This article has been accepted for publication in GUT following peer review.
The definitive copyedited, typeset version is available online at 10.1136/gutjnl

Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis

Andrea C Masi[1], Nicholas D Embleton[2,3], Christopher A Lamb[1,4], Gregory Young[5], Claire L Granger[1,2], Julia A Najera[6], Daniel P Smith[7], Kristi L. Hoffman[7], Joseph F. Petrosino[7], Lars Bode[6], Janet E Berrington[1,2], Christopher J Stewart[1]

[1] Clinical and Translational Research Institute, Newcastle University, NE2 4HH. UK
[2] Newcastle Neonatal Service, Newcastle Hospitals NHS Trust, NE1 4LP. UK
[3] Population Health Sciences Institute, Newcastle University, NE2 4HH. UK
[4] Department of Gastroenterology, Newcastle upon Tyne Hospitals NHS Foundation Trust, NE1 4LP. UK
[5] School of Health & Life Sciences, Northumbria University, NE1 8ST. UK
[6] Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence (MOMI CORE), University of California San Diego, La Jolla, CA, USA.
[7] Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA

Janet Berrington, Lars Bode, and Christopher Stewart share joint correspondence in this work
Dr Janet Berrington: Neonatal unit, Ward 35, Leazes Wing, Royal Victoria Infirmary, Newcastle, NE1 4LP, Email j.e.berrington@ncl.ac.uk
Dr Lars Bode. Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence (MOMI CORE), University of California San Diego, La Jolla, CA, USA. Email. lbode@health.ucsd.edu
Dr Christopher Stewart. Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, 3rd Floor Leech Building, Newcastle, NE2 4HH. Email: christopher.stewart@newcastle.ac.uk

Word count: 4293

List of abbreviations

NEC - Necrotising enterocolitis

MOM - Mothers own milk, MOM
HMO - Human milk oligosaccharide
2’FL - 2’-fucosyllactose
3FL - 3-fucosyllactose
LNnT - Lacto-N-neotetraose
3’SL - 3’-sialyllactose
DFlac - Difucosyllactose
6’SL - 6’-sialyllactose
LNT - Lacto-N-tetraose
LNFP - Lacto-N-fucopentaose
DFLNT - Difucosyl-LNT
LNH - Lacto-N-hexaose
DSLNT - Disialyllacto-N-tetraose
FLNH - Fucosyl-lacto-N-hexaose
DFLNH - Difucosyl-lacto-N-hexaose
FDSLNH - Fucosyl-disialyl-lacto-N-hexaose
DSLNH - Disialyl-lacto-N-hexaose
PGCT - Preterm gut community types
DOL - Day of life
IQR - Interquartile range
ROC - Receiver operating characteristic
SVM - Support Vector Machine
MCCV - Monte-Carlo cross validation
PERMANOVA - Permutational multivariate analysis of variance
MDA - Mean decrease accuracy
DMM - Dirichlet multinomial modelling
PMA - Postmenstrual age
NICU - Neonatal intensive care unit

Abstract

Objective: Necrotising enterocolitis (NEC) is a devastating intestinal disease primarily affecting preterm infants. The underlying mechanisms are poorly understood: mothers own breast milk (MOM) is protective, possibly relating to human milk oligosaccharide (HMO) and infant gut microbiome interplay. We investigated the interaction between HMO profiles and infant gut microbiome development and its association with NEC.

Design: We performed HMO profiling of MOM in a large cohort of infants with NEC (n=33) with matched controls (n=37). In a subset of 48 infants (14 NEC) we also performed longitudinal metagenomic sequencing of infant stool (n=644).

Results: Concentration of a single HMO, disialyllacto-N-tetraose (DSLNT), was significantly lower in MOM received by NEC infants compared to controls. A MOM threshold level of 241 nmol/mL had a sensitivity and specificity of 0.9 for NEC. Metagenomic sequencing before NEC onset showed significantly lower relative abundance of *Bifidobacterium longum* and higher relative abundance of *Enterobacter cloacae* in infants with NEC. Longitudinal development of the microbiome was also impacted by low MOM DSLNT associated with reduced transition into preterm gut community types dominated by *Bifidobacterium* spp. and typically observed in older infants. Random forest analysis combining HMO and metagenome data before disease accurately classified 87.5% of infants as healthy or NEC.

Conclusion: These results demonstrate the importance of HMOs and gut microbiome in preterm infant health and disease. The findings offer potential targets for biomarker development, disease risk stratification, and novel avenues for supplements that may prevent life-threatening disease.

1. What is already known about this subject?

- Necrotising enterocolitis (NEC) is one of the leading causes of death in preterm infants
- Maternal human milk oligosaccharides (HMOs) including disialyllacto-N-tetraose (DSLNT) have been associated with protection from NEC development
- Differences in infant gut microbiome development have been linked to NEC and non-NEC infants, but the causative and protective organisms have not been determined

2. What are the new findings?

- We found for the first time that combined analysis of maternal HMOs and infant gut microbiome can predict NEC
- A specific DSLNT threshold level of 241 nmol/mL had a sensitivity and specificity of 0.9 for NEC and infants receiving milk below this threshold showed abnormal microbiome development
- Infants who developed NEC had significantly lower relative abundance of *Bifidobacterium longum* and significantly higher relative abundance of *Enterobacter cloacae* before disease diagnosis

3. How might it impact on clinical practice in the foreseeable future?

- Our findings demonstrate the importance of maternal HMOs and infant gut microbiome in preterm infants, providing targets for biomarker development, disease risk stratification, and novel avenues for supplementing the infant feed.

Introduction

Necrotising enterocolitis (NEC) is an inflammation-mediated bowel condition that is a leading cause of death and serious morbidity in preterm infants born before 32 weeks gestation (*1*). The mechanisms underlying NEC development are poorly understood and the lack of specificity of symptoms and tests make diagnostic certainty difficult. Infants with NEC have enteral feeds stopped and are treated with broad-spectrum antibiotics, and may need surgery (*2*).

