Understanding how infant factors shape the gut microbiome in preterm infants at risk of necrotising enterocolitis

Lauren Caitlin Beck

A Thesis Submitted for the Degree of Doctor of Philosophy

Translational and Clinical Research Institute

Faculty of Medical Sciences

November 2023

Abstract

Very preterm infants born <32 weeks of gestation are at increased risk of various diseases such as necrotising enterocolitis (NEC), a devastating gastrointestinal disorder affecting up to 10% of preterm infants. Origins of NEC are multifactorial, associated with both a naïve immune system and disturbances to the gut microbiome. This thesis aimed to characterise the gut microbiome of preterm infants at risk of NEC and identify factors that may impact this development. Analysing longitudinal stool samples (*n* = 1431) from infants who did not develop NEC, revealed probiotics to be the main driver of the gut microbiota in this 'healthy' population, driving transition into different preterm gut community types (PGCTs) dependent on probiotic type. Functional analyses identified PGCT-associated stool metabolites and in preterm-derived organoids, sterile faecal supernatants impacted intestinal organoid monolayer gene expression in a PGCT-specific manner. Comparing healthy infants to those who developed NEC (*n* = 75 infants, *n* = 547 samples) revealed that preceding diagnosis, the relative abundance of Proteobacteria such as *Klebsiella variicola* was higher whilst the relative abundance of Actinobacteria, mainly Bifidobacteria was lower, concordant with other studies. Exploring the microbial origins of NEC in the context of probiotics, both the type and administration of probiotics influenced microbial associations with NEC, in particular the colonisation patterns of different probiotic-associated strains. Finally, the complex relationship between temporal development of the gut microbiome, metabolome and circulating T-lymphocytes in the preterm population was characterised. Greatest concordance was found between the gut microbiome and metabolome, with little relationship observed between the gut environment and circulating T-lymphocytes. Together, the results described provide important insights into gut microbiome development in preterm infants during early life, the modifiable factors associated with modulation of the gut microbiome, and how this might be tailored to improve gut health.

Acknowledgements

I'd like to start off with a huge thank-you to Dr. Christopher Stewart, who I have been incredibly fortunate to have as my supervisor. Thank-you for all your help and guidance, for all the opportunities which you have given me, and for your endless support (especially when navigating 'the organoid era'). I am truly grateful for everything. I would also like to thank Dr. Christopher Lamb, who has been equally as supportive throughout and whom I have been very lucky to have as my secondary supervisor. Thank-you both so so much. My thanks also extend to my other co-supervisors, Dr. Janet Berrington and Prof. Nicholas Embleton, for all your clinical wisdom and guidance. It would be a glaring oversight not to mention the NICU staff, in particular, Dr. Janet Berrington for managing all the patient clinical data and Julie Groombridge as well as the other NICU staff for collecting and storing samples. And of course, a special thank-you to the parents and babies who consented to the use of samples, without which none of this work would have been at all possible.

A huge thanks to all Stewart/Lamb lab group members both past and present, for bringing me lots of laughs and being a welcomed distraction when I needed it most. Working with such a nice group of people has made this so much easier and sometimes even enjoyable?

Finally, I want to thank all my friends and family – you have all been instrumental in ensuring I had a life outside of science. Especially big hugs to my parents and my sisters, for always being there for me. And last but by no means least, Curtis, who really has been my rock during this journey. Thank-you so much for always being there for me, I couldn't have done it without you.

Key publications

The following publications have formed part of the introduction and large parts of Chapter 3:

- **Beck, L.C.,** Granger, C.L., Masi, A.C., Stewart, C.J. (2021). Use of omic technologies in early life gastrointestinal health and disease: from bench to bedside. Expert Rev Proteomics 1–13. https://doi.org/10.1080/14789450.2021.1922278

- **Beck, L.C.,** Masi, A.C., Young, G.R., Vatanen, T., Lamb, C.A., Smith, R., Coxhead, J., Butler, A., Marsland, B.J., Embleton, N.D., Berrington, J.E., Stewart, C.J. (2022). Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants. Nature Microbiology 2020 7:10 7, 1525–1535. https://doi.org/10.1038/s41564-022-01213-w

Additional publications:

- Embleton, N.D., Sproat, T., Uthaya, S., Young, G.R., Garg, S., Vasu, V., Masi, A.C., **Beck, L.,** Modi, N., Stewart, C.J., Berrington, J.E. (2023). Effect of an Exclusive Human Milk Diet on the Gut Microbiome in Preterm Infants: A Randomized Clinical Trial. JAMA Netw Open 6, e231165–e231165. https://doi.org/10.1001/JAMANETWORKOPEN.2023.1165

- Powell J, Powell S, Mather M.W, **Beck L,** Nelson A, Palmowski P, Porter A, Coxhead J, Hedley A, Scott J, Rostron A.J., Hellyer T.P., Zaidi F, Davey T, Garnett J.P., Agbeko R, Ward C, Stewart C.J., Taggart C.C., Brodlie M, Simpson J. (2023). Tracheostomy in children is associated with neutrophilic airway inflammation.Thorax, 78:1019-1027.

- Young, G., Berrington, J. E., Cummings, S., Dorling, J., Ewer, A. K., Frau, A., Lett, L., Probert, C., Juszczak, E., Kirby, J., **Beck, L. C.,** Renwick, V. L., Lamb, C., Lanyon, C. V., McGuire, W., Stewart, C., & Embleton, N. (2022). Mechanisms affecting the gut of preterm infants in enteral feeding trials: a nested cohort within a randomised controlled trial of lactoferrin. Archives of Disease in Childhood - Fetal and Neonatal Edition. https://doi.org/10.1136/ARCHDISCHILD-2022-324477

- Granger, C.L., Lamb, C.A., Embleton, N.D., **Beck L.C.,** Masi A.C., Palmer J.M., Stewart C.J., Berrington J.E. et al. (2022). Secretory immunoglobulin A in preterm infants: determination of normal values in breast milk and stool. Pediatr Res 92, 979–986. https://doi.org/10.1038/s41390-021-01930-8

- Granger, C., Dermyshi, E., Roberts, E., **Beck, L. C.,** Embleton, N., & Berrington, J. (2022). Necrotising enterocolitis, late-onset sepsis and mortality after routine probiotic introduction in the UK. Archives of Disease in Childhood - Fetal and Neonatal Edition, 107(4), 352–358. https://doi.org/10.1136/ARCHDISCHILD-2021-32225

Table of Contents

Abbreviations

ANOVA – Analysis of variance

ARG – Antibiotic resistance gene

BHB – 3-hydroxybutyrate

BMF – Breast milk fortifier

CD – Cluster of differentiation

CFU – Colony forming units

CGA – Corrected gestational age

CI – Confidence interval

CMGF – Complete media with growth factors

CyTOF – Cytometry by time of flight

DEG – Differentially expressed gene

DHM – Donor human milk

DMM – Dirichlet multinomial mixtures

DNA – Deoxyribonucleic acid

DOL – Day of life

DSLNT – Disialyllacto-N-tetraose

EC – Enzyme Commission

EDTA – Ethylenediaminetetraacetic acid

ELBW – Extremely low birthweight

ELISA – Enzyme-linked immunosorbent assay

EOS – Early onset sepsis

ESI – Electrospray ionisation

FC – Flow cytometry

FDR – False discovery rate

FIP – Focal intestinal perforation

FoxP3 – Forkhead box-P3

GAGs – Glycosaminoglycans

GC – Gas chromatography

GF – Germ free

GI – Gastrointestinal

GLMM – Generalised linear mixed effects model

glog – Generalised logarithmic

GO – Gene ontology

HILIC – Hydrophilic interaction chromatography

HMO – Human milk oligosaccharides

IgA – Immunoglobulin A

IgG – Immunoglobulin G

IGIP – IgA inducing protein

IL-8 – Interleukin 8

IL18R – Interleukin 18 receptor 1

IL23R – Interleukin 23 receptor 1

INDIGO – Interactions between diet, gut microbes & metabolism in preterm infants

IQR – Interquartile range

LC-MS – Liquid-chromatography mass spectrometry

LEfSe – Linear discriminant analysis effect size

LENG1 – Leukocyte receptor cluster member 1

LMM – Linear mixed effects model

LOESS – Locally estimated scatterplot smoothing

LOS – Late onset sepsis

LPS – Lipopolysaccharide

MAIT – Mucosal-associated invariant T-cells

MAZ – Microbiota for age Z-score

MCIA – Multiple co-inertia analysis

MOM – Mother's own milk

MS – Mass spectrometry

NEC – Necrotising enterocolitis

NICE – National Institute for Health and Care Excellence

NICU – Neonatal intensive care unit

NMDS – Non-metric multidimensional scaling

NMR – Nuclear magnetic resonance

OACC – Organoid anaerobe co-culture

OTU – Operational taxonomic unit

PA – Procrustes analysis

PBMC – Peripheral blood mononuclear cells

PBS – Phosphate buffer saline

PCA – Principal component analysis

PCoA – Principal co-ordinate analysis

PERMANOVA – Permutational multivariate analysis of variance

PGCT – Preterm gut community type

PLS-DA – Partial least-squares discriminant analysis

PN – Parenteral nutrition

PTPT – Preterm T-lymphocyte profile type

QC – Quality control

R^2 – Coefficient of determination

RCT – Randomised controlled trial

RDA – Redundancy analysis

rlog – Regularised logarithm

RNA – Ribonucleic acid

RNA-seq – RNA-sequencing

rRNA – Ribosomal RNA

RP – Reverse phase

RVI – Royal Victoria Infirmary

SCFA – Short chain fatty acids

SERVIS – Supporting Enhanced Research in Vulnerable Infants Study

SIP – Spontaneous intestinal perforation

TCR – T cell receptor

TEER – Transepithelial electrical resistance

Th17 – T helper 17

TLR – Toll-like receptor

TNFSF15 – Tumour-necrosis-factor superfamily member 15

Tregs – Regulatory T-cells

UPLC – Ultra-performance liquid chromatography

VLBW – Very low birthweight

Table of Figures

Table of Tables

1. Introduction

1.1. The human microbiome

Microorganisms have evolved to live in symbiosis with humans, colonising various parts of the human body (i.e., skin and mucosal environments) and forming the human microbiome. The human microbiome spans all domains of life, including viruses, archaea, fungi and bacteria. There are estimated to be roughly 3.8 x 10^{13} bacterial cells associated with the human body, with roughly a 1:1.3 ratio of human:bacterial cells (Sender et al., 2016). Importantly, these estimations do not account for other microorganisms including viruses and fungi, which when combined together likely greatly outnumber human cells. In addition, when considering the functionality of the gut microbiome, there are substantially more bacterial genes (~100X) associated with the human body than there are human genes, offering much more in terms of genetic diversity and flexibility (Gilbert et al., 2018).

Most human-associated microorganisms colonise along the gastrointestinal (GI) tract, which broadly consists of the oral cavity, oesophagus, stomach, small intestine (made up of the duodenum, jejunum, and ileum) and large intestine (made up of the cecum, colon, rectum and anus) (**Figure 1.1)**. Of these, most are attributed to the colon where there are an estimated 10^{11} bacterial cells per ml, and the ileum where there are an estimated 10^{8} bacterial cells per ml (Sender et al., 2016). The gut microbiome therefore refers to the entire community of microorganisms living in the gut, their genomes (i.e., genes) and the GI environment in which they reside.

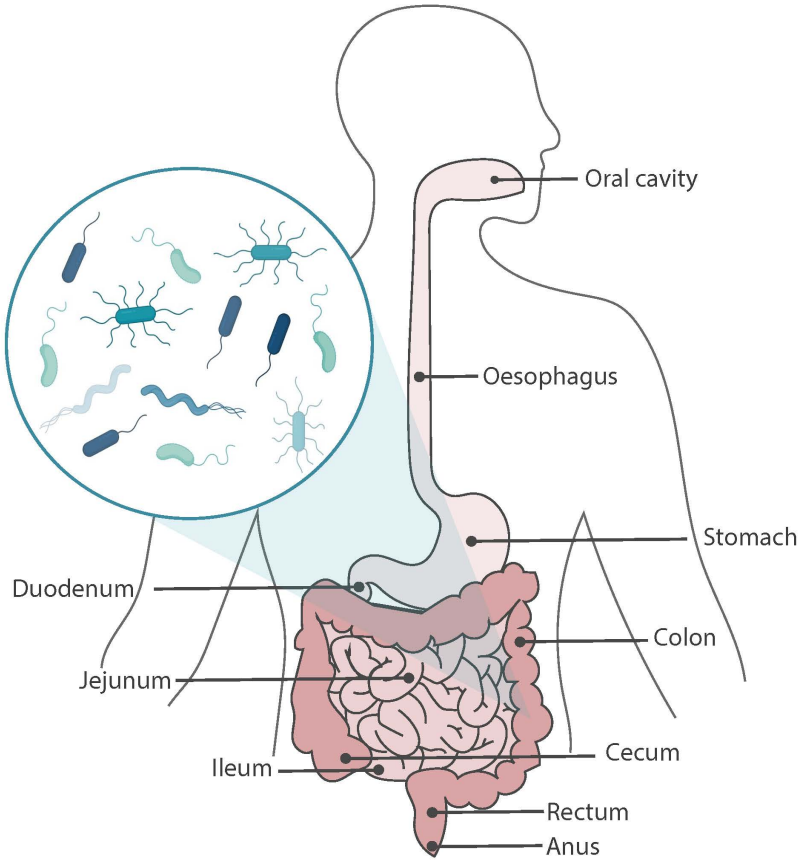

Figure 1.1. Diagram of the human GI tract and accompanying gut microbiota. The highest concentration of microbes can be found in the colon region of the large intestine. Created in part with Biorender.com.

1.2. The role of the gut microbiome in health and disease

The gut microbiome plays a fundamental role in host activities, aiding the digestion and breakdown of dietary compounds, and modulating the immune system. Over years of co-evolution, paralleled with changes in diet and lifestyle (Muegge et al., 2011), an intimate relationship has been forged between the gut microbiota and the immune system. Studies in germ-free (GF) animal models have helped to develop the concept of immune training and imprinting by the microbiota, demonstrating the absence of commensal microbes to be associated with improper immune functioning and general development (Zheng et al., 2020). Of note, GF mice have been found to have reduced levels of immunoglobulin A (IgA) and absence of T helper 17 (Th17) cells, which play important roles in defence against extracellular pathogens, particularly at mucosal sites (Khader et al., 2009). Notably, both IgA and Th17 cells can be restored/induced by the colonisation of microbial commensals (Hapfelmeier et al., 2010; Ivanov et al., 2008). In addition, various other studies have demonstrated evidence of the gut microbiota in

modulating regulatory T cells (Tregs), which help maintain homeostasis and self-tolerance (Kondělková et al., 2010). For example, commensals such as *Bacteroides fragilis* have been shown to direct the development of Tregs by stimulating the expression of the transcription factor: Forkhead box-P3 (FoxP3) (Round & Mazmanian, 2010) and members of the *Clostridium* genus have been shown to promote accumulation of these cells in the colon (Atarashi et al., 2011).

Aside from the roles of the gut microbiota in immune imprinting, multiple studies have evidenced the idea of "colonisation resistance" whereby the gut microbiota can prevent the colonisation of potential pathogens and pathobionts through multiple mechanisms such as via the secretion of antimicrobials, nutrient competition and by supporting gut barrier integrity (Ducarmon et al., 2019).

Notwithstanding, the association of the gut microbiome with health and disease has long been established, disorders of which can lead to a diverse range of GI and systemic diseases. For example, the gut microbiome has been linked to diseases such as inflammatory bowel disease (Yilmaz et al., 2019), necrotising enterocolitis (NEC) (Olm et al., 2019; Stewart et al., 2012), obesity (Davis, 2016), diabetes (Gurung et al., 2020), asthma (Hufnagl et al., 2020) and cardiovascular disease in later life (Zhernakova et al., 2018). Although, in many cases observations are of general 'dysbiosis', or else specific bacteria are not consistently found across studies. The association of the gut microbiome with the occurrence of diseases is inherently complex, not least because a healthy intestinal milieu is the product of a whole network of interactions and associations.

1.3. Early life gut microbiome development in term infants

It is widely accepted that the gut microbiome is formed at birth, following the colonisation of microorganisms along the GI tract. Despite this topic being subject to considerable debate in recent years, with several research groups suggesting the placenta and amniotic fluid to be a reservoir for microbes (Aagaard et al., 2014; Collado et al., 2016), recent evidence has shown neonatal meconium (i.e., stool formed before birth) to have no detectable microbiota (Kennedy et al., 2021) and that the so-called placental microbiome could not be distinguished from kit negative control contamination (Lauder et al., 2016). These data suggest that any microbial signatures

obtained from 'low biomass' foetal samples are instead likely the result of contamination during the clinical procedure, DNA extraction and/or sequencing (Kennedy et al., 2023).

Studies have shown that mother-to-infant microbial seeding accounts for ~58.5% of the infant gut microbiome, with composition being attributed to multiple maternal source communities such as the vagina, faeces, skin and breastmilk (Bogaert et al., 2023). Following colonisation, the human gut microbiome evolves through infancy before becoming increasingly individualised and stable. Preceding this more 'mature' state, the gut microbiome can be considered quite plastic, whereby this dynamic entity is influenced by varying factors. This early life plasticity thereby represents a period of critical importance, impacting on both short- and long-term health. In humans, gestation (i.e., the period of time between conception and birth) is expected to last for 40 weeks, with full-term infants generally being defined as those babies born >37 weeks of gestation. In healthy full-term infants, a well-established set of host and environmental factors have been found to influence the patterns of gut microbiome development, explored further below.

1.4. Factors that impact the term infant gut microbiome

1.4.1. Birth mode

Birth mode is considered one of the first direct determinants of gut microbial composition, providing the initial microbial inoculum for the GI tract both during and immediately after birth. As such, in term infants, vaginally delivered infants are consistently found to have a higher abundance of *Bacteroides* spp., likely reflecting maternal faecal to infant transmission (Mitchell et al., 2020). Conversely, infants born via caesarean section have been found to have disturbed transmission of maternal *Bacteroides* spp., and a higher level of colonisation by hospital-associated opportunistic pathogens such as *Enterococcus*, *Enterobacter* and *Klebsiella* spp. (Shao et al., 2019). Furthermore, some studies have also found vaginally delivered infants to be enriched with *Bifidobacterium* spp. when compared to those born by caesarean (Reyman et al., 2019). These differences are most pronounced during the earlier stages of life, around 1 - 2 weeks after birth (Mitchell et al., 2020; Reyman et al., 2019), following which the delay in colonisation of caesarean born infants somewhat diminishes. After around 1 year of life, the differences in gut microbiota

composition based on birth mode are generally no longer observed with microbial convergence occurring with age (Stewart et al., 2018; Xiao & Zhao, 2023). In addition, during these earlier stages of life, studies have shown that the gut microbiome of infants who are born by caesarean section can at least be partially restored by exposing these infants to maternal vaginal fluids (Dominguez-Bello et al., 2016) , faecal microbiota transplants from the mother (Korpela et al., 2020), and by breastfeeding (Guo et al., 2020).

1.4.2. Breast milk, breast milk components and formula milk

Following birth and during the subsequent postnatal period, one of the biggest factors that governs gut microbiome structure in term infants is diet. This association is well-described, with receipt of breast milk being one of the major reproducible determinants (Stewart et al., 2018). Mother's own milk (MOM) shapes the gut microbiome in various ways including providing a direct source of colonisers from the milk microbiome, as well as from the skin microbiome when infants are fed from the breast directly (Moossavi et al., 2019). In addition, MOM contains several bioactive components that are hypothesised to shape this gut microbial community. One of the most studied bioactive components are human milk oligosaccharides (HMOs). HMOs are the third most abundant component in breast milk, following lactose and lipids, yet cannot be digested by the infant and reach the colon intact (Masi & Stewart, 2022). They act as prebiotic substrates, promoting the growth of specific bacteria that are able to utilise these sugars, namely *Bifidobacterium* spp. As such, studies have shown that infants who are breastfed have a higher abundance of gut *Bifidobacterium* in comparison to those who are formula fed (Ma et al., 2020; Stewart et al., 2018). Conversely, infants who are formula fed have been found to be colonised by a higher abundance of opportunistic pathogens such as *S. aureus, Staphylococcus epidermidis, Klebsiella pneumoniae, Klebsiella oxytoca* and *Clostridium difficile* (Pärnänen et al., 2022).

In addition to the differences observed between infants receiving breast milk or formula milk, previous studies have also found a difference in the milk microbiome specifically of pumped breast milk rather than direct feeding (Moossavi et al., 2019). This likely reflects how flowback of milk from the infant mouth during direct breastfeeding further shapes the milk microbiome with microorganisms from the infant oral cavity. A similar phenomenon also likely occurs with breast pumping, allowing pump-associated

bacteria to enter the milk duct as opposed to from the infant oral cavity (Moossavi et al., 2019). However, little research has been done on the subsequent impact on the infant gut microbiome based on direct vs indirect breastfeeding.

Whilst breast milk-derived HMOs are important in shaping the gut microbial community, they are not the only means by which breast milk is able to shape the gut microbiome. For example, lactoferrin is an iron-binding glycoprotein found in high levels in colostrum (i.e., the first milk) which has also been proposed to modulate the infant gut microbiome by promoting growth of 'beneficial' gut bacteria such as Bifidobacteria, and inhibiting the growth of microbes such as *Staphylococcus aureus* by iron sequestration (Young et al., 2022a). Despite this, recent studies have shown minimal impact of lactoferrin intervention on the preterm infant gut microbiome (Young et al., 2022a). There is also ongoing research into other components such as glycosaminoglycans (GAGs), which are also complex glycans, like HMOs, that pass through the GI tract undigested before being catabolised by bacterial enzymes (Maccari et al., 2016). Although relatively understudied, it has previously been shown that GAGs can inhibit the adhesion of the enteric pathogens *Escherichia coli* and *Salmonella fyris* to human intestinal cells (Coppa et al., 2016). Further work will be needed to see if GAGs can alter the infant gut microbiome.

1.4.3. Antibiotics

Antibiotics impact the gut microbiome of both infants and adults alike. In adults, these effects are generally short-term, with bacterial diversity immediately decreasing coupled with a loss of various species dependent on the type of antibiotic. The gut microbial population of adults generally returns to previous levels prior to antibiotic use, particularly following shorter and less frequent courses (Palleja et al., 2018).

In infants, the effects of antibiotic use have been studied widely since this population can be seen as more vulnerable, with a gut microbial population that has not yet reached a mature state. Short term effects depend on the type of antibiotic and the length of the course, but in general will cause changes to the relative abundance of individual taxa, in particular a decrease in Bifidobacteria (Korpela et al., 2020), and increase in Enterobacteriaceae such as *Klebsiella* and *Enterococcus* spp. (Korpela et al., 2020; Reyman et al., 2022). In this neonatal population, more prolonged effects of

antibiotics can be observed when compared to children and adults, with some studies still seeing the effects up to 12 months after a 48-hour antibiotic course (Reyman et al., 2022). It is hypothesised that this could be due to a “normal” state having not yet been established for the gut microbial population to return to (Reyman et al., 2022). In the neonatal intensive care unit (NICU) antibiotic use is often clinically necessary, but frequent use can delay maturation of the microbiome, in both term and preterm infants alike, although studies have found this to eventually recover (Gasparrini et al., 2019). Moreover, when assessing the impacts of antibiotic use using sequencing methods, DNA from dead microorganisms would still be detected and could potentially skew results. In addition to direct antibiotic use in infants, the impact of maternal antibiotic use on the infant gut microbiome has also been investigated. Studies have shown that intrapartum antibiotic use (i.e., antibiotics given during labour) alters the gut microbial community composition for up to 12 months, with an increase in *Clostridium* and *Enterococcus* and decrease in *Bacteroides* and *Parabacteroides* observed (Azad et al., 2016). Further to this, other studies have shown intrapartum antibiotic use to increase the abundance of antibiotics resistance genes (ARGs) and mobile genetic elements in the infant gut community, suggesting transfer of these genes via bacteria from mother to infant (Pärnänen et al., 2018).

1.4.4. Maternal and other factors

In addition to the factors discussed above, there are additional factors that have been shown to influence the gut community in term infants albeit to varying degrees and with less reproducibility between studies. Geographical location is often found to be a major driver of gut microbiome composition in both infants and adults (Suzuki & Worobey, 2014), although this measure is sometimes potentially confounded with many others such as diet, socioeconomic status, maternal characteristics etc. In addition, studies have found factors such as the presence of siblings and furry pets in the household to influence the infant gut microbial community (Stewart et al., 2018). Aside from infant characteristics and direct infant exposures, maternal characteristics and exposures pre-birth can also impact the subsequent infant gut microbiome. For example, a meta-analysis of 76 studies found maternal intrapartum antibiotic use (discussed previously), maternal obesity and excessive gestational weight gain to all be associated with reduced microbial diversity (Grech et al., 2021).

1.5. Preterm birth

Preterm infants are defined as those babies born <37 weeks of gestation. These infants can be sub-categorised based on their gestational age at birth as follows: moderate to late preterm (32 – 37 weeks), very preterm (28 – 32 weeks) and extremely preterm (<28 weeks). Preterm birth can occur as a result of medical complications that require early induction of labour or caesarean birth, or more commonly they occur spontaneously. Maternal demographic factors (i.e., higher maternal age, higher body mass index, smoking, stress, lower socioeconomic status etc.) as well as other factors including multiple gestations and infections are all associated with preterm delivery. The prevalence of preterm birth varies by region and geographical location, with approximately 10% of births worldwide being preterm (Cao et al., 2022).

1.5.1. The neonatal intensive care unit and practices

Very preterm infants born <32 weeks of gestation will have a vastly different start to life in comparison to their healthy term counterparts. Very preterm infants will initially be cared for on the NICU, a unique setting which plays a crucial role in the acquisition and development of the gut microbiome. Here, infants are limited in their exposure to environmental microbes and will instead be colonised by NICU-acquired microbes, which are typically considered pathobionts (i.e., species which cause harm under certain environmental conditions) such as *Klebsiella* spp. and *E. coli*. In addition, these infants undergo clinical care regimes that will impact microbiome development, including specific feeding practices, antibiotic treatment (often multiple courses), ventilation and intravenous line placement (Ahearn-Ford et al., 2022), not experienced by a healthy term infant.

Much of the clinical care provided to very preterm infants has associations with shaping the gut microbiome. Due to an inability to latch and suckle at the breast and a high energy requirement to support growth, very preterm infants will require support when feeding, with the National Institute for Health and Care Excellence (NICE) guidelines stating parenteral nutrition (PN) should be used for infants born <31 weeks of gestation. PN is the provision of nutrients intravenously (i.e., directly into the bloodstream) thereby bypassing the GI tract which is immature in this population and therefore contributes to their inability to tolerate milk feeds straight away (Binchy et al., 2018). PN is used to encourage growth until the infant is able to tolerate milk feeds,

which are then introduced gradually until reaching what is known as 'full feeds', at which point the infant will receive all their nutrition as milk feeds. These milk feeds are initially introduced by enteral feeding whereby a feeding tube allows the direct intake of milk into the stomach via the nasogastric route (i.e., nose to stomach) or orogastric route (i.e., mouth to stomach), until the infant has developed the suck and swallow mechanism and can feed safely by bottle or breast. These feeding practices greatly differ to healthy full-term infants, for whom the gold standard is direct feeding from the breast. Moreover, PN has been shown to have significant adverse effects on the gut microbiome of preterm infants, associated with a reduced *Bifidobacterium* abundance (Jia et al., 2020).

In terms of antibiotics, these are often used immediately after birth in preterm, very low birthweight (VLBW) and extremely low birthweight (ELBW) infants for at least the first 48 hours of life. Empiric antibiotics are used in order to prevent early onset sepsis (EOS), which preterm infants are more prone to due to their weakened immune systems and underdeveloped skin and gut barriers. Furthermore, during their stay on the NICU, many of these infants will encounter various complications and as a result are often exposed to multiple antibiotic treatments, discussed further in Section 1.6.5.3. In addition, antifungal prophylaxis is recommended by the European Society of Clinical Microbiology and Infectious Diseases for a number of preterm infants in the first 48 – 72 hours after birth (i.e., infants weighing <1000 g, in NICUs where invasive candidiasis infections are frequent etc.), which has been shown to reduce colonisation of *Candida* spp. (Hanna & Mazkereth, 2021). Neonatal intensive care practices such as feeding practices and antibiotic use have been reproducibly shown to impact the developing preterm gut microbiome.

1.5.2. Early life gut microbiome development in preterm infants

All things considered, it is of no surprise that preterm infants often experience a gut microbiome with lower microbial diversity and increased colonisation of pathobionts, compared to healthy term babies (Chernikova et al., 2018). As such, gestational age represents an additional covariate associated with the gut microbiome, with term and preterm infants found to have very distinct gut community compositions (Chernikova et al., 2018; Grier et al., 2017; Hill et al., 2017). For example, preterm infants have

been reported to experience delayed colonisation of *Bifidobacterium* (Korpela et al., 2018) and *Bacteroides* (Stewart et al., 2017) in comparison to term infants.

Aside from gestational age, previous studies specifically investigating preterm gut microbiome development have identified associations that overlap with term infants, including receipt of MOM and antibiotic exposure (Aguilar-Lopez et al., 2021; Cong et al., 2016; Grier et al., 2017). Alike term infants, the opposing effects of MOM and antibiotic exposure on the preterm gut microbiome are reasonably well-recognised.

The impact of other factors on the preterm gut microbiome, such as birth mode and sex are, however, more ambiguous. Specifically, numerous studies have found birth mode not to be significantly associated with the preterm gut microbiome (Hill et al., 2017; Stewart et al., 2016), or to have minimal influence (La Rosa et al., 2014), whilst in term infants, birth mode is thought to have a major impact for at least the first year of life (Bokulich et al., 2016; Reyman et al., 2019). Furthermore, one preterm neonate study has suggested sex-specific differences in preterm infants, including a lower diversity index, higher abundance of Enterobacteriales and lower abundance of Clostridiales in males compared to females (Cong et al., 2016). However, findings like these are not universally observed, and are seemingly specific to certain cohorts (i.e., specific NICUs). These inconsistencies could potentially reflect the use of smaller underpowered cohorts and in some cases a lack of longitudinal sampling. It also underscores the need for a more focused investigation into the factors influencing normal gut microbiome development in healthy preterm infants.

1.5.3. The preterm immune system

Preterm infants are born before their immune system has had chance to develop. During the later stages of pregnancy, mostly in the third trimester of gestation, there is a high level of maternal antibody transfer *in utero* in the form of immunoglobulin G (IgG). Despite the reduction in placental transfer of maternal antibodies, there is evidence of selective transfer of functional antibodies earlier in pregnancy, allowing at least some level of humoral immunity for infants born prematurely (Dolatshahi et al., 2022). Following birth, infants receive maternal antibody protection in the form of IgA from human breast milk, for which there is research surrounding its potential role in protection from NEC (described later in Section 1.6.5.1.).

Neonates rely heavily on the innate immune system, as the adaptive immune system does not fully develop until early childhood. In the preterm population, the innate immune system has a reduced capacity compared to infants born full term and is thought to be less developed, although little is known specifically about lymphocyte populations (Quinello et al., 2014). The immaturity of the innate immune system in preterm infants reduces the ability of these infants to overcome potential pathogenic infections and initiate a sufficient and controlled immune response, or in the case of NEC, an exaggerated un-controlled response (Hackam & Sodhi, 2018). For example, preterm infants have been shown to have attenuated toll-like receptor (TLR) function, which are pattern recognition receptors responsible for mediating an immune response to pathogen-derived ligands. Previous research has shown TLR2 and TLR4 expression to be reduced in preterm infants during the initial weeks following birth, although rapidly increasing during the first months of life (Shen et al., 2013). By contrast, cytokine response from preterm cord blood (i.e., blood remaining in the placenta and umbilical cord after birth) has been previously shown to be imbalanced when compared to term cord blood (Glaser et al., 2023), and that lipopolysaccharide (LPS)-induced cytokine response does not appear to recover like TLR expression (Shen et al., 2013). Further to this, preterm infants have been found to have a reduced pool of monocytes and neutrophils, as well as reduced cytokine production from these cells which in turn limits T-lymphocyte activation (Melville & Moss, 2013). Importantly, studies have shown that the preterm immune system differs to that of term infants early in life, but converges rapidly thereafter, both sharing a common trajectory of immune development (Olin et al., 2018). This immune system adaptation appears to be driven by microbial interactions and is impaired where there is extremely low gut microbiome diversity (Olin et al., 2018).

1.5.4. T-lymphocytes in preterm infants

T-lymphocytes represent an important component of the immune system, responsible, in part, for detecting and dealing with potential pathogens. T lymphocytes originate from hematopoietic stem cells in the bone marrow, after which they migrate to the thymus for maturation, selection, and subsequent migration to the periphery (i.e., outside the brain and spinal cord). From these immature cells, various T lymphocyte lineages arise, which can be defined by differential expression of surface antigen

markers and cytokine profiles. The protocol for identifying cell surface molecules is called the Cluster of Differentiation (CD), whereby surface markers are given CD numbers which are used to identify different immune cell lineages, such as T-lymphocyte sub-populations. The two major T-lymphocyte lineages are CD4+ cells (T-helper cells) and CD8+ cells (cytotoxic T cells). These lineages, and others (i.e., γδ T lymphocytes) can be further divided into smaller sub-populations, each expressing different markers and each with different specific roles (**Figure 1.2**).

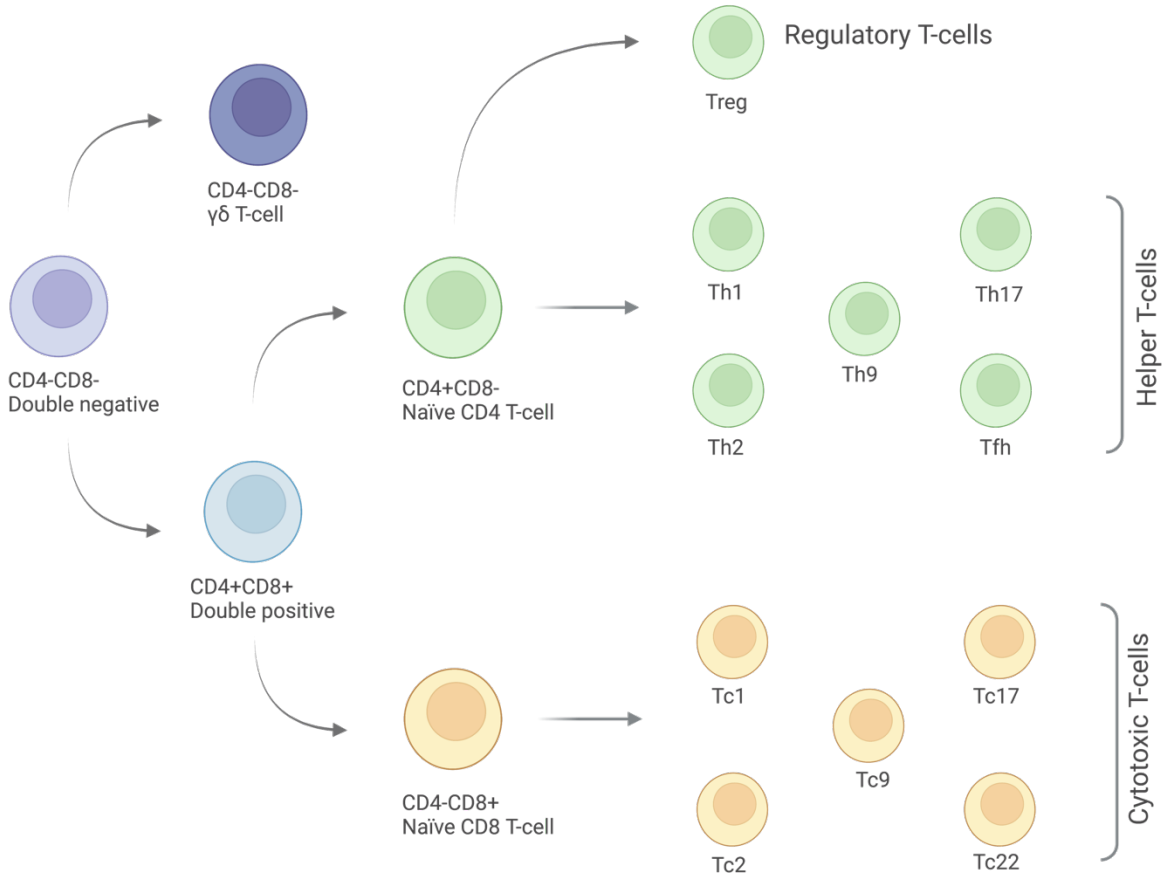

Figure 1.2. Schematic of T-lymphocyte differentiation into major sub-types. Double negative T-lymphocytes (CD4-CD8-) can differentiate into CD4-CD8- or CD4+CD8+. Double positive T-lymphocytes thereafter can differentiate into naïve CD4+CD8- T-cells or naïve CD4-CD8+ T-cells. The expression of different markers gives rise to various regulatory, helper and cytotoxic T-cell subtypes, each with specified roles.

Foetal T-lymphocytes can be detected from around 8 weeks of gestation, but perhaps unsurprisingly, the majority of T-lymphocytes detected in foetal cord blood are naïve (i.e., immature cells that have yet to encounter an antigen and become activated)

(Quinello et al., 2014). Despite these accounting for the majority, there is evidence to support the existence of non-naïve T-lymphocytes in the foetus, including CD4+ T-lymphocytes with an effector memory phenotype, which have been identified in cord blood (Zhang et al., 2014). Furthermore, a memory phenotype has been found to dominate CD4+ T cells in the intestine (Li et al., 2019). Together these data are consistent with the idea of prior antigen exposure *in utero,* despite the womb being considered sterile in normal pregnancies. Ultimately, the nature of this priming is currently unknown (Sproat et al., 2020), and there is much more to explore, although these could be the result of harmless antigen exposure from metabolites transferred across the placenta. Following birth and at around 8 weeks of life, preterm infants have been found to have lower absolute counts of T-lymphocytes, a lower CD4/CD8 ratio and a lower proportion of naïve T-lymphocytes when compared to term infants of the postnatal same age (Berrington et al., 2005), the latter of which could potentially reflect their higher antigen exposure.

On the whole, there is limited data surrounding T-lymphocyte characterisation in preterm infants, as it presents a significant challenge to study. The acquisition of an adequate volume of blood for research purposes is often both ethically and logistically challenging, resulting in a heavy dependence on salvaged samples obtained through routine blood tests. Consequently, numerous studies have resorted to the use of foetal cord blood as a model, despite its inadequacy in representing postnatal immunity as accurately (Olin et al., 2018). Together, this underscores the need for future research to be primarily based on the use of peripheral blood samples, where ethical considerations permit.

1.6. Diseases affecting preterm infants

Infants born prematurely and VLBW are at increased risk of various diseases, many but not all are GI-related, owing to their naïve immune systems and immature intestinal barriers.

1.6.1. Late onset sepsis

Neonatal sepsis is a significant cause of mortality amongst very preterm infants, as well as being associated with increased risk of in-hospital morbidities and poor neurodevelopmental outcomes amongst survivors. Defined as an infection involving

the bloodstream, neonatal sepsis can be subdivided into EOS and late onset sepsis (LOS), based on time after birth (i.e., EOS <72 hours after birth, LOS >72 hours after birth).

Although not strictly a GI disease, in preterm infants, LOS can be caused by translocation of bacteria from the gut into the bloodstream (Pilarczyk-Zurek et al., 2022). Preterm infants are prone to “leaky gut”, whereby the gut epithelial barrier is not fully developed yet, resulting in a great capacity for enteric microorganisms to pass through. In such cases, LOS can be caused by microbes that would typically not be classed as true pathogens and in a more mature gut would likely not cause any issues, evidenced by reported cases of probiotic sepsis in preterm infants (Dani et al., 2015; Jenke et al., 2012). Major risk factors of LOS typically include low birthweight and gestational age, as well as previous antimicrobial exposure (Downey et al., 2010). In terms of prevention, there is limited evidence to suggest lactoferrin decreases LOS in preterm infants (Pammi & Suresh, 2017), as well as early feeds which importantly needs to be balanced against the risks of developing NEC (Downey et al., 2010) (discussed further in Section 1.6.3.).

1.6.2. Spontaneous intestinal perforation

Spontaneous intestinal perforation (SIP), also known as focal intestinal perforation (FIP), is a life-threatening condition that again predominantly affects VLBW, ELBW and preterm infants. It is defined as an isolated perforation (i.e., a hole forming in the organ wall) most commonly occurring in the terminal ileum. SIP has a mortality rate of ~14% (Farrugia et al., 2003), although this varies by study, and a worse prognosis is associated with those who have concomitant sepsis and those who have undergone laparotomy (Fisher et al., 2014). Similar to LOS, survivors often have poorer outcomes such as growth failure and neurodevelopmental delay. Furthermore, complications of SIP include but are not limited to surgical complications, prolonged hospital stays and increased risk of short bowel syndrome (Krishnan & Lotfollahzadeh, 2023). SIP is managed either medically or surgically and is the most common surgical disease in the differential diagnosis of NEC, described further in Section 1.6.3.

The pathogenesis of SIP is unclear and previous studies are limited with many tending to group NEC and SIP together when looking for hallmarks of disease due to both diseases sharing similar clinical presentation.

1.6.3. Necrotising enterocolitis

NEC is a devastating GI disease affecting up to 10% of infants born prematurely. It is responsible for severe mortality in this population, as well as considerable long-term morbidity, such as growth failure and neurodevelopmental delay (Jones & Hall, 2020). Low birthweight and prematurity are well recognised risk factors; however, NEC can also affect infants born full-term, with these cases often co-existing with other co-morbidities (Short et al., 2014).

NEC is characterised by severe inflammation, radiological finding of pneumatosis intestinalis and necrosis of the bowel. Unlike SIP, which is an isolated focal perforation, NEC is an inflammatory disease that usually has systemic with multi-organ affection (Elgendy et al., 2021), with perforation occurring as a complication of disease (Farrugia et al., 2003). Unfortunately, it remains a difficult disease to diagnose, with early clinical presentation often being variable, and typical features (e.g., pneumatosis intestinalis) generally found during the later stages of disease (Beck et al., 2021). Furthermore, there are additional challenges in differentiating NEC from other morbidities that primarily affect this population, such as LOS and SIP, as clinical features are shared amongst these morbidities (Berrington & Embleton, 2022). There is also increasing belief that NEC can be considered more of an umbrella term for what appears to be several different disease subsets (Neu et al., 2018), that manifest in a common endpoint.

The aetiology of NEC is considered ‘multifactorial’, and the pathophysiology remains poorly understood. Considering the likelihood of NEC to encompass different disease endotypes, each with different pathophysiology, this is unsurprising (Neu, 2020; Neu et al., 2018). One of the classically proposed pathways to disease onset is the imbalance in pro-inflammatory signalling pathways. The balance between anti- and pro-inflammatory signalling is mediated by TLR4, which recognises LPS on the cell surface of Gram-negative bacteria (Hackam & Sodhi, 2018) (**Figure 1.3**). There is evidence to suggest that expression of TLR4 is increased and that mutations in the

TLR4 signalling pathway have been observed in preterm infants with NEC (Alganabi et al., 2019). Further evidence to support the role of this signalling pathway in NEC pathophysiology include knockout studies, whereby Δ*TLR4* mice which lack the ability to express TLR4 in epithelial cells do not go on to develop NEC (Liu et al., 2021). In a recent study, the role of *in utero* factors on TLR4 signalling and NEC development has been uncovered. Here, researchers showed that a diet rich in indole-3-carbinole activates the aryl hydrocarbon receptor, reducing TLR4 signalling and preventing NEC in new-born mice (Lu et al., 2021). Indeed, the role of the gut microbiome in NEC pathophysiology has been demonstrated, with GF animals not developing the disease and those receiving faecal microbiota transplants from NEC patients developing intestinal injury (He et al., 2021).

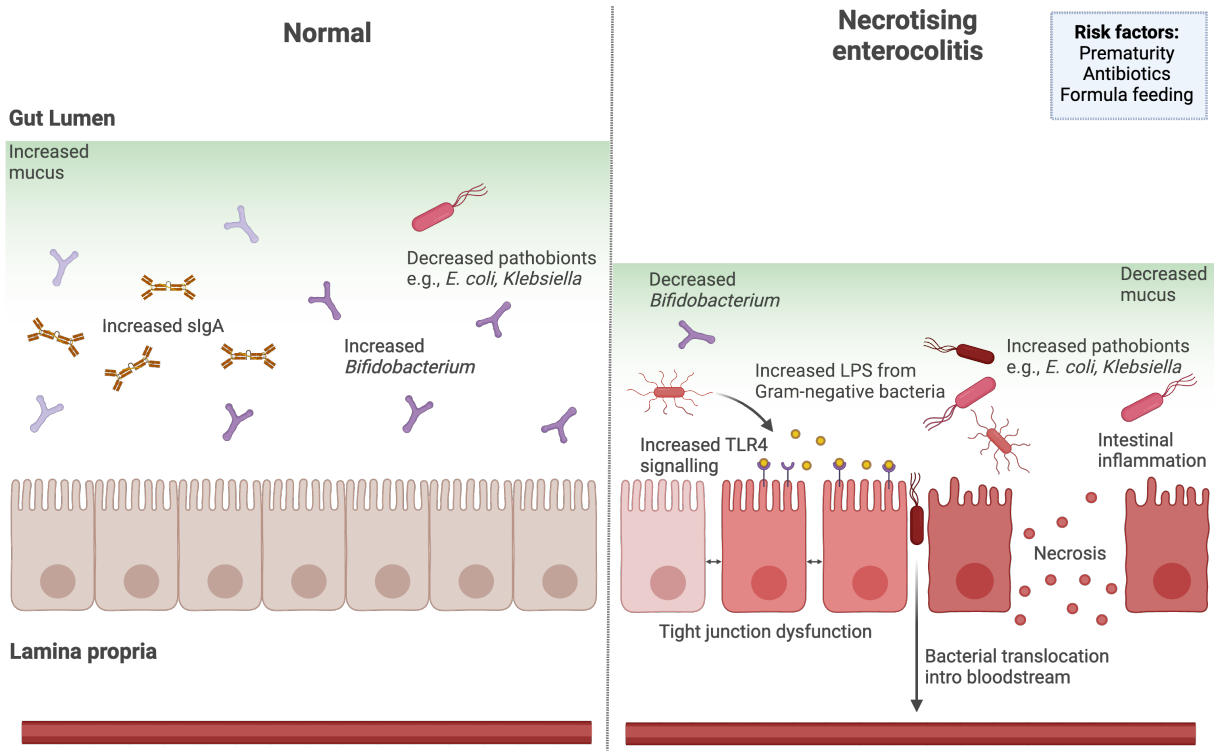

Figure 1.3. Proposed pathophysiology of NEC. A reduction in *Bifidobacterium*, increase in pathobionts and LPS from Gram-negative bacteria increases TLR4 signalling, which is uncontrolled, resulting in intestinal inflammation, intestinal barrier injury and therefore translocation of microbes. The translocation of microbes can result in intestinal ischemia and eventually necrosis.

1.6.4. NEC risk factors

There are various factors thought to be associated with the onset of NEC, including prematurity, antibiotics, feeding and the gut microbiome (Esmaeilizand et al., 2018; Stewart et al., 2016; Sullivan et al., 2010). Prematurity is the only consistently identified NEC risk factor in case-control studies, where disease risk is inversely correlated with gestational age. This can be explained by the immaturity of the intestinal tract at younger gestational ages (Henry & Moss, 2008). A number of observational studies have found that preterm infants who receive an exclusively human breast milk diet have a lower NEC incidence than those who are formula fed, although the mechanism of protection here is not fully understood (Herrmann & Carroll, 2014). Furthermore, previous studies have shown that infants receiving pasteurised donor human milk (DHM) have lower growth rates but reduced risk of developing NEC and feed intolerances when compared to formula-fed infants (Stoltz et al., 2021). In addition, one study showed that lipase digestion of formula milk, but not breast milk, caused death of neutrophils, endothelial and epithelial cells *in vitro,* which could have broad implications in NEC pathophysiology (Penn et al., 2012). However, some infants who receive formula do not go on to develop NEC and some who only ever receive breast milk can unfortunately go on to develop NEC, suggesting the precise composition of breast milk to also be of importance. Aside from milk source, there have also been a number of studies investigating timing of enteral milk feeding on NEC incidence, including whether early introduction of enteral feeding increases the risk of NEC and vice versa. However, according to a 2022 Cochrane meta-analysis of 14 trials, it seems that delaying progressive enteral feeds may not reduce the risk of NEC or death in very preterm or VLBW infants (Young et al., 2022b). Furthermore, whilst delayed introduction may slightly reduce feed intolerance, it probably increases the risk of LOS (Young et al., 2022b) (described in Section 1.6.2.).

The occurrence of NEC has also been associated with antibiotic use in neonates. Previous studies have shown that prolonged empiric antibiotic treatment was associated with increased risk of NEC development (Esmaeilizand et al., 2018). However, this association is complex, and the use of antibiotics is clinically necessary when EOS is suspected, amongst other things. One study found that short empiric antibiotic use of <72 hours actually reduced the risk of NEC when compared to those

who underwent prolonged antibiotic treatment and those who received no antibiotic treatment (Dierikx et al., 2022).

In terms of gut microbiome structure, the most consistent findings associated with disease progression point to lower bacterial diversity, higher Proteobacteria (Taft et al., 2015), and lower *Bifidobacterium* (Olm et al., 2019; Pammi et al., 2017; Stewart et al., 2019). The largest metagenomic study to date analysing samples prior to NEC diagnosis found a significant increase in *Klebsiella* spp., bacteria encoding fimbriae, and bacteria encoding secondary metabolite gene clusters related to quorum sensing and bacteriocin production (Olm et al., 2019). Further to this study, there has been evidence to suggest a higher bacterial diversity and *Bifidobacterium* abundance is protective against NEC development (Stewart et al., 2016). Nevertheless, whilst broad associations have been noted and individual studies have identified taxa of interest, no single causative agent has been consistently associated with disease and it also remains difficult to distinguish between cause and effect.

1.6.5. NEC prophylaxis and prevention

1.6.5.1. Breast milk feeding and human milk oligosaccharides

Due to their association with NEC, the role of the gut microbiome and breast milk-derived HMOs on NEC onset are now active areas of research. Trying to unpick the mechanisms that underly the apparent protection of breast milk, a recent study found that the concentration of a single HMO, disialyllacto-N-tetraose (DSLNT), was significantly lower in MOM received by infants with NEC compared with controls (Masi et al., 2020), and that DSLNT was protective against NEC in rats (Jantscher-Krenn et al., 2012). Furthermore, studies have shown other components of milk, such as IgA, to be protective against the development of NEC (Gopalakrishna et al., 2019). Together these studies highlight the potential for specific breast milk components to be predictive biomarkers of NEC, and guide future clinical decisions surrounding breast milk such as selecting milk with the highest IgA and DSLNT content for infants most at risk of NEC.

1.6.5.2. Probiotics

The association between the gut microbiome and NEC (i.e., lower *Bifidobacterium* and higher Proteobacteria) has led, in part, to an increased use of probiotics across NICUs.

Although, probiotic-using units are still in the minority. Probiotics are live microorganisms, such as bacteria and yeasts, thought to promote health benefits and are generally considered as safe, therefore, have received a lot of interest as an intervention to manipulate the gut microbiome.

In the context of NEC, a number of studies (both observational and randomised controlled trials; RCTs) have been conducted to look at the impact of probiotics on NEC incidence and on the gut microbiome itself. Probiotic treatments are generally focused on the use of various *Lactobacillus* spp. and *Bifidobacterium* spp. *Bifidobacterium* in particular are considered particularly important in the infant gut due to their association with health, and their known ability to utilise HMOs and produce short chain fatty acids (SCFAs) and indole derivatives, which have proven anti-inflammatory effects on intestinal epithelial cells (Ehrlich et al., 2020). For example, *Bifidobacterium* have been shown to do reduce production of pro-inflammatory cytokines, enhance tight junction barrier function between intestinal epithelial cells and reduce permeability across this membrane (Al-Sadi et al., 2021; Bergmann et al., 2013). It is through these means, and other mechanisms, that probiotic strains may provide their protection against NEC.

Despite this, previous studies have yielded inconsistent results, for example a retrospective review of probiotic use in the Newcastle Royal Victoria Infirmary (RVI) NICU by Granger *et al.* (2022) found no change in the risk of NEC, LOS, mortality and no change in the proportion of surgical NEC cases with probiotic use (Granger et al., 2022). However, the study did find a NEC risk reduction in sub-group analysis of infants born >28 weeks of gestation (Granger et al., 2022). Conversely, some retrospective observational studies have noted a reduction in NEC following probiotic use (Robertson et al., 2020). Similarly, some RCTs note a reduction in NEC amongst infants receiving probiotics (Jacobs et al., 2013), whilst others do not (Costeloe et al., 2016). A Cochrane meta-analysis of 56 trials conducted in 2020 concluded that probiotic use in very preterm and VLBW infants may reduce the risk of NEC, serious infection and death – although evidence was of “low certainty” (Sharif et al., 2020). Further to this, a more recent meta-analysis of 70 studies looking at both term and preterm infants, concluded that probiotics significantly reduced NEC, overall mortality, and NEC-related mortality (Wang et al., 2023).

Despite the increased interest in and use of probiotics in the NICU, the efficacy of probiotics in preventing NEC and LOS remains inconclusive. Furthermore, there are very low but important risks reported in the literature from contamination and probiotic bacteraemia/sepsis (Acuna-Gonzalez et al., 2023; Dani et al., 2015; Esaiassen et al., 2016; Jenke et al., 2012) which need to be considered. It is likely that these probiotic species entered the bloodstream via translocation from the gut due to the immaturity of the preterm intestinal barrier, as previously discussed. These concerns have led, in part, to studies evaluating alternative probiotic administrations routes i.e., via the lactating mother, with varying results (Dotterud et al., 2015; Rahkola et al., 2023), suggesting direct administration in fact to have the strongest effect.

1.6.5.3. Antibiotics

Despite prolonged use of antibiotics being associated as a risk factor for NEC, antibiotic use in preterm infants is still a clinically necessary practice. Bacterial infections are a frequent complication of preterm birth and as such, empirical antibiotics are administered to the majority of VLBW infants in the first few days of life (Flannery et al., 2018), as previously discussed. In terms of a preventative treatment for NEC, the hypothesis is that intravenous use shortly after birth would slow down initial bacterial colonisation, potentially reducing the risk of TLR4-mediated NEC by not overwhelming the immune system (Shen et al., 2022). However, as a prophylaxis, there are conflicting data surrounding the effectiveness of antibiotic use, with a number of studies finding no change in NEC prevalence. Nonetheless, as discussed in section 1.6.4., a recent observational study found that short term empiric antibiotic use of <72 hours reduced the risk of NEC (Dierikx et al., 2022). Indeed, there are studies currently ongoing assessing early antibiotic use as a preventative treatment for NEC which aim to optimise early antibiotic use in preterm infants (Shen et al., 2022). Ultimately, finding the balance (i.e., timeframe, antibiotic type etc.) remains a difficult challenge.

1.7. Methodology for microbiome and NEC research

1.7.1. ‘Omic’ technologies

The term ‘omics’ encompasses many technologies, including genomics, metagenomics, transcriptomics, proteomics, and metabolomics. Next generation ‘omic’ technologies have allowed for a much more detailed understanding of host

genetics and functional genetics in relation to disease (Karczewski & Snyder, 2018). In addition, meta-omics studies (i.e., metagenomics, metatranscriptomics, metaproteomics and metabolomics) have increasing use in studying the gut microbiome community, with their ability to provide an insight into strain level taxonomic resolution, functional and metabolic capabilities of the microbiome (Knight et al., 2018).

These technologies have demonstrated powerful potential in clinical research and provided insight into 'omic' signatures of neonatal disease. For instance, improvement in molecular sequencing coupled with rapidly declining costs has enabled research into the differences in the microbiome between individuals, populations, and species. The ability to model multi-omic data and correlate this with clinical findings allow the possibility to identify underlying biomarkers and other modifiable factors that will allow for early detection of disease. Thus, increasingly, 'omic' technologies can be used to target therapies to treat these varied pathologies.

It is noteworthy that experimental design is critical to generating accurate and actionable results. Cohort studies are important to understand neonatal disease such as NEC, where various samples can be analysed using 'omic' pipelines, including stool, serum, urine and resected tissue (Embleton et al., 2017). Stool samples are important non-invasive samples for understanding what the intestine is exposed to; urine includes the host metabolites which are excreted from the body; blood gives a representative overview of the systemic response to disease and resected tissue can give information on the response mounted at the site of disease (D'Adamo et al., 2021; Haas et al., 2017). Notwithstanding, the complexity and individual variation in clinical and environmental exposures can pose significant challenges in data analysis that is ultimately underpinned by the quality and granularity of accompanying metadata.

1.7.2. 16S rRNA gene sequencing and metagenomics

Gene sequencing techniques for microbiome characterisation can be broadly split onto one of two categories: gene amplicon sequencing and shotgun metagenomic sequencing. The former relies on a phylogenetic marker gene, typically the 16S rRNA gene, to provide genus-level identification of bacterial members within the microbiome (Beck et al., 2021). 16S rRNA gene sequencing is relatively inexpensive, it can be used

for low abundance samples due to the amplification during library preparation and analysis can be relatively quick due to smaller number of sequences. Although sometimes referred to as a 'metagenomic' technology, this is not strictly true as 16S rRNA gene sequencing does not involve sequencing of all the genes present in a sample.

Metagenomic sequencing on the other hand, sometimes referred to as shotgun sequencing, involves the untargeted sequencing of host and microbial genes present in a sample, not just targeted sequencing of a single gene (Quince et al., 2017). Despite the increased costs and computational demand, this technique has its advantages over 16S, allowing both species- and strain-level classification (Beck et al., 2021). Furthermore, metagenomics is not limited to bacteria and allows the analysis of all microbial species (i.e., bacteria, viruses, fungi and archaea), and the functional potential of the microbiome, which is important due to the complexity of microbial communities and their genetic content (Beck et al., 2021). The variability of species within a genus and strains within a species is important to note, and these subtleties could impact on health and disease.

There are various sequencing platforms available for metagenomic studies, one of the more popular being Illumina sequencing, which relies on a sequencing by synthesis approach, producing paired read libraries (Bragg & Tyson, 2014). Typically, a metagenomic or 16S rRNA gene sequencing study would comprise the following workflow: (a) sample collection; (b) DNA extraction; (c) sequencing; (d) bioinformatic processing of sequencing reads; (e) taxonomic (and for metagenomics, functional) profiling and statistical analysis, and (f) validation.

It is noteworthy that characterisation using sequencing methods is not necessarily restricted to the genetic capacity of the microbiome. Whilst metagenomics offers some limited insight into the functional capacity (i.e., which genes are present), it fails to capture whether species are active members (i.e., which genes are switched on, which gene products are present and which metabolites are present).

1.7.3. Metabolomics

The complex, dense, community of microorganisms in the GI tract is also responsible for the production of a number of metabolic compounds (Vernocchi et al., 2016).

Metabolites are functional small molecules that represent the intermediate or end product of metabolism. They modulate a variety of signalling pathways which facilitate intestinal mucosa homeostasis (Vernocchi et al., 2016), and form the basis of communication between microbes and host cells. The host metabolome is therefore heavily intertwined with and connected to gut microbial communities. Metabolomics is the study of all the metabolites from a given ecosystem, providing insight into the function of the microbiome, host tissue and the host metabolic profile. Compared to metagenomics where functional annotation is based on capacity and prediction, metabolomics detects the actual end products of cellular reactions. Metabolite profiles can differ between healthy and disease subjects in a range of conditions and pathologies, some of which can be detected at unrelated body sites. For instance, the detection of metabolites in urine in relation to conditions that primarily impact the gut or brain. Such detection of metabolic alterations provides targets for non-invasive biomarkers and enables the mechanisms of disease development to be elucidated.

Metabolomic studies typically utilise Mass Spectrometry (MS) or nuclear magnetic resonance (NMR) based techniques. For MS-based techniques, metabolites are first separated using one of a variety of methods, including but not limited to gas chromatography (GC) and ultra-performance liquid chromatography (UPLC) (Vernocchi et al., 2016). The choice of chromatography will impact which metabolites are retained and subsequently detected by MS. The choice of column used is therefore important and should reflect the study hypotheses or multiple different columns could be used, such as combining reverse-phase liquid chromatography and hydrophilic interaction liquid chromatography (Contrepois et al., 2015). Recent efforts have also focused on combining different chromatograph techniques to maximize the number of metabolites detected. NMR spectroscopy, on the other hand, exploits the local magnetic field that exists around atomic nuclei, allowing the molecular structure of metabolites to be elucidated. MS-based techniques are more sensitive than NMR but require known standards to be run alongside samples in order to identify metabolites of interest (Emwas et al., 2019). Without doing so, any identified feature can only be putative and therefore most features in a liquid-chromatography (LC)-MS experiment will remain unidentified. Further to this, within the context of the gut microbiome, the origin of numerous metabolites remains elusive, as they could potentially arise from

either the microbiota or the host itself. This emphasises the need for further experimental work to better understand these intricacies.

1.7.4. High-dimensional immune profiling

Although perhaps not strictly an 'omic' technology, high-dimensional immune profiling using flow cytometry (FC) or mass cytometry by time of flight (CyTOF) are important techniques to consider when studying preterm infants at risk of NEC. These techniques look to measure protein abundance and can be used to study immune cells by, spanning a broad range (i.e., B-lymphocytes, T-lymphocytes, natural killer cells, monocytes etc.) or homing in on more specific sub-populations of an immune cell type such as the T-lymphocyte population (i.e., γδ T-lymphocytes, Th17 lymphocytes, mucosal-associated invariant T (MAIT) cells etc.). Different immune cells will express different protein markers in varying abundance and these signatures can then be used to identify immune cell populations and sub-populations.

Both techniques rely on the tagging/labelling of a panel of antibodies with either fluorescent dye (for FC) or metal isotopes (for CyTOF), which then give off a signal during the run. Antibody panels are designed to target antigens of interest, that are indicative of specific cell populations. Fluorescence intensities (for FC) or ion counts (for CyTOF) are assumed to be proportional to the expression levels of these antigens of interest (Nowicka et al., 2017).

Conventional analysis of both FC and CyTOF data includes gating, whereby different cell populations can be identified usually through use of scatter or contour plots. Cells

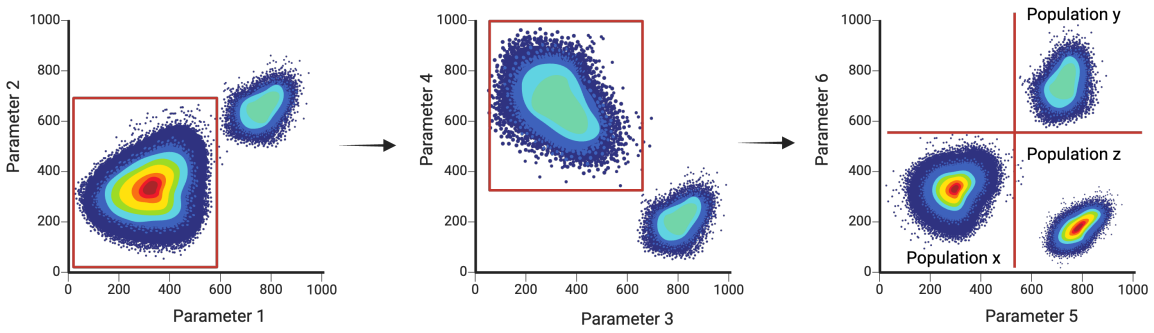

Figure 1.4. Schematic of manual gating of FC and CyTOF data. Cells are plotted in 2 dimensions and cell populations chosen based on different parameters.

can be plotted based on various parameters (i.e., antigen markers) on the X and Y axis and populations can be identified and selected for. The selected cell population can then be further gated based on different parameters, and so on (**Figure 1.4**). Gating is a manual process which is time consuming and can result in marker combinations often going unexamined (Van Gassen et al., 2015), particularly with increasing number of markers.

1.7.5. Multi-omics

Single omics approaches have provided important progress in the understanding of a range of complex diseases. Individually, however, these tools are often unable to capture the true biological complexity of most diseases (Karczewski & Snyder, 2018), including NEC. For instance, changes in specific bacterial species using microbiome sequencing may have little or very profound impacts on the overall function of the microbiome and microbe–host interaction. Advances in systems biology have enabled the integration of multiple types of omics data, termed multi-omics, which allows for a more comprehensive analysis and may provide important advances in the field. This is especially true in the case of diseases like NEC, which arises as a result of the interplay between a range of host and environmental factors. Thus, integrative approaches are advantageous, particularly when studying these kinds of diseases, as they provide a more holistic view. The advent of this more recent integrative approach paves an exciting possibility for the future of NEC research in preterm infants.

1.7.6. Challenges in multi-omic and microbiome data analysis

The advent and development of high-throughput techniques has driven the evolution of bioinformatic pipelines, computational tools and statistical methods used for microbiome characterisation. Without which, the handling and analysis of big data would prove impossible. Yet, omic data poses many challenges, as often big datasets do.

For downstream analysis (i.e., taxonomic and functional profiling) of metagenomic and 16S rRNA gene sequencing data, various statistical approaches and computational packages can be implemented. Microbiome data represents a unique challenge for analysis, in that it is compositional by definition. This is because reads are constrained to the upper bound of sequencing instruments, and that the abundance of any one

fragment is only interpretable relative to another within a given sample. Therefore, the total number of reads for a given stool sample does not reflect the absolute number of microbes present in the gut lumen (Weiss et al., 2017). This is implicitly acknowledged when microbiome data is rarefied, normalised or converted to relative abundance values (Gloor et al., 2017). As such, inferences can be difficult to make and certain analyses, such as correlation analyses between gut members, can be difficult to conduct. These specific challenges are not unique to microbiome data, but also metabolome data.

Some of the further challenges that come with multi-omic experiments involving clinical samples include the sheer complexity and presence of confounding variables. Instead of using standard univariate statistical approaches, such as T-tests, which will fail to account for the complexity of the data, multivariate analyses such as multiple regression models are required (Chen & Li, 2016). Put simply, these models allow multiple independent variables to be incorporated, each of which are controlled for when calculating the associated variance and significance of each individual variable.

There is increasing interest in analysing the temporal development of the infant gut microbiome (Stewart et al., 2017; Stewart et al., 2018; Yassour et al., 2016). Longitudinal datasets introduce repeated measures whereby samples are not independent from one another, which again cannot be analysed using standard statistical approaches. Instead, this data is often dealt with by splitting the dataset into multiple timepoints, and treating each timepoint like cross-sectional data, with 1 sample per patient per timepoint. However, this approach has its limitations, such as losing potentially important data pertaining to within-subject changes, over-simplification and reducing the statistical power within each cross-section. Alternatively, mixed effects models, sometimes referred to as mixed models or linear mixed effects models (LMMs) can be fit to data that has a repeated measures nature. Linear mixed effects models are an extension on simple linear regression models and an extension on multiple regression models, allowing for both fixed and random effects to be fit. Fixed effects are factors of primary interest, and generally include things such as treatments and demographic variables. As with multiple regression models, potential confounders can be controlled for by incorporating them as fixed effects. Random effects tend to refer to factors that are not of primary interest, and almost always include subject identifiers

in order to estimate population variance (Bolker et al., 2009). Therefore, for longitudinal data where there are multiple samples from each patient, the subject identifier (i.e., the variable that will be responsible for much of the variation, but which we are not interested in), can be included as a random effect in the model. Dependent on the error distribution of data, mixed effects models can be described as linear (i.e., when the error distribution is normal/gaussian) or generalised linear (i.e., when the error distribution is non-normal). There are various types of generalised linear mixed models (GLMMs), for example binomial mixed regression models, for which the dependent variable is binary. LMMs and GLMMs can be used to model various kinds of complex microbiome data structures, but in practice are primarily used to model alpha diversity metrics (i.e., diversity related to within a single sample), categorical outcomes such as microbial community clusters (Rozé et al., 2020) and specific taxon abundances.

Manual fitting of models works well for hypothesis-driven statistical testing of microbiome data, and for testing singular feature-wise associations. However, due to the highly sparse nature of microbiome data (i.e., zero-inflation) and high-dimensionality, computational approaches have been optimised to accommodate these challenges. MaAsLin 2 (Microbiome Multivariable Associations with Linear Models) is one such package, which employs a multi-model framework to analyse community data (Mallick et al., 2021). This tool enables the user to identify features, such as specific microbial taxa associated with each covariate in the model from an entire feature table.

Other statistical analyses for microbiome data include beta diversity analysis, which refers to diversity between samples or groups of samples. Specifically, beta diversity measures can be used to determine whether composition vectors of a sample (i.e., the relative abundance of each feature within a sample) differ between samples. One of the most commonly used beta diversity measures is Bray-Curtis dissimilarity, which looks to identify the overall dissimilarity between two composition vectors (Gail et al., 2021). Ordination plots can then be used to visualise beta diversity analyses, reducing multi-dimensional microbiome data usually into 2 or 3-dimensions. Samples that are more similar based on beta-diversity distance measures, are plotted closer together on a scatterplot. Commonly used ordination approaches include non-metric multi-

dimensional scaling (NMDS), principal coordinate analysis (PCoA), principal component analysis (PCA) and redundancy analysis (RDA) (Qian et al., 2020).

