The landscape of antibiotic resistance genes in the nasopharynx of preterm infants: Prolonged signature of hospitalization and effects by antibiotics

Achal Dhariwal[1]†, Polona Rajar[1,2]†, Gabriela Salvadori[1], Heidi Aarø Åmdal[1], Dag Berild[3,4], Ola Didrik Saugstad[5], Drude Fugelseth[2,4], Gorm Greisen[6], Ulf Dahle[7], Kirsti Haaland[2], Fernanda Cristina Petersen[1]*

[1] Institute of Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway.
[2] Department of Neonatal Intensive Care, Division of Pediatric and Adolescent Medicine, Oslo University Hospital Ullevål, Oslo, Norway.
[3] Department of Infectious Diseases, Oslo University Hospital, Oslo, Norway.
[4] Institute of Clinical Medicine, Faculty of Medicine, Oslo University, Oslo, Norway.
[5] Department of Pediatric Research, Oslo University Hospital, University of Oslo, Oslo, Norway.
[6] Department of Neonatology, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark.
[7] Centre for Antimicrobial Resistance, Norwegian Institute of Public Health, Oslo, Norway

†Shared First authorship
*Correspondence: Fernanda Cristina Petersen
Institute of Oral Biology
Faculty of Dentistry,
University of Oslo, Oslo, Norway.
Tel: +47-22840312
E-mail: f.c.petersen@odont.uio.no

ABSTRACT

Respiratory pathogens, commonly colonizing nasopharynx, are among the leading causes of death due to antimicrobial resistance. Yet, antibiotic resistance determinants within nasopharyngeal microbial communities remain poorly understood. Utilizing shotgun metagenomics, we investigated the nasopharynx resistome development in preterm infants, assessed early antibiotic impact on its trajectory, and explored its association with clinical covariates. Our findings revealed widespread nasopharyngeal carriage of antibiotic resistance genes (ARGs) with resistomes undergoing transient changes, including increased ARG diversity, abundance, and composition alterations due to early antibiotic exposure. ARGs associated with the critical nosocomial pathogen *Serratia marcescens* persisted (at 8-10 months of age) for months following discharge, signifying hospitalization signature. The nasopharyngeal resistome strongly correlated with microbiome composition, with inter-individual differences and postnatal age explaining most of the variation. The collateral effects of antibiotics and hospitalization reported here highlights the importance of future studies focusing on this yet rather unexplored reservoir of pathogens and ARGs.

Keywords: respiratory resistome, metagenomics, nasopharynx, microbiome, preterm infants, antibiotics, antimicrobial resistance

INTRODUCTION

Antimicrobial resistant infections are one of the biggest global health threats. Among these, respiratory infections account for more than 1.5 million deaths a year, surpassing the number of deaths by resistant infections in any other human body niches [1]. Main pathogens include *Streptococcus pneumoniae, Staphylococcus aureus, Klebsiella pneumoniae, Enterococcus faecalis* and *Haemophilus influenzae*, all generally present in the nasopharynx of healthy populations. These pathogens share common niches with commensals, which in turn provide colonization resistance against pathogens. However, commensals can also serve as reservoirs of antibiotic resistance genes (ARGs) that can readily be transferred to pathogens [2-4].

Recent studies have highlighted the potential negative impact of antibiotic treatment on the nasopharynx microbiome and its association with an increased risk of developing asthma and respiratory infections in infants [5-7]. These studies have significantly advanced our understanding of the ecological impact of antibiotics on respiratory health and disease. However, as these studies rely on 16S rRNA sequencing technologies, they do not offer a high level of microbial taxonomic resolution and lack information on the overall gene composition, including the presence of ARGs within microbial communities (referred to as the resistome). In our previous study, we developed an optimization strategy for DNA extraction and library preparation that helped to circumvent the challenges posed by low biomass and high host DNA content in nasopharynx samples of preterm infants [8]. In this study, we applied the optimized protocol to nasopharyngeal aspirate samples to characterize the development of the resistome in 36 preterm infants from birth up to 6 months corrected age (8-10 months of chronological age). By analyzing ~1.85 terabytes (Tb) of metagenomic DNA data generated from deep shotgun sequencing, we obtained the most comprehensive information to date on the characteristics and dynamics of resistome development in the respiratory tract of preterm infants. Our hypothesis was that early exposure to broad-spectrum antibiotics for treatment of suspected early onset sepsis (sEONS) in preterm infants may promote prolonged compositional alterations in the respiratory resistome. We also explored the extensive metadata collected to gain insights into additional clinical covariates that may be associated with resistome development.

METHODS

90

91 **Study design and cohort characteristics**

92 The study is a prospective observational cohort study. Between July 2019 and January 2021, we
93 included preterm infants born (or transferred within 48 hours) at the NICU (Oslo University
94 Hospital, Ullevål) who were between 28+0 and 31+6 weeks of gestational age (GA). In total, 66
95 preterm infants were enrolled in the study, and followed from birth until six months corrected age
96 (i.e., adjusted for prematurity) during their stay in the NICU and after discharge. Five of the infants
97 were lost to follow-up. Extensive metadata and covariates from infants and mothers, including
98 mode of delivery, gestational age (GA) at birth, postmenstrual age (PMA), gender, birth weight
99 (BW), postnatal infant antibiotics usage (duration, type, days), maternal antibiotic usage, feeding
100 regimen during hospitalization and after discharge, and other demographic characteristics, were
101 collected through medical records. This data was stored in secured server at the Services for
102 Sensitive Data (TSD), University of Oslo.

103 In our cohort of 66 infants, 33 infants with sEONS received broad-spectrum antibiotics, consisting
104 of a combination of ampicillin (a beta-lactam antibiotic) and gentamicin (an aminoglycoside
105 antibiotic), within 24 hours after birth. Due to the circumstances of premature birth and the
106 instability of the infants, baseline samples could not be collected before antibiotic therapy
107 initiation. Instead, the first nasopharyngeal aspirate samples (baseline) were obtained within 48
108 hours after the initiation of early antibiotic therapy. The average early antibiotic treatment duration
109 was 5 days (SD: 2 days). The sampling time point T2 corresponds to the first available sample
110 collected (except in one infant) after the cessation of early antibiotic treatment. In addition to early
111 antibiotics, one of the included infants received subsequent antibiotics initiated > 72 hours after
112 birth and before discharge from NICU and two other infants received antibiotic therapy between
113 discharge from NICU and 6 months corrected age. A timeline of sample collection and antibiotic
114 administration is represented in Supplementary Fig. 1.

115 To investigate temporal changes in the resistome in response to antibiotics and control for potential
116 baseline differences as confounding factors, this study included only infants who had
117 nasopharyngeal samples collected on the same day as the initiation of early antibiotic treatment (n
118 $= 15$) and antibiotic-naive infants as controls ($n = 21$). Infants exposed to early antibiotics had
119 significantly lower GA and BW, and their mothers had a more prolonged period of ruptured

membranes before delivery compared to antibiotic-naive infants. None of the mothers of naive infants received antibiotics during pregnancy (prenatal antibiotics) before onset of delivery. The baseline characteristics of the infants stratified by antibiotic exposure are presented in Table 1.

Sample collection

Nasopharyngeal aspirate samples were collected from infants at six distinct time points spanning from birth to 6 months corrected age (T1-T6). The days of life (DOL) at which samples were collected for each time point varied as follows: T1 (DOL: 0-3), T2 (DOL: 4-9), T3 (DOL: 12-19), T4 (DOL: 20-31), T5 (DOL: 32-76) and T6 at 6 months corrected age (DOL: 229-288). Additional samples were taken if the antibiotic therapy was initiated or terminated more than 48 hours before/after a predefined sampling time. As a result, in two infants, there was more than one sample available for the defined time points. To facilitate time point-wise analysis, only one sample was included for each of the time points (Supplementary Table 1 & Supplementary Fig. 1). In cases where the sample could not be collected at the scheduled time points (such as with an unstable infant), a new sample was obtained as soon as possible. Infants who were transferred to another hospital before discharged to home adhered to the study protocol. When the infants reached six months of corrected age, they were either invited to an outpatient visit at Ullevål or their local hospital or visited at home for sample collection. Nasopharynx aspirates were collected as previously described [8] and stored at -80°C until further processing. In total, we collected and extracted metagenomic DNA from 198 nasopharyngeal samples obtained from 36 preterm infants, from day of birth until 6 months corrected age. Detailed information on sample exclusion and inclusion statistics at different steps from sampling to whole metagenomic sequencing (WMS) is described in Supplementary Table 1.

Metagenomic DNA extraction and sequencing

Metagenomic DNA was extracted by first thawing the aspirate samples (2 ml) on ice, pelleting them by centrifugation at 10,000 g for 10 minutes at 4°C, and then using MolYsis™ for host DNA depletion (Molzym, Bremen, Germany), following the manufacturer's protocol with some modifications, as previously described [8]. After depletion of human DNA, fresh pellets were spiked with 20 µl of ZymoBIOMICS Spike-in Control II (catalog number: D6321 & D6321-10),

148 and bacterial DNA was extracted using the MasterPure™ Gram Positive DNA Purification Kit
149 (Epicentre, Madison, WI, USA), according to the manufacturer's recommendations. The bacterial
150 DNA was eluted in 35 µl of TE buffer and stored at -80°C until further use. Final DNA
151 concentration was quantified using Qubit dsDNA HS assay kits in a Qubit 4.0 Fluorometer
152 (Invitrogen, Thermo Fisher Scientific, USA). Metagenomic sequencing libraries were constructed
153 using the Nextera DNA Flex library preparation kit (Illumina Inc., CA, USA). All samples were
154 subjected to paired-end sequencing at 2 × 150 bp on an Illumina NovaSeq S4 high-output platform
155 at the Norwegian Sequencing Centre (Oslo, Norway). In total, 6.14 billion raw sequencing reads
156 were obtained with a mean value of sample reads counts of 33.97 million (M) ranging from 4.19
157 to 69.25 M reads (Supplementary Table 2).
158
159 **Controls for low microbial biomass**
160 The study followed recent recommendations for low microbial biomass microbiome research [9].
161 This involved the following steps: (1) collecting samples from the included infants in the two
162 groups within the same time period, (2) using standard protection measures during sample
163 collection (such as clean suits, disposable gloves, and face masks), (3) including sampling and
164 DNA extraction blank controls (negative controls), (4) conducting DNA extraction and library
165 preparation while wearing laboratory coats and disposable gloves in a carefully cleaned
166 environment , and (5) using unique barcodes in library preparation. The negative controls had no
167 or extremely low DNA concentrations to proceed to library preparation as determined by RT-PCR
168 using the Femto Bacterial DNA Quantification Kit from Zymo and were therefore excluded from
169 further analysis ($n = 10$; mean = 0.00043 ng per µL). Additionally, spike-in controls ($n = 6$) were
170 extracted and used as positive controls, and longitudinal samples from each preterm infant were
171 extracted within the same run.
172
173 **Bioinformatics processing**
174 The raw metagenomic sequencing data was pre-processed using our in-house bioinformatics
175 pipelines using a high-performance computing cluster inside a secured environment, i.e., TSD at
176 the University of Oslo. In brief, the Nextera adaptor sequences and low-quality reads were filtered
177 using Trim Galore (v.0.6.1) with default parameters [10]. Next, the human DNA contaminant
178 sequences were identified and removed by mapping the quality-trimmed reads against the human

179 reference genome (GRCh38) using Bowtie2 (v.2.3.4.2) [11] (non-default parameters: *q -N 1 -k 1*
180 *--fr --end-to-end --phred33 --very-sensitive –no-discordant*) along with SAMtools (v.1.9) [12] and
181 BEDTools (v.2.27.1) [13]. On average, 46.71% (range: 1.09 - 97.69%) of total quality-filtered
182 reads were identified as belonging to the human genome. After removal of the human-associated
183 metagenomic reads, a total of 2.98 billion remaining high-quality clean reads across all samples
184 with number of reads ranging from 0.76 to 59.9 M per sample, and with median of 16.4 M reads
185 were subjected to resistome profiling. More detailed results regarding the overall sequence
186 preprocessing are presented in Supplementary Table 2. The sequence quality reports on raw and
187 processed reads were generated using FastQC (v.0.11.9) [14] and summarized using MultiQC
188 (v.1.7) [15].
189
190 **Resistome profiling**

191 To characterize the presence of ARGs, cleaned high-quality reads for each sample were mapped
192 against the *nucleotide_fasta_protein_homolog_model* from the Comprehensive Antibiotic
193 Resistance Database (CARD) (v.3.2.2) [16] using Bowtie2 under parameter *–very-sensitive-local*.
194 The mapped reads from each sample were then filtered, sorted, and indexed using SAMtools. The
195 number of reads mapped to each ARG was calculated using SAMtools *idxstats* and BEDTools.
196 ARGs with a coverage of at least 80% from the alignment were retained for further downstream
197 analyses. The mapped read counts were normalized for bacterial sequence abundances and gene
198 lengths by calculating reads per kilobase of reference gene per million bacterial reads (RPKM) for
199 each CARD reference sequence. The relative abundance of ARGs for each sample was computed
200 by dividing the RPKM by the total sum of the RPKM for each sample. For keeping the acyclic
201 hierarchical annotation structure for accurate resistome profiling at higher functional levels, ARGs
202 were manually re-annotated based on the drug class to which they confer resistance. ARGs
203 belonging to macrolides, lincosamides and streptogramins were congregated into the MLS class.
204 While ARGs belonging to carbapenem, cephamycin, cephalosporin, penem, penam and
205 monobactam were congregated into the Beta-lactam class.

206 Rarefaction analysis was performed using Rarefaction Analyzer [17] to assess whether the samples
207 had sufficient sequencing depth for the comprehensive characterization of the nasopharyngeal
208 resistome composition in preterm infants (Supplementary Fig. 3). This analysis involved the

random subsampling of sequencing reads at proportions varying from 0% to 100% of the total sequencing depth. The subsampled metagenomes were not utilized for any further downstream analysis. Ten samples were excluded from downstream analysis as no reads were assigned to ARGs. Collectively, we assessed the resistome of the nasopharyngeal microbiome in a total of 181 samples collected from 36 preterm infants.

Statistical analysis and data visualization

All statistical analyses were conducted in R (v.4.2.1) within RStudio (v.2022.07.2+576) [18, 19]. The ARG abundance, annotation table, and metadata file were compiled into a single data object using the phyloseq (v.1.40.0) package [20]. Figures, unless stated otherwise, were created using the ggplot2 (v.3.4.0) package [21] and further edited using Adobe Illustrator (v.16.0.0). α- and β-diversity analysis was performed using the vegan (v.2.6.2) [22] and phyloseq R packages. α-diversity was calculated using the Shannon diversity metric, and differences over time points in groups were assessed using a linear mixed-effects (LME) model with infants set as a random effect while correcting for age. The overall resistome composition of nasopharyngeal samples was visualized using principal component analysis (PCA) ordination plots. The ordinates were based on the Euclidean dissimilarity matrix of centered log-ratio (CLR) transformed relative abundance data (i.e., Aitchison distance). Associations of clinical covariates with the dispersion of sample resistome compositions were statistically evaluated with the Permutational multivariate analysis of variance (PERMANOVA) test using the *adonis2* function (vegan package) with 999 permutations. To test the homogeneity of multivariate dispersion, the permutational analysis of multivariate dispersion (PERMDISP) test was used. In case of heterogeneity, the analysis of similarities (ANOSIM) test using the *anosim* function (vegan package) was further used to statistically re-assess the association of variables with the overall resistome composition. In the multivariable, temporal analyses, only the variables that individually showed a significant association with resistome composition were included.

To identify the resistotypes, the Dirichlet multinomial mixture (DMM) model approach was utilized using the DirichletMultinomial (v.1.38.0) R package [23]. Procrustes analysis was performed to examine the effect of underlying microbiota on the resistome. The ARG RPKM

239 abundance matrix and the species-level count per million (CPM) abundance matrix were
240 Hellinger-transformed and ordinated using PCoA (Principal coordinates analysis) on Bray–Curtis
241 dissimilarities. The symmetric Procrustes correlation coefficients between the ordinations and P-
242 values were obtained through the *protest* function from the vegan package. Spearman's correlation
243 was applied to resistome and microbiome relative abundance profiles for paired samples. To avoid
244 the potential bias introduced by Spearman's rank when ranking zero values, we removed ARGs
245 and species that were present in less than half of the samples. The correlation coefficient and
246 corresponding adjusted p-values were calculated using the associate function from the microbiome
247 (v.1.18.0) package [24]. Associations between individual taxa and overall resistome outcomes
248 were calculated using repeated measure correlation analysis from the rmcorr (v.0.5.4) package
249 while controlling for inter-individual variation [25]. Heatmaps were created using the *pheatmap*
250 function from the pheatmap R package (v.1.0.12) [26]. For simple, independent comparisons of
251 group differences, we used a one-way analysis of variance test, Wilcoxon rank-sum test, Kruskal-
252 Wallis test, or chi-square test as appropriate, and we considered $p < 0.05$ to be significant. For all
253 analyses involving multiple comparisons, we applied the Benjamini-Hochberg (BH) method to
254 correct for multiple testing. The statistical analysis scheme used in the study is shown in
255 Supplementary Fig. 2.
256
257 **Ethics statement**
258 Premature infant metadata and all nasopharynx aspirate samples used in this study were collected
259 at the Neonatal Intensive Care Unit (NICU) at Oslo University Hospital, Ullevål, Norway, as part
260 of the Born in the Twilight of Antibiotic project. The study received approval from the Regional
261 Committee for Medical and Health Research Ethics – South East, Norway (2018/1381 REKD),
262 and followed the principles of the Declaration of Helsinki guidelines. Infants were enrolled in the
263 study after obtaining written informed consent from their parents.
264
265 **RESULTS**
266
267 **Distribution and characterization of ARGs in the preterm infant nasopharyngeal**
268 **microbiome**

269 In total, we found 373 ARGs belonging to 15 ARG classes conferring resistance via 5 distinct
270 mechanisms. Multidrug efflux mediated ARGs comprised the majority of the resistome with an
271 average of 27% at T1 to T5 and 17.8% at T6 (six months corrected age) of the total relative
272 abundance. There was a median of 18 (mean: 24) unique ARGs ranging from 1 to 157 per analyzed
273 sample. On average, maximum number of ARGs were detected after the discontinuation of
274 treatment (T2) (mean: 40) in most infants receiving early antibiotics treatment. In the non-treated
275 group, the corresponding mean value at T2 was 19. Only 13 ARGs were both most prevalent
276 (detected in > 40% of all samples) and highly abundant with mean relative abundance of 64.8%
277 (SD: 32.06%) across all preterm infant nasopharyngeal resistome. These core set of resistance
278 genes included multidrug resistance genes (*acrB, oqxB, mexI),* macrolide-lincosamide-
279 streptogramin (MLS) resistance genes (*mel, pmrA, rlmA(II)*), two beta-lactam resistance genes
280 (*blaZ, SST-1*), fluoroquinolone resistance genes (*patA, patB*), tetracycline resistance genes
281 (*tet(41), tetM*), and the aminoglycoside resistance gene *AAC(6')-Ic*.
282
283 In both antibiotic-treated and naive group, the most abundant class of ARG was multidrug (mean:
284 27.85%; SD: 21.90%), followed by fluoroquinolone (mean: 17.41%; SD: 20.18%), beta-lactam
285 (mean: 15.27%; SD 16.99%), tetracycline (mean: 14.70%; SD: 10.75%), MLS (mean: 13.06%;
286 SD: 15.29%) and aminoglycoside (mean: 8.77%; SD: 12.48%) (Fig. 1A). In terms of prevalence,
287 aminoglycoside (76/82) and beta-lactam (75/82) were the most common class present in the
288 samples from antibiotic-treated group, while multidrug (92/99) and beta-lactam (91/99) were the
289 most common class in the naive infants. When classified based on resistance mechanism, antibiotic
290 efflux (mean: 59.56%; SD: 20.97%) was the most abundant category in both the groups across all
291 time points (Supplementary Fig. 3). At baseline (T1), the ARG composition of antibiotic-treated
292 group was found to be more heterogeneous (higher inter-individual variability) and diverse
293 compared to naive infants (Fig. 1B).
294
295 Homology to ARGs posing major current threat for public health, including those identified by
296 recent omics-based framework [27], were identified in nasopharyngeal samples of our cohort.
297 Most of these potential high-risk ARGs (rank I) were detected at multiple time points across infants
298 in both antibiotic-treated and naive group (Supplementary Fig. 4). We also observed a few high
299 risk and clinically relevant ARGs that appeared directly after the early antibiotic treatment at T2.

300 For instance, high-risk ARGs specific to the antibiotics encoding for aminoglycoside-modifying
301 enzymes (*AAC(3)-II*) along with extended-spectrum TEM beta-lactamase (bla_{TEM-1}) were detected
302 in two infants. These ARGs co-occurred with other non-targeted high-risk ARGs such as *qnrS*,
303 *floR* and *aadA* (rank II). Previous studies have shown the co-carriage of these ARGs on multidrug
304 resistance plasmids in Enterobacter species [28]. Also, other clinically relevant Extended Spectrum
305 Beta-Lactamase (ESBL) genes belonging to SHV beta-lactamase (rank II) family were detected in
306 two other infants following antibiotic exposure. In one patient, the plasmid-mediated AmpC-type
307 β-lactamase (ACT-beta lactamase), known to confer resistance against all classes of beta-lactam
308 antibiotics such as penicillins, cephalosporins and carbapenems was detected only after the
309 treatment cessation (T2). These high-risk ARGs were observed mainly in early antibiotics exposed
310 infants whose mothers had also received antibiotics during pregnancy (prenatal). Nonetheless, they
311 did not persist in nasopharyngeal samples collected after discharge at the corrected age of 6
312 months. Amongst ARGs that are found in Gram-positive pathogens, the *Staphylococcus aureus*
313 methicillin-resistance gene i.e., *mecA* was detected in 44 samples from almost all (14/15)
314 antibiotic-treated infants and 10 of the 21 naive infants.
315
316 **Impact of early antibiotics on preterm infant nasopharyngeal resistome**
317 Firstly, we investigated the total ARG abundance (quantified as the sum of RPKM) and ARG α-
318 diversity (quantified by Shannon index) across sampling time points. In both the antibiotic-naive
319 and treated groups, we found that total ARG abundance and diversity increased from T1 to T3-T4,
320 before descending at T5 until 6 month corrected age (T6). In treated infants, total ARG abundance
321 and diversity increased more rapidly upon antibiotic treatment, whereas the changes were
322 generally more gradual over time points in naive infants (Fig. 2A, B). To identify the direct effects
323 of early antibiotic treatment on the total ARG abundance and diversity, we used a generalized
324 linear mixed model with individual set as random effect while correcting for age (DOL). Early
325 antibiotics were significantly associated with increase in both total ARG abundance (LME model:
326 $p.adj = 0.008$) and Shannon diversity ($p.adj = 0.008$) after treatment (T2) as compared to baseline
327 (T1) samples in preterm infant nasopharyngeal resistome. Moreover, the significant effect of
328 antibiotics on increase in total ARG abundance was observed until T3 ($p.adj = 0.001$).
329 Nonetheless, such effects were short-lived, as we did not identify significant differences at any of
330 the following timepoints compared to baseline (T1) in antibiotic-treated group. In naive group of

331 infants, no significant differences appeared in total ARG abundance and α-diversity between any
332 timepoints during the first 6 months of life while correcting for age.
333
334 β-diversity analysis revealed that the overall resistome composition did significantly differ
335 between the two groups (PERMANOVA: $R^2 = 8.2\%$, *p.adj* = 0.02) at baseline (T1). Infants that
336 received early antibiotic treatment had more dissimilar (dispersed) resistome composition than the
337 naive infants (Fig. 3A). Further, we observed that nasopharyngeal samples were clustering more
338 effectively according to birth weight group ($R^2 = 16.4\%$, *p.adj* = 0.012; Supplementary Fig. 5A).
339 No significant difference in baseline resistome composition between antibiotic-treated and naive
340 infants ($R^2 = 0.03$, *p.adj* = 0.319) was found while correcting for birth weight group. However, a
341 minor yet significant compositional difference in resistome was still observed between the two
342 groups ($R^2 = 5.9\%$, *p.adj* = 0.03) directly after the cessation of early antibiotic treatment (T2) (Fig.
343 3B). Moreover, the compositional changes due to early antibiotics at T2 were more discernible in
344 infants whose mothers had received prenatal antibiotics during pregnancy (prenatal + early ($n = 8$)
345 vs naive ($n = 20$): $R^2 = 10.09\%$, *p.adj* = 0.006; Supplementary Fig. 5B). Following treatment,
346 resistome normalized rapidly as no significant differences in composition between naive and
347 treated group were detected from T3 until 6 months of corrected age (T6) (Fig. 3C).
348
349 **Clinical covariates associated with preterm infant nasal resistome composition development**
350 The associations between resistome composition and clinical factors were analyzed using
351 PERMANOVA-tests (all *p.adj* ≤ 0.001) and principal component analysis (PCA) across all
352 nasopharyngeal samples from preterm birth until 6 months corrected age, using centered log-ratio
353 (CLR) transformed relative abundance data (Supplementary Fig. 2). The PCA was explained by
354 two main principal components, describing 18.5% and 17.7% of the variation, respectively.
355 Besides inter-individual variation (PERMANOVA (univariate): $R^2 = 37.7\%$, *p.adj* = 0.001), age
356 was the other significant covariate found to be associated with overall resistome composition in
357 the overarching cohort. A moderate yet statistically significant effect was observed with GA (R^2
358 4%, *p.adj* = 0.001) and BW group (R^2 4.1%, *p.adj* = 0.001), which are highly correlated variables
359 (as indicated by chi-squared test: $p < 2.2e\text{-}16$). Mode of delivery, gender, and intrapartum
360 antibiotic treatment did not appear to shape the overall nasopharyngeal resistome composition
361 (Supplementary Table 3). Covariates that differed between the antibiotic-treated and naive infants

362 at baseline and that also had a significant association with the overall resistome composition (GA
363 and prenatal antibiotics (R^2 4.5%, *p.adj* = 0.001)) were used for correction in the subsequent
364 multivariate analysis.
365
366 Multivariate PERMANOVA analysis showed that the largest effect on overall resistome
367 composition remained attributed to inter-individual variation (R^2 30.03%), followed by PMA (R^2
368 9.7%) and chronological age group (sampling time points) (R^2 8.4%, all *p.adj* = 0.001). Although
369 we observed a high degree of variation between samples with no obvious discrete spatial clustering
370 in relation to these variables upon visual inspection (Fig. 4A-C). We found no significant influence
371 of early antibiotics (R^2 = 0.6%, *p.adj* = 0.31) and its duration (R^2 = 1.2%, *p.adj* = 0.009) after
372 adjusting for confounding variables. However, we observed a significant effect of prenatal
373 antibiotics (R^2 = 3.6%, *p.adj* = 0.001) and its duration (categorized) (R^2 = 4.4%, *p.adj* = 0.001) in
374 shaping the overall nasopharyngeal resistome composition in preterm birth. To account for the
375 likelihood of prenatal antibiotics with early antibiotics exposure in our analyses (Chi-squared test,
376 $p < 0.0003$), we further categorized our early antibiotic-treated group based on whether the mothers
377 had also received antibiotics during pregnancy (prenatal + early, *n = 9*) or not (only early, *n = 6*).
378 Post-hoc analysis revealed significant differences in the overall resistome composition of infants
379 having exposure to both prenatal and early life antibiotics (prenatal + early) compared to infants
380 receiving either only early or naive (no prenatal + no early) (*p.adj* = 0.001; Fig. 4D). No difference
381 in overall resistome composition was observed between the only early antibiotics and naive group
382 (*p.adj* = 0.14), underlining the relevance of prenatal antibiotic exposure in shaping the overall
383 resistome composition of preterm infants during the first six months of life (corrected).
384
385 **Unsupervised clustering classifies nasopharyngeal samples into distinct clusters based on**
386 **composition differences**
387 Next, we determined whether the nasopharyngeal samples were organized into clusters (i.e.,
388 resistotypes) according to their ARG abundance profiles using an unsupervised method, i.e., DMM
389 models. This way, we identified three distinct clusters (or resistotypes) that were driven by specific
390 ARGs. Majority of our samples (~90%) were classified into two main resistotypes, R1 and R2. No
391 significant difference in the sequencing depth, was detected between the two Resistotypes
392 (pairwise Wilcox test: $p = 0.14$), suggesting that sequencing depth was not driving their separation.

393 The first resistotype (R1) was characterized by a predominance of fluoroquinolone resistance
394 genes (*patA, patB*) and *rlmA(II)* (Supplementary Fig. 6). On the other hand, resistotype R2 was
395 driven mainly by relatively higher abundance of the beta-lactam resistance gene *SST-1* and the
396 tetracycline resistance gene *tet(41)*. While resistotype R3 was enriched with the blaZ β-lactamase
397 resistance gene. The identified resistotypes did vary in their overall ARG composition (β-
398 diversity), as shown in a PCA plot (Fig. 5A). Further, PERMANOVA analysis showed that
399 resistotypes were more strongly (larger effect size) and significantly associated with the resistome
400 than other clinical covariates or early-life factors in preterm infants during the first 6 months of
401 life (R^2 = 19.05%, *p.adj* = 0.001).
402
403 **Nasopharyngeal resistome composition of preterm infants is linked to microbial composition**
404 **and carries hospitalization signatures**
405 To investigate the impact of microbial composition on resistome composition, we taxonomically
406 profiled the infant nasopharyngeal samples using MetaPhlAn3 (manuscript under preparation,
407 Rajar et al. 2023). We noticed that microbial community types, as determined by DMM models,
408 were highly linked with the resistotypes (chi-square test: $p < 0.0001$). Resistotype R2 was very
409 specific to the *Serratia* community type. Resistotype R1 was associated with the community types
410 mainly driven by *Streptococcus* and *Gemella* and Resistotype R3 was enriched in community types
411 predominated by *Staphylococcus* and *Cutibacterium* (Fig. 5B). These community types were also
412 significantly associated to the overall composition of the resistome (PERMANOVA: R^2 = 13.01%,
413 *p.adj* = 0.001; Fig. 5C). Additionally, result from Procrustes analysis showed a strong and
414 significant association between the microbial taxa abundance and ARG abundance profiles
415 (correlation coefficient (r) = 0.85, p = 0.001, m^2 = 0.26), underlining that bacterial community
416 composition structured the resistome composition in the nasopharyngeal microbiome of preterm
417 infants. However, the correlation was more robust in the antibiotic-naive infants (r = 0.94, p =
418 0.001, m^2 = 0.11) as compared to antibiotic-treated infants (r = 0.85, p = 0.001, m^2 = 0.27; Fig.
419 6A, B).
420
421 Next, we performed pairwise Spearman's correlation analysis between ARG and microbial species
422 abundances to predict the origin or potential microbial host of ARGs in nasopharyngeal samples.
423 The strongest positive correlation ($r \geq 0.8$, *p.adj* < 0.05) was found between *SST-1, AAC(6')-Ic,*

424 *tet(41)*, *MexI ARGs* and *Serratia marcescens/nematodiphila* (Fig. 6C). It is worth noting that
425 during our study, a clinically detected outbreak of *S. marcescens* occurred in the NICU a few
426 months after infant enrollment had commenced. When present at T1, the carriage of these ARGs
427 persisted in 20 out of 22 preterm infants across both the naïve and treated groups, extending long
428 after hospital discharge as observed at ~8 to 10 months chronological age (6 months corrected age
429 (T6)) (Fig. 6D). Occasional time points in which the ARGs associated with *S. marcescens* showed
430 variations from T1 were observed in few of the samples. Whether these are true variations is
431 difficult to ascertain as metagenomic studies can only detect DNA sequences that are present in
432 the specific sample and to a depth that is possible to be captured by the present analytical
433 approaches.
434
435 The highly prevalent ARGs like *patA, patB and rlmA(II)* within Resistotype 1 were observed to
436 exhibit a strong co-occurrence with *Streptococcus mitis/oralis* and *Gemella*
437 *haemolysans/sanguinis* (Fig. 6C). *Staphylococcus aureus* and *Staphylococcus epidermidis*
438 correlated only with *blaZ* beta-lactam ARG. Interestingly, these associations are in accordance
439 with information in the literature and AMR gene databases on the known microbial host of these
440 ARGs, highlighting the potential of such analysis approach to predict the origin of ARG in
441 metagenomes. However, to reduce the bias due to false ranking of features with many zero values,
442 this analysis is only applicable on the features (ARGs or taxa) present in high prevalence across
443 samples.
444
445 Lastly, we also examined whether some of these strongly associated bacterial species with ARGs
446 would also directly correlate with our overall resistome outcomes. We plotted the relative
447 abundance of each species versus the total ARG abundance (sum of RPKM) and the number of
448 unique ARGs from all samples. Our analysis revealed that only the relative abundance of
449 *Streptococcus mitis* (rmcorr correlation coefficient (r^{rm} = 0.29, *p.adj* = 0.01) showed a moderately
450 linear association with the total ARG abundance while adjusting for inter-individual variability
451 (Supplementary Fig. 7). However, we did not observe a significant correlation between the relative
452 abundance of *S. mitis* and the number of unique ARGs (r^{rm} = 0.19, *p.adj* = 0.12). To statistically
453 assess whether *S. mitits* relative abundance explains the dispersion of resistomes across samples,
454 we conducted PERMANOVA test and found that *S. mitis* significantly explains 11.69% (*p.adj* =

455 0.001) of the variation in the overall resistome composition. Only *S. marcescens* (R^2 = 11.8%,
456 *p.adj* = 0.001) exhibited similar variation in the resistome.
457
458 **DISCUSSION**
459 The nasopharynx represents a highly accessible microbial community and also serves as a crucial
460 diagnostic window in combating respiratory infections and AMR [29]. It is also the primary
461 ecological niche for common respiratory pathogens [30]. To date, one of the closest attempts for
462 an in-depth investigation of ARGs in the respiratory tract using a culture-free approach identified
463 ARGs in 86% of the samples. However, this was based on the microbiomes of the anterior nares
464 and from an adult population (Human Microbiome Project) [31], which are different from the
465 groups and collection sample sites used in our study. In a recent study focusing on infants, ARGs
466 were detected in 64% of nasopharynx samples [32]. Samples in this study were enriched for
467 streptococci, thus precluding general conclusions on the extent of ARG presence in the natural
468 microbiome. In our study, ARGs were present in 95% of the preterm infant samples from the
469 nasopharynx, regardless of whether they were exposed to antibiotics or not. Although our study
470 differs in several aspects from the ones above, the combined findings indicate that the respiratory
471 tract is likely an important reservoir of ARGs and an integral component of human respiratory
472 microbiomes. For preterm infants, the detection of ARGs in the nasopharynx during the first few
473 hours of life might occur due to microbiome colonization from multiple sources such as
474 transgenerational transfer from mothers and the environment [33, 34]. In addition, preterm infants
475 usually have longer stay in the hospital after birth than term infants. This most likely contributes
476 to an increased risk for respiratory tract colonization with drug resistant microorganisms, as
477 recently reported in relation to the gut resistome [35].

478 We found that ARGs conferring multidrug resistance via efflux pumps encompassed a large
479 proportion of the nasopharyngeal resistome in the preterm infants. Similar findings have been
480 previously reported in the gut and oral resistome of preterm and term infants by other metagenomic
481 studies [7, 33, 36, 37]. In addition, we detected many of the same aminoglycoside-modifying
482 enzymes, tetracycline protection proteins, and beta-lactamases ARGs identified previously in the
483 nasopharynx of South African infants, even though the samples in the South African study were
484 culture-enriched for streptococci and were from a country with higher AMR challenges than the
485 origin of samples used in the present study. Such similarities in resistome composition may, at

486 least in part, be explained by the large proportion of streptococci found in the natural microbial
487 communities of the infant nasopharynx. Together, these findings suggest the existence of a core
488 infant nasopharyngeal resistome.

489 *Staphylococcus aureus* is one of the leading pathogens causing neonatal sepsis, and one of the
490 priority drug resistant pathogens by the WHO [38]. We found a strong correlation between the
491 abundance of *Staphylococcus aureus/epidermidis* and the PC1-beta-lactamase (*blaZ*) gene. This is
492 a gene found in the chromosome and mobile plasmids of staphylococci, thus representing high risk
493 ARGs. Another strong correlation was found between *Serratia* species and specific resistance
494 genes such as *SST-1*, which confers resistance against cephalosporins. Other resistance genes
495 correlated with this group included aminoglycoside (*AAC(6')-Ic)* and tetracycline (*tet(41)*)
496 resistance genes. Of note, many clinical isolates of *Serratia spp.*, especially *S. marcescens*, have
497 been reported to harbor these ARGs on their chromosomes or plasmids [39-41]. The beta-
498 lactamase encoding gene *SST-1*, for instance, is found in 33% of the sequenced genomes of *S.*
499 *marcescens*, while the aminoglycoside acetyltransferase *AAC(6')-Ic* and *tet(41*) are found in 96%
500 and 67% of *S. marcescens* available genomes, respectively [16]. This holds particular relevance,
501 given that *S. marcescens* is linked to nosocomial infections and hospital outbreaks, particularly in
502 NICUs [42]. Of note, multidrug resistant *Serratia spp.* are listed among the most critical bacteria
503 by the WHO for which innovative new treatments are urgently needed [43].

504 Our results showed a significant effect of early broad-spectrum antibiotic regimen
505 (ampicillin + gentamicin) on the total abundance, diversity, and composition of ARGs in the
506 nasopharynx of preterm infants. These effects on the resistome were short-lived and only observed
507 directly after the cessation of antibiotic treatment (T2). In a recent study investigating the impact
508 of antibiotics used in the first week of life on intestinal resistome, Reyman et al. reported that three
509 other antibiotic combinations (penicillin + gentamicin, co-amoxiclav + gentamicin, and
510 amoxicillin + cefotaxime) also had a transient impact on the infant resistome [44]. As the effects
511 of antibiotics on microbiome and resistome composition are unique to each individual and
512 dependent on the antibiotics used and the body site, caution should be taken when extending the
513 interpretation of such general findings. Interestingly, the transient effects of
514 ampicillin + gentamicin on the nasopharyngeal resistome were more discernible in preterm infants
515 whose mothers were exposed to antibiotics during pregnancy before delivery (prenatal antibiotics).

516 Moreover, we also observed a more profound effect of prenatal antibiotic exposure on the
517 nasopharyngeal resistome of preterm infants at T2 (after cessation of early antibiotics), which
518 persisted until the last sampling time point before discharge from the NICU (T4-T5), but not at 6
519 months corrected age. Although very few studies have focused on the association between prenatal
520 antibiotic exposure and the resistome, a recent study based on the gut indicates a possible
521 correlation of prenatal antibiotics with increased abundance of ARGs, at least when used 5 to 6
522 weeks before birth [7]. Additionally, in concordance with other studies on the infant gut [33, 45],
523 our data showed no significant impact of intrapartum antibiotic therapy on the overall
524 nasopharyngeal resistome composition. The nasopharyngeal samples of preterm infants at 6
525 months corrected age were still clustered according to the birth weight group (similar to T1,
526 Supplementary Fig. 5C), suggesting a strong but transient influence of other environmental factors,
527 such as antibiotic exposures. Nonetheless, there is a paucity of relevant studies on nasopharynx
528 that can corroborate our findings, as they are mainly restricted to the infant gut resistome [33-35,
529 44-46].

530 One of the strengths of this study is that it presents the most comprehensive information to date on
531 the characteristics and dynamics of resistome development in the nasopharynx, including results
532 for the first six months of life (corrected) – a critical period of airway development in infants. In
533 addition, the naive group was enrolled at the same time as our treated group from the same hospital,
534 thus removing potential confounding differences in enrollment years or sampling duration. On the
535 other hand, the small number of samples at each time point and inter-individual response to
536 antibiotic exposure may be insufficient to unravel significant associations between antibiotic
537 regimen and resistome development. Moreover, the baseline samples for the antibiotic-treated
538 group ideally would represent samples taken before treatment. However, antibiotic treatment
539 initiation for sEONS and clinical stabilization of the preterm infants were a high priority that
540 needed to precede nasopharynx sample collection. To partially address this issue, we included only
541 preterm infants with samples collected no more than 24 hours after antibiotic initiation. Even so,
542 this choice may have resulted in the selection of more clinically stable infants, which could
543 possibly limit the generalizability of our results. Despite this approach, the differences in resistome
544 changes between baseline and after treatment found in our study may underestimate the true impact
545 of antibiotics on the resistome. We rigorously adjusted for such confounding factors and inter-
546 individual variation using multivariable methods or matched analysis to ensure the reliability of

547 our observations as a consequence of early antibiotics. Nevertheless, we cannot isolate early
548 antibiotic effects from other adverse pre-life events that coincide with premature birth, such as
549 prenatal and antepartum maternally administered antibiotics. Furthermore, for low biomass
550 samples such as those used in our study, sensitive approaches such as deep sequencing are
551 necessary to obtain meaningful microbiome information. This increases the risk of detecting
552 potential DNA contaminants. Although it is not possible to eliminate the potential identification
553 of contaminants, we adopted various control procedures, starting from sample collection to
554 analysis, to ensure that the data presented in our study is a reliable representation of the ARGs
555 found in the nasopharynx.
556
557 In conclusion, our findings suggest that early-life treatment with broad-spectrum antibiotics for
558 sEONS can cause perturbations in the nasopharyngeal resistome of preterm infants. However, the
559 overall perturbations were transient, as we did not observe a significant alteration in diversity,
560 abundance, composition, or carriage persistence of high-risk ARGs in treated preterm infants
561 compared to controls at 6 months of corrected age. These are nevertheless potentially relevant,
562 given the vulnerability of preterm infants in their first days of life. The individual variability in the
563 distribution of high-risk ARGs and the changes in the resistome following antibiotic treatment
564 highlight the potential benefits of developing resistome profiling tools for point-of-care use, which
565 could aid in informing and tailoring antibiotic treatment decisions [47]. In addition, prolonged
566 carriage of ARGs associated with nosocomial *S. marcescens* was noted. Future multicenter studies
567 with long-term follow-up are needed to assess the ecological side-effects of hospitalization and
568 different antibiotic regimens, including alternatives to broad-spectrum antimicrobial therapy on
569 infant respiratory resistome development. Understanding the development of the respiratory
570 resistome and the impact of antibiotic therapies is relevant for striking a balance between the
571 conflicting goals of minimizing the risk of serious infectious complications to preterm birth on one
572 side, and minimizing the selection of organisms with ARGs on the other. In general, the
573 disproportionate contribution of respiratory pathogens to drug-resistant deaths at all ages
574 highlights the importance of future studies focusing on this yet rather unexplored reservoir of
575 pathogens and ARGs.
576

577 **REFERENCES**

578 1. Murray, C., et al., *& Naghavi, M.(2022). Global burden of bacterial antimicrobial*
579 *resistance in 2019: a systematic analysis.* The Lancet. **399**(10325): p. 629-655.
580 2. Von Wintersdorff, C.J., et al., *Dissemination of antimicrobial resistance in microbial*
581 *ecosystems through horizontal gene transfer.* Frontiers in microbiology, 2016. **7**: p. 173.
582 3. Kilian, M., et al., *Parallel evolution of Streptococcus pneumoniae and Streptococcus mitis*
583 *to pathogenic and mutualistic lifestyles.* MBio, 2014. **5**(4): p. e01490-14.
584 4. Pailhoriès, H., et al., *Antibiotic resistance in chronic respiratory diseases: from*
585 *susceptibility testing to the resistome.* European Respiratory Review, 2022. **31**(164).
586 5. Teo, S.M., et al., *The infant nasopharyngeal microbiome impacts severity of lower*
587 *respiratory infection and risk of asthma development.* Cell host & microbe, 2015. **17**(5): p.
588 704-715.
589 6. Toivonen, L., et al., *Antibiotic treatments during infancy, changes in nasal microbiota, and*
590 *asthma development: population-based cohort study.* Clinical Infectious Diseases, 2021.
591 **72**(9): p. 1546-1554.
592 7. Li, X., et al., *The infant gut resistome associates with E. coli, environmental exposures, gut*
593 *microbiome maturity, and asthma-associated bacterial composition.* Cell Host & Microbe,
594 2021. **29**(6): p. 975-987. e4.
595 8. Rajar, P., et al., *Microbial DNA extraction of high-host content and low biomass samples:*
596 *Optimized protocol for nasopharynx metagenomic studies.* Frontiers in Microbiology,
597 2022. **13**.
598 9. Eisenhofer, R., et al., *Contamination in low microbial biomass microbiome studies: issues*
599 *and recommendations.* Trends in microbiology, 2019. **27**(2): p. 105-117.
600 10. Krueger, F., *Trim Galore: a wrapper tool around Cutadapt and FastQC.* Trim Galore, 2012.
601 11. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature
602 methods, 2012. **9**(4): p. 357-359.
603 12. Li, H., et al., *The sequence alignment/map format and SAMtools.* bioinformatics, 2009.
604 **25**(16): p. 2078-2079.
605 13. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic*
606 *features.* Bioinformatics, 2010. **26**(6): p. 841-842.
607 14. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010,
608 Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
609 15. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a*
610 *single report.* Bioinformatics, 2016. **32**(19): p. 3047-3048.
611 16. Alcock, B.P., et al., *CARD 2020: antibiotic resistome surveillance with the comprehensive*
612 *antibiotic resistance database.* Nucleic acids research, 2020. **48**(D1): p. D517-D525.
613 17. Doster, E., et al., *MEGARes 2.0: a database for classification of antimicrobial drug, biocide*
614 *and metal resistance determinants in metagenomic sequence data.* Nucleic acids
615 research, 2020. **48**(D1): p. D561-D569.
616 18. Allaire, J., *RStudio: integrated development environment for R.* Boston, MA, 2012.
617 **770**(394): p. 165-171.
618 19. R Core Team, R. and R.C. Team, *R: a language and environment for statistical computing.*
619 *R Foundation for Statistical Computing; 2020*. 2021.

620 20. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive*
621 *analysis and graphics of microbiome census data.* PloS one, 2013. **8**(4): p. e61217.
622 21. Wickham, H., *ggplot2.* Wiley interdisciplinary reviews: computational statistics, 2011.
623 **3**(2): p. 180-185.
624 22. Oksanen, J., et al., *Vegan: Community Ecology Package (R Package Version 2.6-2). 2022*.
625 23. Holmes, I., K. Harris, and C. Quince, *Dirichlet multinomial mixtures: generative models for*
626 *microbial metagenomics.* PloS one, 2012. **7**(2): p. e30126.
627 24. Lahti, L. and S. Shetty, *Introduction to the microbiome R package.* 2018.
628 25. Bakdash, J.Z. and L.R. Marusich, *Repeated measures correlation.* Frontiers in psychology,
629 2017. **8**: p. 456.
630 26. Kolde, R. and M.R. Kolde, *Package 'pheatmap'.* R package, 2018. **1**(10).
631 27. Zhang, A.-N., et al., *An omics-based framework for assessing the health risk of*
632 *antimicrobial resistance genes.* Nature communications, 2021. **12**(1): p. 4765.
633 28. Long, Y., et al., *High Carriage Rate of the Multiple Resistant Plasmids Harboring Quinolone*
634 *Resistance Genes in Enterobacter spp. Isolated from Healthy Individuals.* Antibiotics, 2022.
635 **11**(1): p. 15.
636 29. Flynn, M. and J. Dooley, *The microbiome of the nasopharynx.* Journal of Medical
637 Microbiology, 2021. **70**(6).
638 30. Man, W.H., W.A. de Steenhuijsen Piters, and D. Bogaert, *The microbiota of the respiratory*
639 *tract: gatekeeper to respiratory health.* Nature Reviews Microbiology, 2017. **15**(5): p. 259-
640 270.
641 31. Consortium, H.M.P., *Structure, function and diversity of the healthy human microbiome.*
642 Nature, 2012. **486**(7402): p. 207-214.
643 32. Manenzhe, R.I., et al., *Longitudinal changes in the nasopharyngeal resistome of South*
644 *African infants using shotgun metagenomic sequencing.* PloS one, 2020. **15**(4): p.
645 e0231887.
646 33. Parnanen, K., et al., *Maternal gut and breast milk microbiota affect infant gut antibiotic*
647 *resistome and mobile genetic elements.* Nat Commun, 2018. **9**(1): p. 3891.
648 34. Yassour, M., et al., *Natural history of the infant gut microbiome and impact of antibiotic*
649 *treatment on bacterial strain diversity and stability.* Science translational medicine, 2016.
650 **8**(343): p. 343ra81-343ra81.
651 35. Gasparrini, A.J., et al., *Persistent metagenomic signatures of early-life hospitalization and*
652 *antibiotic treatment in the infant gut microbiota and resistome.* Nature microbiology,
653 2019. **4**(12): p. 2285-2297.
654 36. Rahman, S.F., et al., *Machine learning leveraging genomes from metagenomes identifies*
655 *influential antibiotic resistance genes in the infant gut microbiome.* MSystems, 2018. **3**(1):
656 p. e00123-17.
657 37. Sukumar, S., et al., *The development of the oral resistome during the first decade of life.*
658 2022.
659 38. Tacconelli, E., *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research,*
660 *Discovery, and Development.* 2017.
661 39. Thompson, S.A., et al., *Novel tetracycline resistance determinant isolated from an*
662 *environmental strain of Serratia marcescens.* Applied and environmental microbiology,
663 2007. **73**(7): p. 2199-2206.

40. Matsumura, N., S. Minami, and S. Mitsuhashi, *Sequences of homologous β-lactamases from clinical isolates of Serratia marcescens with different substrate specificities.* Antimicrobial agents and chemotherapy, 1998. **42**(1): p. 176-179.
41. Shaw, K., et al., *Characterization of the chromosomal aac (6')-Ic gene from Serratia marcescens.* Antimicrobial agents and chemotherapy, 1992. **36**(7): p. 1447-1455.
42. Cristina, M.L., M. Sartini, and A.M. Spagnolo, *Serratia marcescens infections in neonatal intensive care units (NICUs).* International journal of environmental research and public health, 2019. **16**(4): p. 610.
43. WHO, I., *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics.* Who, 2017. **2017**: p. 1-7.
44. Reyman, M., et al., *Effects of early-life antibiotics on the developing infant gut microbiome and resistome: a randomized trial.* Nat Commun, 2022. **13**(1): p. 893.
45. Lebeaux, R.M., et al., *The infant gut resistome is associated with E. coli and early-life exposures.* BMC Microbiol, 2021. **21**(1): p. 201.
46. Gibson, M.K., et al., *Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome.* Nature Microbiology, 2016. **1**: p. 16024.
47. Leggett, R.M., et al., *Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens.* Nature Microbiology, 2020. **5**(3): p. 430-442.

DATA AVAILABILITY

The raw sequencing data (after removal of human DNA) for all samples are publicly available at NCBI SRA under BioProject ID: PRJNA1009231.

ACKNOWLEDGEMENTS

We would like to thank the parents of participating infants, and the clinical staff at the Neonatal Intensive Care units at Ullevål hospital for their assistance with sample collection. The sequencing service for this work was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no) and the computations were performed on resources provided by Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

FUNDING

This work was financed by the Norwegian Research Council (NFR) (project number: 273833), the Olav Thon Foundation (project number: 421258), and the Faculty of Dentistry at the University of Oslo.

704 **FIGURES**

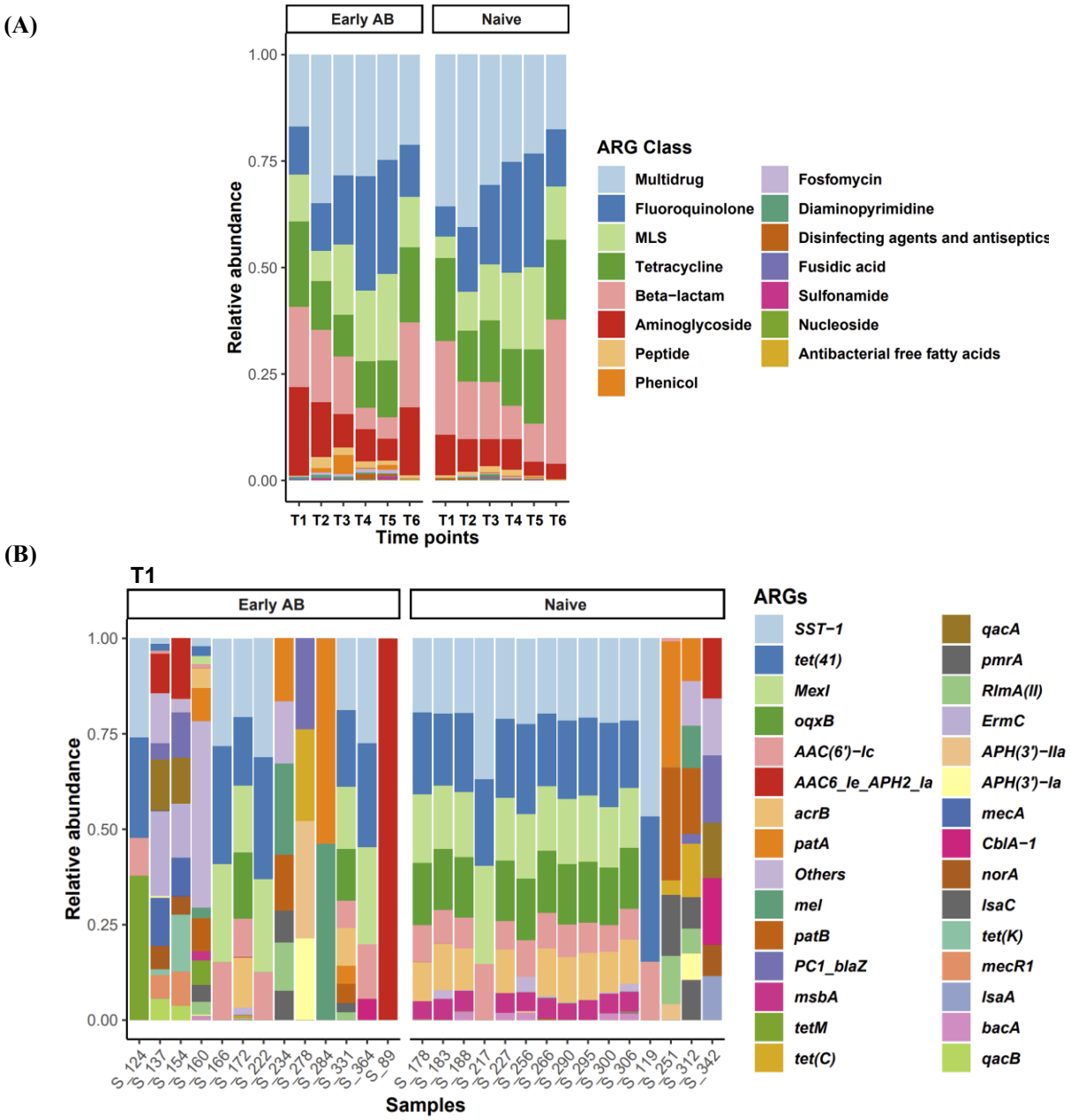


Fig. 1: Resistome composition in the nasopharynx microbiome of preterm infants. Stacked bar plot showing the relative abundance of (A) antibiotic resistance genes (ARGs) clustered at the

Class level for the naive and antibiotic-treated groups over time; (B) top 30 most abundant ARGs identified across samples at baseline (T1) in early antibiotics treated (left) and naive (right) groups. All other ARGs are grouped into the "Other" category. MLS = macrolide-lincosamide-streptogramin.

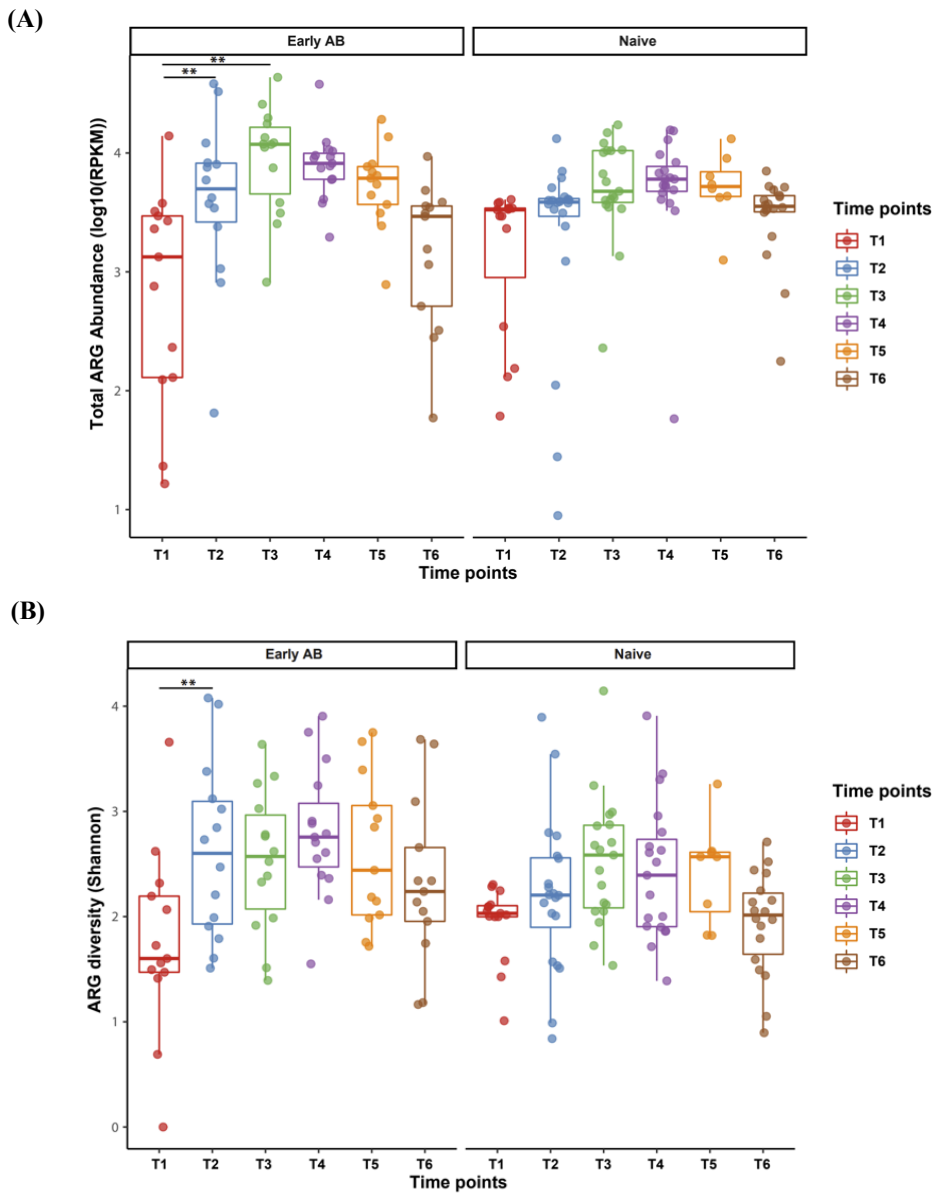
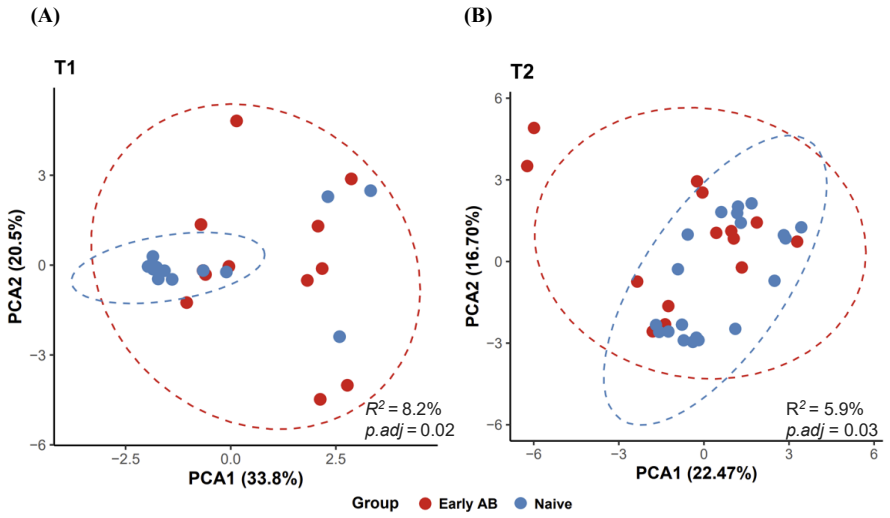


Fig. 2: Impact of early antibiotic treatment on nasopharynx total ARG abundance and ARG α-diversity Boxplots showing the (A) total ARG abundance (sum of RPKM) and (B) ARG Shannon diversity per sample, stratified by antibiotic treatment and time points. The lower and

upper horizontal box lines correspond to the first and third quartiles, with the line in the middle of the boxes representing the median. Whiskers extending from the boxes represent the range of data points within 1.5 times the interquartile range below the first quartile and within 1.5 times the interquartile range above the third quartile. (Statistical significance: LME mixed effect model. Adjusted p-values (p): *p < .05, **p < .01 and ***p < .001). RPKM = reads per kilobase of reference gene per million bacterial reads.



(C)

Time points	R^2 (%)	p.adj
T3	3.4%	0.326
T4	4.4%	0.129
T5	4.99%	0.387
T6	3.58%	0.349

Fig. 3: β-diversity PCA plots of overall nasopharyngeal resistome composition stratified per time point. Principal component analysis (PCA) based on Euclidean distances of CLR-transformed ARG abundances (*i.e.,* Aitchison dissimilarity) between samples, visualizing the overall nasopharyngeal resistome composition stratified for antibiotic-treated infants and naive at time point (A) T1 (baseline) and (B) T2 (immediately after cessation of early antibiotic therapy). Each data point represents the resistome composition of one sample. The ellipses represent the 95% confidence interval around the centroid of resistome composition for each group. Statistical significance (adjusted p-values) and effect size (R^2) of the differences in beta diversity were assessed using PERMANOVA. In panel (C), the results of the PERMANOVA-tests for the later time points are presented. Data points of early antibiotic-treated infants are colored red and those of naive infants are blue. PERMANOVA = permutational multivariate analysis of variance; p.adj = adjusted p-values.

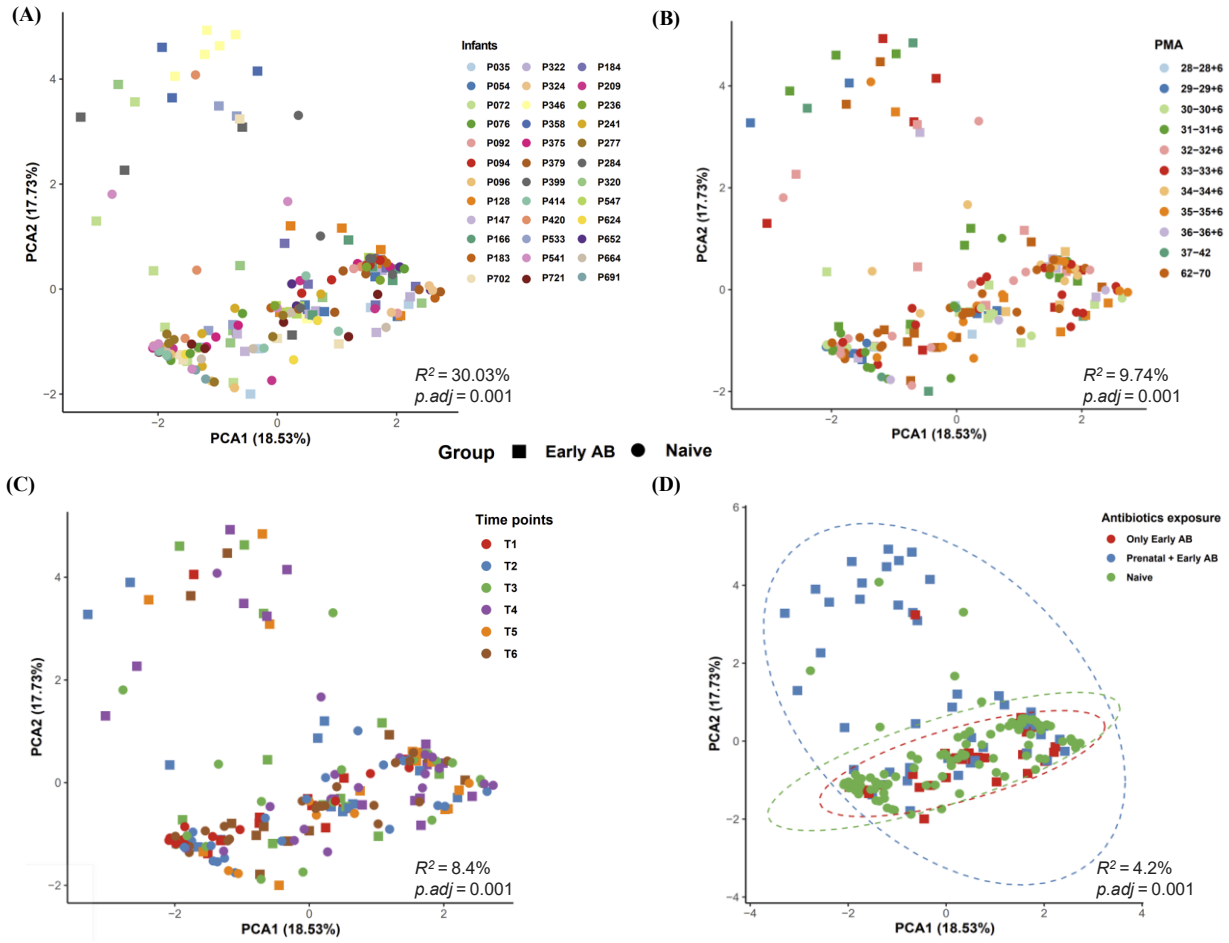


Fig. 4: Factors influencing the overall nasopharyngeal resistome composition in the preterm infants. PCA visualizations of Aitchison dissimilarities between all the nasopharyngeal samples from birth until 6 months corrected age at ARG level. Each data point corresponds to a sample, colored by (A) Infant, (B) PMA, (C) Time points and (D) Prenatal and post-natal antibiotic exposure and shaped according to early antibiotic exposure or naive groups. Effect sizes (R^2) and corresponding adjusted p-values (p.adj, corrected using the Benjamini-Hochberg method) are calculated using PERMANOVA-tests (as shown in the bottom right of each plot). Percentage refers to the percentage of variance explained by the principal component. PMA = postmenstrual age; PC1 = principal component 1; PC2 = principal component 2; PERMANOVA = permutational multivariate analysis of variance; p.adj = adjusted p-values.

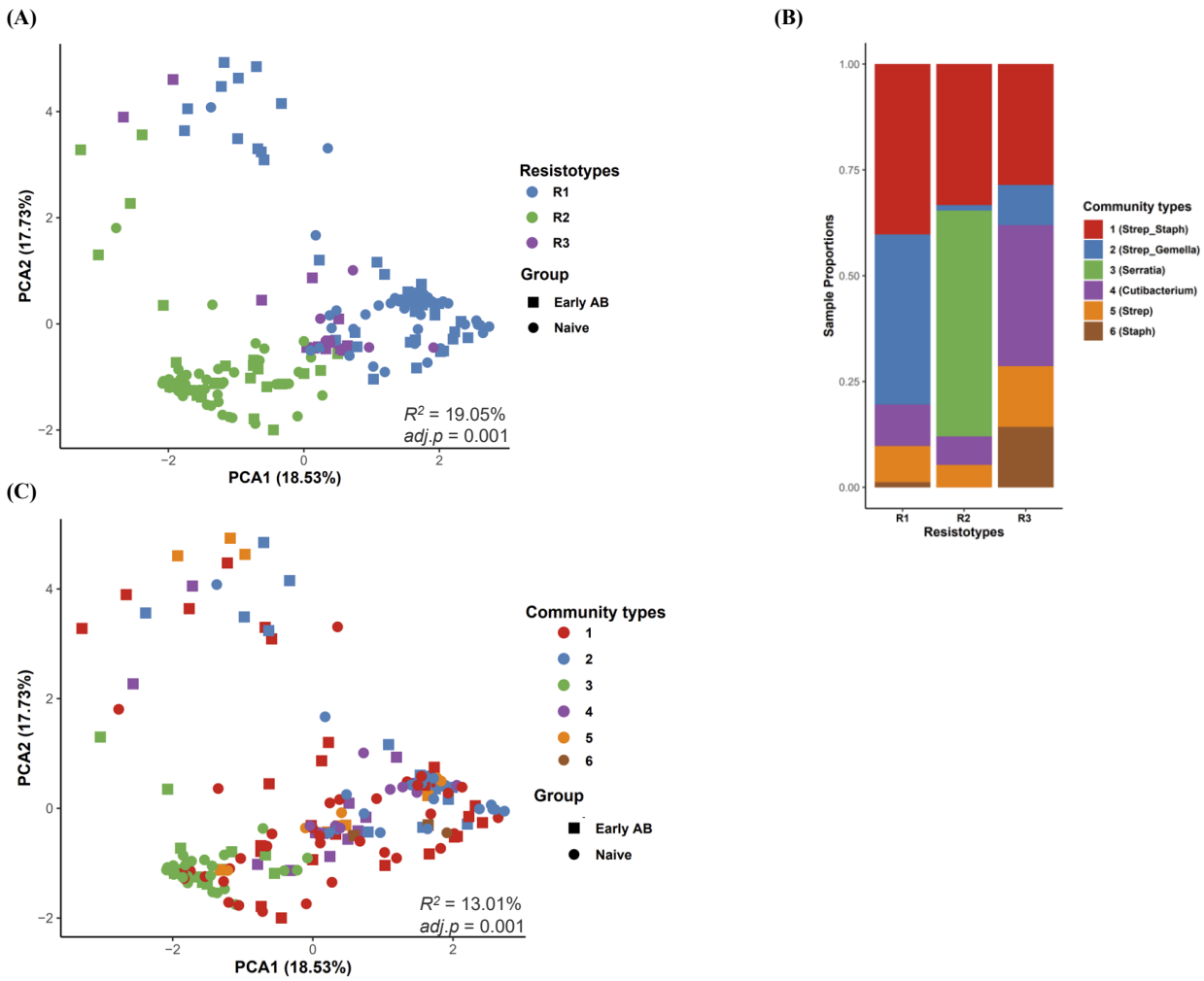


Fig. 5: Association between resistotypes, community types, and ARG abundance profiles. The PCA plot showing the association between overall ARG abundance profile with respect to their (A) resistotypes and (C) community types identified among all the nasopharyngeal samples using the DMM models. R^2 was obtained using the PERMANOVA test. (B) Samples proportions for each resistotype shown as a function of community type. DMM = dirichlet multinomial mixture; Strep = *Streptococcus* ; Staph = *Staphylococcus*.

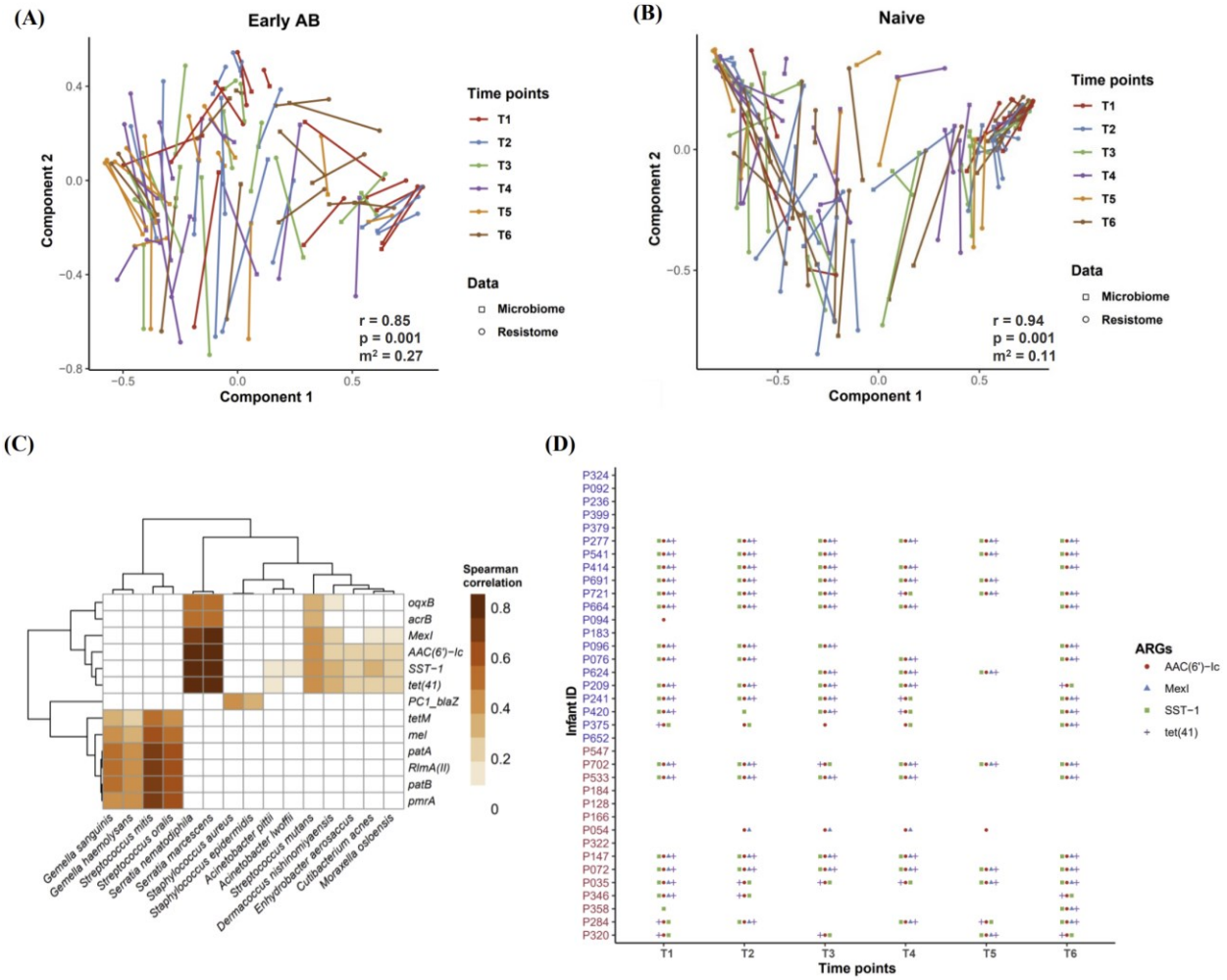
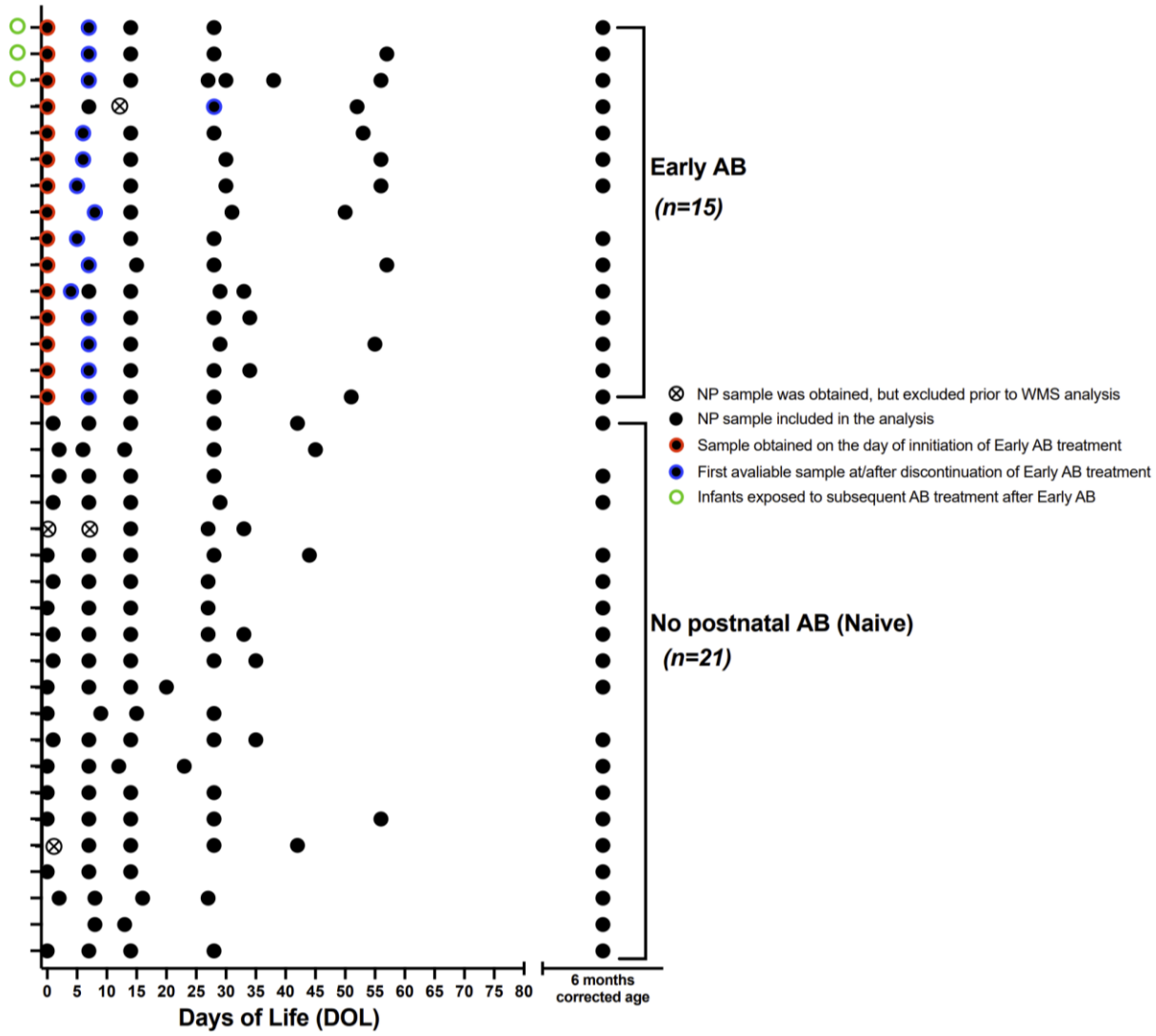
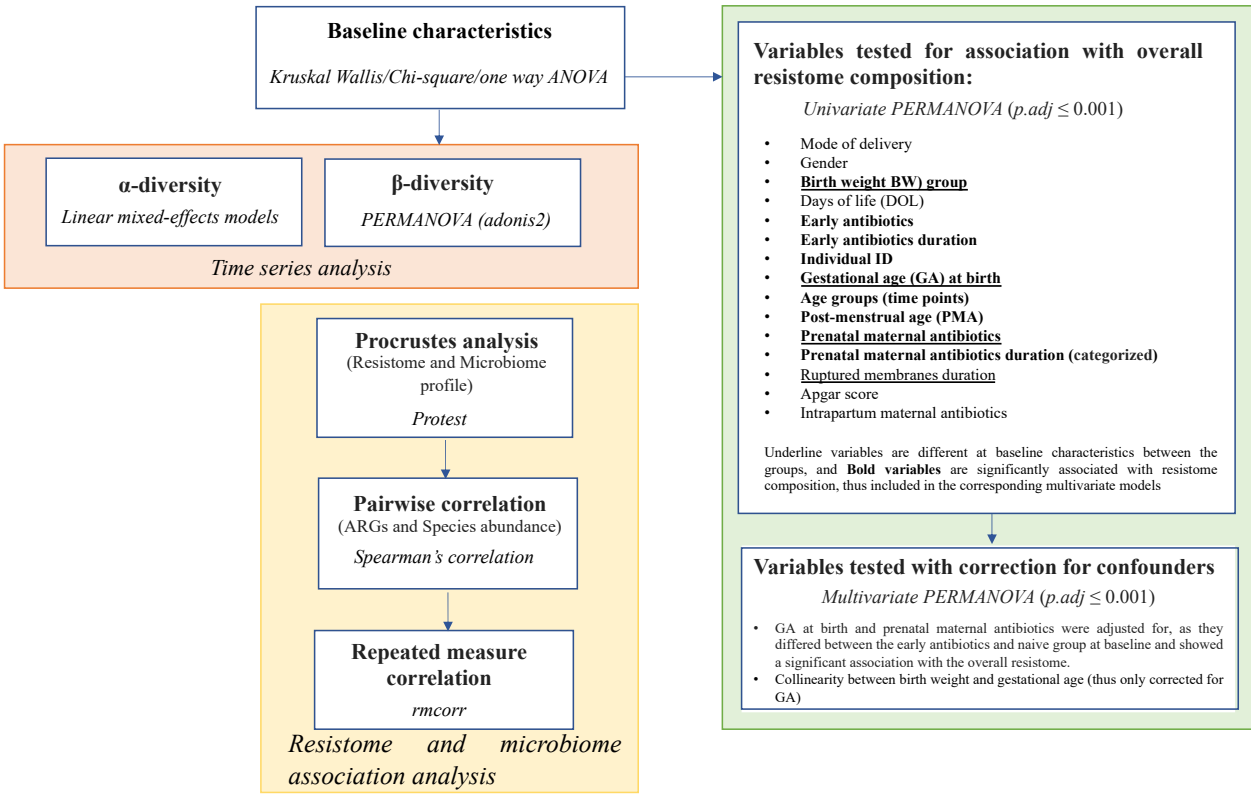


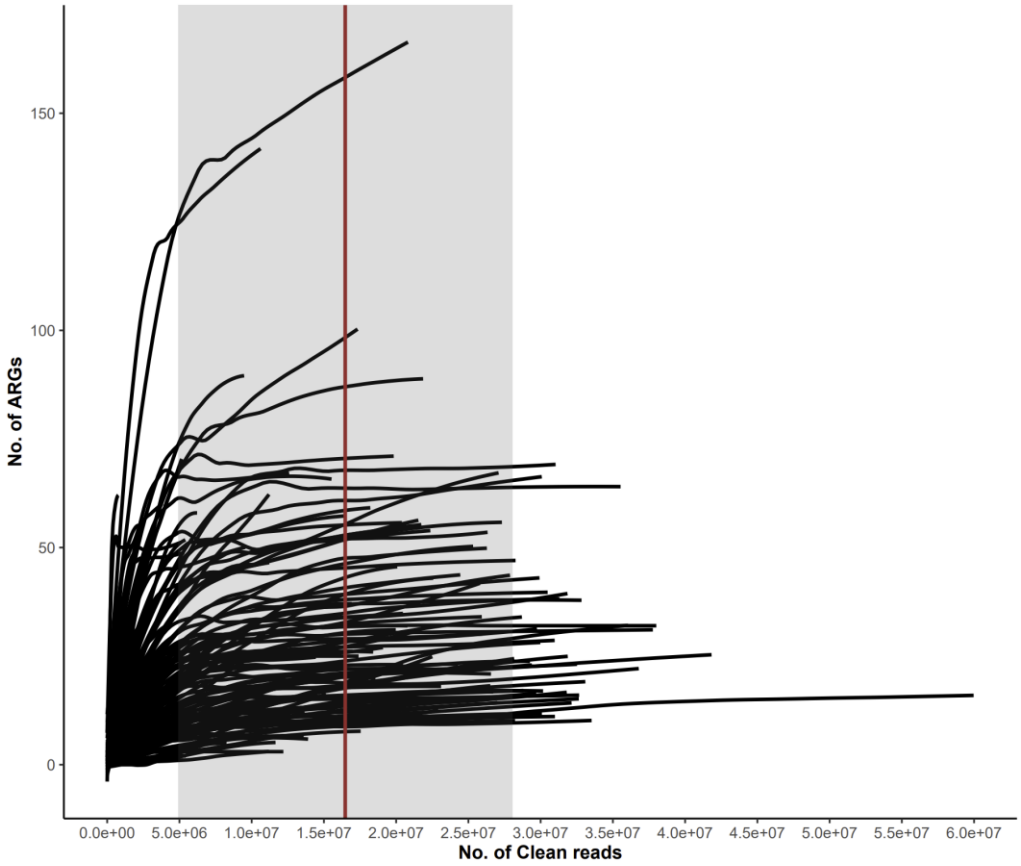
Fig. 6: The resistome variation in nasopharynx is explained by microbiome variation. The ARG and microbial abundance profiles were correlated with each other for antibiotic treated (A) and naive (B) groups using Procrustes analysis. The length of line connecting two points denotes the degree of dissimilarity or distance between the microbiome (filled square) and resistome (filled circle) composition of the same sample. The line and points are colored based on time points in both groups. The correlation coefficients and significance were calculated using the *protest* function in vegan package. (C) Heatmap depicts the Spearman's pairwise correlation coefficients (*r*) between individual ARGs and microbial species from nasopharyngeal samples. Hierarchical clustering using Euclidean distance was performed on both rows and columns. Correlation coefficients (r) was shown only where adjusted p-value < 0.05. (D) Timeline of ARGs associated with *S. marcescens* in the nasopharynx samples of each preterm infant across time points. Each ARG is represented by a different color and shape.



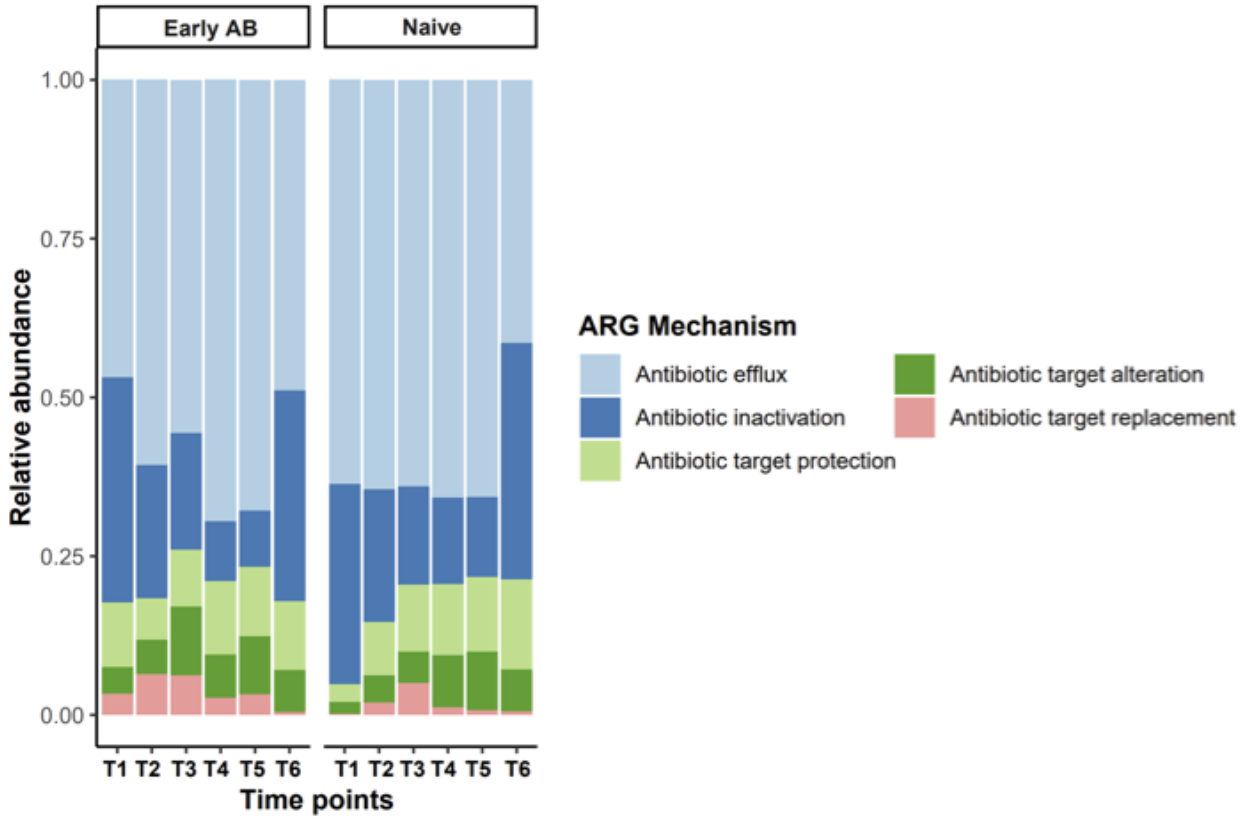
Supplementary Fig. 1: Sample collection and antibiotic exposure for preterm infants in this study. Sampling timeline of 198 nasopharyngeal samples collected from 36 preterm infants over first six months (corrected age), together with information regarding the time of early antibiotic exposure (red and blue colored circles), exposure to subsequent antibiotic treatment (green colored circles) and included/excluded samples. The x-axis represents the days of life (i.e., age) and the y-axis represents the infants. NP = nasopharyngeal; AB = antibiotics.