Receipt of mother's own breast milk (MOM) is the most protective factor against the development of NEC in preterm infants (*3, 4*). However, infants receiving MOM still develop NEC, suggesting the variable composition of nutrients and other components of breast milk may be important. Human milk oligosaccharides (HMOs) are structurally diverse, complex unconjugated sugars that are not usually present in artificial formula milk (*5*). HMOs are indigestible to the infant, reaching the lower gastrointestinal tract intact where they act as growth substrates (i.e. prebiotics) for specific bacteria, notably *Bifidobacterium* spp. thought key to infant health (*6-8*). HMOs may also protect from enteric organism blood stream infections due to anti-microbial activity (*9*), stimulate the immune system (*10*), enhance gut barrier function (*11*), and act as decoy receptors for pathogens (*12*). While >150 HMOs have been described, the 19 most abundant represent >95% of the total HMO content (*13*). HMO profiles are specific to individual mothers and remain relatively stable during lactation (*14*). Presence of an active FUT2 gene, which is involved in the synthesis of α1-2-fucosylated oligosaccharides, is the main determinant of the HMO profile, termed maternal secretor status (*15*).

Recent work has begun to elucidate the potential contribution of HMOs to preterm infant health. In a neonatal rat model, disialyllacto-N-tetraose (DSLNT), a non-fucosylated, but double-sialylated HMO, significantly reduced NEC development and improved NEC-associated mortality rate (*16*). An association of lower DSLNT concentration in MOM and subsequent higher risk of NEC onset in the infant has since been observed in preterm human studies (*17-19*). To date, these studies have

included very small numbers of infants with NEC (between 4 and 8), with a broad range of NEC phenotypes. Thus, validation in a larger cohort is urgently needed.

Altered gut microbiome development has been associated with NEC in preterm infants. While no specific causative microorganism has consistently been identified, studies have reported a higher relative abundance of Enterobacteriaceae, coupled to lower relative abundance of *Bifidobacterium* (*20*-*23*). Instability of the gut microbiome in NEC infants has also been reported in longitudinal studies, with more frequent transitions between different preterm gut community types (PGCT) in NEC (*20*). These findings were replicated at the site of disease in a study using formalin fixed paraffin embedded tissue from NEC infants matched to non-NEC controls (*24*). Previous microbiome studies have largely relied on 16S rRNA gene sequencing of the V4 region, which has limited resolution, especially for emerging key organisms of interest for preterm health (i.e., *Klebsiella* and *Enterobacter* would be classified together as Enterobacteriaceae). Metagenomics may overcome this and recent metagenomic data showed infants who developed NEC had higher relative abundance of *Klebsiella* and higher replication rates in all bacteria before disease onset (*25*).

In this current study, we performed a combined analysis of maternal HMO profiles and longitudinal development of the infant stool/gut microbiome in a large cohort of preterm infants with NEC and healthy controls matched for gestation, birthweight and day of life. We then validated our results in an independent cohort using previously published HMO data (*17*). We hypothesised that differences in maternal HMO profiles and microbiome development may explain why some infants receiving MOM still develop NEC.

Methods

Cohort

This study included 77 preterm infants (born at <32 weeks gestation), who were born in or transferred to a single large tertiary level neonatal intensive care unit (NICU) in Newcastle upon Tyne, UK recruited to the SERVIS study (REC10/H0908/39) with written parental consent covering data and sample collection. 33 infants were diagnosed with definite NEC and 37 non-diseased controls were selected by identifying a healthy infant matched by gestation, birthweight, and having a MOM sample available at a corresponding day of life (DOL) (**table 1**). Detailed information on feeding and antibiotic use are included in **online supplementary table 1 and online supplementary table 2**. Diagnoses were made using an extensive combination of clinical, x-ray and histological findings and blindly agreed by two neonatal clinicians (JEB and NDE). Standard clinical protocols recommended the routine use of supplemental probiotics when more than 30mL/kg/day of MOM was tolerated for at least 1-2 days: all 33 NEC infants received MOM and 31 received probiotics. All 37 controls received probiotics. The probiotics administered were either LaBiNIC (*Lactobacillus acidophilus, Bifidobacterium infantis* and *Bifidobacterium bifidum*) or Infloran (*L. acidophilus* and *B. bifidum*).

The median DOL of NEC diagnosis was 19 (interquartile range; IQR 14-35; **table 1**). A single MOM sample was analysed for each infant, as close to the onset of disease as possible, with control samples matched by day of life (DOL) (**online supplementary figure 1**). The median DOL of MOM from NEC cases was 18 (IQR 13-34) and from controls was also 18 (IQR 12-31). The DOL of the milk sample is the DOL received by the infant, and is not necessarily the same day as the mother expressed the milk due to standard practice that often involves milk storage. Metagenomic sequencing of stool samples (n = 644) was performed longitudinally on a subset of 48 infants (including 14 NEC; **online supplementary figure 1**). These infants were comparable to the full cohort (**online supplementary table 3**).

Full details of the HMO, metagenome, and statistical analysis are described in **online supplementary methods**.

Human milk oligosaccharides analysis

The absolute quantification for the 19 most abundant HMOs was determined by high-performance liquid chromatography (HPLC) following derivatization as per the protocol described by Bode *et al.* (*26*). Maternal secretor (presence of an active FUT2 gene) status was determined by presence or near-absence of 2'FL in the breast milk analysed.

Metagenomes

DNA was extracted from ~0.1g of stool using the DNeasy PowerSoil Kit (QIAGEN) following the manufacturer's protocol and sequencing was performed on the HiSeq X Ten (Illumina) with a read length of 150bp paired end reads. Processed fastq files were mapped against the MetaPhlan2 marker gene database (mpa_v20_m200) (*27*).