1.7.7. Animal models and primary cell lines

Omic studies are limited to identifying associations rather than cause and effect. In order to fully appreciate and characterise the relationship between microorganisms and their host, and to better understand the underlying disease mechanisms, interaction studies are crucial. These types of experiments have typically involved *in vitro, in vivo* or *ex vivo* techniques such as the use of animal models or cell lines. For example, a recent study used Caco-2 cells, which are typically used as a model of the intestinal epithelial barrier, to assess the impact of synthetic HMOs on the intestinal barrier transcriptome (Wu et al. 2022). The study found that transcriptome profiles of cells exposed to synthetic HMOs bearing similar side chains, to cluster with one another, and there to be little overlap in gene regulation shared by all synthetic HMOs (Wu et al., 2022). Further, there are a number of groups using animal models to study NEC specifically, including mice, rats and piglets (Lopez et al., 2023). The induction of NEC in these animal models will vary but typically involve a combination of feeding with infant formula, exposure to hypoxia and/or hypothermia, and the addition of LPS or gut microbes obtained from human infants with NEC. However, there are some limitations to these approaches, including the complexity and ethical issues surrounding the use of animal models and the fact that both animal models and cell culture lines are not necessarily representative of specific human tissue types (Mead & Karp, 2019). Culture lines in particular fail to recapitulate all the intestinal cell types (Foulke-Abel et al., 2014), and animal models have an altered endogenous gut microbiome in comparison to humans. As such, emerging technologies have aimed to address these limitations for studying NEC including the use of human intestinal-derived organoids.

1.7.8. Organoids

Organoids are 3D culture structures generated from stem cells (tissue-derived or induced pluripotent stem cells) and have been used to study various diseases and human physiological processes. Human intestinal-derived organoids have been developed as a model to study the human intestine. The architecture of the human intestinal epithelium comprises (1) the villus; which is composed of enterocytes, goblet,

enteroendocrine and tuft cells, and (2) the crypts; where Paneth and *Lgr5*$^{+}$ stem cells reside (**Figure 1.5A**). The apical surface of intestinal epithelial cells faces the gut lumen, which represents the side that cells are exposed to various microorganisms, microbial products and dietary compounds (Blutt et al., 2018). Human intestinal-derived organoids are generated from the intestinal crypts of tissue (i.e., biopsy sample or resected tissue from surgery) that contains *Lgr5*$^{+}$ stem cells, which are able to generate a continuously expanding, self-organising epithelial structure reminiscent of the intestine (Sato et al., 2009). However, one of the unique challenges introduced by using these organoid models is that the apical surface, where cells would be exposed to microorganisms or metabolites etc. *in vivo,* is enclosed within the organoid structure (**Figure 1.5B**).

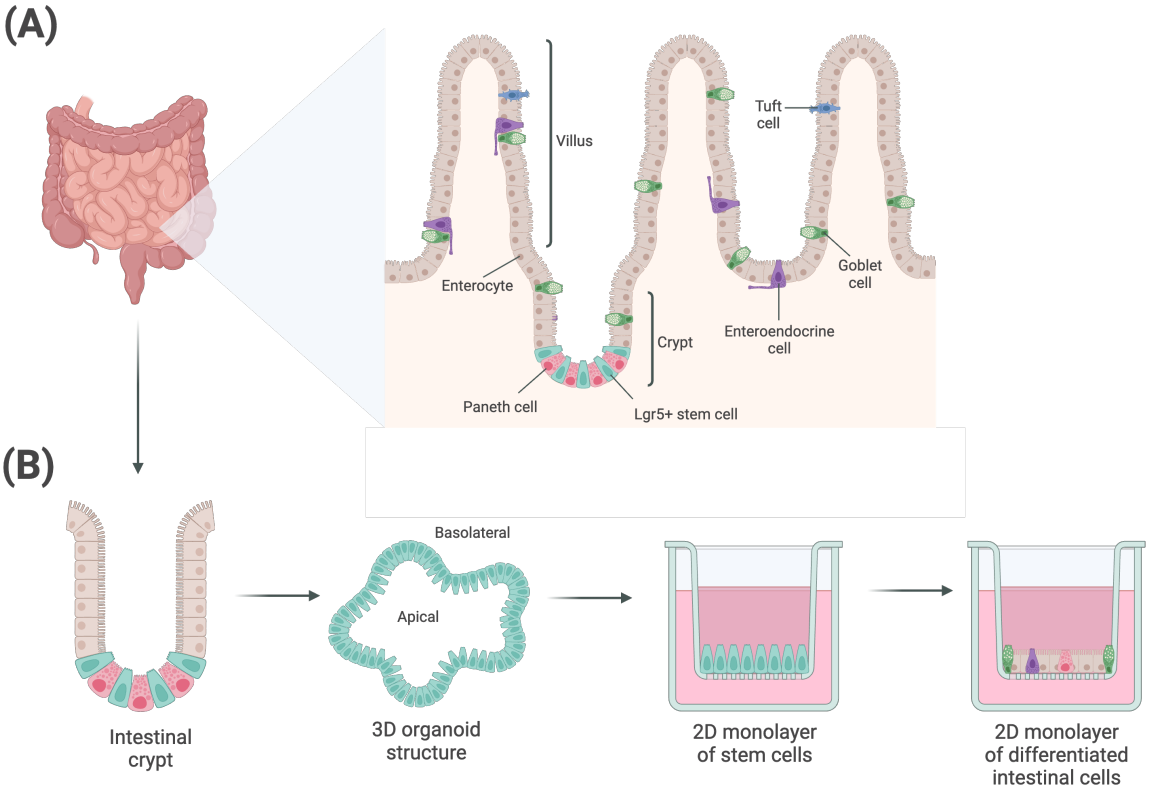

Figure 1.5. Example organoid model for gut microbiome studies. (A) Intestinal epithelium architecture comprising the villus and crypt. **(B)** Typical human intestinal-derived organoid model system set up from isolation of intestinal crypts to dissociation of 3D organoid structure to 2D monolayer of differentiated intestinal cells.

To overcome these challenges, researchers have employed various methods such as dissociating organoid structures into 2D monolayers, then differentiating cells to represent the various intestinal cell types (**Figure 1.5B**). This has the additional

benefits of being optimised to reflect the oxygen condition of the intestinal environment, for example, by having the apical surface exposed to anaerobic conditions to enable the growth of intestinal anaerobic microorganisms (Fofanova et al., 2019). Other methods to overcome these challenges include microinjection to introduce microbes and other products to the organoid 'lumen' and controlling epithelial cell polarity thereby having the apical surface facing outward (Co et al., 2019). Additionally, some groups have looked at seeding 2D monolayers into microfluidic systems, termed the 'intestine-on-chip' model (Chapman & Stewart, 2023). These can be used to provide a fluidic flow of nutrients, metabolites or microbes in a manner that intestinal cells would normally be subject to within the lumen, however, they are labour intensive and complex to use (Chapman & Stewart, 2023). More recently, researchers have established apical-out intestinal organoids as an inflammatory model for NEC (Liebe et al., 2023).

1.8. Aims

The broad aims of this study were to understand how infant factors shape the gut microbiome in preterm infants at risk of NEC. The specific aims were as follows:

- Define infant factors that shape the gut microbiome in healthy preterm infants.
- Identify infant factors and gut microbiome signatures associated with the development of NEC in preterm infants.
- Identify associations and relationships between clinical data, the gut microbiome, metabolome and circulating T-lymphocytes in preterm infants.

Based on these aims, it was hypothesised that the gut microbiome of healthy preterm infants will be influenced by infant factors such as probiotic use and the consumption of breast milk, and that the gut microbial composition of these infants will be distinct to that of infants who go on to develop NEC. Finally, it was hypothesised that there will be associations between infant factors, the neonatal gut ecosystem and systemic immunity.

2. Methods

2.1. Ethics, sample collection and storage

Infants were recruited to the Supporting Enhanced Research in Vulnerable Infants Study (SERVIS) with written parental consent covering data and sample collection. The study protocol was approved by Newcastle Hospitals NHS Foundation Trust, NRES Committee North East and N. Tyneside 2 10/H0908/39, and the research complies with all relevant ethical regulations. All work with clinical samples, including organoids, is covered within these ethical approvals. For Chapter 5, in addition to the ethical cover of SERVIS, infants were recruited to the 'Interactions between the diet and gut microbes and metabolism in preterm infants' (INDIGO; 17/NE/0169) RCT, with written parental consent covering data and sample collection.

All infants were cared for in the NICU of the RVI, Newcastle, with standardised feeding and antibiotic and antifungal guidelines (prophylactic fluconazole). The earliest included infants were born in 2011 and probiotics were introduced into routine use in 2013. Between 2013 and 2016, infants received the probiotic Infloran (*Bifidobacterium bifidum* 1×10^9 colony forming units; CFU, and *Lactobacillus acidophilus* 1×10^9 CFU); then, due to lack of availability, after mid-2016 Labinic (*B. bifidum* 0.67×10^9 CFU., *Bifidobacterium longum* subsp. *infantis* 0.67×10^9 CFU and *L. acidophilus* 0.67×10^9 CFU) was used. Stool samples used in the analysis were collected longitudinally from day 0 until day 120, alongside extensive clinical metadata for each infant, including demographics and treatments such as feed exposures. The clinical variables used vary by chapter and are outlined in section 2.3. Stool and blood samples were collected and stored in the NICU at -20°C before being transferred and stored at -80°C.

2.2. Patient cohorts, sampling overview and study design

Patient demographics pertaining to each of the cohorts are presented within each of the three results chapters. The overall patient cohorts are described below.

2.2.1. Patient cohort: Chapter 3

The study described in Chapter 3 included 123 preterm infants (*n* = 1431 stool samples) born at <32 weeks' gestation without congenital anomaly, EOS, LOS, NEC,

FIP/SIP or other intestinal pathology. These morbidities were excluded because they are most strongly associated with changes in the gut microbiome.

2.2.2. Patient cohort: Chapter 4

The study described in Chapter 4 included 199 preterm infants born at <32 weeks' gestation. Of these infants, 124 (*n* = 1494 stool samples) were from 'healthy control' infants without congenital anomaly, EOS, LOS, NEC, FIP/SIP or other intestinal pathology. Additionally, 75 infants were diagnosed with NEC (*n* = 547 stool samples), of which 24 infants also developed LOS. Diagnoses of NEC were made using a combination of clinical, x-ray and histological findings and blindly agreed by two neonatal clinicians, Dr Janet Berrington and Prof. Nicholas Embleton.

For the cohort used in the 1:1 matched cross-sectional analysis, a single sample from each NEC patient (where possible) was chosen closest to the day of NEC onset. Samples were chosen up to 11 days prior to diagnosis or within four days following diagnosis. A sample from within ± 3 days was then chosen from a healthy control infant, matched to the NEC patient based on gestational age and birthweight. For the 1:1 matched longitudinal analysis, the same matchings were used, and pre-NEC samples that could be matched by day of life (DOL; within ± 3 days) to a healthy control were included in the analysis. For analyses, early NEC was defined as NEC diagnosed before the median day of onset of the matched cohort (≤21 days) whilst late NEC was defined as NEC diagnosed after the day of onset of the matched cohort (≥21 days).

2.2.3. Patient cohort: Chapter 5

The study described in Chapter 5 included 66 preterm infants born <32 weeks of gestation. Stool samples used in this analysis were collected longitudinally (*n* = 266) across 5 timepoints. As part of the INDIGO RCT, infants were randomised to either an exclusive human milk diet (intervention) or standard care (control). The control group consisted of feeding with MOM and the use of preterm formula milk to make up any shortfall MOM supply. The intervention group consisted of MOM with the use of a ready-to-feed pasteurized human milk product (RTF 26, Prolacta Biosciences, Los Angeles, California) to make up any shortfall in MOM. BMF was used and aimed to start within 48 hours of achieving a milk intake of 150 mL/kg per day. Infants in the

control group received commercially available, bovine-derived fortifier (Nutriprem [Nutricia Ltd] or SMA Fortifier [SMA Nutrition UK]) and infants in the intervention group received a pasteurized human milk–derived fortifier (P+6, Prolacta Biosciences). Importantly, the amount of formula, BMF and MOM used will vary widely between infants.

A subset of 101 samples from 56 infants were sent for metabolomics; 85 were the same stool sample as was used for 16S rRNA gene sequencing, 8 were a stool sample from the same day and the remaining 8 were a stool samples from within ± 3 days. A further subset of 41 samples were matched for blood from 24 infants and sent for T-lymphocyte profiling using CyTOF (**Figure 2.1**). Where probiotics were used, infants received Labinic.

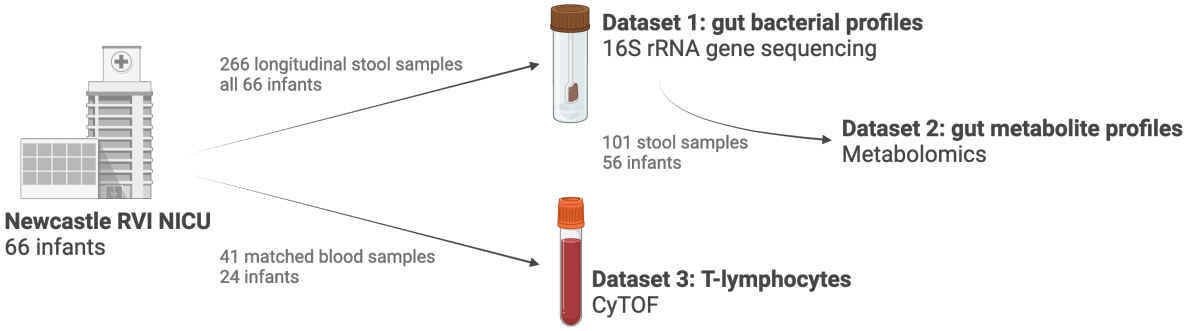

Figure 2.1. Schematic overview of INDIGO study design and sample collection.

2.3. Clinical variables

Clinical data was collected alongside samples to look for associations with data. Variables that are fixed through time (e.g., gestational age, birth mode, sex, etc,), are described on a per-infant basis and thus constant for all samples from a given infant. Other variables were categorised to reflect exposure in relation to time (e.g., antibiotics, receipt of MOM, etc.), and therefore are on a per-sample basis.

2.3.1. Clinical variables: Chapter 3

The clinical variables used in the statistical analyses for Chapter 3 were gestational age at birth (continuous; range 23–31), birthweight (continuous; range 500–2,000 g), birth mode (vaginal/caesarean), sex (male/female), season at birth (winter/spring/summer/autumn), intravenous antibiotics in the past 7 days (no/yes), day

of full feed (continuous; range 6–39), MOM (never/during/after), breast milk fortifier (BMF; never/before/during/after), formula (never/before/during/after), probiotics (no probiotic/Infloran/Labinic) and weight *z*-score difference between birth and discharge (continuous; range –5.4 to 1.1). The variables included are based on variables that have either been identified in previous studies (either term or preterm infant) to be associated with changes in the gut microbiome e.g., various dietary factors, antibiotics, birth mode. Gestational age at birth and birthweight were included to control for the degree of prematurity/health status at birth, season at birth was included as a proxy for changes in the neonatal unit microbial environment over the year and weight z-score difference between birth and discharge was included as a proxy for health status. Additional clinical information, such as ethnicity, were not readily available, but it is important to note that these variables not included, as well as unknown confounders, may be clinically relevant and associated with changes in the gut microbiome.

For the persistence analysis (described further in section 2.12.7.) all co-variates needed to be on a per-infant basis restricting the analysis to gestational age, birthweight, birth mode, sex, season, total number of antibiotic courses, day of full feed, BMF ever (no/yes), formula ever (no/yes), probiotics and weight *z*-score change. MOM could not be included in this particular analysis because there was only one baby who did not receive MOM in this subset.

2.3.2. Clinical variables: Chapter 4

The clinical variables used in the statistical analyses for Chapter 4 were gestational age at birth (continuous; range 23–31), birthweight (continuous; range 500–2,000 g), birth mode (vaginal/caesarean), sex (male/female), intravenous antibiotics in the past 7 days (no/yes), day of full feed (continuous; range 6–120, for infants who died before reaching full feeds, day 120 was used), MOM (never/before/during/after), BMF (never/before/during/after), formula (never/before/during/after), probiotics (no probiotic/Infloran/Labinic) and NEC (no-NEC/pre-NEC/post-NEC) or NEC ever (no/yes). Season at birth and weight z-score were not included in this chapter due to this data not being readily available for a number if infants in this cohort, and these variables being shown to have no significant association with the healthy preterm gut microbiome.

For the persistence analysis (described further in Section 2.12.7.) all co-variates needed to be on a per-infant basis restricting the analysis to gestational age, birthweight, birth mode, sex, total number of antibiotic courses, day of full feed, BMF ever (no/yes), formula ever (no/yes), probiotics and NEC ever (no/yes). The same variables were used for the microbiome stability analysis (described in more detail in Section 2.12.7.), except for day of full feeds.

2.3.3. Clinical variables: Chapter 5

The clinical variables used in the statistical analysis for Chapter 5 were again similar to those used in the previous two chapters: gestational age at birth (continuous; range: 23 – 29.4), birthweight (continuous; range: 475 - 1620), birth mode (vaginal/caesarean), sex (male/female), intravenous antibiotics in the past 7 days (no/yes) and day of full feed (continuous; range: 9 - 47). The feeding variables used were different to those used in Chapter 3 and 4, and instead included % MOM in previous 3 days (continuous; range: 0-100%) and fortifier at time of sample (no/yes).

For the microbiome stability analysis (described in more detail in Section 2.12.7.) all co-variates needed to be on a per-infant basis restricting the analysis to gestational age, birthweight, birth mode, sex, total days on antibiotics, day of full feeds and total days of MOM. Fortifier could not be included in this particular analysis as the model would not converge.

2.4. Stool DNA extraction

DNA was extracted from ~0.1g of stool using the DNeasy PowerSoil Kit (QIAGEN) following the manufacturer's protocol, with minor modifications. Samples were vortexed at maximum speed for 20 min using a Vortex Adapter tube holder and eluted in ~70μl of solution C6 before being stored at -80°C. Subsequent steps were performed according to the protocol. A kit negative control (i.e., with no stool sample) was extracted with every batch of 24 samples.

2.5. Metagenomic shotgun sequencing, taxonomic and functional profiling

Metagenomic sequencing was performed either by Astarte medical or by Baylor College Medicine and subsequent profiling by the Alkek Centre for Metagenomics and

Microbiome Research (Baylor College Medicine, USA). Library prep was performed using the Nextera DNA Flex Kit. Sequencing was performed on the HiSeq X Ten (Illumina) with a target read depth of 10M reads per sample with a read length of 150bp paired end reads. Raw fastq files were quality trimmed and Illumina adapters removed using bbduk (BBMap version 38.69) (Truong et al., 2015). Trimming parameters included kmer length of 19, allowing one mismatch, and a minimum Phred score of 20. Post-trimming, reads with a minimum average Phred <17 and length <50 bp were discarded. Host contamination reads were identified by mapping trimmed fastq files to a combined database containing the hg38 reference human genome and PhiX (standard Illumina spike in) using bbmap (BBMap version 37.58) (Truong et al., 2015) with kmer length of 15, bloom filter enabled, and fast search settings.

Host reads were subsequently removed, and taxonomic profiling of subsequent files was performed by using MetaPhlAn v.2.0 (Segata et al., 2012) (bacterial, archaeal and fungal taxonomic classification) and VirMAP v.1.0 (Ajami et al., 2018) (viral taxonomic classification) based on default settings. Functional profiling was performed using HUMAnN v.2.0 (Franzosa et al., 2018) based on default settings. The Comprehensive Antibiotic Research Database was used to identify antibiotic resistance genes within the metagenomic dataset based on the standard resistance gene identifier protocol with default parameters (McArthur et al., 2013). Microbial enzymes (level-4 enzyme commission (EC) categories) were quantified by Dr. Tommi Vatanen, by summing the abundances of individual gene families mapping to each EC number based on UniRef90-EC mapping from UniProt (Bateman et al., 2017). *B. infantis* HMO genes (as previously described (Casaburi et al., 2021)) were also quantified by Dr. Tommi Vatanen, by first identifying the corresponding UniRef90 gene families and then utilising *B. longum*-stratified gene quantifications (quantifying UniRef90 gene families) from HUMAnN v.2 (Vatanen et al., 2019). Samples with >90% of the genes in these six genomic loci (H1, H2, H3, H4, H5 and a urease gene cluster) were classed as having *B. infantis*.

2.6. 16S rRNA gene sequencing and taxonomic profiling

V4 16S rRNA gene sequencing was performed by NU-OMICS at Northumbria University using the following protocol. The V4 region of the 16S rRNA gene was amplified by polymerase chain reaction using the barcoded Illumina adapter-containing

primers 515F and 806R (Caporaso et al., 2012). Sequencing was performed on the Illumina MiSeq platform, with a target read depth of 10k and a read length of 250 bp paired end. Raw data processing was performed, where merging allowed zero mismatches and a minimum overlap of 50 bases. Merged reads were trimmed at the first base with a *q* less than or equal to 5 and samples were rarefied at 2000 reads per sample. Subsequent fastq files were processed by the Alkek Centre for Metagenomics and Microbiome Research (Baylor College Medicine, USA). 16S rRNA gene sequences were assigned operational taxonomic units (OTUs) based on a similarity cut-off value of 97% using the UPARSE algorithm. OTUs were then mapped to the SILVA Database containing only the 16S V4 region, to determine taxonomies. Abundances were then recovered by mapping the demultiplexed reads to the OTU file.

2.7. Untargeted metabolomics

2.7.1. Untargeted metabolomics: Chapter 3

For the metabolomics in Chapter 3, a subset of 10 stool samples representative of each preterm gut community type (PGCT) and matched serum were selected for LC-MS. As PGCTs were strongly associated with DOL at sampling, samples were primarily chosen to match for DOL between PGCTs to mitigate confounding by age at sampling. Other clinical variables were matched in addition, including gestational age, birthweight, birth mode and sex. Based on these criteria, no clinical variable was significantly different between PGCTs (all $P>0.05$).

Metabolomics was performed by the Marsland Lab in the Department of Immunology and Pathology (Monash University, AU), using the following protocol. Metabolites were extracted using a methanol solvent solution, supplemented with 1 μM MS internal standards (CAPS, CHAPS and PIPES) and 5 μM 2,6-di-*tert*-butyl-4-methylphenol. Serum samples were centrifuged at 800 x*g* for 5 min, supernatants collected, and the solvent solution added at a 4:1 ratio. Samples were shaken for 1 h at 4 °C, centrifuged at 14,000*g* for 10 min and supernatants collected. Liquid from stool samples was evaporated using a Speedvac (Thermo Fisher Scientific) and a solvent solution was added at a ratio of 300 μl per 10 μg. Samples were shaken for 1 h at 4 °C, followed by centrifugation at 14,000*g* for 20 min, and supernatants collected.

The LC–MS data were acquired on a Dionex Ultimate 3000 rapid separation high-performance liquid chromatography system (Thermo Fisher Scientific) coupled with a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific). Chromatographic separation was performed on a ZIC-pHILIC column (5 μm, polymeric, 150 × 4.6 mm^2, SeQuant, Merck). Mobile phase (A) was 20 mM ammonium carbonate and (B) acetonitrile. The gradient programme started at 80% (B) and reduced to 50% (B) over 15 min, then reduced to 5% (B) over 3 min, where washing occurred for 3 min; finally, there was an 80% (B) re-equilibration for 8 min. The flow rate was 0.3 ml min^{-1} and the column compartment temperature was 40 °C. Total run time was 32 min with an injection sample volume of 10 μl. The mass spectrometer operated in positive and negative polarity, switching at 35,000 resolution and 200 *m*/*z* with detection range of 85–1,275 *m*/*z* in full-scan mode. An electrospray ionization source (ESI) was set to 3.5 kV voltage for positive mode and 4.0 kV for negative mode, sheath gas was set to 50 and aux gas to 20 arbitrary units, capillary temperature 300 °C and probe heater temperature 120 °C. Serum samples were analysed as a single batch, as were stool samples. Each sample set was randomised to account for system drift. Mixtures of pure authentic standards containing approximately 320 metabolites were acquired as separate injections and used to confirm retention times.

The raw LC–MS data of both serum and stool samples were independently processed as stated in the metabolome–lipidome–MS-DIAL pipeline using MS-DIAL v.4.8 (Tsugawa et al., 2020). Metabolomic processing was conducted in positive and negative ion mode. Default parameters were applied unless otherwise stated. Peak detection parameters included a minimum peak amplitude of 100,000. Peaks were identified using the MassBank database v.2021.02 (Horai et al., 2010) with a retention time tolerance of 0.1 min, accurate mass tolerance of 0.002 Da and identification score cut-off of 80%. Peaks were aligned using a retention time tolerance of 0.3 min and accurate mass tolerance of 0.002 Da, with gap filling by compulsion. MS/MS was exported and further processed for secondary annotation using the Global Natural Products Social Molecular Networking feature-based molecular networking tool (Nothias et al., 2020).

Peak intensity tables were exported from MS-DIAL and the R package pmp v.1.6.0 (Jankevics et al., 2021) was used for the following quality control (QC) and pre-

processing steps. Peaks were filtered for intensities at least fivefold higher than LC–MS blanks, samples with >80% missing values, features with >20% missing values and peaks filtered based on the percentage of variation in the QC samples with a maximum relative standard deviation of 25%. Based on this, one stool sample from PGCT-2 was excluded. Data were normalised using probabilistic quotient normalisation, followed by Random Forest missing data imputation using the missForest R package v.1.4 (Stekhoven & Bühlmann, 2012) and subsequent generalised logarithmic (glog) transformation. MS1 data were further annotated using the human metabolome database (HMDB, v.4, July 2021) (Wishart et al., 2018), with an AMT of 0.002 Da. Any unannotated features were removed. The remaining dataset was subject to manual feature curation in MS-DIAL, where poor quality spectral features were removed.

2.7.2. Untargeted metabolomics: Chapter 5

Metabolomics was performed on a subset of 101 samples from 56 infants by Metabolon (North Carolina, US) using the following protocol (methods taken from https://www.metabolon.com/support/portal/experimental-procedures/) . Samples were prepared using the automated MicroLab STAR® system from the Hamilton Company. Proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into fractions: two for analysis by separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode ESI, one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, and one for analysis by hydrophilic interaction liquid chromatography (HILIC)/UPLC-MS/MS with negative ion mode ESI. Samples were placed on a TurboVap® (Zymark) to remove the organic solvent and sample extracts were stored overnight under nitrogen before preparation for analysis.

All methods utilised a Waters ACQUITY UPLC and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer. The sample extract was dried then reconstituted in solvents compatible to each of the methods. Each reconstitution solvent contained a series of standards at fixed concentrations. One aliquot was analysed using acidic positive ion conditions, chromatographically optimised for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1×100 mm, 1.7 μm) using water and methanol,

containing 0.05% perfluoropentanoic acid and 0.1% formic acid. Another aliquot was also analysed using acidic positive ion conditions, however it was chromatographically optimised for more hydrophobic compounds. In this method, the extract was gradient eluted from the same C18 column with the addition of acetonitrile and 0.01% FA instead of 0.1%, and was operated at an overall higher organic content. Another aliquot was analysed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analysed via negative ionisation following elution from a HILIC column (Waters UPLC BEH Amide 2.1×150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z.

Raw data was extracted, peak-identified, QC processed and biochemical identifications made using Metabolon's in-house hardware and software. Peaks were quantified using area-under-the-curve.

2.8. Preterm intestinal organoid co-culture system

A human intestinal organoid line was generated from preterm intestinal ileum tissue after surgical resection for NEC (Stewart et al., 2020). The infant was a male born at 24 weeks' gestation and had surgery on DOL 10. Organoids were established and propagated using a method described previously by Stewart *et al.* (Stewart et al., 2020). Dr. Andrea Masi prepared all media and isolated intestinal crypts from tissue. In brief, ~5 mm^2 of tissue was minced and washed with complete chelating solution supplemented with 2.5 μg/ml Amphotericin b (Gibco) and 100 μg/ml Penicillin/Streptomycin (Gibco). To isolate intestinal crypts, tissue was incubated twice in 3 ml complete chelating solution with 0.03 M Ethylenediaminetetraacetic acid (EDTA) then 0.04 M of EDTA shaking for 30 min at 300 rpm (4°C). Supernatant was collected after each incubation and centrifuged for 5 min at 1000 rpm (4°C). Isolated crypts were resuspended in 70 μl Matrigel (Corning), dispensed into a 24-well tissue culture plate and incubated with 500 μl high Wnt media supplemented with 10 μM Y-27632 (Sigma Aldrich). Media was changed every 2-3 days and supplemented with

2.5 μg/ml Amphotericin b and 100 μg/ml Penicillin/Streptomycin for the first 3 passages.

To passage cells, media was removed from wells and Matrigel was disrupted with 500 μl of cold (4°C) Complete Media Growth Factor negative (CMGF-). Organoids were collected and pooled from well-to-well and mechanically disrupted using a 25-gauge needle. Cells were centrifuged for 5 min at 1000 rpm (4°C), resuspended in Matrigel (sufficient to seed ~30 μl in each well) and incubated as above.

To generate monolayers, transwell inserts (6.5 mm, 0.4 μm pore size; Corning) were coated with 100 μl of Matrigel diluted 1:40 in phosphate buffered saline (PBS) and incubated at 37 °C. Organoids were washed in 0.5 mM EDTA, incubated with 1 ml 0.05% Trypsin – 0.5 mM EDTA for 5 min at 37 °C, passed through a 40 μm cell strainer and centrifuged for 5 min at 1200 rpm (4°C). Excess PBS was removed from transwells and inserts were seeded with $5x10^5$ cells suspended in 200 μl CMGF+ media supplemented with 10 μM Y-27632. In addition, 650 μl of media was added to the basolateral side. Transepithelial electrical resistance (TEER) was measured every day using the Millicell® ERS-2 (Electrical Resistance System), and differentiation media was used after TEERs reached ~300 Ohm. The media was renewed every other day.

Intestinal organoid monolayers ($n=3$ technical replicates) were exposed to pooled faecal supernatants representing each PGCT and a control containing no faecal supernatant. Sterile faecal supernatants were prepared using a modified method described elsewhere (Henrick et al., 2021). Briefly, ~0.25 g of stool ($n=10$) was pooled for each PGCT and diluted in 25% (w/v) sterile PBS before being vortexed for 20 min with glass beads. Faecal slurries were centrifuged for 20 min at 1,600*g* (4 °C), the supernatant was re-centrifuged for 10 min at 14,000*g* and 4 °C, and the resulting supernatant was serially filtered (0.45 μm and 0.22 μm). Faecal supernatant was stored at –80 °C until use. Intestinal organoids were seeded as monolayers on 0.4 μm Transwells (Corning) and, after reaching confluence (~2 d), were differentiated for 4 days (Ettayebi et al., 2016).

Co-culture of preterm intestinal organoid monolayers with sterile faecal supernatants was performed for 24 h using the organoid anaerobe co-culture (OACC) model

(Fofanova et al., 2019), after 4 days of monolayer differentiation. The sterile faecal supernatants were added apically, corresponding to the intestinal lumen. The OACC model was used to recapitulate the steep oxygen gradient across the epithelium and mimic the low oxygen gradient of the ileum. A gas mix containing 5% O_2, 5% CO_2 and 90% N_2, was supplied to the basolateral side. TEER was measured at the end of the experiment to confirm that all monolayers remained intact, and cells were contiguous.

2.9. RNA-sequencing

After 24 h of xposure, RNA was extracted from organoid monolayers using the RNeasy kit (QIAGEN) before undergoing RNA-sequencing (RNA-seq) at the Newcastle University Genomics Core Facility. One sample from the PGCT-5 exposure failed QC and was not included in the subsequent analysis. Briefly, stranded messenger RNA-seq libraries were prepared using the TruSeq Stranded mRNA kit (Illumina) and IDT for Illumina TruSeq RNA UD Index adapters following the manufacturer's protocol. All samples had an A260/A280 ratio ~2.0 (purity measure) and an RNA integrity number (RIN) score of >7. Libraries were quantified using a TapeStation 4200 (Agilent Technologies) and Qubit 4 (Thermo Fisher Scientific) and equimolar pooled. The pooled library was sequenced at ~50 million 100 bp single-reads per sample on a NovaSeq 6000 using an S2 100 cycle flow cell (Illumina). Data for individual samples were demultiplexed into separate FASTQ files using Illumina's bcl2fastq software.

QC of raw reads was performed using fastq_quality_trimmer from the FASTX Toolkit v.0.0.14 before being mapped to the human transcriptome (GRCh38.p13) using Salmon v.0.13.1 (Patro et al., 2017) to estimate transcript abundance. Estimated count data were aggregated at the gene level by tximport (Soneson et al., 2016) for downstream analysis. DESeq2 v.1.32.0 (Love et al., 2014) was used to normalize RNA-seq count data and identify differentially expressed genes (DEGs) between PGCT and control replicates. Genes were considered differentially expressed if they displayed an absolute positive or negative fold-change of ≥1.5 and a false discovery rate (FDR)-adjusted $P<0.05$. A Venn diagram of DEGs was produced using the VennDiagram package v.1.7.1 (https://cran.r-project.org/web/packages/VennDiagram/index.html).

2.10. Mass cytometry

CyTOF was performed on 41 matched blood samples from 24 infants. Dr. Christopher Lamb and Dr. Rebecca Payne were involved in the initial design of the antibody panel (**Table 2.1**). Dr. Thomas Sproat prepared all of the samples and the Newcastle University Flow Cytometry Core Facility ran the samples, using the following protocol. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples (~350 – 500 μl blood) by re-suspending samples in HBSS (Sigma-Aldrich®) to a final volume of 1.2 ml. The suspension was pipetted on top of 5ml lymphoprep (Stemcell technologies) before centrifugation at 1200 g for 20 min with an acceleration of 6 and deceleration of 1. The mononuclear layer containing PBMC's was identified and aspirated with a 2 ml Pasteur pipette, washed twice with PBS before being counted. 3-6 x 10^6 PBMC's were isolated and stained with Cell-ID™ Cisplatin (Fluidigm®) and incubated for 5 min before staining with the primary antibody stain with antibodies bound to fluorochromes followed by the secondary antibody stain with remaining cell surface antibodies, both for 30 min. Cells were then fixed for 1 h with 3.2% formaldehyde and stored in freezing media (10% dimethyl sulfoxide in foetal bovine serum) at -80°C. Prior to mass cytometry, cells were defrosted at room temperature, permeabilised with Triton Perm Buffer (PBS + 2% FBS + 0.1% Triton X100), suspended in heparin for 10 min then incubated with intracellular antibodies overnight. Cells were incubated with Maxpar fix and perm buffer with 125 nM Cell-ID™ Intercalator (Fluidigm®) for 1 h, washed, and filtered through a 40 μm filter. All steps were performed with washes in between. Cells were counted using a BD Accuri™ C6 flow cytometer. Samples were then diluted in 10% EQ™ Four Element Calibration Beads (Fluidigm®) milli-Q water to a final concentration of 5 x 105 cells/ml. Cells were analysed on a Helios mass cytometer (Fluidigm®). Sample acquisition occurred at a flow rate of 30 μl/min and each sample was run for 30 min, aiming for an acquisition of 5 x 10^5 events per sample.

Table 2.1. Antibody panel used for CyTOF.

Metal isotope	*Marker*
89Y	CD45
CD113	CD19
141Pr	CD196 (CCR6)
142Nd	CTLA4
143Nd	TCR γδ
144Nd	CD3
145Nd	CD4
146Nd	CD8a
147Sm	CD161
148Sm	CD86
149Sm	CD25
150Nd	CD199 (CCR9)
151Eu	CXCR3
152Sm	CD1d
153Eu	CD56
154Sm	CD49b
155Gd	CD45RA
156Gd	CD335
157Gd	Anti-FITC
158Gd	CD279 (PD-1)
159Tb	CD197 (CCR7)
160Gd	Valpha7.2
161Dy	Tbet
162Dy	FoxP3
163Dy	CD294
164Dy	CD69
165Ho	TIGIT
166Er	Granzyme B
167Er	Gata3
168Er	RORγ
169Tm	CXCR5
170Er	Anti-PE
170Yb	Valpha24
171Yb	CD28
172Yb	CD14
173Yb	CD117
174Yb	HLADR
175Lu	CCR4
176Yb	CD127
209Bi	CD16

2.11. Enzyme-linked immunosorbent-assay

Interleukin (IL)-8 was measured from apical and basolateral media harvested after the co-culture using the DuoSet Enzyme-linked immunosorbent-assay (ELISA) Kit (R&D Systems) following the manufacturer's instructions with some modifications. The volumes used at each step were halved, except for the addition of substrate solution and stop solution which were added at the suggested volumes. The absorbance at 450 nm was measured in a microplate reader. A standard curve was fit to the data using the 'lm' function from the R stats package V.4.1.3., from which sample concentrations were determined. A Kruskal-Wallis test was used to determine whether there was a significant difference between groups.

2.12. Statistical analysis and data visualisation

All statistical analyses were performed in R (https://www.r-project.org/) V.4.0.2. Unless stated otherwise, all visualisations were plotted using the ggplot2 package V.3.3.2 (Wickham, 2016). All appendices in this thesis can be found on GitHub (https://github.com/laurencbeck/supplementary_tables).

2.12.1. Alpha diversity

Shannon diversity and species richness were calculated for each sample using the vegan package V.2.5-7 (https://cran.r-project.org/web/packages/vegan/index.html). Alpha diversity measurements and relative abundance data were modelled using locally estimated scatterplot smoothing (LOESS) regression and plotted with 95% confidence intervals (CI). For 16S data (Chapter 5), whilst alpha diversity analyses were based on OTU, all other microbiome analyses were based on genera.

2.12.2. Determining PGCTs

Dirichlet Multinomial Mixtures (DMM) was used to cluster samples within each cohort/chapter on the basis of microbial community structure (Holmes et al., 2012) to determine PGCTs. Five PGCTs was found to be optimal for samples in Chapter 3 based on the lowest LaPlace approximation score. The same number of PGCTs were used throughout this thesis for consistency. PGCTs were manually ordered youngest (PGCT-1) to oldest (PGCT-5) based on the average DOL of samples within each PGCT. The Linear discriminant analysis Effect Size (LEfSe) (Segata et al., 2011)

method was used to determine the bacterial species that discriminated each cluster based on linear discriminant analysis (LDA), using default parameters.

2.12.3. Determining T-lymphocyte sub-populations

The initial manual gating of the CyTOF data to identify CD45+ CD3+ T-lymphocytes was performed by Dr. Thomas Sproat as follows. Data was processed using CyTOF software (Fluidigm®) and normalised using the EQ™ Calibration beads followed by Gaussian normalisation of the data. The subsequent gating strategy was implemented in FCS Express, with the exclusion of dead cells and doublets followed by positive gating of CD45+CD3+ T-lymphocytes. These populations were then ensured to be CD19, CD66b, and CD14 negative. FCS files containing data on T-lymphocyte cells only were then imported into R where clusters were defined using Flow self-Organising Maps (FlowSOM) (Van Gassen et al., 2015) based on marker expression to identify T-lymphocyte sub-populations. The subsequent sub-populations were then validated by manual gating in FlowJo V.10.8.1. Uniform Manifold Approximation and Projection (UMAP) was used to visualise sub-populations based on 1000 cells per sample using the umap R package V.0.2.8.0.

2.12.4. Determining Preterm Metabolic Profile Types and Preterm T-lymphocyte Profile Types

Metabolite data were clustered initially using DMM as was used to determine PGCTs, these clusters were to be termed Preterm Metabolic Profile Types (PMPTs). Clustering was then attempted using a consensus-based algorithm, three clusters were found to be optimal which were then defined using a hierarchical clustering approach based on complete linkage. The same approach was used to cluster T-lymphocyte data into Preterm T-lymphocyte Profile Types (PTPTs), for which two clusters were found to be optimal.

2.12.5. Permutational multivariate analysis of variance

To determine which clinical co-variates were associated with the various features explored in each chapter while accounting for repeated measures, multiple cross-sectional analyses by permutational analysis of variance (PERMANOVA) using the ‘adonis’ function from the vegan package were performed. Data were split into nine specific time windows based on DOL (Chapter 3 and 4) which were chosen to both

maximize the number of samples within each window and also reflect the progression of enteral feed independence as follows: establishing enteral feeds (0–9), reaching full feeds (10–14), independent of PN (15–19) and maturation on full enteral feeds (20–24, 25–29, 30–34, 35–39, 40–49, 50–69). For Chapter 5, data were split into five pre-defined time-points (A, B, C, D and E) for the 16S rRNA gene sequencing data, into three timepoints including merged timepoints (AB, CD and E) for the metabolomics data and into pre and post fortification for the CyTOF analysis. This was done to reflect the INDIGO study design and how samples were collected, in order to maximise the number of samples in each group. Only a single sample per infant, the earliest available, was included within each time window. The association of clinical variables (defined in Section 2.3.2) on the various feature tables was tested, based on Bray–Curtis dissimilarity. Each test was performed in a stepwise manner and subsequent *P*-values were adjusted for multiple comparisons using FDR adjustment (Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995)).

For Chapter 3, to assess whether there was a statistically significant difference in serum and stool metabolite profiles based on PGCT assignment, PERMANOVA was performed using MetaboAnalyst v.5.0 (Pang et al., 2021).

2.12.6. Ordination

For all data, ordinations were performed using NMDS based on Bray–Curtis dissimilarity matrices using the 'metaMDS' function from the vegan package, unless stated otherwise. For RNA-seq data, euclidean distance on regularized logarithm (rlog), transformed, normalised count data was used. The mean centroid for each group was calculated and plotted. *P*-values were calculated using the 'adonis' or 'envfit' function from the vegan package.

For metabolite analysis of stool and serum samples in Chapter 3, ordinations were performed on all data using partial least-squares discriminant analysis (PLS-DA) using MetaboAnalyst v.5.0 (Pang et al., 2021).

2.12.7. Linear mixed models and generalised linear mixed models

Various LMMs and GLMMs were fit to the data using the glmmTMB package V1.0.2.1(Brooks et al., 2017), or alternatively the logistf package (https://cran.r-project.org/web/packages/logistf/index.html) V.1.24, which was used to fit logistic

regressions using Firth's bias-Reduced penalised-likelihood, when there was quasi-complete or complete separation. To detect separation and infinite maximum likelihood estimates in binomial logistic regression models, the 'detect_separation' and 'check_infinite_estimates' functions from the brglm2 package (https://cran.r-project.org/web/packages/brglm2/index.html) V.0.7.1 were used. Model validity was assessed using diagnostic residual plots, generated by the DHARMa package (https://cran.r-project.org/web/packages/DHARMa/index.html) V.0.3.3.0. Diagnostic residual plots were not generated for models fit by logstif. The general formula for each of the LMMs fitted was as follows:

$$Y \sim X_1 + X_2 + \ldots X_n + (1|SubjectID)$$

To determine which clinical co-variates were significantly associated with the five PGCTs for each chapter, individual mixed-effects binomial logistic regression models were fit, one for each cluster versus all other clusters. Each model contained the same clinical co-variates (outlined in Section 2.3.2) plus DOL as fixed effects and subject ID as a random group intercept.

For Chapter 3, mixed-effects binomial logistic regression models were also fit to assess the prevalence of probiotic species. DOL had an effect on the relative abundance of probiotic species, so the before, during and after probiotic groups are nested in time. To account for this, the control group of samples from infants who had taken no probiotic was subset into three distinct time bins. These specific time bins were based on the mean start DOL for probiotics (8 DOL) and the mean stop DOL for probiotics (44 DOL). Mixed-effects binomial logistic regressions were fit separately within groups (before, during and after probiotics) for each probiotic species. They were also fit separately between groups (Infloran and Labinic) for each species.

To assess the persistence of probiotic species following probiotic cessation, persistence analysis was performed in Chapter 3 and Chapter 4. All infants receiving probiotics that had at least two samples after probiotics were stopped were included in the analyses. Infants were classed as 'non-persister' for a species if there were two consecutive samples with a relative abundance of zero. This criterion was found to be optimal and babies could be separated quite clearly into 'persisters' and 'non-

persisters'. Binomial logistic regressions were fit for each probiotic species to determine which co-variates were significantly associated with persistence, using the logistf package as previously described above. The models included the subject-level co-variates as described in Section 2.3.1.

In Chapter 4 and Chapter 5, to assess the stability of the microbiome between PGCTs and whether this was associated with clinical information, microbiome stability analysis was performed. All infants with at least two microbiome samples were included in the microbiome stability analysis. To find out which covariates were significantly associated with microbiome stability, a binary outcome (stable/unstable) was used. For Chapter 4, infants were classified as stable if their microbiome had made one transition and unstable if their microbiome had made two or more transitions between PGCTs. For Chapter 5, infants were classified as stable if their microbiome had made no transitions and unstable if their microbiome had made one or more transitions between PGCTs. These cut-offs were chosen based on the most optimal way to split the data. For Chapter 5, as this analysis was looking across the entire time-course, the number of transitions between PGCTs as a continuous outcome (0, 1, 2, 3) was also assessed. Mixed effects models were fit using the binomial (binary outcome) and Gaussian distribution (continuous outcome). All covariates included in the 'adonis' analysis plus the number of samples from a given infant were included as fixed effects in the models, and subject ID was included as a random group intercept.

Where used, global *P*-values for fixed effects from the final models were obtained by analysis of variance (ANOVA; Type II Wald Chi-square test) from the car package V.3.0-10 (Fox & Weisberg, 2019). All post-hoc analysis was performed using either pairwise comparisons (Tukey HSD method) or treatment vs control comparisons (Dunnet's test), both adjusting for multiple comparisons, using the emmeans package (https://cran.r-project.org/web/packages/emmeans/index.html) V.1.5.4.

2.12.8. MaAsLin analysis to determine significant features associated with each co-variate

The MaAsLin2 package v.1.2.0 (Casaburi et al., 2021; Mallick et al., 2021) was used to determine features significantly associated with clinical co-variates, while adjusting for potential confounders. All clinical co-variates used in the 'adonis' analysis for each

respective chapter plus DOL were included as fixed effects in the analysis and subject ID was included as a random effect. The arcsin square root transformation was performed on relative abundance data and default MaAsLin2 parameters were used. All *P* values were adjusted by MaAsLin2 for multiple comparisons using FDR adjustment (Benjamini–Hochberg procedure) and the default *q*-value cut-off of 0.25 was used to identify significant results.

2.12.9. RNA-seq enrichment and network analysis

Significance analysis of microarray and metabolites was performed in MetaboAnalyst (Pang et al., 2021) with a Delta threshold of 1.0 to identify specific metabolites discriminating PGCT-3 from PGCT-4/-5 and vice versa in both stool and serum. GO and enrichment analysis were performed using the gprofiler2 package v.0.2.1 (Peterson et al., 2020), with default parameters and a customized genetic background. The top 25 most significant GO biological processes for PGCT-4 and PGCT-5 were reported. A network of interactions was inferred using String V.12.0 (https://string-db.org/), removing nodes that were not connected.

2.12.10. Microbiota age and microbiota-for-age Z score

A random forest regression model was used to determine microbiota maturity, using the 'randomForest' (https://cran.r-project.org/web/packages/randomForest/randomForest.pdf) R package, as previously described (Subramanian et al., 2014). The model was trained on a dataset of 29 'healthy' infants who received no probiotic (n = 462 samples, 22.6% of overall cohort), based on the relative abundance of 19 age-discriminatory species. The model was used to predict age based on DOL. The age of the subject predicted by this model was termed microbiota age and was further used to determine microbiota-for-age Z (MAZ)-scores, as previously described. To generate MAZ scores, the microbiota ages of study members predicted by this model were compared to the median microbiota age of chronologically age-matched children in the healthy reference group.

2.12.11. Multiple co-intertia and generalised Procrustes analysis

Multiple co-inertia analysis (MCIA) was performed using the omicade4 R package V.1.32.0 to integrate the datasets. *P*-values were obtained by pairwise Monte-Carlo Tests on the sum of eigenvalues from the MCIA. Generalised Procrustes analysis

(GPA) was performed using the vegan package using PCoAs generated for the three datasets generated in Chapter 5. Microbiome coordinates were used as the original reference to superimpose the metabolome PCoA onto, then the mean of the superimposed configuration (i.e., the mean between each pair of co-ordinates from the microbiome and metabolome PCoAs) was used as the reference to superimpose the T-lymphocyte PCoA onto. *P*-values were obtained using the 'protest' function from the vegan package.

3. Characterising the healthy preterm gut microbiome

3.1. Abstract

The development of the gut microbiome from birth plays important roles in short- and long-term health, but factors influencing preterm gut microbiome development are poorly understood. Metagenomic sequencing was used to analyse 1431 longitudinal stool samples from 123 very preterm infants (<32 weeks' gestation) who did not develop intestinal disease or sepsis over a study period of ten years. During the study period, one cohort had no probiotic exposure whilst two cohorts were given different probiotic products: Infloran (*B. bifidum* and *L. acidophilus*) or Labinic (*B. bifidum, B. longum* subsp. *infantis*, and *L. acidophilus*). MOM, BMF, antibiotics, and probiotics were significantly associated with the gut microbiome, with probiotics being the most significant factor. Probiotics drove microbiome transition into different PGCTs, each enriched in different *Bifidobacterium* spp. and significantly associated with increased postnatal age. Functional analyses identified stool metabolites associated with PGCT and in preterm-derived organoids, sterile faecal supernatants impacted intestinal organoid monolayer gene expression in a PGCT-specific manner. The current study identifies specific influencers of gut microbiome development in very preterm infants, some of which overlap with those impacting term infants. The results highlight the importance of strain-specific differences in probiotic products, and their impact on host interactions in the preterm gut. The results in this chapter are published.

3.2. Introduction

Host and environmental factors shaping gut microbiome development have been well-defined in term infants (Stewart et al., 2018), but less well defined in significantly preterm infants. In term infants, birth mode (Mitchell et al., 2020; Reyman et al., 2019; Rutayisire et al., 2016; Sordillo et al., 2017) and receipt of breast milk (Azad et al., 2016; Hesla et al., 2014; Sordillo et al., 2017; Stewart et al., 2018) are the main factors influencing the gut microbiome over the first year. Related work in preterm infants has yielded inconsistent results, particularly regarding birth mode and sex, potentially reflecting smaller cohorts and lack of longitudinal sampling (Aguilar-Lopez et al., 2021; Gregory et al., 2016; Stewart et al., 2017a). This ultimately highlights the complexities affiliated with disentangling preterm environmental exposures, clinical factors, and individual variation. Ultimately, these inconsistencies underscore the need for a focused investigation into the factors influencing normal gut microbiome structure and function in preterm infants in the absence of intestinal pathologies such as NEC, or LOS.

Preterm infants born <32 weeks of gestation will initially be cared for on the NICU. This unique setting plays a crucial role in the acquisition and development of the gut microbiome, both directly (i.e., NICU environment) and indirectly (i.e., high antibiotic use). Resultantly, preterm infants experience lower microbial diversity and increased colonisation of pathobionts in comparison to their healthy term counterparts. The gut microbiome of preterm infants has been associated with life-threatening disease including NEC (Olm et al., 2019; Stewart et al., 2016; Torrazza et al., 2013; Warner et al., 2016) and LOS (Stewart et al., 2017b; Taft et al., 2015). This has led to increased interest in and use of probiotics in the NICU, although the efficacy of probiotics in preventing NEC and LOS remains inconclusive (Costeloe et al., 2016) and the potential benefits from probiotic-mediated protection against NEC, LOS or mortality reduction (Sharif et al., 2020; van den Akker et al., 2020) need to be balanced against low but important risks reported in the literature from contamination and probiotic sepsis (Bertelli et al., 2015; Dani et al., 2015; Esaiassen et al., 2016; Jenke et al., 2012). Studies exploring the impact of probiotics on gut microbiome development are few in the preterm population, but have shown that *Bifidobacterium* spp. in particular are able to colonise the gut long-term (Alcon-Giner et al., 2020; van Best et al., 2020; Yousuf et al., 2020).

In the current study of preterm infants in the absence of intestinal disease or LOS, the aims were to 1) characterise the longitudinal development of the preterm gut microbiome throughout their stay on the NICU and 2) determine the influence of co-variates on the developing bacterial community and function during this critical period of early life.

3.3. Results

The current metagenomic analysis included a total of 1431 samples collected longitudinally from 123 very preterm infants born <32 weeks' gestation during their stay in a single United Kingdom NICU (**Figure 3.1**).

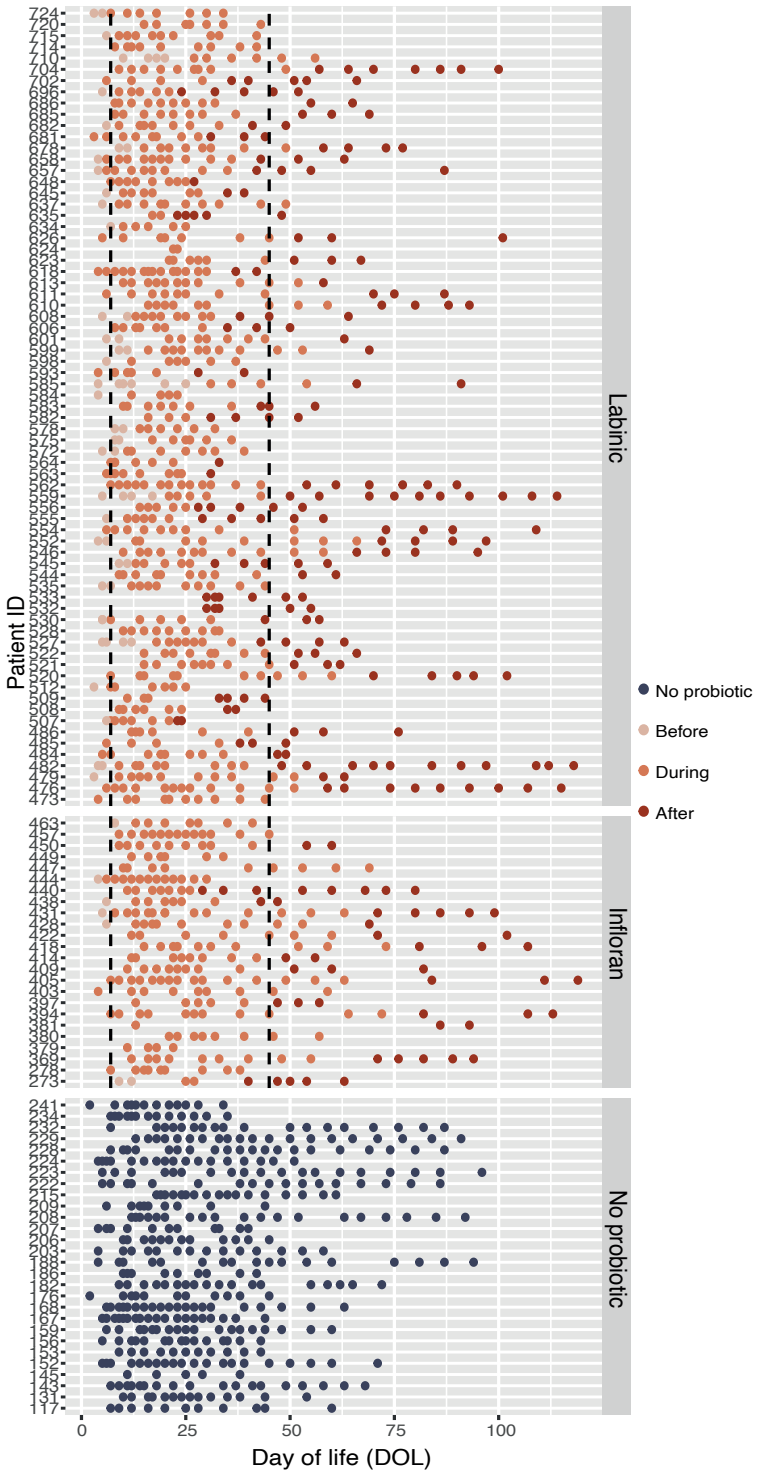

Figure 3.1. Healthy preterm sampling overview. Samples used in the study from birth to day 120. Dashed lines represent the overall mean start and stop day of probiotic treatment.

Samples were collected between birth (DOL 0) and DOL 120, with the median (interquartile range; IQR) DOL for final sample collection occurring on DOL 57 (43 – 77). Infants each contributed a median (IQR) of 11 samples (9-14). Comprehensive demographic information is described in the methods and presented in **Table 3.1**. Most babies received some MOM at some point (92.7%), with receipt of formula increasing with age. All samples had known milk exposure (MOM, formula or both) (**Figure 3.2**) and antibiotic exposure. In order to include infants from before probiotics were introduced, the cohort in this study were admitted over a 10-year period, covering before probiotic introduction, and during two sequentially administered probiotics, as described in methods Section 2.1. Infants born between 2011 and 2013 received no probiotics. Probiotics were then introduced to the NICU in 2013; Infloran (*B. bifidum* $1x10^9$ CFU and *L. acidophilus* $1x10^9$ CFU) was supplemented until mid-2016; after which Labinic (*B. bifidum* 0.67 $x10^9$ CFU*, B. longum* subsp*. infantis* 0.67 $x10^9$ CFU*,* and *L. acidophilus* 0.67 $x10^9$ CFU) has been used.

Table 3.1. Patient demographics of the analytical cohort.

	No probiotics	*Infloran*	*Labinic*	*Overall*
Number of subjects	28	24	71	123
Number of samples	424	253	754	1431
Median no. of samples per subject (IQR)	17 (11.8 - 18.3)	9.5 (8.8 - 13)	10 (9 - 12.5)	11 (9 - 14)
Median gestational age (IQR)	27.2 (25.5 – 28.6)	26.2 (24.8 – 26.9)	27.9 (26 - 29)	27 (25.7 – 28.7)
Median birthweight (g) (IQR)	970 (862 - 1199)	735 (627 - 861)	960 (780 - 1170)	900 (715 - 1138)
Median day of first feed (range)	2 (0 - 11)	2 (1 - 6)	2 (0 - 9)	2 (0 - 11)
Median day of full feed (IQR)	14 (11 - 18)	14 (13 - 19)	13 (12 - 15)	14 (12 - 16)
Median no. of antibiotic courses (IQR)	2 (1 – 3)	3 (2 – 4.25)	2 (1 – 3)	2 (1 – 3)
Median weight Z score change (IQR)	1.5 (2.2 - 0.3)	1 (2 - 0.2)	1.3 (1.8 - 0.75)	1.3 (-2 - 0.55)
MOM ever	25 (89.3%)	23 (95.8%)	66 (93.0%)	114 (92.7%)
Formula ever	15 (62.5%)	42 (59.2%)	19 (67.9%)	76 (61.8%)
Median start day of probiotics (IQR)	/	7 (5 - 8)	7 (5.5 - 9)	7 (5 - 9)
Median stop day of probiotics (IQR)	/	57 (45.8 - 67.8)	41 (28.5 - 50)	45 (31.5 - 55)
Birth mode				
Caesarean	12 (42.9%)	11 (45.8%)	44 (62.0%)	67 (54.5%)
Vaginal	16 (57.1%)	13 (54.2%)	27 (38.0%)	56 (45.5%)
Sex				
Male	20 (71.4%)	8 (33.3%)	40 (56.3%)	68 (55.3%)
Female	8 (28.6%)	16 (66.7%)	31 (43.7%)	55 (44.7%)
Season				
Winter	6 (21.4%)	10 (41.6%)	18 (25.4%)	34 (27.6%)
Autumn	6 (21.4%)	6 (25%)	29 (40.8%)	41 (33.3%)
Summer	10 (35.7.%)	2 (8.3%)	9 (12.7%)	21 (17.1%)
Spring	6 (21.4%)	6 (25%)	15 (21.1%)	27 (22.0%)
Antibiotics in past 7d samples				
No	318 (75%)	176 (69.6%)	549 (72.8%)	1043 (72.9%)
Yes	106 (25%)	77 (30.4%)	205 (27.2%)	388 (27.1%)
MOM samples				
Never	49 (11.6%)	10 (4.0%)	50 (6.6%)	109 (7.6%)
During	318 (75%)	221 (87.4%)	624 (82.8%)	1163 (81.3%)
After	57 (13.4%)	22 (8.7%)	80 (10.6%)	159 (11.1%)
BMF samples				
Never	181 (42.7%)	73 (28.9%)	225 (29.8%)	479 (33.5%)
Before	125 (29.5%)	62 (24.5%)	213 (28.2%)	400 (28.0%)
During	95 (22.4%)	105 (41.5%)	241 (32.0%)	441 (30.8%)
After	23 (5.4%)	13 (5.1%)	75 (9.9%)	111 (7.8%)
Formula samples				
Never	128 (30.2%)	108 (42.7%)	342 (45.4%)	578 (40.4%)
Before	97 (28.9%)	86 (34%)	206 (27.3%)	389 (27.2%)
During	186 (43.9%)	50 (19.8%)	203 (27.0%)	439 (30.7%)
After	13 (3.1%)	9 (3.6%)	3 (0.4%)	25 (1.7%)

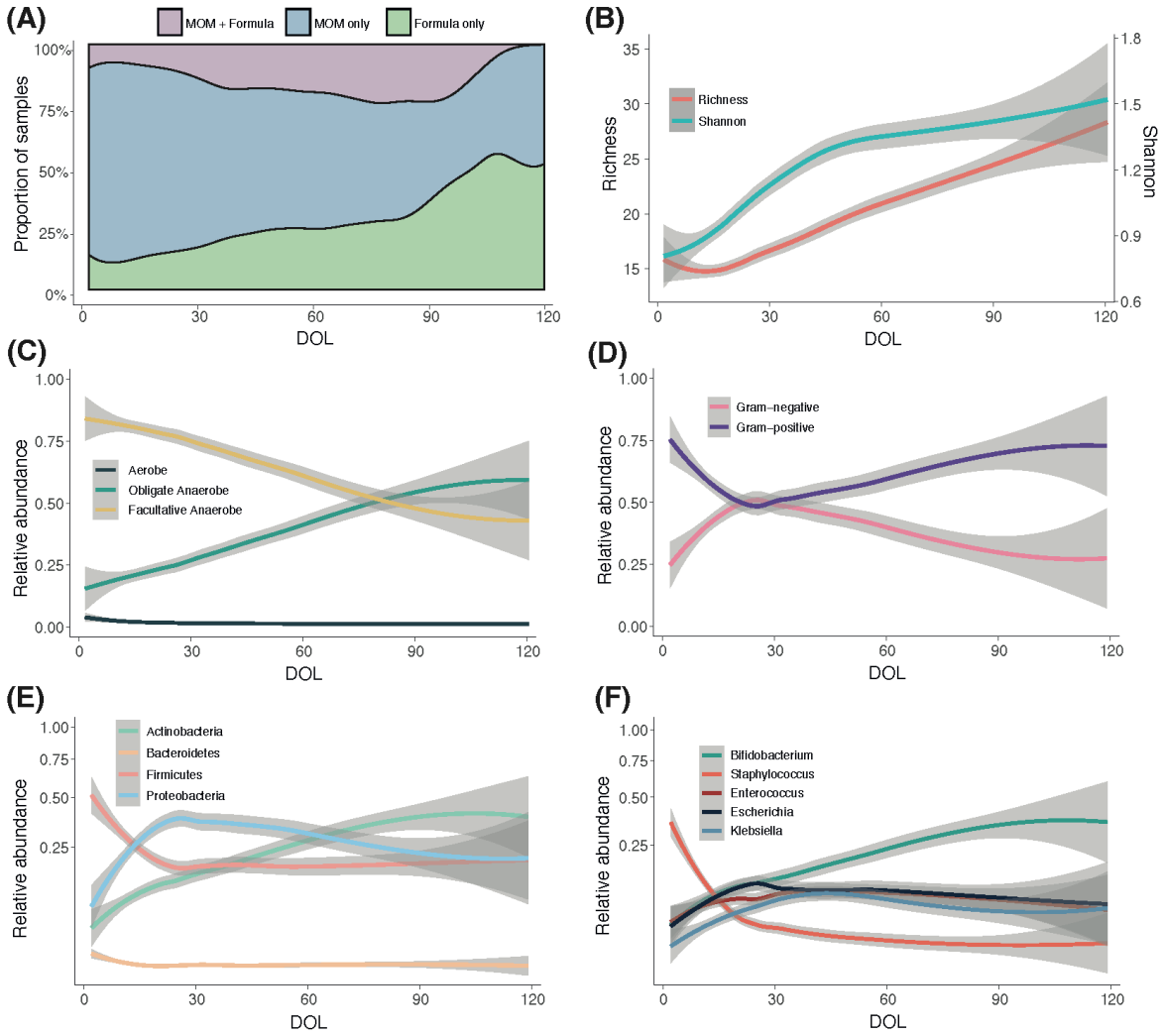

Figure 3.2. Descriptive overview of diet and the healthy preterm gut microbiome in the first 120 days of life. (A) Proportion of samples where infants were receiving MOM, formula or, MOM and formula. **(B-F),** LOESS fit (95% CI shaded in grey) over time for **(B)** richness and Shannon diversity **(C)** aerobic, facultative anaerobic and obligate anaerobic bacteria **(D)** Gram-positive and Gram-negative bacteria **(E)** the top four phyla and **(F)** the top five genera.

3.3.1. Overview of taxonomy

Non-bacterial microbes were explored based on Metagenomic Phylogenetic Analysis (MetaPhlAn; fungi and archaea) and VirMap (virus). No archaea and only 11 fungal species were detected. *Candida albicans* and *Candida glabrata* were the most abundant and prevalent fungi, but only detected in 26 samples (14 infants) and 15 samples (nine infants), respectively. This method allowed detection of DNA viruses, of which only two were detected, *Cytomegalovirus* was found in eight samples from seven infants and *Betapolyomavirus* was detected in two samples from the same infant (data not shown). Importantly, the methods used were tailored towards bacterial profiling (e.g., extraction method and relatively low sequencing depth), which is why

other microorganisms particularly those of a lower biomass (i.e., fungi and archaea) may not have been detected. In total, 394 bacterial species were identified, and thus subsequent analysis was focused on bacteria.

Species richness declined slightly over the first 10 days of life, corresponding to a loss of aerobic bacteria (**Figure 3.2B**). After day 10, species richness increased consistently until day 120 and Shannon diversity increased exponentially from birth until day 45, with a modest increase from day 45 to the end of study (day 120; **Figure 3.2B**). There was a general increase in the relative abundance of obligate anaerobic bacteria from birth until day 80, after which the gut microbiome consisted of approximately 1:1 facultative and obligate anaerobes (**Figure 3.2C**). *Staphylococcus* dominated the earliest samples and accounted for most of the Gram-positive bacteria during the first month of life. Relative abundance of *Bifidobacterium* (Actinobacteria phylum) increased from birth until discharge and from day 30 was the most abundant genera. *Escherichia* and *Klebsiella*, both Gram-negative organisms from the Proteobacteria phylum, increased in relative abundance over the first month of life before gradually declining in relative abundance (**Figure 3.2D, E, F**). Proteobacteria were found to carry significantly more ARGs, whilst the opposite was true of Firmicutes (**Figure 3.3**).

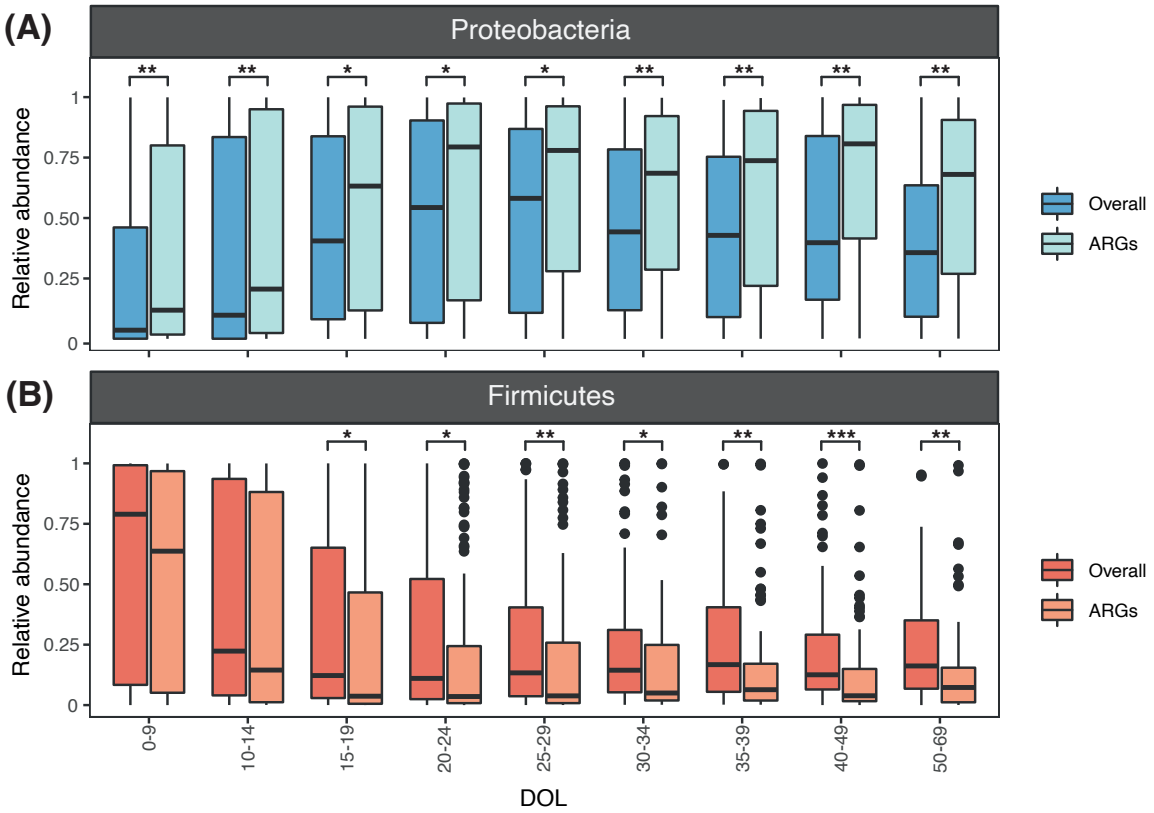

Figure 3.3. Proteobacteria carry the highest number of ARGs. The overall relative abundance of **(A)** ARG-carrying Proteobacteria in relation to other ARG-carrying bacteria and overall Proteobacteria in relation to other phyla, across time and **(B)** ARG-carrying Firmicutes in relation to other ARG-carrying bacteria and overall Firmicutes in relation to other phyla, across time. The centre lines denotes the median, the box limits denote the IQR and whiskers extend to the limits. Points outside the whiskers represent outliers. Only one sample per infant per timepoint is included, *P*-values are based on Mann-Whitney-U tests and are FDR adjusted.

