithelial immune cells and different immune cells and effector proteins that reside in lamina propria, the connective tissue underneath the epithelial layer (Okumura & Takeda, 2016).

1.1.2 The composition and function of the gut microbiota

The human body's internal and external surfaces, including the skin and mucosal surfaces (vagina, oral cavity, nasal cavity, and GI tract), respectively, are habitats for numerous microorganisms called the human microbiota (Kumar & Chordia, 2017). The species of the human microbiota belong predominantly to bacteria (Lloyd-Price, Abu-Ali, & Huttenhower,

2016), but also archaea (Horz, 2015), virus (Cadwell, 2015), protists, and fungi (Parfrey, Walters, & Knight, 2011) are present. Since bacteria are the main microorganism inhabiting all different niches in the body, they are also more extensively studied. The number of bacterial cells inhabiting an average human body is estimated to 4×10^{13}, making bacterial cells to human cells' ratio 1,3 (Sender, Fuchs, & Milo, 2016).

The most substantial part of the human microbiota is the gut microbiota, which is the collection of microorganisms that inhabit the intestines (Ursell, Metcalf, Parfrey, & Knight, 2012). The microbial density and diversity increase both across the length of the GI tract and from the epithelial layer to the intestinal lumen (Sekirov, Russell, Antunes, & Finlay, 2010). The site with the highest microbial density and richness of microbes is consequently the lumen in the large intestine. This distribution is due to the low cell turnover rate, low redox potential, and long transit time in the colon lumen (Hillman, Lu, Yao, & Nakatsu, 2017).

Analysis of the gut content indicates that around 160 bacterial species reside at any given time and that the specie composition varies with around 1150 different bacterial species (Qin et al., 2010). The species belong mainly to the phyla Firmicutes and Bacteroidetes, with a smaller percentage to the phyla Actinobacteria, Proteobacteria, and Verrumicrobia (Tap et al., 2009). Taxonomic diversity is associated with health and disease. A rich and diverse gut microbiota will better withstand external threats, while a microbiota with a lack of diversity is associated with different immune-related noncommunicable diseases (NCDs) such as asthma and allergies (West et al., 2015). Aside from the taxonomical composition, the functional composition is also variable. The human gut microbiome is the collection of genes from the genomes of all species in the microbiota (Lozupone, Stombaugh, Gordon, Jansson, & Knight, 2012). The genes function as predictions for the functionalities of the gut microbiota, and this functional aspect is essential for the interaction between the gut microbiota and the human host.

The relationship between the gut microbes and their respective host are pathogenic, commensal, or symbiotic. Pathogenic microorganisms cause damage to the host, while the two latter conceptions are grouped under mutualism and do not cause damage to the host (Hooper & Gordon, 2001). Commensalism is the collective term for the host-microbe interaction in a healthy individual, but the bacteria have a stronger relationship with the host than the term suggests (Haque & Haque, 2017). According to Haque & Haque in 2017, commensalism is a form of symbiosis whereby one organism gains from its association with another organism, whereas the other is affected in neither a positive nor a deleterious manner.

There are several examples of the strong symbiotic relationship and interplay between the human host and its commensal gut microbiota. The mucus covering the epithelial surface is distributed with an impermeable inner layer to protect the host against the invasion of pathogenic microbes. The outer layer is more permeable than the inner layer and works as both a substrate and a habitat for the commensal microbes (Sicard, Le Bihan, Vogeleer, Jacques, & Harel, 2017). Myolytic bacteria possess the enzymatic activity and collaborate in a community to degrade the mucins. They utilize these endogenous glycans as a source of nutrient and carbon source (Koropatkin, Cameron, & Martens, 2012). Because of the high number of microbes in the large intestine, there is also a higher number of goblet cells producing mucins, making the mucus layer very thick in that part of the gut (Okumura & Takeda, 2017). The commensal bacteria return the favor and use both direct and indirect pathways that protect the host against pathogens. Directly, the commensal bacteria perform colonization resistance by inhibiting pathogen growth due to secreting inhibitory antimicrobial substances and competing for nutrients (Sorbara & Pamer, 2019). Indirectly, they stimulate the intestinal barrier's strength and maintenance of the intestinal tight junctions between the epithelial cells (Hiippala et al., 2018).

The gut commensal bacteria utilize exogenous nutrients from the diet in addition to the endogenously secreted mucins. Parts of the gut microbiota can degrade undigestible, dietary fiber by saccharolytic fermentation in the colon. Some of the fermentative bacteria produce intermediate products such as fumarate, succinate, and lactate that other bacteria can convert further to the final products called short-chain fatty acids (SCFAs) (Rowland et al., 2018). This kind of process, where one organism utilizes the end product of the metabolic pathway of another organism, is called cross-feeding interactions and is famous for the gut microbes (Ríos-Covián et al., 2016). The main SCFAs are acetate, propionate, and butyrate. The primary producers of acetate and propionate are species in the Bacteroidetes phylum, while the primary producers of butyrate are species in the Firmicutes phylum (Rowland et al., 2018). The acids have several beneficial effects for both colonic and overall human health. SCFAs serve as an energy source for the colonocytes and resident bacteria, enhance mineral absorption, reduce luminal pH, and consequently limit pathogen growth (Alexander, Swanson, Fahey, & Garleb, 2019). Other exogenous nutrients that commensal bacteria provide the host are essential vitamins (K and B) and amino acids (Rowland et al., 2018).

1.1.3 The development of the infant gut microbiota

When the colonization of the gut starts is a debated subject. For a long time, the accepted central dogma was that the fetus is sterile *in utero* and that the microbial colonization starts during and after birth (Rodriguez et al., 2015). A considerable number of studies have however challenged this assumption as bacterial genomes have been detected in the placenta (Aagaard et al., 2014), umbilical cord (Jimenez et al., 2005), amniotic fluid (Bearfield, Davenport, Sivapathasundaram, & Allaker, 2002), and in the infant's first feces (meconium) (Jimenez et al., 2008). On the contrary, recent studies argue against that colonization begins *in utero* because there were not any detectable microbial community in the placenta (Leiby et al., 2018), nor the amniotic fluid in healthy term pregnancies (Lim, Rodriguez, & Holtz, 2018). Perez-Muñoz et al. conclude in their review that current scientific evidence does not support the existence of microbiomes within the healthy fetal milieu (Perez-Munoz, Arrieta, Ramer-Tait, & Walter, 2017). In compliance with the latest findings, the correct time for the initial colonization of gut microbiota for healthy children is likely after the rupture of the amniotic membrane during birth (R. E. Moore & Townsend, 2019).

Several factors contribute to shaping the composition of gut microbiota early in life. As previously addressed, the first inoculum of the infant takes place during birth, and the mode of delivery strongly influences the early gut microbial composition (Munyaka, Khafipour, & Ghia, 2014). Vaginally born children are primarily colonized by bacteria from the birth canal and the mother's gut, while cesarean born children are primarily colonized by bacteria from the skin and the environment (Dominguez-Bello et al., 2010). Whether the child is born preterm or at term will affect the development of the early gut microbiota. The gut of preterm infants have delayed colonization and decreased diversity (Henderickx, Zwittink, van Lingen, Knol, & Belzer, 2019). Correlating factors that affect the gut microbiota of preterm infants include that they are more frequently born by cesarean section, are hospitalized more extended period, and are more frequently treated with antibiotics (Milani et al., 2017). Antibiotics during pregnancy and in infancy disrupt the development of the gut microbiota by decreasing bacterial diversity, delay colonization, and increasing antibiotic resistance genes and species (Henderickx et al., 2019).

The infant's GI tract is exposed to numerous microbes with the ingested food in addition to microbes in the environment. *Bifidobacterium* species enrich the gut microbiota of breastfed infants (Turroni et al., 2012). This colonization occurs because breastmilk contains species

belonging to this genus (Soto et al., 2014). Additionally, human milk oligosaccharides (HMOs) shape breastfed infant's gut composition by selectively colonization of species belonging to *Bifidobacterium* and *Bacteroides* that have enzymes to ferment these (Marcobal & Sonnenburg, 2012). HMOs are prebiotic, which is defined by Gibson et al. in 2004 to be a selectively fermented ingredient that allows specific changes, both in the composition and activity in the gut microbiota, that confer benefits upon host well-being and health (Gibson, Probert, Loo, Rastall, & Roberfroid, 2004). The breast milk microbiota and HMO composition are important determinants of the infant gut microbiota (Ho et al., 2018), and the effects are dose-dependent (Pannaraj et al., 2017). Children that are fed formula will contrary possess a higher bacterial diversity resembling that of an adult, including fewer *Bifidobacterium* (Ho et al., 2018). When weaning occurs, and the introduction of solid foods starts, the composition starts resembling the gut microbiota of an adult.

The infant gut microbiota is less diverse, more unstable, and dynamic than the adult microbiota. The shift towards the "adult-like" gut microbiota can be observed from around 12 months of age and will be complete within the first three years of life (Arrieta, Stiemsma, Amenyogbe, Brown, & Finlay, 2014). The gut lumen is an aerobic environment following birth. In parallel with the rapid colonization, the gut gradually becomes anaerobic, and aerobic bacteria such as *Enterobacteriaceae* are replaced with anaerobic bacteria such as *Bifidobacterium*, *Clostridium*, and *Bacteroides* (Arrieta et al., 2014). A common understanding formulated by Matamoros et al. in 2013 is that it is becoming evident that initial microbial colonization and the resulting immune and metabolic programming have a long-lasting influence on the risk for diseases (Matamoros, Gras-Leguen, Le Vacon, Potel, & de La Cochetiere, 2013). With that in mind, numerous researchers are aiming to figure out which interventions can modulate the infant microbiota, such as probiotics, prebiotics, antibiotics, and most invasively, fecal transplantation, giving the infant and later adult the best possible health outcome.

1.2 The human immune system

1.2.1 The innate and adaptive immune system

The immune system divides into the innate and the adaptive one and consists of both immune cells and effector molecules (Nicholson, 2016). The immediate innate immune system includes physical and chemical barriers as the epithelial surface lining the mucosal tissue in the gut and antimicrobial peptides, respectively. It also includes the complement system and unspecific phagocytosis by phagocytes (Nicholson, 2016). The induced innate immune system involves recognition of the pathogen by binding to specific receptors followed by the recruitment of effector molecules called cytokines and interferons. The adaptive immune system involves humoral and cellular responses based on binding of specific receptors to antigens, carried out by B cells and T cells, respectively (Nicholson, 2016).

1.2.2 The development of the immune system

The immune system is relatively immature at birth and evolves throughout all life stages (Simon, Hollander, & McMichael, 2015). Neonatal immunity is dominated by the naïve phenotypes, with impairment of both the innate and adaptive immune system. The innate immune cells have non-optimal functions, and serum concentrations of complement components are considerably lower (Simon et al., 2015). There is reduced efficiency of the adaptive immune system, with weak Th1 and antibody responses (Simon et al., 2015). This composition will shift to more mature phenotypes in parallel to exposure to the environment, such as vaccines, commensal, and infectious microbes (Brodin & Davis, 2017). Immunity at the mucosal surfaces such as the intestines, urogenital tract, and respiratory system are establishing by three months of age. However, maturation varies from 1 to 6 months (Gleeson & Cripps, 2004).

Each immune profile remains stable over time and will return quickly and recover to baseline after perturbation, such as vaccination or infection (Tsang et al., 2014). There is a distinct inter-individual variation in both composition and function that can be many orders of magnitude in healthy individuals (Brodin & Davis, 2017). Both heritable and non-heritable factors contribute to the function and composition of the immune system, but the extent remains unclear. The heritable factors involve genetics of immune cell frequencies, genetics of serum protein concentrations, and gene expression leading to different functional immune responses (Brodin & Davis, 2017). The non-heritable factors are environmental factors such as vaccines, infectious pathogens, and commensal microbiota. Symbiotic and pathogenic microbes

seemingly explain most of the functional and compositional variation in the immune system (Brodin & Davis, 2017).

1.2.3 The gut mucosal immunity

Immune responses in the gut differ from immune responses in the skin. The inflammatory response following a wound in the skin surface is necessary to recruit immune cells since they do not typically reside in the dermis. The damaged tissue can be repaired and restored after inflammation without much risk and problems. On the contrary, inflammation in the gut is very dangerous, and a typical symptom of intestinal diseases. To prevent inflammation, a continuous layer of gut-associated lymphoid tissue (GALT) lines the gut (Randall & Mebius, 2014). Both innate and adaptive immune cells reside in the GALT and will monitor and thereby detect and fight intruders or possible threats very effectively. Additionally, the components of the gut immune system have effective strategies to tolerate commensal bacteria and harmless food antigens (Randall & Mebius, 2014).

Two continuous strategies exist to prevent inflammation and keep the GI tract healthy and free from invading pathogens. The first strategy is called immune suppression and is administrated by the subpopulation of T cells called regulatory T cells (Tregs) that mediates immune tolerance and limits inflammatory response (Lazar et al., 2018). Furthermore, antigen-presenting cells (APCs) induce immunoglobulin A (IgA) class switching in naïve B-cells, an immunoglobulin isotype that does not promote inflammation (Okumura & Takeda, 2016). Immune exclusion is the second strategy, which works to keep the pathogens from entering the body. This strategy involves neutralizing of the pathogens by binding to secretory IgA and thereby anchoring and entrapping them to mucins in the mucus (Mantis, Rol, & Corthesy, 2011).

1.2.4 The effects of gut microbiota on the immune system

Germ-free (GF) animal studies have provided insight into how the gut microbiota influences the host. In comparison to animals colonized by microbiota, GF animals appear to have physiological and functional differences in the gut. The differences include enlargement of the cecum due to accumulation of undegraded mucus, reduced gastrointestinal motility as a consequence to loss of critical digestive functions usually done by the microbiota, as well as aberrant morphology of the epithelial cells, with longer villi and shorter crypts, and reduced amount of antimicrobial peptides (Gensollen, Iyer, Kasper, & Blumberg, 2016). Studies using GF animals have also shown that the microbiota affects the maturation and function of the gut mucosal immune system structurally and functionally. The gut-associated lymphoid follicles

are absent, and the Peyer's patches and mesenteric lymph nodes are smaller in the colon of GF animals (Gensollen et al., 2016). The influence that gut microbiota has on the gut mucosal immune system extends beyond the GI tract and affect the systemic immune system (Kabat, Srinivasan, & Maloy, 2014). Colonization of commensal microbes is especially crucial at infancy to ensure a correct maturation of the immune system (Zheng, Liwinski, & Elinav, 2020). Early life perturbation of gut microbiota can potentially result in disrupted development of the immune system and lead to immune-related diseases later in life (Gensollen et al., 2016).

The gut microbiota influences the immune system in several ways. Animal-based studies have detected that both microbial components and metabolites are mediators of the interplay between the gut microbiota and the immune system. Microbial components from commensal microbiota induce, directly and indirectly, the secretion of antimicrobial peptides from epithelial Paneth cells that ultimately prevents the colonization of new and harmful microbes. This occurs directly by the recognition of lipopolysaccharides by intestinal epithelial cells (Kabat et al., 2014). Indirectly, this occurs by recognition of flagellin by dendritic cells resident in the lamina propria that leads to the activation of innate lymphoid cells (Kabat et al., 2014). The microbial metabolites, SCFAs, promote gut homeostasis by enhancing the epithelial barrier function and promoting immune tolerance. Specifically, microbial production of SCFA leads to increased mucin production, increased secretion of secretory IgA, increased number and function of Tregs, and reduced expression of T-cell activating molecules in APCs (Rooks & Garrett, 2016).

1.3 Methods to study gut microbiota-host interactions

The gut microbial research field was until the late 20th century, dominated by culture-dependent approaches. The conditions and growth medium must replicate the native physiological niche in the gut lumen to successfully cultivate a specie that is normally present in the gut (Hiergeist, Gläsner, Reischl, & Gessner, 2015). Consequently, this way of studying the composition favors species with effective cultivation protocol and limits the species that resist cultivation, giving biased results. An example is that *E.coli* was former wrongly believed to be a highly prevalent member of the human gut microbiota (Lloyd-Price et al., 2016). The first cultivation techniques managed to cultivate less than 20% of the gut microbiota, but the development of strict anaerobic cultivating methods in the late 1960s provided a more representative insight to gut microbiota. The cultural counts increased to 93% of the total microscopic counts in 1974, but the method fails to classify below the genus level for hundreds of the isolates (W. E. Moore & Holdeman, 1974). A fecal sample contains hundreds of different isolates, and a full description of the morphologic, biochemical, and physiologic characters of every isolate in several samples are physically impossible (Rajilić-Stojanović & de Vos, 2014).

The technical limitations of culture-dependent approaches prompted the development of reliable culture-independent techniques (Suau et al., 1999) (Nichols et al., 2010). A culture-independent analysis permits detection of species that resist cultivation or is not discovered yet (Milani et al., 2017). Metagenomic has become an established culture-independent approach that studies metagenomes by high throughput sequencing directly from a complex environmental sample. The metagenomic sequences are used to determine the taxonomic composition, and the microbiotas possible activities and functional roles in the gut are deduced indirectly and directly from the taxonomic composition, and the gene sequences, respectively.

Metagenomic sequencing for microbial identification is possible by different approaches. The 16S ribosomal RNA (rRNA) marker gene technique is a commonly used variant that sequence parts of the prokaryotic 16S rRNA gene and uses the variable regions for taxonomic identification and functional imputation (Rausch et al., 2019). The 16S rRNA technique holds limitations regarding resolution, as it is problematic to separate microbes that are closely related, and reliable classification will only be on the family- or genus level (Earl et al., 2018). Another commonly used technique is the whole genome shotgun approach, which works by sequencing a fragmented metagenome and reconstructing it to complete genomes that allow for both taxonomic and functional identification on the strain level (Rausch et al., 2019). However,

the shotgun method is of high cost and holds challenges by producing massive and complex data that causes bioinformatical challenges (Sharpton, 2014). The novel technique called Reduced Metagenome Sequencing (RMS) is an alternative to the mentioned microbial identification methods that sequence a reduced part of the metagenome after enzymatic restriction cutting. Following any of the mentioned techniques, the determination of the base sequences is done by sequencing technologies separated into three generations. Subsequent sequencing, the sequences are aligned to sequences in a reference database, and the taxonomic and functional information can thereby be determined.

1.3.1 Reduced metagenome technique

RMS is a novel throughput metagenomic method for microbial profiling of the gut. This technique fragments genomic DNA using two restriction enzymes, MseI, that cuts frequently, and EcoRI that cuts infrequently, due to restriction sites of four (5' T|TTAA 3') and six (5' G|GAATTC 3') base pairs, respectively. Adaptors are then ligated to the fragments using ligase. Adaptor ligation creates a universal primer binding site flanking all fragments, making amplification possible. The adaptors contain a core sequence and enzyme-specific sequence. A selective amplification proceeds with polymerase chain reaction (PCR) for the adaptor-ligated sequences cut by both restriction enzymes. Fragments cut by only one of the restriction enzymes will create hairpin loops that terminate the amplification. Figure 1.1, at the end of this subchapter, illustrates the principle for sample preparation in the RMS technique. The illustration is modified and redrawn from Ravi et al. (Ravi et al., 2018).

The fragments make up around 10% of the metagenome in the sample, resulting in manageable amounts of data. The method provides high taxonomic resolution and potential genomic and functional assignment of the gut microbiota similar to what the whole genome shotgun technique does while being as effective and low-cost as the 16S rRNA marker gene technique for large sample sets (Ravi et al., 2018). Unlike the 16S rRNA method, the RMS method is not limited to detecting organisms holding a specific gene and is, therefore, able to capture a wider variety of organisms, e.g., viruses and fungi (Hess et al., 2020). RMS is an alternative to the double digest Restriction Site Associated DNA method that uses the enzyme combination N1aIII and HpyCH4IV (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) (Liu et al., 2017) and Restriction Enzyme-Reduced Representation method that uses the enzyme ApeKI or PstI (Hess et al., 2020). The principle of RMS and the two mentioned methods all originate from a technique called Amplification Fragment Length Polymorphism (AFLP). This method

similarly treats genomic DNA with restriction enzymes, but unlike the sequencing methods, the AFLP technique only analyses the amplified restricted fragments on a gel (Vos et al., 1995).

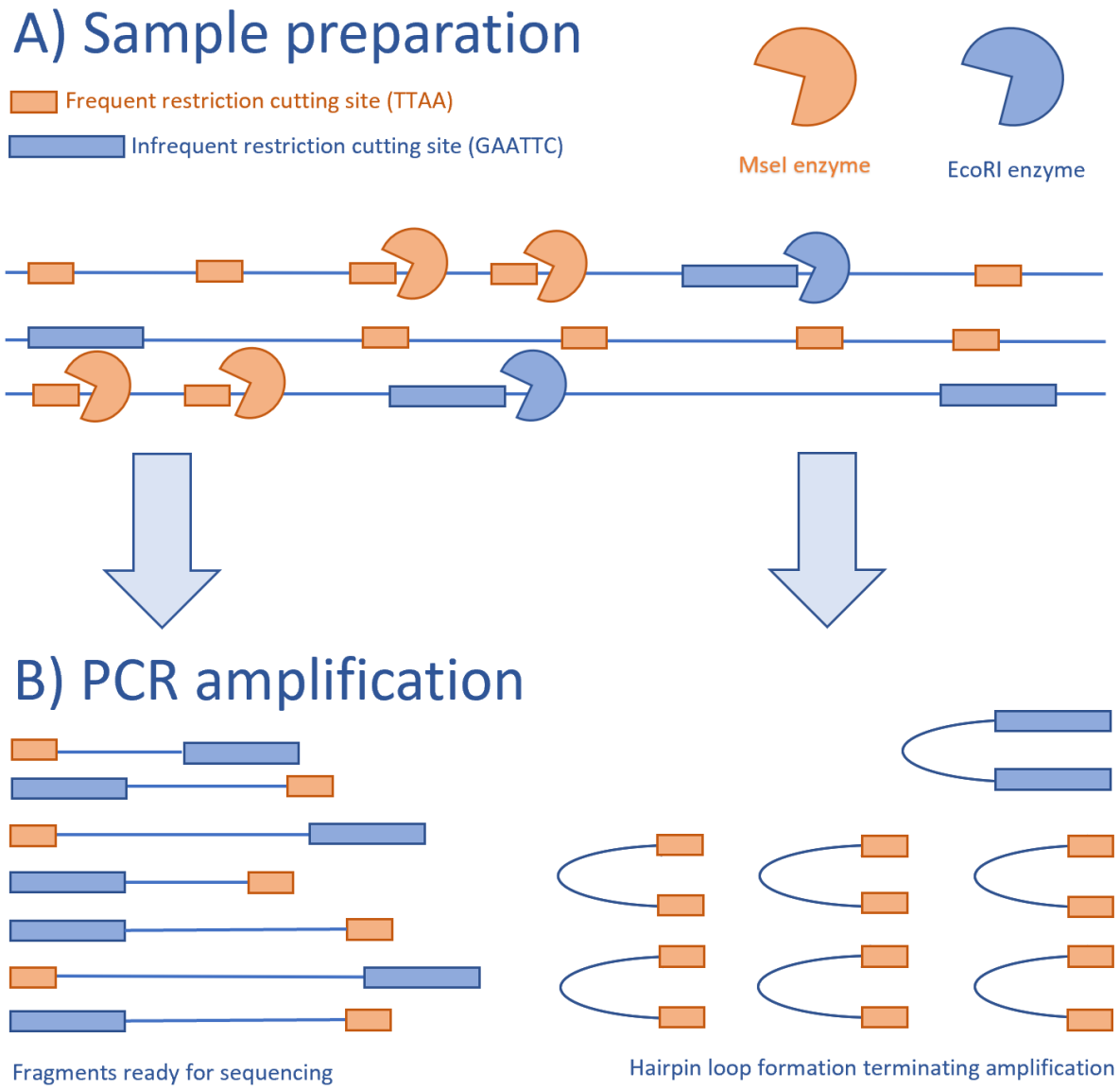

Figure 1.1. **Illustration of the preparation in Reduced Metagenome Sequencing.** Panel A) shows sample preparation where two restriction enzymes cut the genomic DNA, MseI (orange) that cuts a frequent cutting site (orange) and EcoRI (blue) that cuts an infrequent cutting site (blue). B) shows PCR amplification for fragments flanking both cutting sites. These fragments are ready for sequencing. Fragments flanked by the same cutting sites create hairpin loops and are consequently not amplified and sequenced. The illustration is modified and redrawn from Ravi et al. (Ravi et al., 2018).

