Norwegian University
of Life Sciences

Master's Thesis 2022 60 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Factors Associated with Host DNA Methylation in Infant Stool

Marianne Sundet Frøseth
MSc Biotechnology

Factors Associated with Host DNA Methylation in Infant Stool

Norwegian University of Life Sciences (NMBU),
Faculty of Chemistry, Biotechnology and Food Science

Acknowledgments

This thesis was performed at the Faculty of Chemistry, Biotechnology and Food Science at the Norwegian University of Life Sciences. The supervisors were Professor Knut Rudi and Ph.D. student Morten Nilsen. I would like to thank everyone that has helped me along the way to complete my master's degree.

First, I wish to express my deepest gratitude to my main supervisor Knut Rudi for the opportunity to write this thesis and for the guidance you have given me. You have taught me a lot, and I admire your ability to come up with new ideas. I am also extremely grateful for the help from my co-supervisor Morten Nilsen. You have been a great support during the entire process. Thank you for your patience and, in special, help with the data processing.

I would also like to thank all of you in the laboratory. The welcoming and including environment you have made is unique. And to my fellow master's students – I am so grateful for the friendship we have made throughout the years at NMBU.

Thanks to everyone in the PreventADALL study for providing the samples and including me at the symposium.

I am grateful to my better half, Espen, who has been a great support. I also owe thanks to our families – for all your help throughout this year. And of course, to my son, Håvard – for giving the best hugs.

Ås, May 2022

Marianne Sundet Frøseth

Abstract

DNA methylation plays an important role in the regulation of gene expression. There are several factors that can affect DNA methylation, with a complex interplay between the different mechanisms. The maturation of the infant gut microbiota could potentially be one of these factors. During the first year of life, there is shown to be a significant increase in the short chain fatty acid (SCFA) butyrate and a change in bacterial composition. How these changes potentially modulate the DNA methylation pattern is not well established. Therefore, the main aim of this thesis is to study if age, gut bacterial composition, and butyrate are associated with the DNA methylation levels of selected immune-related genes from feces.

Infant fecal samples from the study Preventing atopic dermatitis and allergies (PreventADALL) were used. As a significant increase in butyrate has been observed between 6- and 12-months of age, samples from these two time points were studied. To reveal the DNA methylation level of the selected genes, extracted DNA was bisulfite converted prior to Illumina sequencing. The taxonomic- and short chain fatty acid composition in feces were determined with 16S rRNA gene sequencing and gas chromatography, respectively. The results showed a significant increase in the DNA methylation level of defensin alpha 5 and toll-like receptor 4 with age. In contradiction, interleukin-4 showed a significant decrease in DNA methylation level with age. No significant correlations were found between butyrate and the mean DNA methylation levels, but potential trends were observed between DNA methylation and the bacterial composition.

In conclusion, the results indicate that age is a potential modulator of DNA methylation. The lack of correlations to butyrate indicates that the significant differences between the age groups are not caused by the increase in butyrate. However, trends between the DNA methylation pattern and bacteria suggest that the bacterial composition could be a potential modulator.

Sammendrag

DNA-metylering spiller en viktig rolle for regulering av genuttrykket. Det er flere faktorer som kan påvirke DNA-metylering, og det er et komplekst samspill mellom de ulike mekanismene. Modningen av tarmmikrobiotaen hos spedbarn kan potensielt være en av disse faktorene. I løpet av det første leveåret er det vist å være en signifikant økning av den kortkjeda fettsyren butyrat og en endring i den bakterielle sammensetningen. Hvordan disse endringene potensielt modulerer DNA-metyleringsmønsteret er ikke godt etablert. Hovedmålet med denne oppgaven er derfor å studere om alder, bakteriesammensetning i tarmen og butyrat er assosiert med DNA-metyleringsnivåene av utvalgte immun-relaterte gener fra avføring.

Avføringsprøver fra spedbarn ble hentet fra studien "Preventing atopic dermatitis and allergies (PreventADALL). På grunn av den tidligere observerte økningen av butyrat mellom 6- og 12-måneders alder ble prøver fra disse to tidspunktene studert. For å bestemme DNA-metyleringsnivåene i utvalgte gener ble ekstrahert DNA bisulfitt-konvertert før Illumina-sekvensering. Den taksonomiske sammensetningen og sammensetning av kortkjeda fettsyrer i avføringen ble bestemt med henholdsvis 16S rRNA gensekvensering og gasskromatografi. Resultatene viste en signifikant økning i DNA-metylerings-nivåene av defensin alfa 5 og toll-liknende reseptor 4 med alder. Interleukin-4 viste derimot en signifikant reduksjon i DNA-metyleringsnivå med alder. Ingen signifikante korrelasjoner ble funnet mellom butyrat og de gjennomsnittlige DNA-metyleringsnivåene, men potensielle trender mellom DNA-metylering og den bakterielle sammensetningen ble observert.

For å konkludere viser resultatene at alder er en potensiell modulator av DNA metylering. Mangelen på korrelasjoner til butyrat indikerer at de signifikante forskjellene mellom aldersgruppene ikke skyldes økningen av butyrat. Trendene som ble observert mellom DNA-metyleringsmønsteret og de observerte bakteriene indikerte derimot at den bakterielle sammensetningen kan være en potensiell modulator.

Abbreviations

ANOSIM	Analysis of similarities
ATP	Adenosine triphosphate
bp	Base pair
Caco-2	Cancer coli-2
ddNTP	Dideoxynucleoside triphosphates
DEFA5	Defensin alpha 5 (synonymous with human alpha defensin (*HD5*))
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotides
EC number	Enzyme commission number
EPO	Erythropoietin
EWAS	Epigenome-wide association study
FID	Flame ionization detector
GRCh38.p13	Genome Reference Consortium Human Build 38 patch 13.
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HDACi	Histone deacetylase inhibitor
HMO	Human milk oligosaccharide
HMT	Histone methyltransferase
IFN-γ	Interferon-gamma
Ig	Immunoglobulins
IL-4	Interleukin 4
K	Lysine (K is the amino acid code)
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO-identifiers	KEGG Orthology-identifiers
LPS	Lipopolysaccharides
MeCP	Methyl-CpG-binding proteins
MSA	Multiple sequence alignment
MUSCLE	Multiple Sequence Comparison by Log-Expectation

NCBI	National Center for Biotechnology
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PreventADALL	Preventing atopic dermatitis and allergies
PRR	Pattern recognition receptor
qPCR	Quantitative polymerase chain reaction
quasR	Quantify and annotate short reads in R
RDP	Ribosomal database project
RFU	Relative fluorescence units
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SAM	S-adenyl methionine
SCFA	Short chain fatty acids
Th1/2	T helper 1/2
TLR	Toll-like receptor
TNF-β	Tumor necrosis factor beta
VEGFA	Vascular endothelial growth factor A
zOTU	Zero-radius OTU

Table of contents

1. Introduction

1.1 Maturation of the infant gut microbiota

The infant gut undergoes several maturation processes during the first years of life, including changes in both the gut microbiota and the physical environment (Ximenez & Torres, 2017). There is a symbiotic relationship between the gut microbes and the host – a relationship most often being commensal but with the possibility of becoming mutualistic or parasitic depending on the environmental conditions (Belkaid & Hand, 2014). By an interplay with the immune system and diet, gut microbes contribute to avoiding allergies and other immune-related diseases (Belkaid & Hand, 2014). There is also an association between the microbiota and gene expression, but the mechanisms of regulation are unclear (Nichols & Davenport, 2021).

The human body is composed of approximately equal amounts of human and bacterial cells, most of which are located in the colon (Sender *et al.*, 2016). Bacterial colonization was first thought to start during birth (Milani *et al.*, 2017), and the infant has been shown to be sterile in utero (Rehbinder *et al.*, 2018). However, as reviewed by Milani *et al.* (2017), several studies argue that colonization starts before birth. The colonization depends upon several factors such as delivery mode, breastfeeding, use of antibiotics and environment which all contributes to a different composition of gut bacteria in each individual (Lozupone *et al.*, 2012).

Delivery mode is considered one of the main drivers of the first colonization. While infants born by cesarean section are colonized mainly by skin bacteria, the vaginally born are colonized by bacteria from their mother's vaginal flora (Dominguez-Bello *et al.*, 2010). In addition, vaginally born infants will be exposed to bacteria from the fecal microbiota of their mothers. There is shown that infants delivered by cesarean section acquire a lower abundance of Bacteroidetes than vaginally born infants (Del Chierico *et al.*, 2015). However, independent of delivery mode, the infant gut microbiota is dominated by bacteria belonging to Firmicutes and Proteobacteria during the first days (Del Chierico *et al.*, 2015). Proteobacteria are facultative bacteria and are thus depleted in the adult gut due to the strict anoxic environment (Eckburg *et al.*, 2005).

After birth, feeding mode contributes to the next transition of the gut microbiota. The introduction of solid foods is often recommended around six months of age, with slight variations across countries. The diet also varies with geographical location and differences in the taxonomic

composition have been shown, especially between western- and rural lifestyles (Yatsunenko *et al.*, 2012). As seen in the meta-analysis by Ho *et al.* (2018), breastfed infants are shown to have a lower alpha diversity of the gut microbiota compared to infants that got formula, medicines, or complementary feeding with solid foods. Breastmilk contains human milk oligosaccharides (HMOs), which are not digested by the infant itself but rather utilized as prebiotics for the gut bacteria (Gnoth *et al.*, 2000). *Bifidobacterium longum* subspecies *infantis* are known to utilize these HMOs and will therefore outgrow other bacteria in the gut microbiota of breastfed infants (Bode, 2015). Infants that are not exclusively breastfed are shown to have increased levels of Bacteroidetes and Firmicutes compared with exclusively breastfed infants (Ho *et al.*, 2018). This is a shift towards the adult gut microbiota, which is composed mainly of bacteria belonging to the phyla Firmicutes and Bacteroidetes (Eckburg *et al.*, 2005). As reviewed by Milani *et al.* (2017), the alpha diversity reaches adult levels when the child is approaching three years of age.

The gut bacteria have several beneficial functions for the host, including synthetization of vitamins K and -B and production of short chain fatty acids (SCFAs) (Ramakrishna, 2013).

1.2 Short chain fatty acids and the butyrate metabolism

In a process called fermentation, indigestible carbohydrates are broken down by bacteria in the large intestine to produce essential metabolites such as SCFAs (Blaak *et al.*, 2020). In return, the bacteria gain adenosine triphosphate (ATP), which is an energy source (Louis & Flint, 2009). The main SCFAs include acetate, propionate, and butyrate, which have shown several beneficial effects on gut health (Blaak *et al.*, 2020).

As reviewed by den Besten *et al.* (2013), acetate is the most abundant of the main SCFAs, which are present in an approximate ratio of 20:20:60. SCFAs are absorbed by the colonic mucosa (McNeil *et al.*, 1978), with butyrate being absorbed most rapidly and therefore being considered the main energy source (Clausen & Mortensen, 1995). Acetate is mainly taken up by the liver, while propionate is taken up by peripheral tissues such as the pancreas, brain, muscle, and adipose tissue (Morrison & Preston, 2016).

While acetate is produced by most gut bacteria, there is a smaller subset of propionate- and butyrate producers (Louis & Flint, 2017). Common propionate producers are bacteria belonging to Bacteroidetes, Negativicutes, and *Lachnospiraceae* (Louis & Flint, 2017). Butyrate producers are strictly anaerobic and gram-positive bacteria (Louis & Flint, 2009). The main producers are Firmicutes, and in special *Faecalibacterium prausnitzii*, *Eubacterium rectale* and *Roseburia* spp. (Louis & Flint, 2009). A significant increase in butyrate seen between 6- and 12-months of age has been found to be correlated with *Faecalibacterium prausnitzii* and *Eubacterium rectale* (Nilsen *et al.*, 2020).

There are several pathways for butyrate production, but the acetyl-CoA pathway is the most used (Vital *et al.*, 2014). A bacteria can be in possession of genes for more than one of the butyrate pathways (Vital *et al.*, 2014), and fermentative bacteria are also known for having the ability to use pathways with alternative end products (Louis & Flint, 2009). The activation of a pathway depends on the environmental conditions (Louis & Flint, 2009), and some pathways require different substrates such as amino acids (Vital *et al.*, 2014).

The acetyl-CoA pathway, which converts Acetyl-CoA to butyrate via several steps, is studied in butyrate-producing bacteria belonging to the class clostridia (Bennett & Rudolph, 1995). As reviewed by Bennett and Rudolph (1995), the different steps towards conversion to butyryl CoA are performed by the following enzymes. Acetyl-CoA is first converted to Acetoacetyl-CoA by acetyl-CoA acetyltransferase. Next, 3-hydroxybutyryl CoA is formed in a reaction catalyzed by hydroxybutyryl-CoA dehydrogenase. A dehydration reaction catalyzed by Enoyl-CoA hydratase gives crotonyl-CoA, which is further reduced to butyryl-CoA by Butyryl-CoA dehydrogenase. Two alternative enzymes can catalyze the conversion to the end product butyrate, namely acetate CoA-transferase and butyrate kinase (Vital *et al.*, 2014). It is shown that acetate CoA-transferase, which is encoded by the *but gene* dominates in healthy individuals, while butyrate kinase encoded by the *buk gene* is associated with ulcerative colitis (Vital *et al.*, 2013).

1.3 The gut epithelium

The gastrointestinal tract is separated from the body by only one layer of epithelial cells covered by a mucus layer for protection against both pathogenic and commensal bacteria, which are part of the gut microbiota (Schroeder, 2019). Enterocytes, which are absorptive cells, are the most

abundant epithelial cells (Noah *et al.*, 2011). In the colon, these epithelial cells are called colonocytes (Noah *et al.*, 2011). The colonocytes oxidize SCFAs as an energy source in a process that requires oxygen and thereby maintains anoxic conditions in the gut lumen (Litvak *et al.*, 2018). The SCFAs produced by the gut microbiota improve the gut barrier (Blaak *et al.*, 2020). In addition to the absorptive cells, there are several types of secretory epithelial cells. These include the mucus-producing goblet cells, small intestinal antimicrobial peptide-producing Paneth cells, Tuft cells, and enteroendocrine cells (Noah *et al.*, 2011). The gut epithelial cells have a short lifespan and are replaced every 3 to 5 days (Sender & Milo, 2021). New cells are produced by stem cells located in the crypts of the gut (Noah *et al.*, 2011).

1.4 The immune system

The human immune system is built up of innate- and adaptive immune responses, which are activated when the skin or mucosal barrier is overcome by an invading pathogen. Innate responses are unchanged every time a new pathogen is encountered, while adaptive responses are improved by each activation (Delves & Roitt, 2000). Pattern recognition receptors (PRRs) which are part of the innate immune system, aim to recognize patterns associated with pathogens and cell damage (Kawasaki & Kawai, 2014). The most well-studied PRR is the toll-like receptors (TLRs), which are able to recognize proteins, lipids, lipoproteins, and nucleic acids (Delves & Roitt, 2000). To eradicate a recognized intruder, cytokines and acute-phase proteins are produced by cells belonging to the innate immune defense (Delves & Roitt, 2000). Antimicrobial peptides, called defensins, are also a part of the innate immune system. They comprise α-defensins secreted by Paneth cells of the small intestinal epithelium and neutrophilic granules and β-defensins secreted by epithelial cells of mucosal surfaces (Hazlett & Wu, 2011). The adaptive immune response comprises B-cells which mature in the bone marrow, and T-cells which mature in the thymus. B-cells produce antibodies which are divided into different classes of immunoglobulins (Ig) based on their structure. This makes them able to recognize and bind to different antigens (Delves & Roitt, 2000). T-cells comprise both helper T-cells that activate immune responses and cytotoxic T-cells that kill infected cells (Delves & Roitt, 2000).

Helper T-cells comprise T helper 1 (Th1) cells that produce anti-inflammatory cytokines and T helper 2 (Th2) cells that produce cytokines associated with the IgE-response (Berger, 2000). There

is a paradigm that allergic disorders can be understood by the balance between Th1 and Th2 responses (figure 1.1), with an upregulation of the Th2 responses in atopy (Maggi, 1998). However, this balance is not yet established in newborns who have an immune system that is dominated by Th2 responses (Debock & Flamand, 2014). Cytokines produced by Th-2 cells include interleukin 4 (*IL-4*), and *IL-5*, while Th1-cells produce *IL-2*, interferon-gamma (*IFN-ɣ*), and tumor necrosis factor-beta (*TNF-β*) (Del Prete, 1992). As reviewed by Kidd (2003), the most dominant cytokines are *IFN-ɣ* for the Th1 response and *IL-4* for the Th2-response.

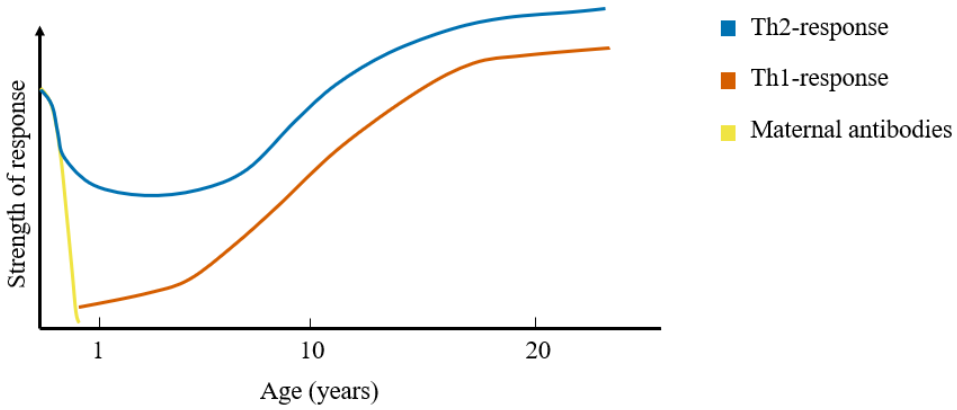

Figure 1.1. The balance between the Th1- and Th2-response. At the beginning of life, the Th2-response decreases while the Th1-response increases before both curves flatten out approximately after the age of 20 years. There is a higher level of maternal antibodies at the beginning of life with a rapid decrease during the first months. The figure is adapted and redrawn from Simon *et al.* (2015).

The contact between the gastrointestinal tract and the external environment of the gut lumen makes it an important part of the immune system (Takiishi *et al.*, 2017). A physical barrier is made by the connection of the epithelial cells by tight junctions, while the mucus layer works as a chemical barrier (Chelakkot *et al.*, 2018). The integrity of the barrier is important to avoid the intrusion of endotoxins such as lipopolysaccharides (LPS) from the microbiota (Chelakkot *et al.*, 2018). Exposure to pathogenic bacteria or their metabolites activates the immune system. To avoid a reaction upon exposure to commensal bacteria, they are recognized by TLRs on the epithelial cells – a function that is important to maintain homeostatic conditions in the gut (Rakoff-Nahoum *et al.*, 2004).

1.5 Epigenetic mechanisms

1.5.1 Epigenetic changes of nucleic acids

Heritable changes that alter gene expression but retain the deoxyribonucleic acid (DNA) sequence are called epigenetics (Wolffe & Matzke, 1999) and include mechanisms such as histone modifications, nucleosome positioning, and DNA methylation (Portela & Esteller, 2010). There is an interplay between the different mechanisms, and the impact on gene expression depends on this interplay (Portela & Esteller, 2010).

The most studied epigenetic mechanism is DNA methylation (Portela & Esteller, 2010). In general, DNA methylation of promoter regions leads to suppression of the transcription (Sarkar *et al.*, 2011). As reviewed by Moore *et al.* (2013), the important functions of DNA methylation include inactivation of one of the X-chromosomes in females, genomic imprinting, and repression of retroviral elements. In DNA methylation, methyl groups are transferred from S-adenyl methionine (SAM) and attached to the fifth carbon atom of cytosine by DNA methyltransferases (DNMTs), resulting in 5-methylcytosine (Moore *et al.*, 2013). DNMTs that catalyze this reaction include DNMT1, DNMT3a, and DNMT3b (Lyko, 2018). Positions in the human genome that are vulnerable to DNA methylation are CpG dinucleotides (Bird, 1980), which are made up of cytosine coupled to guanine by phosphate. CpG dinucleotides are highly present in CpG islands, which are defined as sequence stretches with an average length of 1000 base pairs (bp) and a higher occurrence of cytosine, guanine, and CpG dinucleotides, but with an absence of methylation (Deaton & Bird, 2011). The main sites for DNA methylation are the CpG island shores, which are the approximately two kilobases long area flanking the CpG islands (Irizarry *et al.*, 2009).

The negatively charged DNA is wound around nucleosomes that have positively charged lysine and arginine residues (Kujirai & Kurumizaka, 2020). Nucleosomes are made up of eight histone proteins, namely two of each of the following proteins; H2A, H2B, H3, and H4 (Luger *et al.*, 1997). Histone tails are prone to several post-transcriptional modifications, such as acetylation, methylation, ubiquitination, and phosphorylation (Portela & Esteller, 2010). These modifications can lead to looser or tighter packing of the DNA and thus affect gene expression (Moore *et al.*, 2013). Gene-upregulation or repression is dependent upon the location of the modification, for example, repression by methylation of H3K9 and H3K27 and upregulation by methylation of H3K4

(Hillyar *et al.*, 2020) (figure 1.2). Another example of an upregulating modification is acetylation of H3K27 (Hillyar *et al.*, 2020).

Histone acetylation is coupled to DNA methylation by histone deacetylases (HDACs) that are recruited to the promoter region when DNA is methylated (Sarkar *et al.*, 2011). The mechanism behind the recruitment involves methyl-CpG-binding proteins (MeCPs), which bind to methylated DNA (Lewis *et al.*, 1992). MeCP2 is known for the recruitment of HDACs and has been shown to function as a link between DNA methylation and histone methylation by associating with the histone methyltransferase (HMT) activity of H3K9 (Fuks *et al.*, 2003). There are 18 different HDACs divided into four groups (Seto & Yoshida, 2014). HDACs remove acetyl groups from lysine residues, an action that leads to tighter packing of DNA around histones (figure 1.2) (Park & Kim, 2020). The consequence of tighter packing of the DNA is lower accessibility for the ribonucleic acid (RNA) polymerase and thus lower transcription rates (Park & Kim, 2020).

1.5.2 Effects of butyrate on gene regulation

There are also factors that counteract the effects of HDACs and give increased transcription rates. Unlike HDAC, histone acetyltransferase (HAT) loosens the packing of DNA around histones by addition of acetyl groups that neutralize the positive charge of lysine residues (Park & Kim, 2020). This balance between HDAC and HAT, which is important for gene regulation, can be disturbed by inhibitors (Park & Kim, 2020). Butyrate is one of the HDAC inhibitors and gives an inhibition that leads to hyperacetylation of histones (Kruh, 1981). There is not one specific mechanism underlying the inhibition by HDAC inhibitors (Li & Sun, 2019), and even though the inhibition by butyrate is shown to be non-competitive, the mechanism is unknown (Davie, 2003).

The effect of butyrate on gene expression has been demonstrated on a colonocyte model in an earlier master thesis, showing indications of a transition in metabolism from glycolysis to β-oxidation in cells treated with butyrate (Utheim, 2021). The results suggested that butyrate had a greater impact on gene expression compared to acetate and propionate.

