

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

3

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

7

8 <sup>1</sup> Clinical and Translational Research Institute, Newcastle University, NE2 4HH. UK

9 <sup>2</sup> Newcastle Neonatal Service, Newcastle Hospitals NHS Trust, NE1 4LP. UK

10 <sup>3</sup> Population Health Sciences Institute, Newcastle University, NE2 4HH. UK

11 <sup>4</sup> Department of Gastroenterology, Newcastle upon Tyne Hospitals NHS Foundation Trust, NE1  
12 4LP. UK

13 <sup>5</sup> School of Health & Life Sciences, Northumbria University, NE1 8ST. UK

14 <sup>6</sup> Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of  
15 Research Excellence (MOMI CORE), University of California San Diego, La Jolla, CA, USA.

16 <sup>7</sup> Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology  
17 and Microbiology, Baylor College of Medicine, Houston, TX, USA

18

19 Janet Berrington, Lars Bode, and Christopher Stewart share joint correspondence in this work

20 Dr Janet Berrington: Neonatal unit, Ward 35, Leazes Wing, Royal Victoria Infirmary, Newcastle,  
21 NE1 4LP, Email [j.e.berrington@ncl.ac.uk](mailto:j.e.berrington@ncl.ac.uk)

22 Dr Lars Bode. Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant  
23 Center of Research Excellence (MOMI CORE), University of California San Diego, La Jolla, CA,  
24 USA. Email. [lbode@health.ucsd.edu](mailto:lbode@health.ucsd.edu)

25 Dr Christopher Stewart. Translational and Clinical Research Institute, Faculty of Medical Sciences,  
26 Newcastle University, 3rd Floor Leech Building, Newcastle, NE2 4HH. Email:

27 [christopher.stewart@newcastle.ac.uk](mailto:christopher.stewart@newcastle.ac.uk)

28

29 Word count: 4293

30

### 31 **List of abbreviations**

32 NEC - Necrotising enterocolitis

33 MOM - Mothers own milk, MOM

34 HMO - Human milk oligosaccharide

- 35 2'FL - 2'-fucosyllactose
- 36 3FL - 3-fucosyllactose
- 37 LNnT - Lacto-N-neotetraose
- 38 3'SL - 3'-sialyllactose
- 39 DFlac - Difucosyllactose
- 40 6'SL - 6'-sialyllactose
- 41 LNT - Lacto-N-tetraose
- 42 LNFP - Lacto-N-fucopentaose
- 43 DFLNT - Difucosyl-LNT
- 44 LNH - Lacto-N-hexaose
- 45 DSLNT - Disialyllacto-N-tetraose
- 46 FLNH - Fucosyl-lacto-N-hexaose
- 47 DFLNH - Difucosyl-lacto-N-hexaose
- 48 FDSLNH - Fucosyl-disialyl-lacto-N-hexaose
- 49 DSLNH - Disialyl-lacto-N-hexaose
- 50 PGCT - Preterm gut community types
- 51 DOL - Day of life
- 52 IQR - Interquartile range
- 53 ROC - Receiver operating characteristic
- 54 SVM - Support Vector Machine
- 55 MCCV - Monte-Carlo cross validation
- 56 PERMANOVA - Permutational multivariate analysis of variance
- 57 MDA - Mean decrease accuracy
- 58 DMM - Dirichlet multinomial modelling
- 59 PMA - Postmenstrual age
- 60 NICU - Neonatal intensive care unit