1.3.2 Three generations of sequencing technologies

Accurate and efficient sequencing technology is necessary to determine the nucleic acid sequences when selecting any of the three described metagenome techniques. DNA sequencing technology emerged with first-generation sequencing developed by Dr. Frederik Sanger in 1977. The Sanger sequencing method of single-stranded DNA uses the chain termination

dideoxynucleotide triphosphate method (Sanger, Nicklen, & Coulson, 1977). The sequencing technique is still used due to its high accuracy and long read length (Kircher & Kelso, 2010). However, the method is time-consuming, costly, and low throughput, and is consequently not favorable for large-scale applications (Churko, Mantalas, Snyder, & Wu, 2013).

Second-generation sequencing, commonly called next-generation sequencing (NGS), has revolutionized the world of sequencing since it emerged in the first decade of the twenty-first century. The different technologies all have in common that they are high-throughput, fast, inexpensive, and perform massive parallel sequencing that produces millions to billions of short DNA reads (Levy & Myers, 2016). Several different sequencing platforms exist using different techniques, such as Ion Torrent, SOLiD, Roche 454, and Illumina.

Illumina MiSeq and HisSeq are two popular NGS sequencing platforms. In preparation for sequencing, Illumina adaptors must flank the DNA fragments (Kchouk, Gibrat, & Elloumi, 2017). The adaptors contain complementary sequences to the oligonucleotides on the flow cell were sequencing occurs, and the fragments will be attached to these in the following clonal PCR bridge amplification. This amplification creates around one million identical copies of each fragment, forming a cluster. Sequencing occurs by synthesis when primers, fluorescently modified dNTPs, and DNA polymerase adds to the mix (Kchouk et al., 2017). The nucleotides work as reversible terminators with a fluorophore that occupies the 3' -OH group that must be enzymatically cleaved. The fluorophore assures that the synthesis occurs in a synchronous manner (Heather & Chain, 2016). The clusters emit identical signals that are detected and computationally translated to determine the sequence. The fragments can be read from both ends and produce paired-end sequences (Heather & Chain, 2016). The different platforms differ in output range, run time, reads per run, maximum read length, and price. Illumina sequencing produces high output data, is done at low cost, but give shorter read lengths that can increase errors in assembly (Oulas et al., 2015).

Third-generation sequencing (TGS) is an emerging technology that solves some of the problems of NSG. TGS produces longer read length that simplifies assembly, reduces the price of sequencing, reduces time, and simplifies the preparations by, e.g., excluding the need for PCR amplification (Kchouk et al., 2017). However, TGS have much higher error rates than NGS technologies (Alvarez, Skachkov, Massey, Kalitsov, & Velev, 2015). Oxford Nanopore MinION sequencer and Pacific Biosciences are the two commonly used TGS platforms. In the MinION sequencer, the DNA fragment passes a protein nanopore and generates an ionic current

that translates to the sequence (Kchouk et al., 2017). Pacific Biosciences, conversely, utilizes fluorescent labeling like the NSGs but do not involve amplification and detects in real-time (Kchouk et al., 2017).

1.3.3 Bioinformatic tools for taxonomic and functional classification

The output after metagenomic sequencing is thousands of unclassified sequences that must be taxonomically classified to provide any useful information from the metagenomic sample. Two classification tools used for doing this are the Basic Local Alignment Search Tool (BLAST) (Altschul, Gish, Miller, Myers, & Lipman, 1990) and Kraken tool (Wood & Salzberg, 2014). BLAST assign taxonomy to unknown sequences by finding the best alignment to an extensive database of genomic sequences. BLAST is the most popular program but is not initially intended for metagenomic sequences. On the contrary, Kraken is as accurate as BLAST and other equivalent programs, but are made for metagenomic sequencing data and outcompete them in speed (Wood & Salzberg, 2014).

Kraken is an efficient tool in the taxonomic classification of metagenomic output due to the utilization of K-mers. The standard Kraken database contains k-mers containing 31 bases (K=31) and the Lowest Common Ancestor (LCA) of all organisms with the k-mer in their genomes (Wood & Salzberg, 2014). Kraken approaches the metagenomic sequence by matching all possible k-mers in the metagenomic sequence to k-mers in the database and assigning the taxa by use of LCA belonging to the specific k-mer (Wood & Salzberg, 2014). Metagenomic sequences that do not contain k-mers in the database remain unclassified. It is possible to develop a database with the Kraken tool. Kraken2 HumGut_05 database is an example of this, described in a manuscript by Hiseni et al. (Hiseni, Rudi, Wilson, Hegge, & Snipen, 2020). The HumGut database contains k-mers with K=35 from bacterial genomes that are generally present in a healthy human gut. The collection comprises more than 4779 genomes, representing 1201 unique taxonomy IDs. These genomes are either fully or 95% contained in 2311 healthy human gut metagenomes coming from all around the world.

Metagenomic sequencing provides an exploration of functional information aside from the taxonomical information. The genes present in the metagenome indicates the potential functional properties, and this functional information is useful in studying environmental samples, such as the gut lumen. Functional analysis of metagenomic data encounters several challenges due to the computational problems because of the vast amount of data and short read-lengths during sequencing (Prakash & Taylor, 2012). However, the functional information

can also be inferred by the taxonomical information. The Virtual Metabolic Human (VMH) database is a sequence database that encapsulates the current knowledge of human metabolism in five interconnected resources (Noronha et al., 2019). The gut microbiome is one of these resources, and the database has captured 818 microbes and 632 685 microbial genes (Prakash & Taylor, 2012). With the use of VMH, it is possible to reconstruct the inferred metabolic potential to the gut microbes. The database can consequently be a useful tool in studying the functionalities of the gut microbiota.

1.3.4 Methods to study immune cell composition

A functional immune system is critical to maintaining human health, and the need to understand it makes it a relevant study field. Although blood is not an immunological organ, it holds most of the immune cells circulating the body. Analysis of immune cells in blood samples is consequently a reliable proxy for the human immune system in a non-invasive way (Brodin & Davis, 2017). Reliable methods are necessary to study the immune cells, and the development of those is making substantial progress in immunology (Simoni, Chng, Li, Fehlings, & Newell, 2018).

Single-cell analysis platforms are the current method of choice in immunologic research, and flow cytometry has been the cornerstone technology for decades. The technique enables single-cell analysis by using fluorescently labeled antibodies to measure up to 15 simultaneous parameters (Brodin & Davis, 2017). However, there was an increasing need for a high throughput technique that enables single-cell resolution with high parameterization (Spitzer & Nolan, 2016). This need inspired the fusion of the two technologies flow cytometry and mass spectrometry, creating mass cytometry (Bandura et al., 2009). The technique bases on inductively coupled plasma mass spectrometry and time of flight mass spectrometry. The single-cell suspension is first incubated with antibodies conjugated to a polymer chain of chelating groups bound to stable heavy metal isotopes (Spitzer & Nolan, 2016). The cells are then nebulized into droplets and sent through inductively coupled argon plasma, leading to ionization of the metal-conjugated antibodies (Spitzer & Nolan, 2016). The mass spectrometer analyses signals from the ionized metals. Cytometry time of flight (CyTOF^{TM}) is the current instrument for mass cytometry, and the technique allows for the quantification of approximately 45 simultaneous cellular parameters that enable the assessment of phenotypes and functions (Brodin & Davis, 2017).

1.4 The cohort study: PreventADALL

The increasing numbers of allergic and immune-related NCDs in the Western world are particularly alarming for human health. These diseases are connected to changes in lifestyle and environment, with microbial exposure, diet, physical activity, and antibiotic treatment being some of them (von Hertzen et al., 2015). The Preventing Atopic Dermatitis and ALLergies in children (PreventADALL) cohort study uses allergic diseases as model diseases for understanding NCDs, as allergies develop early in life. The study aims to collect knowledge to prevent the development of NCDs later in life.

Firstly, the main objective is to determine whether primary prevention of allergic diseases is possible through simple and low-cost strategies. Secondly, it is to assess early life factors and exposures, including intrauterine environment, microbiota, and xenobiotics, involved in the development of asthma and allergic diseases or other NCDs, including cardiovascular diseases, obesity, and diabetes (Lodrup Carlsen et al., 2018). Hopefully, the findings will make it possible to identify personalized novel preventative strategies to related microbial diversity, diet, lifestyle, and gene-environment influence on allergic and other NCD development from fetal life (Lodrup Carlsen et al., 2018).

The study is a general mother-child population-based cohort. It includes a healthy population in Norway and Sweden and aims to be representative of this entire population. Information about health and disease in the mother, child, and family and biological samplings such as blood, skin swabs, urine, and feces are collected from mothers and their children in a time period from 18 weeks pregnant to 4 years onward (Lodrup Carlsen et al., 2018).

1.5 Aim of the thesis

Several animal studies have shown that the gut microbiota is essential for modulating and maturing the immune system. However, these associations have not been studied widely in human infants, limiting the complete understanding of gut microbiota and human immunity connections. Thus, did we want to investigate this connection and hypothesize that differences in immune cell composition are associated with the gut microbiota's taxonomic and metabolic composition.

This thesis aimed to determine the relative taxonomic and inferred metabolic composition of the gut microbiota in the first year of life to investigate possible associations of these gut microbial characteristics to the relative immune cell composition at 12 months by an explorative nature. Several sub-goals were included to achieve this.

- Perform Reduced Metagenome Sequencing on fecal samples collected in the PreventADALL cohort study from meconium, 3, 6, and 12 months of age as well as their respective 18-week pregnant mothers to determine the gut microbiotas composition at specie and genus level using the Kraken2 HumGut database.
- Infer the gut microbiotas metabolic potential the first year of life from the taxonomic composition at the species level using the Virtual Metabolic Human database.
- Correlate the gut microbiota's relative inferred metabolic composition and taxonomic composition at the species and genus level the first year of life to immune cell composition at 12 months. The immune cell composition is determined using mass cytometry by group members of the PreventADALL cohort study.

2. Materials and methods

2.1 Clinical samples

In total, 418 fecal samples from mothers at 18 weeks pregnant, and their children sampled from birth to 12 months collected by PreventADALL were analyzed for this study. Fecal samples from 180 participating children at 12 months were initially chosen based on the knowledge that group members of PreventADALL in Stockholm were analyzing blood samples to determine immune cell composition by CyTOF2 Mass Cytometry for the same children. However, results for only 67 of the children were obtainable for analysis in this master thesis. Thus, we chose to include longitudinal samples for the 67 children and their mothers. Because of missing samples at several time points, the number of samples that were analyzed is distributed in the following matter; 60 fecal samples from the infants' first feces (meconium) and corresponding pregnant mother at 18 weeks, 59 fecal samples from 3 and 6 months and 180 fecal samples from 12 months.

Immune cell analysis was conducted by staff at the Brodin lab in Stockholm, Sweden. Brodin, P., and Tadepally, L. handed a manuscript for the protocol of immune cell phenotyping by Mass Cytometry/CYTOF^{TM} and the antibodies and reagents that were used. This manuscript is presented in Appendix A.

The fecal material was suspended in DNA shield buffer (1:10) to prevent DNA degradation, before it was stored in a freezer at -80°C in Oslo, Norway. The fecal samples were transported in a cooler box with cooling elements to Norwegian University of Life Sciences in Ås, Norway, and analyzed there by the master student. The gut microbiota was analyzed by extracting gene sequences cut by the restriction enzymes EcoRI and MseI, according to the RMS protocol (Ravi et al., 2018) on all fecal samples. The sequencing of the metagenomes was done by the Norwegian sequencing center (NSC) using Illumina HiSeq 3000. The flowchart in Figure 2.1 illustrates the work that has been done in this master thesis.

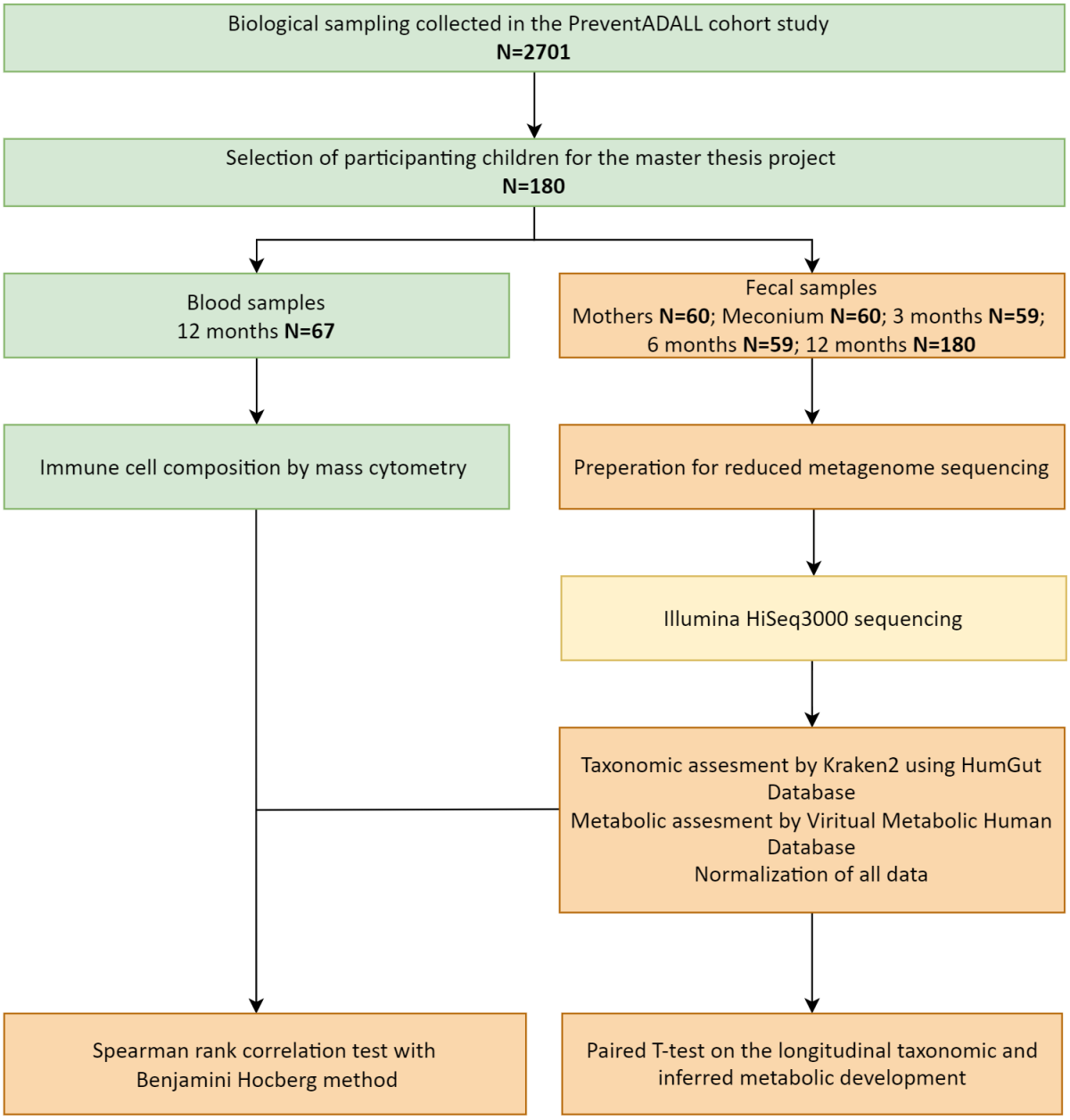

Figure 2.1. **Flowchart.** The flowchart illustrates the workflow of this master thesis. Green boxes have been conducted by group members of PreventADALL, yellow box by the NSC, and orange boxes by the master student, guided by supervisors and laboratory engineers at the Norwegian University of Life Science.

2.2 Sample preparation

All fecal samples were thawed on ice and vortexed for 20 seconds to homogenize the fecal material before any further treatment. Following this step, the samples were pulse centrifuged at 1200 rpm for 8 seconds to separate the homogenized sample from any larger insoluble fecal particles.

2.3 Initial DNA purification

2.3.1 Microbial lysis

The microbial cells in the fecal samples were lysed using a combination of mechanical and chemical stress. The mechanical lysis was done by mixing 200 µl of the homogenized fecal sample in a fast prep tube with 0,2 g acid-washed glass beads (<106 µm Sigma-Aldrich, Germany), 0,2 g acid-washed glass beads (425-600 µm Sigma-Aldrich, Germany) and two large acid-washed beads (2,5-3,5 mm Sigma-Aldrich, Germany). The combination of different bead sizes leads to lysis of both the fragile and more robust cell types present in the sample. Consequently, the combination results in a more representative result after extraction, with higher diversity and less bias (Bakken, 2006). The tubes containing the fecal material and beads were processed twice in FastPrep 96 (MP Biomedicals, USA) at 1800 rpm for 40 sec, followed by centrifugation at 13000 rpm for 5 min. The last centrifugation step aggregates the glass beads and bigger particles, facilitating further purification.

The chemical lysis was done by mixing the supernatant after centrifugation with lysis buffer (Thermo Fisher, USA) and Proteinase K (Thermo Fisher, USA) with the ratio 1:1:0,1, respectively. The samples were placed in the KingFisher Flex robot using "ProteinaseLGCmini" procedure incubating the samples at 55 ºC for 10 min. Proteinase K is an endolytic serine protease that degrades contaminating proteins such as nucleases in the sample, making later PCR amplification more efficient (Crowe et al., 1991).

2.3.2 DNA Extraction

DNA extraction of the stool samples was done using the MagMidi LGC kit (LGC Biosearch Technologies, UK) on a KingFisher Flex Robot (Thermo Fisher Scientific, USA) following the manufacturer's recommendations. This kit uses paramagnetic particles that reversibly bind to DNA via a salt bridge at high salt concentrations (Boom et al., 1990). The samples were mixed with 96 % Ethanol and Mag Particles, followed by three wash steps with buffers containing

salts. DNA from the samples was conclusively eluted with nuclease-free water (VWR, USA), breaking the salt bridges between the Mag Particles and the DNA.

2.4 Library preparation of RMS amplicons

Preparation of the RMS library was done by cutting genomic DNA with the restriction enzymes EcoRI and MseI, ligating adaptors to the fragments, followed by two PCRs and a clean-up using Sera mag beads after both reactions.

2.4.1 Restriction Cutting

The restriction enzymes EcoRI and MseI recognize cutting sites on the genome that occur seldom (5’ G|GAATTC) and often (5’ T|TTAA), respectively. The restriction mix contained 8U EcoRI (New England Biolabs, USA), 4U MseI (New England Biolabs, USA), 1x Cut Smart buffer (New England Biolabs, USA), ~1 ng of extracted gDNA template (10 µL for all samples), and nuclease-free water with a total volume of 20 µL. The restriction mix was incubated at 37 ºC for 1 hour, allowing the enzymes to cut the genomic DNA into fragments. The fragments will either be flanked by EcoRI/EcoRI, EcoRI/MseI, MseI/EcoRI or MseI/MseI cutting sites.

2.4.2 Ligating of Adapter

Following the restriction cutting, there was done an adaptor ligation to make PCR amplification possible. The adaptor sequences contain a core sequence that is identical in the different adaptors, and a sequence-specific to four different flanked cutting sites, forward and reverse for both enzymes. The core sequence works as a common binding site for primers for all fragments. The genomic DNA product after cutting was mixed 1:5 with a ligation mix consisting of 0,5 µM EcoRI adapter mix (Invitrogen, USA), 5 µM MseI adapter mix (Invitrogen, USA), 400U T4 DNA ligase (New England Biolabs, USA), and 1x T4 DNA ligase reaction buffer (New England Biolabs, USA). The mix was made with equal volumes of forward (EcoRI; 5’- CTCGTAGACTGCGTACC-3’, MseI; 5’-GACGATGAGTCCTGAG-3’) and reverse (EcoRI; 5’AATTGGTACGCAGTCTAC-3’, MseI; 5’-TACTCAGGACTCAT-3’) adaptors. The ligation mixture was incubated at 37 ºC for 3 hours.

2.4.3 Polymerase chain reactions

The ligated fragments were then amplified in a qualitative PCR reaction with EcoRI forward and MseI reverse primer complementary to the adaptor and cutting sequence, allowing only the fragment cut by both enzymes (EcoRI/MseI) to be ligated. PCR amplification of the cut and

ligated fragments were done with mixing it 1:5 with 1x HotFirePol® DNA polymerase RTL (Solis BioDyne, Estonia), 0,2 µM EcoRI forward primer (5'-GACTGCGTACCAATTC-3'), 0,2 µM MseI reverse primer (5'GATGAGTCCTGAGTAA-3') and nuclease-free water. The PCR program consisted of a heating step at 95 ºC for 15 minutes followed by 25 cycles with denaturation at 95 ºC for 30 seconds, hybridization at 56 ºC for 1 minute, and elongation at 72ºC for 1 minute.