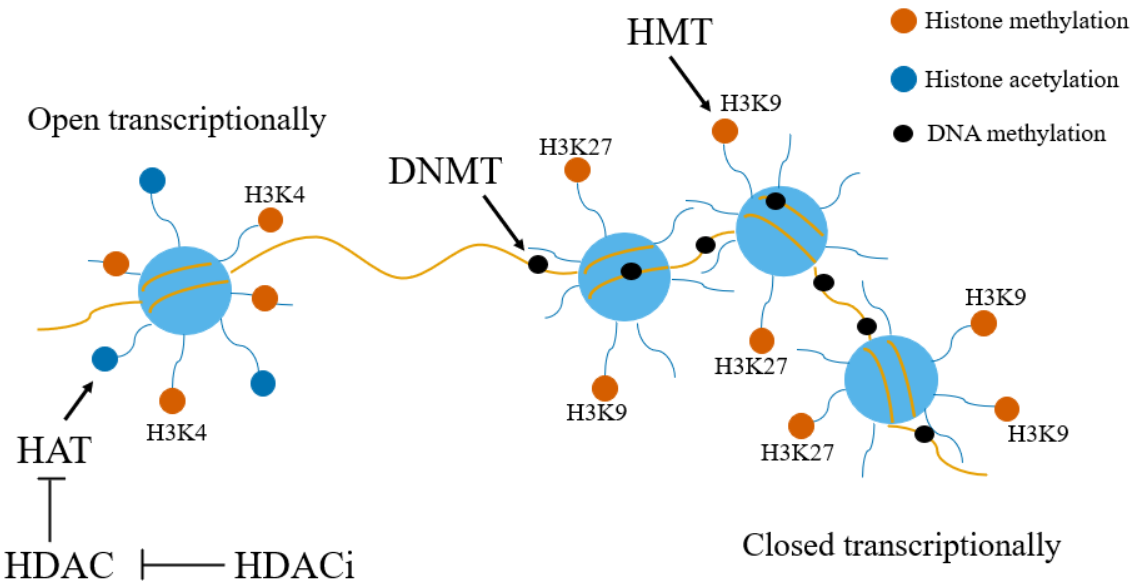

Figure 1.2. The effect of DNA methylation on chromatin structure. The figure is a simplification of the processes of acetylation and methylation of histones and DNA methylation. The blue circles represent histone octamers with histone tails, with the DNA strand wound around them. HAT acetylates histone tails, while HDAC inhibits the function of HAT. Acetylation of histones gives an open structure, while methylation leads to a closed structure. HDAC inhibitors (HDACi), for example, butyrate, inhibit the deacetylase function of HDACs. Cytosines are methylated by DNMTs, while histone tails are methylated by HMTs. The figure is redrawn and adapted from (Hillyar *et al.*, 2020).

1.6 Nucleic acid-based methods

1.6.1 Sequencing technologies

Sanger sequencing was developed in 1977 as one of the first methods to determine the DNA sequence (Sanger *et al.*, 1977). Originally, radiolabelled dideoxynucleotides (ddNTPs) were used, but the technique has been improved by the use of fluorescent ddNTPs to eliminate the need for four different vessels (Heather & Chain, 2016). The sanger technology requires a DNA polymerase that extends the template by adding deoxyribonucleotides (dNTPs) from the primer binding site on the template. As the first ddNTP is added, it will terminate the sequence because it lacks the hydroxyl group necessary for further addition of nucleotides (Sanger *et al.*, 1977). A lower ratio of ddNTPs than dNTPs gives a higher probability of adding a dNTP, and thus there will be fragments of different lengths. The resulting DNA fragments can be detected by capillary electrophoresis,

where fragments are separated by length, and the terminating ddNTP can be read to reveal the sequence (Heather & Chain, 2016).

Next-generation sequencing techniques were developed as a result of the limitations of Sanger sequencing (Muzzey *et al.*, 2015). These technologies are able to parallel sequence millions of fragments and thereby increase efficiency and the amount of data output (Hu *et al.*, 2021). Due to the short read length of next-generation sequencing, it is referred to as short-read sequencing (Hu *et al.*, 2021). There are several next-generation sequencing platforms, ranging from the first commercially available platform from 454 Roche to Ion Torrent produced by Life Technologies (Ambardar *et al.*, 2016) and a range of Illumina platforms such as MiSeq and NovaSeq 6000 (Hu *et al.*, 2021). Illumina is known for its low error rates and is one of the most prevalent technologies today (Hu *et al.*, 2021).

The preparation for Illumina sequencing includes fragmentation of DNA to a fragment length that is optimal for the sequencing platform (Hu *et al.*, 2021), followed by ligation of adapters. These adapters are complementary to oligonucleotides on the flow cell on which the library is loaded, and the fragments will thus bind to the oligonucleotides (Illumina, 2017). Bridge amplification of the fragments results in one cluster for each fragment which is used as a template for sequencing (Illumina, 2017). As an alternative to single-end sequencing, the fragments can be sequenced by paired-end sequencing, meaning they are sequenced from both ends (Hu *et al.*, 2021). The clusters are sequenced in parallel by addition of dNTPs labeled with fluorescent labels, which reveals the nucleotide order by base calling (Illumina, 2017).

Third-generation sequencing technologies have the advantage of read lengths longer than 10 kb and are therefore also referred to as long-read sequencing (Hu *et al.*, 2021). Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT) are the primary third-generation sequencing technologies both using native DNA as input and thus excluding the need for amplification of DNA (Hu *et al.*, 2021). One of the advantages of long reads is avoiding incorrect mapping, as could be a problem for short reads with, for instance, repetitive elements (Pollard *et al.*, 2018). Third-generation sequencing platforms also provide faster sequencing, but a disadvantage is the higher error rates compared to next-generation sequencing (Hu *et al.*, 2021).

1.6.2 Taxonomic markers

A common method to determine the taxonomic composition of the gut microbiota is sequencing of the 16S ribosomal RNA (rRNA) gene followed by division of the resulting sequences into operational taxonomic units (OTUs). Classification of bacteria based on the ribosomal RNA sequences was first proposed by Woese and Fox in 1977 (Woese & Fox, 1977). The about 1500 bp long 16S rRNA gene has a conserved primary structure separated by nine variable regions. Because it is present in all bacteria (Woese, 1987), it is a good reference gene. A threshold of 97 % sequence similarity of the 16S rRNA gene corresponds to a DNA similarity of 70 % or more and is often used to determine bacteria to the species level (Stackebrandt & Goebel, 1994).

1.6.3 Epigenetic markers

A method to reveal epigenetic patterns as histone modifications is chromatin immunoprecipitation with subsequent sequencing (ChIP-seq). This is done by fragmentation of DNA by nucleases followed by precipitation with antibodies that are specific to the modifications of interest (Furey, 2012). For determination of DNA methylation patterns, there are several available techniques, including both techniques to look at the whole genome and those that assess methylation in specific regions (Pajares *et al.*, 2021). Epigenome-wide association studies (EWAS) is a technique that is used to determine the DNA methylation state of genomes, with the Illumina 450k microarray being the most used (Wei *et al.*, 2021). The main methods to look at methylation in specific regions are based on either bisulfite conversion, restriction enzymes, or affinity enrichment, with bisulfite conversion-based methods being the most used (Pajares *et al.*, 2021).

The ability of bisulfite to convert cytosine to uracil by deamination was discovered by both Shapiro *et al.* (1970) and Hayatsu *et al.* (1970) in 1970. The first step in bisulfite conversion is denaturation of DNA due to the requirement of single-stranded DNA for bisulfite conversion (Shapiro *et al.*, 1973). Sodium bisulfite reacts with cytosines by adding to the 5,6 double bonds in a reversible reaction that is controlled by temperature, pH, and bisulfite concentration (Shapiro *et al.*, 1974). Removal of the amin group happens by hydrolysis, with an unknown effect of bisulfite as a catalyst, and uracil- SO3(-) is formed (Hayatsu, 2008). The final step, where removal of the SO3- group gives uracil, requires alkaline conditions (Hayatsu, 2008).

As 5-methylcytosines do not react with bisulfite, they are not converted to uracil (Hayatsu *et al.*, 1970). These findings were later utilized to determine the DNA methylation pattern as done by Frommer *et al.* (1992). Polymerase chain reaction (PCR) amplification of DNA fragments of bisulfite converted DNA gives further conversion of uracil to thymine, while 5-methylcytosine is converted to cytosines (figure 1.3). These fragments can be sequenced to find the methylation pattern (Frommer *et al.*, 1992), a technique known as bisulfite sequencing PCR (Pajares *et al.*, 2021).

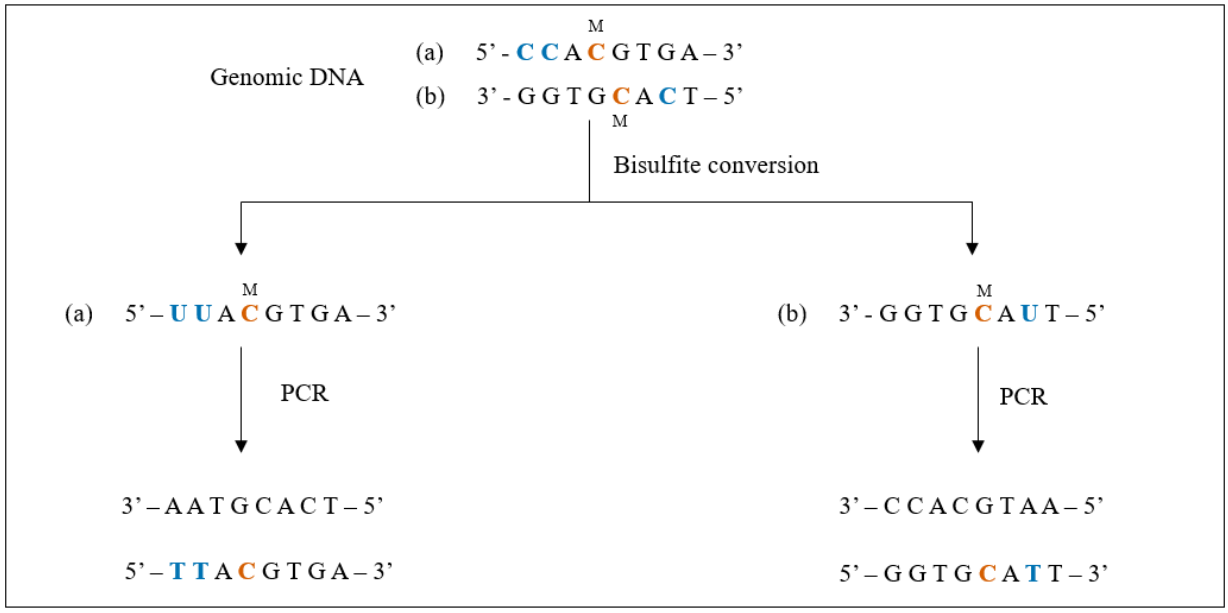

Figure 1.3. Bisulfite conversion and PCR amplification of genomic DNA. Genomic DNA is denatured into two strands that are no longer complementary, shown as strands (a) and (b). Unmethylated cytosines are converted to uracil during bisulfite conversion and are outlined in blue. Methylated cytosines remain unchanged and are outlined in orange. Uracils are further converted to thymine during PCR amplification. The figure is adapted and redrawn from Frommer *et al.* (1992).

As shown in figure 1.3, the two strands of genomic DNA cannot hybridize after bisulfite conversion because they are non-complementary. Two pairs of primers are therefore needed to amplify both strands of bisulfite-converted DNA (Frommer *et al.*, 1992). As illustrated in figure 1.4, the reverse primer is complementary to the bisulfite converted sequence and binds first. Then, the forward

primer is able to bind to the strand that is synthesized by the reverse primer, and the target sequenced is synthesized (ZymoResearch, n.d.-a).

Figure 1.4. Primer binding to bisulfite converted DNA. The reverse primer binds to the bisulfite converted DNA, and a new strand is synthesized (light blue strand). The forward primer binds to the newly synthesized strand, and a new strand is synthesized (yellow strand). The figure is made with inspiration from ZymoResearch (n.d.-a).

An alternative to bisulfite sequencing PCR is methylation-specific PCR. In this technique, the samples undergo PCR amplification in two rounds with primers targeting methylated DNA and unmethylated DNA to determine the methylation pattern in GC-rich regions (Herman *et al.*, 1996). Visualization of the amplified fragments on gel is used to distinguish between methylated and unmethylated samples (Herman *et al.*, 1996).

1.7 Determination of short chain fatty acid profiles

Gas chromatography is an analytical method for separation of compounds based on their chemical properties (Bartle & Myers, 2002). This is the most used method to determine the SCFA composition of fecal samples (Primec *et al.*, 2017). A carrier gas gives a mobile gas phase that carries the sample through a column that contains a stationary phase. Helium and hydrogen are the most used carrier gases and have replaced the use of nitrogen (Bartle & Myers, 2002). The column can be either a packed column that contains particles coated with the stationary phase or a capillary column that is coated with the stationary phase on the inside. As the capillary column is most efficient and requires lower temperatures than the packed columns, it is preferred (Bartle & Myers, 2002).

The interaction between the compounds in the sample and the stationary phase affects how the compounds are separated; thus, it can be controlled by the choice of column (Bartle & Myers, 2002). As the SCFAs are volatile (Primec *et al.*, 2017), they can be separated by non-polar

interactions (Bartle & Myers, 2002). The compounds that emerge from the column are measured by a detector, which is often the universal flame ionization detector (FID) (Bartle & Myers, 2002). Organic carbon is converted to ions in the flame and gives a current that can be measured and converted to a digital signal for analysis (Bartle & Myers, 2002).

1.8 PreventADALL

The diversity of the gut microbiota is one of the factors that are potential causes of noncommunicable diseases such as asthma and allergy (Lødrup Carlsen *et al.*, 2018). A study called preventing atopic dermatitis and allergies in children (PreventADALL) aims to find methods for preventing the development of these diseases which often start with atopic dermatitis in infancy (Lødrup Carlsen *et al.*, 2018). One of the secondary objectives of the PreventADALL study is to assess the role of the microbiota in the development of asthma and allergy (Lødrup Carlsen *et al.*, 2018).

The PreventADALL study cohort included 2397 mother and child pairs, mainly from Norway and Sweden. The majority were vaginally born and had a mean gestational age of 39.2 weeks. The data collected included both electronic questionnaires and biological samples such as feces, blood, urine, skin swabs, vernix caseosa, breast milk, and saliva (Lødrup Carlsen *et al.*, 2018).

1.9 Aim of thesis

The ability to determine the DNA methylation pattern from fecal samples is important to understand the regulation of gene expression in the infant gut. This is challenging due to the small amounts of human DNA in feces but has been accomplished in at least one earlier study (Klerk *et al.*, 2021). The hypothesis is that the maturation of the gut microbiota will lead to alterations of the DNA methylation pattern.

Therefore, the main aim of this work was to study if the factors age, gut bacterial composition, and butyrate affect the DNA methylation level of selected immune-related genes. The following subgoals were included:

- Determine the DNA methylation pattern for the selected genes
- Determine the taxonomic composition
- Study the associations between the DNA methylation pattern and the taxonomic composition
- Determine the SCFA composition to verify an increased butyrate level in the group of 12-month old infants
- Study the associations between the DNA methylation pattern and the SCFAs

2 Materials and methods

The main steps of the study included bisulfite conversion to determine the DNA methylation level of selected genes, 16S rRNA gene sequencing to determine the taxonomic composition, and gas chromatography to determine the SCFA levels in feces (figure 2.1).

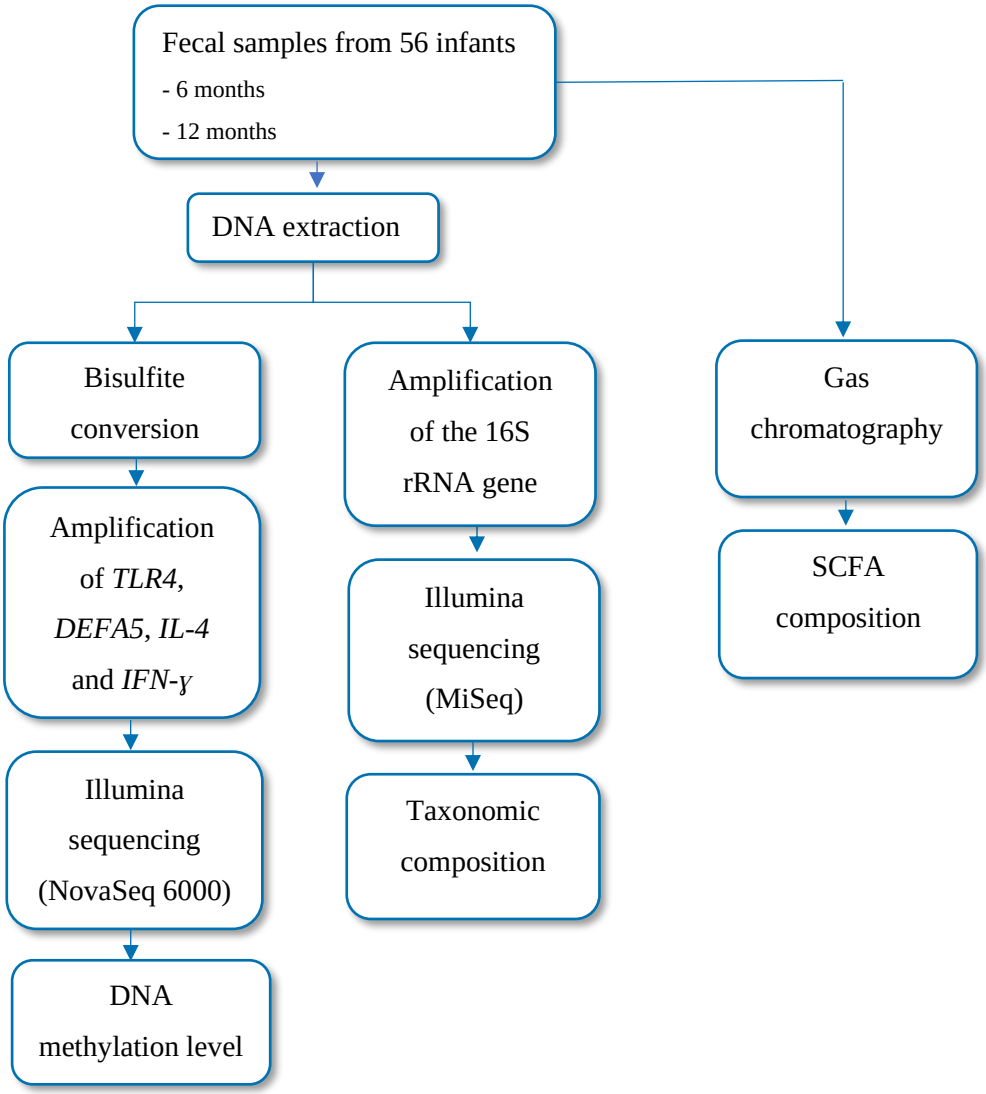

Figure 2.1. Flowchart of the work performed in the main study. The flowchart shows the main steps in the determination of the DNA methylation level, taxonomic composition, and SCFA composition. DNA was extracted from fecal samples originating from the time points 6 months and 12 months of 56 infants. Bisulfite conversion of the DNA followed by Illumina sequencing of specific genes was used to reveal the DNA methylation level. The taxonomic composition was revealed by 16S rRNA gene sequencing, and gas chromatography was used to determine the short chain fatty acid composition.

2.1 Sample preparation

2.1.1 Fecal samples

Fecal samples from 56 randomly selected infants taken from timepoints 6 months and 12 months were used (stored at -80 °C). These time points were chosen based on the earlier shown increase in butyrate production and butyrate-producing bacteria from 6- to 12 months (Nilsen *et al.*, 2020). For the pilot study of bisulfite conversion, samples from 4 infants taken at both 6- and 12 months of age were used. For the main study, all samples underwent bisulfite conversion, 16S rRNA gene sequencing and gas chromatography.

2.1.2 Cell lysis

Cell lysis and DNA extraction of the stool samples and positive controls were performed using the MagPure Stool DNA LQ kit (Angen Biotech, China) following the manufacturer's protocol. Negative controls with nuclease-free water were included. One hundred µl nuclease-free water was added to all samples consisting of 100-150 mg stool to ease the homogenization. A volume of 0.6 ml of both buffer ATL with PV-10 and buffer PCI were added to all samples for chemical lysis. For mechanical lysis of the cells, samples were added to 2 ml tubes prefilled with beads and run at FastPrep-96^{TM} (MP Biomedicals, USA) for 2×40 seconds at 1800 rpm. The samples were incubated at 65°C for 20 minutes to ensure complete lysis of the cells (AngenBiotech, n.d.) and centrifuged at 13 000 rpm for 5 minutes to separate the supernatant, which contained DNA.

2.1.3 DNA extraction

For the pilot study samples, DNA extraction was done manually with use of 400 µl supernatant from each sample. RNA was first removed from the samples by adding 10 µl RNase A. Twenty µl Proteinase K was added to degrade proteins in the samples (BioLabs, n.d.), in addition to 30 µl MagPure Particles N and 600 µl Buffer MLE. The samples were incubated with shaking at 1400 rpm for 10 minutes to let the DNA bind to the paramagnetic beads. The supernatant was removed by placing the tubes on a magnetic stand. To wash the beads, they were resuspended in 600 µl Buffer GW1, and the supernatant containing unwanted substances was removed. An additional wash with 600 µl 75 % ethanol was performed twice, and excess ethanol was removed. To elute DNA, 110 µl elution buffer was added to the tubes and incubated at 55 °C with shaking at 1200

rpm for 10 minutes. DNA from Cancer coli-2 (Caco-2) cells was also extracted manually and used as a positive control for the epigenetic methods. The Caco-2 cells were kindly provided by laboratory technician Karen Utheim.

For the main study samples, DNA was extracted with the program "MagPureStoolDNALQ" on the King Fisher Flex robot (ThermoFisher Scientific, USA) by the lab personnel. The DNA was bound to the paramagnetic beads in the sample plate, which was prepared with 20 µl proteinase K, 450 µl Buffer MLE, 25 µl MagPure Particles N, and 310 µl sample mix consisting of 300 µl sample and 10 µl RNase A. The beads were washed with three different wash plates, one with 600 µl Buffer GW1 and two with 600 µl 75 % ethanol before elution of the DNA with 120 µl elution buffer.

2.1.4 Quantification of DNA

The DNA was quantified with fluorescence measurements of DNA mixed with qubit working solution consisting of 1:200 ratio of reagent and dilution buffer from the Qubit™ dsDNA HS Assay Kit (Invitrogen, USA) following the manufacturer's recommendation. The fluorescent dye in the reagent binds specifically to DNA, which makes it emit fluorescence that can be measured by a fluorometer (ThermoFisherScientific, n.d.). A Varioskan Lux plate reader (ThermoFisher Scientific, USA) was used to quantify DNA by mixing two µl DNA sample with 70 µl qubit working solution (ThermoFisher Scientific, USA). The fluorescence was measured with an excitation wavelength of 497 nm and an emission wavelength of 525 nm, with a relative fluorescence unit (RFU) for each sample as output. To calculate the DNA concentrations, approximately 15 samples representing the range of RFUs were measured with the qubit fluorometer (Invitrogen, USA). A standard curve was made based on the measurements, and the standard curve equation was used to calculate the DNA concentrations in each sample.