61 **Abstract**

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

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

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

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

82

83

84

85

86

87 **1. What is already known about this subject?**

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

93

94 **2. What are the new findings?**

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

103

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

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

108

109 **Introduction**

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

115 Receipt of mother's own breast milk (MOM) is the most protective factor against the development of  
116 NEC in preterm infants (3, 4). However, infants receiving MOM still develop NEC, suggesting the  
117 variable composition of nutrients and other components of breast milk may be important. Human  
118 milk oligosaccharides (HMOs) are structurally diverse, complex unconjugated sugars that are not  
119 usually present in artificial formula milk (5). HMOs are indigestible to the infant, reaching the lower  
120 gastrointestinal tract intact where they act as growth substrates (i.e. prebiotics) for specific bacteria,  
121 notably *Bifidobacterium* spp. thought key to infant health (6-8). HMOs may also protect from enteric  
122 organism blood stream infections due to anti-microbial activity (9), stimulate the immune system  
123 (10), enhance gut barrier function (11), and act as decoy receptors for pathogens (12). While >150  
124 HMOs have been described, the 19 most abundant represent >95% of the total HMO content (13).  
125 HMO profiles are specific to individual mothers and remain relatively stable during lactation (14).  
126 Presence of an active FUT2 gene, which is involved in the synthesis of  $\alpha$ 1-2-fucosylated  
127 oligosaccharides, is the main determinant of the HMO profile, termed maternal secretor status (15).  
128 Recent work has begun to elucidate the potential contribution of HMOs to preterm infant health. In a  
129 neonatal rat model, disialyllacto-N-tetraose (DSLNT), a non-fucosylated, but double-sialylated  
130 HMO, significantly reduced NEC development and improved NEC-associated mortality rate (16).  
131 An association of lower DSLNT concentration in MOM and subsequent higher risk of NEC onset in  
132 the infant has since been observed in preterm human studies (17-19). To date, these studies have  
133 included very small numbers of infants with NEC (between 4 and 8), with a broad range of NEC  
134 phenotypes. Thus, validation in a larger cohort is urgently needed.

135 Altered gut microbiome development has been associated with NEC in preterm infants. While no  
136 specific causative microorganism has consistently been identified, studies have reported a higher  
137 relative abundance of Enterobacteriaceae, coupled to lower relative abundance of *Bifidobacterium*  
138 (20-23). Instability of the gut microbiome in NEC infants has also been reported in longitudinal  
139 studies, with more frequent transitions between different preterm gut community types (PGCT) in  
140 NEC (20). These findings were replicated at the site of disease in a study using formalin fixed paraffin  
141 embedded tissue from NEC infants matched to non-NEC controls (24). Previous microbiome studies  
142 have largely relied on 16S rRNA gene sequencing of the V4 region, which has limited resolution,  
143 especially for emerging key organisms of interest for preterm health (i.e., *Klebsiella* and *Enterobacter*  
144 would be classified together as Enterobacteriaceae). Metagenomics may overcome this and recent  
145 metagenomic data showed infants who developed NEC had higher relative abundance of *Klebsiella*  
146 and higher replication rates in all bacteria before disease onset (25).

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

153

154 **Methods**

155 ***Cohort***

156 This study included 77 preterm infants (born at <32 weeks gestation), who were born in or transferred  
157 to a single large tertiary level neonatal intensive care unit (NICU) in Newcastle upon Tyne, UK  
158 recruited to the SERVIS study (REC10/H0908/39) with written parental consent covering data and  
159 sample collection. 33 infants were diagnosed with definite NEC and 37 non-diseased controls were  
160 selected by identifying a healthy infant matched by gestation, birthweight, and having a MOM sample  
161 available at a corresponding day of life (DOL) (**table 1**). Detailed information on feeding and  
162 antibiotic use are included in **online supplementary table 1 and online supplementary table 2**.  
163 Diagnoses were made using an extensive combination of clinical, x-ray and histological findings and  
164 blindly agreed by two neonatal clinicians (JEB and NDE). Standard clinical protocols recommended  
165 the routine use of supplemental probiotics when more than 30mL/kg/day of MOM was tolerated for  
166 at least 1-2 days: all 33 NEC infants received MOM and 31 received probiotics. All 37 controls  
167 received probiotics. The probiotics administered were either LaBiNIC (*Lactobacillus acidophilus*,  
168 *Bifidobacterium infantis* and *Bifidobacterium bifidum*) or Infloran (*L. acidophilus* and *B. bifidum*).  
169 The median DOL of NEC diagnosis was 19 (interquartile range; IQR 14-35; **table 1**). A single MOM  
170 sample was analysed for each infant, as close to the onset of disease as possible, with control samples  
171 matched by day of life (DOL) (**online supplementary figure 1**). The median DOL of MOM from  
172 NEC cases was 18 (IQR 13-34) and from controls was also 18 (IQR 12-31). The DOL of the milk  
173 sample is the DOL received by the infant, and is not necessarily the same day as the mother expressed  
174 the milk due to standard practice that often involves milk storage. Metagenomic sequencing of stool  
175 samples (n = 644) was performed longitudinally on a subset of 48 infants (including 14 NEC; **online**  
176 **supplementary figure 1**). These infants were comparable to the full cohort (**online supplementary**  
177 **table 3**).