Index PCR was then done in order to later sequence the fragments on Illumina Hiseq 3000. The index primers have a complementary sequence at the 3' end to the adaptor and restriction cutting sequence of each fragment, a complementary sequence at the 5' end to the Illumina flow cell oligonucleotides, and an index sequence that provides traceability to all fragments originating from the same sample. 20 forward and 12 reverse primers were used in the Illumina HiSeq run. Appendix B gives an overview of the primers. The reaction mix contained 1x FirePol Master Mix Ready to Load (Solis BioDyne, Estonia), 0,2 µM EcoRI forward index primer (Invitrogen, USA), 0,2 µM MseI reverse index primer (Invitrogen, USA), with a unique combination for each sample, nuclease-free water, with 2 µL cleaned PCR product making the total volume 25 µL. The PCR program consisted of a heating step at 95 ºC for 5 minutes followed by 25 cycles with denaturation at 95 ºC for 30 seconds, hybridization at 56 ºC for 1 minute, and elongation at 72 ºC for 1 minute.

Both PCRs were done on a 2720 Thermal Cycler (Applied Biosystems, USA). Each 96-well plate had a PCR negative control with nuclease-free water, and positive control with extracted DNA from a culture of *Pseudomonas aeruginosa*, prepared by laboratory engineers at Norwegian University of Life Sciences.

2.4.4 Clean-up of PCR Product

The PCR products after both PCR reactions and the pooled library were cleaned for contaminates such as polymerases, primer dimers, and nucleotides by using SeraMag beads (Thermo Fisher Scientific, USA). The beads work paramagnetically and are suspended in a buffer containing salt. DNA binds to the beads and can, in this way, be separated from the mentioned contaminants. Each sample and pooled library were mixed thoroughly 1,5X with 0,1% SeraMag beads and after that placed on a magnetic rack. The supernatant was then removed, and the beads were washed twice with freshly made 80% ethanol (Antibac, Norway). Finally, the cleaned PCR product was eluted with water.

2.4.5 Normalization and pooling of RMS library

All samples were pooled together in a single library for both sequencing runs. DNA concentrations of each sample were used to determine the amount of each sample in the pooled sample, and the cut-off value for the libraries was 48,5 and 48,6 ng/µl. The maximum volume of each sample was 10 µL, while the minimum volume was 2 µL. Pooling of the libraries was conducted using a Biomek® 3000 robot (Beckman Coulter, USA).

2.5 DNA quantity and quality control

2.5.1 Qubit and Cambrex for quantification

The DNA was quantified using Qubit® dsDNA HS Assay Kit (Invitrogen, USA) on a Qubit™ fluorometer (Life Technologies, USA) and a Cambrex – FLX800 CSE machine (Thermo Fisher, USA) following the manufacture's recommendation. Quant-iT™ working solution was prepared by mixing Quant-iT™ reagent and Quant-iT™ buffer 1:200. The dye in the working solution emits fluorescence when bound to DNA. 2 µL of the sample were then added to 198µL working solution when measuring on the fluorometer, while 2 µL of the sample was added to 70 µL working solution in each well of a Nunc 96 well Nontreated Black Microwell plate (Thermo Fisher, USA) when measuring on the Cambrex machine. Quantification with Cambrex was done by measuring the fluorescence of each sample. DNA concentration from all samples was calculated by making a standard curve based on fluorescence data from ten samples with values ranging from the lowest to the highest and their corresponding DNA concentrations measured with the fluorometer. This Qubit-Cambrex coupled quantification was done after DNA extraction and index PCR, while the pooled libraries were only quantified with the fluorometer.

2.5.2 qPCR for quantification

Quantitative PCR (qPCR) was done after DNA extraction. For each reaction, 1x HOT FIREPol® EvaGreen® qPCR Supermix (Solis BioDyne, Estonia), 0,2 µM of forward (341F) and reverse (806R) primer (Yu, Lee, Kim, & Hwang, 2005), 1 µl extracted genomic DNA, and nuclease-free water was mixed in a total volume of 20 µl. qPCR was executed in LightCycler® 480 II (Roche, Germany) with the following program: Denaturation at 95 °C for 15 minutes followed by 40 cycles of denaturation at 95 °C for 30 seconds, annealing at 60 °C for 30 seconds and elongation at 72 °C for 45 seconds.

2.5.3 Agarose gel electrophoresis for qualification

Agarose gel electrophoresis was done as a qualitative checkpoint after RMS cutting and ligation, PCR reactions, and every clean-up step. The agarose network and electric voltage allow the separation of DNA fragments by size. DNA molecules have a negative charge, and the smallest molecules travel furthest to the positive pole. 1,5 % agarose gel was prepared by dissolving 1,5 % agarose (Invitrogen, USA) with 1x tris-acetate EDTA (TAE) buffer in a microwave. After cooling, 2 µL PeqGreen (PeqLab, Germany) per 50 µL agarose solution was added. PeqGreen is a fluorescent dye that binds to DNA and makes the DNA fragment detectable under UV light. The samples and positive controls were controlled for their expected size and the negative control for any contaminations. A 100 base pair (bp) DNA ladder (Soils BioDyne, Estonia) was used to determine the fragment size of all samples and controls. The gel ran for 35 min with a voltage of 80 V and ampere of 400 amp. The bands were visualized using Molecular Imager® Gel Doc™ XR Imaging Systems with Quantity One 1-D analysis software version 4.6.7 (BioRad, USA).

After PCR reactions, the samples contained a ready to load dye that simplifies application and visualizes the movement of fragments in the gel. The cleaned samples did not, and Loading Dye (Biolabs, Estonia) was mixed 1:5 with these samples. After every PCR reactions, all samples were checked on the gel. After every PCR clean-up, only 12 samples, including positive and negative controls, were checked on the gel.

2.6 Illumina sequencing

Illumina sequencing was done at NSC in Oslo, Norway. Two RMS libraries were submitted to NSC. The first contained 12 months samples, and the second contained samples from the remaining time points, mother, meconium, 3, and 6 months. Both libraries were sequenced on an Illumina HiSeq 3000 platform. NSC preformed quantification, sequencing, and quality control, before delivering FASTQ files and quality control report for each reverse primer.

2.7 Data analysis

2.7.1 Processing of sequence data in Kraken2

Taxonomic classification was done through a developing RMS Kraken2 pipeline made by Snipen, L.G., and described by Lokmic, A, in her unpublished master thesis. (Lokmic, 2019). The first step of the pipeline involves demultiplexing, which utilizes the barcodes in the index sequences to separate the sequences from different samples into separate files. Demultiplexing

produces a FASTQ file pair for each sample. The file pair was then run through the Kraken2 HumGut_05 database for taxonomic classification (Hiseni et al., 2020), both at species and genus level. The HumGut_05 database allows a 95% identity level between reference k-mers and k-mers in the sequences from a sample.

The next step of the pipeline is a correction step. All fragments originating from the same specie and genus were normalized to the total number of available restriction-cut fragments in the genomes of members to that specie and genus. The correction is necessary because every unique genome has a specific amount of possible RMS fragments due to the genome length and number of cutting sites. The signal can be either over- or underestimated if this difference is not considered.

2.7.2 Inferring the metabolic potential by the VMH database

The taxonomic composition at the species level for all time points (meconium, 3, 6, and 12 months and mothers) was matched to the VMH database by Professor Knut Rudi. Each sample was scored based on the inferred metabolic potential of all species in that sample.

2.7.3 Normalization

Normalization in every sample was done to the taxonomic data by dividing the number of sequences for every specie and genus in one sample on the complete number sequences in the sample. The same normalization was done to the inferred metabolic data, namely by dividing the metabolite scores present in one sample on the complete metabolite score in that sample. This number was multiplied by 100%. The specie, genus, and inferred metabolite was removed if its relative abundance summed to zero for all samples.

2.7.4 Paired T-test

The relative longitudinal taxonomic and inferred metabolic development were studied by calculating the average relative percentage of each genus and metabolite, respectively. A paired T-test was used to evaluate significant changes in the relative composition of genus or metabolite between age groups with a relative abundance of ≥5% in at least one age group. The level of significance was set to 5%. The script was made by master student Unni Lise Albertsdóttir Jonsmoen and PhD student Morten Nilsen and was performed by in R studio version 3.5.2 and package stats by the master student.

2.7.5 Spearman correlation

Spearman correlation tests (Spearman, 1904) were used to investigate possible associations between infant and mother gut's relative specie, genus, and inferred metabolite composition the first year of life to immune cell composition at 12 months. The Spearman correlation test produces a rho value ranging from -1 to 1 for each correlation. A positive rho value means a positive correlation, while a negative rho value means a negative correlation between the variables. A positive correlation means that the variables move in tandem, one variable decrease if the other variable decrease or one variable increase if the other variable increase. On the contrary, a negative correlation means that one variable decrease if the other variable increase, and vice versa. An adjustment for the false discovery rate was done using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995) for each immune cell, and the level of significance was set to 5%. The script was made by PhD Student Morten Nilsen and performed in R studio version 3.5.2 and package stats by the master student.

3. Results

3.1 RMS library preparation and sequencing

3.1.1 RMS library preparation

Quantification by qPCR after DNA extraction on all samples resulted in C_q values ranging from 13-32. The positive and negative controls had an average C_q value of 15 and 36, respectively. One mother sample seemingly did not contain fecal material, as the original sample was white, and the extracted sample produced no detectable signal in qPCR. DNA concentrations for all samples were measured to ~ 0,1 ng/µl.

The result after two qualitatively checkpoints with agarose gel electrophoresis for 14 representative samples from 3 months is visualized in Figure 3.1. Gel picture A) is after cutting, ligating, and the first PCR, and gel picture B) is after index PCR. The gel result after cutting, ligating, and the first PCR A) shows that most of the bands vary in fragment size from 100-500 bp for all samples and positive control, while the negative control shows no visual bands. 38 of the 60 meconium samples (63%) did not give visual bands on the gel after cutting, ligating, and the first PCR. The initial PCR reaction was increased with five cycles to a total of 30 cycles for these samples. This adjustment gave clear bands on the gel on all samples, like panel A) in the figure.

The gel result after index PCR B) shows that most of the bands vary in fragments size from 300-500 bp for all samples and positive control, while the negative control does not have fragments in this size area. The range of fragments is reduced after index PCR compared to the first PCR. All samples and both controls have DNA fragments with a size ~ 100 bp, indicating index primer dimer formation.

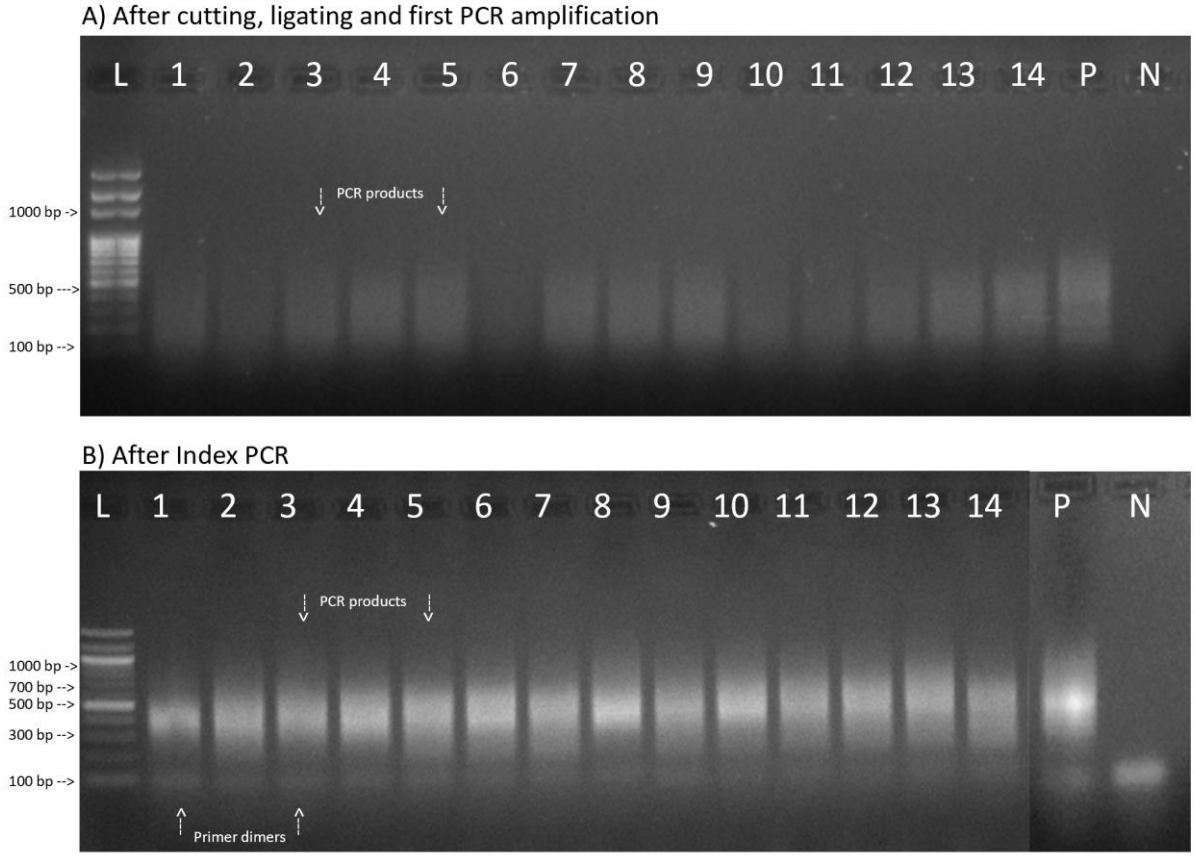

Figure 3.1. **Qualitative check by agarose gel electrophoresis.** The figure shows gel images of PCR products from 14 samples at 3 months (1-14), *P. aeruginosa* as the positive control (P), and nuclease-free water as the negative control (N). The 100 bp ladder (L) was used to determine band size for all fragments. Panel A) is after cutting, ligating, and first PCR amplification, while panel B) is after Index PCR. Primer dimer is observed in B) at 100 bp. The PCR fragments were smeared, and the majority fragment size is 100bp-500 bp in A) and 300-500 bp in B).

3.1.2 RMS sequencing

The first Illumina HiSeq 3000 sequencing run (2x150 bp) of 180 12 months samples produced a total of 234 823 012 sequences for the reverse primers. The read count number in each sample varied from 39-150 012 at the genus level and 51-222 199 at the specie level. The second Illumina HiSeq 3000 sequencing run (2x150 bp) with 59 mothers, 60 meconium, 59 3 months, and 59 6 months samples produced a total of 356 882 086 sequences for the reverse primers. The read count number in each sample varied from 36-12 347 at the genus level and 58-8 878 at the specie level. One reverse file was corrupt and could not be read. The file issue resulted in a reduction to 51 3- and 6-month samples. NSC provided a quality report of the sequences for the reverse primers in each run, and appendix C shows two representatives per base sequence quality plots for all reverse primers. Panel A is for reverse 7 in the sequencing run with samples from 12 months, and panel B is from reverse 2 at the sequence run with samples from mothers,

meconium, 3, and 6 months in Figure C.1 (Appendix C). Both plots display decreasing quality scores in bases from position 60-70 and further towards the end of the read.

3.2 Longitudinal development of the gut microbiota

Determination of the gut microbiota's relative longitudinal taxonomic and inferred metabolic development was done by calculating the average of each genus and metabolite in all samples at each time point. The mother group was set as an adult reference group for both relative taxonomic and inferred metabolic composition. The longitudinal taxonomic development was not analyzed at the species level due to the high total number of 996 species for all time points and because the inferred metabolic composition is based on this species composition.

Paired t-test conducted determination of significant differences in averages of the relative taxonomic and inferred metabolic composition from meconium to 3 months, 3 to 6 months, 6 to 12 months, and 12 months to mothers. Only genera and metabolites with relative abundance ≥5% for at least one time point were checked for significant change, and the level of significance was set to 5%.

3.2.1 Longitudinal taxonomic development at the genus level

The gut microbiotas longitudinal relative taxonomic development at the genus level in the first year of life is illustrated in the bar graph in Figure 3.2 and listed more in detail in Table D.1 (appendix D). All age groups had a total of 65 bacterial genera in common. Nine genera showed ≥5% relative abundance in at least one of the time points and are listed in the figure. The remaining 56 genera are summed and visualized as other genera in the figure. The statistically significant changes in relative abundance between the time points are marked with * in the figure, and the exact p values are listed in Table D.2 (appendix D).

The results from this study found that *Bifidobacterium* was the only genus that showed a significant change in all time point transitions, an increase from meconium to 3 months, and a decrease in relative abundance from 3 to 6 months, 6 to 12 months and 12 months to mothers. Of the remaining eight genera, *Prevotella*, *Allistipes*, *Faecalibacterium*, and *Bacteroides* all significantly decreased in relative abundance from meconium to 3 months. *Sutterella*, *Veillonella*, and *Bacteroides* significantly increased, while *Streptococcus* significantly decreased in relative abundance from 3 to 6 months. *Prevotella*, *Allistipes*, *Faecalibacterium*, *Sutterella*, and *Bacteroides* all significantly increased in relative abundance from 6 to 12 months. *Allistipes*, *Faecalibacterium*, and *Clostridium* significantly increased, while

Veillonella, *Sutterella*, and *Streptococcus* significantly decreased in relative abundance from 12 months to mothers.

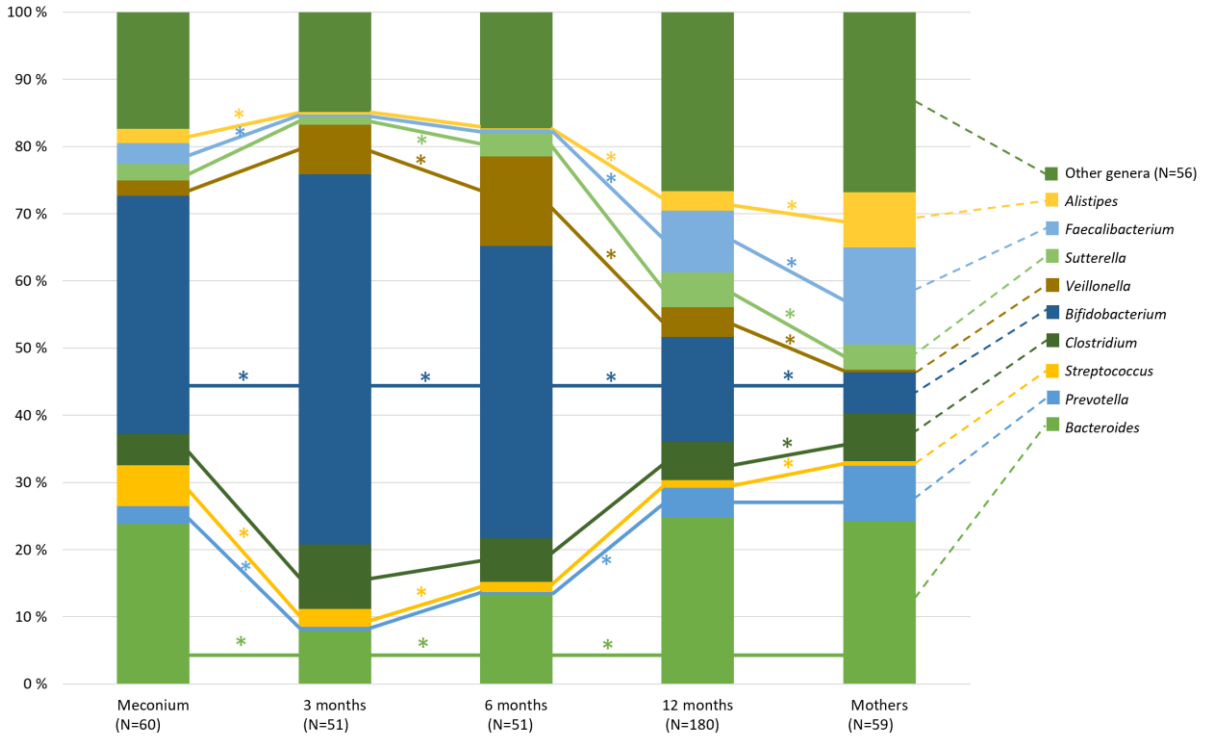

Figure 3.2. **The longitudinal taxonomic development.** The bar graph shows the relative longitudinal taxonomic development for the infants' dominant genera in the first year of life and their mothers. The complete relative abundance data for all genera at all time points are listed in Table D.1 (appendix D). Paired T-tests were done to determine significant differences in relative genus abundance from meconium to 3 months, 3 to 6 months, 6 to 12 months, and 12 months to mothers. Significant changes with a 5% level in the relative abundance of a genus are marked with *, and the exact p-values are listed in Table D.2 (appendix D).

3.2.2 Inferred longitudinal metabolic development from the species level

The gut microbiota's inferred longitudinal metabolic development the first year of life is illustrated in the bar graph in Figure 3.3 and listed in Table E.1 (appendix E). Samples from all age groups were scored to a total of 17 metabolic compounds using VMH on the specie composition data. Nine metabolites showed ≥5% relative abundance in at least one time point and are listed in the figure. The remaining seven metabolites are summed and visualized as other metabolites in the figure. A statistically significant change in relative abundance between two time points is marked with * in the figure, while the exact p-values are listed in Table E.2 (appendix E).

The results from this study find acetic acid as the only metabolite with no significant change in inferred relative abundance in any time point transition. Only formic acid showed significant

changes in the inferred relative abundance in all time point transitions, an increase from meconium to 3 months and 12 months to mother, and a decrease from 3 to 6 months. Of the remaining seven metabolites, succinic acid, and ethanol significantly increased while D-lactic acid and butyric acid significantly decreased in inferred relative abundance from meconium to 3 months. Propionic acid significantly increased, while formic acid, ethanol, and L-lactic acid significantly decreased in inferred relative abundance from 3 to 6 months. Propionic acid, D-lactic acid, butyric acid, and H_2 significantly increased, while formic acid, succinic acid, ethanol, and L-lactic acid significantly decreased in inferred relative abundance from 6 to 12 months. Formic acid, D-lactic acid, and butyric acid significantly increased, while propionic acid, succinic acid, ethanol, H_2, and L-lactic acid significantly decreased in inferred relative abundance from 12 months to mothers.