2.2 Epigenetic methods

2.2.1 Bisulfite conversion of DNA

The extracted DNA was quantified and diluted to 20 µl with nuclease-free water to obtain the optimal DNA concentration of approximately 200-500 ng DNA for bisulfite conversion by the EZ DNA Methylation-Gold Kit (Zymo Research, USA). For the main study, the samples were

quantified by the plate reader, as explained in section 2.1.4. The concentrations ranged from 7 to 85 ng/µl and were diluted to concentrations between 144 and 500 ng/20 µl. The DNA concentrations for the pilot study were measured with the qubit fluorometer as explained in section 2.1.4 and diluted to 123-560 ng DNA per 20 µl sample.

The DNA was bisulfite converted by use of the EZ DNA Methylation-Gold Kit (Zymo Research, USA). A volume of 130 µl CT conversion reagent was mixed with 20 µl DNA in a PCR plate, with a positive and a negative control included for each plate. The samples were heated to 98 °C for 10 minutes, followed by 64 °C for 2.5 hours for denaturation of the DNA and addition of sodium bisulfite to unmethylated cytosines. The samples belonging to each PCR plate were split into two rounds for the spin-column protocol. To make the DNA bind to the silica-based matrix of the Zymo-Spin^{TM} IC Columns (ZymoResearch, n.d.-b) they were prepared with 600 µl M-Binding buffer. All following centrifugation steps were run at 13 000 rpm for 30 seconds. The samples were added to the columns and mixed by inverting the columns prior to centrifugation. One hundred µl of M-Wash Buffer was added to each column and centrifuged. To remove the sulphonate groups, the samples were incubated for 15-20 minutes with 200 µl M-Desulphonation Buffer and centrifuged again. A wash with 200 µl of M-Wash Buffer followed by centrifugation was performed twice. An additional centrifugation step was added after the last wash step for the 12-months samples to ensure sufficient removal of wash buffer. DNA was eluted by adding 10 µl M-Elution Buffer to the column matrix, followed by centrifugation. All samples were frozen at -20 °C before further processing.

2.2.2 Targeting of selected genes

The promoter regions of the genes *TLR4*, *defensin alpha 5* (*DEFA5*), *IL-4*, and *IFN-ɣ* were targeted for determination of the DNA methylation pattern of the CpG-positions. Methylation of the gene promoters of *DEFA5* and *TLR4*, which are part of the innate immune system, have been associated with the gut-disease necrotizing enterocolitis in earlier studies (Klerk *et al.*, 2021). *IL-4* and *IFN-ɣ* were chosen based on their importance for the Th1/Th2 immune response. In an earlier study, they showed opposite methylation patterns of the gene promoters (Berni Canani *et al.*, 2015).

PCRs were performed to amplify the genes of interest. Primers to amplify the promoter regions were taken from Klerk *et al.* (2021) for *TLR4* and *DEFA5* and Berni Canani *et al.* (2015) for *IL-4*

and *IFN-ɣ* (table 2.1). The reaction mixture consisted of 1x HOT FIREPol® DNA Polymerase (Solis BioDyne, Estonia), 0.2 µM forward and reverse primer, 2 µl template DNA, and nuclease-free water to a total volume of 25 µl.

The PCR program used for *DEFA5* and *TLR4* had an initial denaturation step of 15 minutes at 95 °C, followed by 50 cycles of 30s denaturation, 30s annealing, and 30s elongation at 94 °C, 58 °C, and 72 °C, respectively (Klerk *et al.*, 2021). As explained by Klerk *et al.* (2021), 50 cycles were used to compensate for the low amount of human DNA in feces. For *IL-4* and *INF-ɣ*, the same initial denaturation step was used, but the 50 cycles were adjusted to 95 °C for 10 seconds, 61 °C for 10 seconds, and 72 °C for 10 seconds as used by Berni Canani *et al.* (2015). Both programs had a final elongation step of 72 °C for 7 minutes. The promoter regions of the genes erythropoietin (*EPO*) and vascular endothelial growth factor A (*VEGFA*) were included in the pilot study but excluded from the main study due to the short amplicon length of *VEGFA* and discrepancies between the reverse primer of *EPO* and the human genome sequence. The Genome Reference Consortium Human Build 38 patch 13 (GRCh38.p13) from the human genome resources by National Center for Biotechnology (NCBI) were used for verification of correct primers.

Table 2.1. Primer sequences and amplicon length for the targeted genes. The primers for *DEFA5*, *TLR4*, *VEGFA*, and *EPO* are taken from Klerk *et al.* (2021), and the primers for *IL-4* and *IFN-ɣ* are taken from Berni Canani *et al.* (2015). The table is adapted from these articles.

Gene	Primer sequence (5' to 3')	Amplicon length (bp)
DEFA5 F	TAGGAGGTTGAGGTAGGAGAAA	179
DEFA5 R	ACATTATCCTTTAAT	
	TCCATCCATATTATC	
TLR4 F	GTTGAGGTTTATTTT	164
	TAGTTTTGTATGTG	
TLR4 R	AACCTCATTCTA CCTTACATACC	
IL-4 F	AGGTTAGGAGATGGAGATTATTTTG	102
IL-4 R	TAAAACTACAAACACCTACCACCAC	
IFN-ɣ F	GAGTTTTGTTTTGTTATTTAGGTTGG	124
IFN-ɣ R	AATACCTATAATCCCAACTACTC	
*VEGFA F**	GGGAGTAGGAAAGTGAGGT	83
*VEGFA R**	TTCCCCTACCCCCTTCAATAT	
*EPO F**	GGGGGTAGGGGTTGTTATTTGTATG	Unknown
*EPO R**	CCCAAACCTCCTACCCCTACTCTAACC	

**VEGFA* and *EPO* were excluded from the main study

2.2.3 Gel electrophoresis

Gel electrophoresis with a 2 % agarose gel was used to verify presence of the targeted genes prior to and after purification. All samples were run on gel after amplicon PCR, while only randomly selected samples were run on gel after purification. Two % agarose gels were made with 4 g agarose powder, 200 ml 1x Tris-Acetate-EDTA buffer and 8 µl PeqGreen (PeqLab, Germany). Purified samples were added 1x DNA loading dye (New England BioLabs, USA) prior to loading. Four µl 100 bp ladder (Solis Biodyne, Estonia) and 5 µl of each sample were loaded. The gels were run at 80 volts for 50 minutes prior to visualization on a Molecular Imager GelDoc^{TM} XR Imaging System (Bio-Rad, USA) with Quantity One 1-D Analysis Software (v4.6.7).

2.2.4 Manual ampure purification

Purification of the amplicon PCR products from bisulfite converted DNA was performed manually due to the high ampure concentration needed to keep the short fragments (102-179 bp). The amplicon PCR products were mixed with a 0.1 % solution of SpeedBeads^{TM} magnetic carboxylate modified particles (Merck, USA) in a ratio of 1:3.5 and incubated at room temperature for 5 minutes to let the DNA bind to carboxyl groups on the bead surface (Merck, n.d.). To remove the supernatant, the samples were placed on a magnet to let the magnetic beads bind to it. The beads were washed twice with ethanol and air-dried until complete evaporation of excess ethanol. This was done to remove dye and excess components in addition to primer-dimers that are generated during a PCR. Twenty-five µl nuclease-free water was added to break the bonds between the DNA and the beads and incubated for 2 minutes. The samples were placed on a magnet to elute approximately 18 µl of the DNA-containing supernatant.

2.2.5 Index PCR of gene fragments

Index primers for the genes *TLR4*, *DEFA5*, *IL-4*, and *IFN-γ* were designed by modification of 16S index primers and are shown in appendix A. The part of the primers targeting the 16S rRNA gene was removed, and forward- or reverse primer sequences (table 2.1) for each gene were added to make a binding site for the sequencing primers. In addition to the indexes that are used to identify each sample and thus allow for multiplexing of libraries, the index primers comprise sequences that are complementary to oligonucleotides on the flow cell (Illumina, 2017).

Five µl template DNA was mixed with 1x FIREPol (Solis BioDyne, Estonia), 0.2 µM forward primer (Appendix A), 0.2 µM reverse primer (Appendix A), and nuclease-free water to a total volume of 25 µl. Both index PCR programs had an initial denaturation step at 95 °C for 5 minutes. For *TLR4* and *DEFA5*, 15 cycles of 94 °C for 30 seconds, 58 °C for 1 minute, and 72 °C for 30 seconds were performed. For *IL-4* and *IFN-γ*, the program was adjusted to 15 cycles of 95 °C for 10 seconds, 61 °C for 30 seconds, and 72 °C for 10 seconds. Both programs had a final elongation step at 72 °C for 7 minutes. As quality control prior to sequencing, the controls and a selection of 32 samples were run on a 2 % agarose gel as described in section 2.2.3.

2.3 Taxonomic composition

2.3.1 Amplicon PCR of the 16S rRNA gene

16S rRNA gene sequencing was performed to determine the taxonomic composition of the fecal samples. To amplify the V3-V4 region of the 16S rRNA gene, an amplicon PCR was done with a reaction mixture consisting of 2 µl DNA, 0.2 µM forward- and reverse primer (Appendix B)(Yu *et al.*, 2005), 1x HOT FIREPol® DNA polymerase (Solis BioDyne, Estonia) and nuclease-free water to a total volume of 25 µl. The program used had an initial denaturation step at 95 °C for 15 minutes. The amplification was done in 25 cycles of 95 °C for 30 seconds, 55 °C for 30 seconds, and 72 °C for 45 seconds. The program ended with 72 °C for 7 minutes. All samples were run on a 1.5 % agarose gel for 35 minutes. The gels were prepared as described in section 2.2.3. Amplified DNA from *Pseudomonas aeruginosa* was kindly provided by laboratory technician Tonje Nilsen and used as a positive control.

2.3.2 Ampure purification by use of robot

The Biomek®3000 robot (Beckman Coulter™, USA) was used for cleanup of the amplicon PCR product and was programmed to perform the cleanup as explained. A 1:1 ratio of 0.1 % solution SpeedBeads™ magnetic carboxylate modified particles (Merck, USA) was added to ten µl of amplicon PCR product from each sample and incubated for 5 minutes at room temperature to let the nucleic acids bind to the beads (Merck, n.d.). To remove the supernatant, the beads were incubated at a magnet for 2 minutes. The beads were washed with 100 µl ethanol twice and air-dried to remove excess ethanol. Twenty µl of nuclease-free water was added for elution and incubated for 2 minutes to break the bonds between the nucleic acids and the magnetic beads. The samples were incubated on a magnet for 5 minutes to allow elution of 16 µl cleaned sample. To verify successful purification, nine of the samples were randomly selected and run for 35 minutes on a 1.5 % agarose gel that was prepared as described in section 2.2.3.

2.3.3 Index PCR of the 16S rRNA gene fragments

Each of the DNA samples was designated with specific Illumina index primers (Appendix B) by use of an Eppendorf epMotion 5070 robot (Eppendorf, Germany). Two µl template DNA from each sample were mixed with 0.2 µM forward- and reverse primer, 1x FIREPol (Solis BioDyne,

Estonia), and nuclease-free water to a total volume of 25 µl. The PCR program used had an initial denaturation step at 95 °C for 15 minutes. Ten cycles of denaturation, annealing, and elongation were run at 95 °C for 30 seconds, 55 °C for 1 minute, and 72 °C for 45 seconds, respectively. The program had a final elongation step at 72 °C for 7 minutes. All samples were run on a 1.5 % agarose gel for 35 minutes as described in section 2.2.3 to verify the presence of an approximately 500 bp long fragment.

2.4 Gas chromatography to determine the SCFA composition

Gas chromatography of the samples was done with a TRACE™ 1310 Gas Chromatograph (ThermoFisher Scientific, USA) by the laboratory technicians. To prepare the fecal samples, they were dissolved in a 1:10 volume of nuclease-free water and homogenized in a FastPrep 96 (MP Biomedicals, USA) for 40 seconds at 1800 rpm. Some of the samples were homogenized in a MagNA Lyser (Roche, Switzerland) at 2000 rpm for 20 seconds. The samples were spun down, and 300 µl of the supernatant was added to an equal volume of internal standard consisting of 0.4 % formic acid to lower the pH, in addition to 2000 µM 2-methylvaleric acid. A retention time that is known and differs from the fatty acids that are analyzed makes 2-methylvaleric acid a suitable internal standard (Mayhew & Gorbach, 1977). Additionally, 2-methylvaleric is not present in the gut and can thus be used to normalize the quantity of the SCFAs in the samples. The prepared samples were centrifuged at 13 000 rpm for 10 minutes, and 300 µl of the new supernatant were added to spin columns with a 0.2 µm filter. The spin columns were centrifuged at 10 000 rpm for 5 minutes to elute the liquid to use in the gas chromatography vials (Thermo Scientific, USA).

For each sample, 0.2 µl was injected by an autosampler into the liner, where it was evaporated by heat. The gas was further transferred to the fused-silica capillary column (Restek, USA) by split injections. A mix of 2000 µM acid of the various SCFAs present in feces and 0.2 % formic acid was used as external control. The output, given by the program Thermo Scientific™ Dionex™ Chromeleon™ 7 Chromatography Data System Version 7.2 SR4., showed the concentrations of acetate, propionate, butyrate, valerate, isobutyrate, and isovalerate in mmol/kg feces. See Appendix C for details regarding the gas chromatography.

2.5 Sequencing

2.5.1 Preparations for Sanger Sequencing

The bisulfite-converted DNA for the pilot study was sequenced with Sanger sequencing (Eurofins Genomics, Germany). For sequencing preparations, 5 µl bisulfite converted DNA was mixed with 5 µM forward- or reverse primer (table 2.1) to a final volume of 10 µl in separate tubes.

2.5.2 Preparations for Illumina Sequencing

The DNA concentration of the index PCR products of bisulfite converted DNA was measured with a Varioskan Lux plate reader (ThermoFisher Scientific, USA) as explained in section 2.1.4. To minimize the consequences of primer hopping, the 6-month samples and the 12-month samples were pooled in separate libraries. Primer hopping is a misassignment of indexes that can occur during storage of libraries (Illumina, 2018). The libraries were purified with a 3.5x concentration of 0.1 % solution of SpeedBeads^{TM} magnetic carboxylate modified particles (Merck, USA). Forty µl nuclease-free water was used for elution. The cleaned libraries were run on a 2 % agarose gel as described in section 2.2.3 and measured with the qubit fluorometer (Invitrogen, USA) as explained in section 2.1.4. Ten µl nuclease-free water was added to both libraries to obtain theoretical concentrations of 33 ng/µl for the 6-month library and 31 ng/µl for the 12-month library. The libraries were sent to Novogene for Illumina sequencing on the NovaSeq 6000 (Novogene, United Kingdom) with a 2 × 150 bp paired-end run.

Indexed 16S rRNA gene samples were normalized and pooled by the Biomek robot (Biomek®3000, Beckman Coulter^{TM}) based on the DNA concentrations. The pooled library was made with approximately 90 ng DNA from each sample and manually purified with a 4:5 ratio of 0.1 % solution of SpeedBeads^{TM} magnetic carboxylate modified particles (Merck, USA) to DNA, respectively. The purification was done as explained in section 2.2.4, but the volume of nuclease-free water used for elution was adjusted to 40 µl. To verify presence of the indexed 16S rRNA gene fragment and successful cleaning, the library was run on a 2 % agarose gel. The library was measured to a concentration of 66.9 ng/µl by qubit (Invitrogen, USA). A 1:2 dilution of the pooled library was frozen at -20 °C prior to sequencing.

Preparations of the library and sequencing were performed by the laboratory technicians. For quantification of the library, the KAPA library quantification kit for Illumina platforms (Roche,

Switzerland) was used. Tenfold dilutions (10^{-4} to 10^{-7}) of the library, 6 DNA standards, and the negative control were all used in duplicates. Twelve µl quantitative polymerase chain reaction (qPCR) mix consisting of KAPA SYBR FAST qPCR Master Mix and 10x primer premix, and 6 µl nuclease-free water were added to 2 µl of each sample and control. The qPCR program used for quantification had an initial step at 95°C for 5 minutes, followed by cycles of 30 seconds at 95°C and 45 seconds at 60°C. A melt curve analysis at 65°C to 95°C was run. The library was quantified to approximately 4 nM by the KAPA library quantification data analysis template.

The library was diluted with nuclease-free water and 20 % PhiX spike-in to a final concentration of 8pM. PhiX is a virus that is used to make a control library for sequencing runs. Samples with low diversity require a higher concentration of PhiX control to distinguish the clusters from each other (Illumina, 2016). The library was loaded on the Illumina MiSeq platform (Illumina, USA) for a 2 × 300 bp paired-end run.

2.6 Data processing

2.6.1 Determination of the DNA methylation level

Processing of the data in the FASTQ-file from Illumina MiSeq sequencing was done by Ph.D student Morten Nilsen. In total, 322 277 398 raw reads were demultiplexed to assign the reads to the correct sample prior to merging of forward- and reverse reads. Further processing was done with a R-package called quantify and annotate short reads in R (QuasR) (Gaidatzis *et al.*, 2015). Forward- and reverse primers were removed, and the fragments were filtered based on the length of each gene to remove primer dimers and incorrect reads. The cut-off lengths were set to 40 bases for *IL-4*, 100 bases for *TLR4*, 50 bases for *IFN-γ*, and 55 bases for *DEFA5*. All reads were aligned to the human genome sequence of their respective gene taken from GRCh38.p13 (table 2.2), and the number of cytosines and thymines in the CpG-positions were calculated.

Table 2.2. Human genome sequences taken from the Human Genome Resources at NCBI. The table shows the human genome sequences used as a comparison for the bisulfite converted DNA and are the sequences that correspond to the amplified gene fragments.

Gene	Human genome sequence (5' to 3')	Position in the human genome
IL-4	GCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGC	132672456:132672507
TLR4	AGTTTCTTCACAAGAAGGGGCGGGCCAAATTGTGTCCTGCAAAAACCTACATATCGAAGTCCTAACCCCTCTACCTCAGACTATGACTGTATATGGAGAGAGAGCCTTGAAAGA	117703707:117703820
DEFA5	GGCGTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCGGGATCGCACCACTGCACTCCAGCCTGGGCGACAGCAAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAGAAGGAAATTCTTCATTCAT	7058688:7058561
IFN-γ	AGTGCAGTGGCGCAATCTCGGCTCACTGCAAGCTCTGCCTCCCGGATTCACGCCATTCTCCCGCCTCAGCCTCCC	68161473:68161547

2.6.2 Taxonomic composition

A cluster density of 1200 K/mm^2 was obtained by the Illumina MiSeq sequencing, resulting in a FASTQ file with raw data including 6 462 222 reads. These were processed with USEARCH by Ph.D student Morten Nilsen to obtain an OTU table. Due to a drop in quality after the first 80 bases of the forward reads, they were excluded. The reverse reads were cut to a length of 220 bases to avoid the reduction in quality of longer reverse reads. Primer binding sites were removed, and quality filtration was done to remove reads with a mean expected error of more than 1. The reverse reads were further clustered in OTUs based on 97 % sequence similarity, and chimeric sequences were removed. The ribosomal database project (RDP) (version 18) (Maidak *et al.*, 1994) was used for taxonomy annotation of the OTUs, and the samples were rarefied to 10 000 reads per sample. The number of reads belonging to the negative controls ranged between 247 and 690 reads for the 6-month age group and 222 and 474 reads for the 12-month age group and were thus excluded

during rarefaction. To visualize the data, they were imported to Rstudio (version 4.1.2) for processing with the Phyloseq package (McMurdie & Holmes, 2013).

Prediction of function based on the taxonomic composition

The R-based tool "Tax4Fun2" was used to predict the function of the bacteria based on the taxonomy. All OTUs from the 16S rRNA gene sequencing data were first annotated taxonomy through a BLAST search against the reference database "Ref99NR", which has a similarity threshold of 99 % (Wemheuer *et al.*, 2020). The functional prediction gave predictions for both enzymes and pathways, which were further analyzed in the Kyoto Encyclopedia of Genes and Genomes (KEGG) with KEGG Orthology-identifiers (KO-identifiers) as input.

2.7 Statistical tests

Analysis of similarities (ANOSIM)

An ANOSIM was performed in Rstudio (version 4.1.2) to check for statistically significant differences in taxonomy between the 6-month and 12-month age groups. ANOSIM is a non-parametric test that is analogous to one-way ANOVA (Clarke, 1993). The analysis was done using a Bray-Curtis dissimilarity matrix with 9999 permutations. Bray-Curtis distance has been shown to be effective in determining ecological distance (Faith *et al.*, 1987). The test gives an R statistic between 0 and 1. If the null hypothesis, which states that the groups are equal, is true, the R statistic would be close to 0 (Clarke, 1993).

Indicator species analysis

An indicator species analysis was performed in Rstudio (version 4.1.2) to find the indicator species for the 6- and 12-month groups. As reviewed by Cáceres et al., some species can be used as indicator species based on their niche preferences (De Cáceres *et al.*, 2010). A point-biserial correlation was done with 9999 permutations.

Spearman's rho

The non-parametric test Spearman's rho was used to find correlations between the SCFAs, the taxonomic composition, the DNA methylation level and the predicted function. Correlations are given a value between -1 and 1, with negative values indicating negative correlations and positive

values indicating positive correlations. The correlation is stronger the closer it is to 1. All tests were performed in Rstudio (version 4.1.2).

Correction for multiple testing

A Benjamini Hochberg test was performed in Excel to correct for multiple testing following the indicator species analysis. The method aims to avoid type I errors by controlling the false discovery rate (Benjamini & Hochberg, 1995). All Spearman's rho correlations were corrected with Holm's method.

Students t-test

The student's t-test was used to find significant differences in SCFA- and taxonomic composition between the 6- and 12-month groups. The t-test is a parametric test that is used to compare the mean between two groups.

Wilcoxon signed-rank test

The Wilcoxon signed-rank test was used to determine significant differences in DNA methylation patterns between the 6-month age group and the 12-month age group. It was also used to determine the significant differences between the 6-month age group and the 12-month age group for each enzyme in the butyrate metabolism provided by Tax4Fun. For enzymes with several KO-identifiers with the correct enzyme commission number (EC number), all KO-identifiers were included in the test.

3 Results

3.1 The DNA methylation pattern

3.1.1 Selection of genes

The genes *TLR4*, *DEFA5*, *IL-4*, *IFN-ɣ*, *VEGFA*, and *EPO*, were included in the pilot study. Gel electrophoresis of amplified and purified gene fragments of bisulfite converted DNA showed bands slightly above 100 bp for all samples and for the positive controls for *IL-4* and *IFN-ɣ* (figure 3.1, left panel). A weak smear was seen for one of the samples belonging to *TLR4* (indicated with an arrow in figure 3.1, right panel), and the positive control showed a band of approximately 150 bp. For *DEFA5*, a band at approximately 180 bp was shown for the positive control. A weaker band at the same length was seen for one of the samples. For *EPO*, a band at approximately 140 bp was seen for the positive control, while no bands were seen for the samples. Very weak bands slightly lower than 100 bp were seen for both samples and the positive control belonging to *VEGFA*. The gel electrophoresis results for the main study samples are summarized in appendix G.