178

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

181

### 182 *Human milk oligosaccharides analysis*

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

187

### 188 *Metagenomes*

189 DNA was extracted from ~0.1g of stool using the DNeasy PowerSoil Kit (QIAGEN) following the  
190 manufacturer's protocol and sequencing was performed on the HiSeq X Ten (Illumina) with a read  
191 length of 150bp paired end reads. Processed fastq files were mapped against the MetaPhlan2 marker  
192 gene database (mpa\_v20\_m200) (27).

193

### 194 *Statistical analysis*

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

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

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



205 variance (PERMANOVA) was used to determine significance of Bray-Curtis principal coordinate  
206 analysis. MetagenomeSeq was used to assess differential abundance at the phyla and species level.  
207 DMM clusters samples on the basis of microbial community structure (32) and was used to determine  
208 the preterm gut community types (PGCTs) from all samples, as performed previously (33, 34). Five  
209 PGCT was found to be optimal, and these were ordered youngest (PGCT-1) to oldest (PGCT-5) based  
210 on the average DOL of samples within each PGCT. Analysis was performed at specific time windows,  
211 including only a single sample per infant in each time point.  
212 The association of various clinical variables on the HMO and metagenome profiles was tested by  
213 applying the function “adonis” of “vegan” (version 2.5-6) package (35) in R, based on Bray-Curtis  
214 dissimilarity and 10000 permutations. Each test was performed stepwise and P values were adjusted  
215 using Benjamini & Hochberg (29).  
216 Random Forest was used for comparing the performance of classification models built using matched  
217 cross-sectional datasets.

218

## 219 **Results**

### 220 *Association of maternal HMOs and development of NEC in the infant*

221 MOM samples clustered according to maternal secretor status and secretor mothers had a higher total  
222 HMO concentration, a higher HMO Shannon diversity, and a significantly higher concentration of  
223 overall HMO-bound fucose (**online supplementary figure 2**). Thus, where relevant, we have  
224 stratified and adjusted for maternal secretor status in subsequent analyses.  
225 HMO profiles showed significant separation of NEC and control infants (**figure 1a**; 2000  
226 permutations,  $P < 0.001$ ) and this was consistent when secretor and non-secretor samples were  
227 analysed separately (both  $P < 0.001$ ; **online supplementary figure 3a and 3b**). Individually, of the  
228 19 HMOs quantified in this study, only DSLNT was significantly different between NEC and  
229 controls, with a lower concentration in NEC infants (adj.  $P < 0.001$ ; **figure 1b and 1c**). No significant  
230 associations were found in the Shannon diversity of HMOs between NEC and matched controls for

231 the full cohort, or when stratified by maternal secretor status (all  $P > 0.05$ ; **online supplementary**  
232 **figure 4**).

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

239 To test if integration of additional HMOs could improve the classification performance, multivariate  
240 ROC curves built on increasing number of HMOs were performed (**online supplementary figure**  
241 **5a**). Inclusion of 2 HMOs (the minimum in multivariate analysis) resulted in the optimal performance,  
242 with DSLNT being selected as a discriminatory feature in 100% of permutations (**online**  
243 **supplementary figure 5b**). 3FL and LNnT were the 2<sup>nd</sup> and 3<sup>rd</sup> most selected features, with a  
244 selection frequency of around 30%, being more abundant in cases of NEC. However, the integration  
245 of any additional HMOs to DSLNT in the multivariate model resulted in minimal improvement in  
246 performance compared to the univariate model using DSLNT only (AUC of 0.949 and 0.946,  
247 respectively).

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

255

256 *Analysis of HMO profiles stratified by NEC type*

257 We compared medically managed NEC (NEC-M), where infants did not undergo surgery or die from  
258 NEC (i.e. had less severe disease), with NEC infants that underwent surgery (NEC-S). NEC-M and  
259 NEC-S clustered together and were distinct from matched controls (**figure 2a**; 2000 permutations,  $P$   
260  $< 0.001$ ). Two HMOs were found to be significantly different, with DSLNT lower in MOM in both  
261 NEC-M (adj.  $P < 0.001$ ) and NEC-S (adj.  $P < 0.001$ ) compared to controls (**figure 2b**). In addition,  
262 LNnT in MOM was significantly lower in NEC-S in comparison to both NEC-M (adj.  $P = 0.0016$ )  
263 and matched controls (adj.  $P = 0.0423$ ) (**figure 2c**).