Statistical analysis

Statistical analysis of HMO profiles was performed using MetaboAnalyst 3.0 (*28*). For ordinations, HMO data was normalised by logarithmic transformation and 2000 random permutations were used to test the significance. Multivariate ROC curves were generated using linear Support Vector Machine (SVM) classification method coupled with Monte-Carlo cross validation (MCCV).

Correlation between clinical variables and individual HMOs was tested by performing a multivariate adjusted linear model in R (version 3.6.3). HMO concentrations were normalised by log-transformation prior to analysis and P values were adjusted applying the Benjamini & Hochberg correction (*29*).

The cross-sectional cohort of stool samples collected from NEC infants before diagnosis and matched controls was analysed using MicrobiomeAnalyst (*30, 31*). Permutational multivariate

analysis of variance (PERMANOVA) was used to determine significance of Bray-Curtis principal coordinate analysis. MetagenomeSeq was used to assess differential abundance at the phyla and species level.

DMM clusters samples on the basis of microbial community structure (*32*) and was used to determine the preterm gut community types (PGCTs) from all samples, as performed previously (*33*, *34*). Five PGCT was found to be optimal, and these were ordered youngest (PGCT-1) to oldest (PGCT-5) based on the average DOL of samples within each PGCT. Analysis was performed at specific time windows, including only a single sample per infant in each time point.

The association of various clinical variables on the HMO and metagenome profiles was tested by applying the function "adonis" of "vegan" (version 2.5-6) package (*35*) in R, based on Bray-Curtis dissimilarity and 10000 permutations. Each test was performed stepwise and P values were adjusted using Benjamini & Hochberg (*29*).

Random Forest was used for comparing the performance of classification models built using matched cross-sectional datasets.

Results

Association of maternal HMOs and development of NEC in the infant

MOM samples clustered according to maternal secretor status and secretor mothers had a higher total HMO concentration, a higher HMO Shannon diversity, and a significantly higher concentration of overall HMO-bound fucose (**online supplementary figure 2**). Thus, where relevant, we have stratified and adjusted for maternal secretor status in subsequent analyses.

HMO profiles showed significant separation of NEC and control infants (**figure 1a**; 2000 permutations, $P < 0.001$) and this was consistent when secretor and non-secretor samples were analysed separately (both $P < 0.001$; **online supplementary figure 3a and 3b**). Individually, of the 19 HMOs quantified in this study, only DSLNT was significantly different between NEC and controls, with a lower concentration in NEC infants (adj. $P < 0.001$; **figure 1b and 1c**). No

significant associations were found in the Shannon diversity of HMOs between NEC and matched controls for the full cohort, or when stratified by maternal secretor status (all $P > 0.05$; **online supplementary figure 4**).

Given that lower DSLNT was associated with NEC independent of secretor status, the utility of this HMO as a biomarker for NEC development was explored. Univariate ROC curve analysis determined that 241 nmol/mL (or 310.93 µg/mL) was the optimal DSLNT concentration in MOM for distinguishing NEC and control infants (**figure 1c** and **1d**). At this threshold, the area under the curve (AUC) was 0.946 with a sensitivity of 0.9 and a specificity of 0.9, correctly identifying 91% of NEC infants (below threshold) and 86% of control healthy infants (above threshold).

To test if integration of additional HMOs could improve the classification performance, multivariate ROC curves built on increasing number of HMOs were performed (**online supplementary figure 5a**). Inclusion of 2 HMOs (the minimum in multivariate analysis) resulted in the optimal performance, with DSLNT being selected as a discriminatory feature in 100% of permutations (**online supplementary figure 5b**). 3FL and LNnT were the 2nd and 3rd most selected features, with a selection frequency of around 30%, being more abundant in cases of NEC. However, the integration of any additional HMOs to DSLNT in the multivariate model resulted in minimal improvement in performance compared to the univariate model using DSLNT only (AUC of 0.949 and 0.946, respectively).

To validate the 241 nmol/mL threshold defined in the current study in an independent cohort, we analysed data from Autran *et al.* (2018) which contained 8 NEC and 40 matched control infants (*17*). Since this study included temporal sampling before disease, we selected the nearest milk sample to NEC onset for each infant and matched the control samples by sample DOL and included only DSLNT concentration. Using a DSLNT threshold of 241 nmol/mL, the MOM sample for 100% (8/8) NEC infants fell under the threshold, while 60% (24/40) control samples had a DSLNT concentration above 241 nmol/mL (**online supplementary figure 5c**).

Analysis of HMO profiles stratified by NEC type

We compared medically managed NEC (NEC-M), where infants did not undergo surgery or die from NEC (i.e. had less severe disease), with NEC infants that underwent surgery (NEC-S). NEC-M and NEC-S clustered together and were distinct from matched controls (**figure 2a**; 2000 permutations, $P < 0.001$). Two HMOs were found to be significantly different, with DSLNT lower in MOM in both NEC-M (adj. $P < 0.001$) and NEC-S (adj. $P < 0.001$) compared to controls (**figure 2b**). In addition, LNnT in MOM was significantly lower in NEC-S in comparison to both NEC-M (adj. $P = 0.0016$) and matched controls (adj. $P = 0.0423$) (**figure 2c**).

We subsequently investigated the potential association between DSLNT and LNnT concentrations and clinical variables by applying an adjusted linear model. DSLNT was negatively correlated to both disease types, with coefficients equal to -0.60 for NEC-M (adj. $P < 0.001$) and -0.67 for NEC-S (adj. $P < 0.001$) (**figure 2d**). However, LNnT was not associated with disease type following adjusted linear modelling (both adj. $P > 0.05$). DSLNT and LNnT were both significantly higher in secretor mothers (adj. $P = 0.008$, adj. $P < 0.001$, respectively). DSLNT in MOM also positively correlated to gestational age (adj. $P = 0.008$) and negatively to birthweight (adj. $P = 0.008$). Neither HMO correlated to sex, delivery mode, post-menstrual age, or DOL of the MOM sample (**figure 2d** and **online supplementary figure 6**).