Sanger sequencing of amplified and purified *TLR4* gene fragments from conventional DNA followed by a BLAST-search of the sequencing results verified that the correct fragment was successfully targeted. The sequence was used for comparison against bisulfite-converted DNA and gave indications of successful bisulfite conversion. BLAST searches of the sequences from Sanger sequencing of bisulfite converted DNA did not give any results.

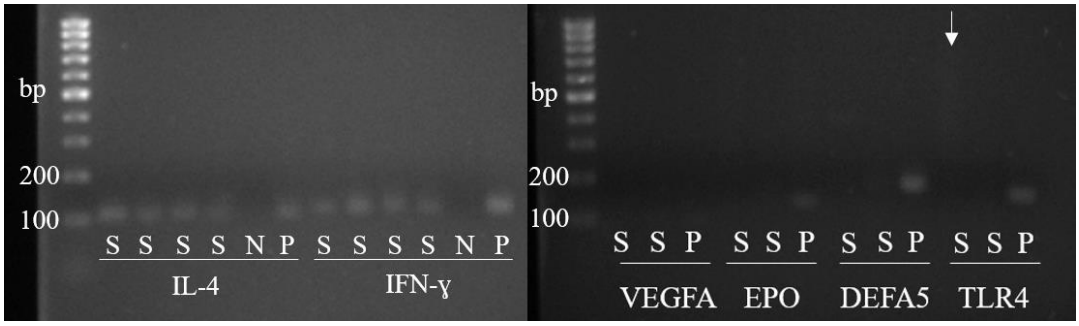

Figure 3.1. Gel electrophoresis of bisulfite converted DNA, targeting the gene fragments of *IL-4*, *IFN-ɣ*, *VEGFA*, *EPO*, *DEFA5* and *TLR4* in the pilot study. The amplified gene fragments were purified prior to gel electrophoresis. The white arrow points towards a weak smear belonging to the first sample of *TLR4*. Samples are marked with an S, and positive- and negative controls are marked with P and N, respectively. A 100 bp ladder was loaded in the first well of both gels.

3.1.2 Differences in the DNA methylation level between the 6-month and 12-month age groups

The DNA methylation level was calculated from the number of cytosines and thymines found by QuasR. A significantly higher mean DNA methylation level was found in the 12-month age group compared to the 6-month age group ($p = 0.01$, Wilcoxon signed-rank test). The mean DNA methylation level was 60.83 % in the 6-month age group compared to 64.89 % in the 12-month age group (figure 3.2). Due to missing sequencing results for the samples belonging to one of the reverse primers, 8 samples were excluded from the 6-month age group of both *IL-4* and *TLR4*.

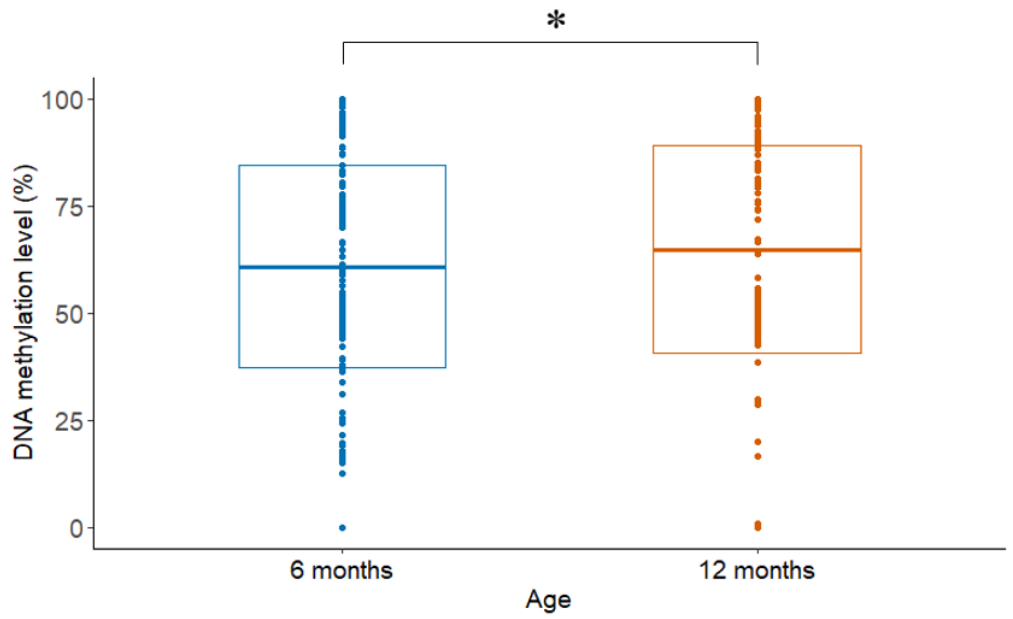

Figure 3.2. The total DNA methylation level shown in percentage for the 6-month age group and 12-month age group. Each dot in the strip chart represents the mean DNA methylation level of all CpG-positions in one gene for one sample. The thick middle line in each box is the mean, and the top and bottom lines are the standard deviation times two. A significantly higher DNA methylation level was seen in the 12-month age group ($p < 0.05$, Wilcoxon signed-rank test), marked with an asterisk in the figure. The code used to make the strip chart in R studio is shown in appendix E.

The DNA methylation level for each gene showed a significant difference between the 6-month age group and the 12-month age group for *TLR4* ($p = 0.01$, Wilcoxon signed-rank test), *IL-4* ($p = 0.04$, Wilcoxon signed-rank test) and *DEFA5* ($p = 0.01$, Wilcoxon signed-rank test). For both *DEFA5* and *TLR4*, there was a higher DNA methylation level for the 12-month group, unlike *IL-*

4, which showed a higher DNA methylation level for the 6-month group (figure 3.3). There was a slightly higher, but non-significant, DNA methylation level in the 12-month age group of *IFN-γ* (p = 0.87, Wilcoxon signed-rank test).

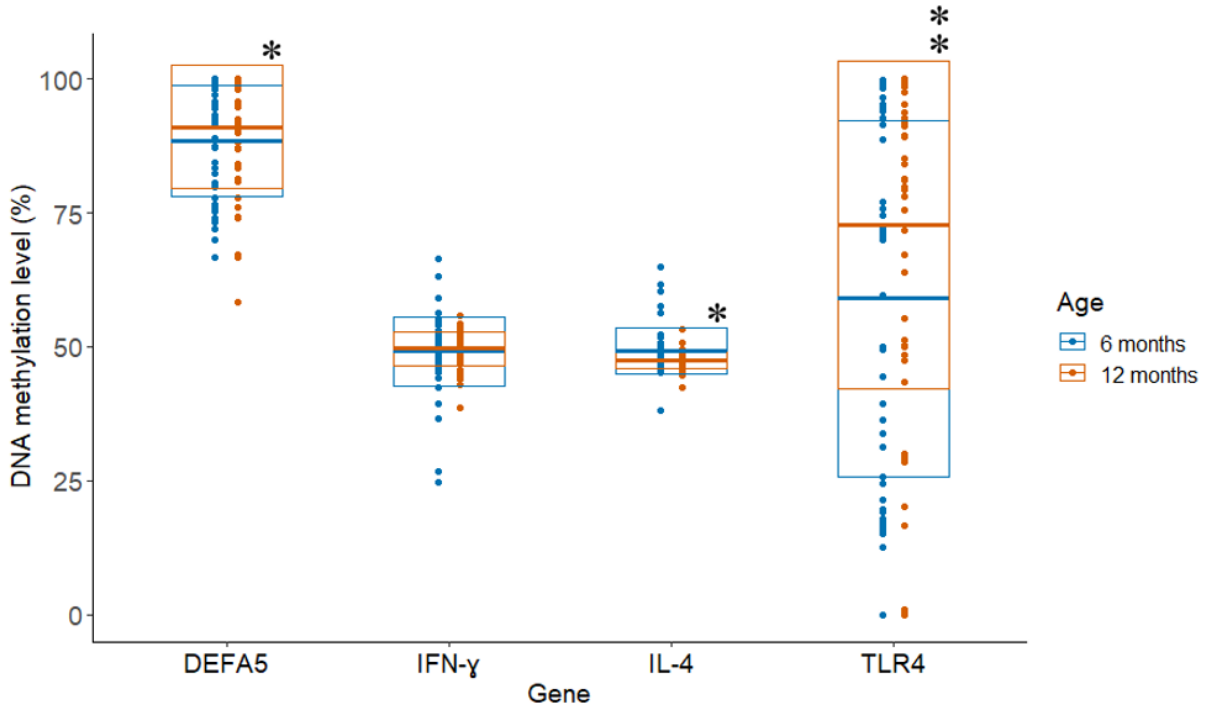

Figure 3.3. The mean DNA methylation level of each gene shown in percentage for the 6-month age group and the 12-month age group. Each dot in the strip chart represents the mean DNA methylation level of all CpG-positions for one sample. The thick middle line in each box plot is the mean, and the top and bottom lines are the standard deviation times two. The genes are separated as indicated on the x-axis. Genes with a significant difference in DNA methylation level between the 6-month age group and the 12-month age group are marked with asterisks in the figure; 1 asterisk for a p-value < 0.05 and 2 asterisks for a p-value < 0.01. A significant difference between the 6-month age group and the 12-month age group was shown for *DEFA5*, *IL-4* (p < 0.05, Wilcoxon signed-rank test), and *TLR4* (p < 0.01, Wilcoxon signed-rank test). The code used to make the strip chart in R studio is shown in appendix E.

Four of the individual CpG-positions showed a significant difference in the DNA methylation level between the 6-month age group and the 12-month age group (figure 3.4). For CpG-position 1 of *TLR4*, there was a significantly higher DNA methylation level in the 12-month age group with 72.74 % DNA methylation compared to 59.06 % DNA methylation in the 6-month group (p =

7.28e-04, Wilcoxon signed-rank test) (figure 3.4A). Even though the difference was not significant for CpG-position 2 of *TLR4* (p = 0.08, Wilcoxon signed-rank test), there was a higher DNA methylation level for the 12-month group for both CpG-positions (figure 3.4A).

CpG-position 5 of *DEFA5* also showed a significantly higher DNA methylation level for the 12-month group with 91.46 % DNA methylation compared to 88.66 % DNA methylation in the 6-month age group (p = 0.01, Wilcoxon signed-rank test) (figure 3.4C). The DNA methylation level of *DEFA5* was higher in the 12-month group for 4 of the in total 6 CpG-positions. However, CpG-position 2 and CpG-position 4 showed a higher DNA methylation level for the 6-months group (figure 3.4C).

For *IL-4*, there was a significantly higher DNA methylation level in the 6-month age group for both CpG-position 1 (p = 0.01, Wilcoxon signed-rank test) and CpG-position 2 (p = 0.01, Wilcoxon signed-rank test) (figure 3.4D), while the mean DNA methylation level of CpG position 3 was approximately equal in both age groups. There were no significant differences for any of the CpG-positions of *IFN-ɣ*, but a slightly higher DNA methylation level for the 12-month age group was seen in CpG-position 1, CpG-position 2, and CpG-position 4 (figure 3.4B).

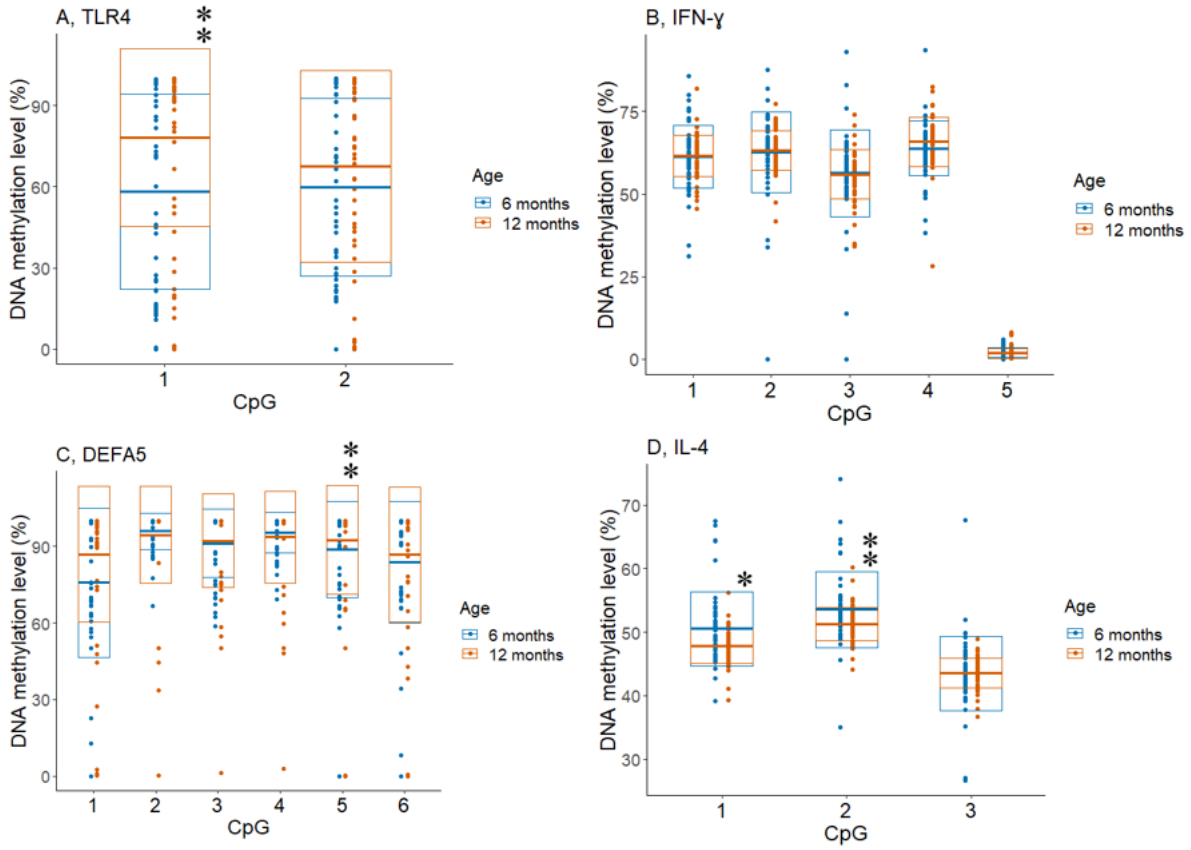

Figure 3.4. DNA methylation level of each CpG-position of (A) *TLR4*, (B) *IFN-γ*, (C) *DEFA5*, and (D) *IL-4* shown in percent for the 6-month age group and the 12-month age group. Each dot in the strip chart represents the DNA methylation level of one CpG-position for one sample. The thick middle line in each box plot is the mean, and the top and bottom lines are the standard deviation times two. CpG-positions with a significant difference between the 6-month age group and the 12-month age group are marked with asterisks on top of the box plot; one asterisk indicates a p-value < 0.05, and two asterisks indicate a p-value < 0.01. A significant difference in DNA methylation level between the 6-month age group and the 12-month age group was seen for *TLR4* CpG-position 1 ($p < 0.01$, Wilcoxon signed-rank test), *DEFA5* CpG-position 5 ($p < 0.01$, Wilcoxon signed-rank test), *IL-4* CpG-position 1 ($p < 0.05$, Wilcoxon signed-rank test) and *IL-4* CpG-position 2 ($p < 0.01$, Wilcoxon signed-rank test). The code used to make the strip chart in R studio is shown in appendix E.

3.1.3 Correlations in the DNA methylation level within and between genes

The mean DNA methylation level of *DEFA5* and *TLR4* was positively correlated to each other (rho = 0.51, adjusted $p < 0.01$, Holm's method) (Appendix D, figure D.2) in the 6-month age group, as the only correlation found in the mean DNA methylation level between genes.

Within each gene, there were positive correlations in the DNA methylation level between the CpG-positions of *TLR4* and *DEFA5*. There was a strong positive correlation between CpG-position 1 and CpG-position 2 of *TLR4* in the 6-month age group (rho = 0.80, adjusted p < 1e-04, Holm's method) (Appendix D, figure D.1A). For *DEFA5*, all CpG-positions except CpG-position 1 were positively correlated to each other in the 6-month age group (Appendix D, figure D.1A).

There were also some non-significant correlations in DNA methylation level between the individual CpG-positions between genes. For *IL-4* and *IFN-γ*, CpG-position 5 of *IFN-γ* were non-significantly correlated to CpG-position 1 (rho = -0.39, p = 0.01, adjusted p = 1, Holm's method) and CpG-position 2 (rho = -0.42, p = 0.01, adjusted p = 1, Holm's method) of *IL-4* in the 6-month age group (Appendix D, figure D.1A).

3.2 The SCFA composition and its correlation to the DNA methylation level

The SCFA composition, as decided by gas chromatography, showed an increase in the total amount of SCFAs from 90.41 mmol/kg feces in the 6-month group to 118.95 mmol/kg feces in the 12-month group (figure 3.5).

Acetate, which was the most abundant SCFA, showed a non-significant increase in acetate from 59.81 mmol/kg feces in the 6-month group to 69.48 mmol/kg feces in the 12-month group (figure 3.5) (p = 0.22, paired t-test). No significant correlations were found between acetate and the mean DNA methylation level. However, there was a non-significant negative correlation between acetate and CpG-position 1 of *IFN-γ* in the 12-month age group (rho = -0.29, p = 0.04 adjusted p = 1) decided with Spearman's correlation and corrected for multiple testing with Holm's method (Appendix D, figure D.3B).

There was a significant increase in the observed concentration of butyrate (p = 3.17e-08, paired t-test) from the 6-month to the 12-month group, with an increase in mean butyrate concentration from 8.43 mmol/kg feces to 19.96 mmol/kg feces (figure 3.5). There were big interindividual variations in the observed butyrate concentrations within each age group. The butyrate concentration ranged from 0 to 32.11 mmol/kg feces in the 6-month group and 3.35 to 96.06 mmol/kg feces in the 12-month group. No significant correlations were found between butyrate and the mean DNA methylation level. However, there was a positive correlation between butyrate

and CpG-position 1 of *IL-4* in the 6-month age group (rho = 0.35, p = 0.03, adjusted p = 1) decided with Spearman's correlation and corrected for multiple testing with Holm's method (Appendix D, figure D.3).

A significant increase from 16.52 mmol/kg feces to 20.34 mmol/kg feces was also seen for the absolute concentrations of propionate ($p < 0.05$, paired t-test) (figure 3.5). No correlations were found between propionate and the DNA methylation levels.

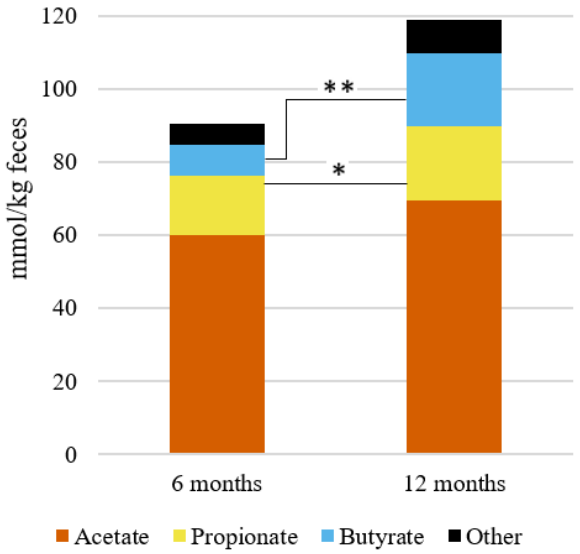

Figure 3.5. The mean SCFA composition shown as absolute concentrations in mmol/kg feces in the 6-month and 12-month age group. The stacked bar charts show the composition of the three main SCFAs acetate, propionate, butyrate, and the group called "other," which consists of isobutyrate, isovalerate, and valerate. The mean concentration for all infants in the respective age group was used to calculate the absolute abundances. There was a significant increase in the observed absolute concentration of butyrate ($p < 0.01$, paired t-tests) and propionate ($p < 0.05$, paired tests). N=56 for the 6-month age group and N=55 for the 12-month age group. Significant differences between the 6-month age group and the 12-month age group are marked with asterisks; 1 asterisk for a p-value < 0.05 and 2 asterisks for a p-value < 0.01.

3.3 The taxonomic composition and its correlation to the DNA methylation level

3.3.1 The taxonomic composition decided by 16S rRNA gene sequencing

There was a significantly higher number of observed OTUs in the 12-month age group compared to the 6-month age group (p = 9.86e-13, paired t-test). In the 6-month age group, there was a mean of 67 observed species, while the mean in the 12-month age group was 103 observed species (figure

3.6). The Simpson Diversity index showed a significantly higher diversity within the 12-month age group (p = 3.54e-08, paired t-test).

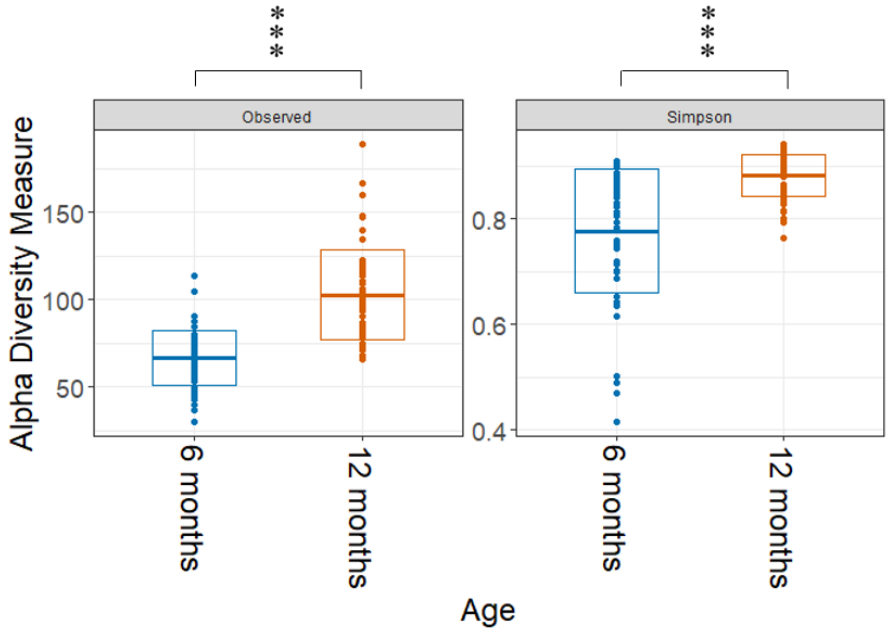

Figure 3.6. Alpha diversity measures showing the observed species and the Simpson D index. The difference in alpha diversity between the 6-month age group and the 12-month age group are shown with the observed species in the left panel and the inverse Simpson's Diversity Index (1-D) in the right panel. The thick middle line in each box is the mean, and the top and bottom lines are the standard deviation times two. The y-axis for the left panel shows the number of observed species. The y-axis for the right panel shows the Simpson's D index (1-D), with a higher number indicating a higher diversity. A higher diversity is shown in the 12-month age group, both by a higher number of observed OTUs (left panel) and a higher Simpson's Diversity index. Significant differences between the 6-month age group and the 12-month age group are marked with 3 asterisks for a p-value < 1e-07. The code used to make the strip chart in R studio is shown in appendix E.