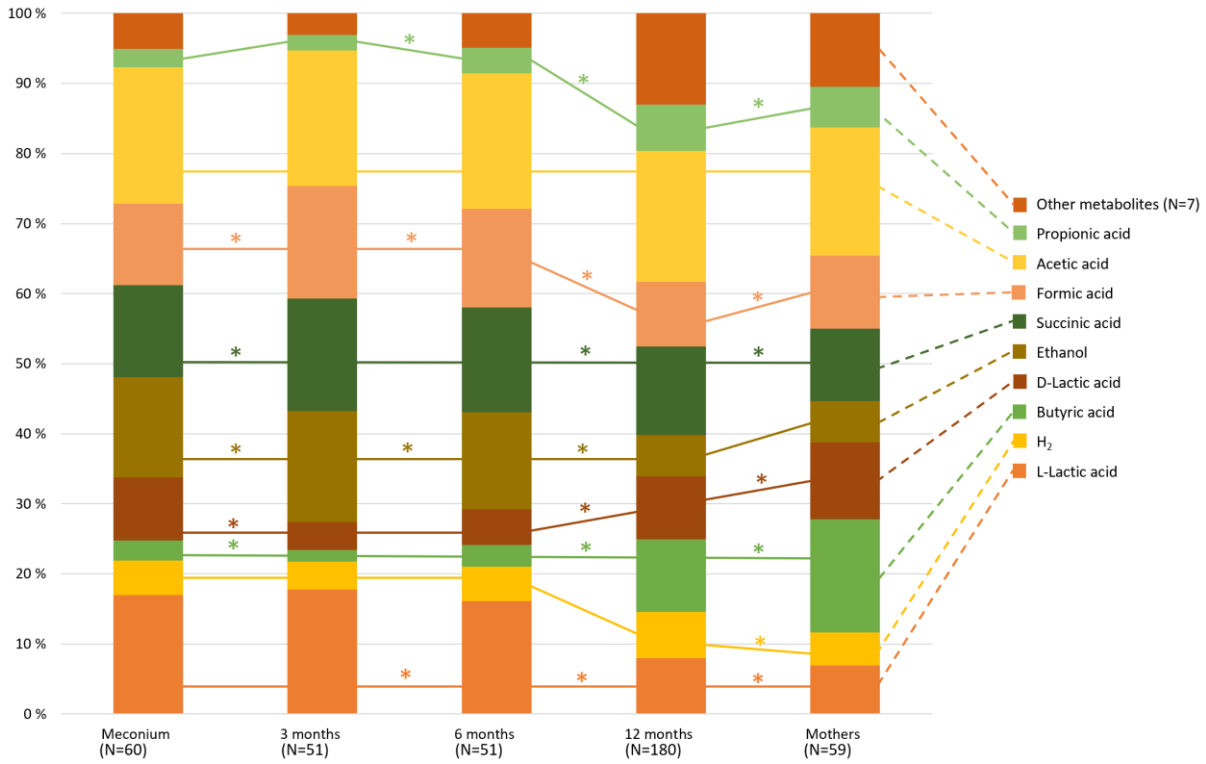

Figure 3.3. **The inferred longitudinal metabolic development.** The bar graph shows the relative inferred longitudinal metabolic development for the infants' dominant metabolites in the first year of life and their mothers. The complete relative abundance data for all metabolites at all time points are listed in Table E.1 (appendix E). Paired T-tests were done to determine the significant difference in relative inferred metabolic abundance from meconium to 3 months, 3 to 6 months, 6 to 12 months, and 12 months to mothers. Significant changes with a 5% level in the relative abundance of a metabolite are marked with *, and the exact p-values are listed in Table E.2 (appendix E).

3.3 Correlation of gut microbial characteristics and immune cells

Correlation between relative immune cell abundance at 12 months to the relative taxonomic abundance at the genus and species level as well as the relative inferred metabolic abundance for at children the first year of life and their mothers was done using the Spearman method. Adjustment for false discovery rate was conducted by the Benjamini-Hochberg procedure. Correlation tests were done for individuals with relative immune cell composition at 12 months and their corresponding mother. The total number of correlation tests were distributed in the following matter: 60 for meconium to 12 months, 51 for 3 to 12 months, 51 for 6 to 12 months, 67 for 12 to 12 months, and 59 for mothers to 12 months. Table F.1 (appendix F) gives an overview of the relative abundance of the 28 different immune cells detected in the 67 children's blood samples at 12 months.

3.3.1 Correlation in relative taxonomic and immune cell abundance

After adjusting for false discovery in this study, the relative abundance of ten species was significantly correlated to the relative abundance of five immune cells. Figure 3.4 visualizes the correlation analysis results, while Figure G.1 (appendix G) lists the associated P and Rho values. Correlating species to immune cells at 12 months originate from all time points except meconium. In the mother group, the relative abundance of *Bifidobacteriaceae bacterium* NR020, *Emergencia timonesis*, and *Clostridium* sp. CAG 508 correlated negatively to the relative abundance of classical monocytes at 12 months. The relative abundance of *Ruminococcaceae bacterium* UBA2 in the mother group correlated negatively to the relative abundance of double-positive T cells, while the relative abundance of *Corpococcus* sp AF21 14LB in the mother group correlated negatively to the relative abundance of translational B cells. The relative abundance of *Streptococcus mutans* at 12 months correlated negatively to the relative abundance of proinflammatory monocytes at 12 months. The relative abundance of *Lactobacillus gasseri* and *uncultured Synergistaceae bacterium* at 6 and 3 months positively correlated to the relative abundance of naïve $CD8^+$ T cells at 12 months. Additionally, the relative abundance of *Firmicutes bacterium* CAG 176 and *Clostridium* sp. CAG 226 at 6 months also correlated positively to the relative abundance of naïve CD8 T cells at 12 months.

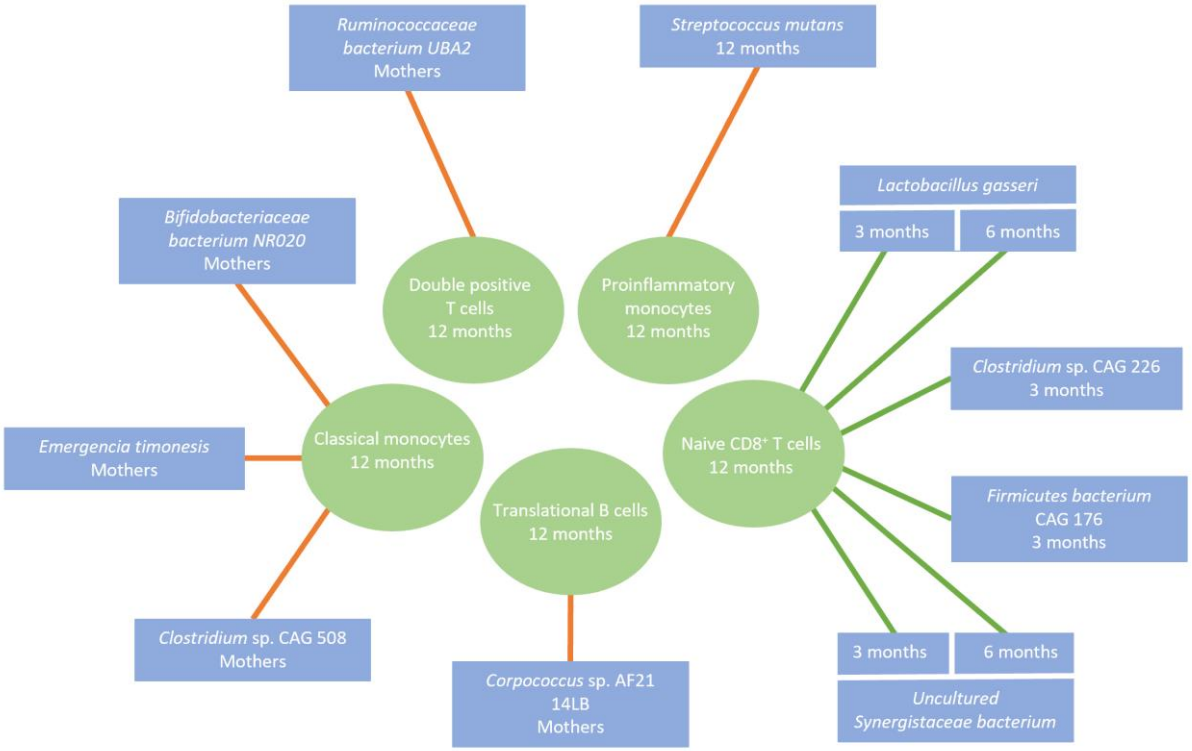

Figure 3.4. **Significant correlation of relative specie and immune cell abundance.** The network plot visualizes the significant correlations between the relative specie and immune cell abundance in this study. Green circles represent the immune cells at 12 months, while blue squares represent the specie and associated time point. Green lines illustrate a positive correlation, while red lines illustrate a negative correlation. Figure G.1 (appendix G) lists the associated P and Rho values to each correlation.

After adjusting for false discovery in this study, the relative abundance of 13 genera was significantly correlated to the relative abundance of three immune cells. Figure 3.5 visualizes the correlation analysis results, while Figure G.2 (appendix G) lists the associated P and Rho values. Correlating genera to immune cells at 12 months are from samples at meconium and 3 months. The relative abundance of *Eggerthella* and *Phascolarctobacterium* at 3 months correlated positively to the relative abundance of central memory $CD8^+$ and $CD4^+$ T cells, respectively. The relative abundance of the two same genera correlated negatively to the relative abundance of nonclassical monocytes at 12 months, as compared to the relative abundance of 11 remaining genera; *Flaconifractor*, *Parasutterella*, *Oscillibacter*, *Gemminger*, *Dorea*, *Holdemania*, *Eubacterium*, *Lactobacillus*, *Clostridium*, *Ruminococcus*, and *Acidaminococcus*.

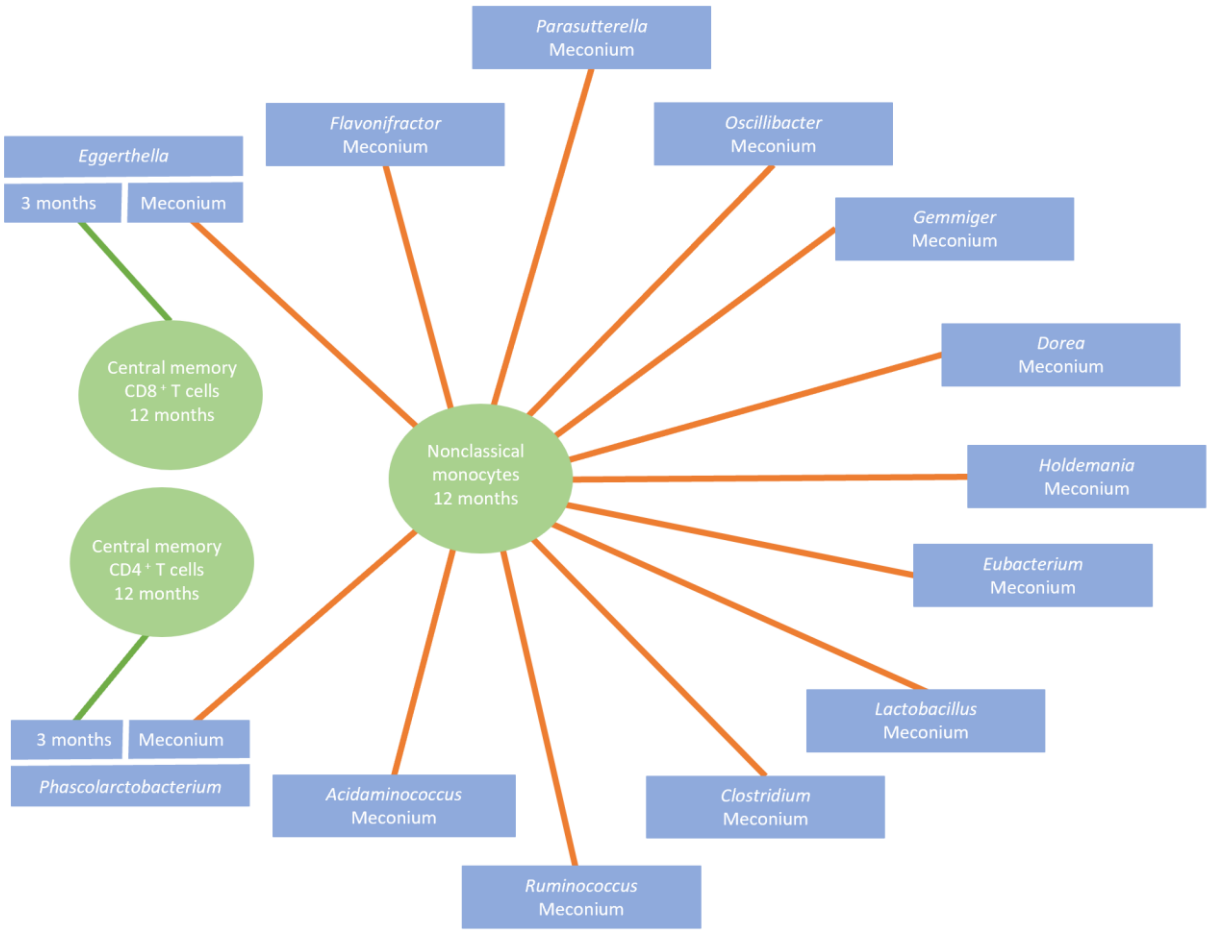

Figure 3.5**. Significant correlation of relative genus and immune cell abundance.** The network plot visualizes the significant correlations between the relative genus and immune cell abundance in this study. Green circles represent the immune cells at 12 months, while blue squares represent the genus and associated time point. Green lines illustrate a positive correlation, while red lines illustrate a negative correlation. Figure G.2 (appendix G) lists the associated P and Rho values to each correlation.

3.3.2 Correlation in relative inferred metabolic and immune cell abundance

After adjusting for false discovery in this study, the relative abundance of seven inferred metabolites was significantly correlated to the relative abundance of four immune cells. Figure 3.6 visualizes the correlation analysis results, while Figure G.3 (appendix G) lists the associated P and Rho values. Correlating metabolites are from mothers, 3, and 12 months samples. In the mother samples, the relative inferred abundance of indole, isovaleric acid, isobutyric acid, and phenylacetic acid negatively correlated to the relative abundance of proinflammatory monocytes at 12 months. The relative inferred abundance of (2R,3R)-2,3-Butanediol from the mother group correlated negatively to the relative abundance of naive CD8 T cells. The relative inferred abundance of butyrate at 12 months correlated negatively to the relative abundance of $CD56^{bright}$ Natural Killer (NK) cells at 12 months, while the relative inferred abundance of CH_4

at 3 months correlated positively to the relative abundance of mucosal-associated invariant T cells at 12 months.

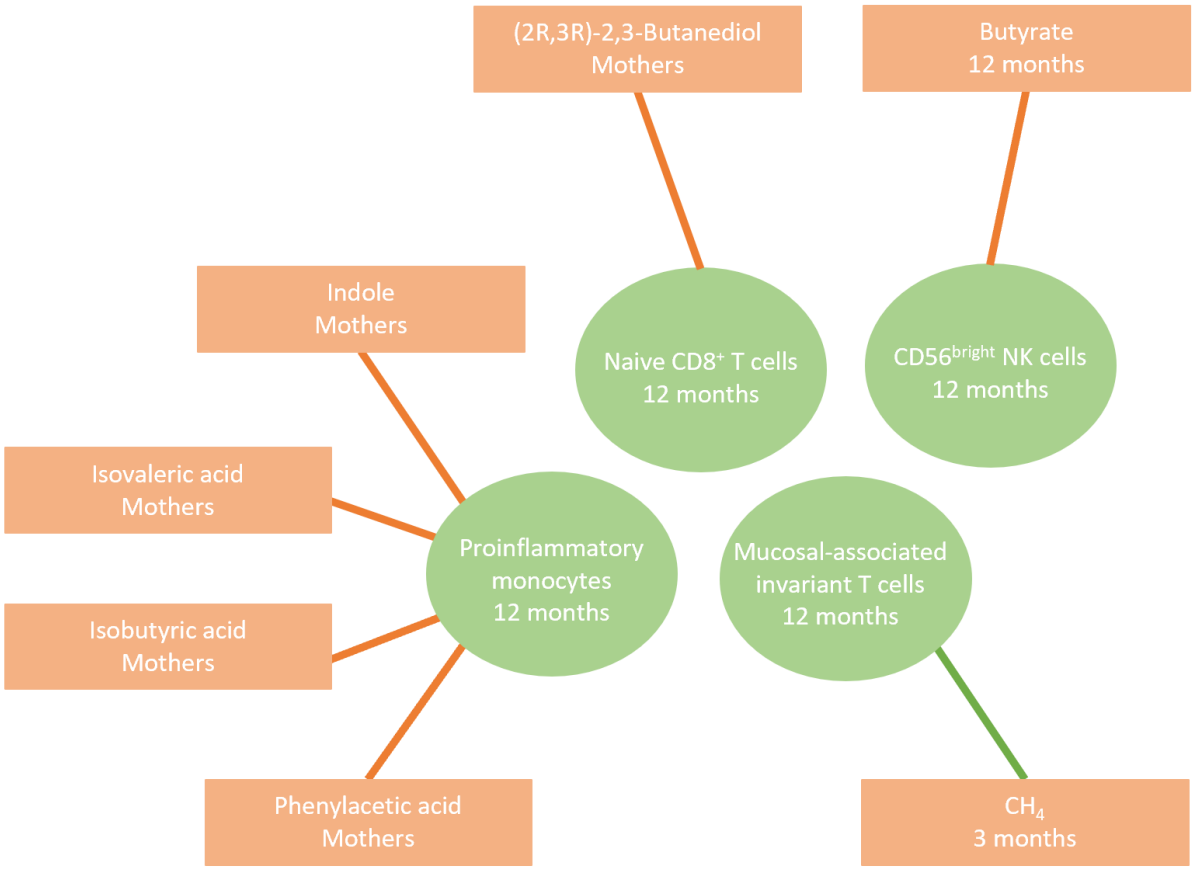

Figure 3.6. **Significant correlation of relative inferred metabolites and immune cell abundance.** The network plot visualizes the significant correlations between the relative inferred metabolite and immune cell abundance in this study. Green circles represent the immune cells at 12 months, while orange squares represent the inferred metabolite and associated time point. Green lines illustrate a positive correlation, while red lines illustrate a negative correlation. Figure G.3 (appendix G) lists the associated P and Rho values to each correlation.

4. Discussion

4.1 Correlation in gut microbial characteristics to immune cells

This study found several associations between the relative abundance of gut microbial taxonomic and inferred metabolic characteristics in the first year of life and the relative abundance of immune cells at 12 months. Most interestingly, the relative abundance of three different subtypes of monocytes showed a strong negative association with both the gut microbiota's relative taxonomic and inferred metabolic composition. These negative correlations and other intriguing findings of the current study are discussed in the following subchapters. In contrast, immune cell correlations to species that are not yet described in the literature and inferred metabolites with no previous association to the immune system are not discussed.

4.1.1 Negative correlation of gut microbial characteristics to monocyte subsets

Monocytes are innate immune cells circulating in the blood and tissue before they differentiate to tissue-resident macrophages or dendritic cells. Monocytes are APCs that can phagocytize and present foreign antigens to adaptive immune cells, secrete chemokines, and proliferate in response to infection and injury (Chiu & Bharat, 2016). Monocytosis is the state of elevated monocyte level and is a result of acute and chronic inflammation (Dutta & Nahrendorf, 2014) that may originate from a viral, bacterial, or parasitic infection. Monocytes divide into different subsets based on the expression of different surface molecules.

A striking observation in the results of the current study is the negative correlation of the relative abundance of 13 different genera in meconium samples to the relative abundance of nonclassical monocytes at 12 months (Figure 3.5). Nonclassical monocytes are a monocyte subtype that is widely accepted as anti-inflammatory because they maintain vascular homeostasis and provide the first line of defense in recognition and clearance of pathogens and cancer cells (Narasimhan, Marcovecchio, Hamers, & Hedrick, 2019). However, several studies reveal that the nonclassical monocyte subtype contributes to the disease pathogenesis by also possessing proinflammatory characteristics, which challenge the prior agreement (Narasimhan et al., 2019). The high number of negative correlating genera in meconium samples may indicate that the relative abundance of these genera similarly affects the relative abundance of nonclassical monocytes. However, the genera are not known to have a common gut microbe-host association. Specific species belonging to the genera *Eggerthella* (Gardiner et al., 2015) and *Ruminococcus* (Henke et al., 2019) have adverse effects on their host. On the contrary,

specific species belonging to the genera *Phascolarctobacterium* (F. Wu et al., 2017), *Eubacterium* (Engels, Ruscheweyh, Beerenwinkel, Lacroix, & Schwab, 2016), *Parasutterella* (Ju, Kong, Stothard, & Willing, 2019), *Oscillibacter* (Gophna, Konikoff, & Nielsen, 2017), *Lactobacillus* (Walter, 2008), and *Clostridium* (Lopetuso, Scaldaferri, Petito, & Gasbarrini, 2013) have positive effects on their host. The genera *Flavonifractor*, *Acidaminococcus*, *Gemmiger*, *Dorea*, and *Holdemania*, are, on the contrary, not entirely determined in terms of their relationship to their host. Furthermore, species belonging to the same genus can differ hugely in their associations to the human host. Very little was found in the literature on the question of the association between gut microbes and nonclassical monocytes in general. Regardless, the combination of findings is highly exciting.

Another interesting finding in the present study is the negative correlation between the relative abundance of *S. mutans* at 12 months and the relative abundance of proinflammatory monocytes at 12 months (Figure 3.4). Proinflammatory monocytes are another monocyte subtype, which, as the name suggests, has proinflammatory properties. Levels of proinflammatory monocytes elevate during pathologic conditions, such as inflammatory and infectious diseases, but the mechanism of the increase is still unclear (Jeng et al., 2017) (Ritz et al., 2011) (Koch, Kucharzik, Heidemann, Nusrat, & Luegering, 2010). *S. mutans* is a member of the oral microbiota known to cause dental plaque formation. The oral cavity is an entry site to the GI tract, and studies indicate that oral bacteria possibly translocate to the gut, change the gut microbiota, and thereby affect the immune system in humans (Atarashi et al., 2017). In a murine model, intravenous administration of specific strains of *S. mutans* caused aggravation of ulcerative colitis (Kojima et al., 2012). However, a positive correlation is expected if this relationship is transferrable to humans rather than the observed negative correlation. Furthermore, it is challenging to determine if this finding applies to humans with the current study results, considering that the abundance of *S. mutans* was detected from tissue samples and not fecal samples.