Association of infant gut microbiome and development of NEC

We included stool microbiome data on a subset of infants with HMO data, where metagenomic sequencing data was available through an on-going independent study (the results of which are not yet published). This included 644 stool samples from 34 controls, and 14 NEC infants (**online supplementary table 3**). To overcome challenges of repeated measures and to compare results with existing published work, we first analysed one stool sample per infant closest to NEC onset (median of 3 days before NEC) and a corresponding control sample matched by DOL (**online supplementary figure 1**). This cross-sectional analysis showed NEC infants had significantly

lower richness (P = 0.027) but comparable Shannon diversity (P = 0.443; **figure 3a**). Bray-Curtis PCoA showed no significant difference between the bacterial profiles of NEC and controls (PERMANOVA P = 0.182; **figure 3b**). Analysis at the phylum level showed significantly lower relative abundance of Actinobacteria (adj. P = 0.034) and higher relative abundance of Proteobacteria (adj. P = 0.034) in NEC infants (**figure 3c**). Correspondingly, at the species level, NEC infants had lower relative abundance of *Bifidobacterium longum* (adj. P = 0.012) and higher relative abundance of *Enterobacter cloacae* (adj. P = 0.012), compared to controls (**figure 3d**).

Integrated analysis of HMO and bacterial profiles

DMM clustering was used to determine preterm gut community types (PGCT) using species level data and five PCGTs was deemed optimal (**figure 4a**). PGCT-1 was characterised by high relative abundance of *Staphylococcus* spp. and *Enterococcus faecalis*, PGCT-2 had high *Escherichia* spp., PGCT-3 had high *Klebsiella* spp., and PGCT-4 and PGCT-5 had high *Bifidobacterium* spp. with *B. breve* notably high in PGCT-5. Using the PGCT clusters, we analysed the temporal transition of an infant's gut microbiome over the first 70 days of life by defining distinct time points and including only one sample per infant at each time point. Based on the distribution of samples across all time points and all clusters, the temporal transition of the microbiome over the first 70 days of life was significantly different in infants in receipt of MOM below the DSLNT threshold of 241 nmol/ml compared to infants above the DSLNT threshold (χ^2 test P < 0.001; **figure 4b**).

The PGCTs were named according to the average age of samples within that cluster, where PGCT-1 contained on average the earliest samples and PGCT-5 on average the latest samples. We compared the number of samples from all time points in only PGCT-1 and PGCT-5 to investigate associations between the MOM DSLNT threshold and gut microbiome development from the typically younger to the typically older PGCTs. Infants receiving MOM with DSLNT level below 241 nmol/mL had significantly more samples remaining within PGCT-1 throughout all time points (78% in PCGT-1 vs. 22% in PGCT-5, χ^2 test P <0.001), whereas infants receiving MOM with DSLNT above this

threshold transitioned from PGCT-1 to PGCT-5 as demonstrated by a similar number of samples in each PGCT across all time points (48% in PCGT-1 vs. 52% in PGCT-5, χ^2 test P = 0.717). In addition to comparing samples from all times points, we next compared samples from the final time point only (i.e., DOL 50-60). After correcting for uneven frequency of sampling between groups, at the final time point infants receiving MOM above the DSLNT threshold were twice as likely to be in PGCT-5 (3/11 samples below vs. 12/22 samples above DSLNT threshold; odds ratio 3.20, 95% CI 0.6657 to 15.3819), which was characterised by high relative abundance of *Bifidobacterium* (**figure 4a**).

Explained variance and random forest classification of HMO and metagenome data

Using the cross-sectional HMO and cross-sectional metagenome dataset, we sought to determine which clinical factors were most associated with the HMO and the bacterial profiles (**figure 5a**). Secretor status explained 56% of the variance within HMO profiles (adj. P < 0.001), but no other co-variate was significantly associated with the HMO profiles. In contrast, the bacterial profiles were significantly associated with both postmenstrual age (R2 0.07; adj. P = 0.006) and day of life (R2 0.07; adj. P = 0.006), as well as receipt of antibiotics at the time of sampling (R2 0.06; adj. P = 0.006) and receipt of probiotics (R2 0.12; adj. P = 0.006), but not maternal secretor status (R2 0.02; adj. P = 0.58). Together, these findings highlight that HMO and bacterial profiles are influenced by numerous non-overlapping factors related to early life in preterm infants.

We compared the performance of random forest classification models built on the cross-sectional subset of HMO profile data, metagenomic sequencing data, and the two datasets combined to classify an infant as NEC or healthy, given that all this information is available before onset of disease and could therefore function as a risk stratification system in clinical practice. The HMO profile alone had a classification error of 0.146, with 21% (3/14) NEC and 12% (4/34) control infants misclassified. DSLNT had the greatest contribution to classification with a Mean Decrease Accuracy (MDA) of 0.11. Other HMOs contributing to classification accuracy included LNH

(MDA = 0.012) and DFLNH (MDA = 0.011), which were non-significantly higher in NEC infants. Random forest generated using the metagenomic sequencing data was characterised by a classification error of 0.229, with 43% (6/14) NEC and 15% (5/34) control infants misclassified. *Enterobacter cloacae* was the most important feature guiding the classification (MDA = 0.036), with higher relative abundance in NEC infants, followed by *Bifidobacterium bifidum* (MDA = 0.024) and *Bifidobacterium longum* (MDA = 0.013) which had higher relative abundance in control infants. Combining HMO and metagenome datasets slightly improved the performance compared to using HMOs alone, with 21% (3/14) NEC infants and (9%) 3/34 controls misclassified. In this combined model, DSLNT was enriched in controls and DSLNH and the relative abundance of *Escherichia unclassified* were higher in NEC infants (**figure 5b**).