The taxonomic composition at the bacterial order level showed approximately even abundances of Bacteroidales, Bifidobacteriales, Clostridiales, and Enterobacterales in the 6-month age group (figure 3.7). In the 12-month age group there was a significant increase in *Clostridiales* (p = 1.75e-12, paired t-test), and a significant decrease in Bifidobacteriales (p = 3.92e-05, paired t-test), Enterobacterales (p = 1.32e-03, paired t-test) and Veillonellales (p = 7.70e-04, paired t-test). A significant difference in the taxonomic composition (figure 3.7) between the 6-month age group and the 12-month age group was also shown by ANOSIM (Appendix E), with a statistic R-value of 0.31 (p = 1e-04).

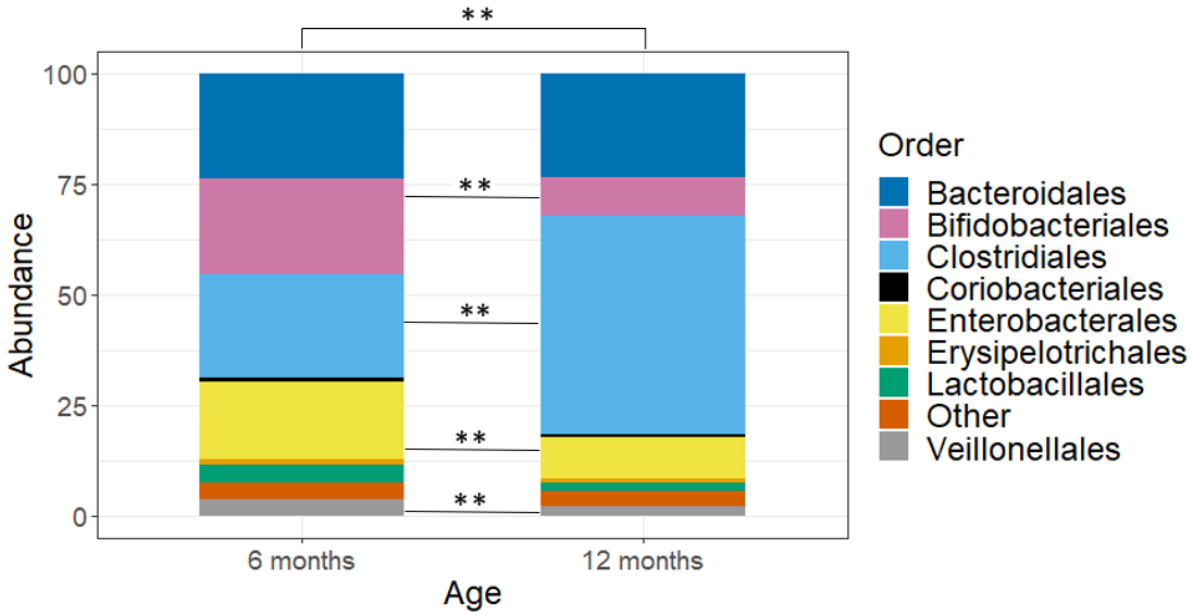

Figure 3.7. The taxonomic composition at the bacterial order level. The stacked bar chart shows the taxonomic composition in the 6-month age group and the 12-month age group as decided by 16S rRNA gene sequencing. The y-axis shows the abundance in percent calculated as the mean abundance of each OTU in all infants of an age group. OTUs with an abundance of less than 0.5 % for each age group are merged in the group called "Other". Significant differences in abundance of a bacterial order between the 6-month age group and the 12-month age group are marked with 2 asterisks for a p-value < 0.01. From 6 months of age to 12 months of age, there was a significant increase in Clostridiales and a decrease in Bifidobacteriales, Enterobacteriales, and Veillonellales ($p < 0.01$, paired t-test). A significant difference in composition between the 6-month age group and the 12-month age group was shown with ANOSIM (p = 1e-04). The code used to make the stacked bar chart in Rstudio is shown in appendix E.

To find which species was associated to each age group and gave the statistically significant difference in taxonomic composition between the age groups, an indicator species analysis was done. The R-code is shown in appendix E. After exclusion of OTUs with a relative abundance lower than 0.5 % and correction for multiple testing, 26 of the in total 1309 OTUs were shown to have a significant association with either the 6-month age group or the 12-month age group ($p < 0.05$, Benjamini Hochberg) (figure 3.8). As shown in figure 3.7, the abundance of Clostridiales increased from 23.38 % in the 6-month group to 49.48 % in the 12-month group. The increase was mainly caused by bacteria belonging to the family *Ruminococcaceae* and *Lachnospiraceae* (figure 3.8B). Bifidobacteriales decreased from 21.61 % in the 6-month group to 8.52 % in the 12-month group (figure 3.7), with a significant difference in the abundance of bacteria belonging to the genus *Bifidobacterium* (figure 3.8A). Enterobacterales decreased from 17.57 % in the 6-month group to

9.55 % in the 12-month group (figure 3.7) and is caused by a significant difference between bacteria belonging to the family *Enterobacteriaceae* (figure 3.8A).

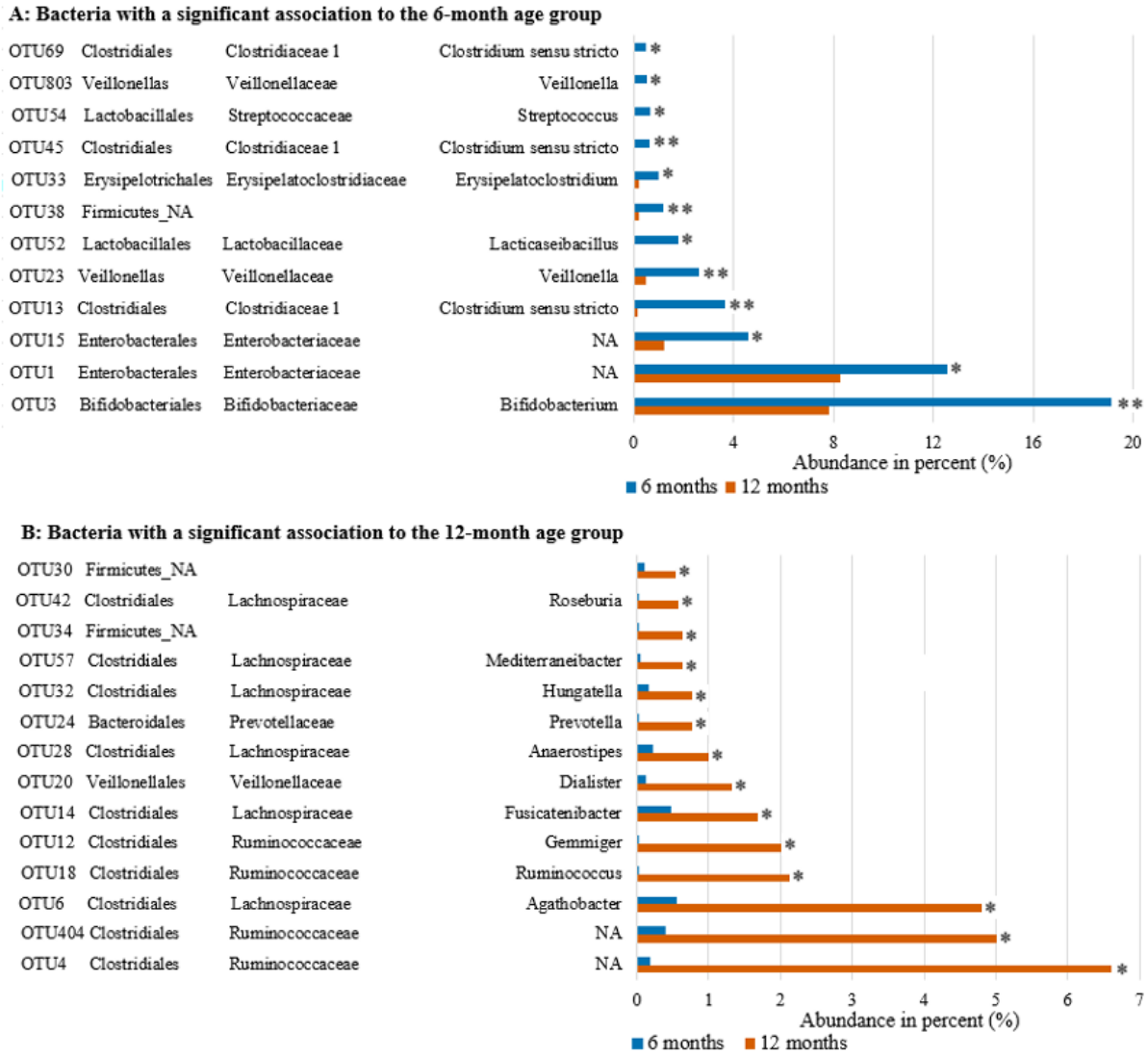

Figure 3.8. Bacteria with a significant association to (A) the 6-month age group and (B) the 12-month age group decided by indicator species analysis. Bacteria with an abundance of more than 0.5 % are included in the bar charts. The OTUs are shown on the y-axis, with bacterial order-, family-, and genus name. Three of the OTUs were not decided to the order level and are named by the phylum level. The abundance shown on the x-axis is the mean percentage of each OTU calculated by the number of reads. All OTUs in the figure are significantly associated with the 6-month age group or the 12-month age group and are marked with asterisks for significance level: 1 asterisk for a p-value < 0.05 and 2 asterisks for a p-value < 0.01 (Benjamini Hochberg).

3.3.2 Correlations between the DNA methylation level and the taxonomic composition

Spearman's correlations between the OTUs with a significant association to one of the age groups (figure 3.8) and the mean DNA methylation level of the genes *IL-4*, *DEFA5,* and *TLR4,* which showed a significant difference between the age groups (figure 3.3) were determined (Appendix D, figure D.4). After correction for multiple testing with Holm's method, none of the correlations were significant. However, some non-significant correlations were found between the mean DNA methylation level of *TLR4* and *DEFA5* and the OTUs in both the 6-month age group and the 12-month age group.

In the 6-month age group, there was a positive correlation between the DNA methylation level of *DEFA5* and *Enterobacteriaceae* (OTU1) (rho = 0.41, p = 0.01, adjusted p = 0.84). There was also a positive correlation between the DNA methylation level of *TLR4* and *Enterobacteriaceae* (OTU 1) (rho = 0.33, p = 0.04, adjusted p = 1). DNA methylation of *DEFA5* was negatively correlated to *Veillonella* (OTU 803) in the 6-month age group (rho = -0.35, p = 0.03, adjusted p = 1). In the 12-month age group, DNA methylation of *DEFA5* was positively correlated to *Ruminococcaceae* (OTU404) (rho = 0.29, p = 0.04, adjusted p = 1) and negatively correlated to *Anaerostipes* (OTU28) (rho = -0.34, p = 0.01, adjusted p = 0.9626). DNA methylation of *TLR4* was positively correlated with *Dialister* (OTU 20) (rho = 0.30, p = 0.03, adjusted p = 1) and negatively correlated with *Fusicatenibacter* (OTU 14) (rho = -0.29, p = 0.04, adjusted p = 1).

As no significant correlations were found between OTUs with a significant association to one of the age groups and the mean DNA methylation level of genes with a significant difference between the age groups, the data was further examined with the inclusion of all bacterial orders (figure 3.7) and the DNA methylation level of all genes (Appendix D, figure D.5). No significant correlations were found, but there were some non-significant trends for *TLR4* in the 12-months age group. *TLR4* was positively correlated to Bacteroidales (rho = 0.30, p = 0.03, adjusted p = 1) and Veillonellales (rho = 0.30, p = 0.03, adjusted p = 1). There were also negative correlations to Lactobacillales (rho = -0.37, p = 0.01, adjusted p = 0.39) and Clostridiales (rho = -0.28, p = 0.04, adjusted p = 1).

3.4 Correlations between SCFAs and bacterial composition and function

Significant correlations were found between the main SCFAs and bacterial orders (figure 3.1). The correlations in table 3.1 represent the significant correlations from the Spearman's correlation plot

in appendix D (Appendix D, figure D.6). In both the 6-month and 12-month age groups, there was a positive correlation between propionate and Bacteroidales (adjusted $p < 0.05$) and between butyrate and Clostridiales (adjusted $p < 0.05$). In the 12-month age group, there was also a negative correlation between butyrate and Enterobacterales and acetate and Enterobacterales ($p < 0.05$, Holm's method). At a lower significance level, there was a negative correlation between propionate and Lactobacillales (adjusted $p < 0.1$). Additionally, there was a positive correlation between butyrate and acetate in both age groups (adjusted $p < 0.05$).

Table 3.1. Spearman's correlations between SCFAs and bacterial orders in (A) the 6-month age group and (B) the 12-month age group. The correlation coefficient rho denotes correlations with a number ranging from 0 to 1, with 0 being no correlation. Blue upward pointing errors indicate positive correlations, and red downward pointing errors indicate negative correlations. Adjusted p-values with a significance level of 5 % corrected by Holm's method are indicated with an asterisk. The correlations are shown as correlation plots in appendix D, figure D.6.

A: The 6-month age group		***P-value***	***Adjusted p-value***	***Correlation coefficient rho***
Butyrate	Clostridiales ↑	<0.0001	<0.0001*	0.60
	Acetate ↑	0.0008	0.0479*	0.44
Propionate	Bacteroidales ↑	0.0005	0.0302*	0.45
	Lactobacillales ↓	0.0010	0.0638	-0.43

B: The 12-month age group		***P-value***	***Adjusted p-value***	***Correlation coefficient rho***
Acetate	Enterobacterales ↓	0.0004	0.0216*	-0.46
Butyrate	Clostridiales ↑	0.0002	0.0117*	0.48
	Acetate ↑	<0.0001	0.0001*	0.59
	Enterobacterales ↓	0.0002	0.0120*	-0.48
Propionate	Bacteroidales ↑	<0.0001	<0.0001*	0.61

Tax4Fun was used to predict the presence of enzymes belonging to the butyrate metabolism. The R-code is shown in Appendix E. All enzymes belonging to the acetyl-CoA pathway of the butyrate metabolism were predicted to be present as shown by the colored boxes in figure 3.9. There was a

significantly higher abundance of EC 4.2.1.17 (p = 0.04, Wilcoxon signed-rank test), EC 1.1.1.35 (p = 2.12e-07, Wilcoxon signed-rank test) and EC 2.8.3.8 (p = 9.90e-04, Wilcoxon signed-rank test) in the 6-month age group. Additionally, there was a non-significant higher abundance of EC 2.3.1.9 (p = 0.22, Wilcoxon signed-rank test). This indicates that there was synthesis of butyrate directly from butanoyl-CoA for 6 months old infants. In the 12-month age group, there was a significantly higher predicted abundance of EC 1.3.8.1 (p = 3.99e-07, Wilcoxon signed-rank test). Additionally, there was a non-significant higher predicted abundance of EC 2.3.1.19 (p = 0.59, Wilcoxon signed-rank test) and EC 2.7.2.7 (p = 0.91, Wilcoxon signed-rank test). This gives an indication of butyrate synthesis via butanoyl-P for 12-months old infants.

Spearman's correlations between the butyrate concentrations as decided by gas chromatography and the predicted presence of the *but*- and *buk genes* showed a significant negative correlation between the predicted presence of the *but gene* and butyrate in the 12-month age group (rho = -0.48, adjusted p = 7e-04).

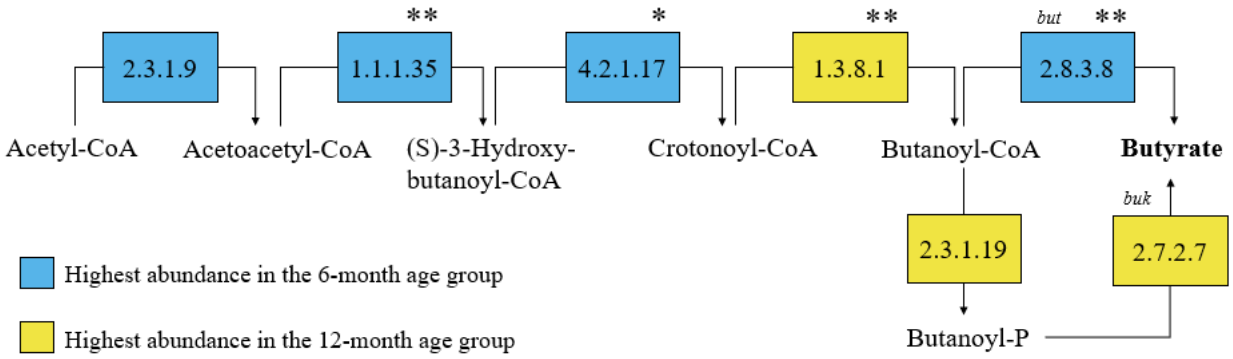

Figure 3.9. Presence of enzymes belonging to the butyrate metabolism as predicted by Tax4Fun. Each box represents an enzyme and is marked with the EC number of the enzyme. The color of the box indicates the age group with a higher predicted abundance of the respective enzyme: blue color denotes higher abundance in the 6-month age group and yellow color denotes higher abundance in the 12-month age group. Enzymes with a significant difference between the age groups are marked with asterisks: 1 asterisk for a p-value < 0.05 and 2 asterisks for a p-value < 0.001. The *but gene* is encoded by EC 2.8.3.8, and the *buk gene* is encoded by EC 2.7.2.7. The figure is adapted and redrawn from KEGG mapper reconstruct tool.

4 Discussion

4.1 Differences in DNA methylation pattern between age groups for the investigated genes

4.1.1 A significantly higher mean DNA methylation level in the 12-month age group

The targeted gene fragments are part of the promoter region of their respective genes, and an assumption is therefore made that an increase in DNA methylation indicates a decrease in gene expression. There was a significantly higher mean DNA methylation level in the 12-month age group (figure 3.2), which indicates downregulation in the expression of the investigated genes as the infants aged. A slowdown of biological processes has been used as an explanation in an earlier DNA methylation study, which also showed a higher DNA methylation level of most genes in the oldest infants (Wikenius *et al.*, 2019). However, this is not well established, and the selection of genes could have an impact on the results. An example of this is the decrease in DNA methylation with age which was seen for *IL-4*. A higher number of genes should therefore be investigated before an assumption is made.

4.1.2 A significant increase in the DNA methylation level of *TLR4* and *DEFA5* with age

The significant increase in DNA methylation level of both *DEFA5* and *TLR4* from the 6-month age group to the 12-month age group (figure 3.3) is assumed to be associated with a decrease in the expression of these genes with age. There was also a significant positive correlation between the mean DNA methylation level of *DEFA5* and *TLR4*, which suggests that the DNA methylation of one of the genes is dependent upon the DNA methylation of the other gene. This is consistent with a study by Menendez *et al.* (2013), which showed a decrease in the gene expression of *DEFA5* with *TLR4* deficiencies.

Comparison of the DNA methylation level of *TLR4* and *DEFA5* to the DNA methylation level found by Klerk *et al.* (2021) showed a slightly higher percentage in the results of this thesis. The DNA methylation level of *DEFA5* was approximately 88 % and 91 % for the 6- and 12-month age groups, respectively (figure 3.3C). In the results by Klerk *et al.* (2021), the mean DNA methylation level was about 74 % for the healthy controls. The mean DNA methylation level for *TLR4* was approximately 59 % and 73 % for the 6- and 12-month age groups, respectively (figure 3.3A). In comparison, the mean DNA methylation level of *TLR4* found by Klerk *et al.* (2021) was about

80 % for the healthy controls. However, a higher number of CpG-positions were included for *DEFA5* in the results of this thesis and could therefore be the reason for the somewhat deviating DNA methylation levels.

According to Klerk *et al.* (2021), the target sequence of the *TLR4* promoter covers regulatory factor X1. The expression of the regulatory factor has shown negative correlations to the expression of *TLR4* and decreased DNA methylation with a decrease in expression of the regulatory factor (Du *et al.*, 2019). This indicates that the higher DNA methylation level of *TLR4* in the 12-month age group could give a decrease in gene expression of *TLR4*, which is consistent with the assumption regarding DNA methylation of promoter regions. This seems to be the case for *DEFA5* as well. According to Klerk *et al.* (2021), the targeted promoter region sequence of *DEFA5* is the same part of the promoter as has shown an increase in DNA methylation in patients with Crohn's disease. The Crohn's disease patients consequently showed a decreased gene expression compared to healthy controls (Cerrillo *et al.*, 2018).

4.1.3 Opposite DNA methylation patterns of *IL-4* and *IFN-ɣ*

Although the difference between the age groups for the mean DNA methylation level of *IFN-ɣ* was non-significant, the results for *IFN-ɣ* and *IL-4* showed an opposite pattern of the mean DNA methylation level based on age group (figure 3.3). The results showed a significantly higher mean DNA methylation level of *IL-4* in the 6-month age group compared to the 12-month age group (figure 3.3D). For *IFN-ɣ*, the mean DNA methylation level was non-significantly higher in the 12-month age group compared to the 6-month age group (figure 3.3B). In the study by Berni Canani *et al.* (2015), an association between DNA methylation of *IL-4* and *IFN-*ɣ and cow's milk allergy in children was found. Children with allergy had a lower DNA methylation level of the gene promoters of *IL-4*, and oppositely, a higher DNA methylation level of the gene promoters of *IFN-ɣ* (Berni Canani *et al.*, 2015). As *IL-4* is a Th2-associated gene and *IFN-ɣ* is Th1-associated, the DNA methylation pattern could be associated with the maturation of the immune system and potentially the gut microbiota. A hypothesis could be that the DNA methylation level of genes in children with allergy resembles the DNA methylation level of the youngest infants because of the immature Th1-response. This hypothesis was not strengthened by the results. However, it would

be interesting to determine the DNA methylation pattern in meconium for comparison and additionally with selection of infants with allergy.

Interestingly, one of the CpG-positions of *IFN-ɣ* was negatively correlated to two CpG-positions of *IL-4* in the 6-month age group. This strengthens the hypothesis of opposing DNA methylation patterns in Th1- and Th2-genes. However, limited knowledge of each specific CpG-positions makes interpretation challenging.

4.2 The impact of SCFAs on the DNA methylation level

4.2.1 Lack of effect of butyrate on the DNA methylation level

There were no significant correlations between the observed SCFAs and the DNA methylation levels. Although both DNA methylation and butyrate had an observed increase with age, no correlations were found when looking into each age group separately. This suggests that butyrate has a limited effect on DNA methylation. In contradiction, an earlier study has shown that butyrate treatment induced hypermethylation of DNA (de Haan *et al.*, 1986). But as discussed by de Haan *et al.* (1986) there are also studies showing hypomethylation after butyrate treatment, and it is therefore likely that different cell lines have different responses to butyrate treatment. Based on the results it is likely that none of the selected immune-related genes are impacted by butyrate.