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

273

#### 274 *Association of infant gut microbiome and development of NEC*

275 We included stool microbiome data on a subset of infants with HMO data, where metagenomic  
276 sequencing data was available through an on-going independent study (the results of which are not  
277 yet published). This included 644 stool samples from 34 controls, and 14 NEC infants (**online**  
278 **supplementary table 3**). To overcome challenges of repeated measures and to compare results with  
279 existing published work, we first analysed one stool sample per infant closest to NEC onset (median  
280 of 3 days before NEC) and a corresponding control sample matched by DOL (**online supplementary**  
281 **figure 1**). This cross-sectional analysis showed NEC infants had significantly lower richness ( $P =$   
282  $0.027$ ) but comparable Shannon diversity ( $P = 0.443$ ; **figure 3a**). Bray-Curtis PCoA showed no

283 significant difference between the bacterial profiles of NEC and controls (PERMANOVA  $P = 0.182$ ;  
284 **figure 3b**). Analysis at the phylum level showed significantly lower relative abundance of  
285 Actinobacteria (adj.  $P = 0.034$ ) and higher relative abundance of Proteobacteria (adj.  $P = 0.034$ ) in  
286 NEC infants (**figure 3c**). Correspondingly, at the species level, NEC infants had lower relative  
287 abundance of *Bifidobacterium longum* (adj.  $P = 0.012$ ) and higher relative abundance of *Enterobacter*  
288 *cloacae* (adj.  $P = 0.012$ ), compared to controls (**figure 3d**).

289

#### 290 *Integrated analysis of HMO and bacterial profiles*

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

301 The PGCTs were named according to the average age of samples within that cluster, where PGCT-1  
302 contained on average the earliest samples and PGCT-5 on average the latest samples. We compared  
303 the number of samples from all time points in only PGCT-1 and PGCT-5 to investigate associations  
304 between the MOM DSLNT threshold and gut microbiome development from the typically younger  
305 to the typically older PGCTs. Infants receiving MOM with DSLNT level below 241 nmol/mL had  
306 significantly more samples remaining within PGCT-1 throughout all time points (78% in PCGT-1 vs.  
307 22% in PGCT-5,  $\chi^2$  test  $P < 0.001$ ), whereas infants receiving MOM with DSLNT above this threshold  
308 transitioned from PGCT-1 to PGCT-5 as demonstrated by a similar number of samples in each PGCT

309 across all time points (48% in PCGT-1 vs. 52% in PGCT-5,  $\chi^2$  test  $P = 0.717$ ). In addition to  
310 comparing samples from all times points, we next compared samples from the final time point only  
311 (i.e., DOL 50-60). After correcting for uneven frequency of sampling between groups, at the final  
312 time point infants receiving MOM above the DSLNT threshold were twice as likely to be in PGCT-  
313 5 (3/11 samples below vs. 12/22 samples above DSLNT threshold; odds ratio 3.20, 95% CI  
314 0.6657 to 15.3819), which was characterised by high relative abundance of *Bifidobacterium* (**figure**  
315 **4a**).

316

### 317 *Explained variance and random forest classification of HMO and metagenome data*

318 Using the cross-sectional HMO and cross-sectional metagenome dataset, we sought to determine  
319 which clinical factors were most associated with the HMO and the bacterial profiles (**figure 5a**).  
320 Secretor status explained 56% of the variance within HMO profiles (adj.  $P < 0.001$ ), but no other co-  
321 variate was significantly associated with the HMO profiles. In contrast, the bacterial profiles were  
322 significantly associated with both postmenstrual age ( $R^2$  0.07; adj.  $P = 0.006$ ) and day of life ( $R^2$   
323 0.07; adj.  $P = 0.006$ ), as well as receipt of antibiotics at the time of sampling ( $R^2$  0.06; adj.  $P = 0.006$ )  
324 and receipt of probiotics ( $R^2$  0.12; adj.  $P = 0.006$ ), but not maternal secretor status ( $R^2$  0.02; adj.  $P =$   
325 0.58). Together, these findings highlight that HMO and bacterial profiles are influenced by numerous  
326 non-overlapping factors related to early life in preterm infants.