Discussion

Receipt of human breast milk and early life gut microbiome development are intrinsically linked and both influence the risk of NEC in preterm infants. Our study represents the largest analysis of HMOs in NEC and the first to integrate HMO and metagenome data. We found DSLNT was present in significantly lower concentrations in MOM fed to infants diagnosed with NEC. Furthermore, lower DSLNT concentrations in MOM were associated with reduced transition into PGCTs typically observed in older infants and lower relative abundance of *Bifidobacterium* spp.

The HMO results from the current study build upon previous findings in humans, showing reduced DSLNT in MOM received by infants developing NEC, independent of maternal secretor status (*17-19*). This is also supported by rodent studies where total and individual HMOs including 2'FL and DSLNT have shown a protective effect against NEC development (*16, 36, 37*). However, 2'FL and mixtures of HMOs (one of which included DSLNT) did not show any protection in NEC piglet models (*38, 39*). Importantly from a clinical perspective, in rats the protection provided by pooled HMOs could be reproduced with DSLNT alone, with specific dependence on its precise structure since closely related sialyllacto-N-tetraose (identical in structure to DSLNT but lacking one sialic

acid residue) did not provide protection, suggesting a highly structure-specific mechanism (*16*). Our findings further extend the evidence for the specificity of DSLNT in the NEC pathway. A threshold level of DSLNT (241 nmol/mL) from a single MOM sample correctly identified 91% of NEC infants (below threshold) and 86% of control healthy infants (above threshold). Of the three infants who developed NEC despite a DSLNT above the threshold, two had not received MOM in the 3 weeks prior to disease onset and the remaining infant had a DSLNT concentration of 248 nmol/ml.

Within the validation dataset (*17*), 100% NEC infants were correctly classified, but only 60% of controls. Making a robust diagnosis of NEC is difficult and it is possible that the specific threshold value of DSLNT we identified will have a different predictive value in other populations or where other criteria are used to determine the presence of disease. Our study contains a large number of cases coded clinically as NEC independently validated by blinded review. In addition, our cohort was more homogenous (predominantly white Caucasian) and the concentration of DSLNT less variable (current study IQR 184-321 nmol/mL vs. Autran *et al.* IQR 122-346 nmol/mL) despite using the same analytical platform. Given HMO composition and DSLNT concentrations may be influenced by genetic factors, geographical location, ethnicity (*40*), and seasonality (*15*), differential thresholds may improve diagnostic performance in other settings. Taken together, this external validation and potential variation in DSLNT concentration by maternal factors underscore the need for large multicentre studies to both refine a universal or stratified threshold for DSLNT concentration in predicting NEC and potentially prospectively identifying milk samples that may benefit from supplementation with synthetically produced DSLNT.

In addition to HMO profiles, our extensive longitudinal stool metagenomic analysis, represents one of the largest datasets to date. This extends our previous work (*20, 33, 41*) where DMM was used to facilitate analysis of temporal microbiome development, and integrate the HMO DSLNT threshold of 241 nmol/ml with infant gut microbiome profiles. We observed a difference in microbiome development between DSLNT groups, with infants receiving MOM with lower DSLNT tending to have delayed progression into the PGCT typically expected in older infants (i.e., PGCT-5). This

supports the theory that concentrations of specific HMOs in MOM are associated with differences in gut microbiome development. On the contrary, transition into PGCT-5 was twice as likely in infants receiving MOM with DSLNT above the threshold, which was characterised by high relative abundance of *Bifidobacterium* spp. *Bifidobacterium* has previously been linked to health in preterm infants (*20, 41, 42*) and our current findings in pre-NEC samples further support the association of reduced *Bifidobacterium* spp., specifically *Bifidobacterium longum*, as a risk factor for NEC. In addition, our species level metagenome data advanced previous associations of Enterobacteriaceae with NEC (*21, 23, 43*), showing *E. cloacae* relative abundance was higher before NEC.

Random forest analysis confirmed the capability of HMO profiles to identify infants who developed NEC and slightly outperformed metagenome profiles by correctly classifying three more NEC cases and one more control. Combining HMO and metagenome data before disease accurately classified 87.5% of infants as healthy or NEC, with DSLNT and the bacterial species identified as important in the random forest analysis being comparable to the unsupervised analysis in the current study and in previous studies. Further work is needed to determine if DSLNT functions via modulation of the microbiome or by acting directly on the host, such as acting in a structure-specific receptor-mediated way to alter immune functioning and reduce inflammation leading to necrosis. In the event of the latter, a microbial community with less DSLNT utilisation could provide an advantage to reducing NEC risk. Taken together, the current findings and recent work highlighting the ability of *Bifidobacterium* spp. to utilise HMOs is strain specific (*7, 8*) underscore the need for further research to better understand the complexity of human milk and other nutritional exposures, including the use of supplements such as prebiotics and probiotics in preterm infants. In addition to therapeutics, the classifiers may provide a basis for the development of biomarkers predicting NEC risk. While additional work is needed, the addition of microbial biomarkers may allow for the most accurate predictions and could inform NEC risk for infants where MOM (and thus HMO information) is not available.

This study involved the largest cohort to date investigating the relationship between HMO composition and NEC development, and includes one of the most extensive longitudinal stool metagenomic analyses of preterm infants. However, there are several limitations and avenues for future work. First, the cross-sectional HMO profiling data precluded assessment of changes within mothers over time and how this may relate to NEC development. The milk sample was selected based on the day of infant feeding and the actual expression of milk may have occurred several days earlier, which may be important clinically. Current published data suggest that the concentration of HMOs, including DSLNT, are relatively stable over time (*14*), but validation in longitudinal preterm cohorts is needed. Second, the amount of MOM an infant receives and tolerates each day is variable, and DSLNT exposure is dependent on both concentration and volume. Although this study identifies DSLNT concentration alone may be useful from both a diagnostic and therapeutic perspective, further studies could consider the volume of milk received in addition to concentration. Thirdly, inclusion of metagenome data was opportunistic based on available data and cost prohibited sequencing all infants in the cohort. As such, the classification accuracy of the model might be impacted by the reduced sample size in comparison to the full cohort, necessitating the need for follow-up analyses in larger cohorts. Despite this, the sample size of 644 including 195 samples from 14 preterm infants who developed NEC makes this dataset one of the largest published to date. Finally, the gene relative abundance data warrants further investigation, in combination with other experimental approaches, to help inform the HMO utilisation capacity of different strains.