DMM modelling of bacterial species determined five clusters to be optimal, herein termed PGCTs. PGCTs were numbered 1-5 based on the average age of samples within that cluster and richness and Shannon diversity expectedly increased through each PGCT (**Figure 3.4A, B**). *Enterococcus* (*E. faecalis* and *E. faecium*) and *Staphylococcus* (*S. epidermidis* and *S. haemolyticus*) discriminated PGCT-1; *Escherichia* (*E. coli* and an unclassified sp.) discriminated PGCT-2; *Klebsiella* (*K. oxytoca* and an unclassified sp.) discriminated PGCT-3; several *Bifidobacterium* (*B. longum*, *B. bifidum*, and *B. animalis*) and *Lactobacillus* (*L. acidophilus* and *L.*

rhamnosus) discriminated PGCT-4; and a single species, *Bifidobacterium breve,* discriminated PGCT-5 (**Figure 3.4C**).

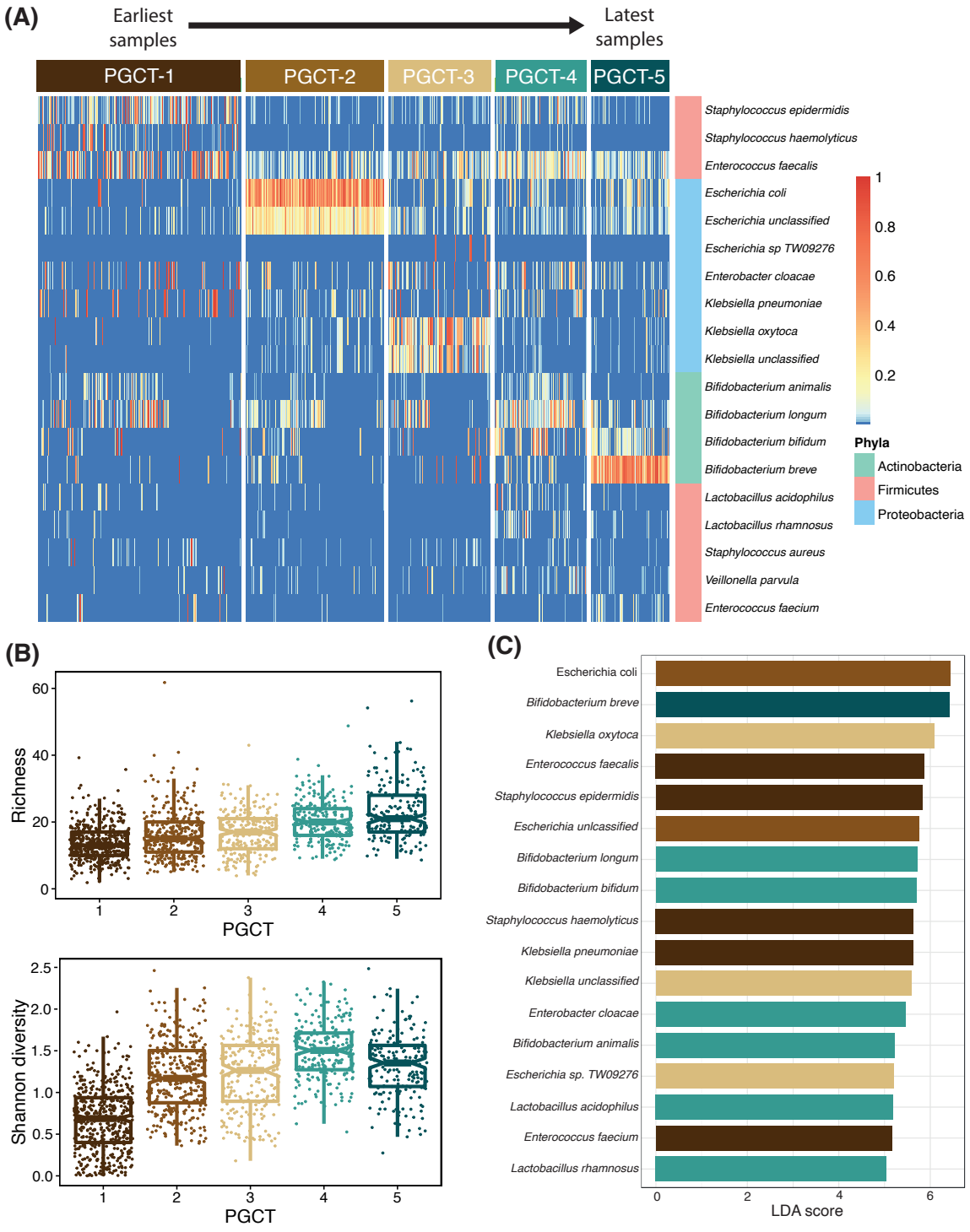

Figure 3.4. DMM clustering into PGCTs. (A) Heatmap of all samples showing the relative abundance of the most dominant species, coloured by phyla, stratified by PGCT. **(B)** Box plots showing the alpha diversity (richness and Shannon diversity) for each PGCT. The centre line denotes the median, the box limits denote the IQR and whiskers extend to the limits. **(C)** LEfSe identifying discriminatory features of each PGCT based on LDA. Coloured bars denote PGCTs.

3.3.2. Factors shaping the preterm gut microbiome

Mixed-effects models were used to assess the association of clinical metadata with Shannon diversity and the total number of ARGs. Shannon diversity was significantly associated with DOL, probiotics (no probiotic/Infloran/Labinic), receipt of MOM (never/during/after), BMF (never/before/during/after), and antibiotics in the past 7 days (no/yes) (**Table 3.2**). The total number of ARGs was significantly associated with DOL, probiotics, MOM, BMF, antibiotics, formula (never/before/during/after) and gestational age (**Table 3.2**). The direction of these effects is described further in later sections. To determine co-variates significantly associated with overall bacterial profiles, univariate PERMANOVA was performed using 'adonis'. DOL explained 4% of the total variance (effect size) in bacterial profiles (P <0.001) and post-conceptional age explained 3.5% (P <0.001), while unique patient identifier explained 1.8% of the variance (P = 0.016).

Table 3.2. Association of clinical co-variates with Shannon diversity and total number of ARGs. Global P-values were calculated based on ANOVA on fitted models.

	Shannon		Total no. ARGs	
	Chisq	**P-value**	**Chisq**	**P-value**
Gestational age	3.156	0.076	5.127	0.024
Birthweight	0.145	0.703	0.39	0.533
Birth mode	2.019	0.155	0.19	0.663
Sex	0.212	0.646	0.739	0.39
Season	2.27	0.518	3.949	0.267
Antibiotics 7d	47.665	<0.001	4.416	0.036
MOM	8.998	0.011	23.486	<0.001
BMF	8.927	0.03	16.852	<0.001
Day full feed	0.006	0.94	0.506	0.477
Formula	0.296	0.961	10.886	0.012
Probiotic	7.047	0.03	6.275	0.043
Weight Z score	0.469	0.494	3.628	0.057
DOL	78.809	<0.001	85.396	<0.001

Antibiotics, MOM, BMF, and probiotics were significantly associated with bacterial taxonomy at one or more time points (**Figure 3.5A**). Probiotics were statistically the most significant (all P < 0.05) and were associated with the bacterial community at all time points, except day 0-9 (P = 0.351) which contained samples collected largely before administration began on day 7 (**Table 3.1**). Complementary analysis on the functional metabolic capacity of the microbiome revealed only probiotics to be significantly associated, at days 10-14, 25-29, 30-34, 35-39 and 50-69 (**Figure 3.5B**).

Notably, gestational age, birthweight, birth mode, formula milk, and sex were not associated with overall bacterial community composition at the taxonomic or functional level.

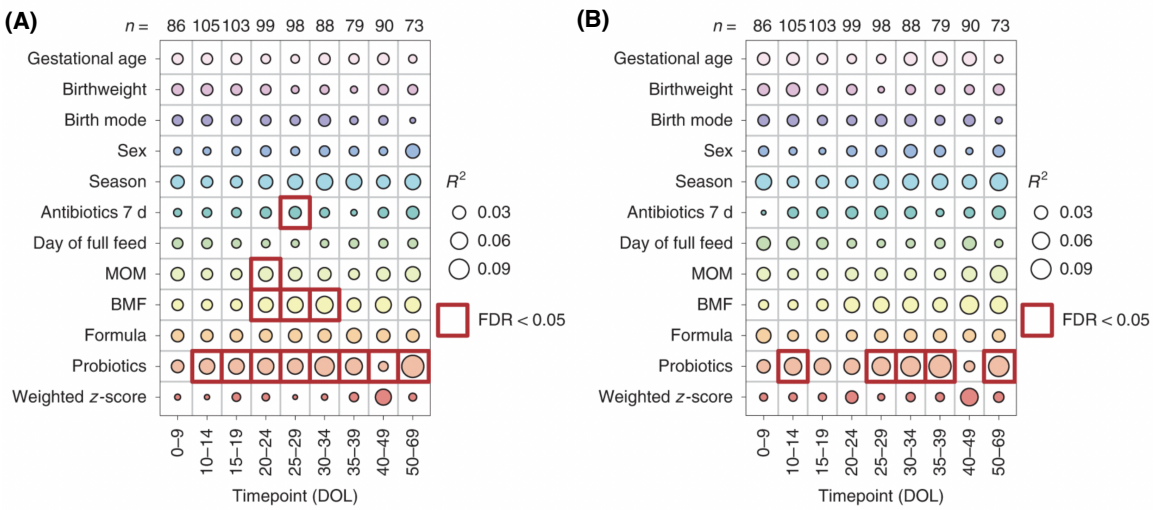

Figure 3.5. Significance and explained variance of 12 clinical co-variates at different timepoints based on taxonomic profiles, modelled by 'adonis'. Bubbles show the amount of variance (R^2) explained by each covariate at a given timepoint, and significant results (FDR < 0.05) are surrounded by a red box. **(A)** Taxonomic profiles at the species level. **(B)** Functional metagenomic capacity at the enzyme level using EC numbers.

To further validate these results, a previously published metagenomic study by Olm *et al*. (Olm et al., 2019) containing 86 control preterm infants (n = 513 stool samples), not receiving probiotics, was used. While feeding information was less granular than in the current study, the results were generally consistent between cohorts with no significant association of any tested co-variate on the gut microbiome (**Figure 3.6**).

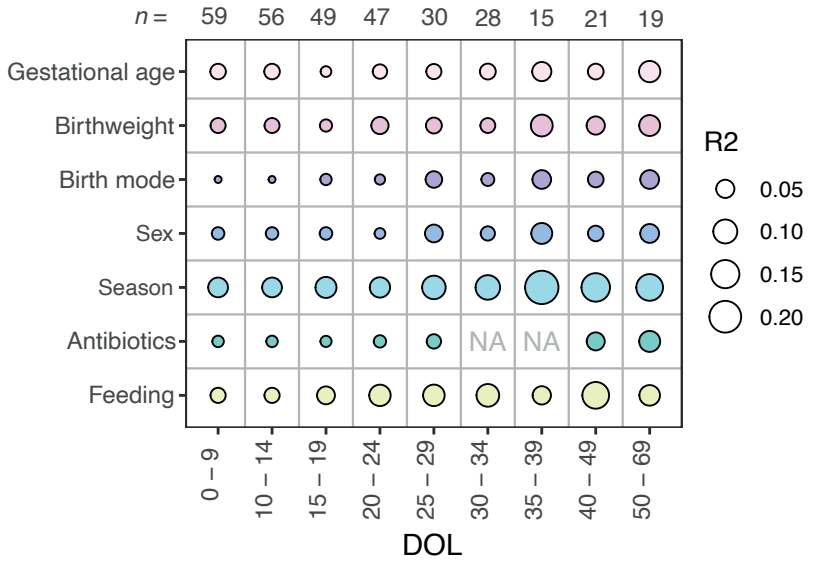

Figure 3.6. Explained variance of 7 clinical co-variates at different timepoints to validate the findings in this study, using a published metagenomic dataset from Olm et al. (Olm et al., 2019)**, modelled by 'adonis'.** Bubbles show the amount of variance (R^2) explained by each covariate at a given timepoint. NA values are used when analyses could not be carried out, due to only one level of the variable existing in that given timepoint. No results were found to be significant based on taxonomic profiles at the species level.

3.3.3. Role of probiotics in shaping the gut community

Binomial mixed-effects models showed that infants who did not receive probiotics were significantly more likely to transition into the *Klebsiella*-enriched PGCT-3 (P = 0.021), which was also associated with a lower gestational age at birth (P = 0.043; **Figure 3.7A** and **Table 3.3**). Infants receiving Infloran were significantly more likely to transition into PGCT-5 and those receiving Labinic to PGCT-4 (both P < 0.001; **Figure 3.7A** and **Table 3.3**). Samples from PGCT-4 and PGCT-5 were from significantly higher DOL (both P < 0.001; **Table 3.3**) and thus reflected the oldest infants. PGCT-5 was dominated by *B. breve* and associated with a higher gestational age (P = 0.008; **Table 3.3.**). PGCT-4 was generally dominated by the species present in Labinic™ including *B. longum*, *B. bifidum* and *L. acidophilus*, but also *B. animalis* (**Figure 3.4**).

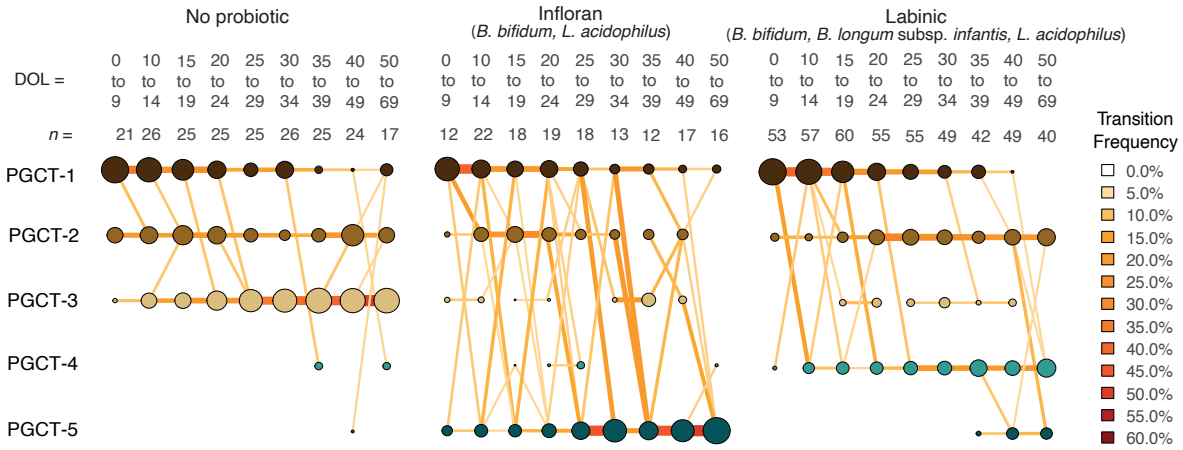

Figure 3.7. Probiotics impact the transition of the preterm gut microbiome over time. Transition model showing the progression of samples through each PGCT from DOL 0 to DOL 69, based on probiotic type. The nodes and edges are sized based on the total counts; nodes are coloured according to PGCT and edges by the transition frequency.

Table 3.3. Association between PGCTs and clinical co-variates in healthy preterm infants. Global *P* values and adjusted odds ratios (aORs) with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect.

	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	0.724 (0.49 - 1.07)	0.105	0.993 (0.57 - 1.73)	0.979	0.598 (0.363 - 0.985)	0.043	1.24 (0.877 - 1.752)	0.223	2.126 (1.216 - 3.716)	0.008
Birthweight	0.998 (0.996 - 1)	0.113	1.002 (0.998 - 1.005)	0.369	1.004 (1.001 - 1.007)	0.012	0.998 (0.996 - 1)	0.035	1.001 (0.997 - 1.004)	0.655
Birth mode		0.109		0.327		0.103		0.343		0.658
Vaginal	1		1		1		1		1	
Caesarean	2.168 (0.842 - 5.577)		0.508 (0.131 - 1.97)		2.672 (0.821 - 8.699)		0.664 (0.284 - 1.549)		0.715 (0.162 - 3.151)	
Sex		0.361		0.401		0.601		0.233		0.006
Male	1		1		1		1		1	
Female	0.667 (0.28 - 1.589)		1.734 (0.48 - 6.262)		0.752 (0.259 - 2.186)		0.62 (0.283 - 1.361)		6.664 (1.727 - 25.712)	
Season										
Winter	1	0.884	1	0.303	1	0.146	1	0.941	1	0.020
Spring	0.857 (0.257 - 2.855)		1.16 (0.204 - 6.591)		1.521 (0.334 - 6.932)		0.835 (0.282 - 2.475)		16.066 (2.195 - 117.6)	
Summer	0.956 (0.258 - 3.54)		0.283 (0.041 - 1.964)		6.325 (1.199 - 33.355)		1.086 (0.331 - 3.56)		1.812 (0.211 - 15.581)	
Autumn	1.334 (0.459 - 3.877)		0.34 (0.069 - 1.684)		2.702 (0.673 - 10.855)		0.796 (0.31 - 2.041)		9.919 (1.627 - 60.47)	
Antibiotics 7d		<0.001		0.138		0.186		<0.001		0.098
No	1		1		1		1		1	
Yes	2.418 (1.582 - 3.696)		0.692 (0.425 - 1.126)		0.701 (0.414 - 1.186)		0.298 (0.163 - 0.544)		0.509 (0.229 - 1.132)	
Day full feed	1.039 (0.95 - 1.135)	0.404	1 (0.88 - 1.137)	0.997	0.945 (0.847 - 1.054)	0.311	0.97 (0.884 - 1.064)	0.522	1.005 (0.882 - 1.145)	0.940
MOM		0.071		0.0855		0.322		0.076		0.290
Never	1		1		1		1		1	
During	0.352 (0.061 - 2.041)		3.693 (0.228 - 59.795)		6.319 (0.571 - 69.884)		0.313 (0.069 - 1.426)		4.738 (0.27 - 83.005)	
After	0.124 (0.018 - 0.842)		1.278 (0.073 - 22.254)		5.381 (0.443 - 65.4)		0.86 (0.184 - 4.012)		8.746 (0.466 - 164.168)	
BMF		0.001		0.004		0.010		0.240		0.315
Never	1		1		1		1		1	
Before	0.561 (0.188 - 1.668)		4.996 (0.991 - 25.188)		0.596 (0.151 - 2.347)		0.455 (0.152 - 1.362)		1.458 (0.25 - 8.523)	
During	0.249 (0.082 - 0.763)		7.449 (1.504 - 36.91)		1.337 (0.353 - 5.067)		0.526 (0.193 - 1.434)		2.531 (0.479 - 13.385)	
After	1.159 (0.268 - 5.006)		1.894 (0.328 - 10.93)		0.407 (0.082 - 2.016)		0.314 (0.095 - 1.041)		4.308 (0.699 - 26.562)	
Formula		0.398		0.677		0.575		0.206		0.252
Never	1		1		1		1		1	
Before	0.709 (0.283 - 1.777)		1.282 (0.334 - 4.921)		0.509 (0.159 - 1.625)		1.479 (0.613 - 3.568)		3.955 (0.848 - 18.45)	
During	0.554 (0.195 - 1.58)		1.558 (0.39 - 6.226)		0.737 (0.224 - 2.428)		0.748 (0.29 - 1.935)		4.716 (1.03 - 21.598)	
After	0.118 (0.009 - 1.638)		0.635 (0.085 - 4.716)		1.377 (0.127 - 14.892)		6.263 (0.475 - 82.594)		3.864 (0.138 - 108.06)	
Probiotics		0.143		0.666		0.021		<0.001		<0.001
No probiotic	1		1		1		1		1	
Infloran	0.473 (0.122 - 1.842)		0.51 (0.072 - 3.629)		0.142 (0.027 - 0.751)		0.614 (0.15 - 2.518)		94.84 (12.865 - 699.2)	
Labinic	1.522 (0.525 - 4.407)		0.493 (0.102 - 2.378)		0.19 (0.052 - 0.694)		10.32 (3.694 - 28.84)		0.485 (0.093 - 2.532)	
Weight Z score	0.94 (0.63 - 1.401)	0.761	1.062 (0.611 - 1.845)	0.832	1.022 (0.629 - 1.658)	0.931	0.893 (0.633 - 1.258)	0.517	1.506 (0.841 - 2.696)	0.169
DOL	0.927 (0.912 - 0.943)	<0.001	1.009 (0.995 - 1.023)	0.203	1.008 (0.993 - 1.023)	0.290	1.031 (1.018 - 1.045)	<0.001	1.055 (1.035 - 1.076)	<0.001

MaAsLin2 analysis confirmed the relative abundance of genera (**Table 3.4**) and species (**Table 3.5**) present in each probiotic were significantly higher in infants receiving that probiotic. Notably, *B. breve* was significantly associated with Infloran ($P < 0.001$, $Q = 0.007$) and *B. animalis* was the most significant taxa associated with Labinic ($P < 0.001$, $Q < 0.001$), despite these species not being named as present in the probiotics (**Table 3.5**). Using culture-based approaches, *B. breve* could not be cultured from Infloran but *B. animalis* was consistently cultured from Labinic. Given these results, *B. animalis* is considered to be present in Labinic in subsequent analysis. Aside from probiotic species, the influence of probiotics on other naturally occurring taxa showed a significant increase in the relative abundance of *Enterococcus faecium* ($P < 0.001$, $Q = 0.004$), and a significant decrease in the relative abundance of *Veillonella parvula* ($P < 0.001$, $Q = 0.022$) and *Propionibacterium acnes* (now known as *Cutibacterium acnes*) ($P = 0.001$, $Q = 0.030$) in infants supplemented with Infloran (**Table 3.5**). No non-probiotic species were significantly increased or decreased in infants supplemented with Labinic (all $Q > 0.05$; **Table 3.5**).

Table 3.4. MaAsLin2 results for significant taxa associated with clinical covariates at the genus level in healthy preterm infants. Mixed-effects linear models using a variance-stabilising arcsin square root transformation on relative abundance phyla data were used to determine the significance. Patient ID was included as a random effect.

	Level	Genus	Coeff	Standard error	pval	qval
Gestational age	Gestational age	*Bifidobacterium*	0.093	0.035	0.009	0.12
Birthweight	Birthweight	*Citrobacter*	0.031	0.011	0.005	0.079
Birth mode	Vaginal	*Enterobacter*	-0.089	0.036	0.014	0.161
Season	Autumn	*Klebsiella*	0.156	0.058	0.008	0.112
Antibiotics 7d	Yes	*Bifidobacterium*	-0.153	0.022	<0.001	<0.001
Antibiotics 7d	Yes	*Staphylococcus*	0.119	0.022	<0.001	<0.001
Antibiotics 7d	Yes	*Enterobacter*	-0.045	0.014	0.001	0.026
Antibiotics 7d	Yes	*Propionibacterium*	-0.013	0.005	0.019	0.178
Antibiotics 7d	Yes	*Lactobacillus*	-0.017	0.007	0.023	0.204
MOM	After	*Staphylococcus*	-0.201	0.066	0.003	0.053
MOM	After	*Bifidobacterium*	0.238	0.086	0.006	0.093
MOM	During	*Bifidobacterium*	0.209	0.083	0.013	0.154
MOM	During	*Staphylococcus*	-0.157	0.063	0.013	0.157
MOM	During	*Lactobacillus*	-0.082	0.034	0.017	0.169
BMF	After	*Clostridium*	0.039	0.009	<0.001	0.001
BMF	Before	*Lactobacillus*	-0.067	0.021	0.002	0.04
BMF	During	*Lactobacillus*	-0.062	0.021	0.004	0.061
BMF	Before	*Enterococcus*	-0.15	0.053	0.005	0.082
BMF	After	*Veillonella*	0.043	0.016	0.01	0.12
BMF	During	*Corynebacterium*	0.016	0.007	0.017	0.165
BMF	After	*Anaerococcus*	-0.014	0.006	0.016	0.165
BMF	During	*Escherichia*	0.171	0.076	0.025	0.216
Formula	During	*Staphylococcus*	-0.109	0.038	0.005	0.073
Formula	During	*Actinomyces*	0.014	0.006	0.019	0.178
Probiotic	Infloran	*Bifidobacterium*	0.414	0.064	<0.001	<0.001
Probiotic	Labinic	*Bifidobacterium*	0.297	0.051	<0.001	<0.001
Probiotic	Infloran	*Propionibacterium*	-0.053	0.012	<0.001	0.001
Probiotic	Labinic	*Lactobacillus*	0.086	0.021	<0.001	0.003
Probiotic	Labinic	*Pseudomonas*	-0.046	0.012	<0.001	0.011
Probiotic	Infloran	*Veillonella*	-0.051	0.016	0.001	0.036
Probiotic	Infloran	*Lactobacillus*	0.085	0.027	0.002	0.04
Probiotic	Infloran	*Pseudomonas*	-0.048	0.016	0.003	0.054
Probiotic	Labinic	*Propionibacterium*	-0.029	0.009	0.003	0.057
Probiotic	Labinic	*Anaerococcus*	0.01	0.004	0.007	0.093
Probiotic	Labinic	*Klebsiella*	-0.158	0.06	0.01	0.12
Probiotic	Infloran	*Staphylococcus*	-0.118	0.047	0.015	0.161
Probiotic	Infloran	*Klebsiella*	-0.171	0.076	0.027	0.225
Weight Z score	Weight Z score	*Pseudomonas*	-0.013	0.005	0.016	0.165
Weight Z score	Weight Z score	*Veillonella*	0.012	0.005	0.022	0.199
Weight Z score	Weight Z score	*Citrobacter*	0.016	0.007	0.024	0.21
DOL	DOL	*Staphylococcus*	-0.098	0.013	<0.001	<0.001
DOL	DOL	*Finegoldia*	0.011	0.001	<0.001	<0.001
DOL	DOL	*Bifidobacterium*	0.084	0.013	<0.001	<0.001
DOL	DOL	*Propionibacterium*	0.019	0.003	<0.001	<0.001
DOL	DOL	*Veillonella*	0.021	0.004	<0.001	<0.001
DOL	DOL	*Clostridium*	0.01	0.002	<0.001	<0.001
DOL	DOL	*Klebsiella*	0.046	0.012	<0.001	0.008
DOL	DOL	*Anaerococcus*	0.005	0.002	0.003	0.053
DOL	DOL	*Streptococcus*	0.012	0.004	0.003	0.053
DOL	DOL	*Actinomyces*	0.005	0.002	0.016	0.165

Table 3.5. MaAsLin2 results for significant taxa associated with clinical covariates at the species level in healthy preterm infants. Mixed-effects linear models using a variance-stabilising arcsin square root transformation on relative abundance phyla data were used to determine the significance. Patient ID was included as a random effect.

	Level	Species	Coeff	Standard error	pval	qval
Gestational age	Gestational age	*Staphylococcus epidermidis*	-0.048	0.018	0.008	0.104
Gestational age	Gestational age	*Clostridium perfringens*	0.009	0.004	0.023	0.205
Gestational age	Gestational age	*Bifidobacterium bifidum*	0.053	0.024	0.026	0.219
Birthweight	Birthweight	*Citrobacter freundii*	0.022	0.007	0.002	0.046
Birthweight	Birthweight	*Citrobacter unclassified*	0.015	0.006	0.008	0.104
Birthweight	Birthweight	*Escherichia unclassified*	0.036	0.016	0.023	0.208
Birthweight	Birthweight	*Klebsiella unclassified*	0.028	0.012	0.025	0.214
Birthweight	Birthweight	*Propionibacterium acnes*	-0.009	0.004	0.026	0.219
Birth mode	Vaginal	*Bifidobacterium animalis*	0.032	0.011	0.006	0.095
Birth mode	Vaginal	*Citrobacter freundii*	-0.025	0.01	0.011	0.122
Birth mode	Vaginal	*Enterobacter cloacae*	-0.089	0.036	0.014	0.143
Birth mode	Vaginal	*Citrobacter unclassified*	-0.017	0.008	0.031	0.243
Sex	Male	*Bifidobacterium breve*	-0.085	0.039	0.032	0.243
Season	Summer	*Staphylococcus aureus*	0.048	0.016	0.004	0.073
Season	Summer	*Citrobacter freundii*	0.036	0.013	0.008	0.104
Season	Summer	*Citrobacter unclassified*	0.029	0.011	0.009	0.105
Season	Autumn	*Klebsiella unclassified*	0.044	0.018	0.019	0.179
Antibiotics 7d	Yes	*Staphylococcus haemolyticus*	0.107	0.013	<0.001	<0.001
Antibiotics 7d	Yes	*Bifidobacterium bifidum*	-0.056	0.012	<0.001	<0.001
Antibiotics 7d	Yes	*Bifidobacterium longum*	-0.063	0.016	<0.001	0.004
Antibiotics 7d	Yes	*Enterobacter cloacae*	-0.045	0.014	0.001	0.026
Antibiotics 7d	Yes	*Pseudomonas aeruginosa*	0.02	0.006	0.002	0.043
Antibiotics 7d	Yes	*Bifidobacterium breve*	-0.046	0.016	0.004	0.067
Antibiotics 7d	Yes	*Enterococcus faecium*	0.022	0.008	0.007	0.097
Antibiotics 7d	Yes	*Staphylococcus warneri*	0.013	0.005	0.009	0.104
Antibiotics 7d	Yes	*Staphylococcus lugdunensis*	0.007	0.003	0.009	0.105
Antibiotics 7d	Yes	*Lactobacillus acidophilus*	-0.014	0.005	0.011	0.125
Antibiotics 7d	Yes	*Corynebacterium tuberculostearicum*	0.002	0.001	0.013	0.137
Day full feed	Day full feed	*Propionibacterium acnes*	0.008	0.003	0.003	0.058
MOM	After	*Lactobacillus rhamnosus*	-0.063	0.017	<0.001	0.009
MOM	After	*Staphylococcus aureus*	-0.079	0.023	0.001	0.026
MOM	During	*Staphylococcus aureus*	-0.074	0.022	0.001	0.03
MOM	During	*Lactobacillus rhamnosus*	-0.05	0.016	0.003	0.058
MOM	During	*Lactobacillus acidophilus*	-0.069	0.023	0.004	0.068
MOM	During	*Staphylococcus haemolyticus*	-0.096	0.036	0.009	0.104
MOM	After	*Klebsiella pneumoniae*	0.187	0.078	0.018	0.174
MOM	During	*Klebsiella pneumoniae*	0.18	0.077	0.021	0.189
BMF	After	*Clostridium perfringens*	0.024	0.008	0.002	0.042
BMF	During	*Corynebacterium kroppenstedtii*	0.021	0.007	0.003	0.058
BMF	During	*Escherichia unclassified*	0.074	0.025	0.004	0.067
BMF	Before	*Enterococcus faecium*	-0.051	0.019	0.007	0.097
BMF	Before	*Staphylococcus haemolyticus*	0.065	0.024	0.007	0.097
BMF	Before	*Lactobacillus acidophilus*	-0.039	0.015	0.009	0.107
BMF	During	*Enterococcus faecium*	-0.046	0.018	0.013	0.137
BMF	After	*Lactobacillus acidophilus*	-0.042	0.017	0.013	0.137
BMF	During	*Bifidobacterium bifidum*	-0.086	0.035	0.014	0.144
BMF	Before	*Bifidobacterium bifidum*	-0.085	0.035	0.016	0.159
BMF	During	*Lactobacillus rhamnosus*	-0.023	0.01	0.025	0.215
BMF	After	*Veillonella parvula*	0.029	0.013	0.029	0.232
BMF	Before	*Bifidobacterium animalis*	-0.03	0.014	0.032	0.243
Formula	After	*Pseudomonas aeruginosa*	-0.102	0.026	<0.001	0.004
Formula	Before	*Staphylococcus epidermidis*	-0.077	0.023	0.001	0.03
Formula	During	*Staphylococcus epidermidis*	-0.08	0.026	0.003	0.058
Formula	During	*Bifidobacterium breve*	0.125	0.043	0.004	0.073

Formula	During	*Lactobacillus acidophilus*	0.031	0.013	0.022	0.197
Formula	During	*Bifidobacterium longum*	-0.068	0.03	0.027	0.219
Formula	During	*Klebsiella pneumoniae*	0.093	0.042	0.03	0.238
Probiotic	Labinic	*Bifidobacterium animalis*	0.115	0.013	<0.001	<0.001
Probiotic	Labinic	*Bifidobacterium longum*	0.258	0.031	<0.001	<0.001
Probiotic	Infloran	*Bifidobacterium bifidum*	0.222	0.044	<0.001	<0.001
Probiotic	Labinic	*Lactobacillus acidophilus*	0.059	0.015	<0.001	0.004
Probiotic	Infloran	*Enterococcus faecium*	0.091	0.023	<0.001	0.004
Probiotic	Infloran	*Bifidobacterium breve*	0.236	0.061	<0.001	0.007
Probiotic	Infloran	*Veillonella parvula*	-0.044	0.013	0.001	0.022
Probiotic	Infloran	*Propionibacterium acnes*	-0.025	0.008	0.001	0.03
Probiotic	Infloran	*Lactobacillus acidophilus*	0.055	0.018	0.003	0.064
Probiotic	Infloran	*Staphylococcus epidermidis*	-0.089	0.032	0.007	0.097
Probiotic	Labinic	*Pseudomonas aeruginosa*	-0.032	0.012	0.008	0.104
Probiotic	Labinic	*Lactobacillus rhamnosus*	0.026	0.01	0.011	0.124
Probiotic	Labinic	*Propionibacterium acnes*	-0.015	0.006	0.012	0.132
Probiotic	Infloran	*Propionibacterium avidum*	-0.026	0.01	0.013	0.137
Probiotic	Labinic	*Bifidobacterium bifidum*	0.084	0.035	0.017	0.168
Probiotic	Infloran	*Veillonella unclassified*	-0.015	0.006	0.018	0.174
Probiotic	Infloran	*Bifidobacterium longum*	0.09	0.039	0.025	0.216
Weight Z score	Weight Z score	*Citrobacter unclassified*	0.008	0.004	0.031	0.242
DOL	DOL	*Staphylococcus epidermidis*	-0.079	0.01	<0.001	<0.001
DOL	DOL	*Finegoldia magna*	0.011	0.001	<0.001	<0.001
DOL	DOL	*Propionibacterium avidum*	0.018	0.002	<0.001	<0.001
DOL	DOL	*Veillonella unclassified*	0.013	0.002	<0.001	<0.001
DOL	DOL	*Streptococcus salivarius*	0.009	0.001	<0.001	<0.001
DOL	DOL	*Bifidobacterium breve*	0.057	0.01	<0.001	<0.001
DOL	DOL	*Lactobacillus rhamnosus*	0.013	0.002	<0.001	<0.001
DOL	DOL	*Lactobacillus acidophilus*	-0.015	0.003	<0.001	<0.001
DOL	DOL	*Streptococcus vestibularis*	0.009	0.002	<0.001	<0.001
DOL	DOL	*Bifidobacterium animalis*	-0.018	0.004	<0.001	0.001
DOL	DOL	*Bifidobacterium bifidum*	0.027	0.007	<0.001	0.009
DOL	DOL	*Veillonella atypica*	0.005	0.002	0.001	0.022
DOL	DOL	*Klebsiella unclassified*	0.011	0.004	0.004	0.067
DOL	DOL	*Staphylococcus warneri*	-0.008	0.003	0.004	0.073
DOL	DOL	*Bifidobacterium longum*	0.027	0.01	0.005	0.089
DOL	DOL	*Veillonella parvula*	0.009	0.003	0.006	0.09
DOL	DOL	*Staphylococcus haemolyticus*	-0.023	0.008	0.006	0.095
DOL	DOL	*Klebsiella pneumoniae*	0.024	0.009	0.007	0.103
DOL	DOL	*Escherichia coli*	0.029	0.011	0.008	0.104
DOL	DOL	*Klebsiella oxytoca*	0.02	0.008	0.008	0.104
DOL	DOL	*Clostridium perfringens*	0.004	0.002	0.028	0.23

Strain-level analyses to detect the presence of *B. longum* subsp. *infantis* and differentiate from non-probiotic *B. longum* strains was conducted using the *B. infantis* HMO gene clusters (H1, H2, H3, H4, H5 and urease), whereby samples with >90% of the genes present in those clusters were classed as having *B. infantis* (Casaburi et al., 2021). *B. infantis* was detected in 672 samples, of which 666 (>99%) were from infants receiving Labinic™. Additional analysis on the *B. infantis* HMO gene clusters identified homologs present in *B. breve, B. bifidum* and *B. pseudocatenulatum* (**Figure 3.8**).

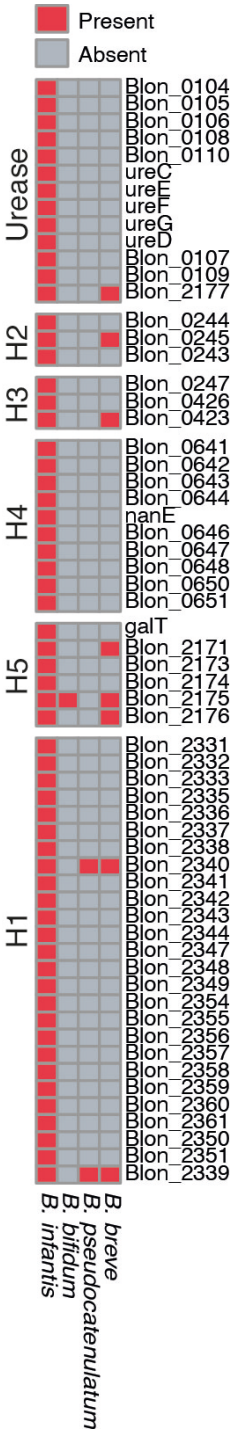

Figure 3.8. Prevalence of the *B. infantis* HMO gene clusters among other species.

Bifidobacterium spp. were also present naturally within the population, with *B. breve*, *B. dentium* and *B. longum* identified in infants who never received probiotics, and before probiotics were ever used on the unit (i.e., no possibility of cross-colonisation) (**Figure 3.9**).

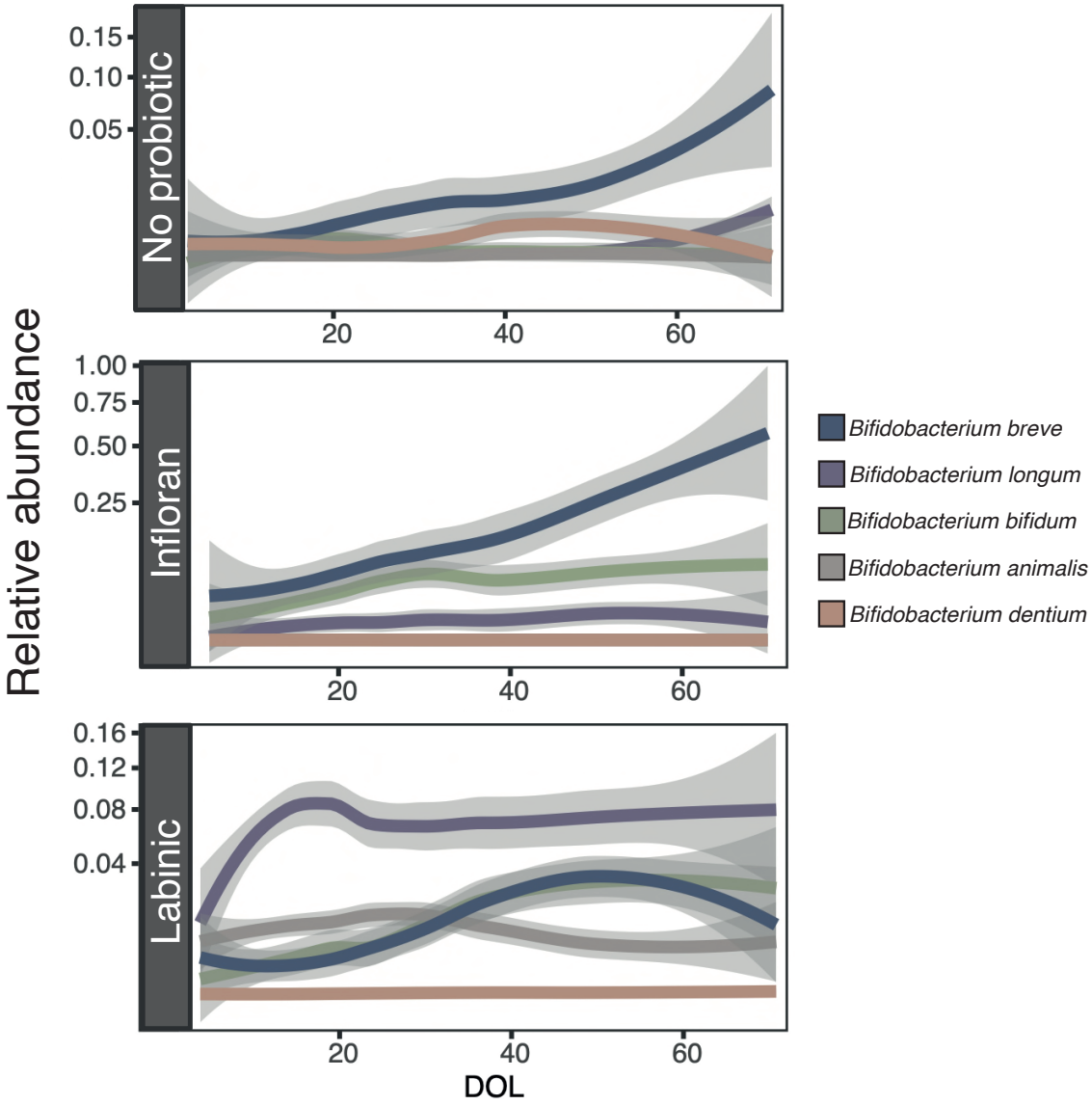

Figure 3.9. Relative abundance of Bifidobacterium species over time. LOESS fit (95% CI shaded in grey) over time for the top 5 most dominant *Bifidobacterium* spp.

The impact of the different probiotics is further demonstrated by significantly higher Shannon diversity and in infants receiving Labinic compared to infants receiving no probiotic (P = 0.035) (**Figure 3.10A**). Furthermore, the total number of ARGs was found to be significantly lower in infants receiving Labinic compared to infants receiving no probiotic (P = 0.033) (**Figure 3.10B**).

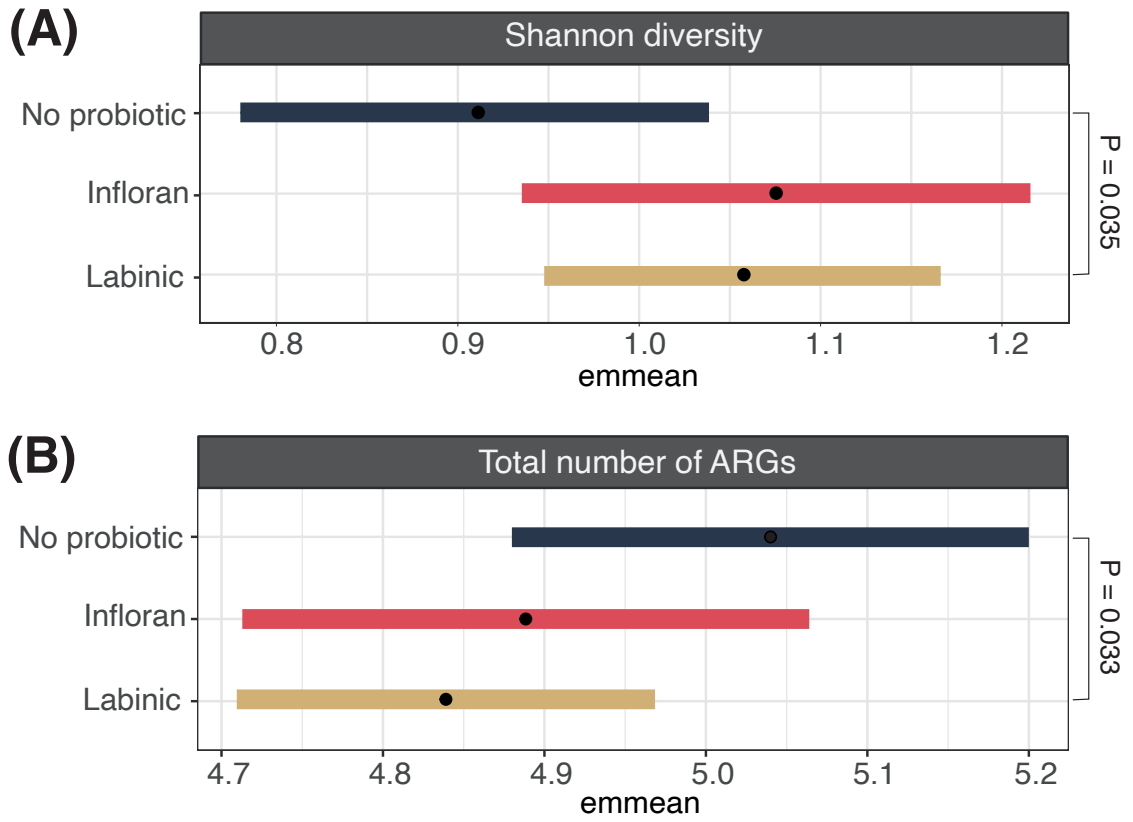

Figure 3.10. Shannon diversity and total number of ARGs differs dependent on probiotic use. Estimated marginal means (95% CIs) representing **(A)** Shannon diversity and **(B)** total number of ARGs for each probiotic type, obtained from the respective LMMs adjusted for gestational age, birthweight, birth mode, sex, season, antibiotics, day of full feed, MOM, BMF, formula, weight z-scores, DOL and patient ID. The statistical significance shown is after adjustment for multiple comparisons using two-tailed Tukey's HSD method.

The impact of clinical data on prevalence (defined as a binary yes/no) and persistence (defined as two consecutive samples where the corresponding species was detected) of the species contained within each probiotic (see Methods) was next explored. Compared to DOL-matched infants who never received probiotics, the probiotic species were significantly more prevalent before, during, and after administration of the respective probiotic (**Figure 3.11**). Comparing between the probiotic groups, the prevalence of *B. bifidum* was significantly higher in Infloran® compared to Labinic during ($P < 0.001$) and after ($P < 0.001$), but *L. acidophilus* prevalence was comparable before, during, and after (all $P > 0.05$). While not present in Infloran®, *B. longum* was significantly more prevalent compared to DOL-matched infants who received no probiotic, during Infloran treatment ($P < 0.001$), and *B. animalis* prevalence was highly specific to Labinic exposure (**Figure 3.11**).

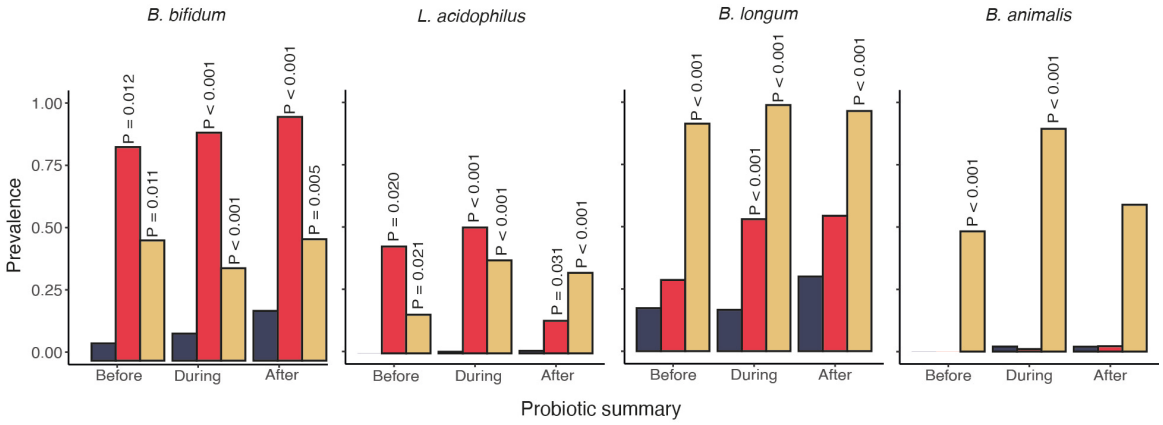

Figure 3.11. Prevalence of probiotic species before, during and after probiotic treatment, stratified by probiotic type. Samples from infants who took no probiotic have been subset into three discrete time bins based on the average start and stop days for probiotic treatment (8 DOL and 44 DOL, respectively). The statistical significance shown is within probiotic summary groups (that is, before, during and after) following adjustment for multiple comparisons using Dunnett's method, whereby samples from infants who took no probiotic were used as the control for each group. Navy represents no probiotic, red represents Infloran and gold represents Labinic.

Analysing persistence of *B. bifidum* and *L. acidophilus* after treatment showed further strain-specific differences between the probiotics, with no clinical covariates other than probiotic type being significantly associated with the persistence of either species (P = 0.001 and P = 0.019, respectively). Analyses also showed increased persistence of *Bifidobacterium* spp. compared to *L. acidophilus. L. acidophilus* did not persist in any infant receiving Infloran and only in a minority of those receiving Labinic, whereas *B. bifidum* persisted in all infants receiving Infloran with non-persistence observed in infants receiving Labinic only (**Figure 3.12**).

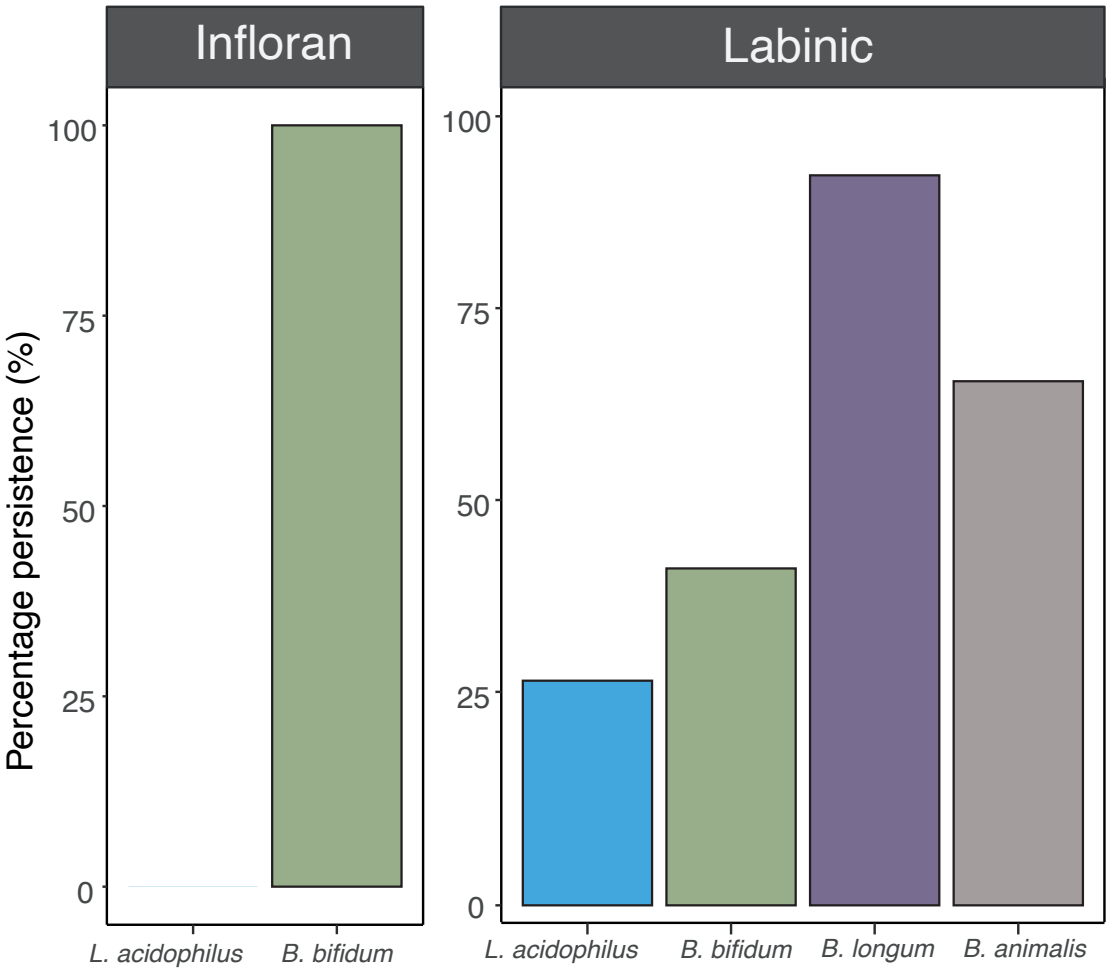

Figure 3.12. Percentage persistence of probiotic species in Infloran and Labinic. *B. animalis* included owing to detection of this species within Labinic.

In accordance with taxonomic profiles, probiotics were found to be the only covariate significantly associated with the overall functional EC profile at any timepoint (**Figure 3.5**). However, unlike taxonomic composition where probiotic groups were more dissimilar to each other than the no probiotic group (**Figure 3.13A**), functional profiles for infants who took any probiotic were more similar than infants who never took probiotics, regardless of which product was received (**Figure 3.13B**). This suggests similar functional potential regardless of *Bifidobacterium*-dominated communities (i.e., functional redundancy between *Bifidobacterium* spp.). MaAslin2 analysis corroborated these findings, with 346/754 (46%) significant EC numbers found to be commonly associated with both probiotic products (https://github.com/laurencbeck/supplementary_tables - **Appendix 1**). Amongst the significantly positively associated EC numbers, a large number of glycosylases and ligases involved in forming carbon-oxygen and carbon-nitrogen bonds were identified.

In contrast, the relative abundance of numerous oxidoreductases acting on a sulphur group and other nitrogenous compounds as donors were found to be negatively associated with probiotics.

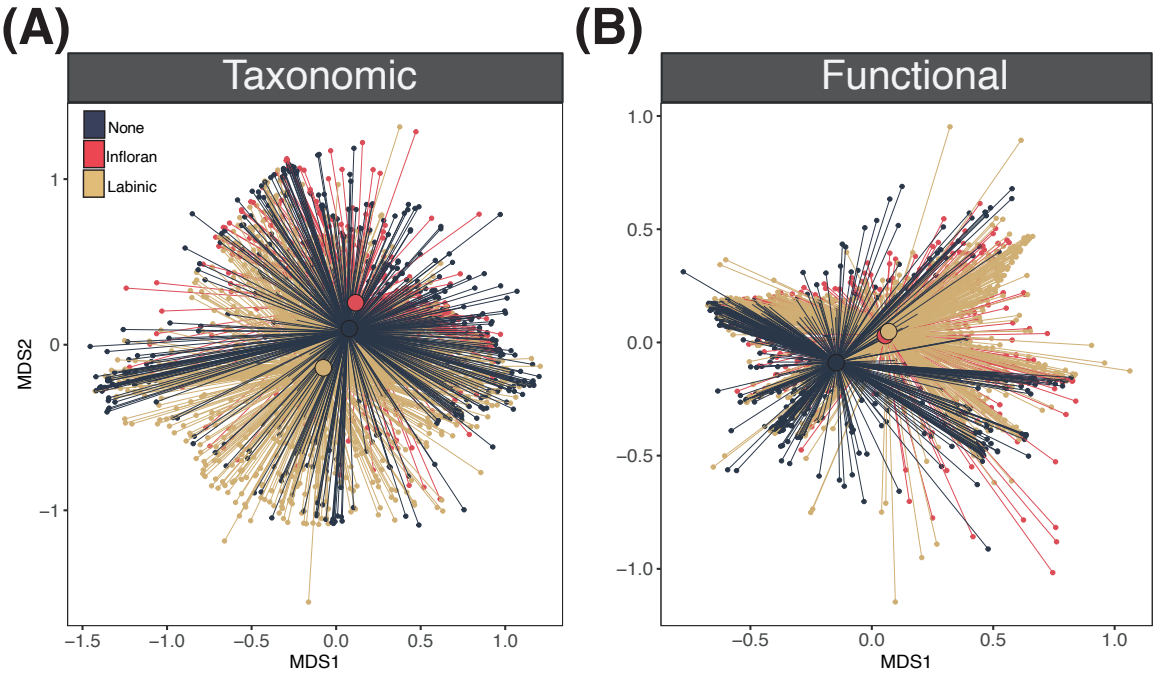

Figure 3.13. Probiotics differentially impact taxonomic profiles but have similar impact on functional potential. NMDS plot of **(A)** taxonomic profiles and **(B)** functional (EC number profiles) for all samples, showing the mean centroid for each probiotic type.

3.3.4. Functional implication and wider significance of PGCTs

To understand how PGCTs relate to samples from full-term infants, the cohort described was compared to the TEDDY study cohort, a study that looked to characterise the microbiome in early life term infants in a large, multi-centre population (Stewart et al., 2018). This was conducted based on the earliest sample from each full-term infant. Samples from both PGCT 4 and in particular PGCT-5, were found to be most similar to those from infants born full-term (**Figure 3.14**). This was true based on all samples **(Figure 3.14A)** and based on just one sample per infant per PGCT (**Figure 3.14B**).

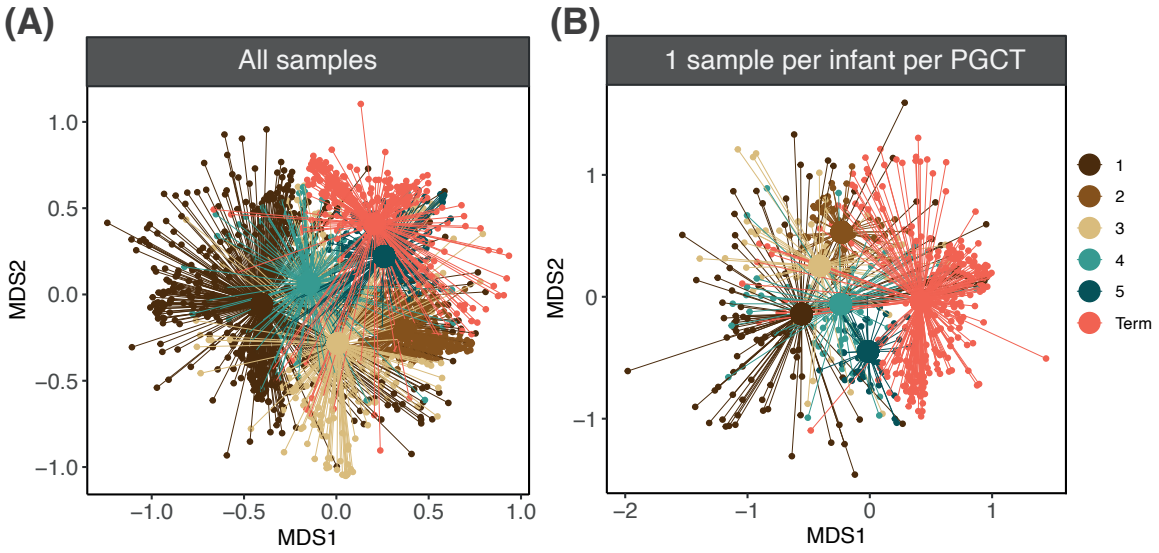

Figure 3.14. PGCT-4 and PGCT-5 are least dissimilar to full-term infant gut communities. NMDS plot of **(A)** all samples and (**B)** 1 sample per PGCT, showing the mean centroid for each PGCT and for samples from infants born full-term from the TEDDY cohort (Stewart et al., 2018). The earliest sample was chosen for each patient from the term cohort.

It was next sought to understand if the PGCTs, defined based on microbial taxonomy, were associated with the functional capacity of the gut microbiome. Analysing EC number showed PGCTs significantly differ in their overall composition (P = 0.001; **Figure 3.15A**), however, no single enzyme or pathway was found to discriminate PGCTs from one another. To further explore the functional impact of PGCTs, a subset of 10 stool samples representative of each PGCT (n = 49; one sample failed QC, see Methods) and matched serum samples (n = 50) were selected for untargeted metabolomics (**Table 3.6**). Overall stool metabolite profiles were found to significantly differ between samples based on PGCT (P = 0.043, **Figure 3.15B**), whereas matched serum metabolite profiles did not (P = 0.151; **Figure 3.15C**).

Table 3.6. Clinical data for samples chosen for metabolomics and organoid experiments. *P*-values are based on Chi-squared test for categorical data and Kruskal-wallis for continuous data.

	PGCT-1	***PGCT-2***	***PGCT-3***	***PGCT-4***	***PGCT-5***	***P-value***
Number of subjects	10	10	10	10	10	
Number of samples	10	10	10	10	10	
Mean DOL (±SD)	53.3 (± 8.3)	53.8 (± 8.9)	51.2 (± 5.9)	51 (± 9.3)	52.9 (± 9.07)	0.885
Median gestational age (IQR)	27 (24.5 – 27.8)	26 (25 – 27.8)	26 (25 – 28)	27 (26.2 – 27)	26.5 (25.2 – 28.8)	0.918
Median birthweight (IQR)	828 (615 – 985)	910 (782 – 970)	928 (782 – 974)	960 (639 – 1055)	880 (775 – 1130)	0.914
Birth mode (% Vaginal)	20	30	60	20	50	0.220
Sex (% Male)	50	30	70	80	60	0.194

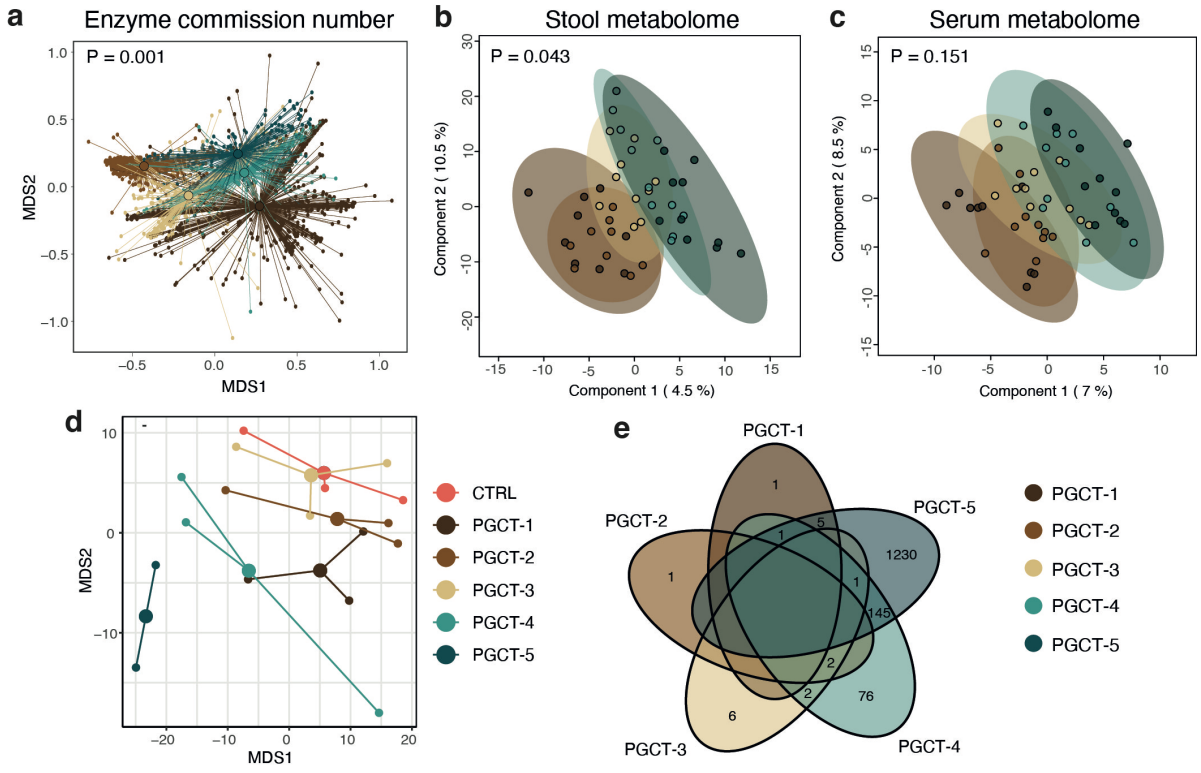

Figure 3.15. Functional implication of PGCTs. (A) NMDS plot of EC number profiles for all samples showing the mean centroid for each PGCT. The statistical significance was based on PERMANOVA, with permutations constrained within the patient. **(B,C)** PLS-DA plots of metabolite profiles ($n = 50$) showing 95% confidence ellipses for each PGCT for stool **(B)** and serum **(C)**. The statistical significance was based on PERMANOVA. **(D)** NMDS plot of preterm intestinal organoid transcriptome profiles showing the mean centroid for each PGCT. CTRL, control. **(E)** Venn diagram showing the number of DEGs compared with control for each PGCT. Zero values were removed for clarity.

Specific metabolites significantly enriched in PGCT-3 (associated with no probiotic infants) compared to PGCT-4/5 (associated with probiotic infants) and vice versa were next explored. In stool, a single unknown metabolite was found to be significantly

enriched in PGCT-3 compared to PGCT-4/5 ($P < 0.001$, $Q = 0.0493$). In serum, a single metabolite, LysoPC 20:3, was found to be significantly enriched in PGCT-4/5 compared to PGCT-3 ($P < 0.001$, $Q = 0.01$) and there was no metabolite correspondingly enriched in both stool and serum.

To explore the impact of small molecules from each PGCT on preterm epithelial barrier function, a preterm intestinal-derived organoid model from an infant at 25 weeks corrected gestation under physiological oxygen conditions was employed. The same 10 stool samples from each PGCT used for metabolomics were used to create sterile faecal supernatants, before being added to differentiated intestinal organoid monolayers for 24 hours. It was confirmed that monolayers remained viable and confluent via the TEER measurements following co-culture (median 3215.5 Ω, IQR 3170.75 – 3265.5 Ω).

Transcriptome profiles from organoids revealed a specific host response to each PGCT faecal supernatant, with PGCT-4 and PGCT-5 clustering distinctly from the other conditions on the X-axis (**Figure 3.15D**). This is further supported by PGCT-4 and PGCT-5 exposed organoid monolayers showing the most DEGs compared to the media control (**Figure 3.15E**). Due to insufficient DEGs being identified for other PGCTs vs control, GO and enrichment analysis was carried out for PGCT-4 and PGCT-5 exposed monolayers only. First grouping genes up-regulated in PGCT-4 and PGCT-5 exposed monolayers by GO revealed various biological processes to be enriched, with a number of cellular and metabolic processes, including cellular protein metabolic processes (**Table 3.7**). Some immune related genes were also commonly differentially expressed, for example amongst those down-regulated were interleukin-23-receptor (IL23R; binds the pro-inflammatory cytokine IL23), CD58 (a ligand of the T-cell co-stimulatory molecule, CD2) and tumour-necrosis-factor superfamily member 15 (TNFSF15; a pro-inflammatory cytokine). Furthermore, other immune related genes were down-regulated in PGCT-5 exposed organoids, only including interleukin-18 receptor 1 (IL18R1; binds the pro-inflammatory cytokine IL18) and IgA-inducing protein homolog (IGIP; enhances IgA secretion from B-cells) Conversely, Leukocyte Receptor Cluster Member 1 (LENG1) was significantly up-regulated for both. Networks of interactions were inferred for both PGCT-4 and PGCT-5, based on their DEGs, and following exclusion of disconnected nodes (**Figure 3.16** and **Figure 3.17**). There was

a significant positive enrichment of genes encoding phosphoproteins for both PGCT-4 and PGCT-5 (both FDR < 0.001), and genes encoding proteins involved in metabolic processes (P = 0.044 and P < 0.001, respectively), as seen based solely on GO. This was particularly true for PGCT-5 (**Figure 3.17**). Interestingly, positively enriched genes common for both PGCT-4 and PGCT-5 exposed monolayers included a number of genes involved in DNA damage response such as growth arrest and DNA damage inducible (GADD45) a and b, and DNA damage inducible transcript 4 (DDIT4) (**Figure 3.16** and **Figure 3.17**).