As reviewed by Kruh (1981), butyrate is a known HDAC inhibitor. HDAC inhibitors prevent the recruitment of HDACs to methylated DNA (Sarkar *et al.*, 2011), which leads to hyperacetylation of histone tails and thus increased transcription. If the DNA methylation level reflected the action of HDAC inhibitors, a decrease in DNA methylation would be expected when the butyrate concentration increased. The lack of negative correlations suggests that butyrate did not function as a HDAC inhibitor for the selected genes.

HDAC inhibitors have shown other mechanisms than solely the inhibition of HDACs. An example is that HDAC inhibitors have shown down-regulating effects on DNMT1 with a reversal of hypermethylated DNA in tumor suppressor genes as a consequence (Sarkar *et al.*, 2011). This should also give a negative correlation between the DNA methylation level and butyrate. The only correlation found between butyrate, and the individual CpG-positions was a positive non-significant correlation between CpG-position 1 of *IL-4* and butyrate. A positive correlation

indicates that the presence of butyrate gives a higher DNA methylation level of this specific CpG-position. This supports the indication of butyrate not having a down-regulating effect on DNMTs.

Butyrate and propionate have been shown to down-regulate transcription of *DEFA5* in concentrations of 9 mM (Sugi *et al.*, 2017). Considering the significantly increased butyrate levels in the 12-month age group (figure 3.5), the concentration should be high enough to potentially suppress transcription. No correlations were found between the DNA methylation level of *DEFA5* to any of the SCFAs. This indicates that presence of butyrate does not increase the DNA methylation level of *DEFA5*, and suggest that the down-regulation seen by Sugi *et al.* (2017) is not caused by an increase in DNA methylation.

4.3 The impact of gut bacteria on the DNA methylation level

The non-significant correlations found between the gut bacteria and the DNA methylation level indicate that there could be a link between them. A possible link is micronutrients – as for example vitamins produced by the gut microbiota, which are used in the production of the methyl donor SAM (Ye *et al.*, 2017).

4.3.1 DNA methylation level of *TLR4* and correlations with gram-negative bacteria

All bacterial OTUs and bacterial orders with a positive correlation to the DNA methylation level of *TLR4* consisted of gram-negative bacteria (*Enterobacteriaceae, Dialister*, Bacteroidales, and Veillonellales). This is interesting as *TLR4* recognizes lipopolysaccharides (Kawasaki & Kawai, 2014) which are part of the outer membrane of gram-negative bacteria. A positive correlation means that a higher abundance of the bacteria is associated with a higher DNA methylation level and vice versa. This indicates that high abundance of these bacteria potentially could down-regulate the gene expression of *TLR4*. In contrast, all bacterial OTUs and bacterial orders with a negative correlation to *TLR4* consisted of gram-positive bacteria (*Fusicatenibacter*, Lactobacillales, and Clostridiales). A possible explanation is that the DNA used to study the DNA methylation level probably originates from epithelial cells in the gut. Because of the frequent exposure to commensal bacteria, the gut epithelial cells are not responsive to these lipopolysaccharides (Takahashi *et al.*, 2009). The expression of *TLR4* has been shown to be downregulated in intestinal epithelial cells, and hence there is no secretion of proinflammatory cytokines (Abreu *et al.*, 2001). The correlations

found in this thesis are consistent independent of age group and could possibly explain one of the mechanisms behind the downregulation of *TLR4* in the gut.

4.3.2 A non-significant correlation between *TLR4* and *DEFA5* and *Enterobacteriaceae*

Although no significant correlations were found between the taxonomic composition and the DNA methylation levels, there were some non-significant findings. Within the 6-month age group, there was a non-significant positive correlation between the DNA methylation level of both *DEFA5* and *TLR4* to the *Enterobacteriaceae* family (OTU 1). This indicates that the higher abundance of *Enterobacteriaceae* in the 6-month age group is possibly involved in downregulating the expression of *DEFA5* and *TLR4* by increasing the DNA methylation level. Although *Enterobacteriaceae* is part of the normal gut microbiota of infants (Del Chierico *et al.*, 2015), certain strains of *Escherichia coli* have been associated with Crohn's disease (Palmela *et al.*, 2018). The correlation between *Enterobacteriaceae* and the DNA methylation level should be further investigated to determine if there could be a link between the DNA methylation level in infancy and the development of Crohn's disease later in life.

4.3.3 The impact of a decrease in abundance of *Lactobacillales* with age

Supplementation with *Lactobacillus rhamnosus* has shown to be associated with a higher DNA methylation level of *IL-4* and a lower DNA methylation level of *IFN-ɣ* compared to controls (Berni Canani *et al.*, 2015). The abundance of Lactobacillales was higher in the 6-month age group (figure 3.7), and the genus *Lacticaseibacillus* was shown to be significantly associated with the 6-month age group (figure 3.8). *Lacticaseibacillus*, which is a newer name of *Lactobacillus*, was not taxonomically determined to the species level and could therefore potentially be *Lactobacillus rhamnosus*. If the 6-month age group is compared with the *Lactobacillus* supplementation group, the DNA methylation level is consistent with the finding of Canani et al., with a higher DNA methylation level of *IL-4* and a lower DNA methylation level of *IFN-ɣ*, with the 12-month age group being considered as a control group. Additionally, depletion of Lactobacillales by antibiotic treatment has also shown to be correlated with lowered gene expression of *DEFA5* (Menendez *et al.*, 2013). The higher DNA methylation level of *DEFA5* in the 12-month age group indicates that a decrease in Lactobacillales is associated with a decrease in gene expression. However, no

correlations were found between Lactobacillales and the DNA methylation level of any of the genes.

4.4 Technical considerations

4.4.1 Targeting of human DNA from infant feces

The small amounts of human DNA in feces make targeting of human genes from fecal DNA challenging. As studied by Klaassen *et al.* (2003), the rate of human DNA in feces relative to the total DNA in feces is approximately 1:100 000.

Which cell types are present in feces is not completely clear, but an earlier study found mRNA from enterocytes, goblet cells, enteroendocrine cells, and Paneth cells (Chapkin *et al.*, 2010). *DEFA5* is expressed by Paneth cells in the small intestine (Sugi *et al.*, 2017), and *TLR4* is expressed by the intestinal epithelium (Hackam & Sodhi, 2018). Determination of the DNA methylation level of *TLR4* and *DEFA5* from fecal samples could therefore, potentially, give an indication of the gene expression of these genes. A small concentration of lymphocytes has also been detected in fecal samples (Chandel *et al.*, 2011), which should give a possibility of targeting *IL-4* and *IFN-ɣ*, which are expressed by T-cells. However, the detected lymphocytes might have been caused by a minor bleeding in the intestines (Chandel *et al.*, 2011).

The presence of a smear of all *TLR4* samples shown by gel visualization gave indications of unspecific binding of the primers to DNA. This was the main reason for the choice of method for read alignment to the target sequence from the human genome. All reads that did not map to the target sequence were excluded. An alternative method of dividing the reads into zero-radius OTUs (zOTU) was first tested. All zOTUs were further aligned with a multiple sequence alignment (MSA) using Multiple Sequence Comparison by Log-Expectation (MUSCLE). This method was rejected as the MSA did not seem correct for all genes, probably due to the presence of different gene fragments.

4.4.2 Bisulfite conversion of DNA

During bisulfite conversion of DNA for the 12-months samples, an additional centrifugation step was added after the wash steps. This was done due to a higher volume of eluate than expected

during bisulfite conversion of 8 of the 6-month samples, most likely caused by insufficient centrifugation after the wash steps. The consequence of this was a dilution of the eluate and the possible presence of substances that could inhibit amplification during PCR. Three of the diluted samples lacked a band on the gel picture after amplicon PCR of *DEFA5*. However, this was not reflected in the number of mapped reads after sequencing. Therefore, the dilution did not seem to have had a considerable impact on the results. To avoid dilution of the samples in future studies an additional centrifugation step is recommended, as was also done by Klerk *et al.* (2021).

4.4.3 The human genome as a reference

An alternative to the use of the human genome (GRCh38.p13) as a reference for the bisulfite converted sequences would be sequencing of the fragments prior to bisulfite conversion. This would, in theory, require modification of the primers to target conventional DNA, but as seen in the pilot study of *TLR4*, there was possible to amplify the fragment with primers meant for bisulfite-converted DNA. The BLAST search showing this was *TLR4* suggests that the desired fragments are successfully targeted and sequenced from the extracted DNA. However, using sequencing for comparison would be too time-consuming, and the human genome was therefore chosen for comparison instead. A disadvantage with using the human genome browser is that it is occasionally updated with new releases, and corrections of the sequences used in this thesis could therefore occur.

4.4.4 The resolution of 16S rRNA gene sequencing

Not all OTUs were taxonomically determined to the genus level by sequencing of the V3-V4 region. A higher resolution could be obtained by sequencing the whole 16S rRNA gene, although sequencing of selected variable regions should be sufficient to assign bacteria to the genus level (Johnson *et al.*, 2019). One of the reasons for the lower resolution than expected is probably the exclusion of forward reads, which excluded most of the V3 region. However, consistent with the expectations, the observed taxonomic composition showed an increase in alpha diversity (figure 3.6) and a significant increase in Clostridiales (figure 3.7) with age. There was also a significant positive correlation between butyrate and Clostridiales (table 3.1), which is consistent with literature which has shown a correlation between bacteria belonging to the order Clostridiales and

butyrate (Nilsen *et al.*, 2020). This suggests that lowered resolution was the only consequence of the exclusion of forward reads.

4.4.5 DNA methylation of individual CpG-positions

To my knowledge, the effect of DNA methylation on one single CpG-position is not clear. Some of the CpG-positions within the same gene showed conflicting results regarding the DNA methylation pattern, as for example for *DEFA5,* where one of the CpG-positions showed an opposite pattern (figure 3.4). Individual CpG-positions could be used for comparison to other studies which have studied the same position, but otherwise, the mean DNA methylation level would probably give a more correct picture.

4.4.6 Tax4Fun for enzyme prediction

Tax4Fun was used for the prediction of enzymes belonging to the acetyl-CoA pathway of the butyrate metabolism. The results gave indications of differences in enzyme distribution between the 6-month age group and the 12-month age group (figure 3.9). These differences could possibly explain the significant increase in butyrate from the 6-months age group to the 12-month age group. However, there was a significant negative correlation between the predicted presence of the *but gene* and butyrate concentration in the 12-month age group. A higher copy number of the *but gene* has shown to be associated with higher butyrate levels (Daskova *et al.*, 2021). This was not consistent with the results (figure 3.9), showing a higher predicted abundance in the 6-month age group. Thus, the results show that predictions of the function based on taxonomic composition should be interpreted with caution.

5 Conclusion and future work

The observed DNA methylation levels showed that there was a difference between the age groups, with a significantly higher mean DNA methylation level in the 12-month age group compared to the 6-month age group. However, the DNA methylation rate of *IL-4* decreased with age, indicating that the impact by the factor age is dependent on the selection of genes. The lack of significant correlations between butyrate and the DNA methylation level gave indications that butyrate is not the main driving force of DNA methylation. Although no significant correlations were found between the bacterial composition and DNA methylation, there were indications of associations. The results showed that gram-negative bacteria could be a potential factor with association to DNA methylation of *TLR4*.

Future work should include actual gene expression studies. This could be done by cultivation of Caco-2 cells in the presence of SCFAs followed by RNA extraction. RNA extraction from fecal samples was done as a pilot study for the work in this thesis (Appendix F) but was unsuccessful and therefore excluded from the main study.

In conclusion, the results suggest that factors associated with the maturation of the infant gut microbiota has the potential to modulate the DNA methylation level of the selected genes. More studies should be done to investigate the impact of the gut bacterial composition and age. The DNA methylation level should also be determined for several time points to investigate if the significant increase with age is consistent over time.

6 References

Abreu, M. T., Vora, P., Faure, E., Thomas, L. S., Arnold, E. T. & Arditi, M. (2001). Decreased expression of Toll-like receptor-4 and MD-2 correlates with intestinal epithelial cell protection against dysregulated proinflammatory gene expression in response to bacterial lipopolysaccharide. *The Journal of Immunology*, 167 (3): 1609-16. doi: 10.4049/jimmunol.167.3.1609.

Ambardar, S., Gupta, R., Trakroo, D., Lal, R. & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, 56 (4): 394-404. doi: 10.1007/s12088-016-0606-4.

AngenBiotech. (n.d.). *MagPure Stool DNA LQ Kit*. Available at: http://www.angenbio.com/upload/pdown/6909190715153738.pdf (accessed: 01.03.22).

Bartle, K. D. & Myers, P. (2002). History of gas chromatography. *TrAC Trends in Analytical Chemistry*, 21 (9): 547-557. doi: 10.1016/S0165-9936(02)00806-3.

Belkaid, Y. & Hand, Timothy W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, 157 (1): 121-141. doi: 10.1016/j.cell.2014.03.011.

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.

Bennett, G. N. & Rudolph, F. B. (1995). The central metabolic pathway from acetyl-CoA to butyryl-CoA in Clostridium acetobutylicum. *FEMS Microbiology Reviews*, 17 (3): 241-249. doi: 10.1111/j.1574-6976.1995.tb00208.x.

Berger, A. (2000). Th1 and Th2 responses: what are they? *British medical journal*, 321 (7258): 424. doi: 10.1136/bmj.321.7258.424.

Berni Canani, R., Paparo, L., Nocerino, R., Cosenza, L., Pezzella, V., Di Costanzo, M., Capasso, M., Del Monaco, V., D'Argenio, V., Greco, L., et al. (2015). Differences in DNA methylation profile of Th1 and Th2 cytokine genes are associated with tolerance acquisition in children with IgE-mediated cow's milk allergy. *Clinical Epigenetics*, 7 (1): 38. doi: 10.1186/s13148-015-0070-8.

BioLabs, N. E. (n.d.). *Proteinase K, Molecular Biology grade*. Available at: https://international.neb.com/products/p8107-proteinase-k-molecular-biology-grade#Product%20Information (accessed: 11.03.22).

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research*, 8 (7): 1499-1504. doi: 10.1093/nar/8.7.1499.

Blaak, E. E., Canfora, E. E., Theis, S., Frost, G., Groen, A. K., Mithieux, G., Nauta, A., Scott, K., Stahl, B., Harsselaar, J. v., et al. (2020). Short chain fatty acids in human gut and metabolic health. *Beneficial Microbes*, 11 (5): 411-455. doi: 10.3920/bm2020.0057.

Bode, L. (2015). The functional biology of human milk oligosaccharides. *Early Human Development*, 91 (11): 619-622. doi: 10.1016/j.earlhumdev.2015.09.001.

Cerrillo, E., Moret, I., Iborra, M., Ramos, D., Busó, E., Tortosa, L., Sáez-González, E., Nos, P. & Beltrán, B. (2018). Alpha-defensins (α-Defs) in Crohn's disease: decrease of ileal α-Def 5 via permanent methylation and increase in plasma α-Def 1-3 concentrations offering biomarker utility. *Clinical and experimental immunology*, 192 (1): 120-128. doi: 10.1111/cei.13085.

Chandel, D. S., Braileanu, G. T., Chen, J.-H. J., Chen, H. H. & Panigrahi, P. (2011). Live Colonocytes in Newborn Stool: Surrogates for Evaluation of Gut Physiology and Disease Pathogenesis. *Pediatric Research*, 70 (2): 153-158. doi: 10.1203/PDR.0b013e3182225ac9.

Chapkin, R. S., Zhao, C., Ivanov, I., Davidson, L. A., Goldsby, J. S., Lupton, J. R., Mathai, R. A., Monaco, M. H., Rai, D., Russell, W. M., et al. (2010). Noninvasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *American journal of physiology. Gastrointestinal and liver physiology*, 298 (5): G582-G589. doi: 10.1152/ajpgi.00004.2010.

Chelakkot, C., Ghim, J. & Ryu, S. H. (2018). Mechanisms regulating intestinal barrier integrity and its pathological implications. *Experimental & Molecular Medicine*, 50 (8): 1-9. doi: 10.1038/s12276-018-0126-x.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18 (1): 117-143. doi: 10.1111/j.1442-9993.1993.tb00438.x.

Clausen, M. R. & Mortensen, P. B. (1995). Kinetic studies on colonocyte metabolism of short chain fatty acids and glucose in ulcerative colitis. *Gut*, 37 (5): 684. doi: 10.1136/gut.37.5.684.

Daskova, N., Heczkova, M., Modos, I., Videnska, P., Splichalova, P., Pelantova, H., Kuzma, M., Gojda, J. & Cahova, M. (2021). Determination of Butyrate Synthesis Capacity in Gut Microbiota: Quantification of but Gene Abundance by qPCR in Fecal Samples. *Biomolecules*, 11 (9): 1303.

Davie, J. R. (2003). Inhibition of histone deacetylase activity by butyrate. *The Journal of nutrition*, 133 (7 Suppl): 2485s-2493s. doi: 10.1093/jn/133.7.2485S.

De Cáceres, M., Legendre, P. & Moretti, M. (2010). Improving indicator species analysis by combining groups of sites. *Oikos*, 119 (10): 1674-1684. doi: 10.1111/j.1600-0706.2010.18334.x.

de Haan, J. B., Gevers, W. & Parker, M. I. (1986). Effects of sodium butyrate on the synthesis and methylation of DNA in normal cells and their transformed counterparts. *Cancer Research*, 46 (2): 713-6.

Deaton, A. M. & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, 25 (10): 1010-1022. doi: 10.1101/gad.2037511.

Debock, I. & Flamand, V. (2014). Unbalanced Neonatal CD4+ T-Cell Immunity. *Frontiers in Immunology*, 5. doi: 10.3389/fimmu.2014.00393.

Del Chierico, F., Vernocchi, P., Petrucca, A., Paci, P., Fuentes, S., Praticò, G., Capuani, G., Masotti, A., Reddel, S., Russo, A., et al. (2015). Phylogenetic and Metabolic Tracking of Gut Microbiota during Perinatal Development. *PLOS ONE*, 10 (9): e0137347. doi: 10.1371/journal.pone.0137347.

Del Prete, G. (1992). Human Th1 and Th2 lymphocytes: their role in the pathophysiology of atopy. *Allergy*, 47 (5): 450-5. doi: 10.1111/j.1398-9995.1992.tb00662.x.

Delves, P. J. & Roitt, I. M. (2000). The Immune System. *New England Journal of Medicine*, 343 (1): 37-49. doi: 10.1056/NEJM200007063430107.

den Besten, G., van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J. & Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of lipid research*, 54 (9): 2325-2340. doi: 10.1194/jlr.R036012.

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. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (26): 11971-5. doi: 10.1073/pnas.1002601107.

Du, P., Gao, K., Cao, Y., Yang, S., Wang, Y., Guo, R., Zhao, M. & Jia, S. (2019). RFX1 downregulation contributes to TLR4 overexpression in CD14+ monocytes via epigenetic mechanisms in coronary artery disease. *Clinical Epigenetics*, 11 (1): 44. doi: 10.1186/s13148-019-0646-9.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, 308 (5728): 1635-1638. doi: 10.1126/science.1110591.

Faith, D. P., Minchin, P. R. & Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69 (1): 57-68. doi: 10.1007/BF00038687.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (5): 1827-1831. doi: 10.1073/pnas.89.5.1827.

Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P. & Kouzarides, T. (2003). The Methyl-CpG-binding Protein MeCP2 Links DNA Methylation to Histone Methylation*. *Journal of Biological Chemistry*, 278 (6): 4035-4040. doi: 10.1074/jbc.M210256200.

Furey, T. S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*, 13 (12): 840-852. doi: 10.1038/nrg3306.
Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. (2015). QuasR: quantification and annotation of short reads in R. *Bioinformatics*, 31 (7): 1130-1132. doi: 10.1093/bioinformatics/btu781.
Gnoth, M. J., Kunz, C., Kinne-Saffran, E. & Rudloff, S. (2000). Human Milk Oligosaccharides Are Minimally Digested In Vitro. *The Journal of Nutrition*, 130 (12): 3014-3020. doi: 10.1093/jn/130.12.3014.
Hackam, D. J. & Sodhi, C. P. (2018). Toll-Like Receptor–Mediated Intestinal Inflammatory Imbalance in the Pathogenesis of Necrotizing Enterocolitis. *Cellular and Molecular Gastroenterology and Hepatology*, 6 (2): 229-238.e1. doi: 10.1016/j.jcmgh.2018.04.001.
Hayatsu, H., Wataya, Y., Kai, K. & Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry*, 9 (14): 2858-2865.
Hayatsu, H. (2008). Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account. *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, 84 (8): 321-330. doi: 10.2183/pjab.84.321.
Hazlett, L. & Wu, M. (2011). Defensins in innate immunity. *Cell and Tissue Research*, 343 (1): 175-188. doi: 10.1007/s00441-010-1022-4.
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.
Herman, J. G., Graff, J. R., Myöhänen, S., Nelkin, B. D. & Baylin, S. B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences*, 93 (18): 9821. doi: 10.1073/pnas.93.18.9821.
Hillyar, C., Rallis, K. S. & Varghese, J. (2020). Advances in Epigenetic Cancer Therapeutics. *Cureus*, 12 (11): e11725. doi: 10.7759/cureus.11725.
Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., Bender, J. M., Azad, M. B., Thompson, A. L., Weiss, S. T., et al. (2018). Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature Communications*, 9 (1): 4169. doi: 10.1038/s41467-018-06473-x.
Hu, T., Chitnis, N., Monos, D. & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82 (11): 801-811. doi: 10.1016/j.humimm.2021.02.012.
Illumina. (2016). *Using a PhiX Control for HiSeq® Sequencing Runs*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/hiseq-phix-control-v3-technical-note.pdf.
Illumina. (2017). *An Introduction to Next-Generation Sequencing Technology*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf (accessed: 15.03.2021).
Illumina. (2018). *Effects of Index Misassignment on Multiplexing and Downstream Analysis*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf.
Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics*, 41 (2): 178-186. doi: 10.1038/ng.298.
Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10 (1): 5029. doi: 10.1038/s41467-019-13036-1.
Kawasaki, T. & Kawai, T. (2014). Toll-Like Receptor Signaling Pathways. *Frontiers in Immunology*, 5. doi: 10.3389/fimmu.2014.00461.
Kidd, P. (2003). Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease. *Alternative medicine review*, 8 (3): 223-246.

Klerk, D. H., Plösch, T., Verkaik-Schakel, R. N., Hulscher, J. B. F., Kooi, E. M. W. & Bos, A. F. (2021). DNA Methylation of TLR4, VEGFA, and DEFA5 Is Associated With Necrotizing Enterocolitis in Preterm Infants. *Frontiers in pediatrics*, 9: 630817. doi: 10.3389/fped.2021.630817.

Klaassen, C. H. W., Jeunink, M. A. F., Prinsen, C. F. M., Ruers, T. J. M., Tan, A. C. I. T. L., Strobbe, L. J. A. & Thunnissen, F. B. J. M. (2003). Quantification of Human DNA in Feces as a Diagnostic Test for the Presence of Colorectal Cancer. *Clinical Chemistry*, 49 (7): 1185-1187. doi: 10.1373/49.7.1185.