In summary, HMO profiling of MOM coupled to metagenomic sequencing of preterm stool showed that the concentration of a single HMO, DSLNT, was lower in milk received by infants who developed NEC. The lower concentration of DSLNT was associated with altered microbiome development, specifically a reduced progression toward the PGCT typically found in the older infants which was abundant in *Bifidobacterium* spp. These results suggest MOM HMO profiling may provide potential targets for biomarker development and disease risk stratification. They may

444 also guide focussed donor milk use (e.g., prioritise high DSLNT for preterm infants) and novel
445 avenues for supplements that may prevent life-threatening disease.
446

References

1. A. C. Masi, C. J. Stewart, The role of the preterm intestinal microbiome in sepsis and necrotising enterocolitis. *Early Hum Dev* **138**, 104854 (2019).
2. J. Neu, W. A. Walker, Necrotizing enterocolitis. *N Engl J Med* **364**, 255-264 (2011).
3. J. Meinzen-Derr *et al.*, Role of human milk in extremely low birth weight infants' risk of necrotizing enterocolitis or death. *J Perinatol* **29**, 57-62 (2009).
4. S. Sullivan *et al.*, 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. *J Pediatr* **156**, 562-567 e561 (2010).
5. L. Bode, Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147-1162 (2012).
6. T. Vatanen *et al.*, The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* **562**, 589-594 (2018).
7. M. A. E. Lawson *et al.*, Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. *ISME J*, (2019).
8. A. Gotoh *et al.*, Sharing of human milk oligosaccharides degradants within bifidobacterial communities in faecal cultures supplemented with Bifidobacterium bifidum. *Sci Rep* **8**, 13958 (2018).
9. D. L. Ackerman *et al.*, Human Milk Oligosaccharides Exhibit Antimicrobial and Antibiofilm Properties against Group B Streptococcus. *ACS Infect Dis* **3**, 595-605 (2017).
10. L. Bode *et al.*, Inhibition of monocyte, lymphocyte, and neutrophil adhesion to endothelial cells by human milk oligosaccharides. *Thromb Haemost* **92**, 1402-1410 (2004).
11. R. Y. Wu *et al.*, Human Milk Oligosaccharides Increase Mucin Expression in Experimental Necrotizing Enterocolitis. *Mol Nutr Food Res* **63**, e1800658 (2019).
12. A. Gustafsson *et al.*, Carbohydrate-dependent inhibition of Helicobacter pylori colonization using porcine milk. *Glycobiology* **16**, 1-10 (2006).
13. S. Ramani *et al.*, Human milk oligosaccharides, milk microbiome and infant gut microbiome modulate neonatal rotavirus infection. *Nat Commun* **9**, 5010 (2018).
14. S. Austin *et al.*, Human Milk Oligosaccharides in the Milk of Mothers Delivering Term versus Preterm Infants. *Nutrients* **11**, (2019).
15. M. B. Azad *et al.*, Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. *J Nutr* **148**, 1733-1742 (2018).
16. E. Jantscher-Krenn *et al.*, The human milk oligosaccharide disialyllacto-N-tetraose prevents necrotising enterocolitis in neonatal rats. *Gut* **61**, 1417-1425 (2012).
17. C. A. Autran *et al.*, Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants. *Gut* **67**, 1064-1070 (2018).
18. E. Van Niekerk *et al.*, Human milk oligosaccharides differ between HIV-infected and HIV-uninfected mothers and are related to necrotizing enterocolitis incidence in their preterm very-low-birth-weight infants. *J Nutr* **144**, 1227-1233 (2014).
19. D. Hassinger, D. M. Clausen, S. Nitka, A. Herdt, I. Griffin, Analysis of Disialyllacto-N-Tetraose (DSLNT) Content in Milk From Mothers of Preterm Infants. *J Hum Lact*, 890334420904041 (2020).
20. C. J. Stewart *et al.*, Temporal bacterial and metabolic development of the preterm gut reveals specific signatures in health and disease. *Microbiome* **4**, 67 (2016).
21. K. Sim *et al.*, Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin Infect Dis* **60**, 389-397 (2015).
22. D. V. Ward *et al.*, Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell Rep* **14**, 2912-2924 (2016).
23. A. L. Morrow *et al.*, Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome* **1**, 13 (2013).