Table 3.7. Top 30 most significantly enriched GO terms based on biological processes for PGCT-4 and PGCT-5.

	ID	Description	Adj. pval	Term size	Gene ratio
PGCT-4	GO:0008150	biological process	1.99E-164	20908	419/437
	GO:0009987	cellular process	3.67E-162	19222	408/437
	GO:0044237	cellular metabolic process	6.72E-131	12747	346/467
	GO:0008152	metabolic process	1.01E-129	14260	356/459
	GO:1901564	organonitrogen compound metabolic process	1.18E-111	6329	245/467
	GO:0002181	cytoplasmic translation	2.98E-107	148	71/464
	GO:0071704	organic substance metabolic process	4.30E-107	13734	332/467
	GO:0044238	primary metabolic process	3.20E-105	12480	317/467
	GO:0044260	cellular macromolecule metabolic process	2.64E-97	5737	221/467
	GO:0043043	peptide biosynthetic process	2.90E-97	738	105/465
	GO:0006518	peptide metabolic process	1.96E-96	895	111/465
	GO:0006412	translation	5.49E-96	713	103/465
	GO:0043604	amide biosynthetic process	3.21E-95	867	109/465
	GO:0006807	nitrogen compound metabolic process	7.10E-95	11970	300/467
	GO:1901566	organonitrogen compound biosynthetic process	1.89E-94	1697	136/465
	GO:0043603	cellular amide metabolic process	7.59E-94	1165	119/465
	GO:0019538	protein metabolic process	1.72E-89	5380	207/467
	GO:0044267	cellular protein metabolic process	2.05E-87	4787	195/467
	GO:0009058	biosynthetic process	2.99E-83	5954	209/467
	GO:0044249	cellular biosynthetic process	8.84E-83	5792	206/467
	GO:1901576	organic substance biosynthetic process	1.20E-82	5867	207/467
	GO:0034645	cellular macromolecule biosynthetic process	3.82E-82	1563	122/467
	GO:0044271	cellular nitrogen compound biosynthetic process	1.25E-74	4733	180/467
	GO:0043170	macromolecule metabolic process	1.21E-73	12227	277/467
	GO:0034641	cellular nitrogen compound metabolic process	3.17E-71	8684	233/467
PGCT-5	GO:0009987	cellular process	4.941e-324	19222	829/903
	GO:0008150	biological process	4.941e-324	20908	824/866
	GO:0008152	metabolic process	3.10E-266	14260	887/1281
	GO:0065007	biological regulation	2.00E-257	12840	833/1276
	GO:0044237	cellular metabolic process	6.17E-251	12747	827/1281
	GO:0050789	regulation of biological process	2.09E-244	12123	797/1276
	GO:0071704	organic substance metabolic process	1.84E-241	13734	844/1281
	GO:0050794	regulation of cellular process	5.81E-239	11057	758/1276
	GO:0044238	primary metabolic process	8.27E-234	12480	799/1281
	GO:0006807	nitrogen compound metabolic process	2.61E-210	11970	755/1281
	GO:1901564	organonitrogen compound metabolic process	3.97E-200	6329	548/1281
	GO:0050896	response to stimulus	1.02E-188	8811	623/1276
	GO:0044260	cellular macromolecule metabolic process	7.54E-185	5737	506/1281
	GO:0043170	macromolecule metabolic process	1.88E-184	12227	729/1281
	GO:0019538	protein metabolic process	4.27E-177	5380	482/1281
	GO:0048518	positive regulation of biological process	1.90E-162	6086	491/1276
	GO:0051716	cellular response to stimulus	8.55E-157	7343	530/1276
	GO:0044267	cellular protein metabolic process	1.74E-156	4787	432/1281
	GO:0048522	positive regulation of cellular process	1.52E-154	5553	461/1281
	GO:0051179	localization	1.76E-143	6414	474/1255
	GO:0031323	regulation of cellular metabolic process	7.59E-138	5919	452/1276
	GO:0019222	regulation of metabolic process	6.93E-137	7208	498/1276
	GO:0071840	cellular component organization or biogenesis	1.51E-129	7917	509/1256
	GO:0009058	biosynthetic process	2.42E-129	5954	445/1293
	GO:0080090	regulation of primary metabolic process	1.14E-126	5692	429/1281

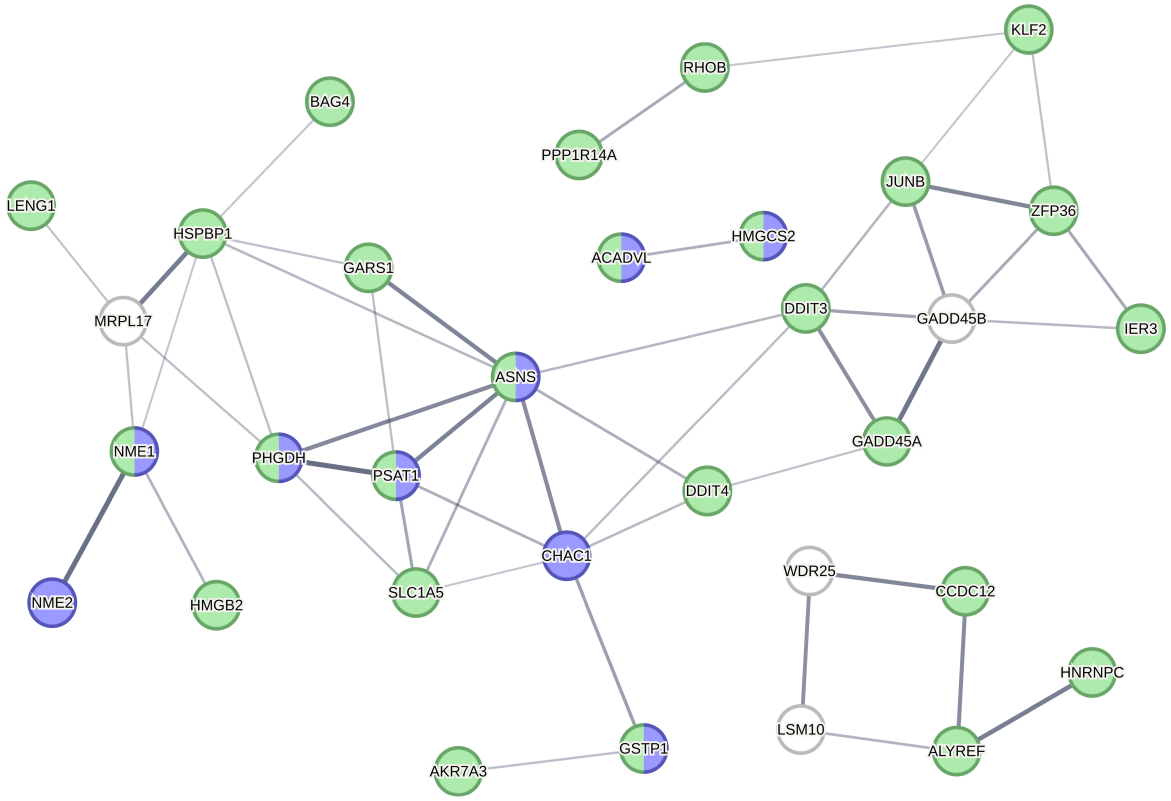

Figure 3.16. Protein-protein network analysis inferred from significantly up-regulated genes from PGCT-4 exposed organoids. Network analysis was performed by STRING V.12.0, based on a minimum interaction score of 0.4 (medium confidence) where edge thickness represents the level of confidence and disconnected nodes have been excluded. Green: phosphoproteins; Blue: proteins involved in metabolic processes.

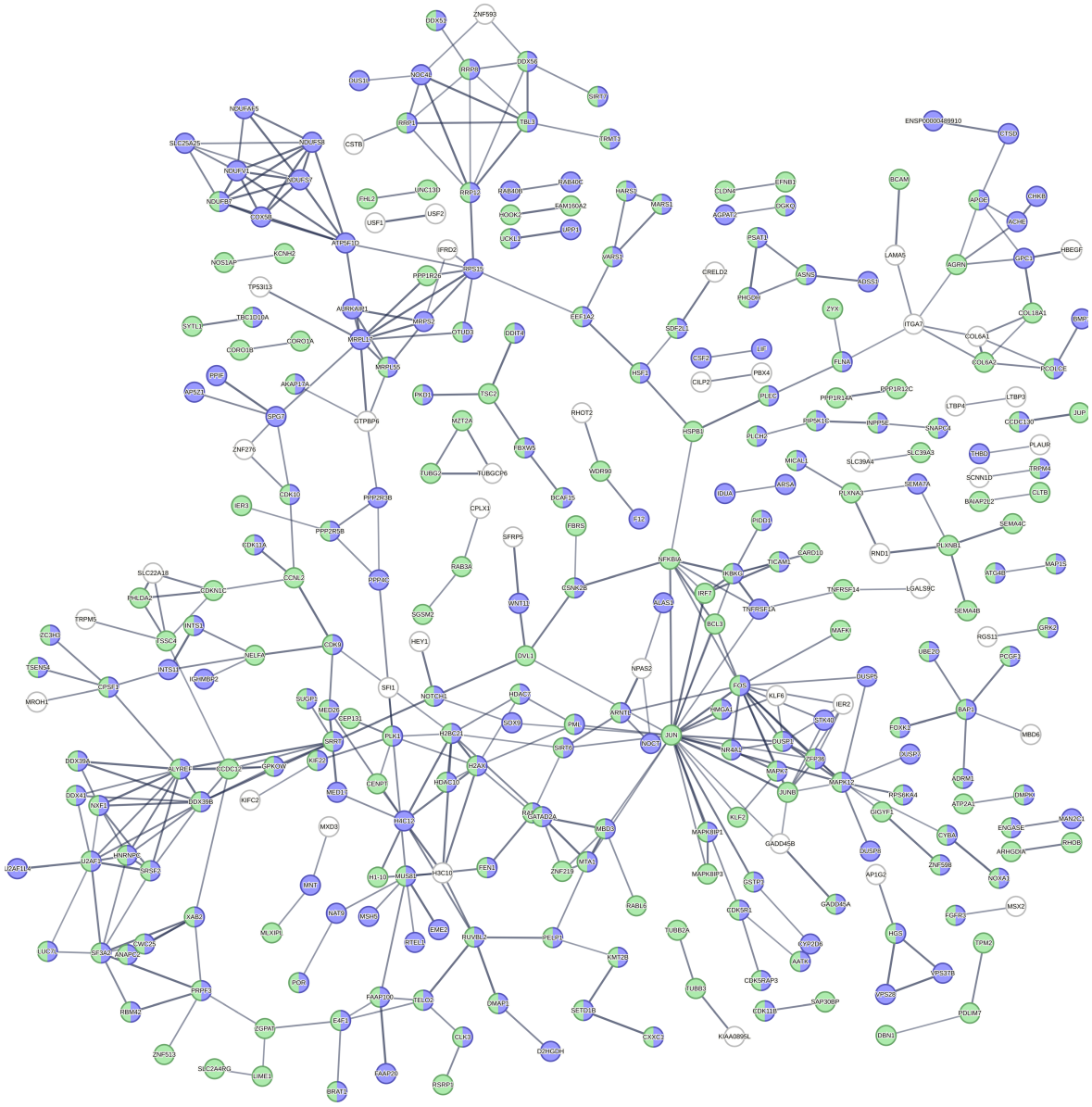

Figure 3.17. Protein-protein network analysis inferred from significantly up-regulated genes from PGCT-5 exposed organoids. Network analysis was performed by STRING V.12.0, based on a minimum interaction score of 0.7 (high confidence) to better visualise the network. Edge thickness represents the level of confidence and disconnected nodes have been excluded. Green: phosphoproteins; Blue: proteins involved in metabolic processes.

In addition to transcriptome profiles, IL-8, a known inflammatory chemokine, was measured from both apical and basolateral media following organoid co-culture. There was no relationship observed between IL-8 production levels and PGCT faecal supernatant exposure in apical media. However, in basolateral media IL-8 production

appeared to be higher for PGCT-3 exposed organoids (**Fig 3.18**), although there was no statistically significant difference in IL-8 concentration between groups (P = 0.251).

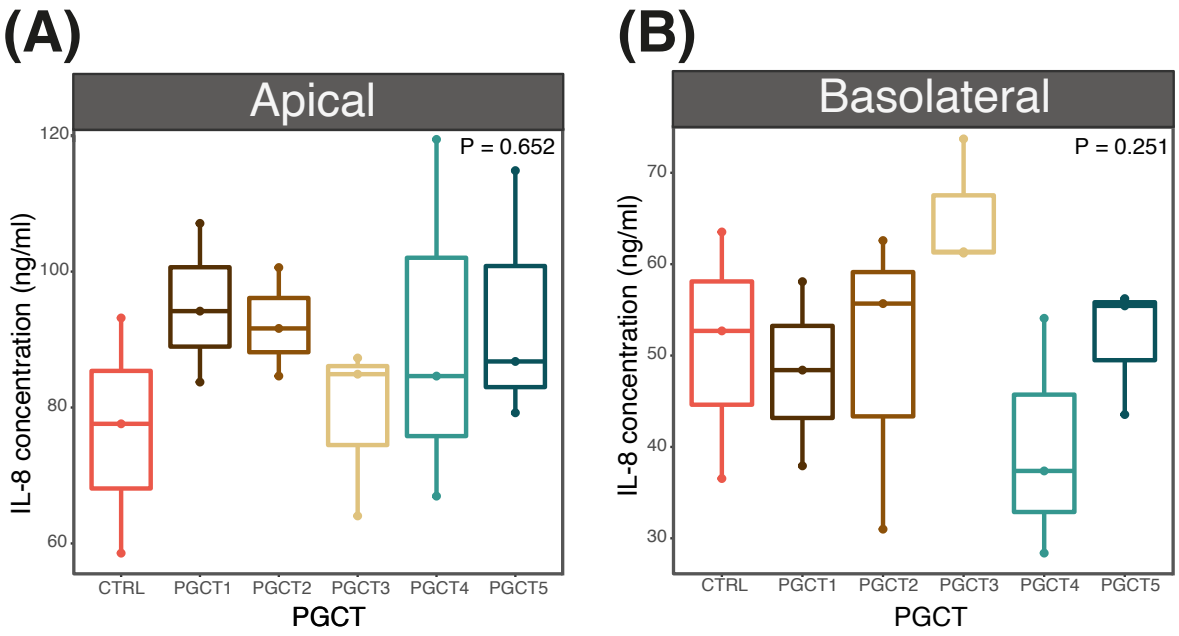

Figure 3.18. Preterm organoids did not differentially produce IL-8 dependent on PGCT faecal supernatant exposure. IL-8 concentration in **(A)** the apical and **(B)** basolateral media following 24-hour incubation with faecal supernatants. IL-8 concentration was measured by ELISA, and P-values were calculated by Wilcoxon rank test.

3.3.5. Modulation of the infant microbiome by diet and antibiotics

Receipt of BMF, receipt of MOM, and antibiotics were significantly associated with Shannon diversity (**Table 3.2**) and the gut microbiome profiles around 1 month of life only (**Figure 3.5**). Shannon diversity was significantly higher after receipt of MOM compared to never receiving MOM (P = 0.012; **Figure 3.19A**) and was significantly reduced in samples where antibiotics had been given in the previous 7 days ($P < 0.001$; **Figure 3.19B**).

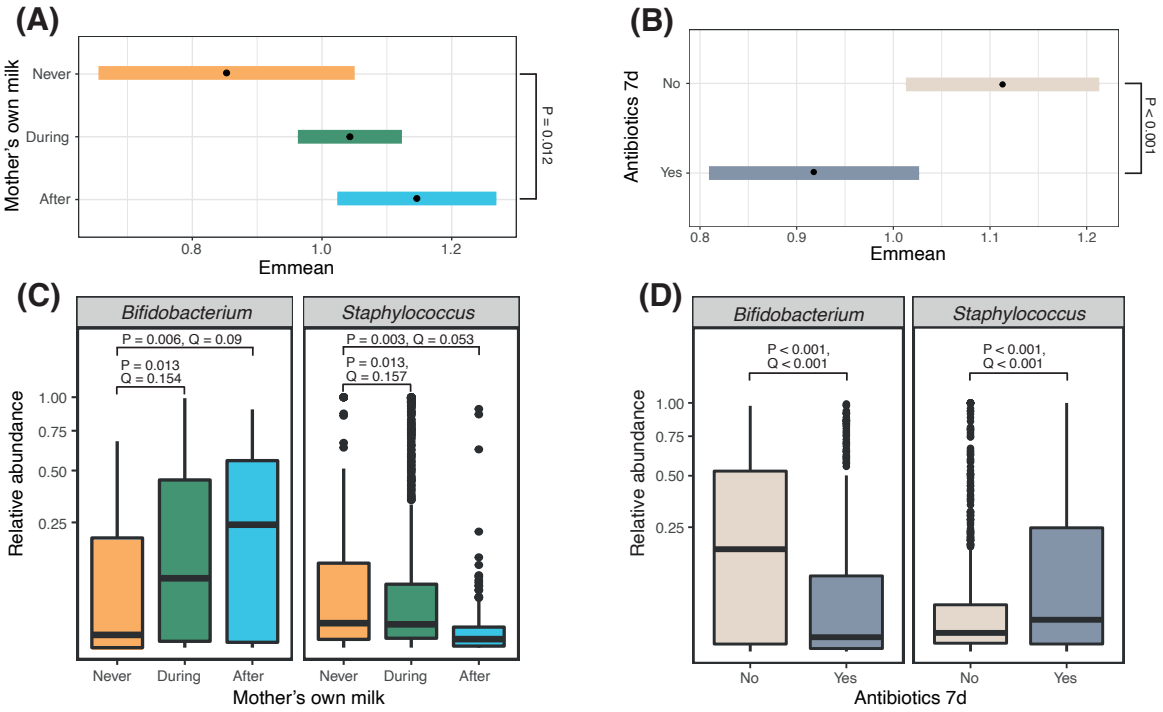

Figure 3.19. MOM and antibiotics are significantly associated with the preterm gut microbiome. (A,B) Estimated marginal means (95% CI) representing Shannon diversity for **(A)** MOM and **(B)** antibiotics, obtained from the Shannon diversity LMMs adjusted for gestational age, birthweight, birth mode, sex, season, day full feed, BMF, formula, probiotics, weight z-scores, DOL, and patient ID. Significance shown is after adjustment for multiple comparisons by Dunnet's method, whereby 'never' or 'no' was used as the control, respectively. **(C,D)** Box plots showing the relative abundance of *Bifidobacterium* and *Staphylococcus* in **(C)** MOM groups and **(D)** antibiotic groups. The centre lines denote the median, the box limits denote the IQR and whiskers extend to the limits. Points outside the whiskers represent outliers. Significance is based on P-values from MaAsLin2.

Samples collected during BMF were more likely to belong to *Escherichia* dominant PGCT-2 (**Table 3.3**), have higher *Escherichia* genus (**Table 3.4**) and an unclassified *Escherichia* sp. (**Table 3.5**). Compared to infants who never received MOM, the relative abundance of *Bifidobacterium* was significantly higher in samples collected during (P = 0.013, Q = 0.154) and after (P = 0.006, Q = 0.09) receipt of MOM, and the relative abundance of *Staphylococcus* was significantly lower during (P = 0.013, Q = 0.157) and after (P = 0.003, Q = 0.053) (**Figure 3.19C** and **Table 3.4**). Analysis at the species level did not find specific *Bifidobacterium* spp. to be significantly enriched with MOM, whereas lower *Staphylococcus* was primarily driven by *S. aureus* (**Table 3.5**). Inverse associations were observed in infants who received antibiotics, where receipt

of antibiotics in the previous 7 days significantly reduced the relative abundance of *Bifidobacterium* ($P < 0.001$, $Q < 0.001$) and increased *Staphylococcus* ($P < 0.001$, $Q < 0.001$) (**Figure 3.19D** and **Table 3.4**). At the species level, *B. bifidum*, *B. longum*, *B. breve*, were significantly reduced (all $P < 0.01$, $Q < 0.1$) and *S. haemolyticus*, *S. warneri*, and *S. lugdunensis* were significantly increased (all $P < 0.01$, $Q < 0.2$); **Table 3.5**).

At a functional level, across all samples, MOM, BMF, and antibiotics were also found to be significantly associated with the total number of ARGs (**Table 3.2**). Post-hoc analysis revealed the total number of ARGs to be significantly higher after receipt of BMF compared to during ($P = 0.001$) and higher during and after ($P < 0.001$) receipt of MOM compared to infants who never received MOM. ARGs were also higher when antibiotics had not been given in the 7 days before compared to if they had ($P = 0.036$).

Furthermore, the relative abundance of a single enzyme, a transaldolase (EC 2.2.1.2), was found to be the most significantly associated EC number both during and after receipt of MOM (https://github.com/laurencbeck/supplementary_tables - **Appendix 1**). The relative abundance of various other transferases were also found to be positively associated during and after receipt of MOM, particularly glycosyltransferases and those involved in the transfer of one-carbon and phosphorus- containing groups. In contrast, the relative abundance of the majority of EC numbers were negatively associated with receipt of antibiotics in the past 7 days, such as enzymes involved in forming carbon-nitrogen bonds and transfer of one-carbon groups (https://github.com/laurencbeck/supplementary_tables - **Appendix 1**). ECs positively associated with antibiotics, include oxidoreductases acting on sulphur groups and those acting on paired donors.

3.4. Discussion

The current study represents the largest longitudinal metagenomic analysis of significantly preterm infants who did not develop intestinal complications or sepsis. Where administered, probiotics were the primary factor influencing the preterm gut microbiome, followed by receipt of antibiotics, MOM, and BMF. Two different probiotic products altered the transition of the microbiome into different PGCTs, both characterised by samples collected at the oldest postnatal ages. The PGCTs were

enriched in different *Bifidobacterium* spp. and showed differences in their functional characteristics and interaction with the host epithelium.

Other findings such as the lack of association with other clinical data, were further validated using the Olm *et al*. cohort(Olm et al., 2019), highlighting important differences in comparison to term infants (Bokulich et al., 2016; Chernikova et al., 2018; Reyman et al., 2019). Birth mode was not associated with the microbiome and the total variance explained by co-variates was around 10-fold lower than observed in term infants (Stewart et al., 2018). This suggests that NICU practices and environment dominate the preterm microbiome, which is important when interpreting findings from different settings. Additionally, it could be that the unit-wide use of antibiotics immediately following birth in this population may mask the changes that would have been seen between vaginal and caesarean born infants.

Over this 10-year observational study, the impact of two different probiotic products that were used during discrete time periods, and prior to probiotics ever being used, was investigated. Once a probiotic was in use, probiotic species were detected in stool before deliberate administration. This 'unit cross-colonisation' has been seen in previous studies (Costeloe et al., 2016; Hickey et al., 2014; Jacobs et al., 2013) and has important implications for probiotic trial design. For example, future RCTs may benefit from separately housing infants based on trial arm or using multi-centre studies that follow similar clinical practices. That said, it is important to note that multi-centre studies would need to carefully consider potential confounders based on clinical practices. The findings of cross-colonisation described here may explain why RCT results have varied so much, and why some probiotic studies have not found any effects on NEC.

Previous studies in preterm infants have shown probiotics to alter the gut microbiome (Abdulkadir et al., 2016; Alcon-Giner et al., 2020; Nguyen et al., 2021; van Best et al., 2020; Yousuf et al., 2020). In the present study, the probiotic product was identified as the main driver in shaping the bacterial community at both a taxonomic and functional level. Supplementation of either Infloran or Labinic was associated with transition into two different *Bifidobacterium* spp. enriched PGCTs (PGCT-4 and -5), both of which reflected samples obtained from the oldest infants. Previous studies have found

Bifidobacterium enriched PGCTs to be associated with positive health outcomes (Stewart et al., 2016), but the functional implications of this have not previously been explored.

To determine the relevance of PGCTs on host-microbial interaction, metabolomics was performed on matched stool and serum samples, and an experimental preterm intestinal organoid model was used. Overall metabolite profiles of stool, but not serum, were associated with the PGCT. In addition, sterile faecal supernatants containing the metabolites and other components of stool were found to impact preterm epithelial response in a PGCT-specific manner. Of note, although a healthy section of tissue with viable crypts was used for organoid generation, intestinal organoid models derived from preterm infants require a patient to have a clinical complication requiring surgery (in this case NEC) and so are not healthy individuals. The intestinal region (i.e., small or large intestine) and maturity of the patient may also impact host transcription (Masi et al., 2021). Although further work is needed to determine the potential biological significance of the functional changes resulting from probiotic administration, this demonstrates that transition into different PGCTs, which was driven by probiotic use, has associated functional implications.

It is important to note that excretion of supplemented strains in stool collected during treatment does not necessarily imply intestinal colonisation. Therefore, an assessment of the persistence of strains after stopping probiotics was included, hypothesising that colonised strains would persist, and lack of persistence would indicate a strain was not truly colonising. Several studies have shown individual difference in probiotic and transient microbe colonisation (Maldonado-Gómez et al., 2016; Zhang et al., 2016; Zmora et al., 2018), as well as differences in the persistence of probiotic species following treatment, particularly higher persistence of bifidobacterial strains compared to lactobacilli (Alcon-Giner et al., 2020; van Best et al., 2020; Yousuf et al., 2020). Individualised differences in probiotic colonisation were also observed. All *Bifidobacterium* species showed higher persistence compared to *L. acidophilus* and the persistence of *B. bifidum* and *L. acidophilus* (i.e., the two strains present in both probiotics) were dependent upon the probiotic used. The lower persistence of *L. acidophilus* may reflect the preterm gut ecosystem not being optimal for this species, as it is not a commonly abundant or persistent member of the preterm gut microbiome

(Abdulkadir et al., 2016; Alcon-Giner et al., 2020; Esaiassen et al., 2018; Yousuf et al., 2020). Alternatively, it could be that because *L. acidophilus* colonises the small intestine, it is less likely to be detected in stool thereafter (Yousuf et al., 2020). These results highlight altered short- and long-term colonisation depending upon the probiotic/strains used, emphasising the importance of better understanding short- and long-term impacts at strain-level.

Despite the apparent importance of probiotics in this population in providing an early source of *Bifidobacterium* spp*.,* natural *Bifidobacterium* colonisers were identified, namely *B. breve, B. dentium*, and *B. longum* subsp. *longum.* It has been widely reported that MOM has a bifidogenic effect through the provision of HMOs (Garrido et al., 2015; Lawson et al., 2020; Ruiz-Moyano et al., 2013; Turroni et al., 2010). All *Bifidobacterium* species detected in this preterm cohort have been previously shown to utilise HMOs for growth, with notable variation at the strain level (Lawson et al., 2020). Notably, MOM was found to be associated with an increased relative abundance of *Bifidobacterium* and decreased relative abundance of *Staphylococcus*, whereas antibiotics showed opposing results, being associated with a decreased relative abundance of *Bifidobacterium* and increased relative abundance of *Staphylococcus*. Although the positive effects of MOM and breastfeeding are indisputable, and it is noted here that MOM is associated with a higher Shannon diversity and the *Bifidobacterium* genus, associations with specific taxa at the species level were not identified. This is likely due to the extensive provision of MOM in this NICU resulting in a lack of power, and also the administration of probiotics which likely mask a lot of the changes. Another limitation is the use of samples from a single NICU, which could affect the generalisability of these findings. However, in doing so, the source of variation in this study was reduced and ensured the standardisation of clinical variables.

The impact of antibiotic use on the emergence of specific ARGs and the emergence of drug-resistant pathogens is well-recognised, but underexplored in preterm infants (Gibson et al., 2016; Nguyen et al., 2021). Here, antibiotic use was found to be associated with a lower total number of ARGs, likely reflecting the reduced bacterial richness as a consequence of antibiotic treatment. Proteobacteria were further found to carry the highest number of ARGs in relation to their relative abundance, a

potentially important finding given the widely reported link between Proteobacteria and NEC onset (Olm et al., 2019; Pammi et al., 2017; Stewart et al., 2019).

3.5. Limitations

This data presented in this study was obtained from a single centre, and therefore the findings may not be entirely relevant or applicable to other cohorts. It is important to consider potential variations in patient demographics, clinical care regimes and environmental factors between centres, as these factors can influence the observed outcomes. Additionally, the data collected was across a 120-day period (the majority being earlier samples) during the infant stay on the NICU. This timeframe may not fully capture the extent of longitudinal changes of the gut microbiome, and it is important to consider changes that may occur beyond this monitored period (i.e., post-discharge). It would be interesting to evaluate the impact of probiotics post-discharge and future studies may consider this. In terms of the collected metadata, information on feeding lacks granularity (i.e., % of enteral milk received) and therefore the precise role of breast milk may not have been fully captured. Furthermore, these data are based on relative abundance which as described in Section 1.7.6, does not give us information on the true abundance of taxa and should therefore be interpreted with caution.

The methods used in the current study and data presented in subsequent chapters, such as the DNA extraction from stool samples and the metagenomic sequencing method, were designed with a primary focus on bacterial profiling. Notably, the relatively low sequencing depth used may have contributed to the limited detection of other microorganisms and although few other microorganisms were detected (e.g., fungal species), this is not to say that these are not present and important in the preterm gut environment.

Finally, the organoid experiments were carried out using a single cell line in triplicate for each condition, and so repeats are technical rather than biological replicates. Consequently, the extent to which the experimental findings can be generalised is restricted, especially since IL-8 production and gene expression profiles are likely to differ somewhat between preterm cell lines. However, the experiment has been run as a proof of concept in order to offer mechanistic evidence to support the primary focus of the study surrounding the computational analysis.

3.6. Conclusions

In summary, this study has shown in a large and extensively longitudinally sampled population of preterm infants, that the choice of probiotic product impacted development of the gut microbiome in different ways, accelerating transition into *Bifidobacterium* dominant PGCT-4 or -5, which reflected bacterial communities of the oldest samples. In addition, these PGCTs showed differences in their functional implications and interaction with the host epithelium. These results help provide a framework and identify important aspects for consideration when designing interventional trials targeting the gut microbiome of preterm infants.

4. Microbial signatures of necrotising enterocolitis in preterm infants

4.1. Abstract

NEC is a devastating GI disorder primarily affecting preterm infants and is responsible for significant morbidity and mortality. Thought to have a multifactorial aetiology, NEC has been associated with changes in the gut microbiome, with previous studies pointing to an increase in Proteobacteria and decrease in *Bifidobacterium* to be associated with disease. This decrease in *Bifidobacterium* has led, in part, to an increased interest in probiotics in the NICU. However, whilst more broad associations with disease have been made, no single causative agent has ever been identified, and the pathogenesis of disease remains poorly understood. Here, metagenomics was performed on extensive longitudinal stool samples (n = 2041) from infants who went onto develop NEC (n = 75) alongside those who did not (n = 123). The findings reveal that before clinical diagnosis of NEC, the relative abundance of Proteobacteria such as *Klebsiella variicola* is higher whilst the relative abundance of Actinobacteria, mainly Bifidobacteria is lower, concordant with other studies. In the first study of its kind, exploring the microbial origins of NEC in the context of probiotic use, both the type and administration of probiotics influenced the microbial association observed with NEC, in particular the colonisation patterns of different probiotic and probiotic-associated strains. The results of this study have provided important insights into microbial changes that occur preceding NEC onset, and informed follow-up studies to further identify potential biomarkers for disease.

4.2. Introduction

NEC is a devastating GI disorder affecting up to 10% of preterm infants and is responsible for significant morbidity and mortality. Thought to have a multifactorial aetiology, alterations in the gut microbiota have been implicated in its pathogenesis, although no definitive causative agent has been identified. Results typically vary between studies and cohort, but in general, previous findings have pointed towards a higher abundance of Proteobacteria, and lower abundance of beneficial Bifidobacteria in NEC cases. Notably, a large metagenomic study by Olm *et al.* (Olm et al., 2019) of 1163 faecal samples from 160 preterm infants (n = 34 NEC infants and 282 NEC samples), found samples prior to NEC diagnosis to have a greater abundance of *Klebsiella* spp. as well as bacteria encoding fimbriae and gene clusters related to quorum sensing and biofilm formation. Other studies have also found a link between NEC and a greater abundance of *Klebsiella* spp. (Casaburi et al., 2022; Coleman et al., 2023), as well as *Enterobacter* spp. (Casaburi et al., 2022) and other Gammaproteobacteria (Warner et al., 2016). Aside from the study by Olm *et al.* (Olm et al., 2019) the majority of previous studies investigating the link between the gut microbiome and NEC onset have largely depended on the use of 16S rRNA gene sequencing. Whilst informative and in many circumstances sufficient, this technique is limited to genus-level identification and therefore will not capture more granular changes in species or strains. In addition, many studies are constrained to smaller cohorts and cross-sectional data, which fails to capture temporal changes and makes it difficult to appreciate the dynamics of the preterm gut microbiome. A longitudinal metagenomic study in a large cohort is therefore necessary to address these limitations.

Understanding the microbial landscape proceeding NEC onset represents an important aspect of study, as identification of abnormal microbial signatures may help to inform future therapeutics and diagnostic tools. The association of the gut microbiome with NEC has led to an increased interest in the use of probiotics as a potential preventative measure. Comparison of microbial characteristics and whether biomarkers of health or disease differ dependent on probiotic consumption, represents a novel element of investigation which is yet to be explored.

Building on the groundwork described in Chapter 3, in the current chapter extensive longitudinal stool samples ($n = 2041$) from preterm infants who went onto develop NEC alongside those who did not, were analysed. This represents the largest metagenomic study of NEC described to date. The aims of this chapter are to identify microbial signatures of NEC before disease diagnosis and compare these between the probiotic groups (i.e., no probiotic/Infloran/Labinic) analysed in Chapter 3. The results of this study will provide important insights into microbial changes that occur preceding NEC onset, which will help inform directions for future research into therapeutic and diagnostic tools.

4.3. Results

The current metagenomic analysis included 2041 samples from 199 preterm infants born <32 weeks of gestation. Of these infants, a total of 75 were diagnosed with NEC (*n* = 547 samples), 24 of whom also developed LOS, and a total of 124 were neither diagnosed with NEC, LOS nor any other GI disease (*n* = 1494 samples) (**Table 4.1**). Of the NEC samples, 252 were collected prior to NEC diagnosis or on the day of NEC diagnosis (pre-NEC) and 295 were obtained following NEC diagnosis (post-NEC; **Figure 4.1**). For infants who developed NEC, the median (IQR) day of onset was 19 (13 – 33) (**Table 4.1**).

Table 4.1. Patient demographics of the analytical cohort. *P*-values are based on Chi-squared test for categorical data and Kruskal-Wallis test for continuous data.

	All	***Healthy***	***NEC***	***P-value***
No. of subjects	199	124	75	
No. of samples	2041	1494	547	
Median no. samples per subject (IQR)	10 (7 – 13.5)	11 (9 – 14.3)	6 (3 – 10)	
Median day of NEC diagnosis (IQR)			19 (13 – 33)	
Median gestational age (IQR)	26.7 (25.3 – 28.1)	27 (25.7 – 28.7)	26 (24.4 – 27.8)	0.007
Median birthweight (g) (IQR)	860 (697.5 – 1065)	902 (718 – 1136)	800 (625 – 970)	0.002
Median day of first feed (range)	2.5 (2 – 4)	2 (2 – 4)	3 (2 – 4)	0.118
Median day of full feed (IQR)	14 (12 – 20)	12 (14 – 16)	18 (13 – 30)	<0.001
Median no. of antibiotic courses (IQR)	3 (2 – 5)	2 (1 – 3)	5 (3 – 6)	<0.001
Mother's own milk ever	189 (95.0%)	115 (92.7%)	74 (98.7%)	0.129
Formula ever	133 (66.8%)	75 (60.5%)	58 (77.3%)	0.048
Median start day of probiotics (IQR)	5 (7 – 10)	5 (7 – 9)	4 (7 – 12)	0.477
Median stop day of probiotics (IQR)	47 (32 – 56)	45 (31 – 54)	49 (36.8 – 60)	0.051
Birth mode				1
Caesarean	109 (54.8%)	68 (54.8%)	41 (54.7%)	
Vaginal	90 (45.2%)	56 (45.2%)	34 (45.3%)	
Sex				0.312
Male	117 (61.6%)	69 (55.6%)	48 (64%)	
Female	82 (43.2%)	55 (44.3%)	27 (36%)	
Probiotic type				0.006
No probiotic	41 (20.6%)	29 (23.4%)	12 (16%)	
Infloran	54 (27.1%)	24 (19.4%)	30 (40%)	
Labinic	104 (52.3%)	71 (57.3%)	33 (44%)	
Antibiotics in the past 7 days samples				<0.001
Yes	656 (32.1%)	411 (27.5%)	245 (44.8%)	
No	1385 (67.9%)	1083 (72.5%)	302 (55.2%)	
MOM samples				<0.001
Never	117 (5.7%)	110 (7.4%)	7 (1.3%)	
Before	11 (0.5%)	7 (0.5%)	4 (7.3%)	
During	1566 (76.7%)	1202 (80.5%)	364 (66.5%)	
After	347 (17.0%)	175 (11.7%)	172 (31.4%)	
BMF samples				<0.001
Never	908 (44.5%)	506 (33.9%)	402 (73.5%)	
Before	498 (24.4%)	418 (28.0%)	80 (14.6%)	
During	487 (23.9%)	447 (30.0%)	40 (7.3%)	
After	148 (7.3%)	123 (8.2%)	25 (4.6%)	
Formula samples				<0.001
Never	720 (35.3%)	603 (40.4%)	117 (21.4%)	
Before	603 (29.5%)	406 (27.2%)	197 (36.0%)	
During	682 (33.4%)	454 (30.4%)	228 (41.7%)	
After	36 (1.8%)	31 (2.1%)	5 (0.9%)	

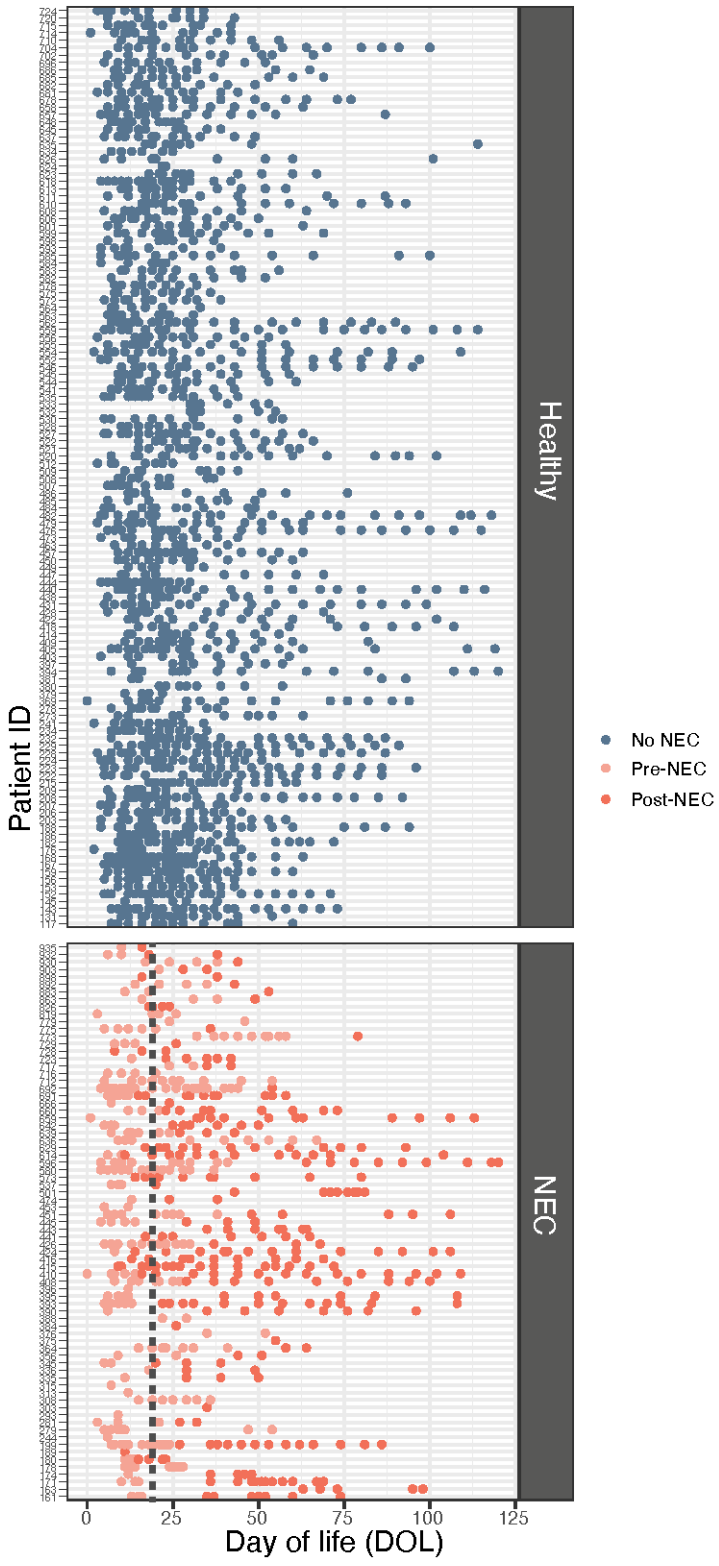

Figure 4.1. Healthy and NEC infant sampling overview. Samples used in this study from birth to day 120, coloured based on the sample timing in relation to NEC diagnosis. Dashed line represents the median day of NEC onset at DOL 19.

4.3.1. Validation of MetaPhlAn v3.0 output

In comparison to Chapter 3, in this chapter, microbiome data was based on an updated version of MetaPhlAn (MetaPhlAn v3.0) and functional data based on an updated version of HUMAnN (HUMAnN v3.0). The same samples analysed in Chapter 3 were re-incorporated here to serve as 'healthy' control samples, but all samples were re-ran using the updated bioBakery tools (i.e., MetaPhlAn and HUMAnN) as described in the methods.

As the 1431 samples analysed in Chapter 3 were re-mapped using the updated MetaPhlAn pipeline, which includes additional marker genes in the curated database, the first aim was to assess comparability with the previous work in Chapter 3. Based on the MetaPhlAn v3.0 output, no archaea and only nine fungal species were detected (contrasting the 11 fungal species identified using MetaPhlAn v2.0 in Chapter 3). Despite the slight reduction, the same fungal species were identified across the two datasets. However, using MetaPhlAn v3.0, *C. parapsilosis* and *M. restricta* were not detected, which were both identified previously, albeit at low abundance. The reads that were previously assigned to these species did not appear to be re-assigned to any other fungal or bacterial species. Similar to the findings of Chapter 3, *C. albicans* was the most abundant and prevalent fungi, but only detected in 22 samples (13 infants). Considering the extremely low prevalence of viruses identified in Chapter 3 from the VirMAP pipeline, this aspect was omitted from the current analysis.

In total, 537 bacterial species were identified with MetaPhlAn v3.0, compared with 394 identified in Chapter 3 using MetaPhlAn v2.0. Due to the dominance of bacterial species in the dataset, subsequent analysis was again focused primarily on bacteria. Consistent with Chapter 3, the most abundant bacterial genera were *Bifidobacterium, Escherichia, Enterococcus, Klebsiella* and *Staphylococcus* and the most abundant species were *E. coli, E. faecalis, B. longum, B. breve* and *K. pneumoniae.* Similar trends were observed longitudinally, with the relative abundance of *Bifidobacterium* increasing over DOL and *Staphylococcus* decreasing, corresponding with the relative abundance of Actinobacteria and Firmicutes, respectively (**Figure 4.2A, B**).

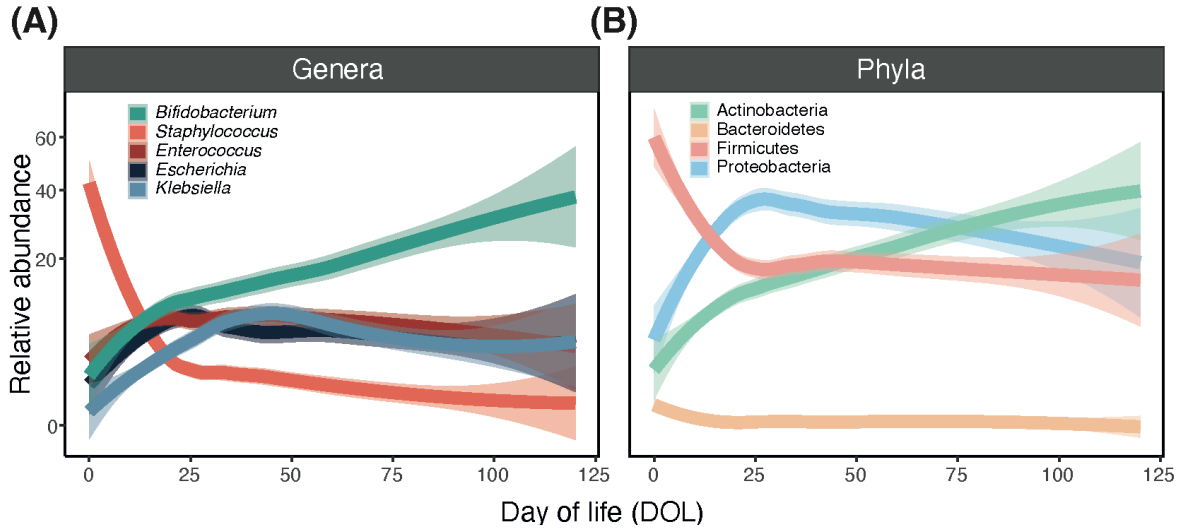

Figure 4.2. Overview of the preterm gut microbiome in healthy and NEC infants. LOESS fits with 95% CI modelling the relative abundance of **(A)** the most abundant genera and **(B)** the most abundant phyla over the first 120 days of life.

4.3.2. Determining preterm gut community types

DMM modelling of bacterial profiles was used to determine PGCTs, as previous. Based on Laplace approximation score, 6 clusters were found to be optimal, however, data were clustered into 5 PGCTs for consistency with Chapter 3. PGCTs were numbered 1–5 based on the average age of samples within that cluster. Species richness expectedly increased through PGCTs ($P < 0.001$), with a significant difference in richness observed between each PGCT ($P < 0.001$) (**Figure 4.3A**). Similarly, Shannon diversity also increased based on PGCTs ($P < 0.001$), with a significant difference observed between each PGCT ($P < 0.001$), apart from PGCT-3 and PGCT-4 (**Figure 4.3B**).

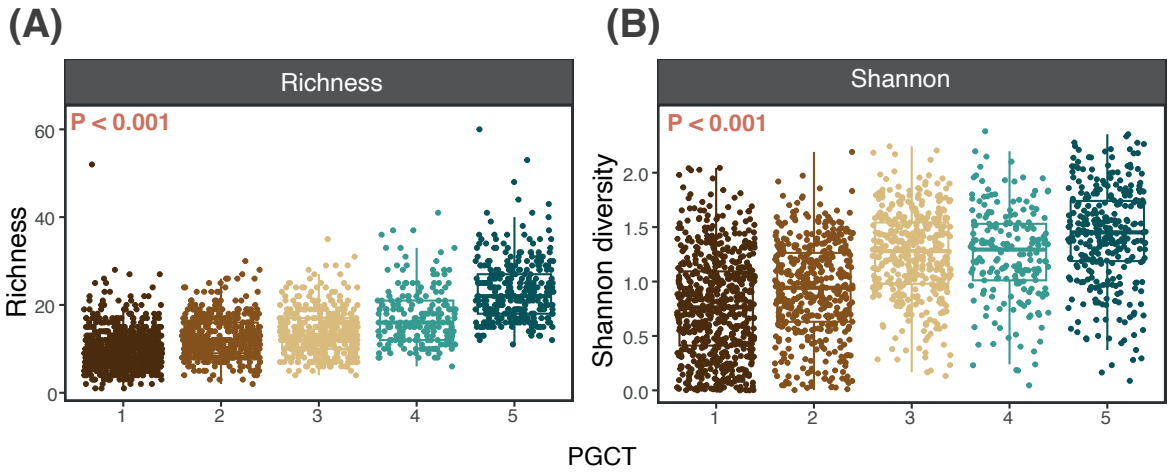

Figure 4.3. Alpha diversity differs based on PGCTs. (A) Shannon diversity and **(B)** species richness. P-values are based on global *P*-values obtained from ANOVAs on fitted LMMs, controlling for patient ID.

LefSE was used to determine the most discriminatory taxa for each PGCT, using a cut-off Log_{10}LDA score of 4. PGCT-1 was discriminated by *Enterococcus faecalis* and Staphylococci (*Staphylococcus epidermidis, S. argenteus* and *S. warneri*); PGCT-2 was discriminated by *Enterobacter cloacae*, *E. coli* and two members of the *K. oxytoca* complex *(K. michiganensis and K. oxytoca*); PGCT-3 was discriminated by various members of the *K. pneumoniae* complex *(K. pneumoniae, K. variicola* and *K. quasipneumoniae); PGCT-4* was discriminated by *B. breve, E. faecium* and *Lactobacillus casei,* and PGCT-5 was discriminated by *B. longum, B. bifidum, Cutibacterium avidum* and *L. rhamnosus* (**Figure 4.4A,B**). PGCTs were generally comparable to those identified in Chapter 3, with the earliest PGCTs being more Staphylocci-dominant and later PGCTs being more *Bifidobacterium*-dominant. Interestingly, members of either the *K. oxytoca* complex or *K. pneumoniae* complex were not found to co-exist (**Figure 4.4A**).

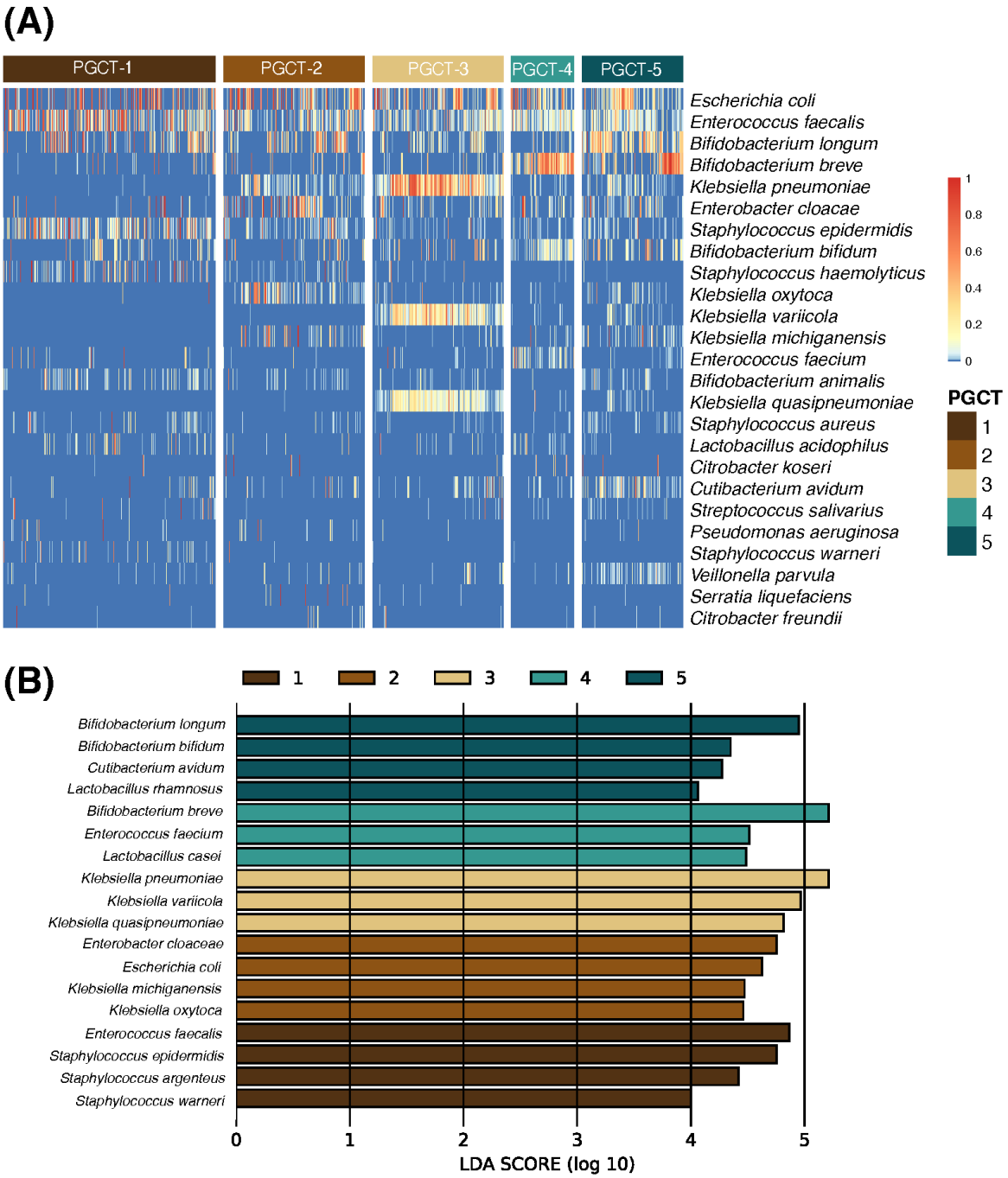

Figure 4.4. PGCT discriminatory features. (A) Heatmap of all samples (n=2041) showing the relative abundance of the most dominant species, stratified by PGCT. **(B)** LEfSe identifying discriminatory features of each PGCT based on LDA. Coloured bars denote PGCTs.

4.3.3. Longitudinal development and stability of the preterm gut microbiome in NEC infants prior to and following diagnosis

The impact of NEC on longitudinal development and stability of the preterm gut microbiome was first explored. Transition plots showed NEC (i.e., healthy vs NEC) to impact the temporal development of the gut microbiome, altering transition through PGCTs (**Figure 4.5**).

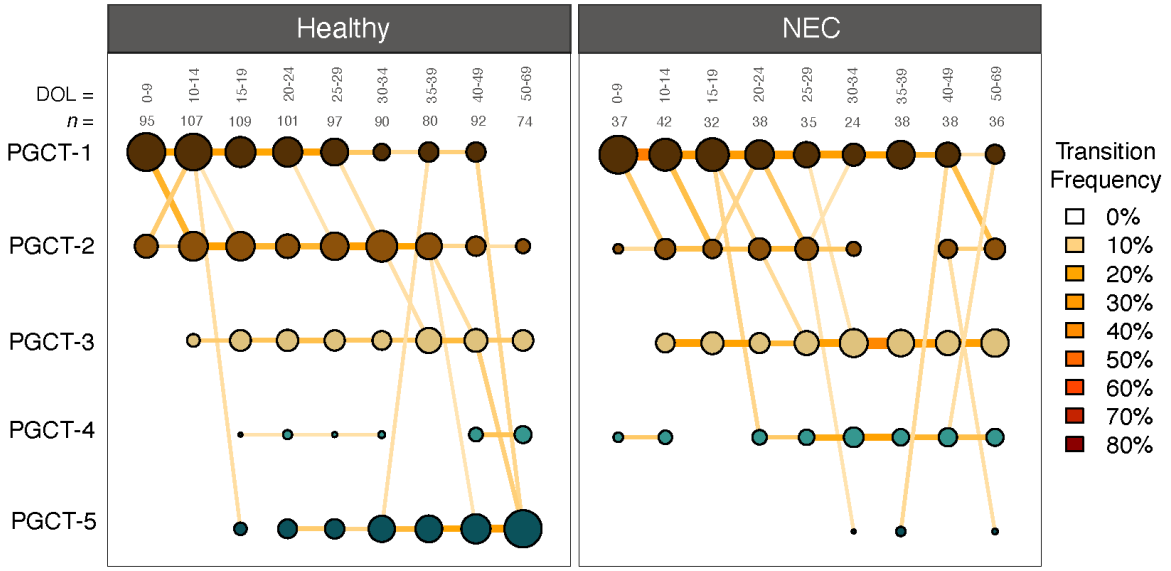

Figure 4.5. NEC impacts the transition of the preterm gut microbiome over time. Transition plot showing the progression of samples through each PGCT from DOL 0 to DOL 69, based on whether an infant developed NEC or not. The nodes and edges are sized based on the total counts; nodes are coloured according to PGCT and edges by the transition frequency.

To elucidate differences in the gut microbiome occurring prior to disease onset to those occurring as a consequence of disease, binomial mixed models were fit to explore the association of NEC (i.e., no-NEC/pre-NEC/post-NEC) with PGCTs. Infants who developed NEC were significantly less likely to transition into PGCT-5 ($P < 0.001$), and significantly more likely to remain in PGCT-1 following NEC diagnosis (i.e, post-NEC) but not prior to NEC diagnosis (i.e., pre-NEC), compared to healthy infants ($P = 0.003$; **Figure 4.5, Table 4.2**). PGCT-1 was also significantly associated with a lower DOL ($P < 0.001$), as well as samples taken before BMF ($P = 0.037$), when antibiotics had been used in the previous 7 days ($P < 0.001$) and when infants had not been fed MOM ($P < 0.001$). Conversely, PGCT-5 was significantly associated with a higher DOL (both $P < 0.001$), as well as Labinic use ($P = 0.004$) and use of MOM ($P = 0.002$). Transition into PGCT-5 was significantly less likely after formula use ($P = 0.019$).

Table 4.2. Association between PGCT and clinical co-variates in healthy and NEC preterm infants. Global *P*-values and aORs with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect.

	PGCT-1		PGCT-2		PGCT-3		PGCT-4		PGCT-5	
	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	0.894 (0.782 - 1.099)	0.288	0.843 (0.675 - 1.054)	0.135	0.932 (0.638 - 1.360)	0.714	1.804 (0.762 - 4.273)	0.180	1.221 (0.964 - 1.547)	0.099
Birthweight	1.000 (1.000 - 1.001)	0.785	1.000 (1.000 - 1.002)	0.185	0.999 (0.997 - 1.002)	0.497	0.998 (0.992 - 1.004)	0.460	1.000 (0.999 - 1.002)	0.791
Birth mode	1.052 (0.584 - 1.895)	0.866	1.137 (0.603 - 2.145)	0.691	1.253 (0.409 - 3.838)	0.694	0.240 (0.020 - 2.928)	0.263	0.858 (0.427 - 1.727)	0.668
Sex	1.446 (0.856 - 2.445)	0.168	0.583 (0.329 - 1.032)	0.064	0.806 (0.294 - 2.208)	0.674	0.690 (0.084 - 5.646)	0.729	1.103 (0.591 - 2.058)	0.758
Antibiotics		<0.001		0.521		0.017		0.022		0.050
No	1		1		1		1		1	
Yes	1.021 (1.005 - 1.038)		0.898 (0.645 - 1.248)		0.612 (0.408 - 0.917)		0.413 (0.194 - 0.879)		0.614 (0.376 - 1.000)	
Day full feeds	1.021 (1.005 - 1.038)	0.012	0.990 (0.971 - 1.008)	0.264	0.973 (0.943 - 1.004)	0.087	0.984 (0.908 - 1.066)	0.690	0.996 (0.972 - 1.020)	0.734
MOM		<0.001		0.217		0.007		0.526		0.002
Never	1		1		1		1		1	
Before	0.048 (0.007 - 0.323)		0.493 (0.031 - 7.848)		5.352 (0.135 - 212)		0.000001 (0 - inf)		121 (9.466 - 1547)	
During	0.044 (0.013 - 0.148)		2.934 (0.762 - 10.53)		26.41 (2.093 - 333)		6.958 (0.012 - 4017)		6.536 (1.364 - 31.32)	
After	0.065 (0.019 - 0.224)		2.443 (0.634 - 9.419)		9.990 (0.768 - 130)		18.1 (0.032 - 10146)		7.629 (1.547 - 37.62)	
BMF		0.037		0.077		0.296		0.099		<0.001
Never	1		1		1		1		1	
Before	1.981 (1.053 - 3.727)		0.637 (0.320 - 1.270)		0.762 (0.238 - 2.448)		0.168 (0.015 - 1.870)		0.465 (0.182 - 1.188)	
During	1.278 (0.658 - 2.482)		0.506 (0.249 - 1.026)		0.803 (0.251 - 2.568)		0.635 (0.060 - 6.694)		1.739 (0.797 - 3.793)	
After	2.227 (0.898 - 5.522)		0.289 (0.111 - 0.752)		0.374 (0.098 - 1.436)		0.789 (0.055 - 11.26)		1.140 (0.460 - 2.825)	
Formula		0.234		0.754		0.157		0.029		0.019
Never	1		1		1		1		1	
Before	1.019 (0.580 - 1.794)		0.877 (0.477 - 1.613)		2.363 (0.799 - 6.992)		0.433 (0.044 - 4.282)		0.861 (0.412 - 1.801)	
During	0.766 (0.393 - 1.491)		0.701 (0.352 - 1.397)		2.807 (0.902 - 8.732)		1.400 (0.131 - 14.93)		1.511 (0.738 - 3.114)	
After	2.553 (0.655 - 9.947)		0.664 (0.133 - 3.318)		8.736 (1.178 - 64.782)		0.007 (0.0001 - 0.472)		0.117 (0.016 - 0.865)	
NEC		0.003		0.574		0.269		0.073		<0.001
No-NEC	1		1		1		1		1	
Pre-NEC	1.301 (0.655 - 2.585)		0.669 (0.304 - 1.469)		2.494 (0.713 - 8.722)		0.146 (0.010 - 2.224)		0.341 (0.112 - 1.056)	
Post-NEC	3.015 (1.486 - 6.115)		0.698 (0.314 - 1.548)		2.716 (0.796 - 9.264)		0.043 (0.003 - 0.738)		0.089 (0.032 - 0.250)	
Probiotics		0.107		<0.001		0.017		<0.001		0.004
None	1		1		1		1		1	
Infloran	1.338 (0.621 - 2.882)		0.157 (0.065 - 0.038)		0.123 (0.028 - 0.538)		11832 (193 - 725297)		0.459 (0.175 - 1.205)	
Labinic	2 (1.021 - 3.918)		0.595 (0.299 - 1.185)		0.472 (0.134 - 1.662)		0.996 (0.979 - 1.013)		1.893 (0.868 - 4.128)	
DOL	0.955 (0.945 - 0.966)	<0.001	0.989 (0.979 - 0.999)	0.032	1.018 - (1.007 - 1.030)	0.002	0.996 (0.979 - 1.013)	0.637	1.051 (1.039 - 1.064)	<0.001

Microbiome development over the initial weeks of life was next assessed, which represents the critical window preceding NEC diagnosis. This was analysed based on the number of unique PGCTs an infant transitions into, with infants who remained in the same PGCT throughout this period being classed as “stable”, and those transitioning to two or more unique PGCTs being classed as “dynamic”. Infants with three or more samples in the first two weeks of life were analysed (n = 91), to see whether NEC or other clinical variables were significantly associated with this measure. Infants who developed NEC were significantly more likely to remain in the same PGCT (i.e., stable) in the first two weeks of life (P = 0.044). These observations were irrespective of the number of samples from that infant within the 2-week period (P = 0.757).

To further explore the impact of NEC and other clinical variables on microbiome maturation and development, microbiota age and MAZ scores were calculated, as previously described (Subramanian et al., 2014). In brief, a model containing 19 age-discriminatory species was used to predict the DOL of samples not used in the training dataset (see Methods). This predicted DOL is referred to as the microbiota age. MAZ scores were generated using the microbiota age of a sample and the median microbiota age from DOL-matched infants in the training dataset. Both the microbiota age and MAZ score was significantly lower in infants who developed NEC (P = 0.003 and P < 0.001, respectively; **Figure 4.6A, B, Table 4.3**). Specifically, MAZ scores were significantly lower in post-NEC samples compared to both no-NEC samples (P < 0.001) and pre-NEC samples (P < 0.001), but not pre-NEC samples compared to no-NEC samples (P = 0.679). This association with NEC remained significant independent of antibiotic use in the previous 7 days (P < 0.001). For microbiota age in the unmatched dataset, NEC cases could not be split into pre-NEC and post-NEC as time is strongly associated with changes in the gut microbiota and pre-NEC and post-NEC are nested in time (i.e., pre-NEC are generally earlier samples whilst post-NEC are generally later samples). Aside from NEC, antibiotic use in the previous 7 days was significantly associated with a decreased microbiota age and MAZ score (P < 0.001 and P = 0.017, respectively) and MOM, BMF and formula use were also significantly associated with both microbiota age and MAZ score (all P < 0.05; **Table 4.3**). Furthermore, increasing gestational age was significantly associated with an increased

MAZ score (P = 0.009) whilst increasing day of full feeds were significantly associated with a decreased MAZ score (P = 0.020; **Table 4.3**).

To further explore the association between microbiota age and NEC onset (i.e., pre-NEC only compared to DOL-matched controls), and to better control for potential confounding variables, the analysis was repeated on a subset of NEC infants who had been matched 1:1 to healthy control infants based on gestational age and birthweight, and their samples matched by DOL (see Methods; **Table 4.4**). Other patient-level clinical data, such as the number of infants from each probiotic group are also described in **Table 4.4**. Based on this stratified analysis of samples prior to NEC diagnosis and healthy matched controls, there was no significant difference in microbiota age (P = 0.376) or MAZ score (P = 0.171). Aside from NEC, BMF and formula use were also significantly associated with both microbiota age and MAZ score (all P < 0.001). Additionally, MOM and antibiotic use in the previous 7 days were both significantly associated with microbiota age (P < 0.001) and gestational age at birth was significantly associated with MAZ score (P = 0.008; **Figure 4.6C, D, Table 4.3**).

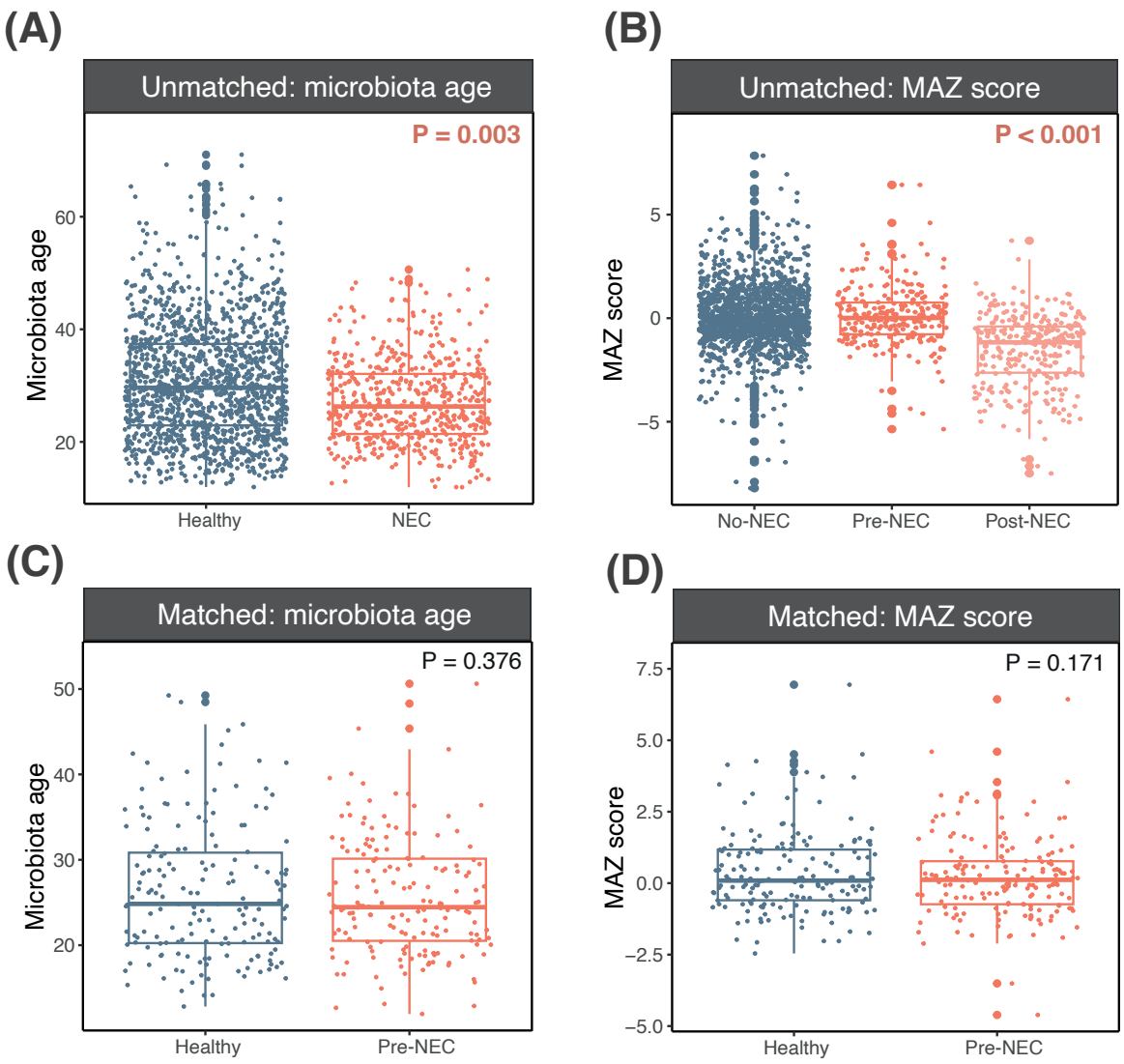

Figure 4.6. MAZ score and microbiota age are significantly lower post-NEC but not pre-NEC. **(A)** microbiota age in the unmatched dataset **(B)** MAZ score in the unmatched dataset, **(C)** microbiota age in the matched dataset and **(D)** MAZ score in the matched dataset. *P*-values are based on LMMs adjusted for gestational age, birthweight, birth mode, sex, season, day of full feed, BMF, formula, probiotics, DOL and patient ID.