327 We compared the performance of random forest classification models built on the cross-sectional  
328 subset of HMO profile data, metagenomic sequencing data, and the two datasets combined to classify  
329 an infant as NEC or healthy, given that all this information is available before onset of disease and  
330 could therefore function as a risk stratification system in clinical practice. The HMO profile alone  
331 had a classification error of 0.146, with 21% (3/14) NEC and 12% (4/34) control infants misclassified.  
332 DSLNT had the greatest contribution to classification with a Mean Decrease Accuracy (MDA) of  
333 0.11. Other HMOs contributing to classification accuracy included LNH (MDA = 0.012) and DFLNH  
334 (MDA = 0.011), which were non-significantly higher in NEC infants. Random forest generated using

335 the metagenomic sequencing data was characterised by a classification error of 0.229, with 43%  
336 (6/14) NEC and 15% (5/34) control infants misclassified. *Enterobacter cloacae* was the most  
337 important feature guiding the classification (MDA = 0.036), with higher relative abundance in NEC  
338 infants, followed by *Bifidobacterium bifidum* (MDA = 0.024) and *Bifidobacterium longum* (MDA =  
339 0.013) which had higher relative abundance in control infants. Combining HMO and metagenome  
340 datasets slightly improved the performance compared to using HMOs alone, with 21% (3/14) NEC  
341 infants and (9%) 3/34 controls misclassified. In this combined model, DSLNT was enriched in  
342 controls and DSLNH and the relative abundance of *Escherichia unclassified* were higher in NEC  
343 infants (**figure 5b**).

344

## 345 **Discussion**

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

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

361 findings further extend the evidence for the specificity of DSLNT in the NEC pathway. A threshold  
362 level of DSLNT (241 nmol/mL) from a single MOM sample correctly identified 91% of NEC infants  
363 (below threshold) and 86% of control healthy infants (above threshold). Of the three infants who  
364 developed NEC despite a DSLNT above the threshold, two had not received MOM in the 3 weeks  
365 prior to disease onset and the remaining infant had a DSLNT concentration of 248 nmol/ml.

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

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

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

394 Random forest analysis confirmed the capability of HMO profiles to identify infants who developed  
395 NEC and slightly outperformed metagenome profiles by correctly classifying three more NEC cases  
396 and one more control. Combining HMO and metagenome data before disease accurately classified  
397 87.5% of infants as healthy or NEC, with DSLNT and the bacterial species identified as important in  
398 the random forest analysis being comparable to the unsupervised analysis in the current study and in  
399 previous studies. Further work is needed to determine if DSLNT functions via modulation of the  
400 microbiome or by acting directly on the host, such as acting in a structure-specific receptor-mediated  
401 way to alter immune functioning and reduce inflammation leading to necrosis. In the event of the  
402 latter, a microbial community with less DSLNT utilisation could provide an advantage to reducing  
403 NEC risk. Taken together, the current findings and recent work highlighting the ability of  
404 *Bifidobacterium* spp. to utilise HMOs is strain specific (7, 8) underscore the need for further research  
405 to better understand the complexity of human milk and other nutritional exposures, including the use  
406 of supplements such as prebiotics and probiotics in preterm infants. In addition to therapeutics, the  
407 classifiers may provide a basis for the development of biomarkers predicting NEC risk. While  
408 additional work is needed, the addition of microbial biomarkers may allow for the most accurate  
409 predictions and could inform NEC risk for infants where MOM (and thus HMO information) is not  
410 available.

411 This study involved the largest cohort to date investigating the relationship between HMO  
412 composition and NEC development, and includes one of the most extensive longitudinal stool



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

430 In summary, HMO profiling of MOM coupled to metagenomic sequencing of preterm stool showed  
431 that the concentration of a single HMO, DSLNT, was lower in milk received by infants who  
432 developed NEC. The lower concentration of DSLNT was associated with altered microbiome  
433 development, specifically a reduced progression toward the PGCT typically found in the older infants  
434 which was abundant in *Bifidobacterium* spp. These results suggest MOM HMO profiling may provide  
435 potential targets for biomarker development and disease risk stratification. They may also guide  
436 focussed donor milk use (e.g., prioritise high DSLNT for preterm infants) and novel avenues for  
437 supplements that may prevent life-threatening disease.