24. C. J. Stewart *et al.*, 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 Microbiol* **19**, 52 (2019).
25. M. R. Olm *et al.*, Necrotizing enterocolitis is preceded by increased gut bacterial replication, Klebsiella, and fimbriae-encoding bacteria. *Sci Adv* **5**, eaax5727 (2019).
26. L. Bode *et al.*, Human milk oligosaccharide concentration and risk of postnatal transmission of HIV through breastfeeding. *Am J Clin Nutr* **96**, 831-839 (2012).
27. D. T. Truong *et al.*, MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12**, 902-903 (2015).
28. J. Xia, I. V. Sinelnikov, B. Han, D. S. Wishart, MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res* **43**, W251-257 (2015).
29. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300 (1995).
30. J. Chong, P. Liu, G. Zhou, J. Xia, Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat Protoc* **15**, 799-821 (2020).
31. A. Dhariwal *et al.*, MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res* **45**, W180-W188 (2017).
32. I. Holmes, K. Harris, C. Quince, Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **7**, e30126 (2012).
33. C. J. Stewart *et al.*, Cesarean or Vaginal Birth Does Not Impact the Longitudinal Development of the Gut Microbiome in a Cohort of Exclusively Preterm Infants. *Front Microbiol* **8**, 1008 (2017).
34. C. J. Stewart *et al.*, Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583-588 (2018).
35. F. G. B. Jari Oksanen, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, vegan: Community Ecology Package. R package version 2.5-6. . *https://CRAN.R-project.org/package=vegan*, (2019).
36. M. Good *et al.*, The human milk oligosaccharide 2'-fucosyllactose attenuates the severity of experimental necrotising enterocolitis by enhancing mesenteric perfusion in the neonatal intestine. *Br J Nutr* **116**, 1175-1187 (2016).
37. C. A. Autran, M. H. Schoterman, E. Jantscher-Krenn, J. P. Kamerling, L. Bode, Sialylated galacto-oligosaccharides and 2'-fucosyllactose reduce necrotising enterocolitis in neonatal rats. *Br J Nutr* **116**, 294-299 (2016).
38. S. O. Rasmussen *et al.*, Human milk oligosaccharide effects on intestinal function and inflammation after preterm birth in pigs. *J Nutr Biochem* **40**, 141-154 (2017).
39. M. S. Cilieborg *et al.*, Minimal short-term effect of dietary 2'-fucosyllactose on bacterial colonisation, intestinal function and necrotising enterocolitis in preterm pigs. *Br J Nutr* **116**, 834-841 (2016).
40. M. K. McGuire *et al.*, What's normal? Oligosaccharide concentrations and profiles in milk produced by healthy women vary geographically. *Am J Clin Nutr* **105**, 1086-1100 (2017).
41. C. J. Stewart *et al.*, Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. *Microbiome* **5**, 75 (2017).
42. R. M. Torrazza *et al.*, Intestinal microbial ecology and environmental factors affecting necrotizing enterocolitis. *PLoS One* **8**, e83304 (2013).
43. V. Mai *et al.*, Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One* **6**, e20647 (2011).

Acknowledgements

The authors wish to thank the neonatal intensive care staff involved in the sample collection. We are grateful to the families for their willingness to help and support research.

Contributions

NDE, JEB, and CJS conceived and designed the study. ACM, CAL, GY, CLG, JEB, and CJS collected the samples and overseen the logistics. JAN and LB performed the HMO profiling. KLH, and JFP performed the bioinformatics on fastq files. ACM, DPS, and CJS performed the analysis. NDE, JEB, and CJS supervised the study. ACM, NDE, CAL, JEB, and CJS co-wrote the manuscript and all authors approved the final submission.

Funding

This work was supported by the MRC Discovery Medicine North Doctoral Training Partnership (to ACM and CJS), the Newcastle University academic career track scheme (to CJS) and Astarte Medical for funding stool sample retrieval and shipment for metagenomic sequencing (to CJS). The funders played no part in the study design, analysis, interpretation, or reporting.

Competing interest

CJS declares performing consultancy for Astarte Medical and honoraria from Danone Early Life Nutrition. NDE declares research funding from Prolacta Biosciences US and Danone Early life Nutrition, and received lecture honoraria from Baxter and Nestle Nutrition Institute, but has no share options or other conflicts. LB is UC San Diego Chair of Collaborative Human Milk Research, endowed by the Family Larsson-Rosenquist Foundation, and serves on the foundation's scientific advisory board. LB is (co-)inventor on patent applications regarding human milk oligosaccharides in prevention of necrotizing enterocolitis and other inflammatory disorders. The other authors declare that they have no competing interests.

Ethical approval

Ethics approval was obtained from the County Durham and Tees Valley Research Ethics Committee (REC10/H0908/39) and parents gave informed consent for stool and data collection.

Data availability statement

Data are available upon reasonable request. All sequencing data generated and analysed in this study have been deposited in the European Nucleotide Archive (ENA) under study accession number PRJEB39610.

Tables

Table 1. Demographics of the analytical cohort with human milk oligosaccharide profile data. Differences between groups were tested applying Chi-square test and Dunn's post-hoc test where applicable.

	Control	NEC	P value
Number of patients	37	33	-
Secretors	25 (68%)	20 (61%)	0.544
Male	14 (38%)	22 (67%)	0.016
Vaginal delivery	25 (68%)	17 (52%)	0.171
Gestational age	25 [24; 26]	25 [24; 27]	0.881
Birthweight	670 [585; 830]	670 [600; 840]	1.000
Probiotics ever	37 (100%)	31 (94%)	0.855
MOM only	3 (8%)	6 (18%)	0.468
MOM + Formula	11 (30%)	12 (37%)	
MOM + BMF	10 (27%)	7 (21%)	
MOM + Formula + BMF	13 (35%)	8 (24%)	
DOL breast milk sample	18 [12; 31]	18 [13; 34]	0.636
DOL disease onset	-	19 [14; 35]	-
NEC surgical	-	16	-

NEC, necrotising enterocolitis; MOM, mother's own breast milk; BMF, breast milk fortifier; DOL, day of life

Figure Legends

Figure 1. Analysis of human milk oligosaccharide (HMO) profiles and DSLNT concentration in necrotising enterocolitis (NEC) and controls. a, Orthogonal partial least squares discriminant analysis (OPLS-DA) of maternal HMO profiles fed to infants diagnosed with NEC and controls. The P value was calculated based on 2000 permutations. **b,** Visual representation of P values obtained from comparison of individual HMOs between NEC and control group. Wilcoxon rank sum test was applied, and P values adjusted with FDR algorithm. The line indicates P value = 0.05**. c,** Univariate receiver operating characteristic (ROC) curve generated on DSLNT concentration

identified 241 nmol/ml as the best threshold for NEC prediction. The performance of the classification is defined by the area under the curve, specificity (false positive rate) and sensitivity (false negative rate). **d**, Box plot showing the concentration of disialyllacto-N-tetraose (DSLNT) between NEC and controls. Blue line represents the 241 nmol/ml threshold.