Table 4.3. Association between MAZ score and microbiota age with clinical covariates. Wald's Chi-squared test statistic for fixed effects are based on Type II ANOVA on the fitted LMMs, with patient ID as a random effect.

	Unmatched: microbiota age		**Unmatched: MAZ score**		**Matched: microbiota age**		**Matched: MAZ score**	
	Chisq	pval	Chisq	pval	Chisq	pval	Chisq	pval
Gestational age	0.507	0.477	6.865	0.009	0.035	0.851	0.718	0.397
Birthweight	0.501	0.479	2.770	0.096	0.494	0.482	2.263	0.133
Birth mode	0.370	0.543	1.030	0.310	0.013	0.909	0.157	0.692
Sex	0.001	0.972	0.199	0.656	0.735	0.391	0.150	0.700
Antibiotics	242.103	<0.001	5.678	0.017	17.475	<0.001	5.015	0.025
Day full feeds	0.033	0.857	5.413	0.020	1.738	0.187	2.646	0.104
MOM	29.028	<0.001	25.288	<0.001	36.537	<0.001	14.742	0.002
BMF	230.505	<0.001	30.888	<0.001	20.680	<0.001	21.510	<0.001
Formula	32.921	<0.001	17.440	<0.001	6.857	0.077	5.324	0.150
NEC	8.562	0.003	157.763	<0.001	0.784	0.376	1.872	0.171
Probiotics	2.721	0.257	3.610	0.165	4.741	0.093	2.680	0.262

Table 4.4. Clinical data for the matched data analysis. *P* values are based on Chi-squared test for categorical data and Kruskal-Wallis test for continuous data.

	Healthy	**NEC**	**P-value**
No. of infants	60	60	
No. of samples	172	172	
Median gestational age (IQR)	26 (25.0 - 27.9)	25.8 (24.4 - 27.9)	0.505
Median birthweight (g) (IQR)	785 (620 - 944)	778 (600 - 882)	0.395
Delivery mode			1
Caesarean	30 (50%)	31 (51.7%)	
Vaginal	30 (50%)	29 (48.3%)	
Sex			0.143
Male	28 (46.7%)	37 (61.7%)	
Female	32 (53.3%)	23 (38.3%)	
Probiotic type			0.058
No probiotic	14 (23.3%)	7 (11.7%)	
Infloran	14 (23.3%)	25 (41.7%)	
Labinic	32 (53.3%)	28 (46.7%)	

Temporal stability was further explored by investigating the impact of NEC (i.e., healthy vs NEC) on intra and inter-individual gut microbiome variation based on Bray-Curtis dissimilarity. There was greater temporal variation within infants who developed NEC compared to those who did not ($P = 0.034$) as well as between those infants ($P < 0.001$; **Figure 4.7A, B**). To further explore whether this was the case prior to diagnosis and

not just a consequence of disease, the same analysis was conducted on the 1:1 longitudinal matched dataset as above. There was no significant difference in intra-individual variability prior to NEC diagnosis (i.e., pre-NEC) compared to matched controls ($P = 0.581$; **Figure 4.7C**). However, there was significantly less inter-individual variability between infants who developed NEC, prior to diagnosis, compared to matched controls ($P < 0.001$; **Figure 4.7D**).

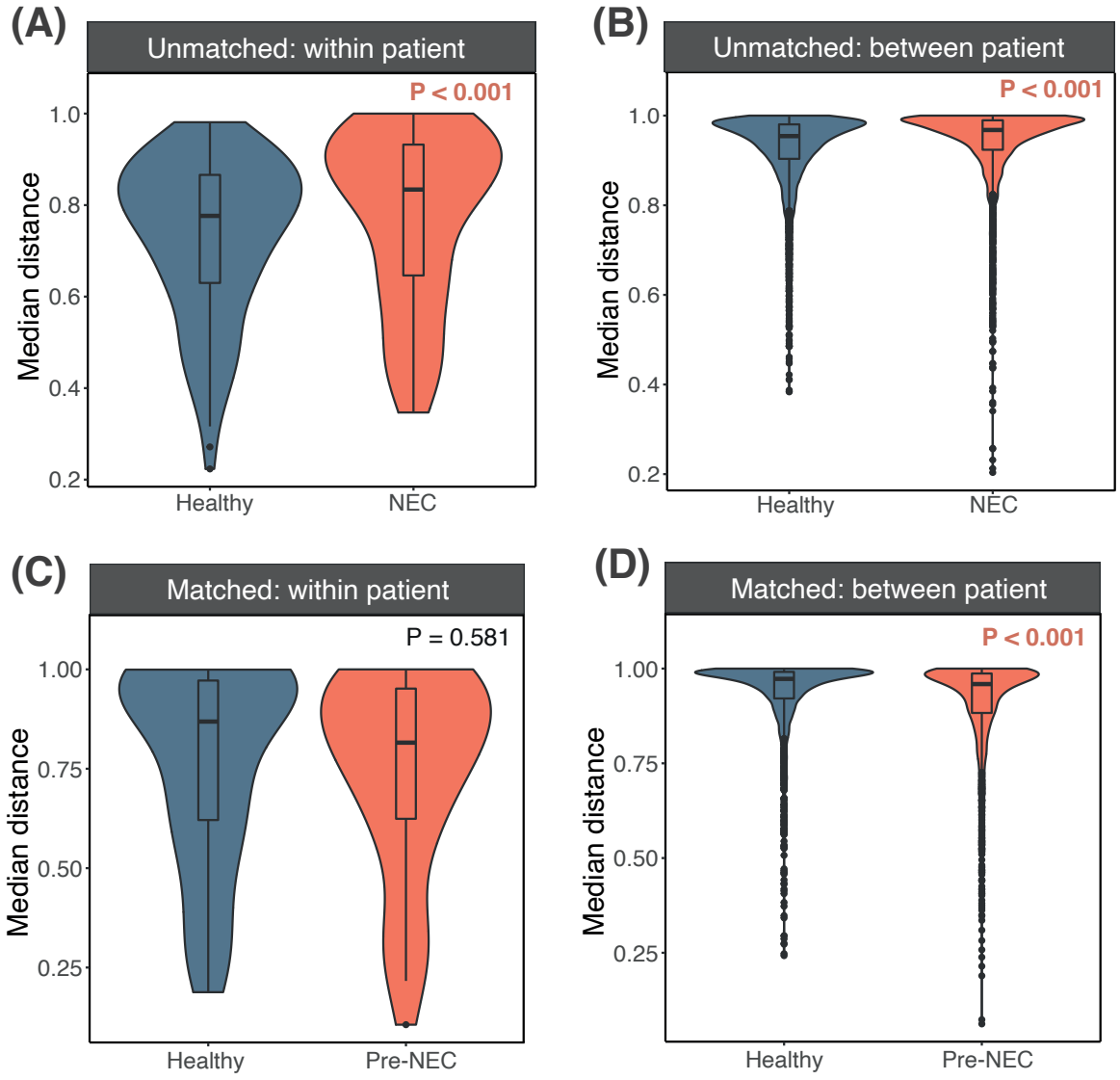

Figure 4.7. Gut microbiome community dissimilarity in healthy and NEC infants. Median Bray-Curtis dissimilarity between samples. **(A)** for the entire cohort (both pre-NEC and post-NEC samples) within patients, **(B)** for the entire cohort (both pre-NEC and post-NEC samples) between patients, **(C)** for the matched 1:1 longitudinal cohort (pre-NEC samples) within patients and **(D)** for the matched 1:1 longitudinal cohort (pre-NEC samples) between patients. P-values are based on Wilcoxon test.

4.3.4. Alpha and beta diversity of the preterm gut microbiome in NEC infants prior to and following diagnosis

Since previous studies have found an association between gut microbial diversity and NEC onset, the impact of NEC (i.e., no-NEC/pre-NEC/post-NEC) on alpha diversity measures was next explored, whilst controlling for confounders. Both species richness and Shannon diversity were significantly associated with NEC ($P < 0.001$). Post-hoc analysis revealed that samples collected post-NEC had a significantly lower Shannon diversity than those collected pre-NEC ($P < 0.001$) or from infants who never developed NEC ($P < 0.001$). However, there was no significant difference in Shannon diversity between pre-NEC and no-NEC samples ($P = 0.780$). On the other hand, species richness was significantly lower in pre-NEC samples ($P < 0.001$) and post-NEC ($P < 0.001$) compared to no-NEC samples, as well as lower post-NEC compared to pre-NEC ($P = 0.009$) (**Figure 4.8**, **Table 4.5**).

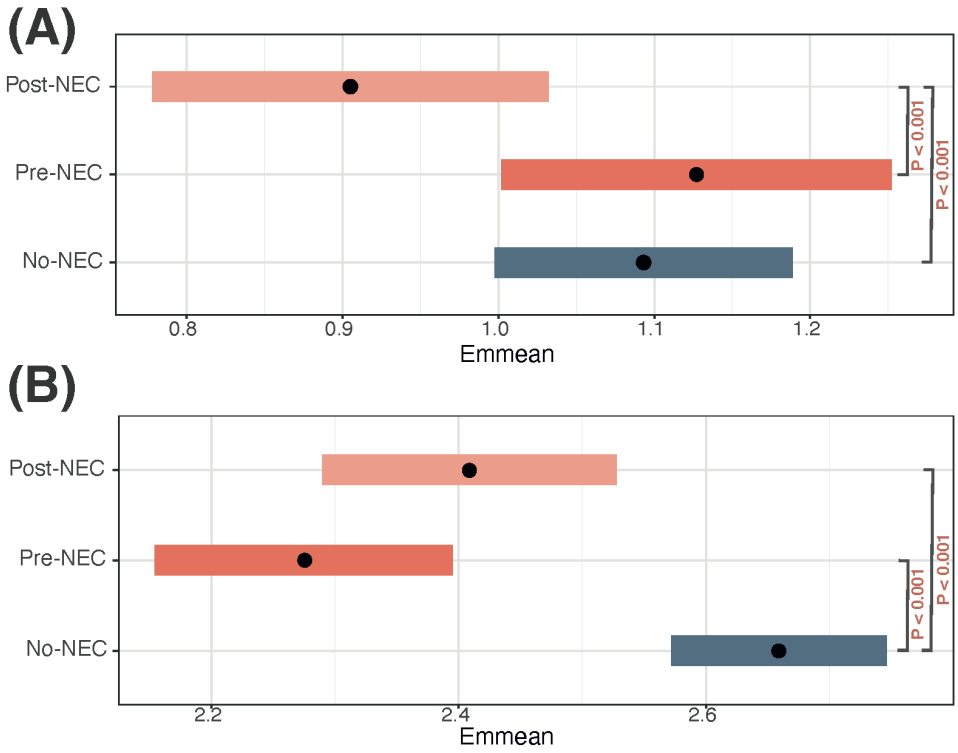

Figure 4.8. NEC is significantly associated with alpha diversity measures. Estimated marginal means (95% CIs) representing **(A)** Shannon diversity for NEC, and **(B)** species richness for NEC obtained from linear mixed-effects models adjusted for gestational age, birthweight, birth mode, sex, season, day of full feed, BMF, formula, probiotics, DOL and patient ID. The statistical significance shown is after adjustment for multiple comparisons using the two-tailed Dunnett's method, whereby 'No-NEC' was used as the control.

Table 4.5. Association of clinical co-variates with Shannon diversity and richness. Global *P*-values and Wald's Chi-squared test statistic for fixed effects are based on Type II ANOVA on the fitted LMMs, with patient ID as a random effect.

	Shannon		**Richness**	
	Chisq	pval	Chisq	pval
DOL	74.371	<0.001	196.616	<0.001
Gestational age	6.266	0.012	2.508	0.113
Birthweight	0.186	0.666	0.028	0.866
Birth mode	2.252	0.133	2.815	0.093
Sex	0.019	0.891	0.258	0.611
Antibiotics	55.326	<0.001	28.025	<0.001
Day full feeds	0.289	0.591	0.007	0.934
MOM	28.325	<0.001	23.645	<0.001
BMF	17.489	<0.001	10.221	0.017
Formula	2.539	0.468	2.714	0.438
NEC	25.509	<0.001	61.464	<0.001
Probiotics	7.999	0.018	18.942	<0.001

To determine whether NEC, or other co-variates were associated with overall bacterial profiles, PERMANOVA was performed using 'adonis'. DOL explained 2.3% variation in bacterial profiles of healthy infants, compared to 1% variation in infants who developed NEC (both $P < 0.001$). Similarly, unique patient identifier explained 3.8% in healthy infants, compared to 2.1% in infants who developed NEC (both $P < 0.001$). Data was split into cross-sectional time-points based on DOL, each with one sample per patient, as done in Chapter 3. Similar to the results from Chapter 3, probiotics were the most significant co-variate, associated at every timepoint after DOL 0-9, and MOM, BMF and antibiotics were also significantly associated at various timepoints (**Figure 4.9A**). In addition, birthweight and gestational age were both significantly associated at various timepoints, and day of full feeds and delivery mode were significantly associated at one timepoint. NEC status was significantly associated at DOL 15-19 only, which is around the median DOL onset for NEC (DOL 19) in this cohort (**Figure 4.9A**).

The 'adonis' analysis was then stratified by control vs NEC. As previous, probiotics were the main driver of the gut microbiome in healthy infants being the most significant factor associated with overall bacterial profiles, significant at every time-point (all FDR = 0.001) except DOL 0-9. There were also additional clinical co-variates significantly associated at various time-points, not identified in Chapter 3 (**Figure 4.9B**). In contrast,

for infants who developed NEC, no clinical co-variates were significantly associated with gut bacterial profiles at any time-point, including probiotics (**Figure 4.9C**).

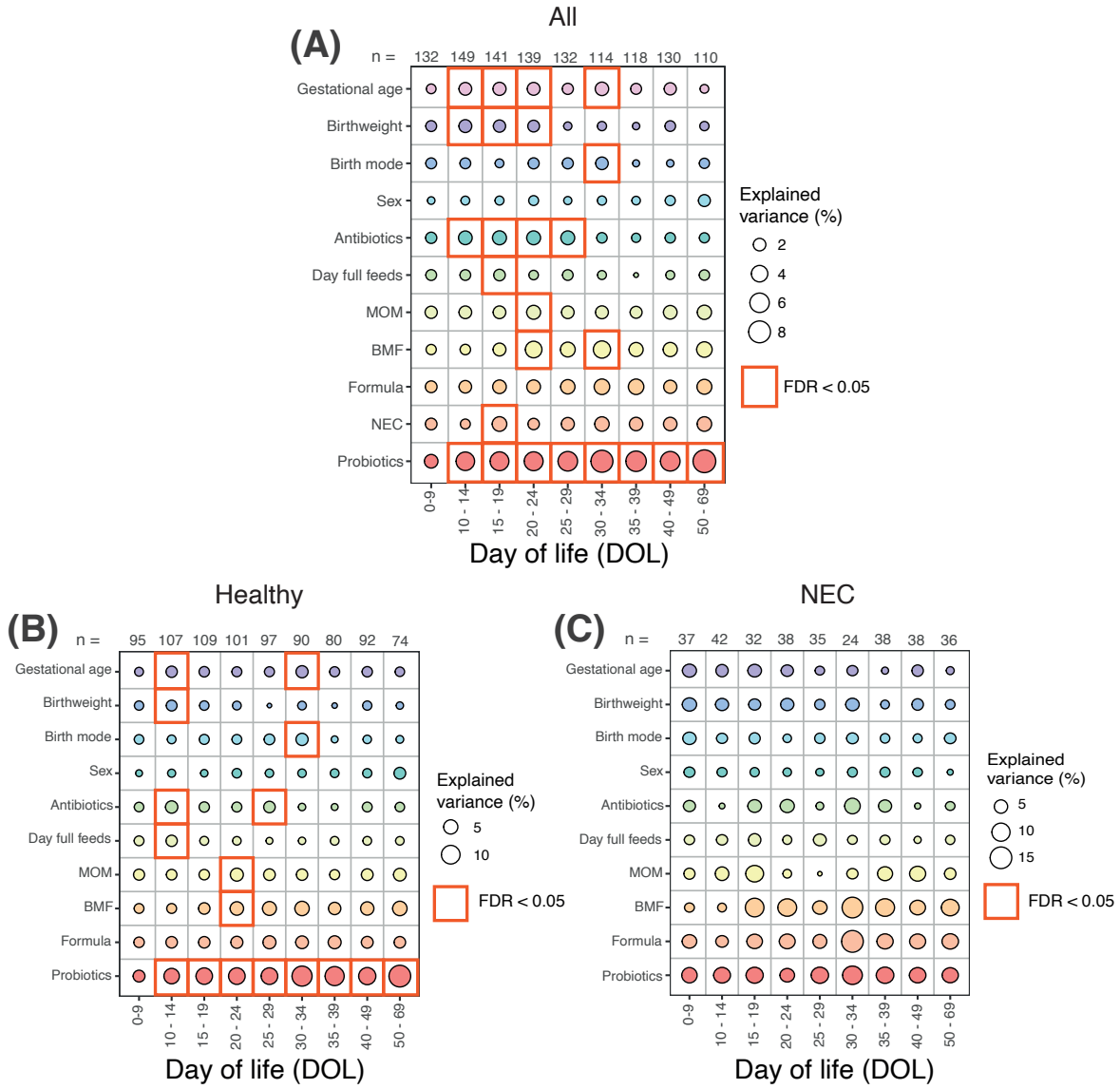

Figure 4.9. Significance and explained variance of clinical co-variates at different timepoints based on taxonomic profiles, modelled by 'adonis'. Bubbles show the amount of variance (R^2) explained by each co-variate at a given timepoint and significant results (FDR<0.05) are surrounded by a red box for **(A)** both healthy and NEC infants **(B)** healthy infants only and **(C)** NEC infants only.

4.3.5. Specific taxonomic signatures of the preterm gut microbiome in NEC infants prior to and following diagnosis

To understand whether there were more specific microbial signatures of NEC, taxonomic analyses were conducted to look for associations at the phylum, genus and species level. At the phylum level, the relative abundance of Proteobacteria was

significantly higher both pre-NEC (P = 0.015, Q = 0.078) and post-NEC (P = 0.002, Q = 0.014) compared to samples from infants who did not develop NEC, whilst the relative abundance of Actinobacteria was significantly lower both pre-NEC (P = 0.011, Q = 0.062) and post-NEC (P < 0.001, Q < 0.001). At the genus level, there was only one significant feature associated with pre-NEC samples compared to samples from infants who did not develop NEC. *Bifidobacterium* relative abundance was significantly lower pre-NEC onset (P = 0.024, Q = 0.178), whilst controlling for confounders such as DOL, probiotics and unique patient identifier. At the species level, *B. animalis* (P = 0.049, Q = 0.23) and *B. longum* (P = 0.035, Q = 0.186) were significantly lower in pre-NEC samples, whereas *L. rhamnosus* (P = 0.022, Q = 0.144) and *K. variicola* (P = 0.056, Q = 0.246) relative abundance were higher in pre-NEC samples, based on the default threshold of Q < 0.25. There were also a number of features identified as significantly associated with post-NEC samples compared to samples from infants who did not develop NEC. It is important to note, however, that there are differences in the number of control vs NEC samples across probiotic eras (**Figure 4.10**), which could be contributing to the results observed for the Labinic-associated probiotic species, *B. animalis* and *B. longum.* As reported in Chapter 3, *B. animalis* and *B. longum* are present in Labinic and their prevalence and abundance are closely linked to the receipt of this probiotic in healthy infants. As there were approximately twice as many Labinic healthy infants than NEC infants (**Figure 4.10**), the analysis was repeated including just 33 randomly selected healthy Labinic infants to match the number of Labinic NEC infants. The differences in the relative abundance of *B. animalis* and *B. longum* were no longer observed.

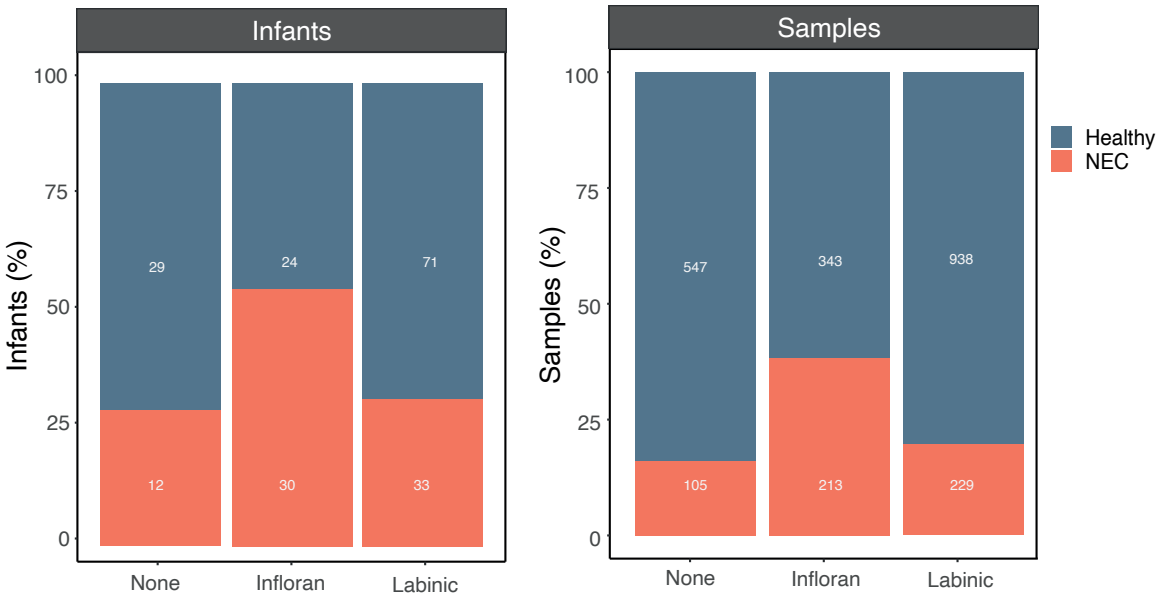

Figure 4.10. Proportion of healthy vs NEC infants and samples across probiotic groups. The raw number of infants and samples in each group is shown.

4.3.6. Investigating the gut microbiome of preterm infants before NEC diagnosis in a cross-sectional and longitudinal 1:1 matched case control cohort

To further probe whether any microbial signatures could be identified around the time of NEC onset, infants from the 1:1 matched dataset were analysed (**Table 4.4**). This analysis was first performed on cross-sectional data, using a single sample per patient close to the day of NEC onset (see Methods) and a matched control sample. Based on this, there was no significant difference in the relative abundance profiles using Bray-Curtis dissimilarity (P = 0.920; **Figure 4.11A**). Following this, it was determined whether there was a significant difference dependent on medically managed NEC and those that underwent surgery (i.e., a recognised proxy for NEC severity). There was also no significant difference in relative abundance profiles when stratifying by NEC severity (P = 0.836; **Figure 4.11B**). Nor was there a significant difference when looking at 'early NEC vs control' (P = 0.473; **Figure 4.11C**) or 'late NEC vs control' (P = 0.182; **Figure 4.11D**). To see whether there was a difference based on more rare taxa, the analysis was repeated using Jaccard distance (i.e., presence or absence of taxa instead of relative abundance). This had little impact and again there was no significant difference in overall profiles in NEC vs control (P = 0.882), NEC severity vs control (P

= 0.808), early NEC (P = 0.471) or late NEC (P = 0.193). Further to this, no species, genus or phylum was significantly associated with NEC or control, even when further stratifying by severity and early/late onset.

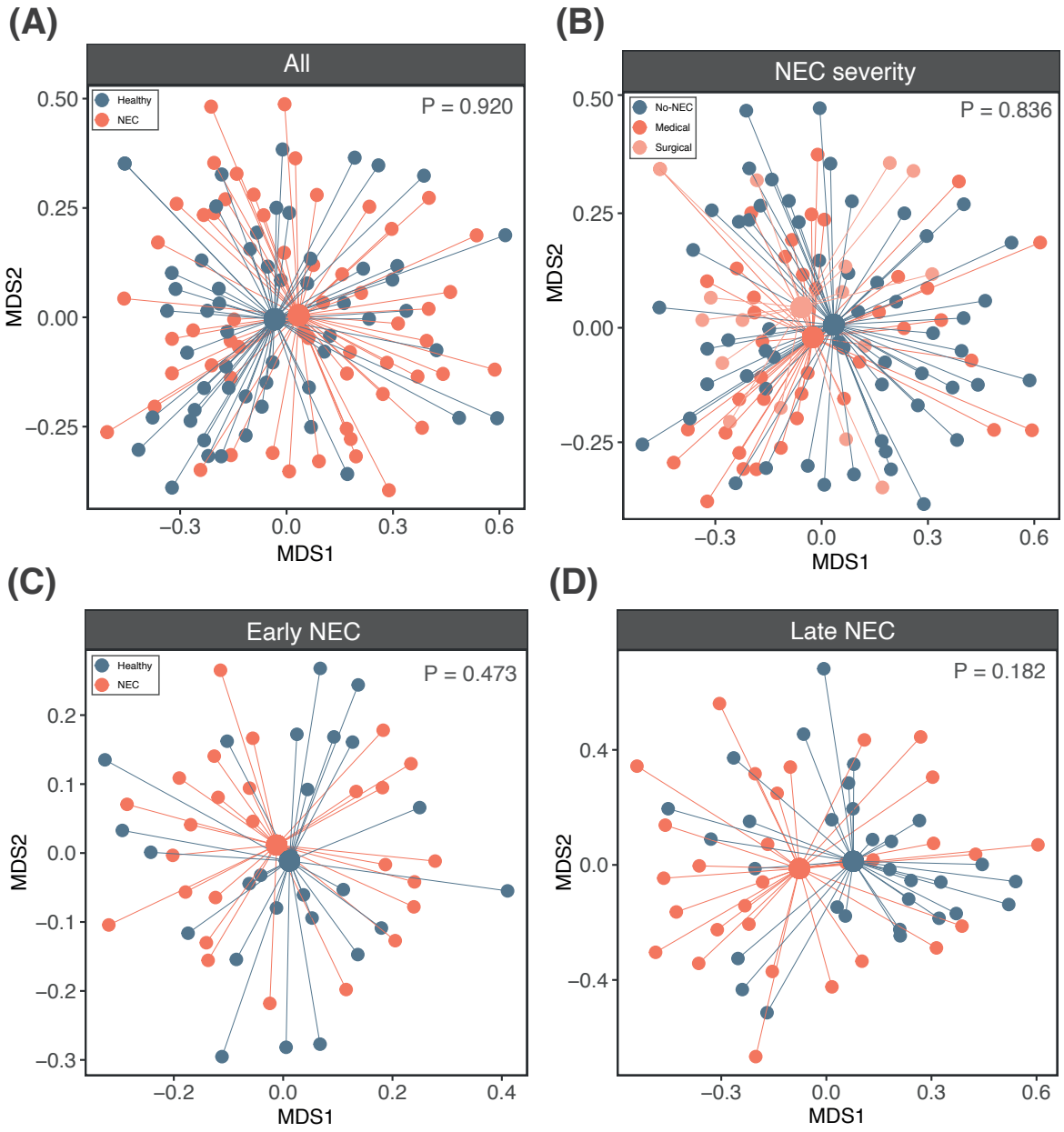

Figure 4.11. There is no significant difference in overall gut bacterial profiles between NEC and control infants at the time of NEC diagnosis based on the cross-sectional 1:1 matched case control cohort. NMDS plots based on Bray-Curtis dissimilarity with the mean centroid plotted and coloured based on **(A)** healthy vs NEC, **(B)** healthy vs medical or surgical NEC, **(C)** early NEC and **(D)** late NEC. *P*-values were calculated using Envfit.

The matched patient analysis was then extended to include multiple samples from each NEC patient prior to disease onset (see Methods), and corresponding control samples matched to the DOL, as used in previous analyses (**Table 4.4**). Based on multiple pre-NEC samples per patient, and matched control samples, there was a significant difference in overall bacterial profiles ($P < 0.001$), whilst controlling for repeated measures. There was also a significant difference when stratifying by NEC severity ($P < 0.001$), early NEC onset ($P = 0.011$) and late NEC onset ($P < 0.001$). As with the previous cross-sectional analysis, no specific genera were significantly associated with NEC or control, even when further stratifying by NEC severity. However, consistent with the non-matched case/control analysis, the relative abundance of the Proteobacteria phylum was significantly higher in all NEC vs control samples ($P = 0.012$, $Q = 0.096$), whilst Actinobacteria was significantly lower in all NEC vs control samples ($P = 0.021$, $Q = 0.125$). Also similar to the non-matched control analysis, *K. variicola* was significantly higher in all NEC vs control samples ($P = 0.010$, $Q = 218$), as well as in the stratified analysis for NEC severity, with the relative abundance of *K. variicola* being significantly higher in medical NEC samples specifically compared to control ($P = 0.003$, $Q = 0.138$), but not surgical NEC samples ($P = 0.121$, $Q = 0.653$). Regarding stratified analysis for 'early NEC vs control' and 'late NEC vs control', the relative abundance of *Streptococcus* genus was significantly lower in early NEC vs control samples ($P < 0.001$, $Q = 0.049$). No other phyla or species were significantly associated with 'early NEC vs control' or 'late NEC vs control'.

Relating to the onset of NEC, microbial patterns were identified, with Proteobacteria and specifically *K. variicola* significantly increasing in the days leading up to NEC diagnosis whilst controlling for patient ID (both $P < 0.001$; **Figure 4.12**). Actinobacteria and *Bifidobacterium* significantly decreased in the days leading up to NEC diagnosis whilst controlling for patient ID (both $P = 0.022$; **Figure 4.12**), but specific *Bifidobacterium* species, *B. animalis* and *B. longum*, identified in the MaAslin2 analysis, did not ($P = 0.210$ and $P = 0.401$, respectively).

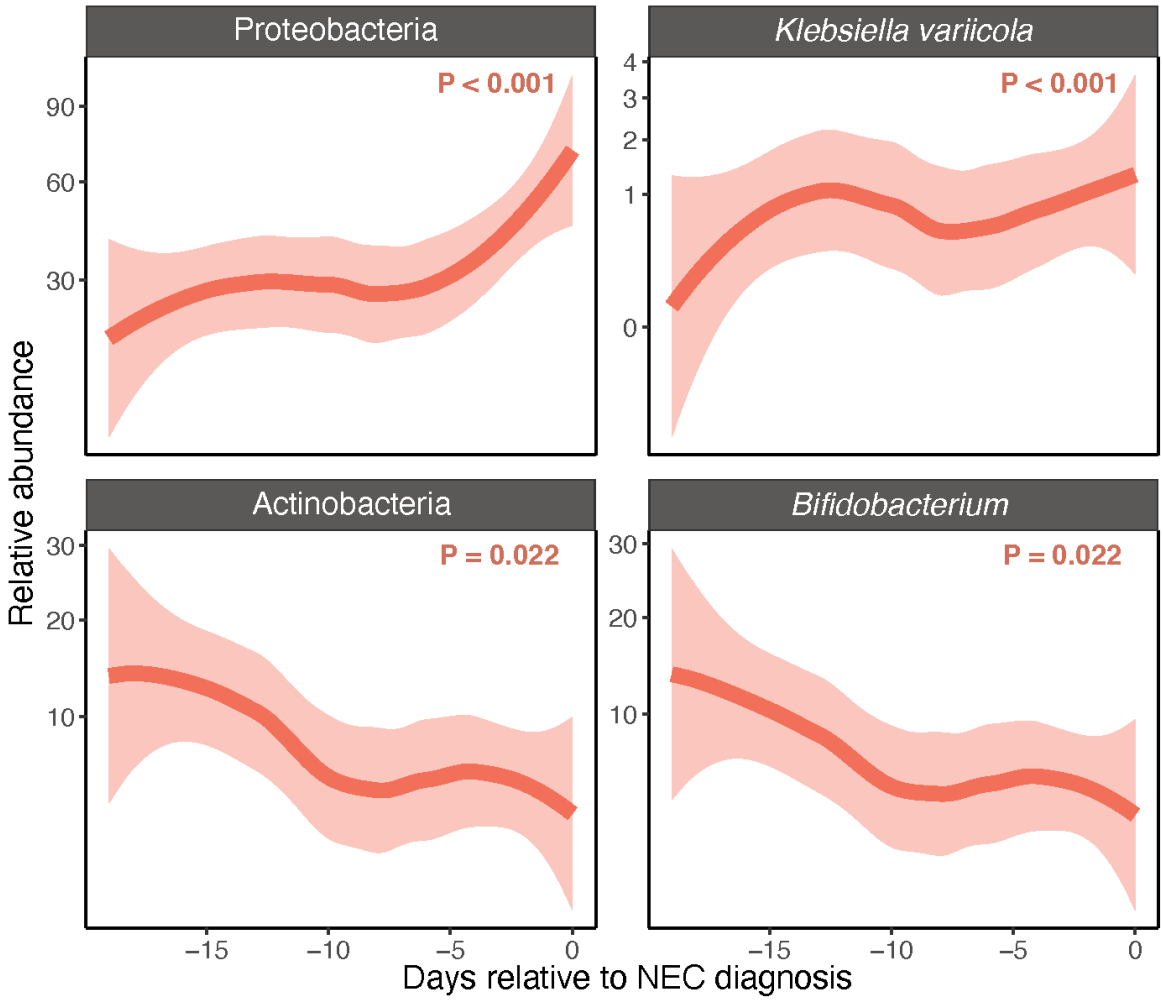

Figure 4.12. Microbial signatures of NEC in the days leading up to diagnosis. The relative abundance of Proteobacteria, *K. variicola*, Actinobacteria and *Bifidobacterium* identified in the MaAslin2 analysis to be significantly higher in pre-NEC samples compared to no-NEC samples, over the 20 days preceding NEC diagnosis. *P*-values are based on linear mixed effects models where patient ID was included as a random effect.

4.3.7. Microbial signatures of NEC prior to and following diagnosis in the context of probiotic use

Since probiotics were significantly associated with the gut microbiome of healthy infants, but not infants who developed NEC (**Figure 4.9B, C**), the next aim was to further explore the association between the gut microbiome and NEC in the context of probiotic use.

Binomial mixed model analysis was used to determine which clinical co-variates were significantly associated with each PGCT, stratifying by healthy vs NEC infants to

investigate whether there was a differential impact of probiotics in these infants. Healthy control infants who received Labinic were significantly more likely to transition into PGCT-5 than those who did not receive any probiotic (P = 0.006). However, NEC infants who received Labinic, were not more likely to transition into PGCT-5 than infants never exposed to probiotics (P = 0.230) (**Figure 4.13**). In fact, of all the infants who received Labinic, only 6.6% of samples from NEC infants were classified as PGCT-5, compared to 23.6% of samples from healthy infants. The same was not true for infants who received Infloran, whereby infants were significantly more likely to transition into PGCT-4 compared to infants who received no probiotic (P < 0.001), regardless of whether or not they developed NEC (**Figure 4.13**).

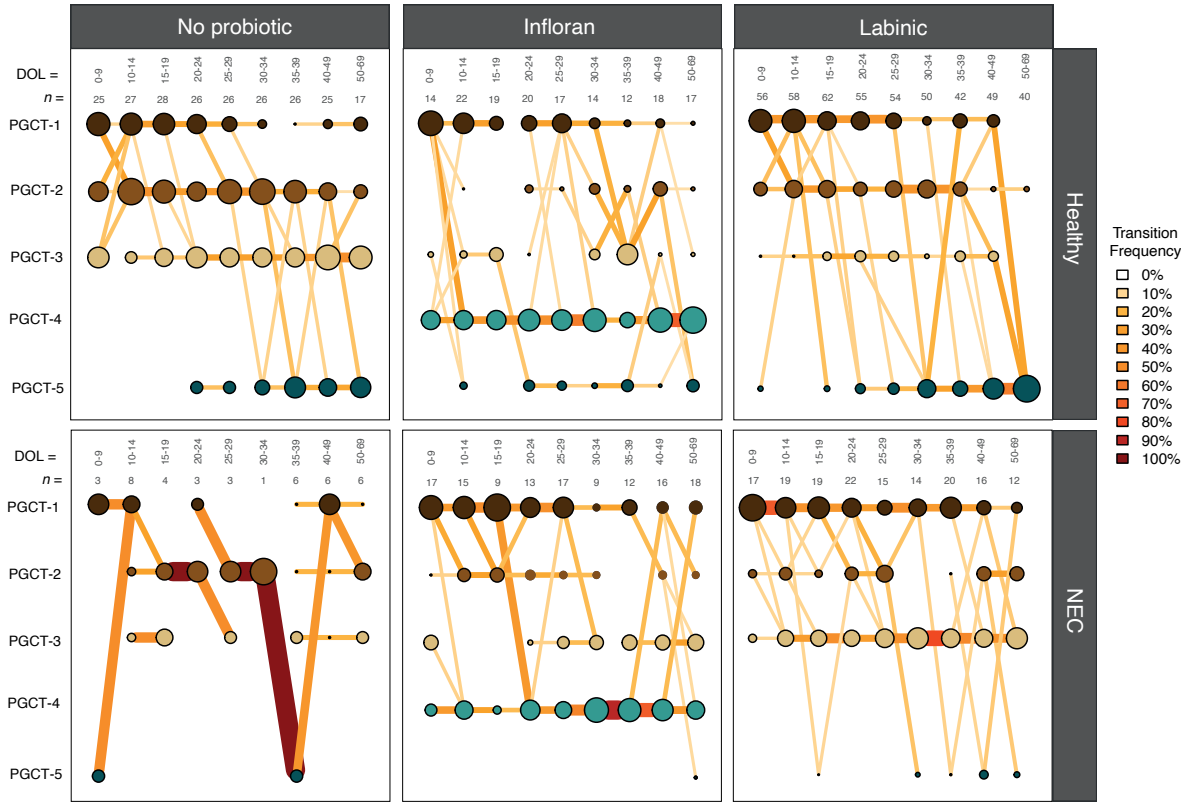

Figure 4.13. NEC differentially impacts the transition of the preterm gut microbiome over time, dependent on probiotic use. Transition plot showing the progression of samples through each PGCT from DOL 0 to DOL 69, within each probiotic group based on whether an infant developed NEC or not. The nodes and edges are sized based on the total counts; nodes are coloured according to PGCT and edges by the transition frequency.

Given the differential impact of Labinic on NEC infants, the analysis was next stratified by probiotic type. This allowed further investigation into the impact of NEC within probiotic groups, and whether differences were observed prior to diagnosis (i.e., pre-

NEC) or as a consequence of diagnosis (i.e., post-NEC). Only one sample from the no probiotic infants and two samples from the Labinic infants were classed as PGCT-4, supporting the significantly strong association observed between Infloran and PGCT-4. Therefore, a PGCT-4 model could not be built for no probiotic or Labinic infants. Similar to the un-stratified analysis, no probiotic infants were significantly more likely to transition into PGCT-1 following NEC diagnosis ($P < 0.001$) compared to infants who did not develop NEC (**Table 4.6**). Despite the strong association between PGCT-4 and Infloran infants (40.7% of Infloran samples), those who developed NEC were also significantly more likely to be in PGCT-1 prior to and following diagnosis, compared to those who did not ($P = 0.03$) (**Table 4.7**). This was not true for infants who received Labinic, where there was no significant association between transition into PGCT-1 and NEC ($P = 0.0818$; **Table 4.8**). Furthermore, in keeping with the un-stratified analysis, the association observed between transition into PGCT-5 and NEC infants was again observed in no probiotic infants and those who received Labinic. Specifically, infants who received no probiotic or received Labinic were significantly less likely to transition into PGCT-5 following NEC diagnosis compared to infants who did not develop NEC ($P = 0.048$ and $P = 0.001$, respectively; **Table 4.6 and 4.8**). In infants who received Infloran, there was no significant association between transition into PGCT-4 and NEC ($P = 0.106$). Consequently, although there were more NEC infants that associated with PGCT-1 (**Table 4.7**), there was no significant impact on transition into PGCT-4, consistent with the stratified analysis of healthy and NEC infants. Finally, compared to healthy controls, infants who received Labinic were significantly more likely to be in PGCT-3 both prior to NEC diagnosis and following diagnosis ($P = 0.029$; **Table 4.8**), which was not the case for infants who received Infloran ($P = 0.288$; **Table 4.7**), and infants never exposed to probiotics ($P = 0.098$; **Table 4.6**).

Table 4.6. Association between PGCT and clinical co-variates in infants who never received probiotics. Global *P*-values and aORs with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect. A model could not be fit for PGCT-4 as only 1 sample was classified as PGCT-4.

	PGCT-1		PGCT-2		PGCT-3		PGCT-5	
	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	1.032 (0.656 - 1.624)	0.891	0.794 (0.526 - 1.199)	0.272	0.915 (0.424 - 1.975)	0.821	1.246 (0.654 - 2.377)	0.504
Birthweight	0.998 (0.994 - 1.001)	0.145	1.003 (1 - 1.006)	0.037	1.001 (0.995 - 1.007)	0.803	0.997 (0.993 - 1.002)	0.316
Birth mode		0.025		0.345		0.337		0.447
Vaginal	1		1		1		1	
Caesarean	0.208 (0.053 - 0.824)		1.847 (0.516 - 6.604)		3.213 (0.296 - 34.855)		0.391 (0.035 - 4.397)	
Sex		0.021		0.478		0.382		0.722
Male	1		1		1		1	
Female	0.209 (0.055 - 0.79)		1.586 (0.443 - 5.671)		2.971 (0.259 - 34.154)		0.687 (0.087 - 5.442)	
Antibiotics		0.677		0.678		0.633		0.027
No	1		1		1		1	
Yes	1.153 (0.59 - 2.251)		1.141 (0.613 - 2.123)		0.825 (0.374 - 1.821)		0.216 (0.055 - 0.843)	
Day full feeds	1.037 (1.006 - 1.068)	0.017	0.987 (0.96 - 1.015)	0.370	0.965 (0.909 - 1.025)	0.247	1.008 (0.958 - 1.061)	0.753
MOM		<0.001		0.173		0.236		0.998
Never	1		1		1		1	
Before	0.027 (0 - 2.226)		0.0000004 (0 – Inf)		419 (0.006 - 27436072)		0.509 (0 – Inf)	
During	0.014 (0.001 - 0.129)		2.569 (0.291 - 22.654)		90.357 (1.172 - 6968.803)		2645521 (0 - Inf)	
After	0.009 (0.001 - 0.086)		6.926 (0.808 - 59.354)		59.098 (0.814 - 4292.35)		2309063 (0 - Inf)	
BMF		0.003		0.368		0.088		0.121
Never	1		1		1		1	
Before	2.858 (0.644 - 12.677)		1.425 (0.336 - 6.045)		0.507 (0.034 - 7.658)		0.15 (0.008 - 2.844)	
During	0.63 (0.121 - 3.287)		2.777 (0.617 - 12.5)		0.405 (0.026 - 6.4)		1.023 (0.072 - 14.488)	
After	7.155 (0.866 - 59.12)		1.73 (0.218 - 13.75)		0.046 (0.002 - 1.036)		0.709 (0.037 - 13.614)	
Formula		0.510		0.112		0.059		0.313
Never	1		1		1		1	
Before	0.932 (0.266 - 3.269)		0.248 (0.073 - 0.849)		14.184 (1.279 - 157.26)		1.279 (0.132 - 12.388)	
During	0.957 (0.243 - 3.775)		0.278 (0.078 - 0.997)		7.318 (0.614 - 87.25)		4.502 (0.578 - 35.062)	
After	0.129 (0.008 - 2.12)		0.798 (0.096 - 6.642)		112.933 (1.831 - 6967.195)		0 (0 - Inf)	
NEC		<0.001		0.160		0.100		0.048
No-NEC	1		1		1		1	
Pre-NEC	0.281 (0.037 - 2.129)		6.44 (0.957 - 43.354)		1.727 (0.058 - 51.046)		0.151 (0.003 - 7.695)	
Post-NEC	7.64 (1.234 - 47.301)		3.375 (0.575 - 19.818)		12.443 (5.416 - 43.055)		0.017 (0 - 0.706)	
DOL	0.952 (0.93 - 0.975)	<0.001	0.973 (0.953 - 0.993)	0.009	1.049 (1.023 - 1.076)	<0.001	1.041 (1.012 - 1.07)	0.005

Table 4.7. Association between PGCT and clinical co-variates in infants who received Infloran. Global *P*-values and aORs with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect. MOM could not be included in the models for PGCT-2, 3 or 5.

	PGCT-1		PGCT-2		PGCT-3		PGCT-4		PGCT-5	
	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	0.688 (0.383 - 1.237)	0.212	0.644 (0.263 - 1.578)	0.336	0.948 (0.393 - 2.287)	0.906	3.034 (0.754 - 12.21)	0.118	1.3 (0.493 - 3.425)	0.596
Birthweight	1.003 (0.999 - 1.007)	0.156	1 (0.993 - 1.006)	0.880	0.995 (0.988 - 1.001)	0.113	0.997 (0.987 - 1.006)	0.473	0.998 (0.991 - 1.006)	0.653
Birth mode		0.069		0.643		0.880		0.131		0.329
Vaginal	1		1		1		1		1	
Caesarean	3.767 (0.904 - 15.70)		0.601 (0.07 - 5.154)		0.841 (0.089 - 7.933)		0.067 (0.002 - 2.23)		0.273 (0.02 - 3.699)	
Sex		0.734		0.168		0.979		0.520		0.302
Male	1		1		1		1		1	
Female	1.224 (0.381 - 3.932)		0.306 (0.057 - 1.645)		0.975 (0.156 - 6.099)		0.418 (0.029 - 5.965)		2.572 (0.428 - 15.47)	
Antibiotics		0.251		0.640		0.589		0.012		0.247
No	1		1		1		1		1	
Yes	1.443 (0.771 - 2.7)		1.235 (0.511 - 2.987)		0.801 (0.358 - 1.793)		0.376 (0.174 - 0.809)		1.972 (0.624 - 6.235)	
Day full feeds	1.008 (0.973 - 1.043)	0.668	0.972 (0.917 - 1.03)	0.335	0.999 (0.945 - 1.056)	0.971	0.983 (0.907 - 1.065)	0.670	0.998 (0.944 - 1.056)	0.956
MOM		0.053		NA		NA		0.188		NA
Never	1		NA		NA		1		NA	
Before	0.135 (0.001 - 12.47)		NA		NA		0.00001 (0 – Inf)		NA	
During	0.048 (0.001 - 1.727)		NA		NA		8.011 (0.003 - 18939)		NA	
After	0.013 (0 - 0.514)		NA		NA		44.66 (0.019 - 102799)		NA	
BMF		0.114		0.366		0.534		0.081		0.306
Never	1		1		1		1		1	
Before	3.829 (1.009 - 14.53)		0.27 (0.033 - 2.206)		3.013 (0.381 - 23.83)		0.064 (0.003 - 1.497)		0.166 (0.01 - 2.714)	
During	1.779 (0.432 - 7.33)		0.146 (0.015 - 1.369)		2.667 (0.333 - 21.36)		0.232 (0.01 - 5.298)		0.469 (0.039 - 5.704)	
After	4.432 (0.464 - 42.29)		0.111 (0.007 - 1.826)		0.679 (0.03 - 15.212)		0.175 (0.005 - 5.937)		0.158 (0.008 - 3.165)	
Formula		0.136		0.368		0.970		0.063		0.916
Never	1		1		1		1		1	
Before	1.632 (0.434 - 6.139)		6.139 (0.824 - 45.717)		0.734 (0.092 - 5.884)		0.336 (0.019 - 5.806)		1.022 (0.146 - 7.16)	
During	0.908 (0.186 - 4.447)		5.777 (0.635 - 52.547)		1 (0.105 - 9.499)		0.63 (0.03 - 13.038)		1.694 (0.228 - 12.56)	
After	16.92 (0.869 - 329.7)		0.000002 (0 – Inf)		0.0000007 (0 – Inf)		0.004 (0 - 0.429)		1.372 (0.012 - 158.4)	
NEC		0.030		0.288		0.537		0.106		0.197
No-NEC			1		1		1		1	
Pre-NEC	5.253 (1.485 - 18.59)		0.193 (0.025 - 1.518)		0.867 (0.124 - 6.075)		0.035 (0.001 - 0.866)		0.172 (0.011 - 2.66)	
Post-NEC	4.807 (1.215 - 19.02)		0.222 (0.027 - 1.847)		1.962 (0.267 - 14.41)		0.099 (0.005 - 2.083)		0.132 (0.013 - 1.288)	
DOL	0.981 (0.963 - 0.999)	0.043	0.993 (0.972 - 1.016)	0.560	0.997 (0.978 - 1.018)	0.796	0.996 (0.979 - 1.014)	0.681	1.037 (1.013 - 1.061)	0.002

Table 4.8. Association between PGCT and clinical co-variates in infants who received Labinic. Global *P*-values and aORs with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect. A model could not be fit for PGCT-4 as only 2 samples were classified as PGCT-4.

	PGCT-1		PGCT-2		PGCT-3		PGCT-5	
	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	0.838 (0.646 - 1.087)	0.183	0.926 (0.707 - 1.213)	0.577	1.022 (0.597 - 1.749)	0.936	1.22 (0.91 - 1.635)	0.184
Birthweight	1 (0.998 - 1.001)	0.554	1.001 (0.999 - 1.002)	0.383	1 (0.997 - 1.003)	0.992	1.001 (0.999 - 1.003)	0.197
Birth mode		0.669		0.516		0.760		0.956
Vaginal	1		1		1		1	
Caesarean	1.19 (0.536 - 2.64)		0.774 (0.358 - 1.674)		1.306 (0.235 - 7.249)		0.977 (0.427 - 2.235)	
Sex		0.124		0.368		0.770		0.857
Male	1		1		1		1	
Female	1.76 (0.856 - 3.62)		0.728 (0.365 - 1.453)		0.795 (0.17 - 3.712)		0.933 (0.439 - 1.983)	0.214
Antibiotics		0.004		0.173		0.028		
No	1		1		1		1	
Yes	1.867 (1.221 - 2.855)		0.731 (0.466 - 1.147)		0.497 (0.266 - 0.929)		0.68 (0.37 - 1.249)	
Day full feeds	1.046 (1.013 - 1.08)	0.006	0.948 (0.895 - 1.005)	0.072	0.973 (0.925 - 1.022)	0.277	0.991 (0.935 - 1.05)	0.753
MOM		<0.001		0.588		0.002		0.064
Never	1		1		1		1	
Before	0.021 (0.002 - 0.278)		0.656 (0.031 - 13.848)		22.223 (0.176 - 2813.158)		25.789 (0.983 - 676.402)	
During	0.052 (0.01 - 0.267)		1.92 (0.391 - 9.437)		30.785 (0.925 - 1024.513)		1.69 (0.304 - 9.387)	
After	0.37 (0.068 - 2.006)		1.187 (0.207 - 6.817)		3.018 (0.084 - 108.474)		3.737 (0.618 - 22.587)	
BMF		0.798		0.103		0.969		0.027
Never	1		1		1		1	
Before	1.035 (0.43 - 2.495)		0.935 (0.401 - 2.182)		0.668 (0.111 - 4.031)		1.05 (0.337 - 3.269)	
During	1.381 (0.564 - 3.382)		0.508 (0.217 - 1.187)		0.701 (0.122 - 4.019)		2.705 (1.049 - 6.975)	
After	1.144 (0.325 - 4.023)		0.318 (0.093 - 1.088)		0.628 (0.085 - 4.641)		1.316 (0.432 - 4.008)	
Formula		0.019		0.363		0.097		0.170
Never	1		1		1		1	
Before	0.601 (0.29 - 1.245)		1.249 (0.618 - 2.524)		2.827 (0.593 - 13.474)		0.955 (0.408 - 2.237)	
During	0.422 (0.166 - 1.068)		0.617 (0.258 - 1.478)		7.352 (1.388 - 38.931)		1.151 (0.486 - 2.723)	
After	7.755 (0.814 - 73.873)		0 (0 – Inf)		4.359 (0.194 - 97.777)		0.063 (0.004 - 0.958)	
NEC		0.818		0.675		0.029		0.001
No-NEC	1		1		1			
Pre-NEC	0.839 (0.324 - 2.172)		0.665 (0.247 - 1.791)		7.681 (1.147 - 51.447)		0.339 (0.085 - 1.359)	
Post-NEC	1.136 (0.419 - 3.079)		0.697 (0.245 - 1.985)		12.506 (1.94 - 80.619)		0.084 (0.022 - 0.323)	
DOL	0.929 (0.913 - 0.945)	<0.001	0.996 (0.982 - 1.011)	0.610	1.027 (1.007 - 1.047)	0.007	1.062 (1.045 - 1.08)	<0.001

It was next sought to look at the association between clinical data and alpha diversity stratified by each of the probiotic groups (i.e., no probiotic/Infloran/Labinic). In infants who received no probiotic, Shannon diversity was significantly lower in samples collected post-NEC than those collected pre-NEC ($P < 0.001$). However, there was no significant difference in Shannon diversity between pre-NEC and no-NEC samples ($P = 0.264$) or post-NEC and no-NEC samples ($P = 0.119$; **Figure 4.14, Table 4.9**). Species richness was significantly lower in post-NEC samples ($P = 0.012$) compared to no-NEC only (**Figure 4.14, Table 4.9**). In infants who received Labinic, Shannon diversity was significantly lower in samples collected post-NEC than those collected pre-NEC ($P = 0.002$) and post-NEC compared to no-NEC ($P = 0.008$; **Figure 4.14, Table 4.10**). However, there was no significant difference in Shannon diversity between pre-NEC and no-NEC samples ($P = 0.843$, **Figure 4.14, Table 4.10**). As seen with the unstratified analysis, for these infants, species richness was significantly lower pre-NEC compared to no-NEC ($P < 0.001$) and post-NEC compared to no-NEC ($P < 0.001$; **Figure 4.14, Table 4.10**). The same was true in infants who received Infloran, whereby species richness was significantly lower in samples collected pre-NEC compared to no-NEC ($P = 0.026$) and post-NEC compared to no-NEC ($P = 0.002$; **Figure 4.14, Table 4.11**). Unlike the no probiotic and Labinic infants, there was no significant association between NEC (no-NEC/pre-NEC/post-NEC) and Shannon diversity ($P = 0.939$) in Infloran infants (**Figure 4.14, Table 4.11**). There was, however, a significant association with antibiotics, whereby samples taken from infants who had been administered antibiotics in the previous 7 days had a significantly lower Shannon diversity than those who did not ($P = 0.003$).

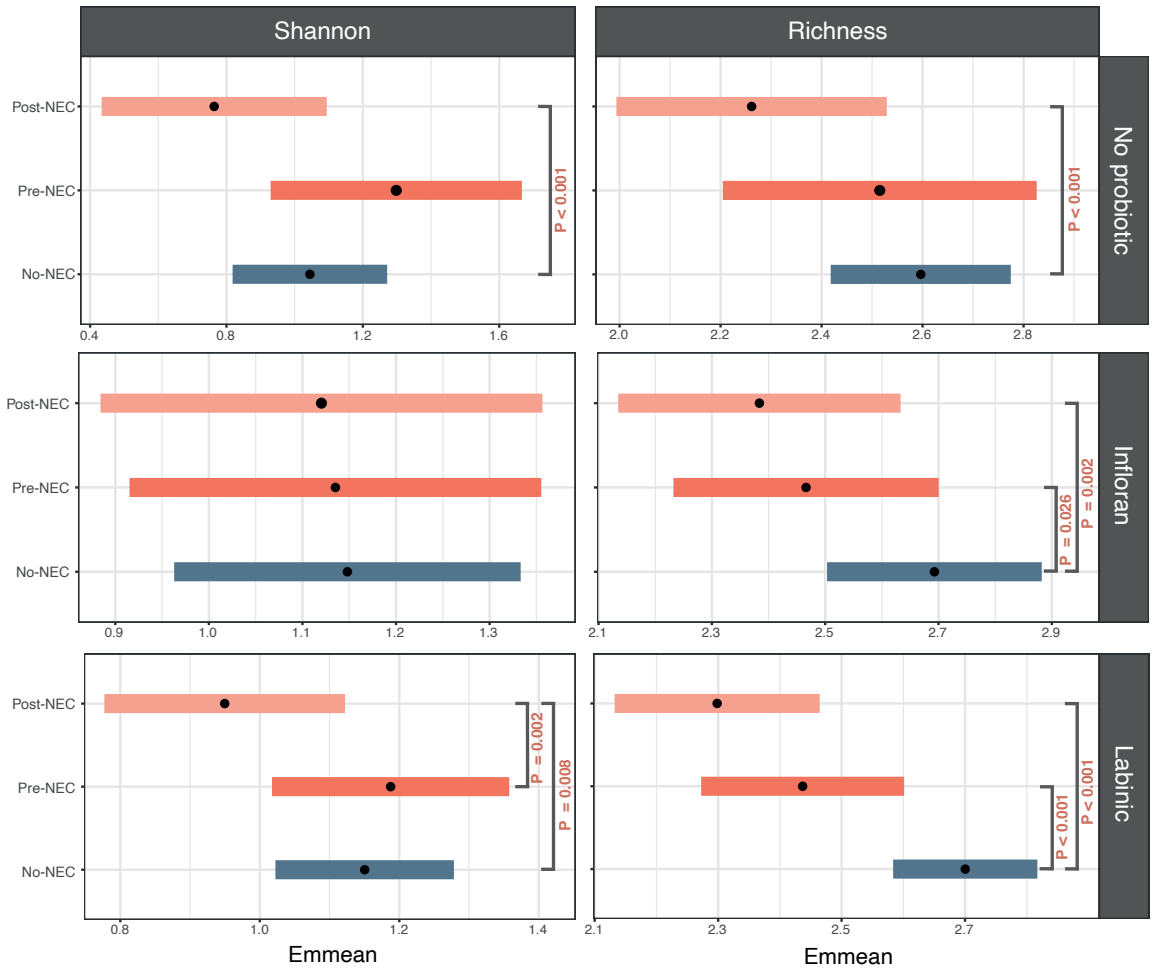

Figure 4.14. NEC is significantly associated with alpha diversity measures, but there are differences dependent on probiotic intake. Estimated marginal means (95% CIs) representing Shannon diversity and species richness for NEC, stratified by probiotic type, obtained from linear mixed-effects models adjusted for gestational age, birthweight, birth mode, sex, season, day of full feed, BMF, formula, probiotics, DOL and patient ID. The statistical significance shown is after adjustment for multiple comparisons using the two-tailed Dunnett's method, whereby 'No-NEC' was used as the control.

Table 4.9. Association of clinical co-variates with Shannon diversity and richness in infants who received no probiotic. Global *P*-values and Wald's Chi-squared test statistic for fixed effects are based on Type II ANOVA on the fitted LMMs, with patient ID as a random effect.

	Shannon		**Richness**	
	Chisq	pval	Chisq	pval
DOL	41.997	<0.001	61.123	<0.001
Gestational age	0.013	0.910	0.326	0.568
Birthweight	3.095	0.079	0.025	0.875
Birth mode	0.060	0.807	0.179	0.672
Sex	1.501	0.221	0.335	0.562
Antibiotics	3.942	0.047	6.485	0.011
Day full feeds	0.541	0.462	0.350	0.554
MOM	16.906	<0.001	32.087	<0.001
BMF	6.944	0.074	4.726	0.193
Formula	2.082	0.556	9.854	0.020
NEC	21.809	<0.001	10.850	<0.001

Table 4.10. Association of clinical co-variates with Shannon diversity and richness in infants who received Labinic. Global *P*-values and Wald's Chi-squared test statistic for fixed effects are based on Type II ANOVA on the fitted LMMs, with patient ID as a random effect.

	Shannon		**Richness**	
	Chisq	pval	Chisq	pval
DOL	48.895	<0.001	128.107	<0.001
Gestational age	7.282	0.007	0.896	0.344
Birthweight	0.0007	0.979	1.591	0.207
Birth mode	2.291	0.130	1.235	0.266
Sex	0.214	0.644	0.019	0.891
Antibiotics	39.452	<0.001	14.523	<0.001
Day full feeds	0.064	0.800	1.184	0.277
MOM	25.485	<0.001	6.290	0.098
BMF	8.477	0.037	3.204	0.361
Formula	4.929	0.178	4.294	0.231
NEC	14.084	<0.001	37.058	<0.001

Table 4.11. Association of clinical co-variates with Shannon diversity and richness in infants who received Infloran. Global *P*-values and Wald's Chi-squared test statistic for fixed effects are based on Type II ANVOVA on the fitted LMMs, with patient ID as a random effect.

	Shannon		**Richness**	
	Chisq	pval	Chisq	pval
DOL	3.430	0.064	31.571	<0.001
Gestational age	1.359	0.244	1.623	0.203
Birthweight	1.672	0.196	1.096	0.295
Birth mode	1.260	0.262	2.013	0.156
Sex	1.092	0.296	0.037	0.848
Antibiotics	9.087	0.003	3.217	0.073
Day full feeds	0.053	0.819	1.400	0.237
MOM	2.021	0.568	6.082	0.108
BMF	4.358	0.225	3.403	0.334
Formula	1.209	0.751	2.366	0.500
NEC	0.126	0.939	12.640	0.002

Following on, stratifying the MaAslin2 analysis into probiotic groups revealed no significant associations with any clinical variable in each of the groups, likely due to over-fitting of the models (i.e., too many clinical variables included as fixed effects for sample size). This analysis was repeated within each probiotic group but only including DOL and NEC (no-NEC/pre-NEC/post-NEC) as fixed effects and patient ID as a random effect, as DOL and patient ID explained the most variation in gut profiles aside from probiotics. In the no probiotic era, there was no significant association between any phyla and pre-NEC samples, although relative abundance of Actinobacteria was significantly lower post-NEC ($P < 0.001$, **Table 4.12**). During both probiotic eras, the relative abundance of Actinobacteria was significantly lower pre-NEC compared to no-NEC (Infloran: $P = 0.118$, $Q = 0.203$, Labinic: $P < 0.001$, $Q < 0.001$; **Table 4.12**), whilst the relative abundance of Proteobacteria was significantly higher pre-NEC compared to no-NEC during the Labinic era only ($P = 0.024$, $Q = 0.036$; **Table 4.12**), as seen in previous analyses. The relative abundance of Bacteroidetes was significantly higher pre-NEC during the Infloran era ($P = 0.05$, $Q = 0.150$; **Table 4.12**).

Table 4.12. MaAsLin2 results for significant taxa associated with NEC at the phylum level. Mixed-effects linear models using a variance-stabilizing arcsin square root transformation on relative abundance phyla data were used to determine the significance. DOL was included as a fixed effect and unique patient ID as a random effect in each of the probiotic models.

	Feature	Variable Level	Coeff	Standard Error	Pval	Qval
No probiotic	Actinobacteria	Post-NEC	-0.361	0.076	<0.001	<0.001
	Firmicutes	Post-NEC	0.242	0.121	0.052	0.155
Infloran	Bacteroidetes	Pre-NEC	0.012	0.006	0.05	0.15
	Actinobacteria	Pre-NEC	-0.111	0.07	0.118	0.203
	Actinobacteria	Post-NEC	-0.24	0.067	0.001	0.003
	Firmicutes	Post-NEC	0.13	0.073	0.08	0.178
Labinic	Actinobacteria	Pre-NEC	-0.219	0.057	<0.001	<0.001
	Proteobacteria	Pre-NEC	0.178	0.078	0.024	0.036
	Actinobacteria	Post-NEC	-0.435	0.059	<0.001	<0.001
	Proteobacteria	Post-NEC	0.227	0.081	0.006	0.01
	Firmicutes	Post-NEC	0.139	0.069	0.045	0.05

There was no genus consistently significantly associated prior to NEC diagnosis across the probiotic groups (i.e., with pre-NEC samples). However, the relative abundance of *Clostridium* spp. was consistently significantly lower post-NEC (no probiotic: $P = 0.05$, $Q = 0.129$; Infloran: $P < 0.001$, $Q < 0.001$; Labinic $P = 0.002$, $Q = 0.009$; **Table 4.13**). In infants never exposed to probiotics, the relative abundance of *Enterococcus* was significantly higher pre-NEC ($P = 0.105$, $Q = 0.248$), whilst *Streptococcus* was significantly lower ($P = 0.009$, $Q = 0.032$; **Table 4.13**). For infants who received Infloran, relative abundance of *Clostridium* was similarly significantly lower pre-NEC as seen post-NEC ($P = 0.07$, $Q = 0.249$) and *Staphylococcus* was significantly higher ($P = 0.003$, $Q = 0.014$; **Table 4.13**). In infants who received Labinic, relative abundance of *Bifidobacterium* was significantly lower pre-NEC ($P < 0.001$, $Q = 0.003$), whilst *Klebsiella* ($P = 0.08$, $Q = 0.171$) and *Citrobacter* ($P = 0.108$, $Q = 0.212$) were significantly higher (**Table 4.13**). To ensure that the differential genera associated with NEC between probiotic groups were not an artifact of the timing surrounding NEC onset (i.e., earlier onset in Infloran infants where *Staphylococcus* is higher pre-NEC, later onset in Labinic infants where *Klebsiella* is higher pre-NEC), the day of NEC onset was compared between probiotic groups. There was no significant difference in day of onset between groups ($P = 0.873$).

Table 4.13. MaAsLin2 results for significant taxa associated with NEC at the genus level. Mixed-effects linear models using a variance-stabilizing arcsin square root transformation on relative abundance genera data were used to determine the significance. DOL was included as a fixed effect and unique patient ID as a random effect in each of the probiotic models.

	Feature	Variable Level	Coeff	Standard Error	Pval	Qval
No probiotic	*Streptococcus*	Pre-NEC	-0.076	0.028	0.009	0.032
	Enterococcus	Pre-NEC	0.164	0.1	0.105	0.248
	Bifidobacterium	Post-NEC	-0.319	0.09	0.001	0.004
	Finegoldia	Post-NEC	-0.032	0.01	0.003	0.013
	Streptococcus	Post-NEC	0.071	0.025	0.006	0.024
	Cutibacterium	Post-NEC	-0.071	0.03	0.019	0.062
	Staphylococcus	Post-NEC	0.224	0.093	0.02	0.062
	Veillonella	Post-NEC	-0.047	0.02	0.023	0.066
	Enterobacter	Post-NEC	0.163	0.079	0.046	0.124
	Clostridium	Post-NEC	-0.029	0.014	0.05	0.129
	Actinomyces	Post-NEC	-0.031	0.017	0.079	0.193
Infloran	*Staphylococcus*	Pre-NEC	0.146	0.048	0.003	0.014
	Clostridium	Pre-NEC	-0.011	0.006	0.073	0.249
	Bifidobacterium	Post-NEC	-0.241	0.068	0.001	0.004
	Clostridium	Post-NEC	-0.015	0.006	0.012	0.043
Labinic	*Bifidobacterium*	Pre-NEC	-0.202	0.058	0.001	0.003
	Klebsiella	Pre-NEC	0.117	0.067	0.08	0.171
	Citrobacter	Pre-NEC	0.029	0.018	0.108	0.212
	Bifidobacterium	Post-NEC	-0.424	0.061	<0.001	<0.001
	Finegoldia	Post-NEC	-0.017	0.004	<0.001	0.001
	Klebsiella	Post-NEC	0.217	0.069	0.002	0.007
	Clostridium	Post-NEC	-0.019	0.006	0.003	0.009
	Veillonella	Post-NEC	-0.04	0.013	0.003	0.009
	Cutibacterium	Post-NEC	-0.032	0.012	0.007	0.02
	Anaerococcus	Post-NEC	-0.015	0.006	0.009	0.027
	Corynebacterium	Post-NEC	-0.015	0.006	0.015	0.041
	Enterococcus	Post-NEC	0.131	0.057	0.023	0.058

Finally, when focusing on specific species, the relative abundance of *B. longum* was consistently lower pre-NEC onset compared to no-NEC in both probiotic eras (Infloran: $P = 0.041$, $Q = 0.155$; Labinic: $P < 0.001$, $Q = 0.002$), but was significantly higher pre-NEC when no probiotic was received ($P < 0.001$, $Q = 0.002$; **Table 4.14**). Although, this could be explained by the low number of NEC infants who never received any probiotic and had at least 1 sample prior to diagnosis (n = 8 infants, compared to n = 24 and n = 28 infants for Infloran and Labinic, respectively). In infants who received no

probiotic, relative abundance of *Streptococcus salivarius* was significantly lower pre-NEC (P = 0.023, Q = 0.078; **Table 4.14**). Aside from *B. longum*, in infants who received Infloran, *S. epidermidis* was significantly higher pre-NEC (P < 0.001, Q = 0.003), as was *L. rhamnosus* (P = 0.039, Q = 0.153; **Table 4.14**). Finally, in infants who received Labinic, *B. bifidum* was significantly lower pre-NEC (P = 0.105, Q = 0.199), as was *C. acnes* (P = 0.109, Q = 0.202; **Table 4.14**). Conversely, relative abundance of various members of the *K. pneumoniae* complex were significantly higher pre-NEC, including *K. variicola* (P = 0.022 Q = 0.054), *K. pneumoniae* (P = 0.025, Q = 0.061) and *K. quasipneumoniae* (P = 0.076, Q = 0.158; **Table 4.14**), in keeping with the binomial mixed model analysis where infants who received Labinic and developed NEC were significantly more likely to transition into PGCT-3 (dominated by *K. pneumoniae* species complex). Additionally, relative abundance of *S. salivarius* (P = 0.017, Q = 0.045) and *E. faecalis* (P = 0.091, Q = 0.176) was also significantly higher pre-NEC (**Table 4.14**). It is important to note that some of the differences observed could be attributable to other confounding variables that were unable to be controlled for, outside of DOL and patient ID, such as antibiotic use or dietary variables. To validate the observation that the relative abundance of *S. epidermidis* was higher in pre-NEC Infloran infants (i.e., a pathobiont known to cause LOS), the number of *S. epidermidis*-associated LOS cases was compared between probiotic groups. Although more of the Infloran NEC infants developed LOS (*n* = 13/54), compared to no probiotic (*n* = 5/41) or Labinic infants (*n* = 6/104), of those cases, the number of *S. epidermidis*-associated LOS cases were comparable between groups (P = 0.654). Furthermore, *S. epidermidis* was detected ubiquitously across numerous non-LOS infants (**Figure 4.4A**).