438

439 **References**

- 440 1. A. C. Masi, C. J. Stewart, The role of the preterm intestinal microbiome in sepsis and necrotising  
 441 enterocolitis. *Early Hum Dev* **138**, 104854 (2019).
- 442 2. J. Neu, W. A. Walker, Necrotizing enterocolitis. *N Engl J Med* **364**, 255-264 (2011).
- 443 3. J. Meinen-Derr *et al.*, Role of human milk in extremely low birth weight infants' risk of necrotizing  
 444 enterocolitis or death. *J Perinatol* **29**, 57-62 (2009).
- 445 4. S. Sullivan *et al.*, An exclusively human milk-based diet is associated with a lower rate of necrotizing  
 446 enterocolitis than a diet of human milk and bovine milk-based products. *J Pediatr* **156**, 562-567 e561  
 447 (2010).
- 448 5. L. Bode, Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147-1162  
 449 (2012).
- 450 6. T. Vatanen *et al.*, The human gut microbiome in early-onset type 1 diabetes from the TEDDY study.  
 451 *Nature* **562**, 589-594 (2018).
- 452 7. M. A. E. Lawson *et al.*, Breast milk-derived human milk oligosaccharides promote Bifidobacterium  
 453 interactions within a single ecosystem. *ISME J*, (2019).
- 454 8. A. Gotoh *et al.*, Sharing of human milk oligosaccharides degradants within bifidobacterial  
 455 communities in faecal cultures supplemented with Bifidobacterium bifidum. *Sci Rep* **8**, 13958 (2018).
- 456 9. D. L. Ackerman *et al.*, Human Milk Oligosaccharides Exhibit Antimicrobial and Antibiofilm Properties  
 457 against Group B Streptococcus. *ACS Infect Dis* **3**, 595-605 (2017).
- 458 10. L. Bode *et al.*, Inhibition of monocyte, lymphocyte, and neutrophil adhesion to endothelial cells by  
 459 human milk oligosaccharides. *Thromb Haemost* **92**, 1402-1410 (2004).
- 460 11. R. Y. Wu *et al.*, Human Milk Oligosaccharides Increase Mucin Expression in Experimental Necrotizing  
 461 Enterocolitis. *Mol Nutr Food Res* **63**, e1800658 (2019).
- 462 12. A. Gustafsson *et al.*, Carbohydrate-dependent inhibition of Helicobacter pylori colonization using  
 463 porcine milk. *Glycobiology* **16**, 1-10 (2006).
- 464 13. S. Ramani *et al.*, Human milk oligosaccharides, milk microbiome and infant gut microbiome modulate  
 465 neonatal rotavirus infection. *Nat Commun* **9**, 5010 (2018).
- 466 14. S. Austin *et al.*, Human Milk Oligosaccharides in the Milk of Mothers Delivering Term versus Preterm  
 467 Infants. *Nutrients* **11**, (2019).
- 468 15. M. B. Azad *et al.*, Human Milk Oligosaccharide Concentrations Are Associated with Multiple Fixed  
 469 and Modifiable Maternal Characteristics, Environmental Factors, and Feeding Practices. *J Nutr* **148**,  
 470 1733-1742 (2018).
- 471 16. E. Jantscher-Krenn *et al.*, The human milk oligosaccharide disialyllacto-N-tetraose prevents  
 472 necrotising enterocolitis in neonatal rats. *Gut* **61**, 1417-1425 (2012).
- 473 17. C. A. Autran *et al.*, Human milk oligosaccharide composition predicts risk of necrotising enterocolitis  
 474 in preterm infants. *Gut* **67**, 1064-1070 (2018).
- 475 18. E. Van Niekerk *et al.*, Human milk oligosaccharides differ between HIV-infected and HIV-uninfected  
 476 mothers and are related to necrotizing enterocolitis incidence in their preterm very-low-birth-weight  
 477 infants. *J Nutr* **144**, 1227-1233 (2014).
- 478 19. D. Hassinger, D. M. Clausen, S. Nitka, A. Herdt, I. Griffin, Analysis of Disialyllacto-N-Tetraose (DSLNT)  
 479 Content in Milk From Mothers of Preterm Infants. *J Hum Lact*, 890334420904041 (2020).
- 480 20. C. J. Stewart *et al.*, Temporal bacterial and metabolic development of the preterm gut reveals specific  
 481 signatures in health and disease. *Microbiome* **4**, 67 (2016).
- 482 21. K. Sim *et al.*, Dysbiosis anticipating necrotizing enterocolitis in very premature infants. *Clin Infect Dis*  
 483 **60**, 389-397 (2015).
- 484 22. D. V. Ward *et al.*, Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic  
 485 E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell Rep* **14**, 2912-2924 (2016).
- 486 23. A. L. Morrow *et al.*, Early microbial and metabolomic signatures predict later onset of necrotizing  
 487 enterocolitis in preterm infants. *Microbiome* **1**, 13 (2013).
- 488 24. C. J. Stewart *et al.*, Using formalin fixed paraffin embedded tissue to characterize the preterm gut  
 489 microbiota in necrotising enterocolitis and spontaneous isolated perforation using marginal and  
 490 diseased tissue. *BMC Microbiol* **19**, 52 (2019).