The relative abundance of proinflammatory monocytes at 12 months is also negatively correlated to the relative abundance of four inferred metabolites in mother samples in the current study, namely indole, isovaleric, isobutyric, and phenylacetic acid (Figure 3.6). The mother's metabolic composition of the gut microbiota may influence the child's immune cell composition at 12 months of age through that the mother's gut microbiota initially shapes the mother's immune system. Antibodies from the mother are then transferred to the fetus through the placenta and to the infant through breast milk and can consequently affect the immune

system (W. Zheng et al., 2020) (Fouda, Martinez, Swamy, & Permar, 2018). Indole derivatives of tryptophan catabolism by bacterial species act directly on cells of the immune system and demonstrate beneficial effects by protecting against liver disease development in humans (Hendrikx & Schnabl, 2019). Phenylacetic acid is an antibiotic product of phenylalanine metabolism that is experimentally showed to have immunostimulatory properties by inducing expression of immune-related genes (Z.-F. Wu et al., 2012). In contrast, an association with the immune system is not found for isovaleric and isobutyric acid in earlier studies. A possible explanation for the negative correlation is that an increased relative abundance of the inferred metabolites in the mother's gut might induce anti-inflammatory changes in the mother's immune system that are transferred to the infant. Ultimately, this leads to a reduction of the relative abundance of proinflammatory monocytes in the infant.

The relative abundance of three species from mother samples, *Bifidobacteriacea bacterium* NR020, *Emergencia timonensis*, *Clostridium* sp. CAG 508, was intriguingly found negatively correlated to the relative abundance of classical monocytes at 12 months (Figure 3.4). Classical monocytes are the third type of monocyte, which are the most prevalently abundant and hold the highest phagocytic capacity among all three subtypes. As the two others, classical monocytes similarly exhibit proinflammatory properties (Sampath, Moideen, Ranganathan, & Bethunaickan, 2018). The species have not yet shown an association with the immune system. However, a similar route from mother to infant, as earlier proposed, may explain this observed negative correlation. In this case, however, both microbial derivatives from the species and their metabolites may impact the mother's and, ultimately, the infant's immune system in an anti-inflammatory manner.

4.1.2 Positive correlation of species to naïve $CD8^+$ T cells

An interesting finding in the present study was the positive correlation between the relative abundance of *L. gasseri* at 3 and 6 months to the relative abundance of naïve $CD8^+$ T cells at 12 months (Figure 3.4). Naïve $CD8^+$ T cells are immune cells that have undergone the positive and negative central selection in the thymus. They circulate the lymph until they are driven to proliferate and differentiate to effector T cells that are able to recognize and eliminate cancerous cells and intracellular pathogens (Pennock et al., 2013). *L. gasseri* is a commensal lactic acid-producing bacterium that occupies mucosal niches of humans. The specie is known to be among the first colonizers of the GI tract and stay persistent throughout adulthood as it tolerates the low pH and bile salts and can successfully adhere to the host epithelium cells (Wall et al., 2007).

L. gasseri obtains probiotic properties in humans due to its antimicrobial activity, bacteriocin production, and the ability to modulate immune responses (Selle & Klaenhammer, 2013). Immunomodulation by probiotic microorganisms occurs through the interaction of microorganism-associated molecular patterns with pattern recognition receptors on APCs and regulates both systemic and mucosal immunity (Lebeer, Vanderleyden, & De Keersmaecker, 2010). The molecular patterns of *L. gasseri* include lipoteichoic acids, glycolipids, and peptidoglycan (Selle & Klaenhammer, 2013) that interact and primarily activate toll-like receptor 2/6 (Stoeker et al., 2011). Murine models have investigated the immune response for the OLL2809 strains of *L. gasseri,* showing that both the bacterium and its microbial RNA had suppressive effects on inflammatory responses by suppressing $CD4^+$ T cell proliferation (Yoshida et al., 2011).

The increased colonization of pathogens might explain the detected positive correlation in infants with a low relative abundance of the commensal bacterium *L. gasseri*. The mucosal immune system detects pathogenic colonization in the gut lumen and induces proliferation and differentiation from naïve cell types to mature effector cell types. The relative abundance of naïve CD8+ T cells may consequently decrease. Conversely, infants with a high relative abundance of *L. gasseri* may have lower levels of pathogenic colonization. The naïve immune cells do not proliferate and differentiate in the same degree, and the relative abundance of naïve CD8+ T cells is consequently high. That a previous study discovered that T cells in pathogen-free mice remain in their naïve state despite colonization by commensal microbes in the GI tract (Belkaid, Bouladoux, & Hand, 2013) strengthens this hypothesis.

The relative abundance of Uncultured *S. bacterium* showed identical correlation patterns at the same time points to the relative abundance of naïve $CD8^+$ T cells at 12 months as *L. gasseri* in the current study (Figure 3.4). This bacterium belongs to the newly proposed phylum Synergistetes (Jumas-Bilak, Roudière, & Marchandin, 2009) and is a part of the human gut microbiota (Duysburgh, Van den Abbeele, Krishnan, Bayne, & Marzorati, 2019). The specie is not intensively studied, and very little was found in the literature on the question of the relationship to its human host. *C.* sp. CAG 176 and *F. bacterium* CAG 176 also show the same positive correlation to naïve $CD8^+$ T cells as the two mentioned species, but only at 3 months. These Co-Abundant Gene groups are assembled as individual genomes based on the clustering of co-abundant genes. Little previous research is consequently available on these species, and no previous association is discovered to the immune system. The possible relationship between

uncultured *S. bacterium*, *C.* sp. CAG 176, and *F. bacterium* CAG 176 to naïve $CD8^+$ T cells may correspond to what is proposed for *L. gasseri*. If this assumption is correct, these species possibly hold similar probiotic properties to what *L. gasseri* has. To validate this hypothesis, a better understanding of the species as members of the gut microbiota is most importantly needed.

4.1.3 Positive correlation of genera to central memory T cells

Interestingly, this study found the relative abundance of *Eggerthella* at 3 months to positively correlate to the relative abundance of central memory $CD8^+$ T cells at 12 months. Furthermore, the study also found the relative abundance of *Phascolarctobacterium* at 3 months to positively correlate to the relative abundance of central memory $CD4^+$ T cells at 12 months (Figure 3.5). Activation by antigen stimulation causes both some $CD8^+$ and $CD4^+$ T cells to develop into memory cells. Central memory T cells, a subtype of the memory T cells, mediates the adaptive immune system's reactive memory. Compared with naïve T cells, the central memory T cells have a higher sensitivity to antigenic stimulation and provide more effective stimulatory feedback to other immune cells and differentiate to effector cells (McKinstry, Strutt, & Swain, 2010) (Martin & Badovinac, 2018).

Species of the genus *Eggerthella* are commonly known members of healthy gut microbiota (Cho et al., 2016). However, members of the genus are not profoundly characterized except *Eggerthella lenta*, due to its pathogenicity (Gardiner et al., 2015) and its ability to inactivate cardiac drug digoxin (Haiser et al., 2013). Moreover, *E. lenta* is not one of the detected species from *Eggerthella* in the current study, and very little was found in the literature of the detected species. A prior study by Tanoue et al. noted that the development of memory $CD8^+$ T cells is associated with gut microbiota. The researchers colonized mice with 11 bacterial strains isolated from healthy human donor feces and detected induction of colonic $IFN\gamma^+$ $CD8^+$ T cells to memory $CD8^+$ T cells (Tanoue et al., 2019). None of the 11 strains belonged to *Eggerthella*, and no relationship between *Eggerthella* and central memory $CD8^+$ T cells are reported otherwhere in the literature. Supposing that species of *Eggerthella* affects the levels of memory $CD8^+$ T cells correspondingly as the 11 strains, the positive correlation may be explained by that an increase in the relative abundance of members of the genus leads to an increase in the relative abundance of the memory $CD8^+$ T cells and vice versa.

Species belonging to *Phascolarctobacterium* are abundant gut bacteria that utilize succinate to produce SCFAs acetate and propionate (F. Wu et al., 2017). A direct association between

Phascolarctobacterium and central memory $CD4^+$ T cells is not yet reported in the literature. However, an earlier study has discovered a reduced abundance of *Phascolarctobacterium* in patients with immune-related inflammatory bowel diseases (Morgan et al., 2012), indicating that the absence of species from the genus negatively affects the immune system. The presence of species from the genus may conversely have a positive effect on the immune system. The same relationship may exist and explain the detected correlation in the current study. The high relative abundance of *Phascolarctobacterium* affects the immune system in an anti-inflammatory way, ultimately leading to a high relative abundance of central memory $CD4^+$ T cells.

4.1.4 Negative correlation of inferred butyric acid to $CD56^{bright}$ NK cells

A surprising finding of the current study is the negative correlation between the inferred relative abundance of butyric acid at 12 months and the relative abundance of $CD56^{bright}$ NK cells at 12 months (Figure 3.6). NK cells are involved in both innate and adaptive immunity and contribute to both maintenances of immune homeostasis and the development of efficient immune responses (Poggi et al., 2019). The subtype $CD56^{bright}$ NK cell type is immunoregulatory and less cytotoxic than the $CD56^{dim}$ NK cell type (Cooper, Fehniger, & Caligiuri, 2001). The SCFA butyric acid is known to modulate human immunity, and the mechanism of immune modulation is widely studied for macrophages (Schulthess et al., 2019), Treg cells (Arpaia et al., 2013), and neutrophils (Vinolo et al., 2011). No association is earlier reported in the literature for butyric acid and $CD56^{bright}$ NK cells. However, a high abundance of butyric acid has been associated with decreased inflammation (Segain et al., 2000), while NK cells are stimulated and induced to proliferate by monocyte-derived cytokines upon infection (Assarsson et al., 2004) (Cooper, Fehniger, Turner, et al., 2001). That high levels relative abundance of butyric acid is confounding with low levels of $CD56^{bright}$ NK cells, and vice versa may be explained by butyric acid's ability to regulate bacterial pathogenesis as reviewed in Sun and O'Riordan in 2013 (Sun & O'Riordan, 2013).

4.2 General aspects related to gut microbiota development in infants

Determining the longitudinal relative taxonomic and inferred metabolic development was two of the sub-goals conducted to explore the main goal of this study. Most of the observed patterns confirmed that of previous findings. The present study also revealed newer patterns concerning the clear taxonomic and minor metabolic development and transmission of genera from mother to infant. These exciting findings are discussed in the following subchapters.

4.2.1 The clear taxonomic and minor metabolic development

A clear taxonomic and minor metabolic development in the gut microbiota, the first year of life, appears when comparing the taxonomic composition with the inferred metabolic composition in the current study. The significant change in the relative abundance of genera such as *Bifidobacterium*, *Bacteroides*, *Faecalibacterium*, *Alistipes*, and *Veillonella* at different time points causes a visible shift in the total relative taxonomic composition (Figure 3.2). This finding broadly supports the work of other studies in this area, linking a change in the taxonomic composition of the gut microbiota with age (Odamaki et al., 2016) (Tsuji et al., 2012).

Even though several inferred metabolites significantly differ in relative abundance between the age points, the inferred longitudinal metabolic development is visually more uniform in every time point than the longitudinal taxonomic development in the current study (Figure 3.3) . The abundance of species with inferred metabolic potential to produce the metabolites is somewhat consistent over time. These findings correspond to the conclusions of other studies investigating the microbiome on a functional level using metabolic reconstruction. The studies have detected that gut microbes share the genes of the common pool and thereby yield a core microbiome at a functional level rather than the organismal level in adults of different ages (Turnbaugh et al., 2009) (Qin et al., 2010). The significant changes in relative abundance found in this study can be explained by the sampling period. The gut develops from an aerobic to an anaerobic environment (Friedman et al., 2018), and the infant diet change from exclusively being breast- or formula-fed to be partly fed solid foods (Pannaraj et al., 2017). Some change in metabolic potential is consequently required.

4.2.2 Transmission of genera from mother to infant

The genera *Bacteroides*, *Prevotella*, *Alistipes*, and *Faecalibacterium* all exhibit a higher relative abundance in the meconium, 12 months, and mothers samples, compared to the relative abundance in 3 and 6 months samples in this study (Figure 3.2). Microbes detected from the first fecal samples reflect the first colonizers of the gut. Maternal transmission might explain the resemblance in the relative abundance of bacteria in meconium and the mother's fecal samples. Several studies demonstrate that microbial strains from multiple maternal body sites transfer to the infant microbiome and that the maternal gut strains are more persistent in the infant's gut than those from other sources (Ferretti et al., 2018) (Makino et al., 2011). Members of *Bacteroides* (Bjerke et al., 2011), *Prevotella* (Könönen et al., 1994), and *Alistipes* (Ferretti et al., 2018), have been demonstrated to be mother-child transmitted by analysis on the strain

and species level. *Faecalibacterium prausnitzii*, the sole bacteria within the *Faecalibacterium* genus, has not demonstrated mother-child transmission in previous studies. However, a study interestingly proposed that colonization by *Faecalibacterium* is accelerated through transfer between siblings (Laursen et al., 2017) Analysis of the species found in the same mothers-children pairs would confirm that the results demonstrate a maternal transfer of bacterial species to children. Due to the time limitations of the current study, this was not conducted.

4.3 Methodological considerations

4.3.1 Fecal sample as a proxy for analyzing gut content in human

Throughout time and using different approaches, fecal material has been the most prevalent sampling method for analyzing the gut content in humans due to the non-invasiveness and convenience of fecal sampling. Prior animal-based studies have investigated samples directly from the GI tract and found that the fecal microbial community is a good proxy for identifying most taxa present in the gut (Yasuda et al., 2015). However, the studies detected phyla bias and inter-individual effects and suggested that determination of gut microbial community from fecal samples should be interpreted with caution (Yan et al., 2019). Several studies have investigated the effects of duration time of storage and temperature of fecal samples used for a culture-independent analysis of gut microbiota. These variables have shown inconsistent results, with both significant and minimal differences in the composition of gut microbiota from the same fecal sample with different handling (Robinson, Brotman, & Ravel, 2016). Minimizing the amount of frozen-thaw cycles and suspending the fecal material with DNA stabilizing buffer, like done in the current study, reduces the variation (W. K. Wu et al., 2019).

4.3.2 RMS library preparation and sequencing

The reduction of short DNA fragments after the index PCR compared to the first PCR is assumedly explained by the Sera Mag clean-up, which was performed between the two gel electrophoresis quality checkpoints (Figure 3.1). The clean-up removes PCR byproducts such as nucleotides, polymerases, primers, and primer dimers and is a crucial part of the Illumina library preparations. An insufficient clean up may conversely lead to sequencing errors downstream. However, a dominating part of PCR fragments ranging from 100-300 bp was presumably also removed. This reduction can introduce bias if the discarded fragments originate from species that are not evenly distributed in size but mainly represented with short fragments after restriction cutting. The ratio of Sera Mag beads to amplified DNA can purposely change from 1,5:1 to 2:1 in future research using the RMS protocol to minimize this bias.

That 63% of the meconium samples did not show visual bands after the first 25 cycle PCR are presumably due to the meconium samples' low biomass microbiome, tar-like texture that complicates the DNA extractions, and high concentrations of PCR inhibitors such as bile salts and acids, glycolipids, and urea (Stinson, Keelan, & Payne, 2018). Stinson, Keelan, and Payne conclude in their study from 2018 that the choice of extraction kit dramatically impacts the ability to extract and detect bacterial DNA in meconium samples and found the MoBio MagAttract PowerMicrobiome kit to yield the highest average total DNA yield. In the present study, the same DNA extraction method was used for samples at all time points, to ensure that the treatment of all samples was as equal as possible. Treating all samples identical is reasonable, but the MagMidi LGC kit used in the current may not be the most efficient for meconium samples. Increasing the first PCR to 30 cycles made the DNA fragment visible on the gel for all meconium samples but may have introduced PCR-induced sequence artifacts and bias (Acinas, Sarma-Rupavtarm, Klepac-Ceraj, & Polz, 2005).

Quality control results after the two Illumina HiSeq 3000 runs were satisfactory. The sequence Q-score was ≥30 for the first 60-70 bases and ≥20 for the first 140 bases, meaning that the inferred base call accuracy was ≥99,9% and ≥99%, respectively (Figure C1, Appendix C). The decrease in per base sequence quality over the read in Illumina HiSeq 3000 run detected in the current study is a normal and well-known phenomenon due to the phasing issues in the sequencing by synthesis technology. Pre-phasing occurs by incorporating two or more nucleotides in one cycle, while post-phasing occurs when blockers are not correctly removed, leading to detection of the old nucleotide a second time (Pfeiffer et al., 2018). With increasing read length, more errors occur and add up to decrease the quality score.

4.3.3 Taxonomic and inferred metabolic assignment

Using the Kraken2 HumGut database for taxonomical assignment of sequences from fecal samples like in the current study is an excellent alternative to the Kraken 2 Standard database because the former one only comprises genomes of gut microbial species, is more specific, and reduces the number of unclassified reads drastically from 50,1% to 10% (Hiseni et al., 2020). However, species that may be present in the sample go undetected if they are not yet detected as members of the human gut microbiota. By time, this limitation will assumedly minimize in parallel with updates on the HumGut collection as the list of publicly available genomes and metagenomes expand (Hiseni et al., 2020).

Determining the accurate metabolic output for species in the gut microbiota is a highly challenging task. Estimating the quantitative abundance of metabolites such as SCFAs in a fecal sample can typically be done with several approaches, such as gas chromatography, liquid chromatography, nuclear magnetic resonance, and capillary electrophoresis (Primec, Micetic-Turk, & Langerholc, 2017). The quantitative metabolic abundance is thereby inferred to resemble the exact content in the gut lumen. Cross-feeding interactions, nutrient uptake by intestinal absorptive cells, and the volatile nature of these acids are all variables that problematize a correct inference and should be considered. By inferring the metabolic potential from species-level composition like in the current study, these variables are removed, which may lead to more precise estimation. However, the genomic potential in a species genome is not continuously utilized, meaning that all genes are not always transcribed, translated, assembled into a functioning protein, and transported to the right place in the cell. Such an inference may accordingly lead to an overestimation of the metabolic potential, and the study results should thereby be interpreted with caution and need further validation. Validation has not been conducted in this study due to time limitations.

How successful the metabolic assignment using the VMH database is also depends on how well all species are evenly represented in the database. Correspondingly as for the Kraken2 HumGut database, this limitation will minimize by time, as the database expands to include genomes from more species present in the human gut.

4.4 Study design

4.4.1 Strengths with the study

The study's biggest strengths are the large sample size of the mother-children cohort analyzed and the standardized nature of the sample collection in the PreventADALL study. The large sample size strengthens the study as the accuracy of the statistical analyzes and power to detect associations between the gut microbiota and immune cells increases.

The longitudinal sample collection with close follow-up visits at 3, 6, and 12 months after birth is a valuable feature of the current study, as it is possible to apprehend the development of the gut microbiota's taxonomic and inferred metabolic composition and study the association of the gut microbial taxonomic and inferred metabolic characteristics in the first year of life to immune cell composition at 12 months.

Another strength of the current study is that it analyzes infant human immunity to gut microbiota connections directly in humans. Although animal-based models have played an

essential role in the gut microbiota and immunology research, it is conversely uncertain if the detected mechanistic connections to immunity derived from animal-based studies are translatable to humans (Nguyen, Vieira-Silva, Liston, & Raes, 2015).

4.4.2 Weaknesses with the study

A weakness of the study was that only 67 of the 180 12-month-year-old children's immune cell composition were available. Enlargement of the sample size to 180 would have given more power to the correlation analysis and would potentially reveal more associations.

A second limitation of the study is that all absolute abundances were converted to relative abundances. This normalization step is reasoned in that the number of sequences found in a sample varied hugely. However, this normalization problematizes the interpretation of the study result, as the relative abundance of each specie, genus, inferred metabolite, or immune cell is dependent on the relative abundance of all other species, genera, inferred metabolites, or immune cells.

The current study did not include the available information about delivery mode (vaginal or cesarean section), gestational age, use of antibiotics, infant diet (breast milk or formula), the timing for the cessation of breastmilk and formula, and introduction of solid foods. These available variables may impact both the gut microbiota's longitudinal taxonomic and inferred composition, as well as the association between these two gut microbial characteristics and the immune cell composition. Further studies that take these variables into account will need to be undertaken to get a more comprehensive grasp of the infant's gut microbiota development and its associations with the immune system.

A limitation of the PreventADALL cohort study is that the participants live only in Norway and Sweden. Regional variation of gut microbiota is detected in several studies (Gupta, Paul, & Dutta, 2017), limiting the study results' application to the Scandinavian population.

5. Conclusion and further research

This study has identified several statistically significant correlations in the gut microbiota's relative taxonomic and inferred metabolic abundance in the first year of life to the relative immune cell abundance at 12 months. The findings indicate that the gut microbial taxonomic and metabolic composition has a definite association with the immune cell composition during infancy in humans.

Most interestingly is the detected negative association between the relative abundance of several species, genera, and inferred metabolites to the relative abundance of three monocyte subtypes. The associated gut microbial taxonomic and inferred metabolic characteristics may induce anti-inflammatory changes to the immune system, as an elevated monocyte level is confounding with inflammation. Additionally, the relative abundance of probiotic *L. gasseri* and three species without a complete biological description at 3 and 6 months were found positively correlated to the relative abundance of naïve $CD8^+$ T cells. This finding may indicate that the presence of these species reduces pathogen colonization in the gut, thereby reducing the differentiation of naïve cell types. Another detected positive correlation was between the relative abundance of *Eggerthella* and *Phascolarctobacterium* at 3 months to the relative abundance of memory T cells. This finding suggests that these genera induce T cell differentiation. Lastly, the inferred relative abundance of inferred butyric acid at 12 months was found negatively correlated to $CD56^{bright}$ NK cells, which might be due to butyric acid's ability to regulate bacterial pathogenesis.

In conclusion, the detected connections in gut microbiota and human immunity from this study may be a valuable contribution to the multidisciplinary research area of human microbiology and immunology. However, further work is highly suggested to explore the biological implications of the identified correlations in greater detail, as the study's scope was limited in terms of its explorative nature.

6. References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science Translational Medicine, 6*(237). doi:ARTN 237ra6510.1126/scitranslmed.3008599

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol, 71*(12), 8966-8969. doi:10.1128/aem.71.12.8966-8969.2005

Alexander, C., Swanson, K. S., Fahey, G. C., Jr, & Garleb, K. A. (2019). Perspective: Physiologic Importance of Short-Chain Fatty Acids from Nondigestible Carbohydrate Fermentation. *Advances in Nutrition, 10*(4), 576-589. doi:10.1093/advances/nmz004

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2

Alvarez, J. R., Skachkov, D., Massey, S. E., Kalitsov, A., & Velev, J. P. (2015). DNA/RNA transverse current sequencing: intrinsic structural noise from neighboring bases. *Front Genet, 6*, 213. doi:10.3389/fgene.2015.00213

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., . . . Rudensky, A. Y. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature, 504*(7480), 451-455. doi:10.1038/nature12726

Arrieta, M. C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., & Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Front Immunol, 5*, 427. doi:10.3389/fimmu.2014.00427

Assarsson, E., Kambayashi, T., Schatzle, J. D., Cramer, S. O., von Bonin, A., Jensen, P. E., . . . Chambers, B. J. (2004). NK Cells Stimulate Proliferation of T and NK Cells through 2B4/CD48 Interactions. *The Journal of Immunology, 173*(1), 174-180. doi:10.4049/jimmunol.173.1.174

Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., . . . Honda, K. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science, 358*(6361), 359-365. doi:10.1126/science.aan4526

Bakken, L. R., Frostegård, Å. (2006). Nucleic Acid Extraction from Soil. *Nannipieri, P., Smalla K. (eds) Nucleic Acids and Proteins in Soil. Soil Biology, 8*.

Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., . . . Tanner, S. D. (2009). Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem, 81*(16), 6813-6822. doi:10.1021/ac901049w

Bearfield, C., Davenport, E. S., Sivapathasundaram, V., & Allaker, R. P. (2002). Possible association between amniotic fluid micro-organism infection and microflora in the mouth. *BJOG, 109*(5), 527-533. doi:10.1111/j.1471-0528.2002.01349.x

Belkaid, Y., Bouladoux, N., & Hand, T. W. (2013). Effector and memory T cell responses to commensal bacteria. *Trends Immunol, 34*(6), 299-306. doi:10.1016/j.it.2013.03.003

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bjerke, G. A., Wilson, R., Storrø, O., Øyen, T., Johnsen, R., & Rudi, K. (2011). Mother-to-child transmission of and multiple-strain colonization by Bacteroides fragilis in a cohort of mothers and their children. *Appl Environ Microbiol, 77*(23), 8318-8324. doi:10.1128/aem.05293-11

Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheim-van Dillen, P. M., & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *J Clin Microbiol, 28*(3), 495-503.

Brodin, P., & Davis, M. M. (2017). Human immune system variation. *Nat Rev Immunol, 17*(1), 21-29. doi:10.1038/nri.2016.125

Cadwell, K. (2015). Expanding the role of the virome: commensalism in the gut. *J Virol, 89*(4), 1951-1953. doi:10.1128/JVI.02966-14

Chiu, S., & Bharat, A. (2016). Role of monocytes and macrophages in regulating immune response following lung transplantation. *Curr Opin Organ Transplant, 21*(3), 239-245. doi:10.1097/mot.0000000000000313

Cho, G. S., Ritzmann, F., Eckstein, M., Huch, M., Briviba, K., Behsnilian, D., . . . Franz, C. M. (2016). Quantification of Slackia and Eggerthella spp. in Human Feces and Adhesion of Representatives Strains to Caco-2 Cells. *Front Microbiol, 7*, 658. doi:10.3389/fmicb.2016.00658

Churko, J. M., Mantalas, G. L., Snyder, M. P., & Wu, J. C. (2013). Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. *Circ Res, 112*(12), 1613-1623. doi:10.1161/circresaha.113.300939

Cooper, M. A., Fehniger, T. A., & Caligiuri, M. A. (2001). The biology of human natural killer-cell subsets. *Trends Immunol, 22*(11), 633-640. doi:10.1016/S1471-4906(01)02060-9

Cooper, M. A., Fehniger, T. A., Turner, S. C., Chen, K. S., Ghaheri, B. A., Ghayur, T., . . . Caligiuri, M. A. (2001). Human natural killer cells: a unique innate immunoregulatory role for the CD56bright subset. *Blood, 97*(10), 3146-3151. doi:10.1182/blood.V97.10.3146

Crowe, J. S., Cooper, H. J., Smith, M. A., Sims, M. J., Parker, D., & Gewert, D. (1991). Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. *Nucleic Acids Research, 19*(1), 184-184. doi:10.1093/nar/19.1.184

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., & Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A, 107*(26), 11971-11975. doi:10.1073/pnas.1002601107

Dutta, P., & Nahrendorf, M. (2014). Regulation and consequences of monocytosis. *Immunol Rev, 262*(1), 167-178. doi:10.1111/imr.12219

Duysburgh, C., Van den Abbeele, P., Krishnan, K., Bayne, T. F., & Marzorati, M. (2019). A synbiotic concept containing spore-forming Bacillus strains and a prebiotic fiber blend

consistently enhanced metabolic activity by modulation of the gut microbiome in vitro. *Int J Pharm X, 1*, 100021. doi:10.1016/j.ijpx.2019.100021

Earl, J. P., Adappa, N. D., Krol, J., Bhat, A. S., Balashov, S., Ehrlich, R. L., . . . Mell, J. C. (2018). Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes. *Microbiome, 6*(1), 190. doi:10.1186/s40168-018-0569-2

Engels, C., Ruscheweyh, H. J., Beerenwinkel, N., Lacroix, C., & Schwab, C. (2016). The Common Gut Microbe Eubacterium hallii also Contributes to Intestinal Propionate Formation. *Front Microbiol, 7*, 713. doi:10.3389/fmicb.2016.00713

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., . . . Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe, 24*(1), 133-145 e135. doi:10.1016/j.chom.2018.06.005

Fouda, G. G., Martinez, D. R., Swamy, G. K., & Permar, S. R. (2018). The Impact of IgG transplacental transfer on early life immunity. *Immunohorizons, 2*(1), 14-25. doi:10.4049/immunohorizons.1700057

Friedman, E. S., Bittinger, K., Esipova, T. V., Hou, L., Chau, L., Jiang, J., . . . Wu, G. D. (2018). Microbes vs. chemistry in the origin of the anaerobic gut lumen. *Proceedings of the National Academy of Sciences, 115*(16), 4170-4175. doi:10.1073/pnas.1718635115

Gardiner, B. J., Tai, A. Y., Kotsanas, D., Francis, M. J., Roberts, S. A., Ballard, S. A., . . . Korman, T. M. (2015). Clinical and microbiological characteristics of Eggerthella lenta bacteremia. *J Clin Microbiol, 53*(2), 626-635. doi:10.1128/jcm.02926-14

Gensollen, T., Iyer, S. S., Kasper, D. L., & Blumberg, R. S. (2016). How colonization by microbiota in early life shapes the immune system. *Science, 352*(6285), 539-544. doi:10.1126/science.aad9378

Gibson, G. R., Probert, H. M., Loo, J. V., Rastall, R. A., & Roberfroid, M. B. (2004). Dietary modulation of the human colonic microbiota: updating the concept of prebiotics. *Nutrition Research Reviews, 17*(2), 259-275. doi:10.1079/NRR200479

Gleeson, M., & Cripps, A. W. (2004). Development of mucosal immunity in the first year of life and relationship to sudden infant death syndrome. *FEMS Immunology & Medical Microbiology, 42*(1), 21-33. doi:10.1016/j.femsim.2004.06.012

Gophna, U., Konikoff, T., & Nielsen, H. B. (2017). Oscillospira and related bacteria – From metagenomic species to metabolic features. *Environmental Microbiology, 19*(3), 835-841. doi:10.1111/1462-2920.13658

Gupta, V. K., Paul, S., & Dutta, C. (2017). Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. *Front Microbiol, 8*, 1162. doi:10.3389/fmicb.2017.01162

Haiser, H. J., Gootenberg, D. B., Chatman, K., Sirasani, G., Balskus, E. P., & Turnbaugh, P. J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science, 341*(6143), 295-298. doi:10.1126/science.1235872

Haque, S. Z., & Haque, M. (2017). The ecological community of commensal, symbiotic, and pathogenic gastrointestinal microorganisms - an appraisal. *Clin Exp Gastroenterol, 10*, 91-103. doi:10.2147/CEG.S126243

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics, 107*(1), 1-8. doi:10.1016/j.ygeno.2015.11.003

Henderickx, J. G. E., Zwittink, R. D., van Lingen, R. A., Knol, J., & Belzer, C. (2019). The Preterm Gut Microbiota: An Inconspicuous Challenge in Nutritional Neonatal Care. *Front Cell Infect Microbiol, 9*, 85. doi:10.3389/fcimb.2019.00085

Hendrikx, T., & Schnabl, B. (2019). Indoles: metabolites produced by intestinal bacteria capable of controlling liver disease manifestation. *Journal of Internal Medicine, 286*(1), 32-40. doi:10.1111/joim.12892

Henke, M. T., Kenny, D. J., Cassilly, C. D., Vlamakis, H., Xavier, R. J., & Clardy, J. (2019). Ruminococcus gnavus, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci U S A, 116*(26), 12672-12677. doi:10.1073/pnas.1904099116

Hess, M. K., Rowe, S. J., Van Stijn, T. C., Henry, H. M., Hickey, S. M., Brauning, R., . . . McEwan, J. C. (2020). A restriction enzyme reduced representation sequencing approach for low-cost, high-throughput metagenome profiling. *PLoS One, 15*(4), e0219882. doi:10.1371/journal.pone.0219882

Hiergeist, A., Gläsner, J., Reischl, U., & Gessner, A. (2015). Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR Journal, 56*(2), 228-240. doi:10.1093/ilar/ilv017

Hiippala, K., Jouhten, H., Ronkainen, A., Hartikainen, A., Kainulainen, V., Jalanka, J., & Satokari, R. (2018). The Potential of Gut Commensals in Reinforcing Intestinal Barrier Function and Alleviating Inflammation. *Nutrients, 10*(8). doi:10.3390/nu10080988

Hillman, E. T., Lu, H., Yao, T., & Nakatsu, C. H. (2017). Microbial Ecology along the Gastrointestinal Tract. *Microbes Environ, 32*(4), 300-313. doi:10.1264/jsme2.ME17017

Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T., & Snipen, L. (2020). HumGut: A comprehensive Human Gut prokaryotic genomes collection filtered by metagenome data. *bioRxiv*, 2020.2003.2025.007666. doi:10.1101/2020.03.25.007666

Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., . . . Kuhn, L. (2018). Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nat Commun, 9*(1), 4169. doi:10.1038/s41467-018-06473-x

Hooper, L. V., & Gordon, J. I. (2001). Commensal host-bacterial relationships in the gut. *Science, 292*(5519), 1115-1118. doi:10.1126/science.1058709

Horz, H. P. (2015). Archaeal Lineages within the Human Microbiome: Absent, Rare or Elusive? *Life (Basel), 5*(2), 1333-1345. doi:10.3390/life5021333

Jeng, Y., Lim, P. S., Wu, M. Y., Tseng, T. Y., Chen, C. H., Chen, H. P., & Wu, T. K. (2017). Proportions of Proinflammatory Monocytes Are Important Predictors of Mortality Risk in Hemodialysis Patients. *Mediators Inflamm, 2017*, 1070959. doi:10.1155/2017/1070959

Jimenez, E., Fernandez, L., Marin, M. L., Martin, R., Odriozola, J. M., Nueno-Palop, C., . . . Rodriguez, J. M. (2005). Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol, 51*(4), 270-274. doi:10.1007/s00284-005-0020-3

Jimenez, E., Marin, M. L., Martin, R., Odriozola, J. M., Olivares, M., Xaus, J., . . . Rodriguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Res Microbiol, 159*(3), 187-193. doi:10.1016/j.resmic.2007.12.007

Ju, T., Kong, J. Y., Stothard, P., & Willing, B. P. (2019). Defining the role of Parasutterella, a previously uncharacterized member of the core gut microbiota. *The ISME Journal, 13*(6), 1520-1534. doi:10.1038/s41396-019-0364-5

Jumas-Bilak, E., Roudière, L., & Marchandin, H. (2009). Description of 'Synergistetes' phyl. nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomonadaceae, phylum 'Firmicutes'. *International Journal of Systematic and Evolutionary Microbiology, 59*(5), 1028-1035. doi:10.1099/ijs.0.006718-0

Kabat, A. M., Srinivasan, N., & Maloy, K. J. (2014). Modulation of immune development and function by intestinal microbiota. *Trends Immunol, 35*(11), 507-517. doi:10.1016/j.it.2014.07.010

Kchouk, M., Gibrat, J.-F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine, 09*. doi:10.4172/0974-8369.1000395

Kircher, M., & Kelso, J. (2010). High-throughput DNA sequencing--concepts and limitations. *Bioessays, 32*(6), 524-536. doi:10.1002/bies.200900181

Koch, S., Kucharzik, T., Heidemann, J., Nusrat, A., & Luegering, A. (2010). Investigating the role of proinflammatory CD16+ monocytes in the pathogenesis of inflammatory bowel disease. *Clin Exp Immunol, 161*(2), 332-341. doi:10.1111/j.1365-2249.2010.04177.x

Kojima, A., Nakano, K., Wada, K., Takahashi, H., Katayama, K., Yoneda, M., . . . Ooshima, T. (2012). Infection of specific strains of Streptococcus mutans, oral bacteria, confers a risk of ulcerative colitis. *Sci Rep, 2*, 332. doi:10.1038/srep00332

Koropatkin, N. M., Cameron, E. A., & Martens, E. C. (2012). How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology, 10*(5), 323-335. doi:10.1038/nrmicro2746

Kumar, A., & Chordia, N. (2017). Role of Microbes in Human Health. *Applied Microbiology: Open Access, 03*. doi:10.4172/2471-9315.1000131

Könönen, E., Saarela, M., Karjalainen, J., Jousimies-Somer, H., Alaluusua, S., & Asikainen, S. (1994). Transmission of oral Prevotella melaninogenica between a mother and her young child. *Oral Microbiology and Immunology, 9*(5), 310-314. doi:10.1111/j.1399-302X.1994.tb00077.x

Laursen, M. F., Laursen, R. P., Larnkjær, A., Mølgaard, C., Michaelsen, K. F., Frøkiær, H., . . . Licht, T. R. (2017). Faecalibacterium Gut Colonization Is Accelerated by Presence of Older Siblings. *mSphere, 2*(6). doi:10.1128/mSphere.00448-17

Lazar, V., Ditu, L. M., Pircalabioru, G. G., Gheorghe, I., Curutiu, C., Holban, A. M., . . . Chifiriuc, M. C. (2018). Aspects of Gut Microbiota and Immune System Interactions

in Infectious Diseases, Immunopathology, and Cancer. *Front Immunol, 9*, 1830. doi:10.3389/fimmu.2018.01830

Lebeer, S., Vanderleyden, J., & De Keersmaecker, S. C. J. (2010). Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nature Reviews Microbiology, 8*(3), 171-184. doi:10.1038/nrmicro2297

Leiby, J. S., McCormick, K., Sherrill-Mix, S., Clarke, E. L., Kessler, L. R., Taylor, L. J., . . . Bushman, F. D. (2018). Lack of detection of a human placenta microbiome in samples from preterm and term deliveries. *Microbiome, 6*(1), 196. doi:10.1186/s40168-018-0575-4

Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics, 17*(1), 95-115. doi:10.1146/annurev-genom-083115-022413

Liao, D. H., Zhao, J. B., & Gregersen, H. (2009). Gastrointestinal tract modelling in health and disease. *World J Gastroenterol, 15*(2), 169-176. doi:10.3748/wjg.15.169

Lim, E. S., Rodriguez, C., & Holtz, L. R. (2018). Amniotic fluid from healthy term pregnancies does not harbor a detectable microbial community. *Microbiome, 6*(1), 87. doi:10.1186/s40168-018-0475-7

Liu, M. Y., Worden, P., Monahan, L. G., DeMaere, M. Z., Burke, C. M., Djordjevic, S. P., . . . Darling, A. E. (2017). Evaluation of ddRADseq for reduced representation metagenome sequencing. *PeerJ, 5*, e3837. doi:10.7717/peerj.3837

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Med, 8*(1), 51. doi:10.1186/s13073-016-0307-y

Lodrup Carlsen, K. C., Rehbinder, E. M., Skjerven, H. O., Carlsen, M. H., Fatnes, T. A., Fugelli, P., . . . study, g. (2018). Preventing Atopic Dermatitis and ALLergies in Children-the PreventADALL study. *Allergy, 73*(10), 2063-2070. doi:10.1111/all.13468

Lokmic, A. (2019). *Undersøkelse av mor-barn assosiasjon av tarmbakterer ved bruk av Redusert Metagenomisk Sekvensering (RMS).* (Master's degree), OlsoMet - storbyuniversitetet,

Lopetuso, L. R., Scaldaferri, F., Petito, V., & Gasbarrini, A. (2013). Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog, 5*(1), 23. doi:10.1186/1757-4749-5-23

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature, 489*(7415), 220-230. doi:10.1038/nature11550

Makino, H., Kushiro, A., Ishikawa, E., Muylaert, D., Kubota, H., Sakai, T., . . . Tanaka, R. (2011). Transmission of Intestinal Bifidobacterium longum subsp. longum Strains from Mother to Infant, Determined by Multilocus Sequencing Typing and Amplified Fragment Length Polymorphism. *Applied and Environmental Microbiology, 77*(19), 6788-6793. doi:10.1128/aem.05346-11

Mantis, N. J., Rol, N., & Corthesy, B. (2011). Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol, 4*(6), 603-611. doi:10.1038/mi.2011.41

Marcobal, A., & Sonnenburg, J. L. (2012). Human milk oligosaccharide consumption by intestinal microbiota. *Clin Microbiol Infect, 18 Suppl 4*(0 4), 12-15. doi:10.1111/j.1469-0691.2012.03863.x

Martin, M. D., & Badovinac, V. P. (2018). Defining Memory CD8 T Cell. *Front Immunol, 9*, 2692. doi:10.3389/fimmu.2018.02692

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., & de La Cochetiere, M. F. (2013). Development of intestinal microbiota in infants and its impact on health. *Trends in Microbiology, 21*(4), 167-173. doi:10.1016/j.tim.2012.12.001

McKinstry, K. K., Strutt, T. M., & Swain, S. L. (2010). The potential of CD4 T-cell memory. *Immunology, 130*(1), 1-9. doi:10.1111/j.1365-2567.2010.03259.x

Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., . . . Ventura, M. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol Mol Biol Rev, 81*(4). doi:10.1128/MMBR.00036-17

Moore, R. E., & Townsend, S. D. (2019). Temporal development of the infant gut microbiome. *Open Biol, 9*(9), 190128. doi:10.1098/rsob.190128

Moore, W. E., & Holdeman, L. V. (1974). Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Appl Microbiol, 27*(5), 961-979.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., . . . Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology, 13*(9), R79. doi:10.1186/gb-2012-13-9-r79

Munyaka, P. M., Khafipour, E., & Ghia, J. E. (2014). External influence of early childhood establishment of gut microbiota and subsequent health implications. *Front Pediatr, 2*, 109. doi:10.3389/fped.2014.00109

Narasimhan, P. B., Marcovecchio, P., Hamers, A. A. J., & Hedrick, C. C. (2019). Nonclassical Monocytes in Health and Disease. *Annual Review of Immunology, 37*(1), 439-456. doi:10.1146/annurev-immunol-042617-053119

Nguyen, T. L., Vieira-Silva, S., Liston, A., & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Dis Model Mech, 8*(1), 1-16. doi:10.1242/dmm.017400

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., . . . Epstein, S. S. (2010). Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. *Applied and Environmental Microbiology, 76*(8), 2445-2450. doi:10.1128/aem.01754-09

Nicholson, L. B. (2016). The immune system. *Essays Biochem, 60*(3), 275-301. doi:10.1042/EBC20160017

Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., . . . Thiele, I. (2019). The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res, 47*(D1), D614-D624. doi:10.1093/nar/gky992

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J.-z., . . . Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to

centenarian: a cross-sectional study. *BMC Microbiology, 16*(1), 90. doi:10.1186/s12866-016-0708-5

Okumura, R., & Takeda, K. (2016). Maintenance of gut homeostasis by the mucosal immune system. *Proc Jpn Acad Ser B Phys Biol Sci, 92*(9), 423-435. doi:10.2183/pjab.92.423

Okumura, R., & Takeda, K. (2017). Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med, 49*(5), e338. doi:10.1038/emm.2017.20

Okumura, R., & Takeda, K. (2018). Maintenance of intestinal homeostasis by mucosal barriers. *Inflamm Regen, 38*, 5. doi:10.1186/s41232-018-0063-z

Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., . . . Iliopoulos, I. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights, 9*, 75-88. doi:10.4137/bbi.S12462

Pannaraj, P. S., Li, F., Cerini, C., Bender, J. M., Yang, S., Rollie, A., . . . Aldrovandi, G. M. (2017). Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr, 171*(7), 647-654. doi:10.1001/jamapediatrics.2017.0378

Parfrey, L. W., Walters, W. A., & Knight, R. (2011). Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. *Front Microbiol, 2*, 153. doi:10.3389/fmicb.2011.00153

Pennock, N. D., White, J. T., Cross, E. W., Cheney, E. E., Tamburini, B. A., & Kedl, R. M. (2013). T cell responses: naïve to memory and everything in between. *Advances in Physiology Education, 37*(4), 273-283. doi:10.1152/advan.00066.2013

Perez-Munoz, M. E., Arrieta, M. C., Ramer-Tait, A. E., & Walter, J. (2017). A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome, 5*(1), 48. doi:10.1186/s40168-017-0268-4

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One, 7*(5), e37135. doi:10.1371/journal.pone.0037135

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., & Mayer, G. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports, 8*(1), 10950. doi:10.1038/s41598-018-29325-6

Poggi, A., Benelli, R., Venè, R., Costa, D., Ferrari, N., Tosetti, F., & Zocchi, M. R. (2019). Human Gut-Associated Natural Killer Cells in Health and Disease. *Front Immunol, 10*, 961. doi:10.3389/fimmu.2019.00961

Prakash, T., & Taylor, T. D. (2012). Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform, 13*(6), 711-727. doi:10.1093/bib/bbs033

Primec, M., Micetic-Turk, D., & Langerholc, T. (2017). Analysis of short-chain fatty acids in human feces: A scoping review. *Anal Biochem, 526*, 9-21. doi:10.1016/j.ab.2017.03.007

Qin, J. J., Li, R. Q., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., . . . Consortium, M. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-U70. doi:10.1038/nature08821

Rajilić-Stojanović, M., & de Vos, W. M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiology Reviews, 38*(5), 996-1047. doi:10.1111/1574-6976.12075

Randall, T. D., & Mebius, R. E. (2014). The development and function of mucosal lymphoid tissues: a balancing act with micro-organisms. *Mucosal Immunology, 7*(3), 455-466. doi:10.1038/mi.2014.11

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., . . . Baines, J. F. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome, 7*(1), 133. doi:10.1186/s40168-019-0743-1