Kruh, J. (1981). Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Molecular and Cellular Biochemistry*, 42 (2): 65-82. doi: 10.1007/BF00222695.

Kujirai, T. & Kurumizaka, H. (2020). Transcription through the nucleosome. *Current Opinion in Structural Biology*, 61: 42-49. doi: 10.1016/j.sbi.2019.10.007.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. & Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, 69 (6): 905-14. doi: 10.1016/0092-8674(92)90610-o.

Li, W. & Sun, Z. (2019). Mechanism of Action for HDAC Inhibitors—Insights from Omics Approaches. *International Journal of Molecular Sciences*, 20 (7). doi: 10.3390/ijms20071616.

Litvak, Y., Byndloss Mariana, X. & Bäumler Andreas, J. (2018). Colonocyte metabolism shapes the gut microbiota. *Science*, 362 (6418): eaat9076. doi: 10.1126/science.aat9076.

Louis, P. & Flint, H. J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS microbiology letters*, 294 (1): 1-8. doi: 10.1111/j.1574-6968.2009.01514.x.

Louis, P. & Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology*, 19 (1): 29-41. doi: 10.1111/1462-2920.13589.

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.

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389 (6648): 251-260. doi: 10.1038/38444.

Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19 (2): 81-92. doi: 10.1038/nrg.2017.80.

Lødrup Carlsen, K. C., Rehbinder, E. M., Skjerven, H. O., Carlsen, M. H., Fatnes, T. A., Fugelli, P., Granum, B., Haugen, G., Hedlin, G., Jonassen, C. M., et al. (2018). Preventing Atopic Dermatitis and ALLergies in Children—the PreventADALL study. *Allergy*, 73 (10): 2063-2070. doi: 10.1111/all.13468.

Maggi, E. (1998). The TH1/TH2 paradigm in allergy. *Immunotechnology*, 3 (4): 233-244. doi: 10.1016/S1380-2933(97)10005-7.

Maidak, B. L., Larsen, N., McCaughey, M. J., Overbeek, R., Olsen, G. J., Fogel, K., Blandy, J. & Woese, C. R. (1994). The Ribosomal Database Project. *Nucleic acids research*, 22 (17): 3485-3487. doi: 10.1093/nar/22.17.3485.

Mayhew, J. W. & Gorbach, S. L. (1977). Internal standards for gas chromatographic analysis of metabolic end products from anaerobic bacteria. *Applied and Environmental Microbiology*, 33 (4): 1002-3. doi: 10.1128/aem.33.4.1002-1003.1977.

McMurdie, P. J. & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8 (4): e61217. doi: 10.1371/journal.pone.0061217.

McNeil, N. I., Cummings, J. H. & James, W. P. (1978). Short chain fatty acid absorption by the human large intestine. *Gut*, 19 (9): 819-22. doi: 10.1136/gut.19.9.819.

Menendez, A., Willing, B. P., Montero, M., Wlodarska, M., So, C. C., Bhinder, G., Vallance, B. A. & Finlay, B. B. (2013). Bacterial Stimulation of the TLR-MyD88 Pathway Modulates the

Homeostatic Expression of Ileal Paneth Cell α-Defensins. *Journal of Innate Immunity*, 5 (1): 39-49. doi: 10.1159/000341630.
Merck. (n.d.). *SpeedBeads™ magnetic carboxylate modified particles*. Available at: https://www.sigmaaldrich.com/NO/en/product/sigma/ge65152105050250.
Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., et al. (2017). The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiology and molecular biology reviews*, 81 (4). doi: 10.1128/mmbr.00036-17.
Moore, L. D., Le, T. & Fan, G. (2013). DNA Methylation and Its Basic Function. *Neuropsychopharmacology*, 38 (1): 23-38. doi: 10.1038/npp.2012.112.
Morrison, D. J. & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*, 7 (3): 189-200. doi: 10.1080/19490976.2015.1134082.
Muzzey, D., Evans, E. A. & Lieber, C. (2015). Understanding the Basics of NGS: From Mechanism to Variant Calling. *Current Genetic Medicine Reports*, 3 (4): 158-165. doi: 10.1007/s40142-015-0076-8.
Nichols, R. G. & Davenport, E. R. (2021). The relationship between the gut microbiome and host gene expression: a review. *Human Genetics*, 140 (5): 747-760. doi: 10.1007/s00439-020-02237-0.
Nilsen, M., Madelen Saunders, C., Leena Angell, I., Arntzen, M., Lødrup Carlsen, K. C., Carlsen, K. H., Haugen, G., Hagen, L. H., Carlsen, M. H., Hedlin, G., et al. (2020). Butyrate Levels in the Transition from an Infant- to an Adult-Like Gut Microbiota Correlate with Bacterial Networks Associated with Eubacterium Rectale and Ruminococcus Gnavus. *Genes (Basel)*, 11 (11). doi: 10.3390/genes11111245.
Nilsen, T. (2021). *The potential for Human Milk Oligosaccharide utilization by Bifidobacterium in 6 months children*. Master's thesis. Brage: Norwegian University of Life Sciences.
Noah, T. K., Donahue, B. & Shroyer, N. F. (2011). Intestinal development and differentiation. *Experimental cell research*, 317 (19): 2702-2710. doi: 10.1016/j.yexcr.2011.09.006.
Pajares, M. J., Palanca-Ballester, C., Urtasun, R., Alemany-Cosme, E., Lahoz, A. & Sandoval, J. (2021). Methods for analysis of specific DNA methylation status. *Methods*, 187: 3-12. doi: 10.1016/j.ymeth.2020.06.021.
Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., Barnich, N., Ng, S. C. & Colombel, J.-F. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Gut*, 67 (3): 574. doi: 10.1136/gutjnl-2017-314903.
Park, S.-Y. & Kim, J.-S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & Molecular Medicine*, 52 (2): 204-212. doi: 10.1038/s12276-020-0382-4.
Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T. & Sandhu, M. S. (2018). Long reads: their purpose and place. *Human Molecular Genetics*, 27 (R2): R234-R241. doi: 10.1093/hmg/ddy177.
Portela, A. & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, 28 (10): 1057-1068. doi: 10.1038/nbt.1685.
Primec, M., Mičetić-Turk, D. & Langerholc, T. (2017). Analysis of short-chain fatty acids in human feces: A scoping review. *Analytical Biochemistry*, 526: 9-21. doi: 10.1016/j.ab.2017.03.007.
Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. (2004). Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis. *Cell*, 118 (2): 229-241. doi: 10.1016/j.cell.2004.07.002.
Ramakrishna, B. S. (2013). Role of the gut microbiota in human nutrition and metabolism. *Journal of Gastroenterology and Hepatology*, 28 (S4): 9-17. doi: 10.1111/jgh.12294.
Rehbinder, E. M., Lødrup Carlsen, K. C., Staff, A. C., Angell, I. L., Landrø, L., Hilde, K., Gaustad, P. & Rudi, K. (2018). Is amniotic fluid of women with uncomplicated term pregnancies free of bacteria? *American Journal of Obstetrics and Gynecology*, 219 (3): 289.e1-289.e12. doi: 10.1016/j.ajog.2018.05.028.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74 (12): 5463-5467.
Sarkar, S., Abujamra, A. L., Loew, J. E., Forman, L. W., Perrine, S. P. & Faller, D. V. (2011). Histone deacetylase inhibitors reverse CpG methylation by regulating DNMT1 through ERK signaling. *Anticancer research*, 31 (9): 2723-32.
Schroeder, B. O. (2019). Fight them or feed them: how the intestinal mucus layer manages the gut microbiota. *Gastroenterology Report*, 7 (1): 3-12. doi: 10.1093/gastro/goy052.
Sender, R., Fuchs, S. & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, 14 (8): e1002533. doi: 10.1371/journal.pbio.1002533.
Sender, R. & Milo, R. (2021). The distribution of cellular turnover in the human body. *Nature Medicine*, 27 (1): 45-48. doi: 10.1038/s41591-020-01182-9.
Seto, E. & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*, 6 (4): a018713. doi: 10.1101/cshperspect.a018713.
Shapiro, R., Servis, R. E. & Welcher, M. (1970). Reactions of Uracil and Cytosine Derivatives with Sodium Bisulfite. *Journal of the American Chemical Society*, 92 (2): 422-424. doi: 10.1021/ja00705a626.
Shapiro, R., Braverman, B., Louis, J. B. & Servis, R. E. (1973). Nucleic acid reactivity and conformation. II. Reaction of cytosine and uracil with sodium bisulfite. *The Journal of biological chemistry*, 248 (11): 4060-4.
Shapiro, R., DeFate, V. & Welcher, M. (1974). Deamination cytosine derivatives by bisulfite. Mechanism of the reaction. *Journal of the American Chemical Society*, 96 (3): 906-912. doi: 10.1021/ja00810a043.
Simon, A. K., Hollander, G. A. & McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. *Proceedings. Biological sciences*, 282 (1821): 20143085-20143085. doi: 10.1098/rspb.2014.3085.
Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44 (4): 846-849. doi: 10.1099/00207713-44-4-846.
Sugi, Y., Takahashi, K., Kurihara, K., Nakano, K., Kobayakawa, T., Nakata, K., Tsuda, M., Hanazawa, S., Hosono, A. & Kaminogawa, S. (2017). α-Defensin 5 gene expression is regulated by gut microbial metabolites. *Bioscience, Biotechnology, and Biochemistry*, 81 (2): 242-248. doi: 10.1080/09168451.2016.1246175.
Takahashi, K., Sugi, Y., Hosono, A. & Kaminogawa, S. (2009). Epigenetic Regulation of TLR4 Gene Expression in Intestinal Epithelial Cells for the Maintenance of Intestinal Homeostasis. *The Journal of Immunology*, 183 (10): 6522. doi: 10.4049/jimmunol.0901271.
Takiishi, T., Fenero, C. I. M. & Câmara, N. O. S. (2017). Intestinal barrier and gut microbiota: Shaping our immune responses throughout life. *Tissue barriers*, 5 (4): e1373208-e1373208. doi: 10.1080/21688370.2017.1373208.
ThermoFisherScientific. (n.d.). *Qubit fluorometric quantification*. Available at: https://www.thermofisher.com/no/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html (accessed: 06.04).
Utheim, K. S. (2021). *Effect of Microbial Metabolites on Mitochondrial Function in Colonocytes*. Master's thesis. Brage: Norwegian University of Life Sciences.
Vital, M., Penton, C. R., Wang, Q., Young, V. B., Antonopoulos, D. A., Sogin, M. L., Morrison, H. G., Raffals, L., Chang, E. B., Huffnagle, G. B., et al. (2013). A gene-targeted approach to investigate the intestinal butyrate-producing bacterialcommunity. *Microbiome*, 1 (1): 8. doi: 10.1186/2049-2618-1-8.
Vital, M., Howe, A. C. & Tiedje, J. M. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *mBio*, 5 (2): e00889-e00889. doi: 10.1128/mBio.00889-14.

Wei, S., Tao, J., Xu, J., Chen, X., Wang, Z., Zhang, N., Zuo, L., Jia, Z., Chen, H., Sun, H., et al. (2021). Ten Years of EWAS. *Advanced Science*, 8 (20): 2100727. doi: 10.1002/advs.202100727.

Wemheuer, F., Taylor, J. A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T. & Wemheuer, B. (2020). Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. *Environmental Microbiome*, 15: 1-12. doi: 10.1186/s40793-020-00358-7.

Wikenius, E., Moe, V., Smith, L., Heiervang, E. R. & Berglund, A. (2019). DNA methylation changes in infants between 6 and 52 weeks. *Scientific Reports*, 9 (1): 17587. doi: 10.1038/s41598-019-54355-z.

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

Woese, C. R. (1987). Bacterial evolution. *Microbiological reviews*, 51 (2): 221-71. doi: 10.1128/mr.51.2.221-271.1987.

Wolffe, A. P. & Matzke, M. A. (1999). Epigenetics: regulation through repression. *Science*, 286 (5439): 481-6. doi: 10.1126/science.286.5439.481.

Ximenez, C. & Torres, J. (2017). Development of Microbiota in Infants and its Role in Maturation of Gut Mucosa and Immune System. *Archives of Medical Research*, 48 (8): 666-680. doi: 10.1016/j.arcmed.2017.11.007.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature*, 486 (7402): 222-227. doi: 10.1038/nature11053.

Ye, J., Wu, W., Li, Y. & Li, L. (2017). Influences of the Gut Microbiota on DNA Methylation and Histone Modification. *Digestive Diseases and Sciences*, 62 (5): 1155-1164. doi: 10.1007/s10620-017-4538-6.

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. *Biotechnology and Bioengineering*, 89 (6): 670-9. doi: 10.1002/bit.20347.

ZymoResearch. (n.d.-a). *Learn more about bisulfite conversion*. Available at: https://www.zymoresearch.com/pages/bisulfite-beginner-guide (accessed: 01.02.22).

ZymoResearch. (n.d.-b). *Zymo-Spin IC Columns*. Available at: https://www.zymoresearch.com/products/zymo-spin-ic-columns.

Appendix A: Primer sequences for index PCR of bisulfite converted DNA

TLR4 Forward (5’-3’)

F1, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaGTTGAGGTTTATTTTTAGTTTTGTATGTG

F2, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccGTTGAGGTTTATTTTTAGTTTTGTATGTG

F3, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaGTTGAGGTTTATTTTTAGTTTTGTATGTG

F4, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccGTTGAGGTTTATTTTTAGTTTTGTATGTG

F5, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagGTTGAGGTTTATTTTTAGTTTTGTATGTG

F6, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcGTTGAGGTTTATTTTTAGTTTTGTATGTG

F7, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaGTTGAGGTTTATTTTTAGTTTTGTATGTG

F8, TLR4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccGTTGAGGTTTATTTTTAGTTTTGTATGTG

TLR4 Reverse (5’-3’)

R1, TLR4R: caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R2, TLR4R: caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R3, TLR4R: caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R4, TLR4R: caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R5, TLR4R: caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R6, TLR4R: caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R7, TLR4R: caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R8, TLR4R: caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTACCTTACATACC

R9, TLR4R: caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R10, TLR4R: caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R11, TLR4R: caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R12, TLR4R: caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R13, TLR4R: caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R14, TLR4R: caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R15, TLR4R: caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

R16, TLR4R: caagcagaagacggcatacgagatGGACGGgtgactggagttcagacgtgtgctcttccgatctAACCTCATTCTA CCTTACATACC

DEFA5 Forward (5’-3’)

F1, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaTAGGAGGTTGAGGTAGGAGAAA

F2, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccTAGGAGGTTGAGGTAGGAGAAA

F3, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaTAGGAGGTTGAGGTAGGAGAAA

F4, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccTAGGAGGTTGAGGTAGGAGAAA

F5, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagTAGGAGGTTGAGGTAGGAGAAA

F6, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcTAGGAGGTTGAGGTAGGAGAAA

F7, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaTAGGAGGTTGAGGTAGGAGAAA

F8, DEFA5F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccTAGGAGGTTGAGGTAGGAGAAA

DEFA5 Reverse (5’-3’)

R17, DEFA5R: caagcagaagacggcatacgagatCTCTACgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R18, DEFA5R: caagcagaagacggcatacgagatGCGGACgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R19, DEFA5R: caagcagaagacggcatacgagatTTTCACgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R20, DEFA5R: caagcagaagacggcatacgagatGGCCACgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R21, DEFA5R: caagcagaagacggcatacgagatCGAAACgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R22, DEFA5R: caagcagaagacggcatacgagatCGTACGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R23, DEFA5R: caagcagaagacggcatacgagatCCACTCgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R24, DEFA5R: caagcagaagacggcatacgagatGCTACCgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R25, DEFA5R: caagcagaagacggcatacgagatATCAGTgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R26, DEFA5R: caagcagaagacggcatacgagatGCTCATgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R27, DEFA5R: caagcagaagacggcatacgagatAGGAATgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R28, DEFA5R: caagcagaagacggcatacgagatCTTTTGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R29, DEFA5R: caagcagaagacggcatacgagatTAGTTGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R30, DEFA5R: caagcagaagacggcatacgagatCCGGTGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R31, DEFA5R: caagcagaagacggcatacgagatATCGTGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

R32, DEFA5R: caagcagaagacggcatacgagatTGAGTGgtgactggagttcagacgtgtgctcttccgatctACATTATCCTTTAAT TCCATCCATATTATC

IL-4 Forward (5’-3’)

F9, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgAGGTTAGGAGATGGAGATTATTTTG

F10, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgAGGTTAGGAGATGGAGATTATTTTG

F11, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggAGGTTAGGAGATGGAGATTATTTTG

F12, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcAGGTTAGGAGATGGAGATTATTTTG

F13, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatAGGTTAGGAGATGGAGATTATTTTG

F14, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcAGGTTAGGAGATGGAGATTATTTTG

F15, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctAGGTTAGGAGATGGAGATTATTTTG

F16, IL-4F: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagAGGTTAGGAGATGGAGATTATTTTG

IL-4 Reverse (5’-3’)

R1, IL-4R: caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R2, IL-4R: caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R3, IL-4R: caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R4, IL-4R: caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R5, IL-4R: caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R6, IL-4R: caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R7, IL-4R: caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R8, IL-4R: caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R9, IL-4R: caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R10, IL-4R: caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R11, IL-4R: caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R12, IL-4R: caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R13, IL-4R: caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R14, IL-4R: caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R15, IL-4R: caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

R16, IL-4R: caagcagaagacggcatacgagatGGACGGgtgactggagttcagacgtgtgctcttccgatctTAAAACTACAAACACCTACCACCAC

IFN-ɣ Forward (5’-3’)

F9, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtttcgGAGTTTTGTTTTGTTATTTAGGTTGG

F10, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcgtacgGAGTTTTGTTTTGTTATTTAGGTTGG

F11, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgagtggGAGTTTTGTTTTGTTATTTAGGTTGG

F12, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctggtagcGAGTTTTGTTTTGTTATTTAGGTTGG

F13, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctactgatGAGTTTTGTTTTGTTATTTAGGTTGG

F14, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgagcGAGTTTTGTTTTGTTATTTAGGTTGG

F15, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctattcctGAGTTTTGTTTTGTTATTTAGGTTGG

F16, IFNF: aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcaaaagGAGTTTTGTTTTGTTATTTAGGTTGG

IFN-ɣ Reverse (5’-3’)

R17, IFNR: caagcagaagacggcatacgagatCTCTACgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R18, IFNR: caagcagaagacggcatacgagatGCGGACgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R19, IFNR: caagcagaagacggcatacgagatTTTCACgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R20, IFNR: caagcagaagacggcatacgagatGGCCACgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R21, IFNR: caagcagaagacggcatacgagatCGAAACgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R22, IFNR: caagcagaagacggcatacgagatCGTACGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R23, IFNR: caagcagaagacggcatacgagatCCACTCgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R24, IFNR: caagcagaagacggcatacgagatGCTACCgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R25, IFNR: caagcagaagacggcatacgagatATCAGTgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R26, IFNR: caagcagaagacggcatacgagatGCTCATgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R27, IFNR: caagcagaagacggcatacgagatAGGAATgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R28, IFNR: caagcagaagacggcatacgagatCTTTTGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R29, IFNR: caagcagaagacggcatacgagatTAGTTGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R30, IFNR: caagcagaagacggcatacgagatCCGGTGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R31, IFNR: caagcagaagacggcatacgagatATCGTGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

R32, IFNR: caagcagaagacggcatacgagatTGAGTGgtgactggagttcagacgtgtgctcttccgatctAATACCTATAATCCCAACTACTC

Appendix B: Primer sequences for 16S rRNA gene sequencing

Illumina primer sequences for amplicon PCR of the 16S rRNA gene (Yu *et al.*, 2005).

PRK341F primer: 5'-CCTACGGGRBGCASCAG-3'

PRK806R primer: 5'-GGACTACYVGGGTATCTAAT-3'

Illumina primer sequences for index PCR of the 16S rRNA gene

Forward (5'-3')

1. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagtcaaCCTACGGGRBGCASCAG

2. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctagttccCCTACGGGRBGCASCAG

3. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctatgtcaCCTACGGGRBGCASCAG

4. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctccgtccCCTACGGGRBGCASCAG

5. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtagagCCTACGGGRBGCASCAG

6. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtccgcCCTACGGGRBGCASCAG

7. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtgaaaCCTACGGGRBGCASCAG

8. aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctgtggccCCTACGGGRBGCASCAG

Reverse (5'-3')

1. caagcagaagacggcatacgagatCGTGATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

2. caagcagaagacggcatacgagatACATCGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

3. caagcagaagacggcatacgagatGCCTAAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

4. caagcagaagacggcatacgagatTGGTCAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

5. caagcagaagacggcatacgagatCACTCTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

6. caagcagaagacggcatacgagatATTGGCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

7. caagcagaagacggcatacgagatGATCTGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

8. caagcagaagacggcatacgagatTCAAGTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

9. caagcagaagacggcatacgagatCTGATCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

10. caagcagaagacggcatacgagatAAGCTAgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

11. caagcagaagacggcatacgagatGTAGCCgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

12. caagcagaagacggcatacgagatTACAAGgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

13. caagcagaagacggcatacgagatTTGACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

14. caagcagaagacggcatacgagatGGAACTgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

15. caagcagaagacggcatacgagatTGACATgtgactggagttcagacgtgtgctcttccgatctGGACTACYVGGGTATCTAAT

Appendix C: Specifications for gas chromatography*

*The specifications are part of the procedure provided by the laboratory technicians

Injector:

Mode: split

Temperature: 250 °C

Carrier gas: Helium

Column flow: 2.5 ml/min

Split flow: 200 ml/min

Purge flow: 3 ml/min

Injection volume: 0.2 µl

Liner: 4mm x 6.3mm x 78.5mm (Catalog# 23311.5, Restek)

Syringe: 10 µl syr FN 50 mm C, Ga 23, cone tip (catalog# 365D3741, ThermoFisher Scientific)

Column:

Stabilwax DA 30m, 0.25 mm ID, 0.25 µm (Restek)

Temperature program: 90 °C to 150 °C (6 min), 150 °C to 245 °C (1.9 min)

Time per sample: 14.9 min

Detector:

Type: FID

Temperature: 275 °C

Hydrogen: 30 ml/min

Air: 300 ml/min

Makeup gas: 30 ml/min

Appendix D: Spearman's correlations

The spearman's correlation plots are made in cooperation with Oda Hamarheim with an R-script adapted from Tonje Nilsen's master thesis (Nilsen, 2021).