Table 4.14. MaAsLin2 results for significant taxa associated with NEC at the species level. Mixed-effects linear models using a variance-stabilizing arcsin square root transformation on relative abundance species data were used to determine the significance. DOL was included as a fixed effect and unique patient ID as a random effect in each of the probiotic models.

	Feature	Variable Level	*Coeff*	Standard Error	*pval*	*qval*
No probiotic	*Streptococcus salivarius*	Pre-NEC	-0.056	0.024	0.023	0.078
	Streptococcus salivarius	Post-NEC	0.072	0.022	0.001	0.007
	Finegoldia magna	Post-NEC	-0.032	0.011	0.004	0.018
	Bifidobacterium breve	Post-NEC	-0.234	0.084	0.008	0.032
	Staphylococcus haemolyticus	Post-NEC	0.102	0.043	0.019	0.067
	Veillonella parvula	Post-NEC	-0.048	0.021	0.028	0.089
	Enterobacter cloacae complex	Post-NEC	0.163	0.083	0.054	0.167
	Bifidobacterium dentium	Post-NEC	-0.058	0.031	0.067	0.194
	Cutibacterium avidum	Post-NEC	-0.05	0.027	0.066	0.194
	Klebsiella michiganensis	Post-NEC	-0.071	0.039	0.077	0.214
Infloran	*Staphylococcus epidermidis*	Pre-NEC	0.134	0.035	<0.001	0.003
	Lactobacillus rhamnosus	Pre-NEC	0.073	0.034	0.039	0.153
	Bifidobacterium longum	Pre-NEC	-0.063	0.031	0.041	0.155
	Bifidobacterium longum	Post-NEC	-0.104	0.029	0.001	0.005
	Bifidobacterium breve	Post-NEC	-0.17	0.074	0.026	0.106
Labinic	*Bifidobacterium longum*	Post-NEC	-0.292	0.051	<0.001	<0.001
	Bifidobacterium longum	Pre-NEC	-0.177	0.049	<0.001	0.002
	Streptococcus salivarius	Pre-NEC	0.032	0.013	0.017	0.045
	Klebsiella variicola	Pre-NEC	0.053	0.023	0.022	0.055
	Klebsiella pneumoniae	Pre-NEC	0.105	0.046	0.025	0.061
	Klebsiella quasipneumoniae	Pre-NEC	0.039	0.022	0.076	0.158
	Enterococcus faecalis	Pre-NEC	0.09	0.053	0.091	0.176
	Bifidobacterium bifidum	Pre-NEC	-0.05	0.031	0.105	0.199
	Cutibacterium acnes	Pre-NEC	-0.007	0.004	0.109	0.202
	Finegoldia magna	Post-NEC	-0.018	0.005	<0.001	0.001
	Bifidobacterium bifidum	Post-NEC	-0.112	0.032	<0.001	0.002
	Clostridium perfringens	Post-NEC	-0.018	0.006	0.004	0.014
	Enterococcus faecalis	Post-NEC	0.157	0.055	0.005	0.017
	Corynebacterium kroppenstedtii	Post-NEC	-0.014	0.005	0.005	0.017
	Klebsiella pneumoniae	Post-NEC	0.134	0.047	0.005	0.018
	Klebsiella variicola	Post-NEC	0.065	0.023	0.006	0.019
	Bifidobacterium breve	Post-NEC	-0.104	0.038	0.007	0.021
	Veillonella parvula	Post-NEC	-0.026	0.01	0.009	0.025
	Cutibacterium avidum	Post-NEC	-0.027	0.01	0.011	0.03
	Bifidobacterium animalis	Post-NEC	-0.041	0.021	0.051	0.119
	Klebsiella quasipneumoniae	Post-NEC	0.044	0.022	0.055	0.123
	Veillonella atypica	Post-NEC	-0.014	0.008	0.091	0.176
	Lactobacillus rhamnosus	Post-NEC	-0.017	0.011	0.126	0.228

4.3.8. The association between NEC and the persistence of probiotic species

Following on from the persistence analysis in Chapter 3, where probiotic type was the only co-variate associated with the persistence of *B. bifidum* and *L. acidophilus*, the same analyses were conducted, including NEC infants. Again, probiotic type was significantly associated with the persistence of both species, with significantly greater persistence of Infloran-derived *B. bifidum* than Labinic-derived ($P < 0.001$), and greater persistence of Labinic-derived *L. acidophilus* than Infloran-derived ($P < 0.001$). In addition, the development of NEC, was significantly negatively associated with the persistence of *B. bifidum* ($P = 0.014$). To further unpick whether this association was reflective of antibiotic use, the analysis was repeated controlling for the total number of antibiotic courses used, after which NEC remained significant ($P = 0.048$). Furthermore, the total number of antibiotic courses used was not itself significantly associated with *B. bifidum* persistence, independent of NEC ($P = 0.105$). Whilst significantly associated with *B. bifidum* persistence, NEC was not significantly associated with the persistence of *L. acidophilus* ($P = 0.539$), although use of formula was ($P = 0.021$), with *L. acidophilus* persisting more in formula-fed infants.

The analyses were then repeated to look specifically at the persistence of Infloran-derived probiotic species and associated species (*B. bifidum, L. acidophilus* and *B. breve*) and the persistence of Labinic-derived probiotic species and associated species (*B. bifidum, B. longum, L. acidophilus* and *B. animalis*). When stratifying to Infloran infants only, no clinical variable was significantly associated with the persistence of *B. bifidum* or *L. acidophilus,* including NEC ($P = 0.555$ and $P = 0.181$, respectively). NEC was also not significantly associated with the persistence of *B. breve*, however, BMF was ($P = 0.049$), with *B. breve* persisting more/less in infants who received BMF. When stratifying to Labinic infants only, no clinical variables were significantly associated with persistence of *B. bifidum* or *B. animalis*, including NEC ($P = 0.083$ and $P = 0.142$). The persistence of *L. acidophilus* in Labinic infants was significantly associated with formula use ($P = 0.044$), with *L. acidophilus* persisting more in formula-fed infants, as seen in the analysis across both probiotic cohorts. The persistence of *B. longum* following Labinic cessation could not be analysed as there were only two infants for which *B. longum* did not persist.

4.3.9. Comparative functional analysis of the gut microbiome in preterm infants with NEC

To determine whether NEC, or other co-variates were associated with overall functional profiles (based on EC number), PERMANOVA was performed using 'adonis'. Similar to taxonomic profiles, DOL explained 2.4% variation in functional profiles of healthy infants, compared to 1.3% variation in infants who developed NEC (both $P < 0.001$). Similarly, unique patient identifier explained 5.3% in healthy infants, compared to 1% in infants who developed NEC (both $P < 0.001$). As previous, data was split into cross-sectional time-points based on DOL, each with one sample per patient. Similar to taxonomic analysis and to the results from Chapter 3, probiotics were the most significant co-variate, associated at every timepoint apart from DOL 0-9, DOL 20-24 and DOL 50-69. MOM and NEC were also significantly associated DOL 40-49 (**Figure 4.15A**). The 'adonis' analysis was then stratified by control vs NEC as done with the taxonomic analysis. Probiotics were the main driver of functional profiles in healthy infants, being the most significant factor associated with overall profiles, significant at all the same timepoints as seen with the above analysis on all infants (**Figure 4.15B**). In addition, gestational age and day of full feeds were significantly associated at DOL 10-14. Similar to the taxonomic analysis, for infants who developed NEC, probiotics were not significantly associated at any timepoint, and only MOM was significantly associated at a single timepoint, DOL 40-49 (**Figure 4.15C**).

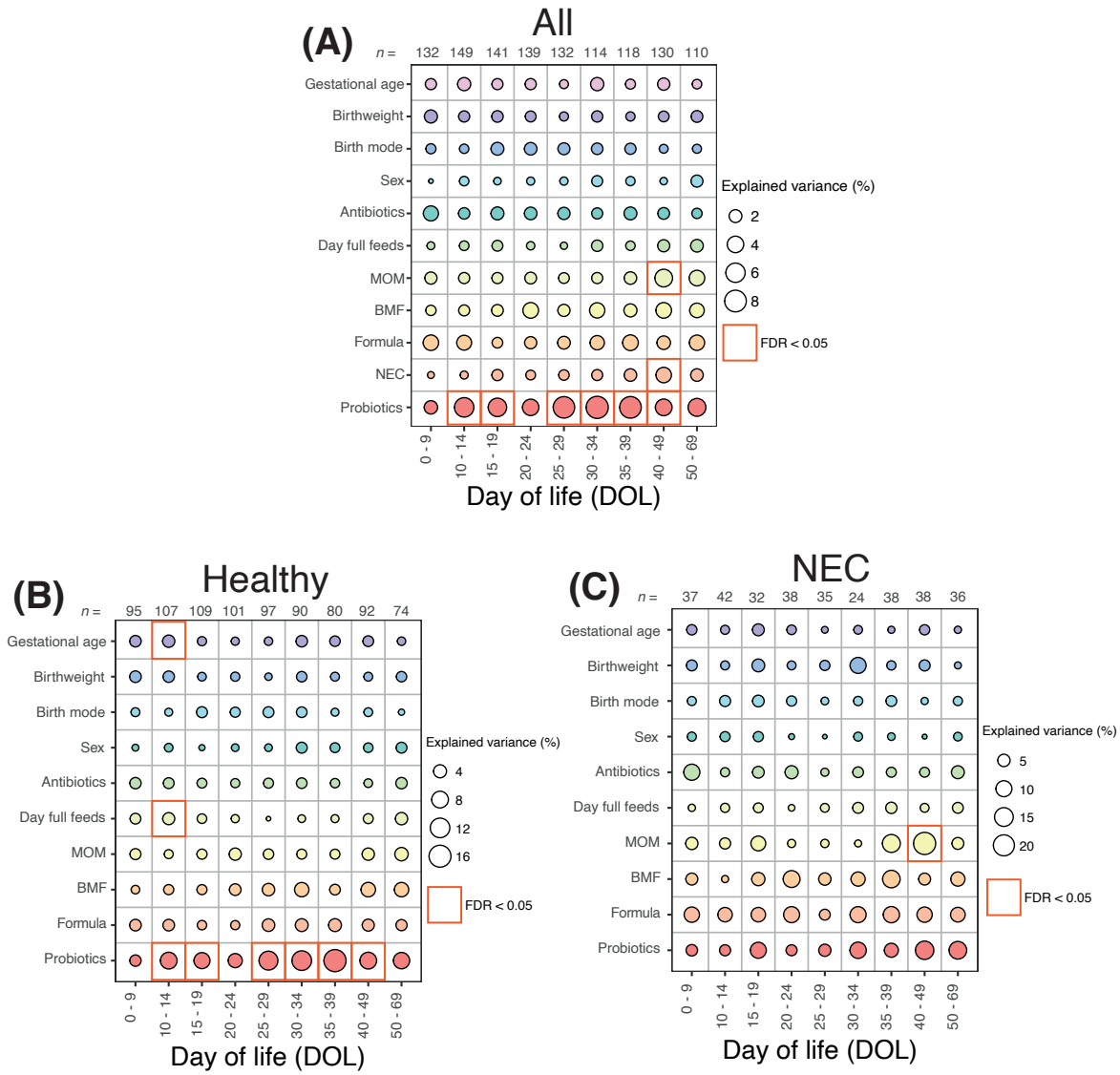

Figure 4.15. Significance and explained variance of clinical co-variates at different timepoints based on functional profiles (EC numbers), modelled by 'adonis'. Bubbles show the amount of variance (R^2) explained by each co-variate at a given timepoint and significant results (FDR<0.05) are surrounded by a red box for **(A)** both healthy and NEC infants **(B)** healthy infants only and **(C)** NEC infants only.

To further explore the functional potential of the gut microbiome in control and NEC infants, MaAsLin2 analysis was used. A total of 56 EC numbers were significantly associated with pre-NEC samples compared to no-NEC samples, whilst controlling for various measures (i.e., all confounders included in the adonis analysis) and subject ID (https://github.com/laurencbeck/supplementary_tables - **Appendix 2**). Of these, all were positively associated. Specifically, the relative abundance of genes encoding EC 2.1.1.17 (phosphatidylethanolamine N-methyltransferase), EC 3.5.1.38 (glutaminase-asparaginase) and EC 2.1.1.177 (23S rRNA (pseudouridine1915-N3)-

methyltransferase) were the most significant. Of the 56 significant features, 12 were phosphotransferases, seven were methyltransferases and five were oxidoreductases acting on the CH-OH group of donors with NAD(+) or NADP(+) as the acceptor. Additionally, 202 EC number were significantly associated with post-NEC samples compared to no-NEC samples (https://github.com/laurencbeck/supplementary_tables - **Appendix 2**). Of these, 91 were negatively associated whilst 111 were positively associated. Specifically, the relative abundance of genes encoding EC 3.2.1.96 (mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase; negative association), EC 2.3.1.191 (UDP-3-O-(3-hydroxyacyl)glucosamine N-acyltransferase; positive association) and EC 6.3.4.6 (urea carboxylase; negative association) were the most significant. 28 of the 56 EC numbers that were significantly positively associated with pre-NEC samples were also significantly positively associated with post-NEC samples.

4.4. Discussion

The present study represents the largest metagenomic longitudinal study in preterm infants who developed NEC. Aiming to gain insights into the gut microbial landscape of infants with NEC in the context of probiotic use, the findings shed light on this previously unexplored aspect of NEC research. Post-diagnosis, NEC was found to have a significant impact on the gut microbial community, which interestingly was not necessarily driven by antibiotic usage. For example, MAZ scores were significantly lower post diagnosis compared to healthy control infants, irrespective of antibiotic use. Notably, more subtle changes were observed preceding clinical diagnosis, including the enrichment of Proteobacteria such as *K. variicola* and a reduction in the relative abundance of Actinobacteria, namely Bifidobacteria. The observed microbial associations with NEC were dependent on whether probiotics were administered, or the type of probiotic that was administered suggesting probiotics to perhaps impact the microbial origins of NEC. Further to this, there were clear differences observed in the engraftment and colonisation patterns of probiotic strains in infants who developed NEC, dependent on probiotic type.

Previous gut microbiome studies in preterm infants with NEC typically point towards a gut community that is lacking in diversity. For example, studies have shown Shannon diversity, which takes into account the evenness of a community, to increase temporally in control infants but decrease in those that developed NEC (Stewart et al., 2016). Furthermore, a number of studies have found various diversity measures to be significantly lower prior to NEC diagnosis (McMurtry et al., 2015; Warner et al., 2016; Zhou et al., 2015). However, this has not been a completely universal finding, with other studies noting no significant difference (Torrazza et al., 2013). In this cohort, Shannon diversity was significantly lower post-diagnosis, but there was no discernible change preceding diagnosis. Instead, Shannon diversity in pre-NEC samples was comparable to no-NEC samples, suggesting the difference in Shannon diversity to be a consequence of disease or related clinical treatments/procedures, as opposed to a cause of disease. This was the case for infants who never received probiotics, and those that received Labinic. On the other hand, species richness was significantly lower prior to diagnosis in infants who received either probiotic but not in infants who received no probiotic. It could be that the probiotic species outcompete or inhibit the growth of other species, thereby reducing species richness, but not thereafter

dominating the community, hence the lack of change in Shannon diversity. In summary, the observed relationship between probiotic use, species richness and NEC diagnosis is intriguing but needs further investigation to provide more insight, such as a RCT combined with in-depth metagenomic analysis. Whilst it is important to characterise the signatures of NEC, it is also important to consider the clinical implications and translation of these findings. Using richness and diversity measures as indicators of clinical susceptibility to NEC would unlikely prove effective.

Aside from alpha diversity measures the temporal development of the gut microbiome was also explored. At the start of life, NEC infants were found to have a more 'stable' gut community, typically transitioning less between PGCTs. Further, in the 1:1 matched case control cohort, infants who developed NEC had a gut microbial community significantly more similar to other infants who developed NEC pre-diagnosis, than matched healthy control infants had to other healthy control infants. These findings corroborate one another, supporting the idea that healthy control infants have a more dynamic gut microbiome at the start of life, and transition more quickly between PGCTs. This contrasts previous findings, suggesting the preterm gut microbiome of NEC infants to be highly dynamic and individual, including prior to diagnosis (Stewart et al., 2016). Additionally, the findings also challenge the 'Anna Karenina principle' for animal microbiomes. According to this principle, the microbial community composition of 'dysbiotic' individuals will vary more than the community composition in healthy individuals (Zaneveld et al., 2017). In contrast, the findings presented here suggest that infants who later go on to develop NEC exhibit a more uniform gut microbiome. These unexpected results may indicate a distinct trajectory in gut microbiome development of NEC infants which is instead characterised by an initial phase pathobiont dominance. Thereafter, the community is unable to develop and mature in the same way that is observed in healthy preterm infants.

The microbial signatures of NEC were consistent with those described in previous studies. Increased *Proteobacteria* such as *K. variicola* was consistently observed in both the unmatched and matched datasets and during the Labinic era. Previous studies have yet to implicate *K. variicola* specifically in NEC, however, this is likely due to the extensive use of 16S rRNA gene sequencing and therefore no species-level identification in previous studies, and also the frequent misidentification of *K. variicola*

as *K. pneumoniae* in clinical settings (Rodríguez-Medina et al., 2019; Seki et al., 2013). Nonetheless, *Klebsiella* spp. have been associated with NEC onset in numerous studies (Coleman et al., 2023; Olm et al., 2019; Paveglio et al., 2020; CStewart et al., 2016), and *K. pneumoniae* in particular has been found to induce NEC-like injury in mice (Zhang et al., 2012). *Klebsiella* are found ubiquitously in the preterm gut, emerging as one of the most abundant and prevalent genera observed in these infants, irrespective of NEC diagnosis (Beck et al., 2022; Underwood et al., 2014). Competition in the preterm gut between members of the *K. pneumoniae* species complex and members of the *K. oxytoca* species complex has recently been described, with the preterm gut being found to be dominated by one group or the other (Coleman et al., 2023). This suggests overlapping niches and competition for luminal resources (Coleman et al., 2023), with the host environment likely playing a role in which species complex dominates and which is excluded. This was similarly observed in the current study, with PGCT-2 being dominated by members of the *K. oxytoca* species complex whilst PGCT-3 was dominated by member of the *K. pneumoniae* species complex, with little co-occurrence observed between the two. Despite infants who received Labinic being significantly more likely to be in PGCT-3 prior to NEC diagnosis, previous work has found members of either complex to be associated with the disease (Coleman et al., 2023; Paveglio et al., 2020).

Interestingly, infants who received Infloran or never received probiotics were not significantly more likely to transition into PGCT-3 prior to diagnosis, unlike Labinic infants. Instead, infants who took Infloran were significantly more likely to transition into PGCT-1 (*Staphylococcus* and *E. faecalis* dominant) prior to diagnosis. Furthermore, the relative abundance of *S. epidermidis* specifically was found to be higher in pre-NEC samples compared to no-NEC samples from Infloran infants. Unlike *Klebsiella* spp., *Staphylococcus* are considered early colonisers and pioneer species of the preterm gut (Rao et al., 2021), and both genera display temporal dynamics in healthy and NEC infants (**Figure 4.2**). Further to this, *K. pneumoniae* has in fact been shown to exploit *S. epidermidis* in order to effectively colonise the preterm gut (Rao et al., 2021). Therefore, it was considered whether the differences in microbial patterns of NEC between probiotic groups was in fact reflective of the day of onset of NEC i.e., whether Infloran infants had earlier onset NEC than Labinic infants. However, this was

not the case, and there was no significant difference in day of NEC onset between probiotic groups.

More historical studies on NEC, have also associated Staphylococci, and coagulase negative staphylococcal infections with the disease (Mollitt et al., 1988). Additionally, *S. epidermidis,* has been found to be the most common cause of LOS (Dong et al., 2018), which in itself is associated with NEC (Wang et al., 2020). Interestingly, more of the Infloran NEC infants developed LOS (n = 13/54), compared to no probiotic (n = 5/41) or Labinic infants (n = 6/104). That said, the number of *S. epidermidis*-associated LOS cases were comparable between groups, and *S. epidermidis* was similarly detected in numerous non-LOS infants and therefore the association with *S. epidermidis* was not confounded by *S. epidermidis* LOS cases. Together, the data suggest that there may be differences in the microbial origin of NEC between these probiotic eras. This is an intriguing possibility that could be a direct consequence of the specific probiotic products themselves, which differ in only one species (*B. longum* subsp. *infantis* present in Labinic but not Infloran), although they do have different strains. For example, if strains of a certain product were able to outcompete specific pathobionts (i.e., by physical displacement, or by altering the metabolic state of the gut environment (Mercer & Arrieta, 2023)). This would potentially give rise to others and allow them to establish a niche instead. This hypothesis further supports the results pertaining to species richness, previously described. Alternatively, this could be more reflective of the time course over which samples have been collected, as probiotics were administered during discrete time periods on the NICU. It stands with reason that various pathobionts would persist in the NICU environment (Hartz et al., 2015), and that this community of environmental microbes would change over time, altering the reservoir of hospital-acquired microbes. For example, a previous study in a newly built neonatal unit found an increase in the relative abundance of environmental *Klebsiella*, *Staphylococcus, Pseudomonas* and *Streptococcus* spp. over time following the introduction of patients (Zachariah et al., 2021). In concordance, whilst the majority of NEC cases are sporadic, NICU-specific outbreaks and clusters of NEC cases over time have been previously described (Meinzen-Derr et al., 2009; Wendelboe et al., 2010).

Also consistent with previous NEC research, the relative abundance of Bifidobacteria and therefore Actinobacteria were consistently lower preceding NEC diagnosis in both the unmatched and matched datasets, as well as during the Labinic era. The mechanisms of protection by *Bifidobacterium* are thought to vary by species, and include reducing the expression of inflammatory cytokines, improving barrier function and altering SCFA production (Al-Sadi et al., 2021; Khailova et al., 2009; Underwood et al., 2014). It is difficult to pinpoint the exact reason for the discrepancies seen in preterm gut Bifidobacterial populations, as some infants have 'lower' *Bifidobacterium* despite receiving breast milk (i.e., a source of HMOs to feed *Bifidobacterium* spp.) and probiotics (i.e., direct source of *Bifidobacterium*). However, it could be speculated that this may reflect differences in the internal host environment in some way, that is either favourable or unfavourable to *Bifidobacterium*. Aside from Actinobacteria and *Bifidobacterium*, in Infloran infants, the relative abundance of *Clostridium* spp. were also significantly lower prior to and following diagnosis, and lower following diagnosis in no probiotic and Labinic infants. The negative association between *Clostridium* spp. and NEC in Infloran infants is intriguing. For example, studies have similarly found the Clostridia class to be negatively associated with NEC, with the relative abundance of Clostridia to be lower preceding NEC diagnosis (McMurtry et al., 2015; Warner et al., 2016). However, it is likely that there are more subtleties to this relationship. By way of explanation, a recent study by Kiu *et al.* found that some *C. perfringens* strains were positively associated with NEC, whilst others were considered hypovirulent or 'commensal-like' (Kiu et al., 2023).

Functional analysis revealed a number of genes encoding various EC numbers to be significantly positively associated with pre-NEC samples. These findings indicate differences in the functional potential of the gut microbiome between control and NEC infants, suggesting a potential imbalance in lipid metabolism, amino acid metabolism and RNA modification, based on the most significant EC numbers. However, it is important to note that these findings surround the functional potential of the gut microbiome, and does not necessarily reflect whether these genes are actively expressed. Further work is also necessary to explore the functional capacity of the gut microbiome and its association with NEC. Whilst the use of metagenomic sequencing has provided more granular analysis of gut microbial species and their function within the community, future work may wish to do strain-level analysis, which is one of the

key advantages of metagenomic sequencing. This would help to separate strains that may be harmful and contribute to disease from those that do not. For example, the study by Kiu *et al.* on *C. perfringens*-associated NEC, found that strains encoding the toxin, perfringolysin O, were more likely to cause disease and were responsible for significantly more cellular damage in both *in vitro* and *in vivo* experiments (Kiu et al., 2023). Furthermore, strain-level analysis could prove a useful tool for further exploring NEC outbreaks within the NICU over the 12-year time-period. It may also be useful to further explore the functional aspect of the microbiome. Whilst metagenomics offers some insight into this, it is of course limited to functional capacity and fails to capture which genes are expressed and what the metabolic landscape looks like. Doing metabolomics on pre-NEC and matched control samples would offer further insight into this. Moreover, exposing organoids to stool-derived metabolites or identified bacterial signatures of NEC (i.e., *K. variicola*) or of health (i.e., *Bifidobacterium* spp.) would allow better understanding of the host response, and help validate the associations found.

Aside from the differential taxa observed, one of the most notable observations were the differences in probiotic species colonisation patterns between NEC and control infants, dependent on probiotic type. Infants who received Infloran were significantly associated with transition into PGCT-4, enriched, in particular, by *B. breve.* This strong association was also observed in infants who developed NEC, although to a lesser extent. Furthermore, development of NEC was not associated with the persistence of *B. breve,* or of other Infloran species *B. bifidum* and *L. acidophilus,* following the cessation of probiotic treatment. Conversely, healthy infants who took Labinic were significantly associated with transition into PGCT-5, whilst those who developed NEC were not. Furthermore, persistence of *B. bifidum* which is present in both products, is significantly impacted by the development of NEC, with a lack of persistence occurring following the cessation of probiotics. It was hypothesised that this impact was driven by lack of persistence in Labinic NEC infants specifically, rather than Infloran NEC infants. Although not reaching significance in the stratified analysis on Labinic infants ($P = 0.083$), the proportion of non-persisters in Labinic infants was greater than for Infloran. Whilst the differential colonisation dynamics of probiotic species has been previously explored (Beck et al., 2022), and investigated in Chapter 3, the impact of NEC development on these colonisation patterns and persistence of probiotic species has not yet been described.

Many of the differences observed were following NEC diagnosis rather than preceding, some of which, but not all, are likely due to antibiotic treatment and other aspects of NEC clinical management including nil by mouth. NEC onset can occur early in life making it often difficult to capture the timeframe leading up to disease. Alternatively, disease onset can occur much later during NICU stay, which despite being easier to analyse the timeframe preceding diagnosis, is likely a different subtype of NEC to those diagnosed earlier, characterised under the same umbrella term. In fact, it is now fairly widely accepted that NEC likely encompasses several disease sub-types (Berrington & Embleton, 2022; Neu et al., 2018), and with a great sample size across multiple sites, stratifying into these different sub-types could be a future possibility to unravel the complexities of disease. Furthermore, whilst some NEC cases are of clear microbial origin, and associated with over-representation of specific pathobionts such as *K. pneumoniae*, others are less-so and may not be of direct microbial origin. For example, blood transfusions have long been associated with the development of NEC (Mohamed & Shah, 2012; Neu et al., 2018). Whilst it is not yet clear whether this association is due to the transfusion itself, or the underlying cause of transfusion i.e., anaemia (Patel et al., 2016), it has been speculated that if it is the latter, the reduced delivery of oxygen to the gut may drive TLR4 expression and signalling resulting in the proinflammatory cascade that leads to NEC (Hackam, 2022).

4.5. Limitations

In addition to the limitations discussed in Chapter 3, surrounding data collection, the main limitation of the results described here are that this was largely an observational study. Observational studies can of course provide insight into the microbial associations with NEC but cannot prove causality. For example, the identified features such as *K. variicola* and *Bifidobacterium* cannot be proven to cause or protect against NEC from the results outlined. To address this, the identified associations would need to be tested either *in vitro* or *in vivo* i.e., by exposing a tissue culture line, organoids or a mouse model to strains. Furthermore, there were multiple confounding variables in the unmatched cohort (**Table 4.1**), which although controlled for in all analyses by including as fixed effects in models, could still have impacted the results. Hence, the use of the 1:1 matched case control cohort. This cohort was better matched, and confounders therefore better controlled for, however, the sample size was much

smaller and therefore may not have had the power to identify nuances and subtle changes. That said, compared to other cross-sectional studies on NEC, this was still a large cohort.

4.6. Conclusions

In summary, this is the largest longitudinal metagenomic study on the gut microbiome of preterm infants with NEC, shedding light on different microbial signatures of NEC, particularly in relation to probiotic use. The findings reveal that before clinical diagnosis of NEC, the relative abundance of Proteobacteria such as *K. variicola* is higher whilst the relative abundance of Actinobacteria, mainly Bifidobacteria is lower, concordant with other studies. The type and administration of probiotics influenced the microbial association observed with NEC, especially the colonisation patterns of different probiotic and probiotic-associated strains and potentially the microbial origins of NEC. The research underscores the complexities of studying NEC, and points towards the need for more focused studies surrounding strain-level analysis, metabolomics and further distinction between disease sub-types, to better understand the role of the gut microbiome in NEC development.

5. Characterising the preterm gut microbiota, metabolome and circulating T-lymphocytes

5.1. Abstract

Preterm infants have immature and naïve immune systems, and are subject to abnormal microbial colonisation, a combination which can increase the risk of disease. Characterising the complex relationship between temporal development of the gut microbiome (both structure and function) and systemic immunity, alongside clinical information, is therefore of significant importance. In this study, 266 longitudinal stool samples from 66 very preterm infants underwent 16S rRNA gene sequencing to analyse gut microbial structure. To further explore the functional status of these gut members, and immune profile in this population, a subset of these samples underwent stool metabolomics (n = 101), and a further subset were matched for blood (n = 41) to explore T-lymphocyte sub-populations using CyTOF.

The strongest associations were found with age for both the gut microbiota ($P < 0.001$) and metabolite profiles ($P < 0.001$). There was no significant association between clinical information and T-lymphocyte profiles, including no association with age ($P > 0.05$). Relationships between the three datasets were next explored, finding the most concordance between the gut microbiome and metabolome, with 706 significant correlations after FDR adjustment identified between the top ten most abundant taxa and all 977 identified metabolites. *Lactobacillus* had the highest number of significant correlations (31%), amongst which was a strong positive correlation with equol sulphate, an oestrogen produced by intestinal bacteria. Despite no correlation between gut microbiota or metabolite profiles and T-lymphocyte sub-populations, significant relationships were found with specific T-lymphocyte markers including a positive correlation between 3-hydroxybutyrate (BHB) and γδ T-cell receptor (TCR).

This study provides an important insight into the potential network of relationships underlying preterm gut microbiome structure and function with the host immune system. Very little correlation was found between the microbiome or metabolome with circulating T-lymphocyte populations, aside from when looking at specific markers.

Further work with a higher number of matched samples is needed to confirm these findings.

5.2. Introduction

Adaptive immunity, also known as acquired immunity, is a component of the immune system that plays a crucial role in protecting against invading pathogens and eliminating infectious agents. Unlike the innate immune system, which is a non-specific defence mechanism, the adaptive immune system is programmed and educated to respond to specific antigens. The neonatal immune system needs to detect, differentiate, and eliminate potentially harmful pathogens, whilst recognising commensal or probiotic microbes that will go on to form the basis of the early life microbiome. This is a delicate task that needs to be achieved for proper immune development. The gut microbiome in infancy is intrinsically linked to the immune system, in particular circulating T-lymphocyte populations which are key in recognising and eliminating pathogens.

CyTOF is an application of mass cytometry which is used for real-time analysis of single cells. In brief, CyTOF relies on the detection of heavy metal ions which are conjugated to a panel of antibodies. This technique can be useful for identifying and characterising specific T-lymphocyte sub-populations, based on a panel of markers which conjugated antibodies are able to bind to. In the context of preterm infants, CyTOF has been successfully used to probe major mucosal immune cell populations of infants with SIP (Olaloye et al., 2023) and NEC (Olaloye et al., 2021) compared to control infants.

The gut microbiome is a complex and dynamic ecosystem involving microbe-microbe and microbe-host interaction. Adding to the complexity, gut microbes produce a wide variety of metabolites, which are small molecules that serve as the basis of communication with other microbes and host cells. These metabolites may play a key role in modulating the immune system, potentially driving the differentiation and proliferation of different T-lymphocyte populations. In previous early life studies, the gut microbiome has been shown to mediate the immune system. For example, a low abundance of *Bifidobacterium* was found to be associated with changes in T-helper 2 (Th2) and Th17 responses as well as *B. infantis*-derived indole-3-lactic acid being shown to upregulate immunoregulatory galectin-1 in Th2 and Th17 cells (Henrick et al., 2021). Despite this, no study has integrated blood CyTOF data with both gut microbiome and metabolome in the preterm population specifically. Therefore,

understanding the complex relationships between gut microbes, metabolites and the host immune system is a novel area of research, especially in the preterm population where normal immune development may be altered due to abnormal bacterial colonisation. Further to this, preterm infants are at increased risk of inflammation and infection, due in part to their naïve immune systems. For instance, this population is more likely to develop NEC which is the leading cause of death in preterm infants and has been hypothesised to develop due to a hyperinflammatory immune response to microbial colonisation of the gut (Mara et al., 2018). A better understanding of the cellular and molecular mechanisms underlying these processes may help to inform future research into targeted interventions.

5.3. Results

5.3.1. Sampling overview

The current study focuses on analysing data that was collected as part of the INDIGO RCT. The aims of INDIGO were to test the effects of an exclusively human milk diet in comparison to a diet containing bovine products. The standard care in the Newcastle RVI NICU at the time of the study was for cow's milk formula to be used should there be a shortfall of MOM, and cow's milk-based fortification if needed after achieving full enteral feeds. Infants enrolled in INDIGO were part of either the control or intervention trial arm. Those in the control trial arm received standard care whereas those in the intervention trial arm were supplemented with fortified human milk should there be a shortfall of MOM, and human milk-based fortification if needed after achieving full enteral feeds.

As part of that trial, extensive clinical information was collected alongside gut microbiota information (16S rRNA gene sequencing) from four sites, including 66 infants from the Newcastle RVI NICU (n = 266 samples). Two of the sites used probiotics routinely which were found to be the main driver of the preterm gut microbiome (Chapter 3), including Newcastle, where all samples were collected during the Labinic era (see Methods). To fulfil the primary objective of the INDIGO trial, the 16S rRNA gene sequencing results were previously published from all units, with a specific focus on the trial (Embleton et al., 2023).

In order to expand on this work, additional data was collected for subsets of the Newcastle babies only, including stool metabolome and blood CyTOF data. For these reasons, for the multi-dataset analysis presented in this thesis chapter, samples from Newcastle infants only were analysed. Clinical information is presented in **Table 5.1.**

Table 5.1. Patient demographics of the analytical cohort, Newcastle samples only. *P*-values are based on Chi-squared test for categorical data and Kruskall-Wallis test for continuous data

	All	***Control***	***Intervention***	***pval***
No. of subjects	66	36	30	
No. of samples	266	152	114	
Median no. samples per subject (IQR)	4 (4 – 5)	5 (4 – 5)	4 (3.25 – 5)	
Median gestational age (IQR)	27.3 (26 – 28.1)	27.1 (26.1 – 28)	27.5 (26 – 28.1)	0.657
Median birthweight (g) (IQR)	912.5 (675 – 1110)	860 (640 – 1008)	1040 (730 – 1155)	0.098
Median day of full feed (IQR)	13 (11 – 18)	13 (11 – 18)	13 (12 – 17.5)	0.479
Median days on antibiotics (IQR)	11 (6 – 16.5)	10 (6 – 18.2)	12.5 (5.25 – 16)	0.949
Median days of MOM (IQR)	46.5 (27.5 – 66.8)	54 (29.2 – 67.5)	42 (26 – 64)	0.395
Birth mode				0.512
Caesarean	37 (56.1%)	22 (61.1%)	15 (50%)	
Vaginal	29 (43.9%)	14 (38.9%)	15 (50%)	
Sex				0.289
Male	36 (54.5%)	17 (47.2%)	19 (63.3%)	
Female	30 (45.5%)	19 (52.8%)	11 (36.7%)	
NEC or LOS				1
No	54 (81.8%)	29 (80.6%)	25 (83.3%)	
Yes	12 (18.2%)	7 (19.4%)	5 (16.7%)	
Antibiotics in the past 7 days				1
No	141 (53.0%)	81 (53.3%)	60 (52.6%)	
Yes	125 (47.0%)	71 (46.7%)	54 (47.4%)	
Median % enteral MOM in previous 3d	100 (37.75 – 100)	100 (93 – 100)	100 (25.8 – 100)	0.069
BMF at time of sample				0.805
No	156 (58.6%)	88 (57.9%)	68 (59.6%)	
Yes	110 (41.4%)	64 (42.1%)	46 (40.4%)	

The datasets making up this analysis included 266 longitudinal 16S rRNA gene sequencing samples (referred to as the microbiome dataset), collected from 66 very preterm infants during their stay at the Newcastle RVI NICU. These samples were collected across five pre-defined timepoints (A, B, C, D and E) reflecting the first sample, a sample after DOL 7, full milk enteral feeds, DOL 21-28 and study end, respectively (**Figure 5.2**). As outlined in the methods, of those 266 samples, 41 samples from 24 infants were matched for blood, taken either pre-fortifier (typically corresponding to time point C) or post-fortifier following full enteral feeds (typically corresponding to time point E). PBMC's isolated from these samples were sent for T-lymphocyte profiling using CyTOF. A different subset of 101 samples from 56 infants were sent for metabolomics; 85 were the same stool sample as was used for 16S rRNA gene sequencing, 8 were a stool sample from the same day and the remaining 8 were

a stool sample from within ± 3 days. 25 of the samples chosen for metabolomics had also been matched for blood so spanned all three datasets. Due to the sparsity of samples after DOL 50 for the microbiome and metabolomics data, all subsequent analysis directly including DOL as a continuous variable was restricted to between DOL 0 and 50 (**Figure 5.1**).

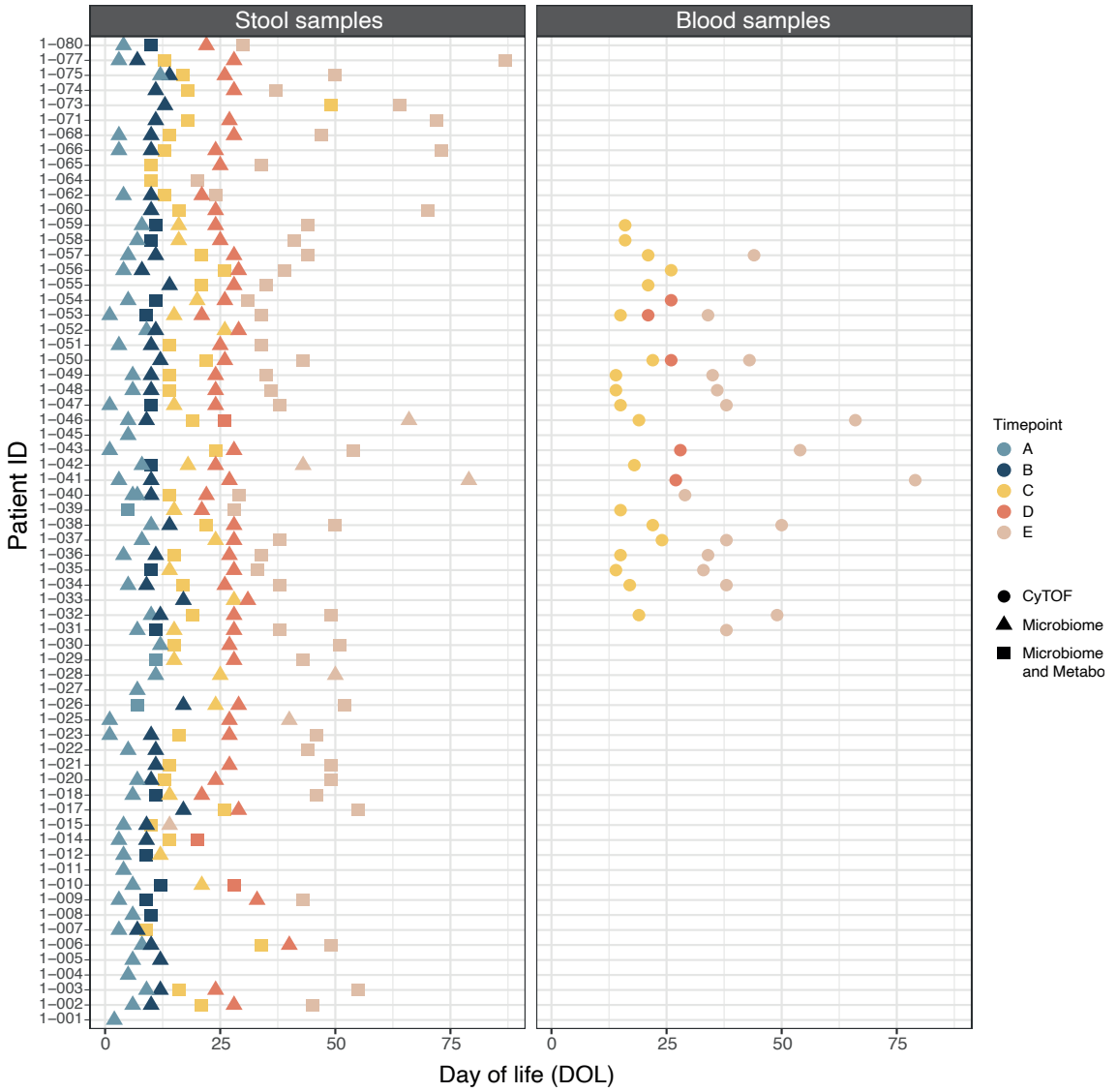

Figure 5.1. INDIGO sampling overview. Samples used in the study from birth to day 90, coloured by pre-defined time-point and shaped based on the dataset(s) the sample was used for.

5.3.2. 16S rRNA gene sequencing data overview

The top five most abundant genera were *Enterobacter/Klebsiella* (cannot be separated based on V4 sequencing), *Bifidobacterium, Escherichia/Shigella* (cannot be separated based on V4 sequencing), *Staphylococcus* and *Enterococcus*. Combined, these five

taxa accounted for a total of 91% of reads across all samples, and a total median relative abundance of 98% (IQR – 89% - 100%). First, relationships between clinical information and overall bacterial profiles were explored. Time had a significant impact on both the overall bacterial profiles. Age measured continuously explained 6.7% and 6.1% of the total variation for DOL and corrected gestational age (CGA), respectively (both $P < 0.001$). Time-point (i.e., categorical grouping) individually explained 10.5% of the total variation ($P < 0.001$). Since the study was designed for a sample to be taken from each patient during each time-point where possible, data was split into these cross-sectional groups, each with no more than one sample per patient. There was no significant association found for any of the clinical variables assessed with overall bacterial profiles at any time-point (i.e., gestational age at birth, birthweight, birth mode, sex, trial arm, antibiotics in the previous 7 days, day of full feeds, % enteral MOM in previous 3 days, BMF at the time of a sample and NEC or LOS disease status).

It was next investigated as to whether there were any associations between clinical data and specific genera, rather than the overall bacterial profile. MaAsLin2 analysis revealed some significant associations between clinical data and specific genera, largely attributable to the specific timepoints, corroborating the above findings (**Table 5.2**) and those from Chapters 3 and 4. Specifically, *Enterobacter/Klebsiella* was found to have a significantly higher relative abundance in time-points D ($P = 0.001$, $Q = 0.044$) and E ($P < 0.001$, $Q = 0.02$) compared to A, whilst the opposite was true of *Staphylococcus,* which was lower in relative abundance in E ($P < 0.001$, $Q = 0.004$) compared to A (**Figure 5.2A**). *Bifidobacterium* relative abundance was higher in time-point C ($P < 0.001$, $Q = 0.036$) compared to A, and was observed consistently across all timepoints after timepoint B, as well as over continuous DOL, likely reflecting that all samples were collected during the Labinic era (product containing *B. bifidum, B. longum* subsp. *infantis and L. acidophilus*) (**Figure 5.2A, B)**. *Veillonella* was also found to be significantly higher in samples where fortifier was used at the time of the sample ($P = 0.001$, $Q = 0.043$). There was no relationship observed between time (either timepoint or DOL) and the relative abundance of *Escherichia/Shigella* or *Enterococcus,* which remained relatively consistent over time (**Figure 5.2A, B**).

Table 5.2. MaAsLin2 results for significant taxa associated with clinical covariates in the INDIGO cohort. Mixed-effects linear models using a variance-stabilizing arcsin square root transformation on relative abundance phyla data were used to determine the significance. Patient ID was included as a random effect.

	Variable level	Feature	Coeff	Standard error	pval	qval
Trial arm	Intervention	*Lactobacillus*	-0.032	0.014	0.023	0.215
Antibiotics 7d	Yes	*Bifidobacterium*	-0.121	0.039	0.002	0.056
Day of full feeds		*Bifidobacterium*	-0.06	0.021	0.007	0.11
% enteral breast milk		*Veillonella*	-0.029	0.009	0.002	0.055
% enteral breast milk		*Lactobacillus*	-0.021	0.007	0.003	0.073
Fortifier	Yes	*Veillonella*	0.063	0.019	0.001	0.044
Fortifier	Yes	*Staphylococcus*	-0.162	0.07	0.023	0.215
NEC or LOS	Yes	*Bifidobacterium*	0.16	0.059	0.009	0.122
Timepoint	E	*Staphylococcus*	-0.401	0.093	<0.001	0.004
Timepoint	E	*Enterobacter/Klebsiella*	0.374	0.099	<0.001	0.017
Timepoint	C	*Bifidobacterium*	0.233	0.067	0.001	0.036
Timepoint	D	*Enterobacter/Klebsiella*	0.307	0.094	0.001	0.044
Timepoint	B	*Bifidobacterium*	0.154	0.055	0.006	0.104
Timepoint	D	*Staphylococcus*	-0.247	0.089	0.006	0.104
Timepoint	E	*Bifidobacterium*	0.175	0.071	0.014	0.182
Timepoint	D	*Bifidobacterium*	0.165	0.068	0.016	0.193
Timepoint	C	*Enterobacter/Klebsiella*	0.223	0.093	0.017	0.193
Timepoint	B	*Escherichia/Shigella*	-0.138	0.059	0.021	0.215

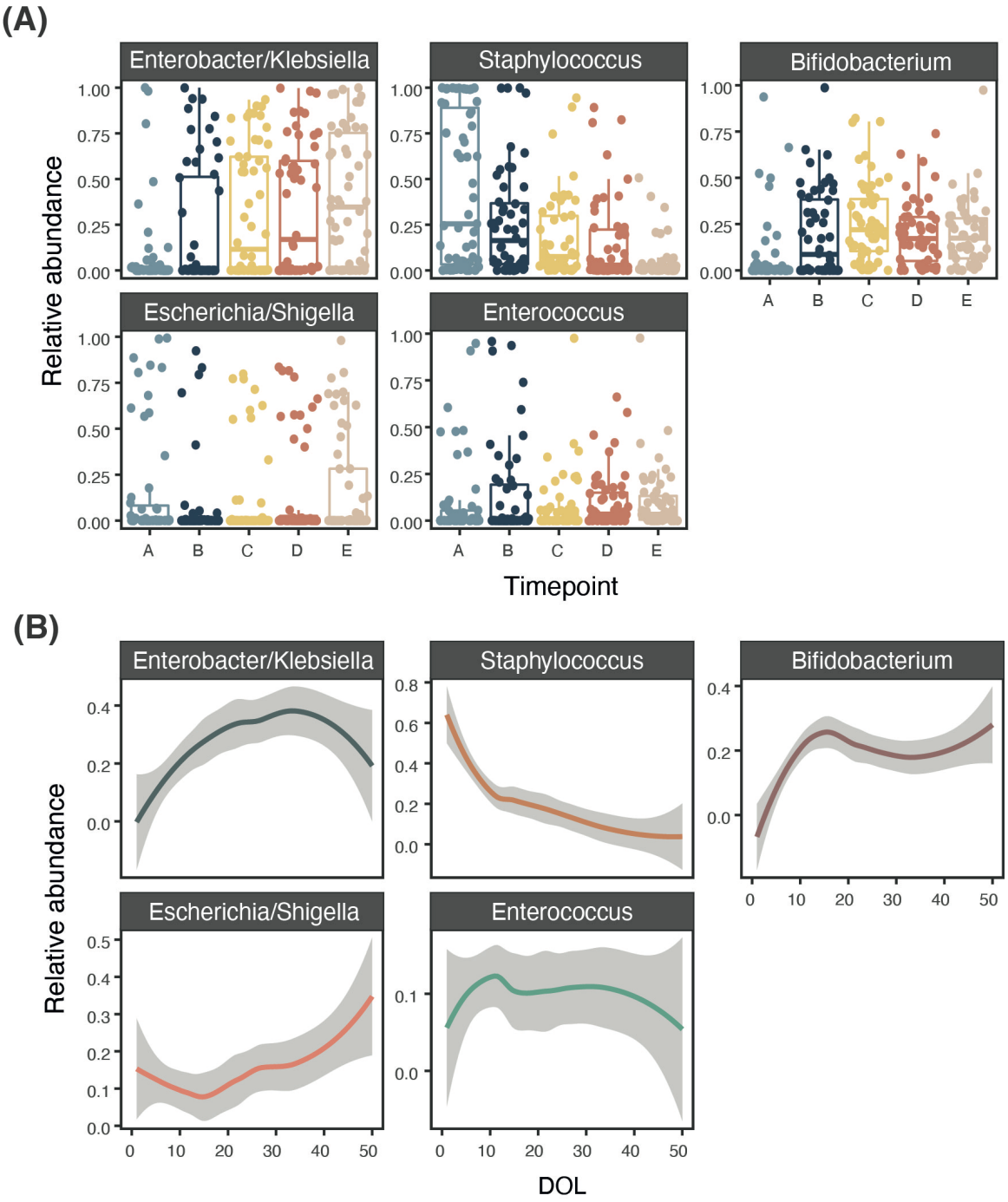

Figure 5.2. Descriptive overview of the most abundant genera. Relative abundance of the five most abundant genera in this cohort across **(A)** categorical time-points and over **(B)** continuous DOL up to DOL 50.

Shannon diversity significantly increased based on both increasing time-point and DOL (both $P < 0.05$) (**Figure 5.3A, B**), but OTU richness was not significantly associated with either time-point ($P = 0.309$) or DOL ($P = 0.506$) (**Figure 3.4C, D**).

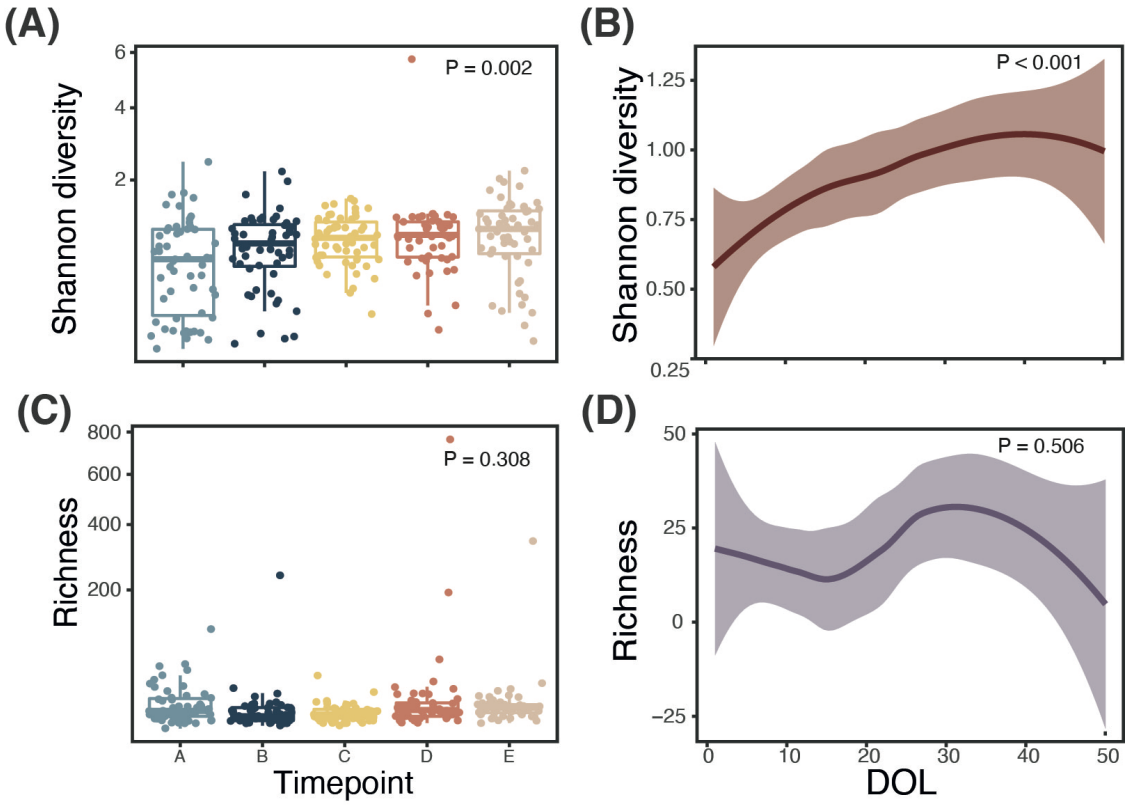

Figure 5.3. Alpha diversity of the gut microbiome in the preterm gut. Shannon diversity based on OTUs across **(A)** categorical timepoints A-E and over **(B)** continuous DOL. Bacterial OTU richness across **(C)** categorical timepoints A-E and over **(D)** continuous DOL.

DMM modelling of bacterial profiles was used to determine PGCTs which were numbered 1–5 based on the average age of samples within that cluster, as previous. LefSE was used to determine the most discriminatory taxa for each PGCT, using a cut-off Log_{10}LDA score of 5. PGCT-1 was discriminated by *Staphylococcus*, PGCT-2 was discriminated by *Enterococcus*, PGCT-3 was discriminated by *Escherichia/Shigella,* and PGCT-4 was discriminated by *Enterobacter/Klebsiella* (**Figure 5.4A, B**). There was no single discriminatory feature using this threshold identified for PGCT-5 (**Figure 5.4A, B**), which was characterised as having a significantly higher Shannon diversity (all adj. $P < 0.001$) in comparison to all other PGCTs (**Figure 5.4C**). This lack of dominance by a single taxon is also evident in the NMDS ordination, where the 95% CI ellipsis of PGCT-5 shows large overlap with the other PGCTs (**Figure 5.4D**). There was little association between clinical variables and PGCT (**Table 5.3**) based on binomial mixed effects models, controlling for patient ID. Namely, samples where fortifier was used at the time of sample were significantly less likely to belong to PGCT-1 ($P = 0.005$) and samples from infants belonging to the control trial arm (cow's milk

formula and fortification if needed) were significantly more likely to belong to PGCT-5 (P = 0.047).

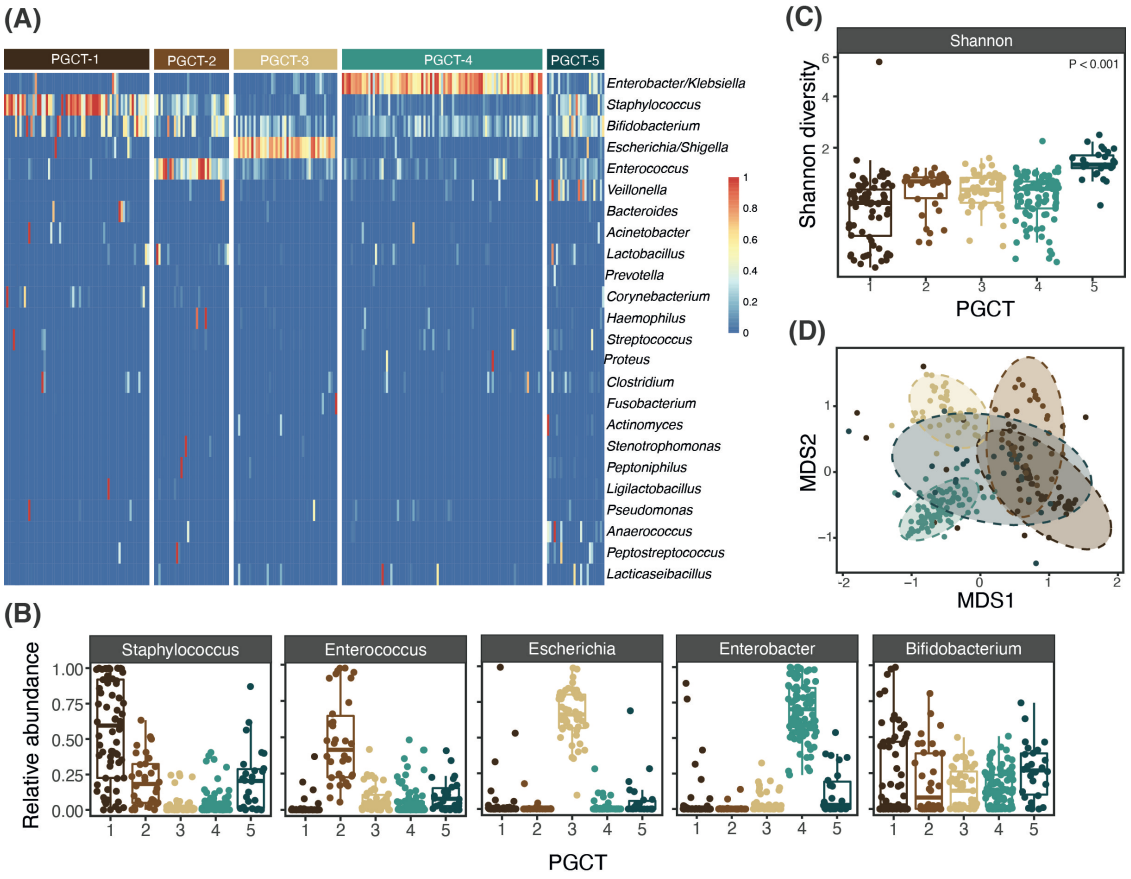

Figure 5.4. Discriminatory features of PGCTs. (A) Heatmap of all samples showing the relative abundance of the most dominant taxa, stratified by PGCT. **(B)** Relative abundance of the top five most abundant taxa across PGCT. **(C)** Box plot showing Shannon diversity for each PGCT. The centre line denotes the median, the box limits denote the IQR and whiskers extend to the limits. **(D)** NMDS ordination based on gut bacterial profiles, with 95% CI ellipses, coloured by PGCT.

Table 5.3. Association between PGCT and clinical co-variates in the INDIGO cohort. Global *P* values and aORs with 95% CIs are based on the fitted mixed-effects logistic regression models, with patient ID as a random effect.

	PGCT-1		PGCT-2		PGCT-3		PGCT-4		PGCT-5	
	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval	aOR (95% CI)	pval
Gestational age	1.017 (0.627 - 1.649)	0.946	0.959 (0.534 - 1.723)	0.888	0.888 (0.352 - 2.238)	0.801	1.27 (0.697 - 2.316)	0.435	0.597 (0.308 - 1.159)	0.128
Birthweight	1.002 (0.999 - 1.005)	0.255	1.001 (0.997 - 1.004)	0.687	0.998 (0.992 - 1.003)	0.452	0.999 (0.995 - 1.002)	0.455	1.001 (0.997 - 1.006)	0.6
Delivery mode		0.326		0.670		0.208		0.369		0.508
Vaginal	1		1		1		1		1	
Caesarean	1.828 (0.549 - 6.09)		1.364 (0.327 - 5.687)		0.216 (0.02 - 2.351)		2.038 (0.431 - 9.623)		0.616 (0.146 - 2.591)	
Sex		0.560		0.493		0.711		0.397		0.616
Male	1		1		1		1		1	
Female	1.42 (0.437 - 4.61)		1.614 (0.411 - 6.327)		0.65 (0.067 - 6.348)		0.524 (0.117 - 2.341)		0.7 (0.174 - 2.82)	
Antibiotics 7d		0.458		0.371		0.265		0.882		0.76
No	1		1		1		1		1	
Yes	1.475 (0.528 - 4.116)		0.582 (0.178 - 1.903)		0.46 (0.118 - 1.799)		1.076 (0.408 - 2.84)		1.201 (0.37 - 3.902)	
Day of full feed	1.007 (0.92 - 1.103)	0.876	1.068 (0.958 - 1.189)	0.235	1.011 (0.844 - 1.211)	0.903	0.982 (0.876 - 1.102)	0.76	0.87 (0.747 - 1.012)	0.071
MOM 3d	1.012 (0.995 - 1.03)	0.164	0.991 (0.974 - 1.009)	0.341	1.002 (0.98 - 1.025)	0.874	0.998 (0.982 - 1.014)	0.771	0.992 (0.973 - 1.012)	0.454
Fortifier		0.005		0.192		0.232		0.201		0.119
No	1		1		1		1		1	
Yes	0.141 (0.036 - 0.55)		0.35 (0.072 - 1.698)		2.793 (0.518 - 15.051)		2.312 (0.641 - 8.348)		3.623 (0.718 - 18.275)	
Trial arm		0.623		0.558		0.783		0.369		0.047
Control	1		1		1		1		1	
Intervention	0.746 (0.232 - 2.402)		0.652 (0.156 - 2.72)		1.356 (0.155 - 11.834)		1.999 (0.442 - 9.049)		0.202 (0.042 - 0.982)	
NEC or LOS		0.815		0.661		0.701		0.746		0.316
No	1		1		1		1		1	
Yes	1.22 (0.23 - 6.464)		0.651 (0.095 - 4.452)		0.554 (0.027 - 11.26)		1.38 (0.196 - 9.691)		2.677 (0.39 - 18.364)	
Timepoint		0.025		0.642		0.291		0.056		0.145
A	1		1		1		1		1	
B	0.981 (0.325 - 2.961)		0.965 (0.245 - 3.793)		0.202 (0.029 - 1.388)		4.196 (0.818 - 21.539)		1.557 (0.254 - 9.557)	
C	1.183 (0.267 - 5.236)		0.768 (0.13 - 4.52)		0.29 (0.036 - 2.322)		10.278 (1.508 - 70.058)		0.499 (0.05 - 4.996)	
D	0.155 (0.024 - 0.983)		0.831 (0.142 - 4.866)		0.787 (0.109 - 5.68)		20.208 (2.765 - 147.666)		0.394 (0.039 - 3.943)	
E	0.083 (0.01 - 0.716)		0.27 (0.035 - 2.093)		0.864 (0.104 - 7.169)		14.896 (1.949 - 113.834)		2.668 (0.276 - 25.76)	

5.3.3. Stool metabolome overview

A total of 977 unique metabolites were identified, of these, 833 had confirmed identity using standards (85.3%). Based on super-pathways, the majority of identified metabolites were amino acids. The most abundant super-pathways and all sub-pathways were found to remain relatively constant over time, apart from the xenobiotics super-pathway which increased up to around DOL 20 before decreasing, and the sub-pathway "food component/plant" (i.e., part of the xenobiotics super-pathway) which followed the same trend (**Figure 5.5A, B**).

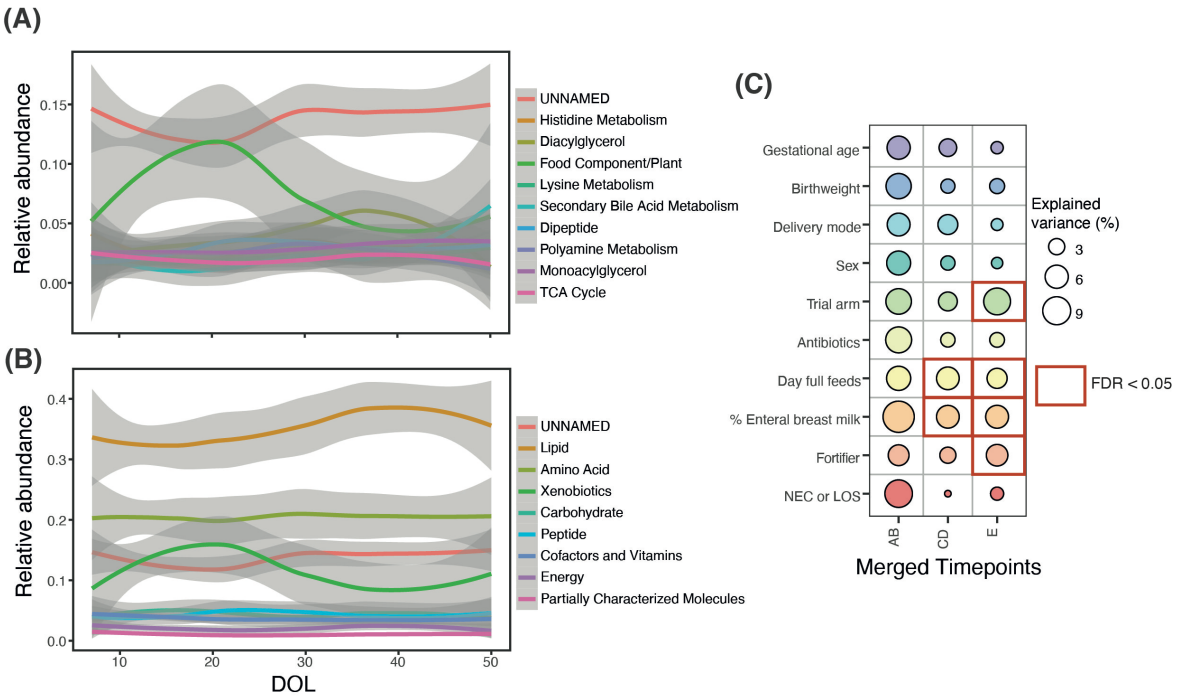

Figure 5.5. Overview of gut metabolome profiles. Relative abundance based on LOESS fits (95% CI) over DOL of (A) the most abundance sub-pathways and (B) the most abundant super-pathways. (C) The explained variance of 10 clinical co-variates at different time-points based on overall gut metabolomic profiles, modelled by 'adonis'. Bubbles show the amount of variance (R^2) explained by each covariate at a given timepoint and significant results (FDR < 0.05) are surrounded by a red box.

First, relationships between clinical information and the metabolome were explored. Comparable with bacterial profiles, time had a significant impact on the metabolome with DOL individually explaining 3% of the overall variation in metabolite profiles ($P < 0.001$) and CGA explaining 2.4% ($P < 0.001$). As with bacterial profiles, time-point also explained the most variation in the overall metabolite profiles, individually explaining 6% of the overall variation ($P = 0.002$). However, due to lower sample numbers in

timepoints A ($n = 3$) and D ($n = 3$), timepoints A and B were combined, as were C and D. For two of the samples in timepoint D, these infants already had a sample in timepoint C and so the samples from timepoint D were removed from the analysis. The combined timepoints (AB, CD and E) remained significantly associated with overall metabolite profiles ($P < 0.001$). Of the clinical variables assessed, % enteral MOM in the previous 3 days and day of full feeds were significantly associated with changes in the overall metabolome at timepoints CD, and E (all FDR $P < 0.05$). Additionally, fortifier at the time of a sample and trial arm were significantly associated with changes in the overall metabolome at timepoint E (both FDR $P < 0.05$) (**Figure 5.5C**). As with the bacterial analysis, it was next determined what specific metabolites were significantly associated with the extensive clinical data using MaAsLin2. The majority (42%) of significant metabolites were attributable to trial arm (i.e., human vs bovine-based fortifier), including a number of unnamed metabolites (https://github.com/laurencbeck/supplementary_tables - **Appendix 3**). Alpha-tocopherol and alpha-tocopherol acetate were amongst the most significantly enriched metabolites in the control trial arm (cow's milk formula and fortification if needed) (https://github.com/laurencbeck/supplementary_tables - **Appendix 3**). Day of full feed had the second highest number of significant metabolites (26%), then timepoint (17%) (https://github.com/laurencbeck/supplementary_tables - **Appendix 3**). These findings contrast the bacterial profiling results where the majority of significant associations were attributed to timepoint (80%), compared to the metabolite profiles where trial arm had the highest percentage of associations (42%) (**Figure 5.6**). Furthermore, nine co-variates returned significantly associated metabolites, which is considerably more than bacterial genera which only had significant features for age and fortifier.