- 491 25. M. R. Olm *et al.*, Necrotizing enterocolitis is preceded by increased gut bacterial replication,  
492 Klebsiella, and fimbriae-encoding bacteria. *Sci Adv* **5**, eaax5727 (2019).
- 493 26. L. Bode *et al.*, Human milk oligosaccharide concentration and risk of postnatal transmission of HIV  
494 through breastfeeding. *Am J Clin Nutr* **96**, 831-839 (2012).
- 495 27. D. T. Truong *et al.*, MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12**,  
496 902-903 (2015).
- 497 28. J. Xia, I. V. Sinelnikov, B. Han, D. S. Wishart, MetaboAnalyst 3.0--making metabolomics more  
498 meaningful. *Nucleic Acids Res* **43**, W251-257 (2015).
- 499 29. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach  
500 to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289-300  
501 (1995).
- 502 30. J. Chong, P. Liu, G. Zhou, J. Xia, Using MicrobiomeAnalyst for comprehensive statistical, functional,  
503 and meta-analysis of microbiome data. *Nat Protoc* **15**, 799-821 (2020).
- 504 31. A. Dhariwal *et al.*, MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and  
505 meta-analysis of microbiome data. *Nucleic Acids Res* **45**, W180-W188 (2017).
- 506 32. I. Holmes, K. Harris, C. Quince, Dirichlet multinomial mixtures: generative models for microbial  
507 metagenomics. *PLoS One* **7**, e30126 (2012).
- 508 33. C. J. Stewart *et al.*, Cesarean or Vaginal Birth Does Not Impact the Longitudinal Development of the  
509 Gut Microbiome in a Cohort of Exclusively Preterm Infants. *Front Microbiol* **8**, 1008 (2017).
- 510 34. C. J. Stewart *et al.*, Temporal development of the gut microbiome in early childhood from the TEDDY  
511 study. *Nature* **562**, 583-588 (2018).
- 512 35. F. G. B. Jari Oksanen, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlenn, Peter R.  
513 Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene  
514 Wagner, vegan: Community Ecology Package. R package version 2.5-6. . [https://CRAN.R-](https://CRAN.R-project.org/package=vegan)  
515 [project.org/package=vegan](https://CRAN.R-project.org/package=vegan), (2019).
- 516 36. M. Good *et al.*, The human milk oligosaccharide 2'-fucosyllactose attenuates the severity of  
517 experimental necrotising enterocolitis by enhancing mesenteric perfusion in the neonatal intestine.  
518 *Br J Nutr* **116**, 1175-1187 (2016).
- 519 37. C. A. Autran, M. H. Schoterman, E. Jantscher-Krenn, J. P. Kamerling, L. Bode, Sialylated galacto-  
520 oligosaccharides and 2'-fucosyllactose reduce necrotising enterocolitis in neonatal rats. *Br J Nutr* **116**,  
521 294-299 (2016).
- 522 38. S. O. Rasmussen *et al.*, Human milk oligosaccharide effects on intestinal function and inflammation  
523 after preterm birth in pigs. *J Nutr Biochem* **40**, 141-154 (2017).
- 524 39. M. S. Ciliborg *et al.*, Minimal short-term effect of dietary 2'-fucosyllactose on bacterial colonisation,  
525 intestinal function and necrotising enterocolitis in preterm pigs. *Br J Nutr* **116**, 834-841 (2016).
- 526 40. M. K. McGuire *et al.*, What's normal? Oligosaccharide concentrations and profiles in milk produced  
527 by healthy women vary geographically. *Am J Clin Nutr* **105**, 1086-1100 (2017).
- 528 41. C. J. Stewart *et al.*, Longitudinal development of the gut microbiome and metabolome in preterm  
529 neonates with late onset sepsis and healthy controls. *Microbiome* **5**, 75 (2017).
- 530 42. R. M. Torrazza *et al.*, Intestinal microbial ecology and environmental factors affecting necrotizing  
531 enterocolitis. *PLoS One* **8**, e83304 (2013).
- 532 43. V. Mai *et al.*, Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One* **6**,  
533 e20647 (2011).