Ravi, A., Avershina, E., Angell, I. L., Ludvigsen, J., Manohar, P., Padmanaban, S., . . . Rudi, K. (2018). Comparison of reduced metagenome and 16S rRNA gene sequencing for determination of genetic diversity and mother-child overlap of the gut associated microbiota. *J Microbiol Methods, 149*, 44-52. doi:10.1016/j.mimet.2018.02.016

Ríos-Covián, D., Ruas-Madiedo, P., Margolles, A., Gueimonde, M., de Los Reyes-Gavilán, C. G., & Salazar, N. (2016). Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front Microbiol, 7*, 185. doi:10.3389/fmicb.2016.00185

Ritz, B. W., Alexander, G. M., Nogusa, S., Perreault, M. J., Peterlin, B. L., Grothusen, J. R., & Schwartzman, R. J. (2011). Elevated blood levels of inflammatory monocytes (CD14+ CD16+) in patients with complex regional pain syndrome. *Clin Exp Immunol, 164*(1), 108-117. doi:10.1111/j.1365-2249.2010.04308.x

Robinson, C. K., Brotman, R. M., & Ravel, J. (2016). Intricacies of assessing the human microbiome in epidemiologic studies. *Ann Epidemiol, 26*(5), 311-321. doi:10.1016/j.annepidem.2016.04.005

Rodriguez, J. M., Murphy, K., Stanton, C., Ross, R. P., Kober, O. I., Juge, N., . . . Collado, M. C. (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis, 26*, 26050. doi:10.3402/mehd.v26.26050

Rooks, M. G., & Garrett, W. S. (2016). Gut microbiota, metabolites and host immunity. *Nat Rev Immunol, 16*(6), 341-352. doi:10.1038/nri.2016.42

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr, 57*(1), 1-24. doi:10.1007/s00394-017-1445-8

Sampath, P., Moideen, K., Ranganathan, U. D., & Bethunaickan, R. (2018). Monocyte Subsets: Phenotypes and Function in Tuberculosis Infection. *Front Immunol, 9*, 1726. doi:10.3389/fimmu.2018.01726

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*(12), 5463-5467. doi:10.1073/pnas.74.12.5463

Schulthess, J., Pandey, S., Capitani, M., Rue-Albrecht, K. C., Arnold, I., Franchini, F., . . . Powrie, F. (2019). The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages. *Immunity, 50*(2), 432-445.e437. doi:10.1016/j.immuni.2018.12.018

Segain, J. P., Raingeard de la Bletiere, D., Bourreille, A., Leray, V., Gervois, N., Rosales, C., . . . Galmiche, J. P. (2000). Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. *Gut, 47*(3), 397-403. doi:10.1136/gut.47.3.397

Sekirov, I., Russell, S. L., Antunes, L. C., & Finlay, B. B. (2010). Gut microbiota in health and disease. *Physiol Rev, 90*(3), 859-904. doi:10.1152/physrev.00045.2009

Selle, K., & Klaenhammer, T. R. (2013). Genomic and phenotypic evidence for probiotic influences of Lactobacillus gasseri on human health. *FEMS Microbiology Reviews, 37*(6), 915-935. doi:10.1111/1574-6976.12021

Sender, R., Fuchs, S., & Milo, R. (2016). Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell, 164*(3), 337-340. doi:10.1016/j.cell.2016.01.013

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci, 5*, 209. doi:10.3389/fpls.2014.00209

Sicard, J. F., Le Bihan, G., Vogeleer, P., Jacques, M., & Harel, J. (2017). Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. *Frontiers in Cellular and Infection Microbiology, 7*. doi:ARTN 38710.3389/fcimb.2017.00387

Simon, A. K., Hollander, G. A., & McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci, 282*(1821), 20143085. doi:10.1098/rspb.2014.3085

Simoni, Y., Chng, M. H. Y., Li, S., Fehlings, M., & Newell, E. W. (2018). Mass cytometry: a powerful tool for dissecting the immune landscape. *Curr Opin Immunol, 51*, 187-196. doi:10.1016/j.coi.2018.03.023

Sorbara, M. T., & Pamer, E. G. (2019). Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunology, 12*(1), 1-9. doi:10.1038/s41385-018-0053-0

Soto, A., Martín, V., Jiménez, E., Mader, I., Rodríguez, J. M., & Fernández, L. (2014). Lactobacilli and bifidobacteria in human breast milk: influence of antibiotherapy and other host and clinical factors. *J Pediatr Gastroenterol Nutr, 59*(1), 78-88. doi:10.1097/mpg.0000000000000347

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology, 15*(1), 72-101. doi:10.2307/1412159

Spitzer, M. H., & Nolan, G. P. (2016). Mass Cytometry: Single Cells, Many Features. *Cell, 165*(4), 780-791. doi:10.1016/j.cell.2016.04.019

Stinson, L. F., Keelan, J. A., & Payne, M. S. (2018). Comparison of Meconium DNA Extraction Methods for Use in Microbiome Studies. *Front Microbiol, 9*, 270. doi:10.3389/fmicb.2018.00270

Stoeker, L., Nordone, S., Gunderson, S., Zhang, L., Kajikawa, A., LaVoy, A., . . . Dean, G. A. (2011). Assessment of Lactobacillus gasseri as a Candidate Oral Vaccine Vector. *Clinical and Vaccine Immunology, 18*(11), 1834-1844. doi:10.1128/cvi.05277-11

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D., & Doré, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol, 65*(11), 4799-4807.

Sun, Y., & O'Riordan, M. X. (2013). Regulation of bacterial pathogenesis by intestinal short-chain Fatty acids. *Adv Appl Microbiol, 85*, 93-118. doi:10.1016/b978-0-12-407672-3.00003-4

Tanoue, T., Morita, S., Plichta, D. R., Skelly, A. N., Suda, W., Sugiura, Y., . . . Honda, K. (2019). A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature, 565*(7741), 600-605. doi:10.1038/s41586-019-0878-z

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., . . . Leclerc, M. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol, 11*(10), 2574-2584. doi:10.1111/j.1462-2920.2009.01982.x

Tsang, J. S., Schwartzberg, P. L., Kotliarov, Y., Biancotto, A., Xie, Z., Germain, R. N., . . . Consortium, C. H. I. (2014). Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell, 157*(2), 499-513. doi:10.1016/j.cell.2014.03.031

Tsuji, H., Oozeer, R., Matsuda, K., Matsuki, T., Ohta, T., Nomoto, K., . . . Yamashiro, Y. (2012). Molecular monitoring of the development of intestinal microbiota in Japanese infants. *Benef Microbes, 3*(2), 113-125. doi:10.3920/BM2011.0038

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., . . . Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature, 457*(7228), 480-U487. doi:10.1038/nature07540

Turroni, F., Peano, C., Pass, D. A., Foroni, E., Severgnini, M., Claesson, M. J., . . . Ventura, M. (2012). Diversity of Bifidobacteria within the Infant Gut Microbiota. *PLoS One, 7*(5). doi:ARTN e3695710.1371/journal.pone.0036957

Ursell, L. K., Metcalf, J. L., Parfrey, L. W., & Knight, R. (2012). Defining the human microbiome. *Nutr Rev, 70 Suppl 1*(Suppl 1), S38-44. doi:10.1111/j.1753-4887.2012.00493.x

Vinolo, M. A. R., Rodrigues, H. G., Hatanaka, E., Sato, F. T., Sampaio, S. C., & Curi, R. (2011). Suppressive effect of short-chain fatty acids on production of proinflammatory mediators by neutrophils. *The Journal of Nutritional Biochemistry, 22*(9), 849-855. doi:doi:10.1016/j.nutbio.2010.07.009

von Hertzen, L., Beutler, B., Bienenstock, J., Blaser, M., Cani, P. D., Eriksson, J., . . . de Vos, W. M. (2015). Helsinki alert of biodiversity and health. *Ann Med, 47*(3), 218-225. doi:10.3109/07853890.2015.1010226

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., . . . et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res, 23*(21), 4407-4414.

Wall, R., Fitzgerald, G., Hussey, S., Ryan, T., Murphy, B., Ross, P., & Stanton, C. (2007). Genomic diversity of cultivable Lactobacillus populations residing in the neonatal and adult gastrointestinal tract. *FEMS Microbiology Ecology, 59*(1), 127-137. doi:10.1111/j.1574-6941.2006.00202.x

Walter, J. (2008). Ecological Role of Lactobacilli in the Gastrointestinal Tract: Implications for Fundamental and Biomedical Research. *Applied and Environmental Microbiology, 74*(16), 4985-4996. doi:10.1128/aem.00753-08

Walton, K. D., Freddo, A. M., Wang, S., & Gumucio, D. L. (2016). Generation of intestinal surface: an absorbing tale. *Development, 143*(13), 2261-2272. doi:10.1242/dev.135400

West, C. E., Renz, H., Jenmalm, M. C., Kozyrskyj, A. L., Allen, K. J., Vuillermin, P., . . . in, F. M. I. G. (2015). The gut microbiota and inflammatory noncommunicable diseases: associations and potentials for gut microbiota therapies. *J Allergy Clin Immunol, 135*(1), 3-13; quiz 14. doi:10.1016/j.jaci.2014.11.012

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol, 15*(3), R46. doi:10.1186/gb-2014-15-3-r46

Wu, F., Guo, X., Zhang, J., Zhang, M., Ou, Z., & Peng, Y. (2017). Phascolarctobacterium faecium abundant colonization in human gastrointestinal tract. *Exp Ther Med, 14*(4), 3122-3126. doi:10.3892/etm.2017.4878

Wu, W. K., Chen, C. C., Panyod, S., Chen, R. A., Wu, M. S., Sheen, L. Y., & Chang, S. C. (2019). Optimization of fecal sample processing for microbiome study - The journey from bathroom to bench. *J Formos Med Assoc, 118*(2), 545-555. doi:10.1016/j.jfma.2018.02.005

Wu, Z.-F., Liu, G.-L., Zhou, Z., Wang, G.-X., Xia, L., & Liu, J.-L. (2012). Induction of Immune-related Gene Expression in Ctenopharyngodon idella Kidney Cells by Secondary Metabolites from Immunostimulatory Alcaligenes faecalis FY-3. *Scandinavian Journal of Immunology, 76*(2), 131-140. doi:10.1111/j.1365-3083.2012.02722.x

Yan, W., Sun, C., Zheng, J., Wen, C., Ji, C., Zhang, D., . . . Yang, N. (2019). Efficacy of Fecal Sampling as a Gut Proxy in the Study of Chicken Gut Microbiota. *Front Microbiol, 10*, 2126. doi:10.3389/fmicb.2019.02126

Yasuda, K., Oh, K., Ren, B., Tickle, T. L., Franzosa, E. A., Wachtman, L. M., . . . Morgan, X. C. (2015). Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque. *Cell Host Microbe, 17*(3), 385-391. doi:10.1016/j.chom.2015.01.015

Yoshida, A., Yamada, K., Yamazaki, Y., Sashihara, T., Ikegami, S., Shimizu, M., & Totsuka, M. (2011). Lactobacillus gasseri OLL2809 and its RNA suppress proliferation of CD4(+) T cells through a MyD88-dependent signalling pathway. *Immunology, 133*(4), 442-451. doi:10.1111/j.1365-2567.2011.03455.x

Yu, Y., Lee, C., Kim, J., & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng, 89*(6), 670-679. doi:10.1002/bit.20347

Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Research*. doi:10.1038/s41422-020-0332-7

Zheng, W., Zhao, W., Wu, M., Song, X., Caro, F., Sun, X., . . . Kasper, D. L. (2020). Microbiota-targeted maternal antibodies protect neonates from enteric infection. *Nature, 577*(7791), 543-548. doi:10.1038/s41586-019-1898-4

Appendix

Appendix A: Method for Mass Cytometry, Antibodies and reagents

"Cryopreserved and stabilized whole blood (blood mixed with 'Stabilizer' component of Whole blood processing kit; Cytodelics AB, Sweden) collected from 67 patients sampled during the study period were thawed, and cells were fixed and RBCs lysed using Wash # 1 and # 2 buffers (Whole blood processing kit; Cytodelics AB, Sweden) as per the manufacturer's recommendations. This was performed a few days prior to barcoding and staining of cells. Post fix/lysis of cells, ~1x106 cells/sample were plated onto a 96 well 'U' bottom plate using standard cryoprotective solution (10% DMSO and 90% FBS) and cryopreserved at -80oC.

On the day of barcoding and staining of cells, cells were thawed at 37oC using RPMI medium supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin and benzonase (Sigma-Aldrich, Sweden). Briefly, cells were barcoded using automated liquid handling robotic system (Agilent technologies, Santa Clara, CA, USA (REF Mikes et al, Methods Mol Biol, 2019) using the Cell-ID 20-plex Barcoding kit (Fluidigm Inc.) as per the manufacturer's recommendations. Following cell pooling batch wise (with samples from placebo and treatment groups equally represented in each batch), cells were washed, FcR blocked using blocking buffer (in-house developed recipe) for 10 min at room temperature, following which cells were incubated for another 30 min at 4°C after addition of a cocktail of metal conjugated antibodies targeting the surface antigens. Following two washes with CyFACS buffer, cells were fixed overnight using 4% formaldehyde made in PBS (VWR, Sweden). For acquisition by CyTOF (within 2 days after staining), cells were stained with DNA intercalator (0.125 μM Iridium-191/193 or MaxPar® Intercalator-Ir, Fluidigm) in 4% formaldehyde made in PBS for 20 min at room temperature. After multiple washes with CyFACS, PBS and milliQ water, cells were filtered through a 35μm nylon mesh and diluted to 750,000 cells/ml. Cells were acquired at a rate of 300-500 cells/s using a super sampler (Victorian Airship, USA) connected to a CyTOF2 (Fluidigm) mass cytometer, CyTOF software version 6.0.626 with noise reduction, a lower convolution threshold of 200, event length limits of 10-150 pushes and a sigma value of 3 and flow rate of 0.045 ml/min.

Purified antibodies for mass cytometry were obtained in carrier/protein-free buffer and then coupled to lanthanide metals using the MaxPar antibody conjugation kit (Fluidigm Inc.) as per the manufacturer's recommendations. Following the protein concentration determination by measurement of absorbance at 280 nm on a nanodrop, the metal-labelled antibodies were diluted in Candor PBS Antibody Stabilization solution (Candor Bioscience, Germany) for long-term storage at 4°C."

Appendix B: RMS Illumina primer sequences for Index PCR

EcoRI forward (5’-3’):

1. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaGACTGCGTACCAATTC
2. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccGACTGCGTACCAATTC
3. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaGACTGCGTACCAATTC
4. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccGACTGCGTACCAATTC
5. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagGACTGCGTACCAATTC
6. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcGACTGCGTACCAATTC
7. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaGACTGCGTACCAATTC
8. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccGACTGCGTACCAATTC
9. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgGACTGCGTACCAATTC
10. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgGACTGCGTACCAATTC
11. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggGACTGCGTACCAATTC
12. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcGACTGCGTACCAATTC
13. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatGACTGCGTACCAATTC
14. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcGACTGCGTACCAATTC
15. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctGACTGCGTACCAATTC
16. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagGACTGCGTACCAATTC

MseI reverse (5’-3’):

1. caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

2. caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

3. caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

4. caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

5. caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

6. caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

7. caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

8. caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

9. caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

10. caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

11. caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

12. caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

13. caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

14. caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

15. caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

16.caagcagaagacggcatacgagatGGACGGgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

17. caagcagaagacggcatacgagatCTCTACgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

18.caagcagaagacggcatacgagatGCGGACgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

19. caagcagaagacggcatacgagatTTTCACgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

20. caagcagaagacggcatacgagatGGCCACgtgactggagttcagacgtgtgctcttccgatctGATGAGTCCTGAGTAA

Appendix C: Illumina HiSeq 3000 sequencing quality results

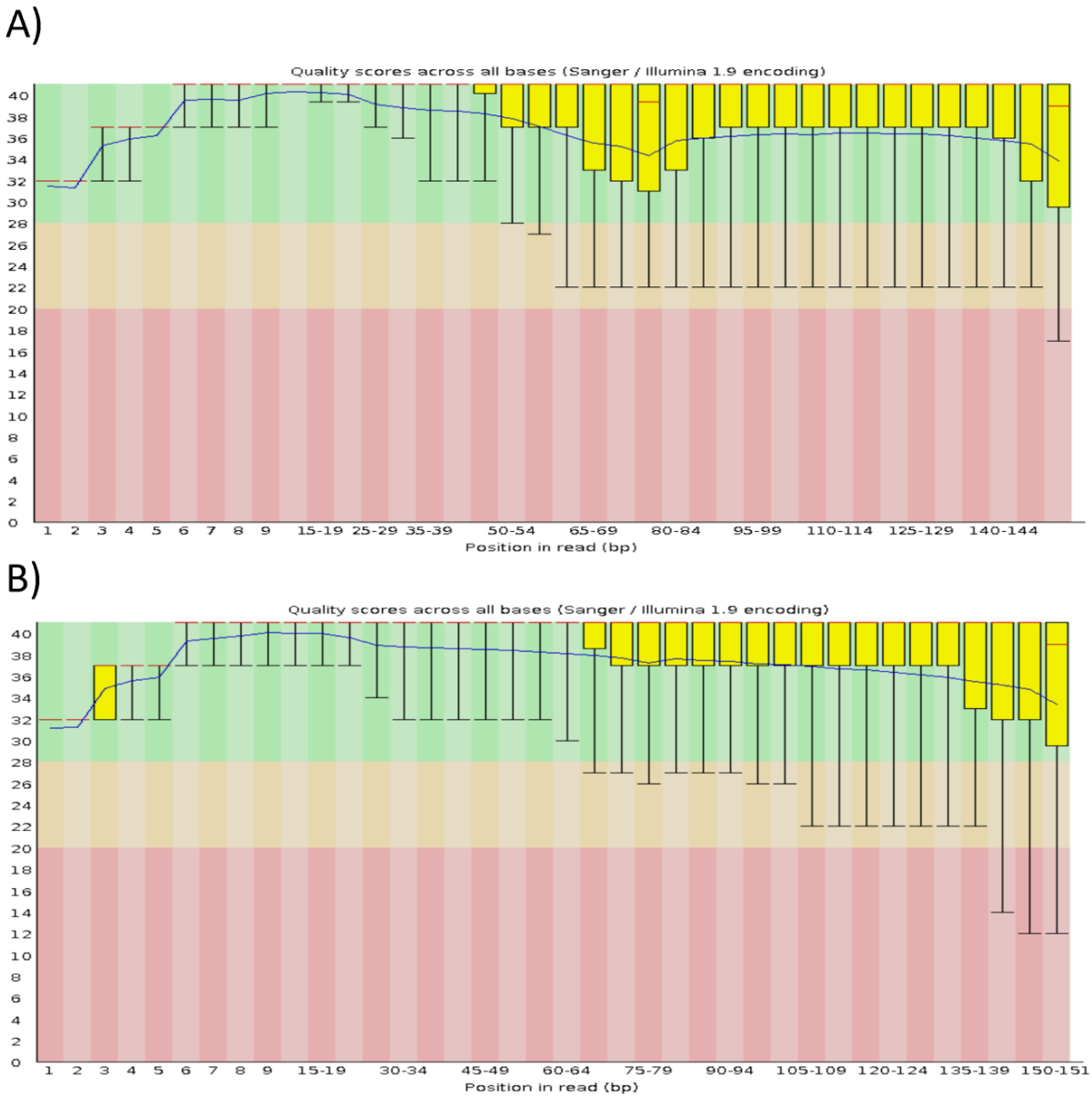

Figure C.1. **Per base sequence quality.** The graph shows per base sequence quality for A) reverse 2 from the sequencing run with mothers, meconium, 3 , and 6 months samples and B) reverse 7 for the 12 months samples. The red line is the median and the blue line is the mean of the quality values. The background of the graph is divided in three colour codes, green, yellow and red, associated with high, medium and low quality.

Appendix D: The relative longitudinal taxonomic composition at genus level

Table D.1. **The relative longitudinal taxonomic composition at genus level.** The table shows the average percentage and standard error of each bacterial genus at the time points meconium, 3, 6, and 12 months and mothers. The data is based on 60, 51, 51, 180 and 59 individual taxonomic profiles on the genus level, respectively. The nine first rows marked in bold are the bacterial genus with ≥5% abundance in at least one time point.