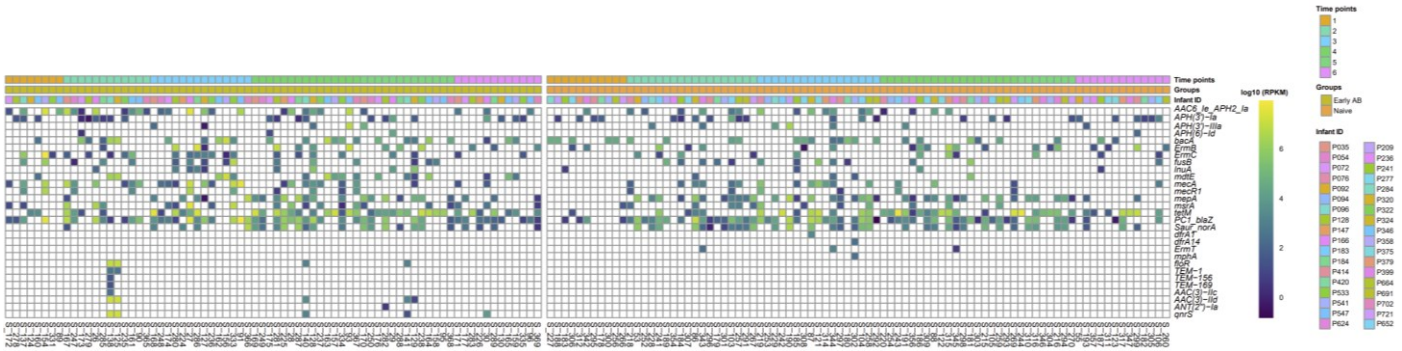
Supplementary Fig. 2: Statistical analysis design used in the study. ARGs = antibiotic resistance genes; PERMANOVA = permutational multivariate analysis of variance.



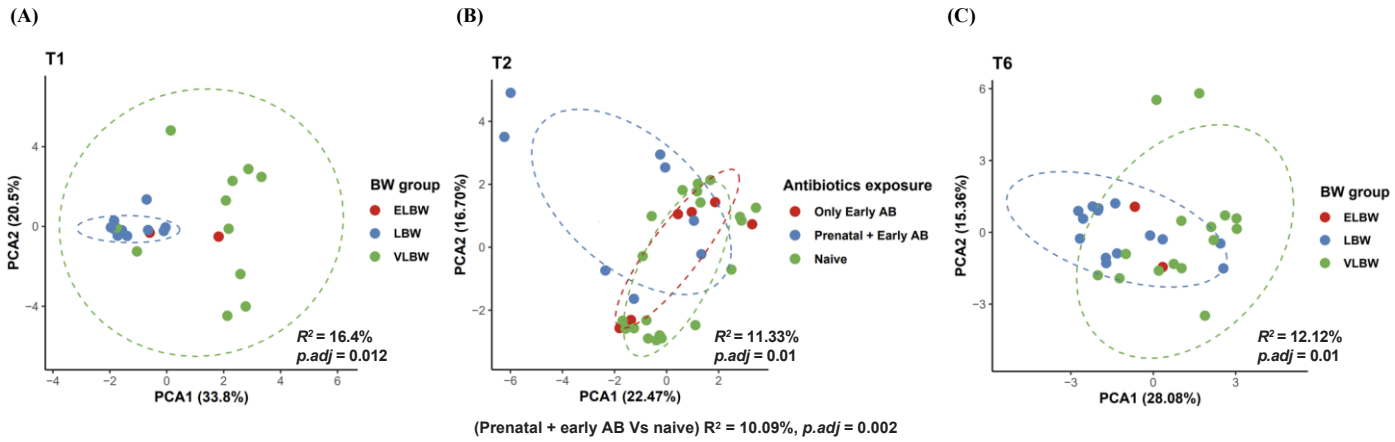
Supplementary Fig. 3: Rarefaction curve for the number of unique ARGs identified at subsampled sequencing depth. The red line indicates the median sequencing depth for all samples, while the gray shading represents one standard deviation.



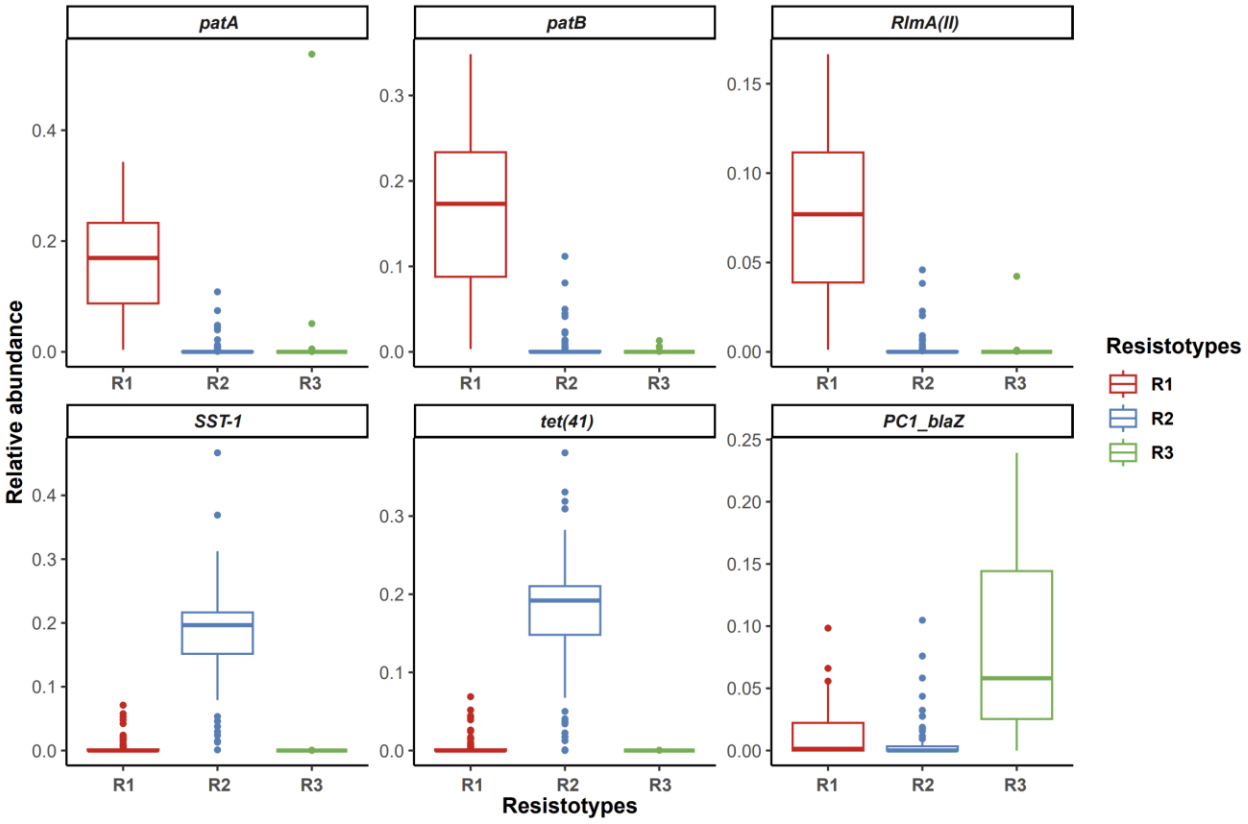
Supplementary Fig. 4: Mean normalized relative abundances of different ARG mechanisms found in the nasopharynx of early antibiotic treated and naive preterm infants at different time points (T1-T6).



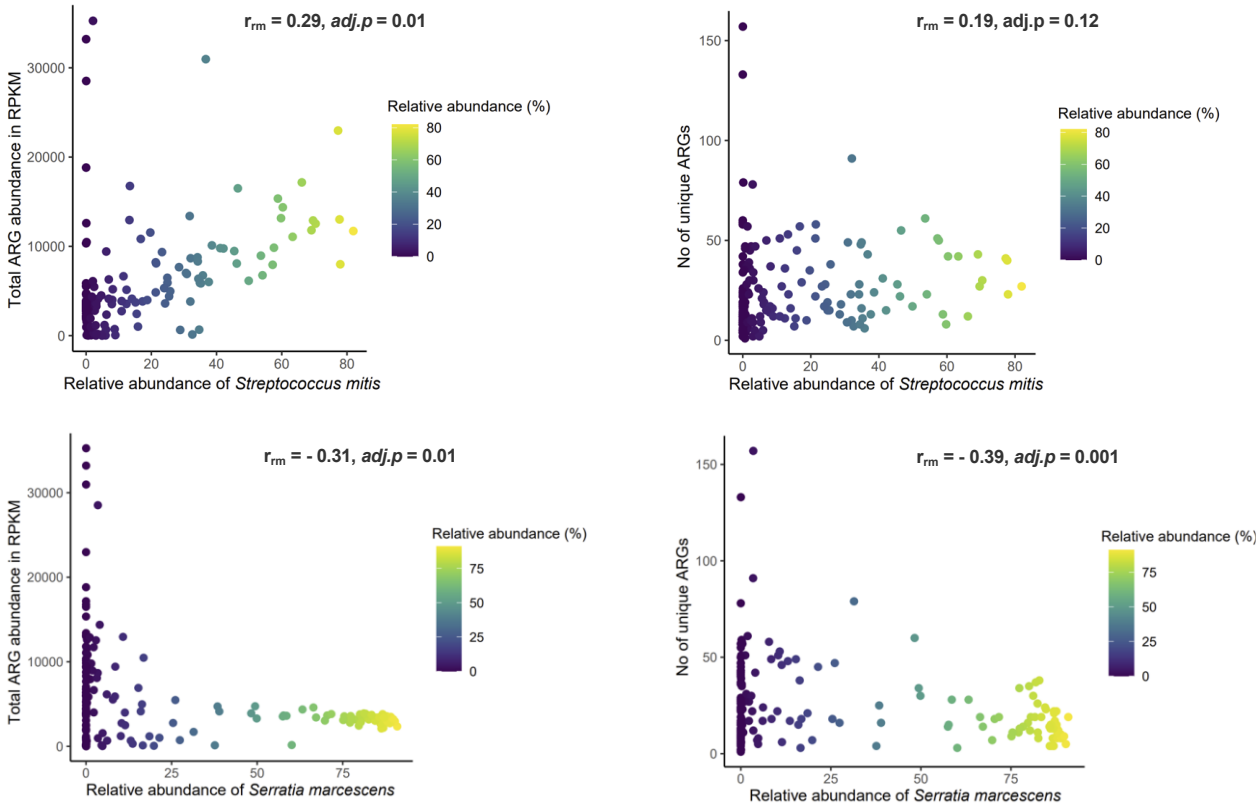
Supplementary Fig. 5: Heatmap showing the log10-transformed RPKM abundance of potential high-risk antibiotic resistance genes (ARGs) identified in the nasopharynx of all preterm infants across samples. The x-axis represents the names of samples, which are arranged by treatment group and time points. RPKM = reads per kilobase of reference gene per million bacterial reads.



Supplementary Fig. 6: PCA visualizations of beta diversity analysis using the Aitchison distances at time points: (A) T1, (B) T2 and (C) T6 (6 month corrected age). Each point corresponds to a sample colored based on birth weight group (A, C) and prenatal and postnatal antibiotic exposure (B). Ellipses indicate 95% confidence intervals (CI). Adjusted p-values and effect size (R^2) are based on PERMANOVA test. CLR = centered-log ratio; BW= Birth Weight.



Supplementary Fig. 7: Box plots of the relative abundances of the main ARG contributors are shown for each Resistotype cluster identified across all the preterm infants' nasopharyngeal resistomes using DMM models. The X-axis displays the three identified clusters, and the Y-axis represents the relative abundance of the ARGs. DMM = dirichlet multinomial mixture.



Supplementary Fig. 8: Scatterplots showing the association between correlated species and overall resistome outcomes across all infant samples from birth to 6 months corrected age. The total ARG abundance in RPKM was positively correlated with the relative abundance of *Streptococcus mitis* (A) and non-linearly negatively associated with the relative abundance of *Serratia marcescens* (C). Plots (B) and (D) depict the relationship between the number of unique ARGs and the relative abundance of *Streptococcus mitis* and *Serratia marcescens*, respectively. Plot (A) and (C) are colored by the relative abundance of *Streptococcus mitis*, while plots (B) and (D) are colored by the relative abundance of *Serratia marcescens*. RPKM = reads per kilobase of reference gene per million bacterial reads.

NP cohort	Naive	Early-antibiotics	p
n	21	15	
Male, n (%)	12 (57%)	11 (73%)	0.484
GA, weeks (mean, SD)	31, 6/7	29 5/7, 1	**<0.001**
BW, grams (mean, SD)	1562, 188	1306, 277	**<0.05 (0.0022)**
BW, group, n (%)			**<0.05 (0.013)**
ELBW	0 (0%)	2 (13%)	
VLBW	8 (38%)	10 (67%)	
LBW	13 (62%)	3 (20%)	
Mode of delivery, n (%)			0.694
Vaginal	4 (19%)	4 (27%)	
C-section	17 (81%)	11 (73%)	
Antibiotics during pregnancy, n (%)			**<0.001**
None	21 (100%)	6 (40%)	
< 10 days	0	7 (47%)	
>= 10 days	0	2 (13%)	
Intrapartum antibiotics (IAP), n (%)			0.50
None documented	2 (10%)	0 (0%)	
Given	19 (90%)	15 (100%)	
ROM, hours (means, SD)	3, 12	432, 733	**<0.05 (0.0108)**
Apgar10 (mean, SD)	9, 1	8, 1	0.3039
Nutrition (discharge to 6 months CA)			**<0.05 (0.039)**
fully breastfed	7 (33%)	4 (29%)	
at least 50% mother milk, but also some formula	3 (14%)	6 (42%)	
mostly formula and some mothers milk	6 (29%)	0 (0%)	
formula	2 (10%)	4 (29%)	
Lost to follow up	3 (14%)	1 (7%)	0.626

Table 1: Baseline characteristics of preterm infants analyzed in this study

Time points	T1	T2	T3	T4	T5	T6	Total
Days of life (DOL)	0-3	4-9	12-19	20-31	32-76	229-288	
Extracted samples (before library prep)	35	37	36	35	23	32	**198**
Exclude samples (Low DNA yield)	2	1	1	0	0	0	**4**
Exclude samples (Low DNA yield) (%)	5.7%	2.7%	2.8%	0%	0%	0%	**2%**
Sequenced samples	33	36	35	35	22	32	**194**
Samples failed during Resistome profiling	5	1	2	0	1	1	**10**
Samples with double sampling time points	0	1	0	1	1	0	**3**
Included in the final analysis	28	34	33	34	21	31	**181**

Supplementary Table 1. Sample flowchart. Flowchart showing number of nasopharyngeal aspirate samples from preterm infants (*n=36*) available at several steps in the study including at sampling, DNA extraction, sequencing and bioinformatics analysis.