534

535

536 **Acknowledgements**

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

539

540 **Contributions**

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

546

547 **Funding**

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

552

553 **Competing interest**

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

562

563 **Ethical approval**

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

566

567 **Data availability statement**

568 Data are available upon reasonable request. All sequencing data generated and analysed in this  
569 study have been deposited in the European Nucleotide Archive (ENA) under study accession  
570 number PRJEB39610.  
571

572 **Tables**

573 **Table 1. Demographics of the analytical cohort with human milk oligosaccharide profile data.**

574 Differences between groups were tested applying Chi-square test and Dunn’s post-hoc test where  
575 applicable.

	<b>Control</b>	<b>NEC</b>	<b>P value</b>
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%)	
MOM + Formula	11 (30%)	12 (37%)	0.468
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	-

576 NEC, necrotising enterocolitis; MOM, mother’s own breast milk; BMF, breast milk fortifier;

577 DOL, day of life

578

579

580

581 **Figure Legends**

582

583 **Figure 1. Analysis of human milk oligosaccharide (HMO) profiles and DSLNT concentration**  
584 **in necrotising enterocolitis (NEC) and controls. a,** Orthogonal partial least squares discriminant  
585 analysis (OPLS-DA) of maternal HMO profiles fed to infants diagnosed with NEC and controls. The  
586 P value was calculated based on 2000 permutations. **b,** Visual representation of P values obtained  
587 from comparison of individual HMOs between NEC and control group. Wilcoxon rank sum test was  
588 applied, and P values adjusted with FDR algorithm. The line indicates P value = 0.05. **c,** Univariate  
589 receiver operating characteristic (ROC) curve generated on DSLNT concentration identified 241

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

594

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

605

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

612

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

623

624 **Figure 5. Modelling of cross-sectional human milk oligosaccharide (HMO) and infant stool**  
625 **metagenomic profiles using Adonis and random forest. a,** horizontal bar plots showing the  
626 variance ( $r^2$ ) in maternal HMO and infant stool metagenomic profiles explained by clinical covariates  
627 as modelled by univariate Adonis. Variables with an FDR  $P < 0.05$  are shown in red. DOL, day of  
628 life; PMA, post-menstrual age. **b,** Feature importance from combined HMO and metagenome random  
629 forest classification model. Mean decrease accuracy (MDA) value defines the contribution given by  
630 a certain feature to classification process.  
631



632 **Online supplementary figure Legends**

633

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

639

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

650

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

659

660

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

664

665

666 **Online supplementary figure 5. Human milk oligosaccharide (HMO) profiles were predictive**  
667 **of necrotising enterocolitis (NEC) status. a,** Receiver operating characteristic (ROC) curves  
668 generated using linear support vector machine (SVM) classification of HMO profiles between NEC  
669 and control groups. Increasing numbers of HMO were included in the model and performance was  
670 described by the AUC. Two HMOs model gave the optimal performance. **b,** Feature importance for  
671 the two HMOs model. Disialyllacto-N-tetraose (DSLNT) was selected as discriminatory feature in  
672 100% of the permutations. **c,** Box plot showing the concentration of disialyllacto-N-tetraose  
673 (DSLNT) between NEC and controls from the validation dataset, Autran *et al.* (2018). Blue line  
674 represents the 241 nmol/ml threshold.

675

676

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

682