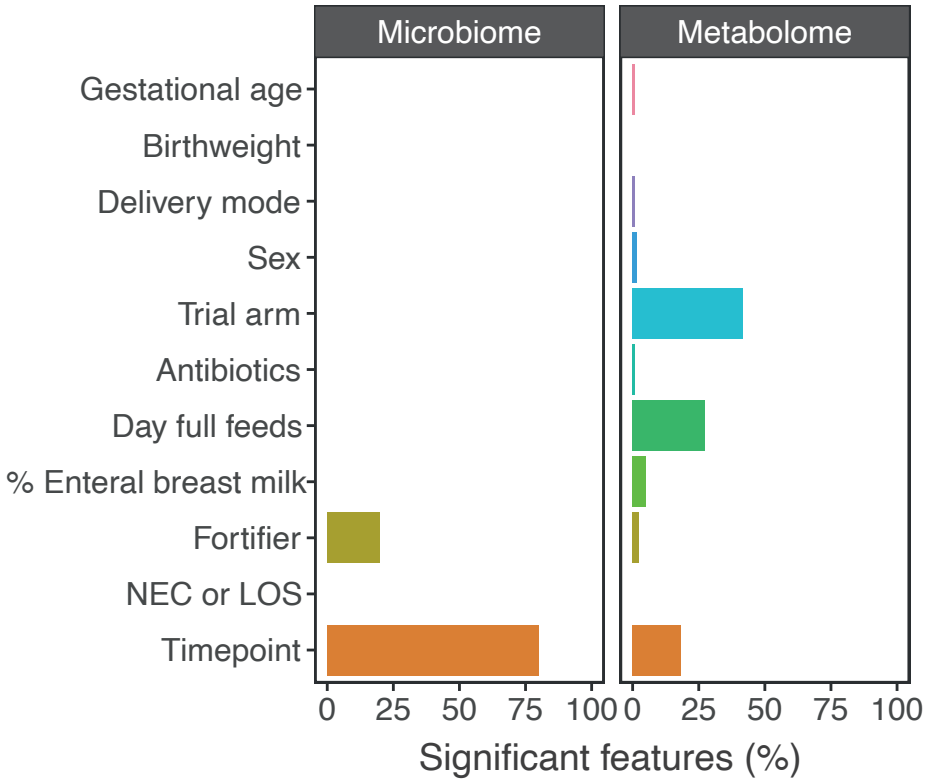

Figure 5.6. Percentage of significant features identified by MaAslin2 analysis attributed to each clinical variable. Based on the microbiome (gut bacterial profiles) and metabolome (gut metabolite profiles). The majority of significant associations with gut bacterial profiles were attributed to age at sampling (timepoint), compared to metabolite profiles where trial arm had the highest percentage of associations.

Due to fortifier being the only clinical variable found to be significantly associated with bacterial profiles other than timepoint (specifically, *Veillonella* was higher during fortifier use), the specific metabolites associated with fortifier was further probed. The 3 metabolites significantly associated with fortifier use were all acylcarnitines (myristoylcarnitine (C14), margaroylcarnitine (C17) and 3-hydroxyoleoylcarnitine) and were all positively associated.

Akin to the microbiome analysis, DMM modelling was used in an attempt to cluster samples into PMPTs based on their metabolite profiles (see Methods), DMM modelling was used. However, all samples were found to cluster together and the lowest Laplace approximation was one (i.e., the optimal number of clusters was one). Clustering was then attempted using a consensus-based algorithm, three clusters were found to be

optimal which were then defined using a hierarchical clustering approach based on complete linkage. However, one sample was found to cluster by itself, and another cluster contained just six of the 103 samples. This likely reflects the relatively comparable metabolite profiles between samples. Due to the limited utility of these clusters, no further analysis was performed based on PMPTs.

5.3.4. Blood T-lymphocyte profile overview

After positive gating of CD45+ CD3+ T-lymphocytes, FlowSOM was used to define 20 sub-populations based on marker expression which were then validated by manual gating (**Figure 5.7A, B**). The median ratio of CD4:CD8 T-lymphocytes in samples was 2.687 (IQR; 1.955 - 3.604), with the majority of cells being classified as either naïve CD4+ (47.96%) or naïve CD8+ (23.96%). The majority of cells were found to have a somewhat naïve phenotype, expressing naïve markers (CD45RA and CCR7) to a relatively high degree, with T-lymphocyte subsets lacking expression of both CD45RA and CCR7 accounting for only ~7% of cells.

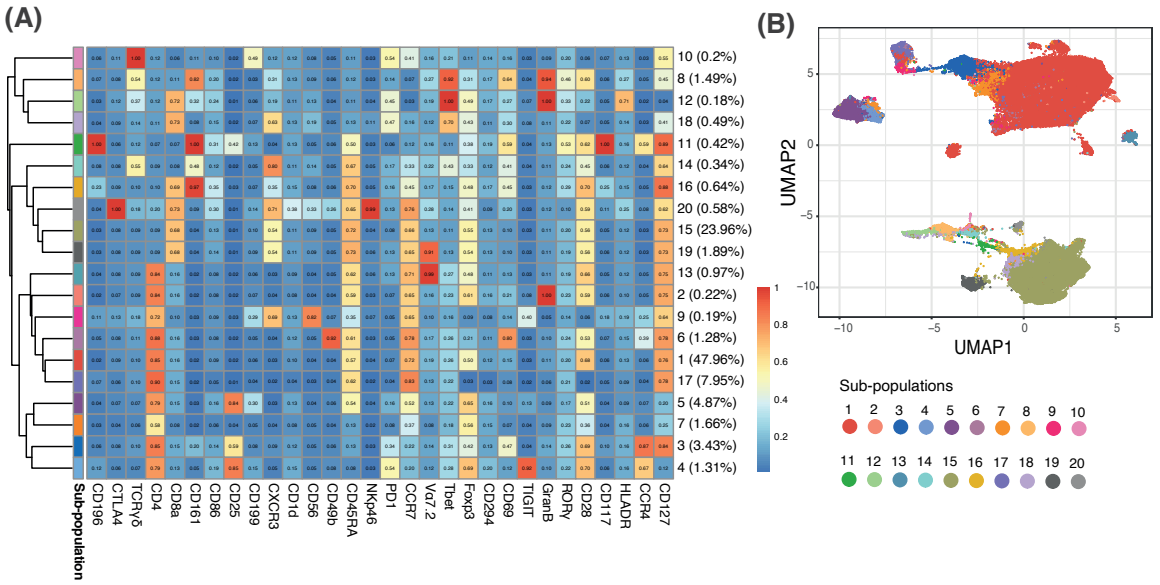

Figure 5.7. Defining circulating T-lymphocyte populations. (A) Heatmap depicting relative signal intensity of the antibody markers by T-lymphocyte sub-populations that were defined using FlowSOM. High expression denoted by red and lower by blue. **(B)** UMAP plot based on 1000 cells per sample, coloured by FlowSOM clusters.

As with bacterial and metabolite profiles, relationships between clinical information and T-lymphocyte profiles were explored. Unlike the other datasets, there was no impact of time (DOL, CGA, or time-point) on the T-lymphocyte profile of samples (P = 0.581,

P = 0.6564, and P = 0.881, respectively), or the overall marker expression of samples (P = 0.854, P = 0.863, and P = 0.968, respectively). As time was not a significant driver of blood T-lymphocyte profiles, instead, T-lymphocyte data was split into pre-fortifier treatment and post-fortifier treatment groups, each with one sample per patient, reflecting the study design for blood sample collection in the INDIGO study. Pre- and post-fortifier treatment were also not significantly associated with T-lymphocyte profiles (P = 0.525), Assessing the same clinical co-variates as for the other datasets, there was no significant association found with overall T-lymphocyte profiles (all $P > 0.05$) or overall marker expression (all $P > 0.05$). There was also no significant association with the ratio of CD4:CD8 T-lymphocytes and any of the clinical variables (including DOL, CGA, or time-point) whilst controlling for patient ID (all $P > 0.05$). More in-depth analysis on changes in specific T-lymphocyte subsets and marker expression with clinical data using MaAsLin2 further revealed no significant associations.

Following on from this, as with bacterial and metabolomic profiles, samples were further clustered based on their overall T-lymphocyte profiles (i.e., based on the proportion of the different 20 identified T-lymphocyte sub-populations) to characterise PTPTs (see Methods). Using a consensus-based algorithm, two clusters were found to be optimal which were then defined using a hierarchical clustering approach based on complete linkage. A binomial mixed effects model found that no clinical variable that was tested for, was associated with PTPTs whilst controlling for patient ID.

5.3.5. Stability of the microbiome and T-lymphocyte profile over time

Infants with two or more samples were analysed for microbiome stability ($n = 61$). Infant PGCT was found to change once on average across all timepoints, and expectedly increased with the number of samples given (P = 0.031). Aside from sample number, microbiome stability was only found to be associated with trial arm (P = 0.008), with infants in the control trial arm (cow's milk formula and fortification if needed) being found to have a significantly more 'unstable' microbiome. A second model was fit to look at the number of transitions between PGCTs as an outcome, with both sample number (P = 0.038) and trial arm (P = 0.006) again being significantly associated. To better control for the number of samples given by each infant, since this was significantly associated, rather than rely on adjustment within the model, the same analyses were carried out on infants who had a sample in all five of the timepoints (*n*

= 30 infants, 45% of the cohort). The same clinical co-variates were included, and none were found to be significantly associated with microbiome stability based on the binary outcome (stable/unstable) or the number of transitions between PGCTs. Specifically, trial arm, which was associated with microbiome stability and the number of transitions when adjusting for the number of samples given in the model, was no longer significant (microbiome stability P = 1; number of transitions P = 0.086). However, it is important to note that there were expectedly far fewer infants classed as 'stable' (*n* = 2) for the microbiome stability model, as the analysis spanned all time points, which limits the reliability and applicability of this model.

Overall T-lymphocyte profiles were found to be stable over time within infants, with PTPTs remaining stable post versus pre-fortification. Specifically, of the 14 infants who had two or more CyTOF samples, 13 of them were found to have the same PTPT pre- and post-fortification.

5.3.6. Associations between datasets

5.3.6.1. Gut microbiome and stool metabolome

To determine if there were any associations between the three datasets, first focusing on the gut microbiome and stool metabolome, PERMANOVA was used. Based on this, PGCT had a significant association with the overall metabolome, explaining 2.35% of variation in the data whilst stratifying by patient ID (P = 0.038; **Figure 5.8**). Specifically, 82 metabolites were significantly associated with the different PGCTs when PGCT-1 (*Staphylococcus* dominant) was used as a reference (including both positive and negative coefficients), based on MaAsLin2 analysis (**Table 5.4**). Amongst the most significant for each PGCT included three amino acids for PGCT-2 (*Enterococcus* dominant), two involved in tyrosine metabolism; tyramine O-sulphate and tyramine, and one involved in lysine metabolism, 5-hydroxylysine. For PGCT-3 (*Escherichia/Shigella* dominant), there were 41 significantly associated metabolites, including 20 amino acids, six of which were involved in leucine, isoleucine and valine metabolism, and 11 lipids. There were 13 metabolites significantly associated with PGCT-4 (*Enterobacter/Klebsiella* dominant), five of which were lipids including lysophospholipids and phosphatidylethanolamines. There were 93 metabolites significantly associated with PGCT-5 (diverse cluster), the majority of which (60%) were attributable to unknown/unnamed metabolites.

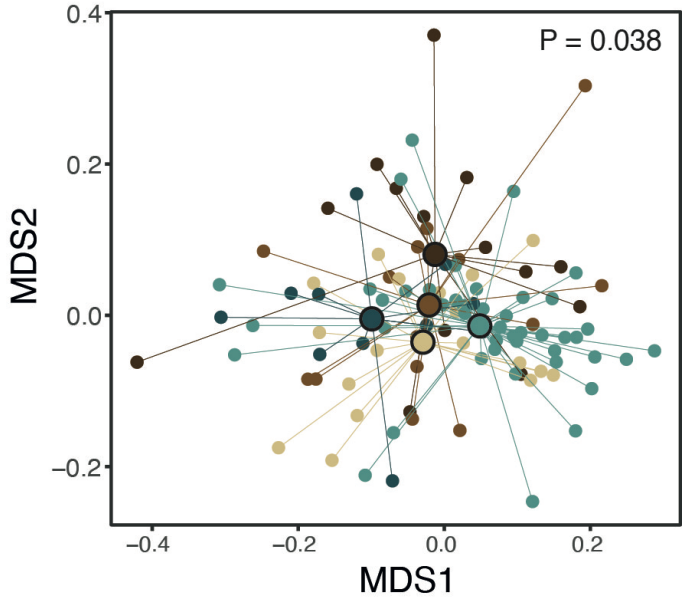

Figure 5.8. Gut metabolite profiles significantly differ based on gut PGCT. NMDS based on gut metabolite profiles, coloured by PGCT. P-value is based on PERMANOVA.

Table 5.4. MaAsLin2 results for significant metabolites associated with PGCTs. Mixed-effects linear models using a variance-stabilising arcsin square root transformation on metabolite data were used to determine the significance. Patient ID was included as a random effect.

PGCT	Coeff	stderr	pval	qval	Super pathway	Sub pathway	Chemical name
2	0.016	0.005	0.001	0.044	Amino Acid	Tyrosine Metabolism	tyramine
2	0.018	0.005	0.001	0.033	Amino Acid	Lysine Metabolism	5-hydroxylysine
2	0.032	0.007	<0.001	0.003	Amino Acid	Tyrosine Metabolism	tyramine O-sulfate
3	0.021	0.004	<0.001	0.002	Amino Acid	Polyamine Metabolism	putrescine
3	0.02	0.005	<0.001	0.016	Energy	TCA Cycle	fumarate
3	0.009	0.002	0.001	0.033	Amino Acid	Leucine, Isoleucine and Valine Metabolism	isoleucine
3	0.009	0.002	<0.001	0.026	Amino Acid	Leucine, Isoleucine and Valine Metabolism	leucine
3	-0.017	0.005	<0.001	0.027	Amino Acid	Glycine, Serine and Threonine Metabolism	betaine
3	-0.045	0.011	<0.001	0.013	Amino Acid	Alanine and Aspartate Metabolism	asparagine
3	0.02	0.004	<0.001	0.005	Amino Acid	Tyrosine Metabolism	tyramine
3	0.034	0.009	<0.001	0.03	Amino Acid	Phenylalanine Metabolism	phenethylamine
3	0.019	0.005	0.001	0.039	Lipid	Phosphatidylethanolamine (PE)	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
3	0.024	0.005	<0.001	0.005	Amino Acid	Lysine Metabolism	cadaverine
3	-0.028	0.007	<0.001	0.011	Nucleotide	Pyrimidine Metabolism, Cytidine containing	2'-deoxycytidine
3	-0.05	0.012	<0.001	0.007	Carbohydrate	Aminosugar Metabolism	glucuronate
3	0.031	0.008	<0.001	0.015	Amino Acid	Tryptophan Metabolism	3-indoxyl sulfate
3	0.015	0.004	0.001	0.032	Lipid	Fatty Acid, Monohydroxy	3-hydroxymyristate
3	0.069	0.013	<0.001	0.001	Lipid	Secondary Bile Acid Metabolism	7-ketodeoxycholate
3	0.041	0.011	<0.001	0.03	Lipid	Secondary Bile Acid Metabolism	7-ketolithocholate
3	0.019	0.004	<0.001	<0.001	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N-acetylisoleucine
3	0.015	0.004	<0.001	0.029	Amino Acid	Leucine, Isoleucine and Valine Metabolism	alpha-hydroxyisovalerate
3	0.018	0.005	0.001	0.048	Amino Acid	Leucine, Isoleucine and Valine Metabolism	2-hydroxy-3-methylvalerate
3	0.026	0.004	<0.001	<0.001	Lipid	Lysophospholipid	1-palmitoyl-GPE (16:0)
3	0.027	0.005	<0.001	0.001	Lipid	Lysophospholipid	1-oleoyl-GPE (18:1)
3	0.015	0.004	<0.001	0.023	Lipid	Corticosteroids	cortisone 21-sulfate
3	0.025	0.006	<0.001	0.005	Cofactors and Vitamins	Pantothenate and CoA Metabolism	pantoate
3	0.026	0.007	<0.001	0.016	Lipid	Phospholipid Metabolism	trimethylamine N-oxide
3	0.038	0.011	0.001	0.043	Amino Acid	Tyrosine Metabolism	tyrosol
3	0.011	0.002	<0.001	0.002	Xenobiotics	Food Component/Plant	indolin-2-one
3	0.016	0.003	<0.001	<0.001	Amino Acid	Tryptophan Metabolism	N-formylanthranilic acid
3	0.021	0.005	<0.001	0.01	Lipid	Lysophospholipid	1-palmitoyl-GPG (16:0)*
3	0.024	0.007	0.001	0.032	Amino Acid	Tyrosine Metabolism	tyramine O-sulfate
3	0.03	0.007	<0.001	0.011	Amino Acid	Glutathione Metabolism	2-hydroxybutyrate/2-hydroxyisobutyrate
3	0.017	0.005	0.001	0.048	Energy	TCA Cycle	2-methylcitrate/homocitrate
3	0.017	0.004	<0.001	0.016	Lipid	Phosphatidylethanolamine (PE)	1,2-dipalmitoyl-GPE (16:0/16:0)*
3	0.018	0.005	<0.001	0.027	Amino Acid	Polyamine Metabolism	(N(1) + N(8))-acetylspermidine
3	0.013	0.004	0.001	0.031	Lipid	Corticosteroids	11-dehydrocorticosterone sulfate

3	0.027	0.007	<0.001	0.019	Amino Acid	Phenylalanine Metabolism	N-succinyl-phenylalanine
3	0.029	0.007	<0.001	0.014	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N-succinyl-leucine
3	0.066	0.013	<0.001	0.001	Amino Acid	Polyamine Metabolism	N-carbamoylputrescine
3	0.063	0.011	<0.001	<0.001	Unknown	Unknown	12216
3	0.027	0.005	<0.001	<0.001	Unknown	Unknown	13729
3	-0.018	0.004	<0.001	0.002	Unknown	Unknown	22162
3	-0.033	0.009	0.001	0.032	Unknown	Unknown	23654
4	0.012	0.003	0.001	0.033	Amino Acid	Polyamine Metabolism	1,3-diaminopropane
4	0.018	0.005	<0.001	0.02	Lipid	Phosphatidylethanolamine (PE)	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)
4	-0.041	0.01	<0.001	0.01	Carbohydrate	Aminosugar Metabolism	glucuronate
4	0.012	0.003	<0.001	0.011	Amino Acid	Leucine, Isoleucine and Valine Metabolism	N-acetylisoleucine
4	0.017	0.003	<0.001	0.002	Lipid	Lysophospholipid	1-palmitoyl-GPE (16:0)
4	0.018	0.005	<0.001	0.014	Lipid	Lysophospholipid	1-oleoyl-GPE (18:1)
4	-0.01	0.003	0.001	0.043	Cofactors and Vitamins	Tocopherol Metabolism	gamma-CEHC glucuronide*
4	0.022	0.004	<0.001	0.001	Lipid	Lysophospholipid	1-palmitoyl-GPG (16:0)*
4	0.015	0.004	<0.001	0.014	Lipid	Phosphatidylethanolamine (PE)	1,2-dipalmitoyl-GPE (16:0/16:0)*
4	-0.017	0.003	<0.001	0.001	Unknown	Unknown	22162
4	0.019	0.004	<0.001	0.004	Unknown	Unknown	25837
4	-0.017	0.005	0.001	0.038	Unknown	Unknown	25950
4	0.012	0.003	<0.001	0.015	Unknown	Unknown	25955
5	0.015	0.003	<0.001	0.004	Nucleotide	Purine Metabolism, Adenine containing	1-methyladenine
5	0.037	0.009	<0.001	0.005	Amino Acid	Polyamine Metabolism	N-acetylputrescine
5	0.044	0.01	<0.001	0.005	Amino Acid	Phenylalanine Metabolism	phenethylamine
5	0.073	0.02	0.001	0.032	Energy	TCA Cycle	tricarballylate
5	0.026	0.007	<0.001	0.019	Amino Acid	Lysine Metabolism	N6-acetyllysine
5	0.018	0.004	<0.001	0.001	Amino Acid	Urea cycle; Arginine and Proline Metabolism	dimethylarginine (SDMA + ADMA)
5	0.027	0.007	<0.001	0.025	Amino Acid	Tyrosine Metabolism	tyramine O-sulfate
5	0.03	0.007	<0.001	0.004	Peptide	Dipeptide	cyclo(pro-sulfo-tyr)*
5	0.02	0.006	0.001	0.044	Amino Acid	Histidine Metabolism	1-methyl-5-imidazolelactate
5	0.024	0.004	<0.001	<0.001	Lipid	Primary Bile Acid Metabolism	chenodeoxycholic acid sulfate (2)
5	0.012	0.003	<0.001	0.014	Unknown	Unknown	12708
5	0.011	0.003	0.001	0.049	Unknown	Unknown	13553
5	0.022	0.005	<0.001	0.014	Unknown	Unknown	13729
5	0.032	0.007	<0.001	0.005	Unknown	Unknown	19220
5	0.026	0.006	<0.001	0.005	Unknown	Unknown	19917
5	0.027	0.006	<0.001	0.003	Unknown	Unknown	19921
5	0.027	0.007	0.001	0.031	Unknown	Unknown	19928
5	0.088	0.026	0.001	0.043	Unknown	Unknown	23662
5	0.026	0.008	0.001	0.037	Unknown	Unknown	23908
5	0.026	0.007	<0.001	0.03	Unknown	Unknown	24474
5	0.057	0.013	<0.001	0.004	Unknown	Unknown	25053
5	0.017	0.004	<0.001	0.019	Unknown	Unknown	25823
5	0.021	0.005	<0.001	0.014	Unknown	Unknown	25830
5	0.096	0.024	<0.001	0.013	Unknown	Unknown	25832
5	0.013	0.003	<0.001	0.014	Unknown	Unknown	25853

Spearman's rank correlation analyses between all the 977 metabolites and the 10 most abundant taxa (*Enterobacter/Klebsiella, Bifidobacterium, Escherichia/Shigella, Staphylococcus*, *Enterococcus, Veillonella, Lactobacillus, Bacteroides, Acinetobacter* and *Ligilactobacillus*) revealed 706 significant relationships after FDR adjustment (427 positive and 279 negative). The highest number of significant correlations were attributable to *Lactobacillus* (31%), the most significant of which were three xenobiotics (equol sulphate, ferulic acid 4-sulphate, enterolactone sulphate) which were all positively correlated to *Lactobacillus* relative abundance. *Staphylococcus* also had a high number of significant correlations (26%), both positive and negative. The three most significant were negative correlations with phospholipids including lysophospholipid and phosphatidylethanolamine (all $P < 0.001$). In contrast, these metabolites were significantly positively correlated to *Enterobacter/Klebsiella* relative abundance, and correspondingly were all also found to be significantly enriched in PGCT-4 (*Enterobacter/Klebsiella* dominant). There were 37 metabolites that were significantly correlated with *Veillonella* relative abundance, both positive and negative, with a positive correlation being identified with one of the acylcarnitines significantly associated with fortifier use.

5.3.6.2. Gut microbiome and blood T-lymphocyte profiles

There was no significant association between PGCT and overall T-lymphocyte profile ($P = 0.2809$), based on PERMANOVA whilst stratifying by patient ID. There was also no significant association between PGCT and CD4:CD8 ratio based on LMMs ($P = 0.779$), whilst controlling for patient ID. However, there was a significant association between PGCTs and T-lymphocyte marker expression profile ($P = 0.025$), whilst stratifying by patient ID. Based on MaAsLin2 analysis, RORγ was significantly higher in PGCT-3 ($P < 0.001$, $Q = 0.011$) and PGCT-4 ($P< 0.001$, $Q = 0.011$) associated samples, compared to PGCT-1 (**Figure 5.9**). CD49b and CD16 were also significantly lower in PGCT-4 ($P < 0.001$, $Q = 0.011$ and $P = 0.001$, $Q = 0.046$, respectively) associated samples compared to PGCT-1 (**Figure 5.9**).

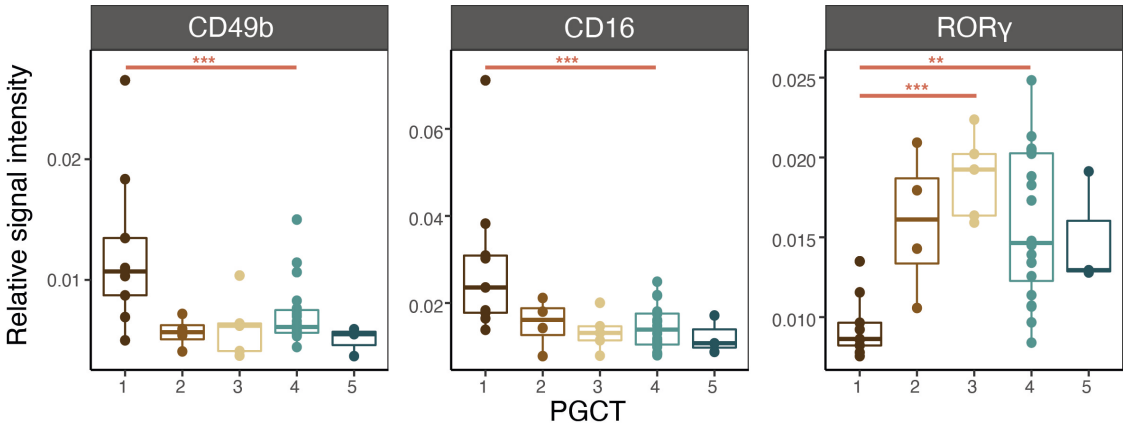

Figure 5.9. Relative signal intensity of T-lymphocyte markers across matched gut PGCTs. Significance is based on MaAsLin2 analysis, where (**) denotes $P < 0.01$ and (***) denotes $P < 0.001$.

Spearman's rank correlation analysis was carried out between the 10 most abundant taxa identified in patients with CyTOF data (*Enterobacter/Klebsiella, Bifidobacterium, Escherichia/Shigella, Staphylococcus*, *Enterococcus, Veillonella, Ligilactobacillus, Bacteroides, Acinetobacter* and *Clostridium)*, and the 20 T-lymphocyte subsets. No significant correlations were identified after adjustment for multiple comparisons. The same analyses were carried out to look for correlations between the same top 10 taxa and specific markers, as well as CD4:CD8 ratio, further showing no significant correlations after adjustment for multiple comparisons.

5.3.6.3. Stool metabolome and blood T-lymphocyte profiles

Associations between the metabolome and T-lymphocyte profiles were next explored. Spearman's rank correlation analyses between all the 977 metabolites and the 20 T-lymphocyte subsets revealed no significant associations after adjustment for multiple comparisons. Looking at correlation between metabolites and the CD4:CD8 ratio also revealed no significant correlations after adjustment for multiple comparisons.

There was, however, five significant associations between all the metabolites and marker expression, after adjustment for multiple comparisons. These included a strong positive correlation between phosphate and CXCR3 ($P = 0.033$) and a strong negative correlation between 1-methylhypoxanthine and CXCR3 ($P = 0.034$) (**Figure 5.10A, B**). A strong positive correlation was observed between N-propionylmethionine and CD69

(P = 0.012), whilst a strong negative correlation between ribotol and CD69 (P = 0.012) was identified (**Figure 5.10C, D**). Finally, a strong positive correlation between BHB and γδ TCR (P < 0.001) was identified (**Figure 5.10E**).

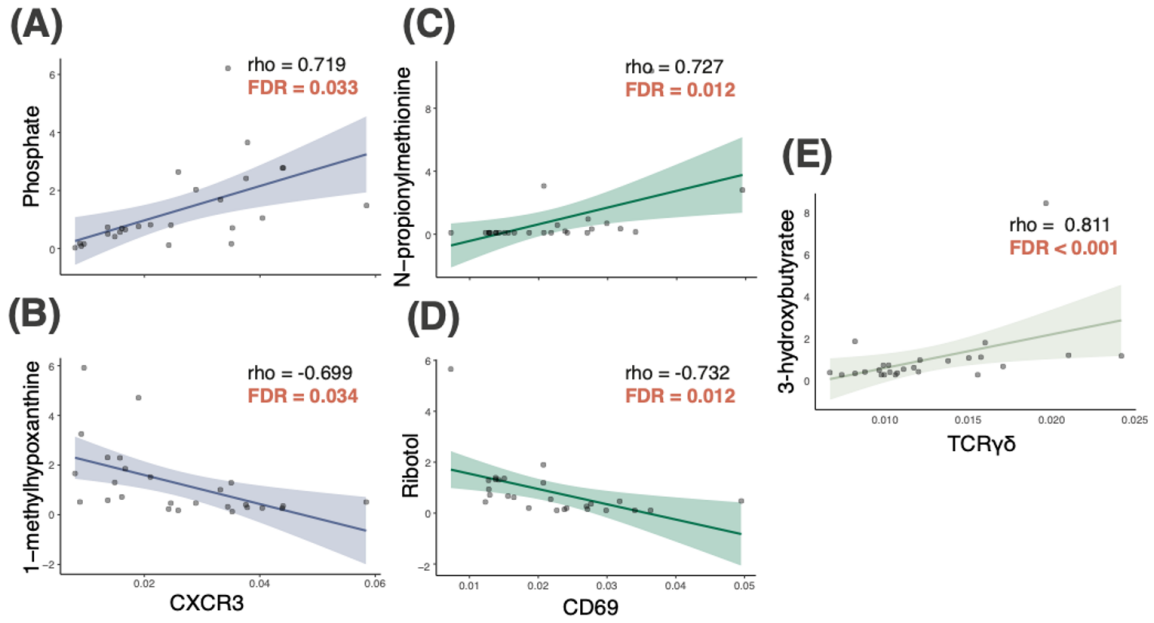

Figure 5.10. Specific gut metabolites are significantly correlated with specific T-lymphocyte markers. Correlation plots for the relative abundance and relative signal intensity (respectively) of **(A)** Phosphate and CXCR3, **(B)** 1-methylhypoxanthine and CXCR3, **(C)** N-propionylmethionine and CD69, **(D)** Ribotol and CD69, and **(E)** 3-hydroxybutyrate (BHB) and TCRγδ.

5.3.6.4. Multiple co-inertia analysis and Procrustes analysis

Finally, MCIA and PA were used to integrate the three datasets (bacterial, metabolite and T-lymphocyte marker profiles) and visualise any relationships between them. MCIA is a multivariate co-inertia analysis that looks to identify co-relationships between multiple datasets, transforming and plotting datasets onto the same projection. For the MCIA, the figure shows the projection of all samples from the different datasets onto the first two principal components of the MCIA, where the datasets have been transformed into the same projection (**Figure 5.11A**). The coordinates of the samples in each of the three datasets are connected by edges, the length of which indicates the divergence (i.e., the shorter the line, the higher the level of concordance) between the datasets for a particular sample/matched sample (i.e., stool sample and matched blood). Pairwise Monte-Carlo Tests on the sum of eigenvalues from the MCIA revealed no significant correlation between any of the datasets (bacterial – metabolite: RV coefficient = 0.140, P = 0.995; bacterial – T-lymphocytes: RV coefficient = 0.355, P =

0.264; metabolite – T-lymphocytes: RV coefficient = 0.191, P = 0.510). Although samples and matched samples were found to cluster somewhat into PGCT groups (**Figure 5.11A**).

PA is similar to MCIA, and again can be useful for evaluating relationships between multiple datasets. For the PA, PCoA coordinates for each of the three datasets were transformed and superimposed onto the same space (**Figure 5.11B**). Again, edges connect the three datasets from a centroid. In contrast to the MCIA, pairwise correlation analyses between the transformed PCoA coordinates for each of the three datasets revealed a significant correlation between bacterial and metabolite profiles (dissimilarity parameter m^2 = 0.398, P = 0.002) but not between bacterial and T-lymphocyte marker profiles (dissimilarity parameter m^2 = 0.457, P = 0.519) or metabolite and T-lymphocyte marker profiles (dissimilarity parameter m^2 = 0.534, P = 0.617).

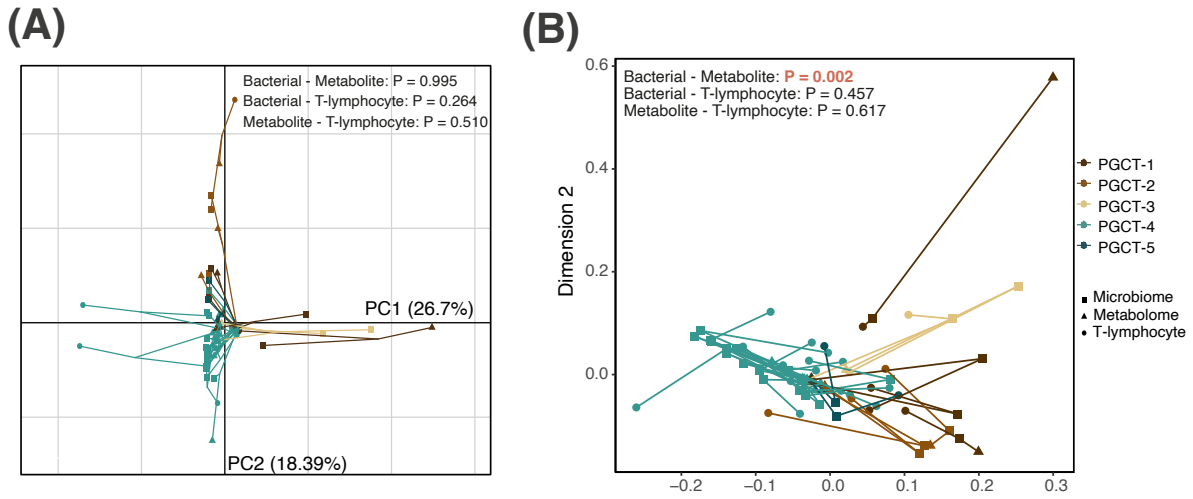

Figure 5.11. Gut bacterial and metabolite profiles are significantly correlated, but neither are correlated with circulating T-lymphocytes. (A) MCIA plot of the three datasets where edges connect bacterial profiles (square), metabolite profiles (triangle) and T-lymphocyte profiles (triangle) for each individual sample or matched sample, coloured by PGCT. Samples have been projected into the same dimensional space based on the first two principal components, each explaining 26.7% and 18.39% of variation in the data, respectively. Shorter edges depict greater similarity between datasets. **(B)** PA based on superimposition of coordinates from PCoA of bacterial profiles (square), metabolite profiles (triangle) and T-lymphocyte profiles (circle). Samples and matched samples are connected by edges with the shorter edges representing greater similarity between datasets.

5.4. Discussion

In this longitudinal multi-omic study, the relationship between clinical data, the microbiome, metabolome and circulating T-lymphocytes in preterm infants was investigated. Additionally, the correlation between the three datasets and how they might be interrelated was explored. The aims of this research were to contribute to an enhanced understanding of the link between preterm infant clinical characteristics, their gut microbiome, metabolome and systemic immunity.

As part of the clinical information, the impact of time (DOL, CGA and time-point) on the microbiome, metabolome and T-lymphocyte populations was examined. The results showed that whilst time had no significant impact on changes in the overall T-lymphocyte populations of preterm infants, it did impact the microbiome and metabolome. Time-point was found to explain the most variation in both datasets, likely due to the focus on discrete time-points in the INDIGO study design and sample collection based around these (rather than regular temporal sampling). It is important to note that in both instances, DOL explained more variation than CGA, which was similarly observed in Chapter 3. There is an existing debate as to whether time since birth or CGA (taking into account gestational age at birth) is best used as a measure of time in longitudinal studies of preterm infants. The current results highlight that DOL may be a more reliable measure of time in comparison to CGA, when measuring preterm gut bacteria and luminal metabolites. It is also worth highlighting that although no significant association between time and circulating T-lymphocytes was identified, this could be due to the limited number of samples (n = 41) that were mostly in two timepoints (C and E). Therefore, this could reflect that there were not enough samples spanning the breadth of time to robustly analyse the impact and any changes over this shorter time frame could not be captured. It could be that a difference may have been found when analysing samples taken at the start of life with a later sample.

When analysing other clinical factors aside from time, there was no significant impact on overall bacterial profiles at the different timepoints. This is consistent with previous research in the Newcastle cohort, where the main drivers of microbial structure in healthy preterm infants were time and probiotic use (Beck et al., 2022) (Chapter 3). Additionally, no clinical factors were found to be significantly associated with the healthy preterm gut in the Olm *et al.* cohort (Olm et al., 2019), a US based cohort of

preterm infants who did not take probiotics. All samples in this cohort were taken during the Labinic era, so the impact of probiotics was not evaluated.

Similarly, no clinical factors were found to be significantly associated with T-lymphocyte sub-population or marker expression before or after fortifier treatment. On the other hand, a significant association between clinical data and the overall metabolite profiles was found at merged timepoint CD and in particular time-point E. All of the significant covariates were surrounding feeding in some way (e.g., % enteral MOM in previous 3 days, day of full feeds, fortifier at time of sample and trial arm). Although these factors were identified to be significantly associated with changes in the overall metabolome at one or more timepoints, the explained variance was found to be much lower than that observed for the microbiome or T-lymphocytes, despite none of these being identified as significant.

Upon investigation of the association between clinical data and the relative abundance of bacteria and metabolites, several significant associations were identified, particularly with specific metabolites. Of the five significant associations with bacterial genera, fortifier use was the only clinical factor found to be significantly associated with any bacteria aside from time. Use of fortifier was found to be significantly associated with an increase in the relative abundance of *Veillonella,* as well as the abundance of three specific acylcarnitines. One of these acylcarnitines was found to be directly correlated with the relative abundance of *Veillonella*, as well as five additional acylcarnitines. *Veillonella* has been shown to produce hydrogen sulphide (Washio et al., 2014), which has been reported to inhibit the transport of acylcarnitines into the mitochondria for beta oxidation (Giangregorio et al., 2016). This could have led to an accumulation of acylcarnitines which have been subsequently excreted and detected in the stool, potentially explaining the correlation between *Veillonella* and acylcarnitines. Aside from this, the majority of the 120 significant associations between clinical data and the metabolome were attributable to trial arm. It is likely, however, that these differences in metabolites are in fact reflective of the nutritional differences between cows' milk and human milk-based fortifier which provide a direct source of metabolites to the gut.

The study found 706 significant relationships between the top 10 most abundant taxa in the gut and all 977 metabolites, after adjusting for multiple comparisons. The majority

of these associations (31%) were observed with *Lactobacillus*, the three most significant of which were positive correlations with xenobiotics. It is interesting to note that equol, which is an oestrogen known to be produced by intestinal bacteria, is positively correlated with *Lactobacillus* in its sulphate form, as *Lactobacillus* is a known producer of equol (Kwon et al., 2018). However, although this study is one of the largest combining bacterial and metabolite profiling in preterm infants, it should be noted that many of these significant correlations may be influenced by outliers.

Analysis revealed limited association between T-lymphocyte populations or the ratio of CD4:CD8 T-lymphocytes and the microbiome or metabolome. However, a number of significant associations when looking at T-lymphocyte markers were found. Microbial PGCT was found to be significantly associated with T-lymphocyte marker expression, with some markers being significantly associated with specific PGCTs. RORγ, for example, was found to be significantly positively associated with PGCT-3 and PGCT-4 when compared to PGCT-1. RORγ is usually expressed on Th17 cells, which play a role in maintaining mucosal homeostasis and clearing extracellular pathogens from these mucosal sites (Khader et al., 2009). Whilst the microbes dominating these PGCTs have not been previously implicated in relation to RORγ, it is worth mentioning that previous studies have established a link between the gut microbiota and RORγ T-lymphocytes (Ohnmacht et al., 2015).

Regarding specific metabolites associated with T-lymphocyte markers, five significant associations between specific metabolites and marker expression were identified. One particularly interesting finding was that BHB, which is a ketone, had a strong positive correlation with TCRγδ. Previous research has shown that ketogenic diets can activate protective γδ T-lymphocyte responses (Goldberg et al., 2019). Furthermore, that the elevation of ketone bodies in the circulatory system has been implicated in immunological barriers, impacting the microbiota, mucosal layer and mucosal tissue by various means (Qi et al., 2022). BHB is involved in the formation of milk fats in the mammary gland and known to be secreted in breast milk, having been detected in previous human milk metabolomic studies (Alexandre-Gouabau et al., 2019). Receiving BHB through breast milk may therefore have an impact on systemic γδ T-lymphocyte proliferation, but this would need to be confirmed experimentally in future work.

Finally, the three datasets were integrated, looking for interconnections and how they might be linked together. Despite the lack of significance from the Monte-Carlo tests on the sum of eigenvalues from the MCIA, clustering into the different PGCTs across all three datasets was observed. PA analysis further revealed significant concordance between bacterial and metabolite profiles, but not between bacterial or metabolite profiles and T-lymphocyte markers. Further to this, the data demonstrated clear clustering into PGCTs across all three datasets. The results suggest a link between microbes and metabolites in the gut, which is supported by previous results in preterm infants, as shown in Chapter 3 (Beck et al., 2022). When considering circulating T-lymphocytes it is perhaps unsurprising that there is less correlation with the contents of the gut lumen, which again is supported by previous research and results presented in Chapter 3, suggesting little correlation between the gut microbiome and blood metabolome (Beck et al., 2022). This could be due to the lack of stability and dynamic nature of the preterm gut. In adult populations, where the gut microbiome is thought to be more stable, a much greater correlation has been seen between the gut lumen and blood contents (Diener et al., 2022).

5.5. Limitations

Limitations in the immunological aspect of this study include the sole focus on T-lymphocyte populations, without investigating other immune components such as B cells, natural killer cells or macrophages. Biases toward the study of this specific population of immune cells however could be explained by the study design, whereby a marker panel that was specifically designed to analyse T-lymphocyte sub-populations was set up. This specific panel would not have been sufficient for investigating other immune cell sub-populations. In this study, there was only a relatively minor impact on T-lymphocyte populations by clinical data, microbiome and metabolome, which could be due to blood reflecting only systemic and not gut-specific impacts, with direct gut impacts perhaps not being captured in blood. However, it is possible other immune components are impacted by these factors. To overcome this, future studies should design panels that include additional markers to allow more in-depth analysis and comprehensive overview of other immune components.

Further to this, if considering the study of additional immune cells, consideration may be necessary regarding blood sample collection and storage, depending on the targeted leukocyte population. This is because some blood cells, mostly those of myeloid origin such as macrophages and neutrophils, have a half-life time of less than 24 hours (Diks et al., 2019). Furthermore, different leukocytes populations have different preferred storage conditions, including temperatures and buffer preferences. In addition, the number of matched blood samples available for the microbiome was low (n = 41) and even lower when including the metabolome (n = 21). This aspect of the study was therefore not powered to find robust associations with circulating T-lymphocytes. Instead of casting a broader net with additional markers as discussed above, future work may consider remaining focused on T-lymphocyte populations but increasing the sample size instead. This could be done in combination with using an alternative method to CyTOF altogether, such as FC, which offers greater resolution whilst targeting fewer markers.

Regarding the gut microbiome, this was based on 16S rRNA gene sequencing data and so was restricted to genus-level identification. It is possible that a deeper analysis using shotgun metagenomics to allow species and strain level identification may have revealed species or strain-specific interactions with circulating T-lymphocytes.

Finally, this analysis is descriptive in nature, and cause or effect cannot be determined, as with previous chapters. Furthermore, from these datasets it cannot be determined whether metabolites were host, diet or microbially derived. Future work may therefore consider experimentally determining whether specific metabolites of interest were microbially derived (i.e., whether preterm bacterial strains produce specific metabolites in culture), and where relevant, these metabolites could be cultured with immune cell lines to determine whether they impact T-lymphocyte or other immune populations. Ultimately, these data therefore provide an important basis for future experiments and analysis.

5.6. Conclusions

In summary, age was found to have the biggest impact on both the gut microbiome and metabolome but was not found to impact circulating T-lymphocytes, over the time frame of the current study. The greater similarity between gut microbiome and gut

metabolome was continuously observed, and these datasets showed the most concordance, likely due to the direct proximity of these gut/stool-based data. In contrast, fewer relationships were found with circulating T lymphocytes, particularly specific sub-populations, for which no significant associations were observed. This could reflect that the relationships between blood and the gut lumen in preterm infants are too dynamic to capture, especially in relation to their spatial proximity. Looking at specific features of the data rather than overall profiles, no significant associations between specific T-lymphocyte markers and PGCT as well as various specific luminal metabolites were found, suggesting there may be some very specific connections between the gut lumen and systemic immunity. To confirm these findings, further work would be needed with a higher number of CyTOF samples. Additionally, experimental validation of the associations identified would be needed to determine cause or effect.

Overall, this chapter demonstrates the potential to perform and integrate multiple omic datasets from a range of longitudinal preterm infant samples, within the framework of a RCT. Metabolomics revealed the most significant features associated with the RCT trial arm. However, it is important to note that this may ultimately reflect the nutritional differences, which introduce a direct source of distinct metabolites. Incorporating temporal sampling and multi-omic analysis into future RCTs in preterm infants holds immense potential for mechanistic exploration and ensure added valuable insights when evaluating novel clinical interventions.

6. General discussion

6.1. Discussion

The gut microbiome is a complex community of microorganisms residing in the GI tract, that has been implicated in preterm health and disease, including its association with the onset of NEC (Coleman et al., 2023; McMurtry et al., 2015; Olm et al., 2019; Pammi et al., 2017; Stewart et al., 2012; Stewart et al., 2016; Warner et al., 2016). Whilst there is substantial knowledge regarding host and environmental factors that influence the gut microbiome of infants born full-term, our understanding of what drives gut microbial changes in preterm infants is comparatively more ambiguous. Delineating infant factors that influence the preterm gut microbiome is therefore increasingly important.

The studies in Chapter 3 and 4 represent two of the largest longitudinal metagenomic studies in a healthy preterm cohort and a NEC cohort, respectively. Previous studies have found conflicting results, which may owe to smaller cohorts, cross-sectional data and use of 16S rRNA gene sequencing. Therefore, this data allows a much more in-depth exploration of microbial changes, otherwise difficult to capture cross-sectionally, due to the highly dynamic nature of the preterm gut microbiome.

Infant factors shaping the gut microbiome in preterm infants who did not go onto develop NEC were identified and explored, revealing the specific features underpinning those associations. Probiotics were found to be the main driver of changes to the gut microbiome, and different probiotic products were found to have differential impact on the taxonomic composition of the communities. Regardless, both products drove transition into *Bifidobacterium*-rich communities, that were more functionally comparable than infants who did not receive probiotics. The study provided a highly necessary foundation and will help provide a framework for identifying important aspects for consideration when designing interventional trials targeting the gut microbiome of preterm infants.

Comparing this ‘healthy’ cohort to infants who went on to develop NEC revealed differential patterns in microbiome development at the start of life and microbial signatures prior to NEC onset such as an increase in the relative abundance of *K. variicola* and a decrease in the relative abundance of *Bifidobacterium* spp. Notably, *K.*

variicola has been noted as an emerging human pathogen (Rodríguez-Medina et al., 2019), likely previously misidentified as *K. pneumoniae*. Whilst not specifically implicated in disease onset in previous studies, *Klebsiella* spp. and in particular *K. pneumoniae* have previously been found to be significantly higher in the days leading up to disease onset (Coleman et al., 2023; Olm et al., 2019; Paveglio et al., 2020; C. Zhang et al., 2012). That said, *Klebsiella* were one of the most abundant genera in the healthy cohort and so are not solely responsible for NEC based on their high abundance, in fact the onset of NEC is clearly much more complex, hence being described a multifactorial disease (Neu, 2020). Conversely, the reduction in the relative abundance of *Bifidobacterium* supports the notion that *Bifidobacterium* have more health promoting benefits, and are potentially protective against NEC in preterm infants, although a mechanism of protection is not fully understood. Whilst these results do support those associations found in previous studies (Coleman et al., 2023; Olm et al., 2019; Pammi et al., 2017; Stewart et al., 2016; Warner et al., 2016), it is also important to note that the results described here all rely on relative abundance data. Relative abundance data refers to the proportion of each microbe in relation to the total number of microbes. This means that absolute numbers cannot be deduced, and it is therefore unclear whether *K. variicola,* for example*,* truly increases prior to disease onset or whether another microbe decreases (i.e, *Bifidobacterium*) or vice versa. To overcome this, absolute quantification of significant features identified, such as qPCR could be used to validate findings. Alternatively, future studies where data is yet to be acquired could look to use internal standards, for example the spike-in method (either DNA-based or cell-based), where known quantities of synthetic DNA or DNA/cells from rare gut microbes are added to samples (Rao et al., 2021; Zaramela et al., 2022).

Differences were observed dependent on whether probiotics were received, or the type of probiotic received. This is the first-time microbial signatures of NEC have been explored within the context of probiotics and it was evident that the two products showed differences in persistence and engraftment of their respective species in infants who developed disease. This in itself represents an interesting point for discussion regarding targeted probiotic intervention, and further exemplifies how probiotics unlikely follow a ‘one-size-fits-all’ convention, with microbial changes being highly individualised. The only consistent predicters of probiotic persistence was the type of probiotic used (i.e., the different strains between the different products) and

whether an infant developed NEC. Whilst the latter cannot feasibly be used clinically to target probiotic intervention, since there is no specific indication preceding diagnosis, this still warrants further research.

Probiotics were found to significantly influence alpha diversity measures in healthy infants, with Shannon diversity being significantly higher in Labinic infants. Interestingly, whilst Shannon diversity has been found to be significantly lower before NEC diagnosis in other studies (McMurtry et al., 2015; Warner et al., 2016; Zhou et al., 2015), this was not the case in this cohort. Instead, whilst species richness was significantly lower prior to diagnosis, Shannon diversity was similar to no-NEC controls. This indicates that although there was a reduction in the number of species (richness) pre-NEC, the distribution of remaining species was unaffected. Whilst the gut microbiome clearly plays an important role in the development of NEC, demonstrated by some of the associations described here and elsewhere, it is evident that other factors (i.e., probiotics, age, individual subject) explain more variation in the gut microbial population of preterm infants. This is perhaps unsurprising, with NEC now being commonly accepted as an umbrella term for multiple disease subtypes (Berrington & Embleton, 2022; Neu et al., 2018), alongside the difficulty in being able to fully capture the period of time during which more subtle microbial changes occur prior to disease onset. Interestingly, the microbial origins of NEC appeared different dependent on probiotic use. Whilst these differences were observed, suggesting probiotics to potentially affect the microbial origins of NEC, it is important to note numerous confounding clinical variables could also explain the changes observed. In fact, this is potentially more reflective of the different discrete time periods during which different probiotics were administered. Regardless, this poses an intriguing avenue for follow-up study.

NEC can be considered an immune-mediated inflammatory disease, making the immune system in preterm infants an infinitely important aspect of consideration. In Chapter 5, data surrounding preterm systemic immunity was integrated with gut microbial and metabolomic data, in order to better understand the relationships between these parameters and to unpick this complex interplay within the population. Whilst previous studies have explored the microbial-immune axis in preterm infants (Lemme-Dumit et al., 2022; McDavid et al., 2022), no study to date has integrated

metagenomics, metabolomics and T-lymphocyte data. In fact, a recent study exploring gut microbiome and immune development in preterm infants suggested the use of metagenomics and metabolomics to better resolve microbial species and functionality (McDavid et al., 2022). Therefore, incorporating multiple high-dimensional datasets, allowed for a more holistic understanding of this complex system and a more comprehensive view of the interactions that occur between microbes, the host, and the environment. The gut microbiome and metabolome showed the greatest concordance, with little robust correlation seen between the circulating host immune system and the gut environment, consistent with the stool and serum metabolomics data presented in Chapter 3. It is likely that the majority of microbiota-mediated immune effects are site-specific and that these interactions between the gut microbiome and the immune system remain localised, meaning the impact on circulating T-lymphocytes would be minimal. Despite this, working with multiple high-dimensional datasets in this population still provides an important groundwork and direction for future studies. Together, this highlights the need to look at potential associations between the gut microbiome and immune system at more specific sites i.e., using resected intestinal tissue.

Whilst potentially important bacterial features of NEC were identified, consistent with previous research (Coleman et al., 2023; Olm et al., 2019; Pammi et al., 2017; Stewart et al., 2016; Warner et al., 2016), these identified associations would need to be incorporated into models of disease such as organoids. This would help to validate the identified associations and whether certain taxa offer anti-inflammatory, pro-inflammatory or other effects on host cells. In addition, building on identified associations with these models may offer mechanistic insights as to how these microbes may influence disease state. In summary, whilst metagenomic studies offer a huge amount of information, and a view of potential associations between the gut microbiome and NEC, following up with disease models is essential to transition from observational correlations to a deeper understanding of causality and mechanism, paving the way for therapeutic developments.

6.2. Future work

Since the focus of the results presented are largely based on observational studies and identified associations, future work should focus primarily on validating these

findings and confirming causality. Following on from the findings in Chapter 4 in particular, whereby potential biomarkers of NEC were identified, future work should focus on building on these identified associations. For example, these studies may wish to test specific bacteria on preterm organoids or other models of NEC and assessing whether they exacerbate or reduce inflammation, impact barrier integrity or alter gene expression profiles. This would help to better understand the pathophysiology of NEC and whether these identified associations are cause or effect. In addition, the correlations observed between specific metabolites and T-lymphocyte markers in Chapter 5 warrants validation using an immune cell model.

Furthermore, the work described in this thesis pertains to a single NICU and it is therefore important to consider differences across sites. To validate these findings, external cohorts could be used such, as done in Chapter 3.

One of the major drawbacks when integrating stool metabolomics and metagenomics, is that it cannot be easily deduced using current methodology whether metabolites are microbially derived, or whether they are derived from the host or as a direct dietary source. Although this is a near impossible challenge to fully deduce the origin of gut metabolites, future work may wish to use functional metagenomic data in parallel to see whether some of the metabolites identified could potentially be produced by the gut microbiota, and which members at least have the capacity of producing these. This may help to further unpick the complex interplay between diet, microbiota, metabolome and the immune system. Further to this, the associations identified with specific T-lymphocyte markers could be tested experimentally on T-lymphocyte lines in the lab i.e., exposing cells to the metabolites and pooled gut communities and assessing marker expression.

Finally, much of the work relies on the use of relative abundance, which as described can be misleading and is not necessarily an accurate representation of true abundance. To quantify significant taxa, qPCR could be used to confirm increases/decreases in relative abundance between groups. Additionally, future studies may wish to consider quantification using a spike-in method, as previously described.

7. References

Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science Translational Medicine*, *6*(237), 237ra65-237ra65. https://doi.org/10.1126/scitranslmed.3008599

Abdulkadir, B., Nelson, A., Skeath, T., Marrs, E. C. L., Perry, J. D., Cummings, S. P., Embleton, N. D., Berrington, J. E., & Stewart, C. J. (2016). Routine Use of Probiotics in Preterm Infants: Longitudinal Impact on the Microbiome and Metabolome. *Neonatology*, *109*(4), 239–247. https://doi.org/10.1159/000442936

Acuna-Gonzalez, A., Kujawska, M., Youssif, M., Atkinson, T., Grundy, S., Hutchison, A., Tremlett, C., Clarke, P., & Hall, L. J. (2023). Bifidobacterium bacteraemia is rare with routine probiotics use in preterm infants: A further case report with literature review. *Anaerobe*, *80*, 102713. https://doi.org/10.1016/J.ANAEROBE.2023.102713

Aguilar-Lopez, M., Dinsmoor, A. M., Ho, T. T. B., & Donovan, S. M. (2021). A systematic review of the factors influencing microbial colonization of the preterm infant gut. *Gut Microbes*, *13*(1), 1–33. https://doi.org/10.1080/19490976.2021.1884514

Ahearn-Ford, S., Berrington, J. E., & Stewart, C. J. (2022). Development of the gut microbiome in early life. *Experimental Physiology*, *107*(5), 415–421. https://doi.org/10.1113/EP089919

Ajami, N. J., Wong, M. C., Ross, M. C., Lloyd, R. E., & Petrosino, J. F. (2018). Maximal viral information recovery from sequence data using VirMAP. *Nature Communications*, *9*(1), 1–9. https://doi.org/10.1038/s41467-018-05658-8

Alcon-Giner, C., Dalby, M. J., Caim, S., Ketskemety, J., Shaw, A., Sim, K., Lawson, M. A. E., Kiu, R., Leclaire, C., Chalklen, L., Kujawska, M., Mitra, S., Fardus-Reid, F., Belteki, G., McColl, K., Swann, J. R., Kroll, J. S., Clarke, P., & Hall, L. J. (2020). Microbiota Supplementation with Bifidobacterium and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An Observational Study. *Cell Reports Medicine*, *1*(5), 100077. https://doi.org/10.1016/j.xcrm.2020.100077

Alganabi, M., Lee, C., Bindi, E., Li, B., & Pierro, A. (2019). Recent advances in understanding necrotizing enterocolitis [version 1; referees: 2 approved]. In *F1000Research* (Vol. 8). F1000 Research Ltd. https://doi.org/10.12688/f1000research.17228.1

Al-Sadi, R., Dharmaprakash, V., Nighot, P., Guo, S., Nighot, M., Do, T., & Ma, T. Y. (2021). Bifidobacterium bifidum Enhances the Intestinal Epithelial Tight Junction Barrier and Protects against Intestinal Inflammation by Targeting the Toll-like Receptor-2 Pathway in an NF-κB-Independent Manner. *International Journal of Molecular Sciences*, *22*(15). https://doi.org/10.3390/IJMS22158070

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, I. I., Umesaki, Y., Itoh, K., & Honda, K. (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. *Science*, *331*(6015), 337–341. https://doi.org/10.1126/SCIENCE.1198469/SUPPL_FILE/ATARASHI_SOM.PDF

Azad, M. B., Konya, T., Persaud, R. R., Guttman, D. S., Chari, R. S., Field, C. J., Sears, M. R., Mandhane, P. J., Turvey, S. E., Subbarao, P., Becker, A. B., Scott, J. A., & Kozyrskyj, A. L. (2016). Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: A prospective cohort study. In *BJOG: An International Journal of Obstetrics and*

Gynaecology (Vol. 123, Issue 6, pp. 983–993). Blackwell Publishing Ltd. https://doi.org/10.1111/1471-0528.13601

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., … Zhang, J. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. https://doi.org/10.1093/NAR/GKW1099

Beck, L. C., Granger, C. L., Masi, A. C., & Stewart, C. J. (2021). Use of omic technologies in early life gastrointestinal health and disease: from bench to bedside. *Expert Review of Proteomics*, 1–13. https://doi.org/10.1080/14789450.2021.1922278

Beck, L. C., Masi, A. C., Young, G. R., Vatanen, T., Lamb, C. A., Smith, R., Coxhead, J., Butler, A., Marsland, B. J., Embleton, N. D., Berrington, J. E., & Stewart, C. J. (2022). Strain-specific impacts of probiotics are a significant driver of gut microbiome development in very preterm infants. *Nature Microbiology 2020 7:10*, *7*(10), 1525–1535. https://doi.org/10.1038/s41564-022-01213-w

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. http://www.jstor.org/stable/2346101

Bergmann, K. R., Liu, S. X. L., Tian, R., Kushnir, A., Turner, J. R., Li, H. L., Chou, P. M., Weber, C. R., & De Plaen, I. G. (2013). Bifidobacteria Stabilize Claudins at Tight Junctions and Prevent Intestinal Barrier Dysfunction in Mouse Necrotizing Enterocolitis. *The American Journal of Pathology*, *182*(5), 1595. https://doi.org/10.1016/J.AJPATH.2013.01.013

Berrington, J. E., Barge, D., Fenton, A. C., Cant, A. J., & Spickett, G. P. (2005). Lymphocyte subsets in term and significantly preterm UK infants in the first year of life analysed by single platform flow cytometry. *Clinical and Experimental Immunology*, *140*(2), 289–292. https://doi.org/10.1111/J.1365-2249.2005.02767.X

Berrington, J., & Embleton, N. D. (2022). Discriminating necrotising enterocolitis and focal intestinal perforation. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *107*(3), 336–339. https://doi.org/10.1136/ARCHDISCHILD-2020-321429

Bertelli, C., Pillonel, T., Torregrossa, A., Prod'hom, G., Julie Fischer, C., Greub, G., & Giannoni, E. (2015). Bifidobacterium longum Bacteremia in Preterm Infants Receiving Probiotics. *Clinical Infectious Diseases*, *60*(6), 924–927. https://doi.org/10.1093/CID/CIU946

Binchy, Á., Moore, Z., & Patton, D. (2018). Feeding Intervals in Premature Infants ≤1750 g: An Integrative Review. *Advances in Neonatal Care*, *18*(3), 168–178. https://doi.org/10.1097/ANC.0000000000000486

Blutt, S. E., Crawford, S. E., Ramani, S., Zou, W. Y., & Estes, M. K. (2018). Engineered Human Gastrointestinal Cultures to Study the Microbiome and Infectious Diseases. In *CMGH* (Vol. 5, Issue 3, pp. 241–251). Elsevier Inc. https://doi.org/10.1016/j.jcmgh.2017.12.001

Bogaert, D., van Beveren, G. J., de Koff, E. M., Lusarreta Parga, P., Balcazar Lopez, C. E., Koppensteiner, L., Clerc, M., Hasrat, R., Arp, K., Chu, M. L. J. N., de Groot, P. C. M., Sanders, E. A. M., van Houten, M. A., & de Steenhuijsen Piters, W. A. A. (2023). Mother-to-infant microbiota transmission and infant microbiota

development across multiple body sites. *Cell Host & Microbe*, *31*(3), 447-460.e6. https://doi.org/10.1016/J.CHOM.2023.01.018

Bokulich, N. A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., Lieber, A. D., Wu, F., Perez-Perez, G. I., Chen, Y., Schweizer, W., Zheng, X., Contreras, M., Dominguez-Bello, M. G., & Blaser, M. J. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Science Translational Medicine*, *8*(343). https://doi.org/10.1126/scitranslmed.aad7121

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. In *Trends in Ecology and Evolution* (Vol. 24, Issue 3, pp. 127–135). Elsevier Current Trends. https://doi.org/10.1016/j.tree.2008.10.008

Bragg, L., & Tyson, G. W. (2014). Metagenomics using next-generation sequencing. *Methods in Molecular Biology*, *1096*, 183–201. https://doi.org/10.1007/978-1-62703-712-9_15

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, *9*(2), 378–400. https://doi.org/10.32614/rj-2017-066

Cao, G., Liu, J., & Liu, M. (2022). Global, Regional, and National Incidence and Mortality of Neonatal Preterm Birth, 1990-2019. *JAMA Pediatrics*, *176*(8), 787–796. https://doi.org/10.1001/JAMAPEDIATRICS.2022.1622

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal Advance Online Publication*. https://doi.org/10.1038/ismej.2012.8

Casaburi, G., Duar, R. M., Brown, H., Mitchell, R. D., Kazi, S., Chew, S., Cagney, O., Flannery, R. L., Sylvester, K. G., Frese, S. A., Henrick, B. M., & Freeman, S. L. (2021). Metagenomic insights of the infant microbiome community structure and function across multiple sites in the United States. *Scientific Reports*, *11*(1), 1472. https://doi.org/10.1038/s41598-020-80583-9

Casaburi, G., Wei, J., Kazi, S., Liu, J., Wang, K., Tao, G. Z., Lin, P. Y., Dunn, J. C. Y., Henrick, B. M., Frese, S. A., & Sylvester, K. G. (2022). Metabolic model of necrotizing enterocolitis in the premature newborn gut resulting from enteric dysbiosis. *Frontiers in Pediatrics*, *10*, 893059. https://doi.org/10.3389/FPED.2022.893059/BIBTEX

Chapman, J. A., & Stewart, C. J. (2023). Methodological challenges in neonatal microbiome research. *Https://Doi.Org/10.1080/19490976.2023.2183687*, *15*(1), 2183687. https://doi.org/10.1080/19490976.2023.2183687

Chen, E. Z., & Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, *32*(17), 2611–2617. https://doi.org/10.1093/bioinformatics/btw308

Chernikova, D. A., Madan, J. C., Housman, M. L., Zain-ul-abideen, M., Lundgren, S. N., Morrison, H. G., Sogin, M. L., Williams, S. M., Moore, J. H., Karagas, M. R., & Hoen, A. G. (2018). The premature infant gut microbiome during the first 6 weeks of life differs based on gestational maturity at birth. *Pediatric Research*, *84*(1), 71–79. https://doi.org/10.1038/s41390-018-0022-z

Co, J. Y., Margalef-Català, M., Li, X., Mah, A. T., Kuo, C. J., Monack, D. M., & Amieva Correspondence, M. R. (2019). Controlling Epithelial Polarity: A Human Enteroid Model for Host-Pathogen Interactions. *Cell Reports*, *26*. https://doi.org/10.1016/j.celrep.2019.01.108

Coleman, S., Unterhauser, K., Rezaul, K., Ledala, N., Lesmes, S., Caimano, M. J., Zhou, Y., Jackson, E., Gratalo, D., Driscoll, M. D., & Matson, A. P. (2023). High-resolution microbiome analysis reveals exclusionary Klebsiella species competition in preterm infants at risk for necrotizing enterocolitis. *Scientific Reports 2023 13:1*, *13*(1), 1–11. https://doi.org/10.1038/s41598-023-34735-2

Collado, M. C., Rautava, S., Aakko, J., Isolauri, E., & Salminen, S. (2016). Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Scientific Reports*, *6*(1), 1–13. https://doi.org/10.1038/srep23129

Cong, X., Xu, W., Janton, S., Henderson, W. A., Matson, A., McGrath, J. M., Maas, K., & Graf, J. (2016). Gut Microbiome Developmental Patterns in Early Life of Preterm Infants: Impacts of Feeding and Gender. *PLOS ONE*, *11*(4), e0152751. https://doi.org/10.1371/journal.pone.0152751

Contrepois, K., Jiang, L., & Snyder, M. (2015). Optimized analytical procedures for the untargeted metabolomic profiling of human urine and plasma by combining hydrophilic interaction (HILIC) and reverse-phase liquid chromatography (RPLC)-mass spectrometry. *Molecular and Cellular Proteomics*, *14*(6), 1684–1695. https://doi.org/10.1074/mcp.M114.046508

Coppa, G. V., Facinelli, B., Magi, G., Marini, E., Zampini, L., Mantovani, V., Galeazzi, T., Padella, L., Marchesiello, R. L., Santoro, L., Coscia, A., Peila, C., Volpi, N., & Gabrielli, O. (2016). Human milk glycosaminoglycans inhibit in vitro the adhesion of Escherichia coli and Salmonella fyris to human intestinal cells. *Pediatric Research*, *79*(4), 603–607. https://doi.org/10.1038/PR.2015.262

Costeloe, K., Bowler, U., Brocklehurst, P., Hardy, P., Heal, P., Juszczak, E., King, A., Panton, N., Stacey, F., Whiley, A., Wilks, M., & Millar, M. R. (2016). A randomised controlled trial of the probiotic Bifidobacterium breve BBG-001 in preterm babies to prevent sepsis, necrotising enterocolitis and death: the Probiotics in Preterm infantS (PiPS) trial. *Health Technology Assessment (Winchester, England)*, *20*(66), vii–83. https://doi.org/10.3310/HTA20660

Costeloe, K., Hardy, P., Juszczak, E., Wilks, M., & Millar, M. R. (2016). Bifidobacterium breve BBG-001 in very preterm infants: A randomised controlled phase 3 trial. *The Lancet*, *387*(10019), 649–660. https://doi.org/10.1016/S0140-6736(15)01027-2

D'Adamo, G. L., Widdop, J. T., & Giles, E. M. (2021). The future is now? Clinical and translational aspects of "Omics" technologies. *Immunology and Cell Biology*, *99*(2), 168–176. https://doi.org/10.1111/IMCB.12404

Dani, C., Coviello, C., Corsini, I., Arena, F., Antonelli, A., & Rossolini, G. M. (2015). Lactobacillus Sepsis and Probiotic Therapy in Newborns: Two New Cases and Literature Review. *AJP Reports*, *6*(1), e25–e29. https://doi.org/10.1055/S-0035-1566312/ID/JR150044-24

Davis, C. D. (2016). The gut microbiome and its role in obesity. *Nutrition Today*, *51*(4), 167–174. https://doi.org/10.1097/NT.0000000000000167

Diener, C., Dai, C. L., Wilmanski, T., Baloni, P., Smith, B., Rappaport, N., Hood, L., Magis, A. T., & Gibbons, S. M. (2022). Genome–microbiome interplay provides insight into the determinants of the human blood metabolome. *Nature*

Metabolism 2022 4:11, *4*(11), 1560–1572. https://doi.org/10.1038/s42255-022-00670-1

Dierikx, T. H., Deianova, N., Groen, J., Vijlbrief, D. C., Hulzebos, C., de Boode, W. P., d'Haens, E. J., Cossey, V., Kramer, B. W., van Weissenbruch, M. M., de Jonge, W. J., Benninga, M. A., van den Akker, C. H., van Kaam, A. H., de Boer, N. K. H., Visser, D. H., Niemarkt, H. J., & de Meij, T. G. J. (2022). Association between duration of early empiric antibiotics and necrotizing enterocolitis and late-onset sepsis in preterm infants: a multicenter cohort study. *European Journal of Pediatrics*, *181*(10), 3715–3724. https://doi.org/10.1007/S00431-022-04579-5/TABLES/4

Diks, A. M., Bonroy, C., Teodosio, C., Groenland, R. J., de Mooij, B., de Maertelaere, E., Neirynck, J., Philippé, J., Orfao, A., van Dongen, J. J. M., & Berkowska, M. A. (2019). Impact of blood storage and sample handling on quality of high dimensional flow cytometric data in multicenter clinical research. *Journal of Immunological Methods*, *475*, 112616. https://doi.org/10.1016/J.JIM.2019.06.007

Dolatshahi, S., Butler, A. L., Pou, C., Henckel, E., Bernhardsson, A. K., Gustafsson, A., Bohlin, K., Shin, S. A., Lauffenburger, D. A., Brodin, P., & Alter, G. (2022). Selective transfer of maternal antibodies in preterm and fullterm children. *Scientific Reports 2022 12:1*, *12*(1), 1–16. https://doi.org/10.1038/s41598-022-18973-4

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Hoashi, M., Rivera-Vinas, J. I., Mendez, K., Knight, R., & Clemente, J. C. (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Medicine 2016 22:3*, *22*(3), 250–253. https://doi.org/10.1038/nm.4039

Dong, Y., Speer, C. P., & Glaser, K. (2018). Beyond sepsis: Staphylococcus epidermidis is an underestimated but significant contributor to neonatal morbidity. *Virulence*, *9*(1), 621. https://doi.org/10.1080/21505594.2017.1419117

Dotterud, C. K., Avershina, E., Sekelja, M., Simpson, M. R., Rudi, K., Storrø, O., Johnsen, R., & Eien, T. (2015). Does maternal perinatal probiotic supplementation alter the intestinal microbiota of mother and child? *Journal of Pediatric Gastroenterology and Nutrition*, *61*(2), 200–207. https://doi.org/10.1097/MPG.0000000000000781

Downey, L. C., Smith, P. B., & Benjamin, D. K. (2010). Risk Factors and Prevention of Late Onset Sepsis in Premature Infants. *Early Human Development*, *86*(Suppl 1), 7. https://doi.org/10.1016/J.EARLHUMDEV.2010.01.012

Ducarmon, Q. R., Zwittink, R. D., Hornung, B. V. H., van Schaik, W., Young, V. B., & Kuijper, E. J. (2019). Gut Microbiota and Colonization Resistance against Bacterial Enteric Infection. *Microbiology and Molecular Biology Reviews*, *83*(3). https://doi.org/10.1128/MMBR.00007-19/ASSET/8CF6E364-1C0B-404C-B843-9AA28A898434/ASSETS/GRAPHIC/MMBR.00007-19-F0003.JPEG

Ehrlich, A. M., Pacheco, A. R., Henrick, B. M., Taft, D., Xu, G., Huda, M. N., Mishchuk, D., Goodson, M. L., Slupsky, C., Barile, D., Lebrilla, C. B., Stephensen, C. B., Mills, D. A., & Raybould, H. E. (2020). Indole-3-lactic acid associated with Bifidobacterium-dominated microbiota significantly decreases inflammation in intestinal epithelial cells. *BMC Microbiology*, *20*(1), 1–13. https://doi.org/10.1186/S12866-020-02023-Y/FIGURES/6

Elgendy, M. M., Othman, H. F., Heis, F., Qattea, I., & Aly, H. (2021). Spontaneous intestinal perforation in premature infants: a national study. *Journal of*

Perinatology 2021 41:5, *41*(5), 1122–1128. https://doi.org/10.1038/s41372-021-00990-2

Embleton, N. D., Berrington, J. E., Dorling, J., Ewer, A. K., Juszczak, E., Kirby, J. A., Lamb, C. A., Lanyon, C. V., McGuire, W., Probert, C. S., Rushton, S. P., Shirley, M. D., Stewart, C. J., & Cummings, S. P. (2017). Mechanisms Affecting the Gut of Preterm Infants in Enteral Feeding Trials. *Frontiers in Nutrition*, *4*, 14. https://doi.org/10.3389/FNUT.2017.00014/BIBTEX

Embleton, N. D., Sproat, T., Uthaya, S., Young, G. R., Garg, S., Vasu, V., Masi, A. C., Beck, L., Modi, N., Stewart, C. J., & Berrington, J. E. (2023). Effect of an Exclusive Human Milk Diet on the Gut Microbiome in Preterm Infants: A Randomized Clinical Trial. *JAMA Network Open*, *6*(3), e231165–e231165. https://doi.org/10.1001/JAMANETWORKOPEN.2023.1165

Emwas, A. H., Roy, R., McKay, R. T., Tenori, L., Saccenti, E., Nagana Gowda, G. A., Raftery, D., Alahmari, F., Jaremko, L., Jaremko, M., & Wishart, D. S. (2019). NMR Spectroscopy for Metabolomics Research. *Metabolites 2019, Vol. 9, Page 123*, *9*(7), 123. https://doi.org/10.3390/METABO9070123

Esaiassen, E., Cavanagh, P., Hjerde, E., Simonsen, G. S., Støen, R., & Klingenberg, C. (2016). Bifidobacterium longum Subspecies infantis Bacteremia in 3 Extremely Preterm Infants Receiving Probiotics. *Emerging Infectious Diseases*, *22*(9), 1664. https://doi.org/10.3201/EID2209.160033

Esaiassen, E., Hjerde, E., Cavanagh, J. P., Pedersen, T., Andresen, J. H., Rettedal, S. I., Støen, R., Nakstad, B., Willassen, N. P., & Klingenberg, C. (2018). Effects of Probiotic Supplementation on the Gut Microbiota and Antibiotic Resistome Development in Preterm Infants. *Frontiers in Pediatrics*, *6*, 347. https://doi.org/10.3389/fped.2018.00347

Esmaeilizand, R., Shah, P. S., Seshia, M., Yee, W., Yoon, E. W., Dow, K., Investigators, C. N. N., Shah, P. S., Harrison, A., Synnes, A., Sokoran, T., Yee, W., Aziz, K., Kalapesi, Z., Sankaran, K., Seshia, M., Alvaro, R., Shivananda, S., Da Silva, O., … Lee, S. K. (2018). Antibiotic exposure and development of necrotizing enterocolitis in very preterm neonates. *Paediatrics & Child Health*, *23*(4), e56–e61. https://doi.org/10.1093/PCH/PXX169

Ettayebi, K., Crawford, S. E., Murakami, K., Broughman, J. R., Karandikar, U., Tenge, V. R., Neill, F. H., Blutt, S. E., Zeng, X. L., Qu, L., Kou, B., Opekun, A. R., Burrin, D., Graham, D. Y., Ramani, S., Atmar, R. L., & Estes, M. K. (2016). Replication of human noroviruses in stem cell-derived human enteroids. *Science (New York, N.Y.)*, *353*(6306), 1387–1393. https://doi.org/10.1126/SCIENCE.AAF5211

Farrugia, M. K., Morgan, A. S., McHugh, K., & Kiely, E. M. (2003). Neonatal gastrointestinal perforation. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *88*(1), F75–F75. https://doi.org/10.1136/FN.88.1.F75

Fisher, J. G., Jones, B. A., Gutierrez, I. M., Hull, M. A., Kang, K. H., Kenny, M., Zurakowski, D., Modi, B. P., Horbar, J. D., & Jaksic, T. (2014). Mortality associated with laparotomy-confirmed neonatal spontaneous intestinal perforation: A prospective 5-year multicenter analysis. *Journal of Pediatric Surgery*, *49*(8), 1215–1219. https://doi.org/10.1016/J.JPEDSURG.2013.11.051

Flannery, D. D., Ross, R. K., Mukhopadhyay, S., Tribble, A. C., Puopolo, K. M., & Gerber, J. S. (2018). Temporal Trends and Center Variation in Early Antibiotic Use Among Premature Infants. *JAMA Network Open*, *1*(1), e180164–e180164. https://doi.org/10.1001/JAMANETWORKOPEN.2018.0164

Fofanova, T. Y., Stewart, C. J., Auchtung, J. M., Wilson, R. L., Britton, R. A., Grande-Allen, K. J., Estes, M. K., & Petrosino, J. F. (2019). A novel human enteroid-anaerobe co-culture system to study microbial-host interaction under physiological hypoxia. *BioRxiv*, 555755. https://doi.org/10.1101/555755

Foulke-Abel, J., In, J., Kovbasnjuk, O., Zachos, N. C., Ettayebi, K., Blutt, S. E., Hyser, J. M., Zeng, X. L., Crawford, S. E., Broughman, J. R., Estes, M. K., & Donowitz, M. (2014). Human enteroids as an ex-vivo model of host–pathogen interactions in the gastrointestinal tract. *Experimental Biology and Medicine*, *239*(9), 1124–1134. https://doi.org/10.1177/1535370214529398

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (Third). Sage.