	Meconium (N=60)	**3 months (N=51)**	**6 months (N=51)**	**12 months (N=180)**	**Mothers (N=59)**
			Average ± SE (%)		
Alistipes	2 ± 2	0,43 ± 0,06	0,17 ± 0,02	2,9 ± 0,5	8,2 ± 0,4
Faecalibacterium	3 ± 3	0,38 ± 0,05	0,63 ± 0,08	9,2 ± 0,6	14 ± 1
Sutterella	2 ± 6	1,1 ± 0,1	3,4 ± 0,4	5,1 ± 0,4	3,7 ± 0,7
Veilonella	2 ± 2	7 ± 1	13 ± 2	4,5 ± 0,4	0,4 ± 0,6
Bifidobacterium	35 ± 19	55 ± 7	44 ± 6	16 ± 1	6 ± 2
Clostridium	5 ± 10	10 ± 1	6,5 ± 0,8	5,7 ± 0,4	7,1 ± 0,7
Streptococcus	6 ± 13	2,6 ± 0,3	1,5 ± 0,2	1,2 ± 0,1	0,7 ± 0,1
Prevotella	3 ± 3	0,62 ± 0,08	0,38 ± 0,05	4,4 ± 0,8	8,2 ± 0,6
Bacteroides	24 ± 21	8 ± 1	13 ± 2	25 ± 1	24 ± 3
Treponema	0,0007 ± 0,0008	9E-05 ± 1 E-05	0,00038 ± 5E-05	0,00075 ± 0,0001	0,0014 ± 0,0001
Campylobacter	0,02 ± 0,2	0,022 ± 0,003	0,018 ± 0,002	0,007 ± 0,001	0,0143 ± 0,0009
Neisseria	0,3 ± 0,5	0,027 ± 0,004	0,033 ± 0,004	0,041 ± 0,009	0,886 ± 0,005
Haemophilus	3 ± 13	0,63 ± 0,08	0,53 ± 0,07	0,8 ± 0,1	0,2 ± 0,1
Porphyromonas	0,06 ± 0,08	0,018 ± 0,002	0,018 ± 0,002	0,055 ± 0,009	0,078 ± 0,007
Roseburia	0,8 ± 0,8	1,2 ± 0,2	1,1 ± 0,1	3,5 ± 0,3	3,1 ± 0,4
Oxalobacter	0,02 ± 0,03	0,0013 ± 0,0002	0,0017 ± 0,0002	0,011 ± 0,001	0,029 ± 0,001
Fusobacterium	0,1 ± 0,3	0,10 ± 0,01	0,078 ± 0,01	0,063 ± 0,008	0,031 ± 0,008
Desulfovibrio	0,2 ± 0,3	0,058 ± 0,007	0,033 ± 0,004	0,07 ± 0,01	0,510 ± 0,009

Acidaminococcus	0,02 ± 0,03	0,0024 ± 0,003	0,0048 ± 0,0006	0,026 ± 0,009	0,044 ± 0,003
Megasphaera	0,2 ± 0,2	0,052 ± 0,007	0,27 ± 0,03	0,33 ± 0,03	0,09 ± 0,04
Peptostreptococcus	0,005 ± 0,01	0,013 ± 0,002	0,0041 ± 0,005	0,00081 ± 9E-05	0,0028 ± 0,0001
Ruminococcus	1 ± 1	2,0 ± 0,3	2,6 ± 0,3	4,4 ± 0,3	3,1 ± 0,6
Gemella	0,03 ± 0,1	0,009 ± 0,001	0,014 ± 0,002	0,0063 ± 0,0004	0,0021 ± 0,0008
Atopobium	0,1 ± 0,3	0,043 ± 0,006	0,030 ± 0,004	0,019 ± 0,002	0,020 ± 0,002
Bacillus	0,3 ± 2	0,0025 ± 0,0003	0,0091 ± 0,001	0,0018 ± 0,001	0,0066 ± 0,0002
Lactobacillus	0,4 ± 0,5	0,43 ± 0,06	0,17 ± 0,02	0,22 ± 0,05	0,34 ± 0,03
Actinomyces	0,4 ± 0,8	2,0 ± 0,1	0,36 ± 0,05	0,028 ± 0,002	0,036 ± 0,004
Corynebacterium	0,3 ± 1,0	0,010 ± 0,001	0,0056 ± 0,0007	0,0033 ± 0,0003	0,0030 ± 0,0004
Eubacterium	0,6 ± 0,6	0,051 ± 0,007	0,38 ± 0,05	1,6 ± 0,1	2,4 ± 0,2
Mycoplasma	0,05 ± 0,2	0,0037 ± 0,0005	0,0023 ± 0,0003	0,0049 ± 0,006	0,0700 ± 0,0006
Phascolarctobacterium	0,04 ± 0,04	0,020 ± 0,002	0,018 ± 0,002	0,03 ± 0,01	0,228 ± 0,004
Coprococcus	0,2 ± 0,2	0,051 ± 0,007	0,049 ± 0,006	0,16 ± 0,2	1,32 ± 0,02
Dialister	0,2 ± 0,3	0,029 ± 0,004	0,040 ± 0,005	0,49 ± 0,06	0,89 ± 0,06
Holdemania	0,004 ± 0,003	0,00014 ± 2E-05	0,00054 ± 7E-05	0,016 ± 0,004	0,025 ± 0,002
Eggerthella	0,05 ± 0,04	0,030 ± 0,004	0,09 ± 0,01	0,056 ± 0,005	0,260 ± 0,007
Coprobacillus	0,5 ± 0,6	2,3 ± 0,3	2,4 ± 0,3	1,9 ± 0,2	0,3 ± 0,2
Collinsella	3 ± 3	3,3 ± 0,4	3,2 ± 0,4	2,8 ± 0,3	2,8 ± 0,4
Granulicatella	0,02 ± 0,07	0,0056 ± 0,0007	0,023 ± 0,003	0,021 ± 0,002	0,0054 ± 0,003
Solobacterium	0,8 ± 2	0,18 ± 0,02	0,57 ± 0,07	0,56 ± 0,04	0,79 ± 0,07
Olsenella	0,5 ± 1,0	0,071 ± 0,009	0,072 ± 0,009	0,17 ± 0,01	0,15 ± 0,02
Catenibacterium	0,005 ± 0,003	0,0037 ± 0,0005	0,0028 ± 0,0004	0,011 ± 0,002	0,026 ± 0,001
Finegoldia	0,1 ± 0,4	0,040 ± 0,005	0,012 ± 0,002	0,0057 ± 0,0005	0,0138 ± 0,0007
Peptoniphilus	0,04 ± 0,07	0,022 ± 0,003	0,015 ± 0,002	0,0063 ± 0,0007	0,0116 ± 0,0008

Cetobacterium	0,0003 ± 0,001	2,1-05 ± 3E-06	3,0E-05 ± 4	5E-05 ± 2E-05	4E-05 ± 6E-05
Dorea	0,3 ± 0,3	0,52 ± 0,07	1,1 ± 0,1	0,70 ± 0,04	1,12 ± 0,09
Gemmiger	0,09 ± 0,08	0,016 ± 0,002	0,024 ± 0,003	0,30 ± 0,03	0,58 ± 0,04
Marvinbryantia	0,002 ± 0,002	0,0046 ± 0,0006	0,0028 ± 0,0004	0,0042 ± 0,0004	0,0101 ± 0,0005
Parabacteroides	1 ± 2	0,30 ± 0,04	0,69 ± 0,09	0,46 ± 0,04	1,24 ± 0,06
Aggregatibacter	0,05 ± 0,3	0,0018 ± 0,0002	0,0012 ± 0,0002	0,0080 ± 0,001	0,001 ± 0,001
Adlercreutzia	0,008 ± 0,01	0,00028 ± 4E-05	0,00012 ± 2E-05	0,003 ± 0,002	0,0420 ± 0,0004
Oscillibacter	0,2 ± 0,2	0,051 ± 0,007	0,061 ± 0,008	0,17 ± 0,01	0,91 ± 0,02
Parvimonas	0,003 ± 0,1	0,0017 ± 0,0002	0,0019 ± 0,0003	0,0005 ± 0,0001	0,00148 ± 6E-05
Blautia	1 ± 1	1,1 ± 0,1	1,4 ± 0,2	3,3 ± 0,2	3,3 ± 0,4
Parasutterella	0,01 ± 0,01	0,00041 ± 5E-05	0,00050 ± 6E-05	0,045 ± 0,006	0,063 ± 0,006
Butyricicoccus	0,2 ± 0,2	0,056 ± 0,007	0,045 ± 0,006	0,20 ± 0,03	0,53 ± 0,03
Succinatimonas	0,0253189	0,009 ± 0,001	0,015 ± 0,002	0,012 ± 0,002	0,007 ± 0,002
Negativicoccus	0,0002 ± 0,0009	0,0011 ± 0,0001	0,00028 ± 4E-05	0,0008 ± 0,0005	7E-05 ± 0,0001
Flavonifractor	0,2 ± 0,3	0,42 ± 0,06	1,3 ± 0,2	2,9 ± 0,4	0,5 ± 0,4
Pseudoflavonifractor	0,03 ± 0,03	0,011 ± 0,001	0,013 ± 0,002	0,08192103	0,20 ± 0,01
Lachnoclostridium	0,03 ± 0,06	0,025 ± 0,003	0,09 ± 0,01	0,25 ± 0,03	0,09 ± 0,03
Holdemanella	0,02 ± 0,02	0,0021 ± 0,0003	0,0061 ± 0,0008	0,018 ± 0,002	0,064 ± 0,002
Sanguibacteroides	0,008 ± 0,01	0,0032 ± 0,0004	0,0020 ± 0,0003	0,0015 ± 0,0002	0,0062 ± 0,0002
Hungatella	0,3 ± 0,5	0,56 ± 0,07	0,22 ± 0,03	0,8 ± 0,1	0,1 ± 0,1
Anaeromassilibacillus	0,02 ± 0,02	0,0066 ± 0,0009	0,016 ± 0,002	0,046 ± 0,006	0,081 ± 0,006
Massilimicrobiota	0,003 ± 0,004	0,0015 ± 0,0002	0,0010 ± 0,0001	0,007 ± 0,002	0,0010 ± 0,0009

Table D.2. **Paired t-test on taxonomic composition.** The table shows the p-value results after paired t-test on the development of the taxonomic composition on the genus level. The data used for this test was from meconium 3, 6 and 12 months and mothers, with the sample size 60, 59, 59, 180 and 59, respectively. The 9 genera with an average abundance ≥5% are listed in the table. Significant values with 5% level of significance are marked with *.

	Genus								
Time point transition	*Allistipes*	*Faecalibacterium*	*Sutterella*	*Veilonella*	*Bifidobacterium*	*Clostridium*	*Streptococcus*	*Prevotella*	*Bacteroides*
Meconium – 3 months	3,9E-06*	8,7E-12*	0,14	0,0036*	0,00021*	0,092	0,045*	2,7E-05*	1,2E-06
3 – 6 months	0,24	0,28	0,024*	0,020*	0,050*	0,32	0,031*	0,31	0,035*
6 – 12 months	8,4E-09*	6,9E-33*	0,086	1,5E-05*	1,4E-09*	0,63	0,20	4,6E-07*	1,2E-05*
12 months - Mothers	3,2E-09*	1,6E-05*	0,0033*	3,0E-18*	1,7E-12*	0,018*	0,00052*	0,061	0,77

Appendix E: The relative inferred longitudinal metabolic composition

Table E.1. **Relative inferred longitudinal metabolic composition**. The table shows the average percentage and standard error of each metabolite at the time points meconium, 3, 6, and 12 months and mothers. The data is based on 60, 51, 51, 180 and 59 individual metabolic profiles, respectively. The nine first rows in the first column are marked in bold and are the metabolites with ≥5% abundance in at least one time point.

	Meconium (N=60)	**3 months (N=51)**	**6 months (N=51)**	**12 months (N=180)**	**Mothers (N=59)**
			Average ± SE (%)		
Propionic acid	2,5 ± 0,3	2,2 ± 0,4	3,5 ± 0,3	6,4 ± 0,2	5,7 ± 0,3
Acetic acid	18,8 ± 0,5	19,0 ± 0,2	19,0 ± 0,4	18,1 ± 0,3	17,9 ± 0,3
Formic acid	11,2 ± 0,5	15,80 ± 0,5	13,8 ± 0,6	8,9 + 0,3	10,2 ± 0,3
Succinic acid	12,7 ± 0,6	15,9 ± 0,5	14,61 ± 0,5	12,3 ± 0,2	10,1 ± 0,3
Ethanol	13.9 ± 0,6	15,5 ± 0,6	13,6 ± 0,6	5,9 ± 0,3	5,7 ± 0,4
D-Lactic acid	8.7 ± 0,7	3,9 ± 0,5	4,9 ± 0,5	8,9 ± 0,2	10,7 ± 0,4
Butyric acid	2,8 ± 0,3	1,6 ± 0,4	3,1 ± 0,6	10,1 ± 0,3	15,7 ± 0,6
H_2	4,8 ± 0,7	4,0 ± 0,5	4,8 ± 0,5	6,4 ± 0,2	4,6 ± 0,2
L-Lactic acid	16,4 ± 0,6	17,5 ± 0,3	15,8 ± 0,6	7,6 ± 0,3	6,8 ± 0,4
(R)-Acetoin	1,0 ± 0,2	0,8 ± 0,2	0,50 ± 0,09	0,48 ± 0,05	0,33 ± 0,03
Indole	3,4 ± 0,5	1,7 ± 0,3	1,9 ± 0,3	3,1 ± 0,1	2,5 ± 0,2
Isovaleric acid	1,5 ± 0,2	0,7 ± 0,1	1,6 ± 0,3	4,9 ± 0,2	3,8 ± 0,2
Isobutyric acid	1,4 ± 0,2	0,63 ± 0,1	1,5 ± 0,2	4,6 ± 0,2	3,4 ± 0,2
(2R,3R)-2,3-Butanediol	0,17 ± 0,2	0,5 ± 0,2	0,22 ± 0,05	0,031 ± 0,006	0,05 ± 0,01
Phenylacetic acid	0,8 ± 0,1	0,39 ± 0,9	1,1 ± 0,2	2,6 ± 0,1	1,7 ± 0,2
1-Butanol	0,0049 ± 0,0006	0,002 ± 0,001	0,006 ± 0,002	0,022 ± 0,003	0.084 ± 0,006
Methane	0,05 ± 0,01	0.0014 ± 0,004	0,00012 ± 0,00003	0	0,8 ± 0,3

Table E.2. **Paired t-test on inferred metabolic composition.** The table shows the p-value results after paired t-test on the development of the metabolic composition that has been extrapolated from the taxonomic composition at the species level by the use of Virtual Metabolic Human database. The data used for this test was from meconium 3, 6 and 12 months and mothers, with the sample size 60, 59, 59, 180 and 59, respectively. The 9 metabolites with an average abundance ≥5% are listed in the table. Significant values with 5% level of significance are marked with *.

Time point transition	Metabolite								
	Propionic acid	Acetic acid	Formic acid	Succinic acid	Ethanol	D-Lactic acid	Butyric acid	H_2	L-Lactic acid
Meconium – 3 months	0,54	0,78	3,7E-08*	0,00012*	0,054	1,8E-07*	0,33	0,26	0,16
3 – 6 months	0,014*	0,97	0,020*	0,12	0,033*	0,16	0,059	0,28	0,023*
6 – 12 months	1,6E-09*	0,083	5,9E-10*	0,00017*	9,6E-18*	1,0E-11*	2,4E-14*	0,0065*	8,5E-19
12 months - Mothers	0,035*	0,47	0,0065*	1,3E-09*	0,99	2,0E-7*	4,3E-12*	7,6E-10*	0,060

Appendix F: Overview of immune cell composition at 12 months

Table F.1. **Immune cell composition**. The table shows the average percentage and standard error of each measured immune cell at 12 months of age. The data is based on 67 individual immune cell profiles.

12 months (N=67)			
Immune cell	**Average ± SE (%)**	**Immune cell**	**Average ± SE (%)**
IgD positive memory B cells	0,19 ± 0,02	Naïve CD8 T cells	7,73 ± 0,003
IgD negative memory B cells	1,00 ± 0,05	Eosinophils	2,37 ± 0,002
Naïve B cells	12,40 ± 0,04	Mucosal associated invariant T cells	0,2787 ± 0,0002
Transitional B cells	0,024 ± 0,002	Classical monocytes	6,147 ± 0,002
Plasmablasts	0,00090 ± 0,00005	Nonclassical monocytes	0,6546 ± 0,0008
Central memory CD4 T cells	1,00 ± 0,04	Proinflammatory monocytes	0,05209 ± 0,00003
Effector memory CD4 T cells	2,8 ± 0,1	CD56 bright natural killer cells	0,4713 ± 0,0002
Naïve CD4 T cells	20,9 ± 0,7	CD56dim natural killer cells	2,20 ± 0,001
Naïve T regulatory cells	1,6 ± 0,06	Neutrophils	30,79 ± 0,01
Memory T regulatory cells	0,55 ± 0,02	Others	4,340 ± 0,002
Activated $CD8^+$ T cells	0,2749 ± 0,0006	Basophils	0,4369 ± 0,0002
Central memory $CD8^+$ T cells	0,3227 ± 0,0002	CD161 positive gamma delta T cells	0,8590 ± 0,0006
Double positive T cells	0,0531 ± 0,00003	CD161 negative gamma delta T cells	0,7152 ± 0,0004
Effector memory CD8 T cells	1,73 ± 0,02	Plasmacytoid dendritic cells	0,16985 ± 0,00009

Appendix G: Spearman correlations

Specie	Immune cell (12 months)	P value (adj.)	Rho value
Bifidobacteriaceae bacterium NR020 (Mothers)	Classical monocytes	0,044	-0,50
Emergencia timonensis (Mothers)	Classical monocytes	0,044	-0,48
Clostridium sp. CAG508 (Mothers)	Classical monocytes	0,044	-0,48
Ruminococcaceae bacterium UBA2 (Mothers)	Double positive T cells	0,041	-0,51
Corpococcus sp. AF21 14LB (Mothers)	Translational B cells	0,045	-0,51
Streptococcus mutans (12 months)	Proinflammatory monocytes	0,024	-0,49
Uncultured Synergistaceae bacterium (6 months)	Naive CD8 T cells	0,013	0,56
Lactobacillus gasseri (6 months)	Naive CD8 T cells	0,036	0,52
Firmicutes bacterium CAG 176 (3 months)	Naive CD8 T cells	0,015	0,54
Clostridium sp. CAG 226 (3 months)	Naive CD8 T cells	0,028	0,51
Uncultured Synergistaceae bacterium (3 months)	Naive CD8 T cells	0,015	0,54
Lactobacillus gasseri (3 months)	Naive CD8 T cells	0,015	0,55

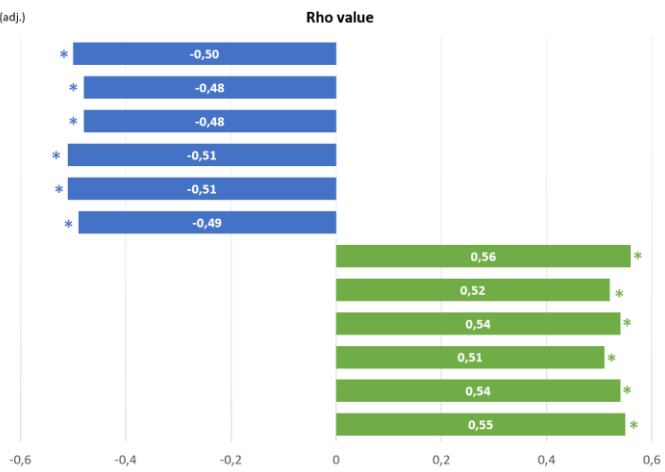

Figure G.1. **Spearman correlations between relative specie and immune cell composition.** The graph illustrates the results after spearman correlation between relative specie and immune cell composition, their associated time points, adjusted p-value and rho value. A total of 996 species and 28 immune cells were tested for correlation. All correlations in the graph are significant with a level of significance of 5% and are marked with *.

Genus	Immune cell (12 months)	P value (adj.)	Rho value
Eggerthella (3 months)	Central memory CD8 T cells	0,0037	0,54
Phascolarctobacterium (3 months)	Central memory CD4 T cells	0,027	0,48
Flaconifractor (Meconium)	Nonclassical monocytes	0,015	-0,42
Parasutterella (Meconium)	Nonclassical monocytes	0,040	-0,34
Oscillibacter (Meconium)	Nonclassical monocytes	0,036	-0,36
Gemminger (Meconium)	Nonclassical monocytes	0,021	-0,39
Dorea (Meconium)	Nonclassical monocytes	0,039	-0,34
Eggerthella (Meconium)	Nonclassical monocytes	0,036	-0,35
Holdemania (Meconium)	Nonclassical monocytes	0,015	-0,46
Phascolarctobacterium (Meconium)	Nonclassical monocytes	0,036	-0,35
Eubacterium (Meconium)	Nonclassical monocytes	0,036	-0,35
Lactobacillus (Meconium)	Nonclassical monocytes	0,015	-0,42
Clostridium (Meconium)	Nonclassical monocytes	0,040	-0,34
Ruminococcus (Meconium)	Nonclassical monocytes	0,022	-0,39
Acidaminococcus (Meconium)	Nonclassical monocytes	0,015	-0,43

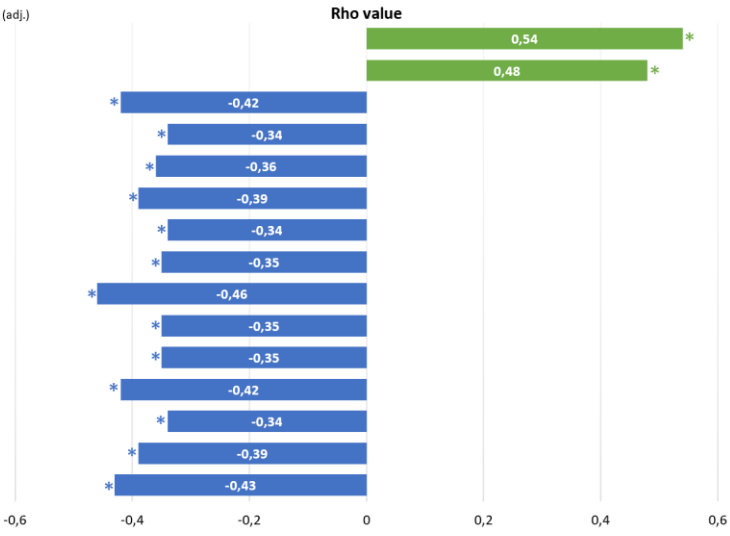

Figure G.2. **Spearman correlations between relative genus and immune cell composition**. The graph illustrates the results after spearman correlation between relative genus and immune cell composition, their associated time points, adjusted p-value and rho value. A total of 65 genera and 28 immune cells were tested for correlation. All correlations in the graph are significant with a level of significance of 5% and are marked with *.

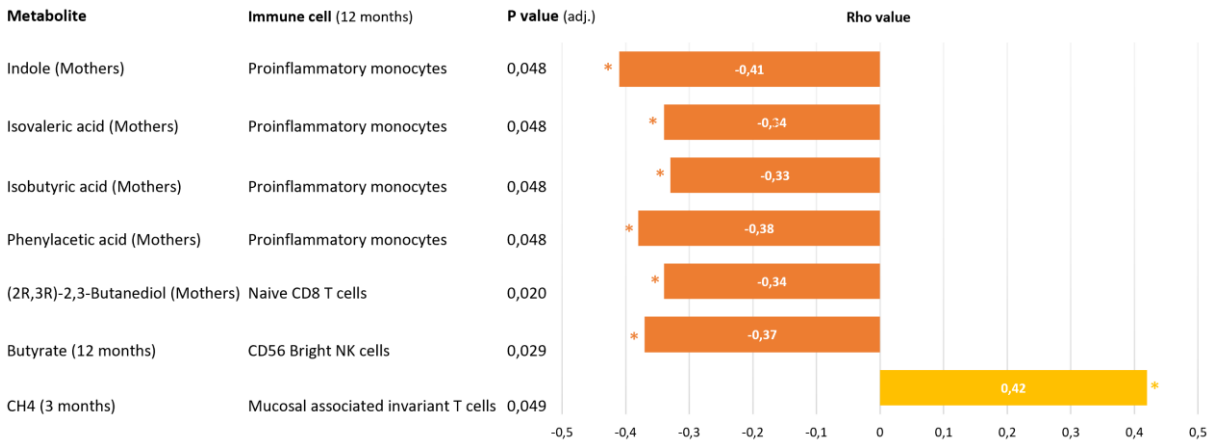

Figure G.3. **Spearman correlations between relative inferred metabolite and immune cell composition.** The graph illustrates the results after spearman correlation between relative metabolite and immune cell composition, their associated time points, adjusted p-value and rho value. A total of 17 metabolites and 28 immune cells were tested for correlation. All correlations in the graph are significant with a level of significance of 5% and are marked with *.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway