



HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2020 March 09.

Published in final edited form as:

Nat Microbiol. 2019 December ; 4(12): 2285–2297. doi:10.1038/s41564-019-0550-2.

Metagenomic signatures of early life hospitalization and antibiotic treatment in the infant gut microbiota and resistome persist long after discharge

Andrew J. Gasparri¹, Bin Wang^{1,2}, Xiaoqing Sun^{1,2}, Elizabeth A. Kennedy¹, Ariel Hernandez-Leyva¹, I. Malick Ndao³, Phillip I. Tarr^{3,4}, Barbara B. Warner³, Gautam Dantas^{1,2,4,5,*}

¹The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, Saint Louis, MO, USA

²Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, MO, USA

³Department of Pediatrics, Washington University in St. Louis School of Medicine, St. Louis, MO, USA

⁴Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, Saint Louis, MO, USA

⁵Department of Biomedical Engineering, Washington University in St. Louis, Saint Louis, MO, USA

Summary

Because hospitalized preterm infants are vulnerable to infection, they receive frequent and often prolonged exposures to antibiotics. It is not known if the short-term effects of antibiotics on the preterm infant gut microbiota and resistome persist after discharge from neonatal intensive care units. Here, we use complementary metagenomic, culture based, and machine learning techniques

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to G.D. dantas@wustl.edu.

Author Contributions

A.J.G and G.D. conceived and designed the study. P.I.T, B.B.W, and I.M.N. assembled the cohorts, collected the biologics and maintained the database, oversaw transfer of biologics and clinical metadata, provided clinical insight. A.J.G. and B.W. extracted metagenomic DNA from stools and prepared shotgun metagenomic sequencing libraries. A.J.G., B.W., and A.H-L. performed stool culturing experiments and isolate genomic DNA extraction. A.J.G. and B.W. prepared isolate genome sequencing libraries. E.A.K. performed *Enterococcus* phenotyping experiments. X.S. created functional metagenomic libraries, performed functional selections, and prepared functional metagenomic sequencing libraries. A.J.G. analyzed clinical metadata, shotgun metagenomic sequencing data, isolate genome sequencing data, and functional metagenomic data. A.J.G. wrote the manuscript with input from G.D., B.B.W., and P.I.T.

Competing interests

P.I.T. is a member of the Scientific Advisory Board of, holds equity in, and is a consultant to MediBeacon, Inc. P.I.T. is a coinventor on a patent application to test intestinal permeability in humans which might generate royalty payments. This involvement is not directly relevant to this manuscript.

Data Availability

Assembled functional metagenomic contigs, shotgun metagenomic reads, shotgun genomic reads, and assemblies have been deposited to NCBI GenBank and SRA under BioProject ID PRJNA489090.

Code availability

The software packages used in this study are free and open source. Analysis scripts employing these packages (and associated usage notes) are available from the authors upon request.

to interrogate the gut microbiota and resistome of antibiotic-exposed preterm infants, during and after hospitalization, and compare these readouts to antibiotic-naïve healthy infants sampled synchronously. We find a persistently enriched gastrointestinal antibiotic resistome, prolonged carriage of multidrug resistant *Enterobacteriaceae*, and distinct antibiotic-driven patterns of microbiota and resistome assembly in extremely preterm infants who received early life antibiotics. The collateral damage of early life antibiotic treatment and hospitalization in preterm infants is long-lasting. We urge development of strategies to reduce these consequences in highly vulnerable neonatal populations.

Introduction

Gut microbes play important roles in host health and disease throughout life, particularly in infancy¹. Infant gut microbiota (IGM) assembly accelerates in the first months of life, following inoculation by organisms from mothers and the environment², but stabilizes by approximately three years of age³. Antibiotics in this interval may disproportionately damage the host-microbiota ecosystem^{3–5}. Indeed, emerging data suggest that early-life gut microbial alterations correlate with chronic metabolic and immune disorders later in life^{1,4,6–16}, including allergies¹⁷, psoriasis¹⁸, adiposity¹⁹, diabetes²⁰ and inflammatory bowel disease^{21–23}. For most of these disorders, a causal link between antibiotic-mediated microbiota disruption and onset of pathology is lacking. However, antibiotics in infancy are associated with permanent immune alterations^{18,24} and inflammatory bowel disease in childhood²³, highlighting the damaging long-term potential of early life antibiotic treatment.

Over 11% of live births worldwide occur preterm²⁵, and preterm birth and its sequelae are prominent causes of childhood morbidity and mortality worldwide²⁶. Because bacterial infections are frequent complications of preterm birth²⁷, 79% of very low birthweight and 87% of extremely low birthweight infants in US NICUs receive antibiotics within three days of birth²⁹. The gastrointestinal tracts of even healthy infants harbor a diverse antibiotic resistome³⁰, which is shaped by factors including antibiotics, diet, and environment^{31–33}. Preterm IGM perturbation immediately following antibiotic treatment is characterized by decreased alpha diversity, increased *Enterobacteriaceae* abundance, and antibiotic-specific enrichment of antibiotic resistance genes (ARGs) and multidrug resistant organisms (MDROs)³⁴. Because microbiota perturbation during infancy may be disproportionately damaging^{35,36}, it is imperative to study the lasting effects of antibiotics and hospitalization on the preterm IGM. Prior studies of preterm infants report IGM recovery concomitant with NICU discharge^{37–39}. However, these studies rely on culture or amplicon sequencing (e.g. 16S rRNA) based analysis, which focus on taxa in the microbiota rather than the functions they collectively encode. Here, we analyze ~1.2 terabases of metagenomic DNA from 437 infant stools, culture and sequence 530 bacterial isolates, and functionally select 300 gigabases of metagenomic DNA for antibiotic resistance, to investigate the long-term consequences of antibiotic treatment on the preterm IGM.

Metagenomic based analysis of the effect of antibiotics on the preterm IGM

To understand the long-term effects of prematurity and associated early life hospitalization and antibiotic therapy on the IGM, we performed whole metagenome shotgun sequencing of 437 fecal samples from 58 infants over the first 21 months of life (Supplementary Fig. 1; Supplementary Fig. 2). Our cohort included 41 preterm infants sampled in the NICU at St. Louis Children's Hospital and following discharge to home. One subset of this cohort (n=9) received scant antibiotic therapy neonatally (each received a single concurrent course of gentamicin and ampicillin for <7 days). The remaining 32 preterm infants received extensive antibiotics over the first 21 months (median (interquartile range (IQR)) 8 courses (6,10.3) and 29.5 days (41.63,68.3) antibiotic therapy). All infants in this cohort were classified as being born preterm ((median (IQR) gestational age at birth of 26 weeks (25,27)) with very low birth weights (median (IQR) 840 g (770,960)). Additionally, we included 17 antibiotic-naïve, healthy early term⁴⁰ or late preterm⁴¹ (median (IQR) gestational age at birth 36 weeks (36,37 weeks); "near-term") infants of the same chronological age range, sampled synchronously with the preterm cohort.

We inferred bacterial taxonomic composition using MetaPhlAn⁴². Across all infants, Shannon diversity increased during a developmental phase before stabilizing (Fig. 1a). In near-term infants, microbiota Shannon diversity increases rapidly in the first month before plateauing, while preterm infant microbiota diversity increases more gradually and with greater variation (Fig. 1a). *Enterobacteriaceae* and *Enterococcaceae* dominate the preterm IGM in the first months of life. In contrast, early colonization by *Enterobacteriaceae* in near-term infants precedes robust colonization with *Bifidobacteriaceae* (Fig. 1b; Supplementary Fig. 3). *Enterococcaceae* is significantly less abundant early in life (<4 months chronological age) in near-term compared to preterm infants ($p<0.001$, Wilcoxon), and *Prevotellaceae* is similarly less abundant later in infancy (>8 months chronological age) in preterm infants compared to near-term infants ($p<1\times 10^{-10}$, Wilcoxon). Despite these differences, we observed predicted microbiota functional stability both over time and between groups (Fig. 1c), as inferred by HUMAnN2⁴³. While it is likely that greater variation exists when finer functional category or predicted hosts of functions are taken into account, the invariability of microbiota functional capacity at this high level suggests that while prematurity, early life hospitalization, and antibiotic treatment drastically perturb taxonomic composition of the microbiota, a core set of microbial functions remains conserved across hosts⁴⁴.

Low gut microbiota diversity is often associated with adverse health in infants^{45–47}, children⁵, and adults⁴⁸. To identify features associated with microbiota diversity, we regressed Shannon diversity on clinical variables (Methods, Supplementary Table 1) using a generalized linear mixed model with subject defined as individual effect. All variables in Supplementary Table 1 were included in initial modeling, and a final model was fit via backwards elimination of variables. After correcting for multiple comparisons, day of life was significantly associated with increased Shannon diversity ($p<0.001$), while recent (within 30 days of sample collection) administration of vancomycin ($p<0.001$), ampicillin ($p<0.001$), meropenem ($p=0.009$), or cefepime ($p=0.012$) was significantly associated with decreased diversity (Fig. 1d). No clinical variable included in the model other than antibiotic treatment was significantly associated with IGM diversity. The sparse model explained 57%

of the variance in Shannon diversity by the fixed effects (day of life and antibiotic treatments) alone, and an additional 12% by subject. The magnitude of the model estimate for day of life (0.002) was substantially less than those for recent antibiotics (vancomycin, -0.34; ampicillin, -0.60; meropenem, -0.42; cefepime, -0.46; oxacillin, -0.70). Thus, a single recent course of antibiotics has an effect on diversity of the same magnitude as the diversity increase observed over ~5–12 months of life. Across all infants in our cohort through the first 110 days of life, Shannon diversity of the microbiota was significantly lower in infants who received >1 course of antibiotics in the prior month (Fig. 1e). Accordingly, recent antibiotic treatment appears to be a key driver of microbiota diversity early in life.

Partial microbiota recovery following NICU discharge

While the taxonomic composition of the preterm infant microbiota clustered by both gestational age at birth and antibiotic treatment status (Adonis, $p < 0.001$, Bray–Curtis), chronological age was a major driver of microbiota composition across all infants (Fig. 2a). We hypothesized that after observed early-life perturbation, the composition of the preterm microbiota would converge towards that of age-matched healthy, antibiotic-naïve, near-term infants within the first 21 months of life, but that microbiota ‘scars’ from this early-life disruption (e.g., enriched ARGs and MDROs) would persist.

To quantify the extent of this perturbation, we used random forests to regress the relative abundances of species in the microbiota of infants against their chronological age as previously described⁵¹. The minimum number of variables required for accurate prediction was 50 (Fig. 2b). We trained a model consisting of the 50 most informative predictors on the antibiotic-naïve, near-term infant subset to model healthy microbiota development, and subsequently refined and validated the model. The top age discriminatory taxa in the IGM of antibiotic-naïve, near-term infants were *Faecalibacterium prausnitzii*, *Subdoligranulum* sp., *Ruminococcus gnavus*, and *Oscillobacter* sp. (Fig. 2c). We used this sparse model to predict infant chronological age using the relative abundance of these 50 species. This prediction, or ‘microbiota age,’ approximates relative microbiota maturity⁵¹. We observed a linear relationship between the chronological and microbiota ages of antibiotic-naïve, near-term infants, suggesting that the model accurately predicts near-term infant age. For preterm infants, however, predicted microbiota ages were younger than chronological age across several stages of development, indicating that microbiota development is disrupted in these infants. To better quantify the extent of disruption, we computed a microbiota for age Z-score (MAZ) for each metagenome, as previously described⁵¹. Using a Z-score to compare age bins is necessary because this value reflects the variance of predicted age across infant microbiota development. Preterm infants who receive antibiotic treatment have significantly lower MAZs than near-term infants in the first months of life (Fig. 2e). However, by months 12–15 of life, the MAZs of hospitalized preterm infants closely resemble those of healthy, antibiotic-naïve, near-term infants (Fig. 2f). Thus, despite transient delays in the development of the preterm IGM, the bacterial taxonomic composition converges on common structures with those of healthy, antibiotic-naïve, infants within the first 21 months of life (Fig. 2d).

Antibiotic resistome of preterm IGM

We next characterized the antibiotic resistome encoded in the IGM of our cohort. We conducted functional metagenomic analysis⁵³ of 217 preterm and term infant stools, selected to encompass the diversity in clinical variables in our cohort. We constructed 22 functional metagenomic libraries totaling 396 gigabase pairs (Supplementary Fig. 4; Supplementary Table 1; see Methods) with an average insert size of 2–3 kilobase pairs, selected libraries on sixteen antibiotics relevant to infants and children (Supplementary Table 3), and recovered resistant transformants for each antibiotic except meropenem (Supplementary Fig. 4). We found that the infant gut metagenome encoded transferrable resistance even to antibiotics rarely or never used in neonates, such as ciprofloxacin and chloramphenicol, and those that represent last lines of defense against MDROs, such as tigecycline and colistin. Only one of eight libraries constructed from stools of antibiotic-naïve, near-term infants encoded ciprofloxacin resistance (mediated by loci other than *gyrA* or *parC*), compared to six out of fourteen libraries constructed from preterm infant stools. This observation, given the scarce use of ciprofloxacin in neonates²⁸, suggests either that acquired ciprofloxacin resistance occurs naturally in preterm infant gut bacterial communities, or that organisms resistant to ciprofloxacin are co-selected by other antibiotics to which they are resistant.

We sequenced resistance-conferring metagenomic inserts and assembled 874 unique ARGs. The median identity of these functionally-selected ARGs to the NCBI nr database was 94.4%, while their median identity to the Comprehensive Antibiotic Resistance Database (CARD)⁵⁴ was 32.0% (Fig. 3a). Hence, while most resistance determinants discovered in our functional selections have been previously sequenced, they have frequently not been assigned resistance functions, a discordance we have previously noted³⁴. Functionally-selected ARGs with low identity to CARD, while not canonical resistance genes widespread in the clinical setting at this point, represent candidates for or progenitors of clinical resistance genes given opportunity, mobilization, or evolution^{53,58}. The predicted sources of resistance conferring ORFs (determined by best BLAST hit to the NCBI non-redundant protein database) were predominantly uncultured bacteria or *Enterobacteriaceae* (Fig. 3b). The identification of *Enterobacteriaceae* as likely hosts of ARGs in the IGM is consistent with current understanding of *Enterobacteriaceae* as prolific hosts and traffickers of ARGs^{55–57}. Additionally, the identification of uncultured (189 ORFs) and unclassified (24 ORFs) bacteria as sources of ARGs highlights the value of functional metagenomics as a culture- and sequence-unbiased method for characterizing resistomes⁵⁸.

Highlighting the potential for lateral ARG exchange within the infant microbiome, 225 contigs (6.4% of all contigs) recovered in functional selections encoded a mobile genetic element (MGE, Supplementary Fig. 5a–e). MGEs were most commonly observed in tetracycline selections (Supplementary Fig. 5f), but were also commonly observed in β -lactam, chloramphenicol, gentamicin, and ciprofloxacin selections. We observed enrichment for MGEs on amoxicillin/clavulanate ($p < 0.01$, hypergeometric test), tetracycline ($p < 0.01$, hypergeometric test), and gentamicin ($p < 0.001$, hypergeometric test). The synteny of functionally-selected ARGs with MGEs suggests the possibility of a mobilizable resistome in the IGM.

We extended our resistome analysis using ShortBRED⁵⁹ to quantify translated ARG abundance in all sequenced metagenomes with a custom database that included all ARGs from CARD as well as functionally-selected ARGs identified here. Resistomes clustered according to gestational age at birth and antibiotic treatment status (Fig. 3c, $p < 0.001$, Adonis). The gut metagenomes of preterm infants encoded fewer unique ARGs than those of near-term infants ($p < 0.01$, Wilcoxon, Fig. 3d). However, the cumulative resistome relative abundance was significantly higher in the IGM of preterm infants with early-plus-subsequent antibiotic treatment compared to preterm infants with early-only antibiotic treatment and to antibiotic-naïve, near-term infants ($p < 0.05$, Wilcoxon, Fig. 3e). There was a weak inverse correlation between taxonomic alpha diversity and cumulative resistome burden across all metagenomes ($R^2 = 0.09$, Supplementary Fig. 6a), indicating that resistome-enriched microbiota are dominated by a few species. Indeed, in 41 of the 54 metagenomes with a cumulative resistome of reads per kilobase per million mapped reads (RPKM) > 5000 , a single species comprised $> 50\%$ of microbiota relative abundance. In 25 of these samples, the dominant species was *E. coli* (Supplementary Fig. 6b). Other dominant species were *Enterococcus faecalis* ($n = 5$), *Klebsiella pneumoniae* ($n = 2$), *Staphylococcus epidermidis* ($n = 2$), *Enterobacter aerogenes* ($n = 2$), *Bifidobacterium breve* ($n = 2$), *Pseudomonas aeruginosa*, *Bifidobacterium longum*, and *Citrobacter koseri* ($n = 1$ each). Thus, it appears that extreme prematurity, associated hospitalization, and antibiotic treatment select for one or two MDROs that dominate the IGM rather than enriching for a greater diversity of resistant organisms.

To define the developmental progression of the infant gastrointestinal resistome over the first months of life, we regressed the abundance of ARGs in a subset of antibiotic-naïve, near-term infant gut metagenomes against the day of life for these infants using random forests⁵². We constructed a sparse model using the 50 most informative ARGs. The sparse model was subsequently applied to preterm samples to predict ‘resistome age.’ A clear developmental trajectory based on these 50 ARGs was evident in near-term infants (Fig. 3g). The developmental trajectory of the preterm infant gut resistome deviates from that of the antibiotic-naïve, near-term infants in prolonged carriage of some ARGs (e.g., *oqxA*, *oqxB*, *catI*, *fosA5*, *cdeA*), near absence of others (e.g., *abeM*), and a general increase in the normalized abundance of these genes in the gut across all timepoints (Fig. 3g). Overall, we found that the model only modestly predicted the chronological age of preterm infants ($R^2 = 0.62$, Fig. 3f), suggesting that distinct patterns of resistome development emerge based on antibiotic treatment status and gestational age at birth.

Persistence of multidrug resistant Enterobacteriaceae in the preterm IGM

Whole metagenome shotgun sequencing is a powerful method for describing gross microbiota composition and function but is less well equipped to elucidate strain level variation. The gut has been established as an early reservoir of bacteria that cause late onset bloodstream infections in neonates⁶⁰ and is dominated by multidrug resistant (MDR) *Proteobacteria*³⁴, but the extent to which these early colonizing strains persist in the IGM is poorly defined. We hypothesized that early life hospitalization and antibiotic treatment in preterm infants might create a gastrointestinal niche for such *Proteobacteria* that is not relinquished after discharge from the NICU. To better understand the persistence of specific

bacterial strains in the microbiota of infants in our cohort, we cultured pairs of stools collected 8–10 months apart from 15 infants (nine preterm and six near-term) on a series of selective agars (see Methods). We optimized culture conditions to isolate opportunistic extraintestinal pathogens known to be highly prevalent and abundant in the IGM as well as those that are frequently MDR. In total, we cultured 530 isolates from these 30 samples. We whole genome sequenced, assembled, and annotated 277 and 253 isolates from the preterm and near-term sets, respectively.

The species most frequently isolated by this direct selection were *E. coli* (n=139), *K. pneumoniae* (n=62), *E. faecalis* (n=50), *Enterobacter cloacae* (n=42), *E. faecium* (n=22), *C. freundii* (n=15), and *K. oxytoca* (n=14). We identified within-infant persistence of nearly identical strains of *E. coli*, *E. cloacae*, and *K. variicola* in samples collected from both preterm and near-term infants. These highly similar, persistent isolate pairs from preterm infants included isolates from samples collected both while in the NICU and following discharge (Fig. 4). Among the persistent isolates recovered were strains of *E. coli* ST405 and *E. cloacae* ST108, both of which are high-risk lineages known to encode extended-spectrum β -lactamases and NDM-family carbapenemases^{61–63}. Each of the *E. coli* strains encoded a TEM-1 β -lactamase as well as an aac(3)-IId aminoglycoside acetyltransferase with predicted resistance to aminoglycosides, and each *E. cloacae* strain encoded an AmpC type β -lactamase. The *K. variicola* strains each encoded *oqxAB*, the RND-type multidrug efflux pump⁶⁴, and the chromosomal *Klebsiella* β -lactamase *blaOKP-B-1*⁶⁵ (Supplementary Table 4). We isolated nearly identical MDR *Enterobacteriaceae*, as suggested by average nucleotide identity >99.997% (Fig. 4b,d,f) and core gene single nucleotide polymorphism distances (Supplementary Table 4), from the preterm IGM both in the NICU and following discharge. These data support an enduring and transmissible pathological microbiome “scar” associated with preterm birth, early life hospitalization, and antibiotic treatment.

Because *Enterococcus* species are prevalent and abundant in the preterm infant gut³⁴, often MDR⁶⁶, and cause nosocomial blood stream infections in preterm infants⁶⁷, we investigated their resistance and virulence phenotypes. A particular concern among hospitalized populations is vancomycin resistant *Enterococci* (VRE)⁶⁸. Of the 15 unique *Enterococcus* strains we isolated, ten were *E. faecalis* and five were *E. faecium* (Supplementary Fig. 7a). No *E. faecalis*, and two *E. faecium* isolates were resistant to vancomycin. However, no *E. faecium* strain formed a biofilm, while four of the *E. faecalis* strains formed robust biofilms at room temperature, and an additional six formed biofilms at 37°C (Supplementary Fig. 7b). Interestingly, all biofilm forming strains were isolated from preterm infant stool. This is consistent with the prevailing understanding that early colonizers of the preterm infant gut are largely surface adapted strains that are prevalent in the NICU environment⁶⁹. *E. faecalis* biofilm formers, while susceptible to vancomycin when planktonic, were resistant to this antibiotic when in biofilms (Supplementary Fig. 7c). Thus, despite the apparent tradeoff between vancomycin resistance and biofilm formation observed among *Enterococcus* strains, nearly all have evolved strategies for surviving vancomycin treatment. This is concerning given widespread usage of vancomycin (Table 1) and prevalence of *Enterococcus* colonization (Fig. 1b) in the NICU.

Persistent metagenomic signature of antibiotic treatment in premature infants

To understand if prematurity, hospitalization, and antibiotic treatment have persistent effects on gut microbial content and function, we sought to identify metagenomic features that distinguish post-NICU discharge samples in preterm infants from age-matched samples from antibiotic-naïve, near-term infants. We used a supervised learning approach to classify samples as originating from a hospitalized preterm infant (including both early-only and early-plus-subsequent antibiotic treatment groups) or an antibiotic-naïve near-term infant residing at home, based on the relative abundance of bacterial taxa and ARGs in their IGM. Using a support vector machine, we identified the fifteen most informative features and constructed a model consisting of only these variables, which correctly classified all preterm and 15 of the 17 near-term samples (96.4% accuracy, Fig. 5a). Of the fifteen variables most important to model performance, six were ARGs and nine were bacterial taxa (Fig. 5b). The ARGs important to classification were the class A β -lactamase *cfxA670*, and five genes functionally selected on piperacillin or tetracycline, respectively. The highest identity BLAST hit of four of the functionally-selected ARGs was an ABC transporter, while the other was a MATE family efflux transporter. The predictive species were members of the order *Clostridiales* (*Eubacterium rectale*, *Ruminococcus obeum*, *R. lactaris*, *Dorea formicigenerans*, *E. ventriosum*, *E. ramulus*, *E. eligens*) and *Bacteroidales* (*Prevotella copri*, *Barnesiella intestinihominis*). Our model accurately identified if a preterm infant was hospitalized and received early life antibiotic treatment based on metagenome composition following NICU discharge despite high level architectural recovery.

Conclusion

By combining metagenomic sequencing, selective and differential stool culture paired with isolate sequencing, functional metagenomics, and machine learning, we demonstrate persistent metagenomic signatures of early life antibiotic treatment and hospitalization in preterm infants. This is manifest in an enriched gut resistome and persistent carriage of MDR *Enterobacteriaceae*, despite apparent recovery in microbiota maturity. Regardless of prematurity or antibiotic exposure, we observed little variation in the functional capacity of the microbiota, albeit when metagenomic reads are binned in broad functional categories. Our work highlights the need to integrate sequencing- and culture-based approaches for interrogating microbiota to reveal underappreciated effects of perturbations. These complementary methods provide data supporting a persistent metagenomic signature of early life hospitalization and antibiotic treatment associated with prematurity in the dynamic microbial community housed in the infant gut.

We were unable to isolate the effects of antibiotics from those of other adverse early life events coinciding with prematurity, such as extended hospitalization and illness. While an interventional study to probe these variables in neonates is infeasible, future animal studies could provide important insights into their relative contributions. Additionally, it is probable that yet to be defined environmental variables play a role in the co-development of the immune system and the microbiota, which would need to be addressed in future studies.

Despite these caveats, we supply compelling evidence for the underappreciated lasting effect of prematurity and associated hospitalization and antibiotic treatment on the microbiome. These perturbations may play a role in chronic pathologies associated with prematurity for which the etiology is unclear. From a clinical standpoint, our findings emphasize a necessity for alternatives to broad-spectrum antimicrobial therapy for managing infection in the NICU. This should entail therapeutic approaches such as narrow-spectrum antibiotics and probiotic therapies⁷¹, but also improved accuracy and speed of diagnostics to reduce unnecessary courses of antibiotics. It is unclear if these results are generalizable across NICUs. Future multicenter studies are important to reveal the effect of neonatal antibiotic stewardship practices in IGM development. While the metagenomic scars we identified may be implicated in sequelae of preterm birth such as neurodevelopmental^{72–74}, metabolic^{75,76}, cardiac^{77,78}, and respiratory^{79,80} defects, further experiments with model systems including gnotobiotic animals are needed to link these enduring dysbioses and lasting pathologies.

Methods

Sample and metadata collection

All samples and patient metadata used in this study were collected as part of the Neonatal Microbiome and Necrotizing Enterocolitis Study (P.I.T., P.I.) or the St. Louis Neonatal Microbiome Initiative (B.B.W., P.I.) at Washington University School of Medicine and approved by the Human Research Protection Office (approval numbers 201105492 and 201104267, respectively). Samples were obtained from infants after parents provided informed consent. Because very few hospitalized preterm infants are antibiotic naïve, we stratified our cohort for sample analysis by antibiotic exposure and gestational age at birth, with a subset of individuals with early antibiotic exposure only (N=9) with no antibiotic exposure outside the first week of life, a subset of individuals with early and subsequent antibiotic exposure (N=32), and a subset of late preterm or early term infants (N=17) who were not hospitalized and were antibiotic-naïve over the first months of life (Table 1). All stools produced were collected and stored as previously described^{1,2}. In total, 437 samples collected longitudinally from 58 infants were shotgun sequenced and included in all metagenomic analysis.

Metagenomic DNA extraction

Metagenomic DNA was extracted from approximately 100 mg of stool samples using the PowerSoil DNA Isolation Kit (MoBio Laboratories) following the manufacturer's protocol with the following modification: samples were lysed by two rounds of two minutes of bead beating at 2.5k oscillations per minute for 2 minutes followed by 1 minute on ice and 2 additional minutes of beadbeating using a Mini-Beadbeater-24 (Biospec Products). DNA was quantified using a Qubit fluorometer dsDNA BR Assay (Invitrogen) and stored at -20°C .

Metagenomic sequencing library preparation

Metagenomic DNA was diluted to a concentration of 0.5 ng/ μL prior to sequencing library preparation. Libraries were prepared using a Nextera DNA Library Prep Kit (Illumina) following the modifications described in Baym *et al*, 2015³. Libraries were purified using

the Agencourt AMPure XP system (Beckman Coulter) and quantified using the Quant-iT PicoGreen dsDNA assay (Invitrogen). For each sequencing lane, 10 nM of approximately 96 samples were pooled three independent times. These pools were quantified using the Qubit dsDNA BR Assay and combined in an equimolar fashion. Samples were submitted for 2×150 bp paired-end sequencing on an Illumina NextSeq High-Output platform at the Center for Genome Sciences and Systems Biology at Washington University in St. Louis with a target sequencing depth of 2.5 million reads per sample.

Rarefaction analysis

To determine the appropriate sequencing depth necessary to fully characterize infant gut microbiota composition and function, 17 representative metagenomes that were sequenced most deeply were subsampled at the following read depths: 8000000, 7000000, 6000000, 5000000, 4000000, 3000000, 2000000, 1000000, 100000, and 10000. Subsampled metagenomes were profiled using MetaPhlan 2.0⁴ to determine species richness at each depth. Rarefaction was only used to establish an appropriate sequencing depth, and subsampled metagenomes were not used for any downstream analyses.

Metagenome profiling

Prior to all downstream analysis, Illumina paired-end reads were binned by index sequence. Adapter and index sequences were trimmed and sequences were quality filtered using Trimmomatic v0.36⁵ using the following parameters: *java -Xms2048m -Xmx2048m -jar trimmomatic-0.33.jar PE -phred33 ILLUMINACLIP:NexteraPE-PE.fa:2:30:10:1:true SLIDINGWINDOW:6:10 LEADING:13 TRAILING:13 MINLEN:36*. Relative abundance of species was calculated using MetaPhlan 2.0⁴ (repository tag 2.2.0). Relative abundance tables were merged using the *merge_metaphlan_tables.py* script. Abundance of metabolic pathways was determined using HUMAnN2⁶. Raw count values were normalized for sequencing depth, collapsed by ontology, and tables were merged using the *humann2_renorm_table*, *humann2_regroup_table*, and *humann2_join_tables* utility scripts.

Construction of metagenomic libraries from infant gut samples for functional selection

We constructed 22 functional metagenomic libraries by pooling metagenomic DNA from 9–10 stools per library, encompassing 396 gigabase pairs (Gb) of metagenomic DNA with an average library size of 18 Gb (Supplementary Fig. 4; Supplementary Table 1) and an average insert size of 2–3 kilobase pairs (kb). Approximately 5 µg purified extracted total metagenomic DNA was used as starting material for metagenomic library construction. To create small-insert metagenomic libraries, DNA was sheared to a target size of 3,000bp using the Covaris E210 sonicator following manufacturer's recommended settings (http://covarisinc.com/wp-content/uploads/pn_400069.pdf). Sheared DNA was concentrated by QIAquick PCR Purification Kit (Qiagen) and eluted in 30 µl nuclease-free H₂O. Then the purified DNA was size-selected by using BluePippin instrument (Sage Science) to a range of 1000–6000 bp DNA fragment through a premade 0.75% Pippin gel cassette. Size selected DNA was then end-repaired using the End-It DNA End Repair kit (Epicentre) with the following protocol

1. Mix the following in a 50 μ l reaction volume: 30 μ l of purified DNA, 5 μ l dNTP mix (2.5 mM), 5 μ l 10X End-Repair buffer, 1 μ l End-Repair Enzyme Mix and 4 μ l nuclease-free H₂O.
2. Mix gently and incubate at room temperature for 45 min.
3. Heat-inactivate the reaction at 70°C for 15 min.

End-repaired DNA was then purified using the QIAquick PCR purification kit (Qiagen) and quantified using the Qubit fluorometer BR assay kit (Life Technologies) and ligated into the pZE21-MCS-1 vector at the HincII site. The pZE21 vector was linearized at the HINCII site using inverse PCR with PFX DNA polymerase (Life Technologies)

1. Mix the following in a 50 μ l reaction volume: 10 μ l of 10X PFX reaction buffer, 1.5 μ l of 10 mM dNTP mix (New England Biolabs), 1 μ l of 50 mM MgSO₄, 5 μ l of PFX enhancer solution, 1 μ l of 100 pg μ l 21 circular pZE21, 0.4 μ l of PFX DNA polymerase, 0.75 μ l forward primer (5' GAC GGT ATC GAT AAG CTT GAT 3'), 0.75 μ l reverse primer (5' GAC CTC GAG GGG GGG 3') and 29.6 μ l of nuclease free H₂O to a final volume of 50 μ l.
2. PCR cycle temperature as follows: 95°C for 5 min, then 35 cycles of [95°C for 45 s, 55°C for 45 s, 72°C for 2.5 min], then 72°C for 5 min.