Supplementary Table 2: Sequence reads statistics

Sample	Raw reads	Reads after Trimming	Human reads	Remaining (Clean) reads	Reads aligning to CARD database	% of clean reads aligned to CARD
S_103	26080501	26067155	16513464	9553691	17435	0.1825
S_104	22764460	22752830	20748076	2004754	10262	0.5119
S_105	20970229	20964496	3419609	17544887	506	0.0029
S_106	28332496	28321717	15003675	13318042	6043	0.0454
S_119	27475272	27457851	25839804	1618047	105	0.0065
S_120	22266609	22253052	20695679	1557373	82	0.0053
S_121	26201196	26188557	22868983	3319574	11179	0.3368
S_122	33790528	33772729	28966515	4806214	6608	0.1375
S_123	28686631	28676796	1679787	26997009	22095	0.0818
S_124	36678862	36580731	32369824	4210907	330	0.0078
S_125	37839541	37794781	27178195	10616586	302677	2.851
S_127	27688873	27682506	593223	27089283	181147	0.6687
S_128	31005030	30983206	27167769	3815437	15888	0.4164
S_129	28691309	28680175	17478848	11201327	17478	0.156
S_130	33163819	33076882	19478139	13598743	401	0.0029
S_137	31372046	31329501	14342527	16986974	13164	0.0775
S_138	21667917	21662586	846609	20815977	504320	2.4228
S_140	24503529	24497786	2632209	21865577	26203	0.1198
S_141	32823359	32806877	26585210	6221667	27575	0.4432
S_142	26789831	26776365	22169024	4607341	3279	0.0712
S_154	33191448	33139128	27972056	5167072	39874	0.7717
S_156	26884193	26874230	546850	26327380	586046	2.226
S_157	37364064	37340342	32241091	5099251	30575	0.5996
S_158	21680884	21674017	651114	21022903	25474	0.1212
S_159	32047093	32031086	26619128	5411958	4165	0.077

S_160	37292025	37257965	16861312	20396653	12196	0.0598
S_161	45235114	45148815	41399286	3749529	1241	0.0331
S_162	67632600	67539141	62221411	5317730	28064	0.5277
S_163	38901430	38869031	36914900	1954131	22451	1.1489
S_164	44585351	44483261	8951337	35531924	288242	0.8112
S_165	25628648	25540450	10013320	15527130	61843	0.3983
S_166	25531087	25512936	21798951	3713985	3399	0.0915
S_167	23680492	23663333	6734120	16929213	116337	0.6872
S_168	45431318	45401429	35864464	9536965	59284	0.6216
S_169	55896867	55855030	52825510	3029520	9542	0.315
S_170	30624654	30570075	19822065	10748010	26921	0.2505
S_171	27481415	27469324	11777603	15691721	2894	0.0184
S_172	37856150	37836322	25830529	12005793	16131	0.1344
S_173	32620409	32598666	2520144	30078522	144584	0.4807
S_174	29843735	29836353	8305978	21530375	41500	0.1928
S_175	28853583	28843865	8770011	20073854	42924	0.2138
S_176	40254779	40230614	38052826	2177788	4703	0.216
S_177	35925057	35889012	19426456	16462556	18834	0.1144
S_178	20274695	20263845	4157655	16106190	27812	0.1727
S_179	28497251	28482477	685432	27797045	64980	0.2338
S_180	23519660	23515651	1775004	21740647	87325	0.4017
S_181	23879524	23867846	23099772	768074	4218	0.5492
S_182	27132318	27127749	3087081	24040668	45811	0.1906
S_183	31972863	31964880	2266406	29698474	50764	0.1709
S_184	31108627	31101186	1126947	29974239	57650	0.1923
S_185	31622571	31616517	3735586	27880931	60664	0.2176
S_186	46757985	46730502	44193952	2536550	4820	0.19
S_187	30402571	30396836	3688615	26708221	47750	0.1788
S_188	33765529	33756376	1114356	32642020	62286	0.1908
S_189	33083199	33073575	2904430	30169145	57677	0.1912

S_190	30060393	30040023	26669757	3370266	18826	0.5586
S_191	33250280	33222469	16861645	16360824	101214	0.6186
S_193	47659319	47614810	14786449	32828361	56496	0.1721
S_214	52537220	52508963	49589032	2919931	5637	0.1931
S_215	29571723	29543447	8765633	20777814	47796	0.23
S_216	35053190	35033550	6769117	28264433	76971	0.2723
S_217	39517504	39487561	38163679	1323882	625	0.0472
S_218	30675257	30665911	1270181	29395730	56564	0.1924
S_219	27547953	27529357	972788	26556569	111748	0.4208
S_220	33157900	33149181	484150	32665031	66541	0.2037
S_221	21346517	21337082	831108	20505974	31847	0.1553
S_222	26286828	26273899	25340090	933809	919	0.0984
S_223	33944591	33939850	414140	33525710	66865	0.1994
S_224	40663366	40642802	38200609	2442193	2399	0.0982
S_225	31198454	31171888	11201529	19970359	96199	0.4817
S_226	28486798	28471399	21921787	6549612	11262	0.1719
S_227	23681140	23676886	758241	22918645	45401	0.1981
S_228	22840259	22827956	12959279	9868677	18766	0.1902
S_229	32833037	32813316	29804790	3008526	4748	0.1578
S_230	32797647	32777151	27045604	5731547	8726	0.1522
S_234	23368571	23345781	15243726	8102055	2282	0.0282
S_235	29631358	29616146	9188612	20427534	924	0.0045
S_236	24316377	24307110	10727801	13579309	5898	0.0434
S_237	31936082	31906503	3105384	28801119	81934	0.2845
S_238	27183251	27160033	9761467	17398566	61606	0.3541
S_241	21273396	21265742	2856634	18409108	35437	0.1925
S_242	37922507	37895061	34772369	3122692	17337	0.5552
S_243	49645886	49619252	46731003	2888249	14157	0.4902
S_244	66190561	66123529	63015475	3108054	9000	0.2896
S_245	24627154	24616003	12421581	12194422	31173	0.2556

S_247	42327108	42314690	484772	41829918	230967	0.5522
S_248	47197213	47176488	26244459	20932029	30315	0.1448
S_249	51501346	51474520	13469371	38005149	93014	0.2447
S_250	32355912	32328838	13242506	19086332	8192	0.0429
S_251	27317957	27289175	17759907	9529268	914	0.0096
S_252	61163417	61088642	47722474	13366168	36220	0.271
S_253	45402778	45377662	39162898	6214764	9007	0.1449
S_254	32164676	32143130	6335296	25807834	88170	0.3416
S_255	39009857	38993377	25088212	13905165	25987	0.1869
S_256	44877026	44838912	37951047	6887865	9660	0.1402
S_257	34093993	34087358	3603064	30484294	64242	0.2107
S_258	21659477	21650879	1178683	20472196	111693	0.5456
S_259	33243536	33235008	1377165	31857843	84420	0.265
S_26	30270838	30256072	26479542	3776530	14701	0.3893
S_260	35892679	35854245	17290178	18564067	36661	0.1975
S_266	33878209	33871694	778911	33092783	52270	0.1579
S_267	29882106	29867961	3938949	25929012	127336	0.4911
S_268	32997828	32992621	2002198	30990423	54318	0.1753
S_269	48795323	48743359	18996479	29746880	175072	0.5885
S_27	49542796	49483919	35050025	14433894	172882	1.1978
S_270	30056925	30048511	785470	29263041	22344	0.0764
S_271	51738432	51716699	48204756	3511943	5587	0.1591
S_278	29095206	29082097	17430590	11651507	147	0.0013
S_279	29498360	29486657	2167755	27318902	214738	0.786
S_28	37996875	37978461	36894669	1083792	1639	0.1512
S_280	39383489	39340791	14014208	25326583	176136	0.6955
S_281	20907270	20894105	2682760	18211345	106040	0.5823
S_282	31922829	31901011	7462386	24438625	93232	0.3815
S_283	37846534	37832711	1040092	36792619	139836	0.3801
S_284	29863131	29846583	24399378	5447205	67	0.0012

S_285	34228242	34214505	14192796	20021709	73436	0.3668
S_286	41765816	41726214	30517072	11209142	66965	0.5974
S_287	32011138	32000174	2080267	29919907	149226	0.4988
S_288	34869195	34833718	19684842	15148876	58922	0.389
S_289	39559968	39542758	1775686	37767072	105574	0.2795
S_29	38045611	38021180	34100541	3920639	5300	0.1352
S_290	37264718	37245276	26298029	10947247	18539	0.1693
S_291	32072825	32065914	1094390	30971524	56532	0.1825
S_292	63816388	63807235	3838178	59969057	109957	0.1834
S_293	29500665	29493003	5380549	24112454	41039	0.1702
S_294	33573848	33566767	1050194	32516573	68572	0.2109
S_295	29125110	29118244	889487	28228757	46684	0.1654
S_296	34647449	34641389	2501224	32140165	55314	0.1721
S_297	27768840	27758959	1480206	26278753	145877	0.5551
S_298	33622191	33613344	2308277	31305067	72539	0.2317
S_299	22661274	22654821	287602	22367219	66621	0.2979
S_30	28267450	28256137	5690538	22565599	6201	0.0275
S_300	20301912	20296120	296296	19999824	37094	0.1855
S_301	24829638	24825480	540037	24285443	50436	0.2077
S_302	31874316	31867866	833258	31034608	76222	0.2456
S_303	39962863	39916697	18377375	21539322	92682	0.4303
S_304	26868511	26861459	297886	26563573	62060	0.2336
S_305	32243814	32235699	429181	31806518	71735	0.2255
S_306	28410746	28404637	5017057	23387580	48409	0.207
S_307	27609393	27603414	1085057	26518357	56119	0.2116
S_308	32232733	32227923	351812	31876111	63333	0.1987
S_309	26565516	26560390	766174	25794216	57990	0.2248
S_310	36884546	36844010	16364759	20479251	43776	0.2138
S_311	24334581	24314320	1204496	23109824	50555	0.2188
S_312	31069699	31062500	976160	30086340	1325	0.0044

S_313	36354561	36333646	28072594	8261052	178	0.0022
S_314	39811266	39782498	37746037	2036461	205	0.0101
S_315	54249580	54222485	51664787	2557698	7333	0.2867
S_316	44061982	44039723	40399113	3640610	7186	0.1974
S_317	44116478	44095092	42622292	1472800	633	0.043
S_331	29644028	29624401	13828912	15795489	31192	0.1975
S_332	39672217	39644362	38196480	1447882	827	0.0571
S_333	26467120	26441980	13851310	12590670	97356	0.7732
S_334	28679182	28648634	11319308	17329326	105763	0.6103
S_335	29706172	29701470	1533135	28168335	46960	0.1667
S_342	4190563	4185620	997864	3187756	873	0.0274
S_343	26076305	26063824	20907404	5156420	4327	0.0839
S_344	28905903	28893140	19415003	9478137	8856	0.0934
S_345	36832919	36821990	776032	36045958	136753	0.3794
S_346	61044890	61004312	59371452	1632860	2000	0.1225
S_347	69251411	68623791	64379680	4244111	1229	0.029
S_354	27433235	27422512	16225329	11197183	53	0.0005
S_356	27229735	27221424	4719264	22502160	86767	0.3856
S_357	27546962	27535322	18225311	9310011	27877	0.2994
S_364	64520057	64429899	62001636	2428263	979	0.0403
S_365	24055590	24045067	10406711	13638356	24613	0.1805
S_366	55484020	55321634	52415053	2906581	4562	0.157
S_367	28470266	28444380	8614772	19829608	96588	0.4871
S_368	20964782	20957580	7515333	13442247	25918	0.1928
S_369	31720586	31709062	3007642	28701420	42042	0.1465
S_53	34434233	34414250	32019563	2394687	7072	0.2953
S_55	26470453	26463326	7666968	18796358	2006	0.0107
S_86	28423053	28408167	23773990	4634177	13285	0.2867
S_87	66938909	66905387	64680249	2225138	6394	0.2874
S_88	30976709	30958246	4982552	25975694	72811	0.2803

S_89	25774477	25760044	24970857	789187	46	0.0058
S_90	46594816	46555570	45060476	1495094	395	0.0264
S_91	36674226	36641370	35795529	845841	1172	0.1386
S_93	53178273	53147304	51682818	1464486	2822	0.1927
S_95	33533344	33517467	32593464	924003	1464	0.1584
S_96	26810898	26796813	22625616	4171197	492	0.0118

Supplementary Table 3: PERMANOVA results

Univariate analysis					
Factor	**Df**	**SumOfSqs**	**R2**	**F**	**p.adj**
Early antibiotics	1	3574	0.02283	4.1813	0.001
Individual ID	35	59041	0.37705	2.5075	0.001
Time points (age groups)	5	12832	0.08195	3.1241	0.001
Mode of Delivery	1	600	0.00383	0.6884	0.756
Gender	1	1856	0.01185	2.1468	0.02
Birth weight (BW) group	2	6393	0.04083	3.7885	0.001
Days of life (DOL)	1	2466	0.01575	2.8645	0.005
Post-menstrual age (PMA)	10	15136	0.09666	1.8191	0.001
Gestational age (GA) at birth	3	6313	0.04032	2.4787	0.001
Prenatal maternal antibiotics duration (categorized)	2	8162	0.05212	4.8942	0.001
Prenatal maternal antibiotics	1	7008	0.04475	8.386	0.001
Early antibiotics duration	1	4640	0.02963	5.4662	0.001
Ruptured membranes duration	1	1692	0.0108	1.9551	0.046
Apgar score	1	1412	0.00902	1.6285	0.082
Intrapartum maternal antibiotics	1	2311	0.01476	2.6813	0.007

Multivariate analysis					
Factor	**Df**	**SumOfSqs**	**R2**	**F**	**p.adj**
Individual ID	31	47028	0.30033	2.255	0.001
Early antibiotics	1	947	0.00605	1.1577	0.31
Early antibiotics duration	1	2009	0.01283	2.4664	0.009
Time points (age groups)	5	13213	0.08438	3.4531	0.001
Post-menstrual age (PMA)	10	15261	0.09746	1.959	0.001
Prenatal maternal antibiotics	1	5700	0.0364	6.9391	0.001
Prenatal maternal antibiotics duration (categorized)	2	6961	0.04445	4.25	0.001