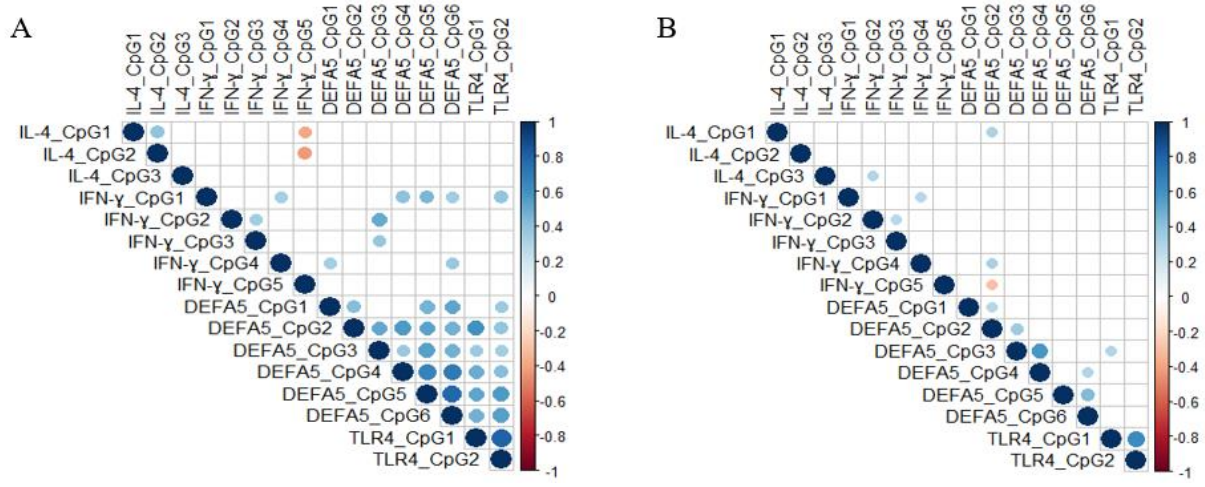

Figure D.1. Spearman's correlation between all CpG-positions in (A) the 6-month age group and (B) the 12-month age group. Significant correlations (unadjusted $p < 0.05$) are shown as blue dots for positive correlations and red dots for negative correlations, and the strength of the color is proportional to the correlation.

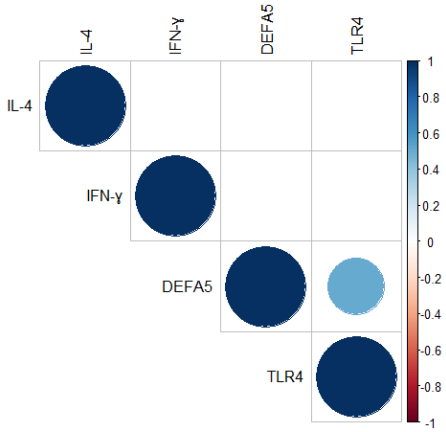

Figure D.2. Spearman's correlation between the mean DNA methylation level of all genes in the 6-month age group. A significant correlation (unadjusted $p < 0.05$) is shown as a blue dot for a positive correlation, and the strength of the color is proportional to the correlation.

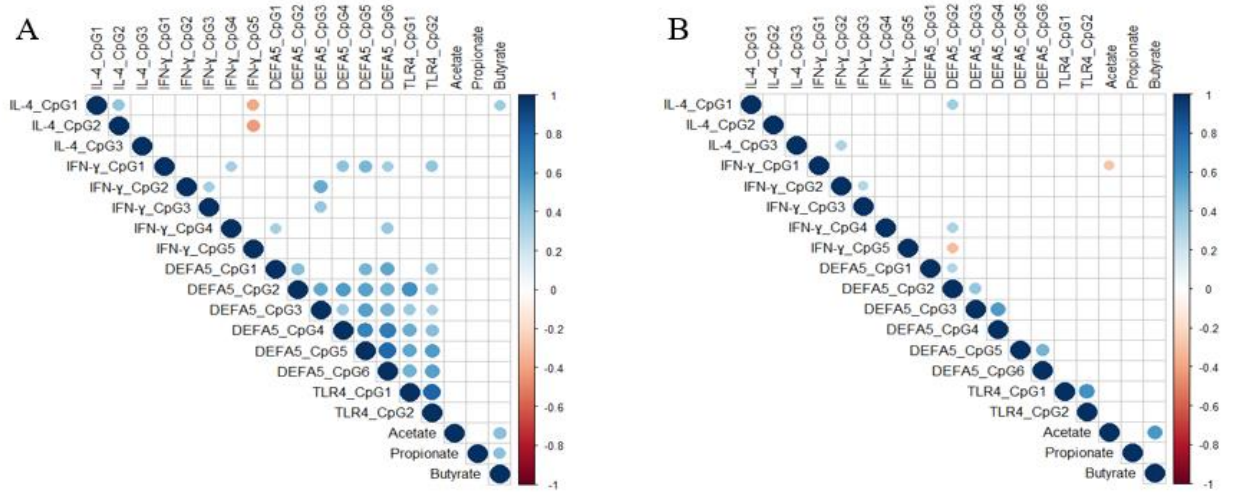

Figure D.3. Spearman's correlation between SCFAs (mmol/kg) and CpG-positions in (A) the 6 month-group and (B) the 12-month group. Significant correlations (unadjusted $p < 0.05$) are shown as blue dots for positive correlations and red dots for negative correlations, and the strength of the color is proportional to the correlation.

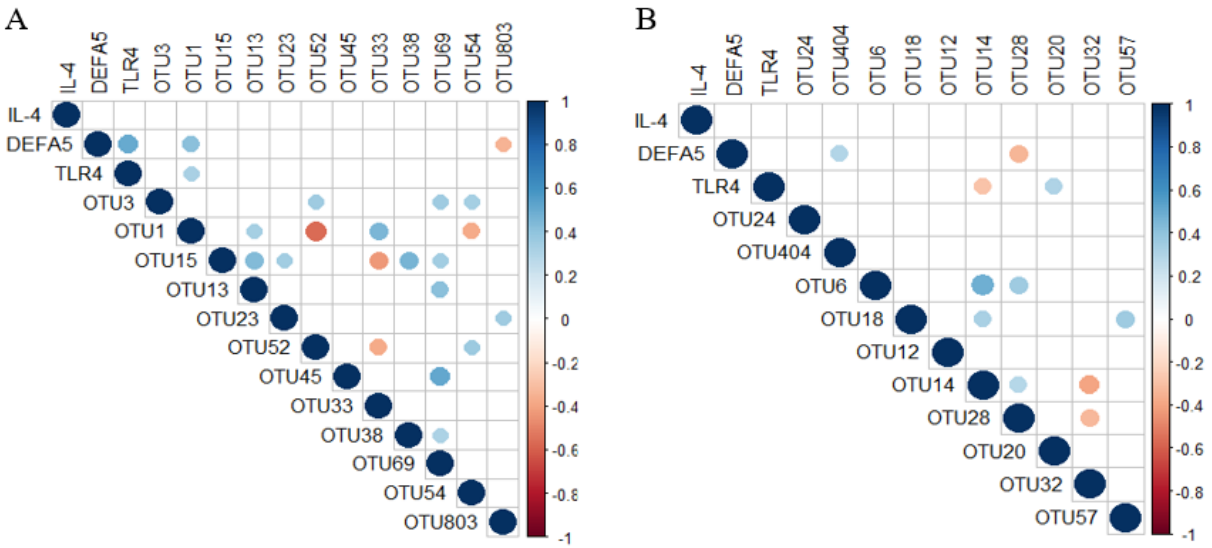

Figure D.4. Spearman's correlation between the genes with a significant difference in DNA methylation level between the age group and OTUs with a significant association to (A) the 6-month age group and (B) the 12-month age group. Significant correlations (unadjusted $p < 0.05$) are shown as blue dots for positive correlations and red dots for negative correlations, and the strength of the color is proportional to the correlation.

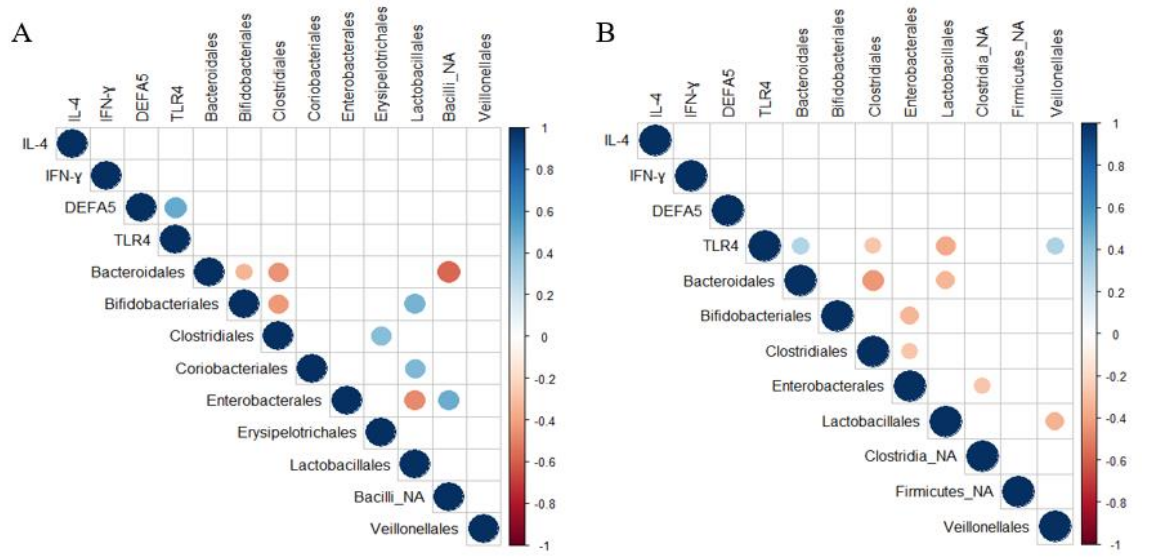

Figure D.5. Spearman's correlation between bacterial orders and the mean DNA methylation level of each gene in (A) the 6 month-group and (B) the 12-month group. OTUs with an abundance lower than 0.5 % were excluded before dividing into orders. Significant correlations (unadjusted $p < 0.05$) are shown as blue dots for positive correlations and red dots for negative correlations, and the strength of the color is proportional to the correlation. Bacteria that are not decided to the order level are indicated with "_NA". The DNA methylation level is calculated as the mean of all CpG-positions in all infants.

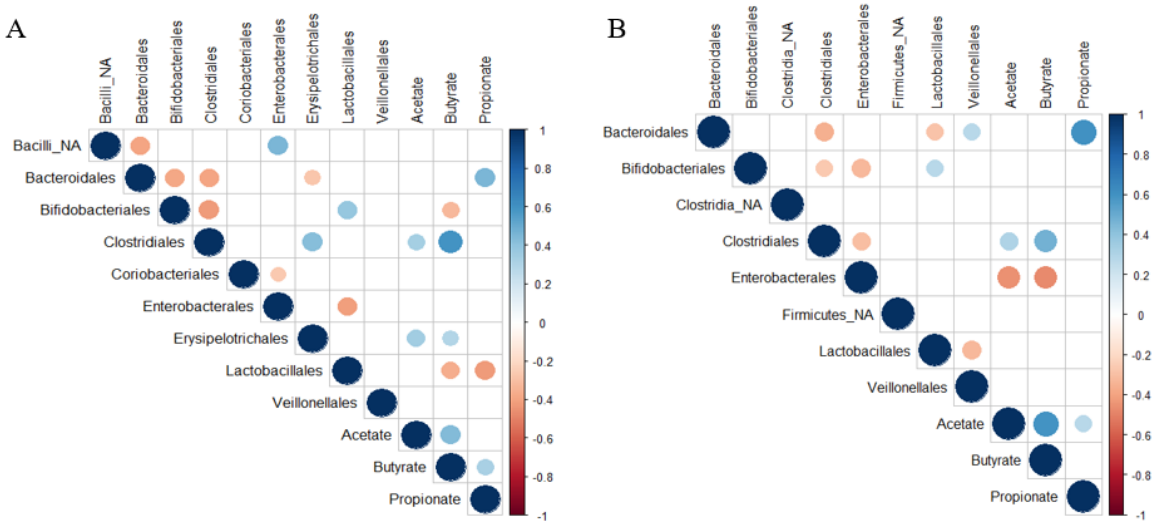

Figure D.6. Spearman's correlation between SCFAs (mmol/kg) and bacterial orders in (A) the 6 month-group and (B) the 12-month group. OTUs with an abundance lower than 0.5 % were excluded before dividing into orders. Significant correlations (unadjusted $p < 0.05$) are shown as blue dots for positive correlations and red dots for negative correlations, and the strength of the color is proportional to the correlation. Bacteria that are not decided to the order level are indicated with "_NA".

Appendix E: R-codes

Data input for the ANOSIM and indicator species analysis were prepared with help from Ph.D student Morten Nilsen, and the analyses were done with the R-codes below.

```
library(vegan)

ano = anosim(Tabell_uten_alder, Tabell_med_alder$Age, distance = "bray",

permutations = 9999)
ano
```

```
library(indicspecies)

inv = multipatt(Tabell_uten_alder, Tabell_med_alder$Age, func = "r.g",

control = how(nperm=9999))
```

Preprocessing of the data with the Phyloseq package to make the alpha diversity plot and the stacked bar chart was done with help from Ph.D student Morten Nilsen. The R-code used for visualization of the data is shown below.

```
library(ggplot2)
physeq1@sam_data$Age <- factor(physeq1@sam_data$Age, levels = c("6 mths", "12
mths"))

alpha <- plot_richness(physeq = physeq1, x = "Age", measures=c("Observed", "S
impson"), color = "Age")

alpha +scale_color_manual(values=c("#0072B2", "#D55E00"))+ theme(axis.text.x
= element_text(color="#000000",

size=18)) +
  theme(axis.title = element_text(size = 18))+
  theme(legend.text = element_text(size = 14))+ #6 og 12
  theme(legend.title = element_text(size = 16))+
  theme(plot.title = element_text(size = 18))+
  theme(axis.text.y = element_text(size = 14))+
scale_x_discrete(breaks=c("6 mths", "12 mths"),
                 labels=c("6 months", "12 months"))
```

```r
GPr  = transform_sample_counts(physeq1, function(x) x / sum(x) *100 )

taxglom <- tax_glom(GPr, taxrank = "Order")
melt <- psmelt(taxglom)
order_level <- aggregate(Abundance ~ Age + Order, data = melt, FUN = mean)

order_level$Age <- factor(order_level$Age, levels = c("6 mths", "12 mths"))

over_05 <- which(order_level$Abundance > 0.5)
over_05 <- order_level[over_05,]
all_12 <- which(over_05$Age == "12 mths")
Sum_12 <- sum(over_05$Abundance[all_12])
Other_12 <- 100 - Sum_12

All_6 <- which(over_05$Age == "6 mths")
Sum_6 <- sum(over_05$Abundance[All_6])
Other_6 <- 100 - Sum_6

Twelve <- c("12 mths", "Other", Other_12)
Six <- c("6 mths", "Other", Other_6)

Test_tabell_plot2 <- rbind(over_05, Twelve)
plot2 <- rbind(Test_tabell_plot2, Six)
plot2$Abundance <- as.numeric(plot2$Abundance)

library(ggplot2)
t <- ggplot(plot2, aes(x = Age, y = Abundance, fill = Order)) + geom_bar(stat
= "identity", width = 0.6) +
  theme_bw() + scale_fill_manual(values=c("#0072B2", "#CC79A7", "#56B4E9","#0
00000","#F0E442", "#E69F00","#009E73","#D55E00", "#999999"))



t+theme(axis.title = element_text(size = 18))+
  theme(legend.text = element_text(size = 18))+
  theme(legend.title = element_text(size = 18))+
  theme(plot.title = element_text(size = 18))+
  theme(axis.text.y = element_text(size = 14))+
  scale_x_discrete(breaks=c("6 mths", "12 mths"),
                   labels=c("6 months", "12 months"))+
  theme(axis.text.x = element_text(size = 14))
```

The Spearman's correlations were done as shown in the example code below which is adapted from the R-code in the master's thesis of Tonje Nilsen (Nilsen, 2021).

```
library(RcmdrMisc)

library(corrplot)

scfatax6 <- rcorr.adjust(correlation_scfa_bacterialorders_6mths,

type = c("spearman"))

corrplot(scfatax6$R$r,
         method = "circle",
         type = "upper",
         p.mat = scfatax6$R$P,
         sig.level = 0.05,
         insig = "blank",
         tl.col = "black",
)
```

The strip charts for visualization of the DNA methylation level were made as shown in the example R-code below.

```
library(ggplot2)

CpG <- factor(IFN_stripchart$`CpG-position`, levels = c("1", "2","3","4","5")
)

pIL <- ggplot(IFN_stripchart, aes(x=`CpG`, y=`DNA methylation level (%)`,

color=Age)) +
  geom_jitter(position=position_dodge(0.2))+
  labs(title="B, IFN-ɣ")+
  theme_classic()+
  stat_summary(fun.data="mean_sdl", fun.args = list(mult=1), geom="crossbar",
width=0.5)


pIL+scale_color_manual(values=c("#0072B2", "#D55E00"))+

 theme(axis.text.x = element_text(color="#000000",

size=18)) +
  theme(axis.title = element_text(size = 18))+
  theme(legend.text = element_text(size = 14))+
```

```
  theme(legend.title = element_text(size = 16))+
  theme(plot.title = element_text(size = 18))+
  theme(axis.text.y = element_text(size = 14))
```

Data input for KEGG was made by use of Tax4Fun with help from Ph.D student Morten Nilsen. A reference database was built using the functions buildReferenceData() and buildDependencies() prior to running the R-code below.

```
library(Tax4Fun2)


runRefBlast(
  path_to_otus = "test/otus.fa",
  path_to_reference_data = "database/Tax4Fun2_ReferenceData_v2",
  path_to_temp_folder ="temp/",
  database_mode = "Ref99NR",
  use_force = T,
  num_threads = 4)

makeFunctionalPrediction(path_to_otu_table = "temp/otutabny_ferdig.txt",

path_to_reference_data = "database/Tax4Fun2_ReferenceData_v2",

path_to_temp_folder = "temp/", database_mode = "Ref99NR",

normalize_by_copy_number = TRUE,

min_identity_to_reference = 0.97, normalize_pathways = TRUE)
```

Appendix F: RNA extraction

This part of the work was done as an attempt to study the gene expression from infant feces. This was not successful and was therefore excluded from the main study. In the first attempt, RNA was extracted from 5 infant fecal samples stored in a DNA/RNA shield (Zymo Research, USA). This was done with the MagMAX™-96 Total RNA Isolation Kit (Invitrogen™, USA) following the manufacturer's procedure, except for the reduction of elution buffer volume to 30 µl.

FIREScript® RT cDNA synthesis MIX (Solis BioDyne, Estonia) was used to synthesize cDNA using the extracted RNA. The reaction mixture consisted of 10 µl template RNA, 1x RT Reaction Premix with Random primers, 1.5 µl FIREScript Enzyme Mix and nuclease-free water to a total volume of 20 µl. For synthesis, primer annealing was done at 25 °C for 8 minutes, reverse transcription at 50 °C for 60 minutes, and enzyme inactivation at 85 °C for 5 minutes. The newly synthesized cDNA was measured with Qubit™ dsDNA HS Assay Kit (Invitrogen, USA) and run on a 2 % agarose gel. To further check for presence of cDNA, qPCR with primers targeting GAPDH was done. The reaction mixture consisted of 3 µl cDNA, 1x HOT FIREPol® EvaGreen® qPCR supermix (Solis BioDyne, Estonia), 0.5 µM forward primer (5'- CCA CAT CGC TCA GAC ACC AT -3') 0.5 µM reverse primer (5'- GCG CCC AAT ACG ACC AAA T -3') and nuclease-free water to a final volume of 10 µl. The program used had an initial step at 94 °C for 10 minutes, followed by 40 cycles at 94 °C for 19 seconds and 60 °C for 60 seconds. A melt curve analysis was done with an initial step at 95 °C for 15 seconds, followed by an increase in temperature from 60 °C to 94 °C with an increase rate of 1 minute per °C.

As the first attempt did not give high enough concentrations to be measured by qubit or to give a band on the visualized gel, the method for RNA extraction was changed. In the second attempt, RNA was extracted from fecal samples from 2 6-month-old infants by use of TRIzol™ Reagent (Invitrogen™, USA). The fecal samples, which were stored in DNA/RNA shield at -80 °C, were thawed on ice and homogenized prior to centrifugation of 0.5 ml sample in Eppendorf tubes with open lids in a Savant SpeedVac concentrator (Thermo Scientific, USA). This was done to concentrate the samples before proceeding with the RNA extraction using TRIzol™ Reagent (Invitrogen, Thermo Fisher Scientific) and PureLink® RNA Mini Kit (Invitrogen, USA). The fecal samples were reduced to 0.25 ml after approximately 12 hours in the SpeedVac concentrator. To lyse the cells, 0.75 TRIzol® Reagent were added to each sample and homogenized by pipetting.

Further RNA isolation was done according to the protocol provided by the manufacturer of the kit, with use of 0.15 ml chloroform per sample. Thirty µl were eluted from each sample in 3 sequences and the RNA concentration was measured with Qubit™ RNA HS Assay Kit (Invitrogen™, USA). Additionally, 5 µl of each eluate was run on a 1.5 % agarose gel with 4 µl 1kB ladder (Solis Biodyne). RNA was not detected by either the qubit measurement or gel electrophoresis.

Appendix G: Technical aspects regarding sequencing of bisulfite converted DNA

There were bands in some of the negative controls after gel electrophoresis of bisulfite converted DNA. In all negative controls belonging to *TLR4*, there was a weak smear both after amplicon PCR and index PCR. For *DEFA5*, one of the in total three negative controls had a band at the same length as the samples after amplicon PCR. After index PCR, a band with a shorter base-pair length was seen for the two remaining negative controls of *DEFA5*. No bands were seen in the negative controls of *IL-4* and *IFN-γ* after amplicon PCR, but bands with a shorter base-pair length were seen after index PCR. The positive controls from Caco-2 cells showed bands at expected band lengths for all genes, but with an additional smear for *TLR4* and some additional bands for *IL-4*. The samples of *IFN-γ*, IL-4, and *DEFA5* showed bands at the expected band length, while all samples of TLR4 showed a smear at the gel.

The sequencing reads were aligned to the target sequence to identify reads from the correct gene sequence. The percentage of reads that mapped to the target sequence varied from approximately 1 to 38 % (table G.1). Reads that did not map to the target sequence were excluded. The percentage of reads in the negative controls was low, except for one of the negative controls in the 6-month age group of *DEFA5* (Table G.2).

Table G.1. The percentage of reads that mapped to the target sequence for the 6-month age group and the 12-month age group. The total number of reads belonging to each age group of a gene ranged from 764 543 reads to 20 398 129 reads.

	Percentage of mapped reads	
	6 months	12 months
TLR4	38 %	11 %
DEFA5	2 %	1 %
IL-4	31 %	30 %
IFN-γ	1 %	1 %

Table G.2. The percentage of reads in the negative controls. The negative controls were compared to the positive controls by calculating the mean number of mapped reads of the negative controls to the mean number of mapped reads of the positive controls. The higher number of reads in the negative controls of the 6-month age group for *DEFA5* was caused by contamination in one of the three negative controls.

	Percentage of reads in the negative controls	
	6 months	12 months
TLR4	1 %	0.1 %
DEFA5	31 %	2 %
IL-4	0.2 %	0.1 %
IFN-γ	0.2 %	0.2 %

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

Postboks 5003
NO-1432 Ås
Norway