Figure 2. Analysis of human milk oligosaccharide (HMO) profiles with stratification of necrotising enterocolitis (NEC) into medical (NEC-M) and surgical (NEC-S). a, Partial least squares discriminant analysis (PLS-DA) of HMO profiles from control, NEC-M, and NEC-S infants. NEC-M and NEC-S cluster together and separately from controls ($P < 0.001$). P values were calculated based on 2000 permutations. Box plots of (**b**) disialyllacto-N-tetraose (DSLNT) and (**c**) lacto-N-neotetraose (LNnT) concentration between control, NEC-M, and NEC-S infants. Kruskal-Wallis followed by Dunn's test using Bonferroni adjustment was applied. **d**, Adjusted linear regression model for DSLNT and LNnT including potential clinical confounders. P values were corrected by FDR. Significant variables are indicated by asterisks: *** denotes FDR $P < 0.001$; ** denotes FDR $P > 0.01$. DOL, day of life; PMA post-menstrual age; GA, gestational age.

Figure 3. Cross-sectional analysis of preterm stool metagenome profiles between necrotising enterocolitis (NEC) and matched controls. Analysis includes the sample closest NEC onset (median of 3 days prior to NEC) and a corresponding control sample matched by day of life. **a,** Alpha-diversity based on observed species (richness) and Shannon diversity. **b,** Bray-Curtis principal coordinate analysis. **c,** Box plots showing the relative abundance of significant phyla. **d,** Box plots showing the relative abundance of significant species.

Figure 4. Analysis of preterm gut community types (PGCTs) by infants receiving maternal milk above or below the 241 nmol/mL DSLNT threshold. The entire dataset of 644 samples formed five distinct clusters based on lowest Laplace approximation following Dirichlet multinomial clustering. **a,** Heatmap showing the relative abundance of dominant bacterial species within each PGCT cluster. The phyla for each species are also shown. **b**, Transition model showing the progression of samples through each PGCT, from day of life 0 to 60 across eight distinct time points. Plots are separated based on whether the concentration of disialyllacto-N-tetraose (DSLNT) in maternal milk was above or below the 241 nmol/mL threshold. Nodes and edges are sized based on the total counts. Nodes are coloured according to DMM cluster number and edges are coloured by the transition frequency. Transitions with less than 5% frequency are not shown.

Figure 5. Modelling of cross-sectional human milk oligosaccharide (HMO) and infant stool metagenomic profiles using Adonis and random forest. a, horizontal bar plots showing the variance (r2) in maternal HMO and infant stool metagenomic profiles explained by clinical covariates as modelled by univariate Adonis. Variables with an FDR $P < 0.05$ are shown in red. DOL, day of life; PMA, post-menstrual age. **b,** Feature importance from combined HMO and metagenome random forest classification model. Mean decrease accuracy (MDA) value defines the contribution given by a certain feature to classification process.

Online supplementary figure Legends

Online supplementary figure 1. Sampling schematic for the entire cohort. Only samples collected in the first 100 days of life are shown. Shapes represent sample timing in relation to the diagnosis of disease; control (n = 37) and necrotising enterocolitis (NEC; n = 33). Colours indicate if the sample on that day of life was a maternal breast milk human milk oligosaccharide (HMO) or infants stool metagenome, or if both sample/data types were generated from each sample collected on that day.

Online supplementary figure 2. Comparison of human milk oligosaccharide (HMO) profiles by maternal secretor status. All infants were included (n = 77). **a,** Principal component analysis (PCA) showing the clustering of HMO profiles based on secretor status. **b**, Visual representation of P values for comparison of individual HMOs between secretor and non-secretor groups. 16 of the 19 HMOs were different between the two groups. Wilcoxon rank test was applied, and P values were adjusted using FDR algorithm. **c**, HMO Shannon diversity was higher in breast milk from secretor mothers compared to non-secretors. Wilcoxon rank test was applied. **d**, Stacked bar plot of HMOs concentrations describing HMO profile of each breast milk sample analysed. 2'FL used for identifying secretor status is almost absent in non-secretor breast milks, and present in relatively high concentration in samples from secretor mothers.

Online supplementary figure 3. Human milk oligosaccharide (HMO) profiles and disialyllacto-N-tetraose (DSLNT) are different between necrotising enterocolitis (NEC) and control group independent of maternal secretor status. Orthogonal partial least squares discriminant analysis (OPLS-DA) of breast milk HMO profiles from secretors (**a**) and non-secretors (**b**). P values were calculated performing 2000 permutations. Comparison of DSLNT concentration in milk received by NEC or controls separated by (**c**) secretor and (**d**) non secretor status. DSLNT concentration is lower in milk received by the NEC group independently of secretor status. Group comparison was performed applying Wilcoxon rank test and P values adjusted using FDR.

Online supplementary figure 4. Shannon diversity of human milk oligosaccharides (HMOs) was not associated with NEC development. Shannon diversity of (**a**) overall cohort, (**b**) secretor group, and (**c**) non-secretor group.

Online supplementary figure 5. Human milk oligosaccharide (HMO) profiles were predictive of necrotising enterocolitis (NEC) status. a, Receiver operating characteristic (ROC) curves generated using linear support vector machine (SVM) classification of HMO profiles between NEC and control groups. Increasing numbers of HMO were included in the model and performance was described by the AUC. Two HMOs model gave the optimal performance. **b**, Feature importance for the two HMOs model. Disialyllacto-N-tetraose (DSLNT) was selected as discriminatory feature in 100% of the permutations. **c**, Box plot showing the concentration of disialyllacto-N-tetraose (DSLNT) between NEC and controls from the validation dataset, Autran *et al.* (2018). Blue line represents the 241 nmol/ml threshold.

Online supplementary figure 6. Disialyllacto-N-tetraose (DSLNT) and lacto-N-neotetraose (LNnT) concentrations were not influenced by the day of life (DOL) of sample. Plot of (**a**) DSLNT and (**b**) LNnT concentrations in relation to DOL of the sample. P values were calculated by linear regression and DSLNT and LNnT concentrations were not related to the DOL the breast milk was fed to the infant.