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, *15*(11), 962–968. https://doi.org/10.1038/s41592-018-0176-y

Gail, M. H., Wan, Y., & Shi, J. (2021). Power of Microbiome Beta-Diversity Analyses Based on Standard Reference Samples. *American Journal of Epidemiology*, *190*(3), 439–447. https://doi.org/10.1093/aje/kwaa204

Garrido, D., Ruiz-Moyano, S., Lemay, D. G., Sela, D. A., German, J. B., & Mills, D. A. (2015). Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Scientific Reports*, *5*(1), 1–18. https://doi.org/10.1038/srep13517

Gasparrini, A. J., Wang, B., Sun, X., Kennedy, E. A., Hernandez-Leyva, A., Ndao, I. M., Tarr, P. I., Warner, B. B., & Dantas, G. (2019). Persistent metagenomic signatures of early-life hospitalization and antibiotic treatment in the infant gut microbiota and resistome. *Nature Microbiology 2019 4:12*, *4*(12), 2285–2297. https://doi.org/10.1038/s41564-019-0550-2

Giangregorio, N., Tonazzi, A., Console, L., Lorusso, I., De Palma, A., & Indiveri, C. (2016). The mitochondrial carnitine/acylcarnitine carrier is regulated by hydrogen sulfide via interaction with C136 and C155. *Biochimica et Biophysica Acta*, *1860*(1 Pt A), 20–27. https://doi.org/10.1016/J.BBAGEN.2015.10.005

Gibson, M. K., Wang, B., Ahmadi, S., Burnham, C. A. D., Tarr, P. I., Warner, B. B., & Dantas, G. (2016). Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nature Microbiology*, *1*, 16024. https://doi.org/10.1038/nmicrobiol.2016.24

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine*, *24*(4), 392. https://doi.org/10.1038/NM.4517

Glaser, K., Kern, D., Speer, C. P., Schlegel, N., Schwab, M., Thome, U. H., Härtel, C., & Wright, C. J. (2023). Imbalanced Inflammatory Responses in Preterm and Term Cord Blood Monocytes and Expansion of the CD14+CD16+ Subset upon Toll-like Receptor Stimulation. *International Journal of Molecular Sciences*, *24*(5), 4919. https://doi.org/10.3390/IJMS24054919/S1

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. In *Frontiers in Microbiology* (Vol. 8, Issue NOV, p. 2224). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2017.02224

Goldberg, E. L., Molony, R. D., Kudo, E., Sidorov, S., Kong, Y., Dixit, V. D., & Iwasaki, A. (2019). Ketogenic diet activates protective γδ T cell responses

against influenza virus infection. *Science Immunology*, *4*(41). https://doi.org/10.1126/SCIIMMUNOL.AAV2026/SUPPL_FILE/AAV2026_SM.PDF

Gopalakrishna, K. P., Macadangdang, B. R., Rogers, M. B., Tometich, J. T., Firek, B. A., Baker, R., Ji, J., Burr, A. H. P., Ma, C., Good, M., Morowitz, M. J., & Hand, T. W. (2019). Maternal IgA protects against the development of necrotizing enterocolitis in preterm infants. *Nature Medicine*, *25*(7), 1110–1115. https://doi.org/10.1038/s41591-019-0480-9

Granger, C., Dermyshi, E., Roberts, E., Beck, L. C., Embleton, N., & Berrington, J. (2022). Necrotising enterocolitis, late-onset sepsis and mortality after routine probiotic introduction in the UK. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *107*(4), 352–358. https://doi.org/10.1136/ARCHDISCHILD-2021-322252

Grech, A., Collins, C. E., Holmes, A., Lal, R., Duncanson, K., Taylor, R., & Gordon, A. (2021). Maternal exposures and the infant gut microbiome: a systematic review with meta-analysis. *Gut Microbes*, *13*(1), 1–30. https://doi.org/10.1080/19490976.2021.1897210

Gregory, K. E., Samuel, B. S., Houghteling, P., Shan, G., Ausubel, F. M., Sadreyev, R. I., & Walker, W. A. (2016). Influence of maternal breast milk ingestion on acquisition of the intestinal microbiome in preterm infants. *Microbiome*, *4*(1), 68. https://doi.org/10.1186/s40168-016-0214-x

Grier, A., Qiu, X., Bandyopadhyay, S., Holden-Wiltse, J., Kessler, H. A., Gill, A. L., Hamilton, B., Huyck, H., Misra, S., Mariani, T. J., Ryan, R. M., Scholer, L., Scheible, K. M., Lee, Y. H., Caserta, M. T., Pryhuber, G. S., & Gill, S. R. (2017). Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. *Microbiome*, *5*(1), 158. https://doi.org/10.1186/s40168-017-0377-0

Guo, C., Zhou, Q., Li, M., Zhou, L., Xu, L., Zhang, Y., Li, D., Wang, Y., Dai, W., Li, S., & Zhang, L. (2020). Breastfeeding restored the gut microbiota in caesarean section infants and lowered the infection risk in early life. *BMC Pediatrics*, *20*(1), 1–6. https://doi.org/10.1186/S12887-020-02433-X/TABLES/1

Gurung, M., Li, Z., You, H., Rodrigues, R., Jump, D. B., Morgun, A., & Shulzhenko, N. (2020). Role of gut microbiota in type 2 diabetes pathophysiology. In *EBioMedicine* (Vol. 51). Elsevier B.V. https://doi.org/10.1016/j.ebiom.2019.11.051

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S. J., & Ralser, M. (2017). Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. *Current Opinion in Systems Biology*, *6*, 37–45. https://doi.org/10.1016/J.COISB.2017.08.009

Hackam, D. J. (2022). Anemia, blood transfusions, and necrotizing enterocolitis in premature infants. *Pediatric Research 2022 91:6*, *91*(6), 1317–1319. https://doi.org/10.1038/s41390-022-02007-w

Hackam, D. J., & Sodhi, C. P. (2018). Toll-Like Receptor–Mediated Intestinal Inflammatory Imbalance in the Pathogenesis of Necrotizing Enterocolitis. In *CMGH* (Vol. 6, Issue 2, pp. 229-238.e1). Elsevier Inc. https://doi.org/10.1016/j.jcmgh.2018.04.001

Hanna, M., & Mazkereth, R. (2021). Which patient in the neonatal intensive care unit should receive antifungal prophylaxis therapy? *Israel Medical Association Journal*, *23*(2), 116–120.

Hapfelmeier, S., Lawson, M. A. E., Slack, E., Kirundi, J. K., Stoel, M., Heikenwalder, M., Cahenzli, J., Velykoredko, Y., Balmer, M. L., Endt, K., Geuking, M. B., Curtiss, R., McCoy, K. D., & Macpherson, A. J. (2010). Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science*, *328*(5986), 1705–1709. https://doi.org/10.1126/science.1188454

Hartz, L. E., Bradshaw, W., Brandon, D. H., & Gregory, K. E. (2015). Potential NICU Environmental Influences on the Neonate's Microbiome: A Systematic Review. *Advances in Neonatal Care : Official Journal of the National Association of Neonatal Nurses*, *15*(5), 324. https://doi.org/10.1097/ANC.0000000000000220

He, Y., Du, W., Xiao, S., Zeng, B., She, X., Liu, D., Du, H., Li, L., Li, F., Ai, Q., He, J., Song, C., Wei, H., Zhao, X., & Yu, J. (2021). Colonization of fecal microbiota from patients with neonatal necrotizing enterocolitis exacerbates intestinal injury in germfree mice subjected to necrotizing enterocolitis-induction protocol via alterations in butyrate and regulatory T cells. *Journal of Translational Medicine*, *19*(1), 1–12. https://doi.org/10.1186/S12967-021-03109-5/FIGURES/6

Henrick, B. M., Rodriguez, L., German, J. B., Frese, S. A., Lakshmikanth, T., Pou, C., Henckel, E., Arzoomand, A., Olin, A., Wang, J., Mikes, J., Tan, Z., Chen, Y., Ehrlich, A. M., Bernhardsson, A. K., Mugabo, C. H., Ambrosiani, Y., Gustafsson, A., Chew, S., … Brodin, P. (2021). Bifidobacteria-mediated immune system imprinting early in life In brief Bifidobacteria-mediated immune system imprinting early in life. *Cell*, *184*, 3884-3898.e11. https://doi.org/10.1016/j.cell.2021.05.030

Henry, M. C. W., & Moss, R. L. (2008). *Necrotizing Enterocolitis*. https://doi.org/10.1146/annurev.med.60.050207.092824

Herrmann, K., & Carroll, K. (2014). An Exclusively Human Milk Diet Reduces Necrotizing Enterocolitis. *Breastfeeding Medicine*, *9*(4), 184. https://doi.org/10.1089/BFM.2013.0121

Hesla, H. M., Stenius, F., Jäderlund, L., Nelson, R., Engstrand, L., Alm, J., & Dicksved, J. (2014). Impact of lifestyle on the gut microbiota of healthy infants and their mothers - the ALADDIN birth cohort. *FEMS Microbiology Ecology*, *90*(3), 791–801. https://doi.org/10.1111/1574-6941.12434

Hickey, L., Garland, S. M., Jacobs, S. E., O'Donnell, C. P. F., & Tabrizi, S. N. (2014). Cross-colonization of infants with probiotic organisms in a neonatal unit. *Journal of Hospital Infection*, *88*(4), 226–229. https://doi.org/10.1016/j.jhin.2014.09.006

Hill, C. J., Lynch, D. B., Murphy, K., Ulaszewska, M., Jeffery, I. B., O'Shea, C. A., Watkins, C., Dempsey, E., Mattivi, F., Tuohy, K., Paul Ross, R., Anthony Ryan, C., O' Toole, P. W., & Stanton, C. (2017). Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome*, *5*(1), 1–18. https://doi.org/10.1186/s40168-016-0213-y

Holmes, I., Harris, K., & Quince, C. (2012). Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE*, *7*(2), 30126. https://doi.org/10.1371/journal.pone.0030126

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M. Y., Nakanishi, H., Ikeda, K., … Nishioka, T. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry : JMS*, *45*(7), 703–714. https://doi.org/10.1002/JMS.1777

Hufnagl, K., Pali-Schöll, I., Roth-Walter, F., & Jensen-Jarolim, E. (2020). Dysbiosis of the gut and lung microbiome has a role in asthma. In *Seminars in*

Immunopathology (Vol. 42, Issue 1, pp. 75–93). Springer. https://doi.org/10.1007/s00281-019-00775-y

Ivanov, I. I., Frutos, R. de L., Manel, N., Yoshinaga, K., Rifkin, D. B., Sartor, R. B., Finlay, B. B., & Littman, D. R. (2008). Specific Microbiota Direct the Differentiation of IL-17-Producing T-Helper Cells in the Mucosa of the Small Intestine. *Cell Host and Microbe*, *4*(4), 337–349. https://doi.org/10.1016/j.chom.2008.09.009

Jacobs, S. E., Tobin, J. M., Opie, G. F., Donath, S., Tabrizi, S. N., Pirotta, M., Morley, C. J., & Garland, S. M. (2013). Probiotic effects on late-onset sepsis in very preterm infants: A randomized controlled trial. In *Pediatrics* (Vol. 132, Issue 6, pp. 1055–1062). American Academy of Pediatrics. https://doi.org/10.1542/peds.2013-1339

Jankevics, A., LLoyd, G. R., & Weber, R. J. M. (2021). *pmp: Peak Matrix Processing and signal batch correction for metabolomics datasets. R package version 1.6.0*.

Jantscher-Krenn, E., Zherebtsov, M., Nissan, C., Goth, K., Guner, Y. S., Naidu, N., Choudhury, B., Grishin, A. V., Ford, H. R., & Bode, L. (2012). The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats. *Gut*, *61*(10), 1417–1425. https://doi.org/10.1136/gutjnl-2011-301404

Jenke, A., Ruf, E. M., Hoppe, T., Heldmann, M., & Wirth, S. (2012). Bifidobacterium septicaemia in an extremely low-birthweight infant under probiotic therapy. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *97*(3), F217–F218. https://doi.org/10.1136/ARCHDISCHILD-2011-300838

Jia, J., Xun, P., Wang, X., He, K., Tang, Q., Zhang, T., Wang, Y., Tang, W., Lu, L., Yan, W., Wang, W., Hu, T., & Cai, W. (2020). Impact of Postnatal Antibiotics and Parenteral Nutrition on the Gut Microbiota in Preterm Infants During Early Life. *Journal of Parenteral and Enteral Nutrition*, *44*(4), 639–654. https://doi.org/10.1002/JPEN.1695

Jones, I. H., & Hall, N. J. (2020). Contemporary Outcomes for Infants with Necrotizing Enterocolitis—A Systematic Review. *Journal of Pediatrics*, *220*, 86-92.e3. https://doi.org/10.1016/j.jpeds.2019.11.011

Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. *Nature Reviews Genetics 2018 19:5*, *19*(5), 299–310. https://doi.org/10.1038/nrg.2018.4

Kennedy, K. M., De Goffau, M. C., Elisa Perez-Muñoz, M., Arrieta, M.-C., Bäckhed, F., Braun, T., Bushman, F. D., Dore, J., De Vos, W. M., Earl, A. M., Eisen, J. A., Elovitz, M. A., Ganal, S. C., Konnikova, L., Lebeer, S., Macpherson, A. J., Massey, R. C., Mchardy, A. C., Koren, O., … O'mahony, L. (2023). Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies Check for updates. *Curtis Huttenhower*, *613*, 61. https://doi.org/10.1038/s41586-022-05546-8

Kennedy, K. M., Gerlach, M. J., Adam, T., Heimesaat, M. M., Rossi, L., Surette, M. G., Sloboda, D. M., & Braun, T. (2021). Fetal meconium does not have a detectable microbiota before birth. *Nature Microbiology*, 1–9. https://doi.org/10.1038/s41564-021-00904-0

Khader, S. A., Gaffen, S. L., & Kolls, J. K. (2009). Th17 cells at the crossroads of innate and adaptive immunity against infectious diseases at the mucosa. *Mucosal Immunology*, *2*(5), 403–411. https://doi.org/10.1038/MI.2009.100

Khailova, L., Dvorak, K., Arganbright, K. M., Halpern, M. D., Kinouchi, T., Yajima, M., & Dvorak, B. (2009). Bifidobacterium bifidum improves intestinal integrity in a rat

model of necrotizing enterocolitis. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *297*(5), 940–949. https://doi.org/10.1152/AJPGI.00141.2009/ASSET/IMAGES/LARGE/ZH30110955130007.JPEG

Kiu, R., Shaw, A. G., Sim, K., Acuna-Gonzalez, A., Price, C. A., Bedwell, H., Dreger, S. A., Fowler, W. J., Cornwell, E., Pickard, D., Belteki, G., Malsom, J., Phillips, S., Young, G. R., Schofield, Z., Alcon-Giner, C., Berrington, J. E., Stewart, C. J., Dougan, G., … Hall, L. J. (2023). Particular genomic and virulence traits associated with preterm infant-derived toxigenic Clostridium perfringens strains. *Nature Microbiology 2023 8:6*, *8*(6), 1160–1175. https://doi.org/10.1038/s41564-023-01385-z

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L. I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology 2018 16:7*, *16*(7), 410–422. https://doi.org/10.1038/s41579-018-0029-9

Kondělková, K., Vokurková, D., Krejsek, J., Borská, L., Fiala, Z., & Ctirad, A. (2010). Regulatory T cells (TREG) and their roles in immune system with respect to immunopathological disorders. *Acta Medica (Hradec Kralove)*, *53*(2), 73–77. https://doi.org/10.14712/18059694.2016.63

Korpela, K., Blakstad, E. W., Moltu, S. J., Strømmen, K., Nakstad, B., Rønnestad, A. E., Brække, K., Iversen, P. O., Drevon, C. A., & de Vos, W. (2018). Intestinal microbiota development and gestational age in preterm neonates. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-20827-x

Korpela, K., Helve, O., Kolho, K. L., Saisto, T., Skogberg, K., Dikareva, E., Stefanovic, V., Salonen, A., Andersson, S., & de Vos, W. M. (2020). Maternal Fecal Microbiota Transplantation in Cesarean-Born Infants Rapidly Restores Normal Gut Microbial Development: A Proof-of-Concept Study. *Cell*, *183*(2), 324-334.e5. https://doi.org/10.1016/J.CELL.2020.08.047

Korpela, K., Salonen, A., Saxen, H., Nikkonen, A., Peltola, V., Jaakkola, T., de Vos, W., & Kolho, K. L. (2020). Antibiotics in early life associate with specific gut microbiota signatures in a prospective longitudinal infant cohort. *Pediatric Research 2020 88:3*, *88*(3), 438–443. https://doi.org/10.1038/s41390-020-0761-5

Krishnan, P., & Lotfollahzadeh, S. (2023). Spontaneous Intestinal Perforation Of The Newborn. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK585031/

Kwon, J. E., Lim, J., Kim, I., Kim, D., & Kang, S. C. (2018). Isolation and identification of new bacterial stains producing equol from Pueraria lobata extract fermentation. *PLOS ONE*, *13*(2), e0192490. https://doi.org/10.1371/JOURNAL.PONE.0192490

La Rosa, P. S., Warner, B. B., Zhou, Y., Weinstock, G. M., Sodergren, E., Hall-Moore, C. M., Stevens, H. J., Bennett, W. E., Shaikh, N., Linneman, L. A., Hoffmann, J. A., Hamvas, A., Deych, E., Shands, B. A., Shannon, W. D., & Tarr, P. I. (2014). Patterned progression of bacterial populations in the premature infant gut. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(34), 12522–12527. https://doi.org/10.1073/pnas.1409497111

Lauder, A. P., Roche, A. M., Sherrill-Mix, S., Bailey, A., Laughlin, A. L., Bittinger, K., Leite, R., Elovitz, M. A., Parry, S., & Bushman, F. D. (2016). Comparison of

placenta samples with contamination controls does not provide evidence for a distinct placenta microbiota. *Microbiome*, *4*(1), 1–11. https://doi.org/10.1186/s40168-016-0172-3

Lawson, M. A. E., O'Neill, I. J., Kujawska, M., Gowrinadh Javvadi, S., Wijeyesekera, A., Flegg, Z., Chalklen, L., & Hall, L. J. (2020). Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. *ISME Journal*, *14*(2), 635–648. https://doi.org/10.1038/s41396-019-0553-2

Lemme-Dumit, J. M., Song, Y., Lwin, H. W., Hernandez-Chavez, C., Sundararajan, S., Viscardi, R. M., Ravel, J., Pasetti, M. F., & Ma, B. (2022). Altered Gut Microbiome and Fecal Immune Phenotype in Early Preterm Infants With Leaky Gut. *Frontiers in Immunology*, *13*, 815046. https://doi.org/10.3389/FIMMU.2022.815046/BIBTEX

Li, N., van Unen, V., Abdelaal, T., Guo, N., Kasatskaya, S. A., Ladell, K., McLaren, J. E., Egorov, E. S., Izraelson, M., Chuva de Sousa Lopes, S. M., Höllt, T., Britanova, O. V., Eggermont, J., de Miranda, N. F. C. C., Chudakov, D. M., Price, D. A., Lelieveldt, B. P. F., & Koning, F. (2019). Memory CD4+ T cells are generated in the human fetal intestine. *Nature Immunology 2019 20:3*, *20*(3), 301–312. https://doi.org/10.1038/s41590-018-0294-9

Liebe, H., Schlegel, C., Cai, X., Golubkova, A., Loerke, C., Leiva, T., & Hunter, C. J. (2023). Apical-Out Enteroids as an Innovative Model for Necrotizing Enterocolitis. *Journal of Surgical Research*, *283*, 1106–1116. https://doi.org/10.1016/J.JSS.2022.11.048

Liu, T., Zong, H., Chen, X., Li, S., Liu, Z., Cui, X., Jia, G., & Shi, Y. (2021). Toll-like receptor 4-mediated necroptosis in the development of necrotizing enterocolitis. *Pediatric Research 2021 91:1*, *91*(1), 73–82. https://doi.org/10.1038/s41390-021-01457-y

Lopez, C. M., Sampah, M. E. S., Duess, J. W., Ishiyama, A., Ahmad, R., Sodhi, C. P., & Hackam, D. J. (2023). Models of necrotizing enterocolitis. *Seminars in Perinatology*, *47*(1), 151695. https://doi.org/10.1016/J.SEMPERI.2022.151695

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Lu, P., Yamaguchi, Y., Fulton, W. B., Wang, S., Zhou, Q., Jia, H., Kovler, M. L., Salazar, A. G., Sampah, M., Prindle, T., Wipf, P., Sodhi, C. P., & Hackam, D. J. (2021). Maternal aryl hydrocarbon receptor activation protects newborns against necrotizing enterocolitis. *Nature Communications*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-21356-4

Ma, J., Li, Z., Zhang, W., Zhang, C., Zhang, Y., Mei, H., Zhuo, N., Wang, H., Wang, L., & Wu, D. (2020). Comparison of gut microbiota in exclusively breast-fed and formula-fed babies: a study of 91 term infants. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-72635-x

Maccari, F., Mantovani, V., Gabrielli, O., Carlucci, A., Zampini, L., Galeazzi, T., Galeotti, F., Coppa, G. V., & Volpi, N. (2016). Metabolic fate of milk glycosaminoglycans in breastfed and formula fed newborns. *Glycoconjugate Journal*, *33*(2), 181–188. https://doi.org/10.1007/S10719-016-9655-5

Maldonado-Gómez, M. X., Martínez, I., Bottacini, F., O'Callaghan, A., Ventura, M., van Sinderen, D., Hillmann, B., Vangay, P., Knights, D., Hutkins, R. W., & Walter, J. (2016). Stable Engraftment of Bifidobacterium longum AH1206 in the Human Gut Depends on Individualized Features of the Resident Microbiome.

Cell Host & Microbe, *20*(4), 515–526. https://doi.org/10.1016/J.CHOM.2016.09.001

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Paulson, J. N., Franzosa, E. A., Corrada Bravo, H., & Huttenhower, C. (2021). Multivariable Association Discovery in Population-scale Meta-omics Studies 3. *BioRxiv*, 2021.01.20.427420. https://doi.org/10.1101/2021.01.20.427420

Mara, M. A., Good, M., & Weitkamp, J. H. (2018). Innate and adaptive immunity in necrotizing enterocolitis. *Seminars in Fetal & Neonatal Medicine*, *23*(6), 394. https://doi.org/10.1016/J.SINY.2018.08.002

Masi, A. C., Embleton, N. D., Lamb, C. A., Young, G., Granger, C. L., Najera, J., Smith, D. P., Hoffman, K. L., Petrosino, J. F., Bode, L., Berrington, J. E., & Stewart, C. J. (2020). Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. *Gut*. https://doi.org/10.1136/gutjnl-2020-322771

Masi, A. C., Fofanova, T. Y., Lamb, C. A., Auchtung, J. M., Britton, R. A., Estes, M. K., Ramani, S., Cockell, S. J., Coxhead, J., Embleton, N. D., Berrington, J. E., Petrosino, J. F., Stewart, C. J., & Andrea Masi, T. C. (2021). Distinct gene expression profiles between human preterm-derived and adult-derived intestinal organoids exposed to Enterococcus faecalis: a pilot study. *Gut*, *0*(0), gutjnl-2021-326552. https://doi.org/10.1136/GUTJNL-2021-326552

Masi, A. C., & Stewart, C. J. (2022). Untangling human milk oligosaccharides and infant gut microbiome. *IScience*, *25*(1). https://doi.org/10.1016/J.ISCI.2021.103542

McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., Bhullar, K., Canova, M. J., De Pascale, G., Ejim, L., Kalan, L., King, A. M., Koteva, K., Morar, M., Mulvey, M. R., O'Brien, J. S., Pawlowski, A. C., Piddock, L. J. V., Spanogiannopoulos, P., … Wright, G. D. (2013). The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3348–3357. https://doi.org/10.1128/AAC.00419-13

McDavid, A., Laniewski, N., Grier, A., Gill, A. L., Kessler, H. A., Huyck, H., Carbonell, E., Holden-Wiltse, J., Bandyopadhyay, S., Carnahan, J., Dylag, A. M., Topham, D. J., Falsey, A. R., Caserta, M. T., Pryhuber, G. S., Gill, S. R., & Scheible, K. M. (2022). Aberrant newborn T cell and microbiota developmental trajectories predict respiratory compromise during infancy. *IScience*, *25*(4). https://doi.org/10.1016/J.ISCI.2022.104007

McMurtry, V. E., Gupta, R. W., Tran, L., Blanchard, E. E., Penn, D., Taylor, C. M., & Ferris, M. J. (2015). Bacterial diversity and Clostridia abundance decrease with increasing severity of necrotizing enterocolitis. *Microbiome*, *3*(1), 1–8. https://doi.org/10.1186/S40168-015-0075-8/FIGURES/3

Mead, B. E., & Karp, J. M. (2019). All models are wrong, but some organoids may be useful. In *Genome Biology* (Vol. 20, Issue 1, pp. 1–3). BioMed Central Ltd. https://doi.org/10.1186/s13059-019-1677-4

Meinzen-Derr, J., Morrow, A. L., Hornung, R. W., Donovan, E. F., Dietrich, K. N., & Succop, P. A. (2009). Epidemiology of Necrotizing Enterocolitis Temporal Clustering in Two Neonatology Practices. *The Journal of Pediatrics*, *154*(5), 656-661.e2. https://doi.org/10.1016/J.JPEDS.2008.11.002

Melville, J. M., & Moss, T. J. M. (2013). The immune consequences of preterm birth. *Frontiers in Neuroscience*, *7*(7 MAY). https://doi.org/10.3389/FNINS.2013.00079

Mercer, E. M., & Arrieta, M. C. (2023). Probiotics to improve the gut microbiome in premature infants: are we there yet? *Gut Microbes*, *15*(1). https://doi.org/10.1080/19490976.2023.2201160

Mitchell, C. M., Mazzoni, C., Hogstrom, L., Bryant, A., Bergerat, A., Cher, A., Pochan, S., Herman, P., Carrigan, M., Sharp, K., Huttenhower, C., Lander, E. S., Vlamakis, H., Xavier, R. J., & Yassour, M. (2020). Delivery Mode Affects Stability of Early Infant Gut Microbiota. *Cell Reports Medicine*, *1*(9), 100156. https://doi.org/10.1016/j.xcrm.2020.100156

Mohamed, A., & Shah, P. S. (2012). Transfusion Associated Necrotizing Enterocolitis: A Meta-analysis of Observational Data. *Pediatrics*, *129*(3), 529–540. https://doi.org/10.1542/PEDS.2011-2872

Mollitt, D. L., Tepas, J. J., & Talbert, J. L. (1988). The role of coagulase-negative Staphylococcus in neonatal necrotizing enterocolitis. *Journal of Pediatric Surgery*, *23*(1), 60–63. https://doi.org/10.1016/S0022-3468(88)80542-6

Moossavi, S., Sepehri, S., Robertson, B., Bode, L., Goruk, S., Field, C. J., Lix, L. M., de Souza, R. J., Becker, A. B., Mandhane, P. J., Turvey, S. E., Subbarao, P., Moraes, T. J., Lefebvre, D. L., Sears, M. R., Khafipour, E., & Azad, M. B. (2019). Composition and Variation of the Human Milk Microbiota Are Influenced by Maternal and Early-Life Factors. *Cell Host and Microbe*, *25*(2), 324-335.e4. https://doi.org/10.1016/J.CHOM.2019.01.011

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., Henrissat, B., Knight, R., & Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, *332*(6032), 970–974. https://doi.org/10.1126/science.1198719

Neu, J. (2020). Necrotizing Enterocolitis: The Future. In *Neonatology* (Vol. 117, Issue 2, pp. 240–244). S. Karger AG. https://doi.org/10.1159/000506866

Neu, J., Modi, N., & Caplan, M. (2018). Necrotizing enterocolitis comes in different forms: Historical perspectives and defining the disease. In *Seminars in Fetal and Neonatal Medicine* (Vol. 23, Issue 6, pp. 370–373). W.B. Saunders Ltd. https://doi.org/10.1016/j.siny.2018.07.004

Nguyen, M., Holdbrooks, H., Mishra, P., Abrantes, M. A., Eskew, S., Garma, M., Oca, C.-G., McGuckin, C., Hein, C. B., Mitchell, R. D., Kazi, S., Chew, S., Casaburi, G., Brown, H. K., Frese, S. A., & Henrick, B. M. (2021). Impact of Probiotic B. infantis EVC001 Feeding in Premature Infants on the Gut Microbiome, Nosocomially Acquired Antibiotic Resistance, and Enteric Inflammation. *Frontiers in Pediatrics*, *9*, 618009. https://doi.org/10.3389/fped.2021.618009

Nothias, L. F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A. A., Alka, O., Allard, P. M., Barsch, A., Cachet, X., Caraballo-Rodriguez, A. M., Da Silva, R. R., Dang, T., … Dorrestein, P. C. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nature Methods 2020 17:9*, *17*(9), 905–908. https://doi.org/10.1038/s41592-020-0933-6

Nowicka, M., Krieg, C., Crowell, H. L., Weber, L. M., Hartmann, F. J., Guglietta, S., Becher, B., Levesque, M. P., & Robinson, M. D. (2017). CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Research*, *6*, 748. https://doi.org/10.12688/F1000RESEARCH.11622.3

Ohnmacht, C., Park, J. H., Cording, S., Wing, J. B., Atarashi, K., Obata, Y., Gaboriau-Routhiau, V., Marques, R., Dulauroy, S., Fedoseeva, M., Busslinger, M., Cerf-Bensussan, N., Boneca, I. G., Voehringer, D., Hase, K., Honda, K., Sakaguchi, S., & Eberl, G. (2015). The microbiota regulates type 2 immunity

through RORγt+ T cells. *Science*, *349*(6251), 989–993. https://doi.org/10.1126/SCIENCE.AAC4263/SUPPL_FILE/OHNMACHT-SM.PDF
Olaloye, O., Eke, C., Jolteus, A., & Konnikova, L. (2023). Single cell analysis via mass cytometry of spontaneous intestinal perforation reveals alterations in small intestinal innate and adaptive mucosal immunity. *Frontiers in Immunology*, *14*, 365. https://doi.org/10.3389/FIMMU.2023.995558/BIBTEX
Olaloye, O. O., Liu, P., Toothaker, J. M., McCourt, B. T., McCourt, C. C., Xiao, J., Prochaska, E., Shaffer, S., Werner, L., Gringauz, J., Good, M., Goldsmith, J. D., An, X., Wang, F., Snapper, S. B., Shouval, D., Chen, K., Tseng, G., & Konnikova, L. (2021). CD16+CD163+ monocytes traffic to sites of inflammation during necrotizing enterocolitis in premature infants. *The Journal of Experimental Medicine*, *218*(9). https://doi.org/10.1084/JEM.20200344
Olin, A., Henckel, E., Chen, Y., Lakshmikanth, T., Pou, C., Mikes, J., Gustafsson, A., Bernhardsson, A. K., Zhang, C., Bohlin, K., & Brodin, P. (2018). Stereotypic Immune System Development in Newborn Children. *Cell*, *174*(5), 1277-1292.e14. https://doi.org/10.1016/j.cell.2018.06.045
Olm, M. R., Bhattacharya, N., Crits-Christoph, A., Firek, B. A., Baker, R., Song, Y. S., Morowitz, M. J., & Banfield, J. F. (2019). Necrotizing enterocolitis is preceded by increased gut bacterial replication, Klebsiella, and fimbriae-encoding bacteria. *Science Advances*, *5*(12), 5727–5738. https://doi.org/10.1126/sciadv.aax5727
Palleja, A., Mikkelsen, K. H., Forslund, S. K., Kashani, A., Allin, K. H., Nielsen, T., Hansen, T. H., Liang, S., Feng, Q., Zhang, C., Pyl, P. T., Coelho, L. P., Yang, H., Wang, J., Typas, A., Nielsen, M. F., Nielsen, H. B., Bork, P., Wang, J., … Pedersen, O. (2018). Recovery of gut microbiota of healthy adults following antibiotic exposure. *Nature Microbiology 2018 3:11*, *3*(11), 1255–1265. https://doi.org/10.1038/s41564-018-0257-9
Pammi, M., Cope, J., Tarr, P. I., Warner, B. B., Morrow, A. L., Mai, V., Gregory, K. E., Simon Kroll, J., McMurtry, V., Ferris, M. J., Engstrand, L., Lilja, H. E., Hollister, E. B., Versalovic, J., & Neu, J. (2017). Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: A systematic review and meta-analysis. *Microbiome*, *5*(1). https://doi.org/10.1186/s40168-017-0248-8
Pammi, M., & Suresh, G. (2017). Enteral lactoferrin supplementation for prevention of sepsis and necrotizing enterocolitis in preterm infants. *The Cochrane Database of Systematic Reviews*, *2017*(6). https://doi.org/10.1002/14651858.CD007137.PUB5
Pang, Z., Chong, J., Zhou, G., De Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P. É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research*, *49*(W1), W388–W396. https://doi.org/10.1093/NAR/GKAB382
Pärnänen, K., Karkman, A., Hultman, J., Lyra, C., Bengtsson-Palme, J., Larsson, D. G. J., Rautava, S., Isolauri, E., Salminen, S., Kumar, H., Satokari, R., & Virta, M. (2018). Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nature Communications 2018 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-06393-w
Pärnänen, K. M. M., Hultman, J., Markkanen, M., Satokari, R., Rautava, S., Lamendella, R., Wright, J., McLimans, C. J., Kelleher, S. L., & Virta, M. P. (2022). Early-life formula feeding is associated with infant gut microbiota alterations and an increased antibiotic resistance load. *The American Journal of Clinical Nutrition*, *115*(2), 407. https://doi.org/10.1093/AJCN/NQAB353

Patel, R. M., Knezevic, A., Shenvi, N., Hinkes, M., Keene, S., Roback, J. D., Easley, K. A., & Josephson, C. D. (2016). Association of Red Blood Cell Transfusion, Anemia, and Necrotizing Enterocolitis in Very Low-Birth-Weight Infants. *JAMA*, *315*(9), 889–897. https://doi.org/10.1001/JAMA.2016.1204

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods 2017 14:4*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Paveglio, S., Ledala, N., Rezaul, K., Lin, Q., Zhou, Y., Provatas, A. A., Bennett, E., Lindberg, T., Caimano, M., & Matson, A. P. (2020). Cytotoxin-producing Klebsiella oxytoca in the preterm gut and its association with necrotizing enterocolitis. *Emerging Microbes & Infections*, *9*(1), 1321. https://doi.org/10.1080/22221751.2020.1773743

Penn, A. H., Altshuler, A. E., Small, J. W., Taylor, S. F., Dobkins, K. R., & Schmid-Schönbein, G. W. (2012). Digested formula but not digested fresh human milk causes death of intestinal cells in vitro: implications for necrotizing enterocolitis. *Pediatric Research 2012 72:6*, *72*(6), 560–567. https://doi.org/10.1038/pr.2012.125

Peterson, H., Kolberg, L., Raudvere, U., Kuzmin, I., & Vilo, J. (2020). gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research*, *9*. https://doi.org/10.12688/F1000RESEARCH.24956.2

Pilarczyk-Zurek, M., Majka, G., Skowron, B., Baranowska, A., Piwowar, M., & Strus, M. (2022). The Multi-Component Causes of Late Neonatal Sepsis—Can We Regulate Them? *Nutrients*, *14*(2). https://doi.org/10.3390/NU14020243

Qi, J., Gan, L., Fang, J., Zhang, J., Yu, X., Guo, H., Cai, D., Cui, H., Gou, L., Deng, J., Wang, Z., & Zuo, Z. (2022). Beta-Hydroxybutyrate: A Dual Function Molecular and Immunological Barrier Function Regulator. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/FIMMU.2022.805881

Qian, X. B., Chen, T., Xu, Y. P., Chen, L., Sun, F. X., Lu, M. P., & Liu, Y. X. (2020). A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. *Chinese Medical Journal*, *133*(15), 1844–1855. https://doi.org/10.1097/CM9.0000000000000871

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. In *Nature Biotechnology* (Vol. 35, Issue 9, pp. 833–844). Nature Publishing Group. https://doi.org/10.1038/nbt.3935

Quinello, C., Silveira-Lessa, A. L., Ceccon, M. E. J. R., Cianciarullo, M. A., Carneiro-Sampaio, M., & Palmeira, P. (2014). Phenotypic Differences in Leucocyte Populations among Healthy Preterm and Full-Term Newborns. *Scandinavian Journal of Immunology*, *80*(1), 57–70. https://doi.org/10.1111/SJI.12183

Rahkola, E. N., Rautava, S., Hiltunen, H., Ross, C., Lahti, L., & Isolauri, E. (2023). The preterm gut microbiota and administration routes of different probiotics: a randomized controlled trial. *Pediatric Research 2023*, 1–8. https://doi.org/10.1038/s41390-023-02560-y

Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R., & Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature 2021 591:7851*, *591*(7851), 633–638. https://doi.org/10.1038/s41586-021-03241-8

Reyman, M., van Houten, M. A., van Baarle, D., Bosch, A. A. T. M., Man, W. H., Chu, M. L. J. N., Arp, K., Watson, R. L., Sanders, E. A. M., Fuentes, S., & Bogaert, D.

(2019). Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-13014-7

Reyman, M., van Houten, M. A., Watson, R. L., Chu, M. L. J. N., Arp, K., de Waal, W. J., Schiering, I., Plötz, F. B., Willems, R. J. L., van Schaik, W., Sanders, E. A. M., & Bogaert, D. (2022). Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial. *Nature Communications 2022 13:1*, *13*(1), 1–12. https://doi.org/10.1038/s41467-022-28525-z

Robertson, C., Robertson, C., Savva, G. M., Clapuci, R., Jones, J., Maimouni, H., Brown, E., Minocha, A., Hall, L. J., Clarke, P., & Clarke, P. (2020). Incidence of necrotising enterocolitis before and after introducing routine prophylactic Lactobacillus and Bifidobacterium probiotics. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *105*(4), 380–386. https://doi.org/10.1136/ARCHDISCHILD-2019-317346

Rodríguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J., & Garza-Ramos, U. (2019). Klebsiella variicola: an emerging pathogen in humans. *Emerging Microbes & Infections*, *8*(1), 973. https://doi.org/10.1080/22221751.2019.1634981

Round, J. L., & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12204–12209. https://doi.org/10.1073/PNAS.0909122107/SUPPL_FILE/PNAS.200909122SI.PDF

Rozé, J. C., Ancel, P. Y., Marchand-Martin, L., Rousseau, C., Montassier, E., Monot, C., Le Roux, K., Butin, M., Resche-Rigon, M., Aires, J., Neu, J., Lepage, P., & Butel, M. J. (2020). Assessment of Neonatal Intensive Care Unit Practices and Preterm Newborn Gut Microbiota and 2-Year Neurodevelopmental Outcomes. *JAMA Network Open*, *3*(9), 2018119. https://doi.org/10.1001/jamanetworkopen.2020.18119

Ruiz-Moyano, S., Totten, S. M., Garrido, D. A., Smilowitz, J. T., Bruce German, J., Lebrilla, C. B., & Mills, D. A. (2013). Variation in consumption of human milk oligosaccharides by infant gut-associated strains of bifidobacterium breve. *Applied and Environmental Microbiology*, *79*(19), 6040–6049. https://doi.org/10.1128/AEM.01843-13

Rutayisire, E., Huang, K., Liu, Y., & Tao, F. (2016). The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: A systematic review. *BMC Gastroenterology*, *16*(1), 86. https://doi.org/10.1186/s12876-016-0498-0

Sato, T., Vries, R. G., Snippert, H. J., Van De Wetering, M., Barker, N., Stange, D. E., Van Es, J. H., Abo, A., Kujala, P., Peters, P. J., & Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature*, *459*(7244), 262–265. https://doi.org/10.1038/nature07935

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, *9*(8), 811–814. https://doi.org/10.1038/nmeth.2066

Seki, M., Gotoh, K., Nakamura, S., Akeda, Y., Yoshii, T., Miyaguchi, S., Inohara, H., Horii, T., Oishi, K., Iida, T., & Tomono, K. (2013). Fatal sepsis caused by an unusual Klebsiella species that was misidentified by an automated identification system. *Journal of Medical Microbiology*, *62*(PART5), 801–803. https://doi.org/10.1099/JMM.0.051334-0/CITE/REFWORKS

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. https://doi.org/10.1371/JOURNAL.PBIO.1002533

Shao, Y., Forster, S. C., Tsaliki, E., Vervier, K., Strang, A., Simpson, N., Kumar, N., Stares, M. D., Rodger, A., Brocklehurst, P., Field, N., & Lawley, T. D. (2019). Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature 2019 574:7776*, *574*(7776), 117–121. https://doi.org/10.1038/s41586-019-1560-1

Sharif, S., Meader, N., Oddie, S. J., Rojas-Reyes, M. X., & McGuire, W. (2020). Probiotics to prevent necrotising enterocolitis in very preterm or very low birth weight infants. *Cochrane Database of Systematic Reviews*, *2020*(10). https://doi.org/10.1002/14651858.CD005496.PUB5/EPDF/ABSTRACT

Shen, C. M., Lin, S. C., Niu, D. M., & Kou, Y. R. (2013). Development of monocyte Toll-like receptor 2 and Toll-like receptor 4 in preterm newborns during the first few months of life. *Pediatric Research 2013 73:5*, *73*(5), 685–691. https://doi.org/10.1038/pr.2013.36

Shen, R., Embleton, N., Lyng Forman, J., Gale, C., Griesen, G., Sangild, P. T., Uthaya, S., & Berrington, J. (2022). Early antibiotic use and incidence of necrotising enterocolitis in very preterm infants: a protocol for a UK based observational study using routinely recorded data. *BMJ Open*, *12*(11), e065934. https://doi.org/10.1136/BMJOPEN-2022-065934

Short, S. S., Papillon, S., Berel, D., Ford, H. R., Frykman, P. K., & Kawaguchi, A. (2014). Late onset of necrotizing enterocolitis in the full-term infant is associated with increased mortality: Results from a two-center analysis. *Journal of Pediatric Surgery*, *49*(6), 950. https://doi.org/10.1016/J.JPEDSURG.2014.01.028

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research 2016 4:1521*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Sordillo, J. E., Zhou, Y., McGeachie, M. J., Ziniti, J., Lange, N., Laranjo, N., Savage, J. R., Carey, V., O'Connor, G., Sandel, M., Strunk, R., Bacharier, L., Zeiger, R., Weiss, S. T., Weinstock, G., Gold, D. R., & Litonjua, A. A. (2017). Factors influencing the infant gut microbiome at age 3-6 months: Findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *Journal of Allergy and Clinical Immunology*, *139*(2), 482-491.e14. https://doi.org/10.1016/j.jaci.2016.08.045

Sproat, T., Payne, R. P., Embleton, N. D., Berrington, J., & Hambleton, S. (2020). T Cells in Preterm Infants and the Influence of Milk Diet. *Frontiers in Immunology*, *11*, 1035. https://doi.org/10.3389/FIMMU.2020.01035

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112–118. https://doi.org/10.1093/BIOINFORMATICS/BTR597

Stewart, C. J., Ajami, N. J., O'Brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., Ross, M. C., Lloyd, R. E., Doddapaneni, H. V., Metcalf, G. A., Muzny, D., Gibbs, R. A., Vatanen, T., Huttenhower, C., Xavier, R. J., Rewers, M., Hagopian, W., Toppari, J., Ziegler, A. G., … Petrosino, J. F. (2018). Temporal development

of the gut microbiome in early childhood from the TEDDY study. *Nature*, *562*(7728), 583–588. https://doi.org/10.1038/s41586-018-0617-x

Stewart, C. J., Embleton, N. D., Clements, E., Luna, P. N., Smith, D. P., Fofanova, T. Y., Nelson, A., Taylor, G., Orr, C. H., Petrosino, J. F., Berrington, J. E., & Cummings, S. P. (2017a). Cesarean or Vaginal Birth Does Not Impact the Longitudinal Development of the Gut Microbiome in a Cohort of Exclusively Preterm Infants. *Frontiers in Microbiology*, *8*(JUN), 1008. https://doi.org/10.3389/fmicb.2017.01008

Stewart, C. J., Embleton, N. D., Marrs, E. C. L., Smith, D. P., Fofanova, T., Nelson, A., Skeath, T., Perry, J. D., Petrosino, J. F., Berrington, J. E., & Cummings, S. P. (2017b). Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. *Microbiome*, *5*(1), 75. https://doi.org/10.1186/s40168-017-0295-1

Stewart, C. J., Embleton, N. D., Marrs, E. C. L., Smith, D. P., Nelson, A., Abdulkadir, B., Skeath, T., Petrosino, J. F., Perry, J. D., Berrington, J. E., & Cummings, S. P. (2016). Temporal bacterial and metabolic development of the preterm gut reveals specific signatures in health and disease. *Microbiome*, *4*(1), 67. https://doi.org/10.1186/s40168-016-0216-8

Stewart, C. J., Estes, M. K., & Ramani, S. (2020). Establishing Human Intestinal Enteroid/Organoid Lines from Preterm Infant and Adult Tissue. *Methods in Molecular Biology (Clifton, N.J.)*, *2121*, 185–198. https://doi.org/10.1007/978-1-0716-0338-3_16

Stewart, C. J., Fatemizadeh, R., Parsons, P., Lamb, C. A., Shady, D. A., Petrosino, J. F., & Hair, A. B. (2019). Using formalin fixed paraffin embedded tissue to characterize the preterm gut microbiota in necrotising enterocolitis and spontaneous isolated perforation using marginal and diseased tissue. *BMC Microbiology*, *19*(1), 52. https://doi.org/10.1186/s12866-019-1426-6

Stewart, C., Marrs, E., Magorrian, S., Nelson, A., Lanyon, C., Perry, J., Embleton, N., Cummings, S., & Berrington, J. (2012). The preterm gut microbiota: changes associated with necrotizing enterocolitis and infection. *Acta Paediatrica*, *101*(11), 1121–1127. https://doi.org/10.1111/j.1651-2227.2012.02801.x

Stoltz, T., Jones, A., Rogers, L., Sidhu, H., Patel, R., Shamsuddin, A., Lee, C., Graham, C., Costa, A., & Rosner, J. (2021). 51 Donor Milk in the NICU: A Community Pediatrics Perspective. *Paediatrics & Child Health*, *26*(Supplement_1), e36–e36. https://doi.org/10.1093/PCH/PXAB061.040

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M. A., Benezra, A., Destefano, J., Meier, M. F., Muegge, B. D., Barratt, M. J., VanArendonk, L. G., Zhang, Q., Province, M. A., Petri, W. A., Ahmed, T., & Gordon, J. I. (2014). Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature 2014 510:7505*, *510*(7505), 417–421. https://doi.org/10.1038/nature13421

Sullivan, S., Schanler, R. J., Kim, J. H., Patel, A. L., Trawöger, R., Kiechl-Kohlendorfer, U., Chan, G. M., Blanco, C. L., Abrams, S., Cotten, C. M., Laroia, N., Ehrenkranz, R. A., Dudell, G., Cristofalo, E. A., Meier, P., Lee, M. L., Rechtman, D. J., & Lucas, A. (2010). An Exclusively Human Milk-Based Diet Is Associated with a Lower Rate of Necrotizing Enterocolitis than a Diet of Human Milk and Bovine Milk-Based Products. *Journal of Pediatrics*, *156*(4). https://doi.org/10.1016/j.jpeds.2009.10.040

Suzuki, T. A., & Worobey, M. (2014). Geographical variation of human gut microbial composition. *Biology Letters*, *10*(2), 20131037. https://doi.org/10.1098/RSBL.2013.1037

Taft, D. H., Ambalavanan, N., Schibler, K. R., Yu, Z., Newburg, D. S., Deshmukh, H., Ward, D. V., & Morrow, A. L. (2015). Center variation in intestinal microbiota prior to late-onset sepsis in preterm infants. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0130604

Torrazza, R. M., Ukhanova, M., Wang, X., Sharma, R., Hudak, M. L., Neu, J., & Mai, V. (2013). Intestinal microbial ecology and environmental factors affecting necrotizing enterocolitis. *PLoS ONE*, *8*(12), 83304. https://doi.org/10.1371/journal.pone.0083304

Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., Okahashi, N., Yamada, Y., Tada, I., Bonini, P., Higashi, Y., Okazaki, Y., Zhou, Z., Zhu, Z. J., Koelmel, J., Cajka, T., Fiehn, O., Saito, K., Arita, M., & Arita, M. (2020). A lipidome atlas in MS-DIAL 4. *Nature Biotechnology 2020 38:10*, *38*(10), 1159–1163. https://doi.org/10.1038/s41587-020-0531-2

Turroni, F., Bottacini, F., Foroni, E., Mulder, I., Kim, J. H., Zomer, A., Sánchez, B., Bidossi, A., Ferrarini, A., Giubellini, V., Delledonne, M., Henrissat, B., Coutinho, P., Oggioni, M., Fitzgerald, G. F., Mills, D., Margolles, A., Kelly, D., Van Sinderen, D., & Ventura, M. (2010). Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(45), 19514–19519. https://doi.org/10.1073/pnas.1011100107

Ty, F., Cj, S., 2#, A., Jm, Rl, W., Ra, B., Kj, G.-A., Mk, E., & Jf, P. (n.d.). *A novel human enteroid-anaerobe co-culture system to study microbial-host interaction under physiological hypoxia*. https://doi.org/10.1101/555755

Underwood, M. A., Arriola, J., Gerber, C. W., Kaveti, A., Kalanetra, K. M., Kananurak, A., Bevins, C. L., Mills, D. A., & Dvorak, B. (2014). Bifidobacterium longum subsp. infantis in experimental necrotizing enterocolitis: alterations in inflammation, innate immune response, and the microbiota. *Pediatric Research 2014 76:4*, *76*(4), 326–333. https://doi.org/10.1038/pr.2014.102

van Best, N., Trepels-Kottek, S., Savelkoul, P., Orlikowsky, T., Hornef, M. W., & Penders, J. (2020). Influence of probiotic supplementation on the developing microbiota in human preterm neonates. *Gut Microbes*, *12*(1), 1–16. https://doi.org/10.1080/19490976.2020.1826747

van den Akker, C. H. P., van Goudoever, J. B., Shamir, R., Domellöf, M., Embleton, N. D., Hojsak, I., Lapillonne, A., Mihatsch, W. A., Berni Canani, R., Bronsky, J., Campoy, C., Fewtrell, M. S., Fidler Mis, N., Guarino, A., Hulst, J. M., Indrio, F., Kolaček, S., Orel, R., Vandenplas, Y., … Szajewska, H. (2020). Probiotics and Preterm Infants: A Position Paper by the European Society for Paediatric Gastroenterology Hepatology and Nutrition Committee on Nutrition and the European Society for Paediatric Gastroenterology Hepatology and Nutrition Working Group for Probiotics and Prebiotics. *Journal of Pediatric Gastroenterology and Nutrition*, *70*(5), 664–680. https://doi.org/10.1097/MPG.0000000000002655

Van Gassen, S., Callebaut, B., Van Helden, M. J., Lambrecht, B. N., Demeester, P., Dhaene, T., & Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry. Part A : The Journal of the International Society for Analytical Cytology*, *87*(7), 636–645. https://doi.org/10.1002/CYTO.A.22625

Vatanen, T., Plichta, D. R., Somani, J., Münch, P. C., Arthur, T. D., Hall, A. B., Rudolf, S., Oakeley, E. J., Ke, X., Young, R. A., Haiser, H. J., Kolde, R., Yassour, M., Luopajärvi, K., Siljander, H., Virtanen, S. M., Ilonen, J., Uibo, R., Tillmann, V., … Xavier, R. J. (2019). Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. *Nature Microbiology*, *4*(3), 470–479. https://doi.org/10.1038/s41564-018-0321-5

Vernocchi, P., Del Chierico, F., & Putignani, L. (2016). Gut microbiota profiling: Metabolomics based approach to unravel compounds affecting human health. *Frontiers in Microbiology*, *7*(JUL), 1144. https://doi.org/10.3389/FMICB.2016.01144/BIBTEX

Wang, H., Meng, X., Xing, S., Guo, B., Chen, Y., & Pan, Y. Q. (2023). Probiotics to prevent necrotizing enterocolitis and reduce mortality in neonates: A meta-analysis. *Medicine (United States)*, *102*(8), E32932. https://doi.org/10.1097/MD.0000000000032932

Wang, Z. L., An, Y., He, Y., Hu, X. Y., Guo, L., Li, Q. Y., Liu, L., & Li, L. Q. (2020). Risk factors of necrotizing enterocolitis in neonates with sepsis: A retrospective case-control study. *Https://Doi.Org/10.1177/2058738420963818*, *34*. https://doi.org/10.1177/2058738420963818

Warner, B. B., Deych, E., Zhou, Y., Hall-Moore, C., Weinstock, G. M., Sodergren, E., Shaikh, N., Hoffmann, J. A., Linneman, L. A., Hamvas, A., Khanna, G., Rouggly-Nickless, L. C., Ndao, I. M., Shands, B. A., Escobedo, M., Sullivan, J. E., Radmacher, P. G., Shannon, W. D., & Tarr, P. I. (2016). Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: A prospective case-control study. *The Lancet*, *387*(10031), 1928–1936. https://doi.org/10.1016/S0140-6736(16)00081-7

Washio, J., Shimada, Y., Yamada, M., Sakamaki, R., & Takahashi, N. (2014). Effects of pH and lactate on hydrogen sulfide production by oral Veillonella spp. *Applied and Environmental Microbiology*, *80*(14), 4184–4188. https://doi.org/10.1128/AEM.00606-14

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 1–18. https://doi.org/10.1186/s40168-017-0237-y

Wendelboe, A. M., Smelser, C., Lucero, C. A., & McDonald, L. C. (2010). Cluster of necrotizing enterocolitis in a neonatal intensive care unit: New Mexico, 2007. *American Journal of Infection Control*, *38*(2), 144. https://doi.org/10.1016/J.AJIC.2009.06.009

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., … Scalbert, A. (2018). HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Research*, *46*(D1), D608–D617. https://doi.org/10.1093/NAR/GKX1089

Wu, R. Y., Li, B., Horne, R. G., Ahmed, A., Lee, D., Robinson, S. C., Zhu, H., Cadete, M., Alganabi, M., Filler, R., Johnson-Henry, K. C., Delgado-Olguin, P., Pierro, A., & Sherman, P. M. (2022). Structure–Function Relationships of Human Milk Oligosaccharides on the Intestinal Epithelial Transcriptome in Caco-2 Cells and

a Murine Model of Necrotizing Enterocolitis. *Molecular Nutrition & Food Research*, *66*(4), 2100893. https://doi.org/10.1002/MNFR.202100893

Xiao, L., & Zhao, F. (2023). Microbial transmission, colonisation and succession: from pregnancy to infancy. *Gut*, *72*(4), 772–786. https://doi.org/10.1136/GUTJNL-2022-328970

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A. M., Härkönen, T., Ryhänen, S. J., Franzosa, E. A., Vlamakis, H., Huttenhower, C., Gevers, D., Lander, E. S., Knip, M., & Xavier, R. J. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Science Translational Medicine*, *8*(343), 343ra81-343ra81. https://doi.org/10.1126/scitranslmed.aad0917

Yilmaz, B., Juillerat, P., Øyås, O., Ramon, C., Bravo, F. D., Franc, Y., Fournier, N., Michetti, P., Mueller, C., Geuking, M., Pittet, V. E. H., Maillard, M. H., Rogler, G., Abdelrahman, K., Ademi, G., Aepli, P., Thomas, A., Anderegg, C., Antonino, A. T., … Macpherson, A. J. (2019). Microbial network disturbances in relapsing refractory Crohn's disease. *Nature Medicine*, *25*(2), 323–336. https://doi.org/10.1038/s41591-018-0308-z

Young, G., Berrington, J. E., Cummings, S., Dorling, J., Ewer, A. K., Frau, A., Lett, L., Probert, C., Juszczak, E., Kirby, J., Beck, L. C., Renwick, V. L., Lamb, C., Lanyon, C. V., McGuire, W., Stewart, C., & Embleton, N. (2022a). Mechanisms affecting the gut of preterm infants in enteral feeding trials: a nested cohort within a randomised controlled trial of lactoferrin. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. https://doi.org/10.1136/ARCHDISCHILD-2022-324477

Young, L., Oddie, S. J., & McGuire, W. (2022b). Delayed introduction of progressive enteral feeds to prevent necrotising enterocolitis in very low birth weight infants. *Cochrane Database of Systematic Reviews*, *2022*(1). https://doi.org/10.1002/14651858.CD001970.PUB6/MEDIA/CDSR/CD001970/IMAGE_N/NCD001970-CMP-001.07.SVG

Yousuf, E. I., Carvalho, M., Dizzell, S. E., Kim, S., Gunn, E., Twiss, J., Giglia, L., Stuart, C., & Hutton, E. K. (2020). Persistence of Suspected Probiotic Organisms in Preterm Infant Gut Microbiota Weeks After Probiotic Supplementation in the NICU. *Frontiers in Microbiology*, *11*, 574137. https://doi.org/10.3389/fmicb.2020.574137

Zachariah, P., Rozenberg, F. D., Stump, S., Moscoso, D. I., Krishnamurthy, G., Saiman, L., Uhlemann, A. C., & Freedberg, D. E. (2021). Evolution of the environmental microbiota of a new neonatal intensive care unit (NICU) and implications for infection prevention and control. *Infection Control & Hospital Epidemiology*, *42*(2), 156–161. https://doi.org/10.1017/ICE.2020.396

Zaramela, L. S., Tjuanta, M., Moyne, O., Neal, M., & Zengler, K. (2022). synDNA—a Synthetic DNA Spike-in Method for Absolute Quantification of Shotgun Metagenomic Sequencing. *MSystems*, *7*(6). https://doi.org/10.1128/MSYSTEMS.00447-22/SUPPL_FILE/MSYSTEMS.00447-22-S0010.XLSX

Zhang, C., Derrien, M., Levenez, F., Brazeilles, R., Ballal, S. A., Kim, J., Degivry, M. C., Quéré, G., Garault, P., Van Hylckama Vlieg, J. E. T., Garrett, W. S., Doré, J., & Veiga, P. (2016). Ecological robustness of the gut microbiota in response to ingestion of transient food-borne microbes. *The ISME Journal*, *10*(9), 2235–2245. https://doi.org/10.1038/ISMEJ.2016.13

Zhang, C., Sherman, M. P., Prince, L. S., Bader, D., Weitkamp, J. H., Slaughter, J. C., & McElroy, S. J. (2012). Paneth cell ablation in the presence of Klebsiella pneumoniae induces necrotizing enterocolitis (NEC)-like injury in the small intestine of immature mice. *Disease Models & Mechanisms*, *5*(4), 522. https://doi.org/10.1242/DMM.009001

Zhang, X., Mozeleski, B., Lemoine, S., Deŕiaud, E., Lim, A., Zhivaki, D., Azria, E., Le Ray, C., Roguet, G., Launay, O., Vanet, A., Leclerc, C., & Lo-Man, R. (2014). CD4 T cells with effector memory phenotype and function develop in the sterile environment of the fetus. *Science Translational Medicine*, *6*(238). https://doi.org/10.1126/SCITRANSLMED.3008748/SUPPL_FILE/6-238RA72_SM.PDF

Zheng, D., Liwinski, T., & Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. In *Cell Research* (Vol. 30, Issue 6, pp. 492–506). Springer Nature. https://doi.org/10.1038/s41422-020-0332-7

Zhernakova, D. V., Le, T. H., Kurilshikov, A., Atanasovska, B., Bonder, M. J., Sanna, S., Claringbould, A., Võsa, U., Deelen, P., Franke, L., de Boer, R. A., Kuipers, F., Netea, M. G., Hofker, M. H., Wijmenga, C., Zhernakova, A., & Fu, J. (2018). Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nature Genetics*, *50*(11), 1524–1532. https://doi.org/10.1038/s41588-018-0224-7

Zhou, Y., Shan, G., Sodergren, E., Weinstock, G., Walker, W. A., & Gregory, K. E. (2015). Longitudinal Analysis of the Premature Infant Intestinal Microbiome Prior to Necrotizing Enterocolitis: A Case-Control Study. *PLOS ONE*, *10*(3), e0118632. https://doi.org/10.1371/JOURNAL.PONE.0118632

Zmora, N., Zilberman-Schapira, G., Suez, J., Mor, U., Dori-Bachash, M., Bashiardes, S., Kotler, E., Zur, M., Regev-Lehavi, D., Brik, R. B. Z., Federici, S., Cohen, Y., Linevsky, R., Rothschild, D., Moor, A. E., Ben-Moshe, S., Harmelin, A., Itzkovitz, S., Maharshak, N., … Elinav, E. (2018). Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell*, *174*(6), 1388-1405.e21. https://doi.org/10.1016/j.cell.2018.08.041