Linearized pZE21 was size-selected (~2,200bp) on a 1% low melting point agarose gel (0.5X TBE) stained with GelGreen dye (Biotium) and purified by QIAquick Gel Extraction Kit (Qiagen). Pure vector was dephosphorylated using calf intestinal alkaline phosphatase (CIP, New England BioLabs) by adding 1/10th reaction volume of CIP, 1/10th reaction volume of New England BioLabs Buffer 3, and nuclease-free H₂O to the vector elute and incubating at 37°C overnight before heat inactivation from 15 min at 70°C. End-repaired metagenomic DNA and linearized vector were ligated together using the Fast-Link Ligation Kit (Epicentre) at a 5:1 ratio of insert:vector using the following protocol

1. Mix the following in a 15 μ l reaction volume: 1.5 μ l 10X Fast-Link buffer, 0.75 μ l ATP (10 mM), 1 μ l FastLink DNA ligase (2 U/ μ l), 5:1 ratio of metagenomic DNA to vector, and nuclease-free H₂O to final reaction volume.
2. Incubate at room temperature overnight.
3. Heat inactivate for 15 min at 70°C.

After heat inactivation, ligation reactions were dialyzed for 30 min using a 0.025 μ m cellulose membrane (Millipore catalogue number VSWP09025) and the full reaction volume used for transformation by electroporation into 25 μ l *E. coli* MegaX (Invitrogen) according to manufacturer's recommended protocols (http://tools.invitrogen.com/content/sfs/manuals/megax_man.pdf). Cells were recovered in 1 ml Recovery Medium (Invitrogen) at 37°C for one hour. Libraries were titered by plating out 0.1 μ l and 0.01 μ l of recovered cells onto Luria–Bertani (LB) agar plates containing 50 μ g/ml kanamycin. For each library, insert size distribution was estimated by gel electrophoresis of PCR products obtained by amplifying the insert from 36 randomly picked clones using primers flanking the HincII site of the multiple cloning site of the pZE21 MCS1 vector (which contains a selectable marker for kanamycin resistance). The average insert size across all libraries was determined to be 3 kb,

and library size estimates were calculated by multiplying the average PCR-based insert size by the number of titered colony forming units (CFUs) after transformation recovery. The rest of the recovered cells were inoculated into 50 ml of LB containing 50 µg/ml kanamycin and grown overnight. The overnight culture was frozen with 15% glycerol and stored at –80°C for subsequent screening.

Functional selections for antibiotic resistance

Each metagenomic library was selected for resistance to each of 16 antibiotics (at concentrations listed in Supplementary Table 3 plus 50 µg/ml kanamycin for plasmid library maintenance) was performed using LB agar. Of note, as our library host, *E. coli*, is intrinsically resistant to vancomycin, we are unable to functionally screen for loci conferring resistance to this antibiotic. Further, the use of kanamycin as the selective marker for the metagenomic plasmid library results in low-level cross-resistance with other aminoglycoside antibiotics, resulting in a higher required minimum inhibitory concentration for gentamicin. For each metagenomic library, the number of cells plated on each antibiotic selection represented 10x the number of unique CFUs in the library, as determined by titers during library creation. Depending on the titer of live cells following library amplification and storage, the appropriate volume of freezer stocks were either diluted to 100 µl using MH broth + 50 µg/ml kanamycin or centrifuged and reconstituted in this volume for plating. After plating (using sterile glass beads), antibiotic selections were incubated at 37°C for 18 hours to allow the growth of clones containing an antibiotic resistance conferring DNA insert. Of the 352 antibiotic selections performed, 296 yielded antibiotic-resistant *E. coli* transformants (Supplementary Fig. 4). After overnight growth, all colonies from a single antibiotic plate (library by antibiotic selection) were collected by adding 750 µl of 15% LB-glycerol to the plate and scraping with an L-shaped cell scraper to gently remove colonies from the agar. The slurry was then collected and this process was repeated a second time for a total volume of 1.5 mL to ensure that all colonies were removed from the plate. The bacterial cells were then stored at –80°C before PCR amplification of antibiotic-resistant metagenomic fragments and Illumina library creation.

Amplification and sequencing of functionally-selected fragments

Freezer stocks of antibiotic-resistant transformants were thawed and 300 µl of cells pelleted by centrifugation at 13,000 revolutions per minute (r.p.m.) for two minutes and gently washed with 1 mL of nuclease-free H₂O. Cells were subsequently pelleted a second time and re-suspended in 30 µl of nuclease-free H₂O. Re-suspensions were then frozen at –20°C for one hour and thawed to promote cell lysis. The thawed re-suspension was pelleted by centrifugation at 13,000 r.p.m. for two minutes and the resulting supernatant was used as template for amplification of resistance-conferring DNA fragments by PCR with Taq DNA polymerase (New England BioLabs)

1. Mix the following for a 25 µl reaction volume: 2.5 µl of template, 2.5 µl of ThermoPol reaction buffer (New England BioLabs), 0.5 µl of 10 mM deoxynucleotide triphosphates (dNTPs, New England Biolabs), 0.5 µl of Taq polymerase (5 U/µl), 3 µl of a custom primer mix, and 16 µl of nuclease-free H₂O.

2. PCR cycle temperature as follows: 94°C for 10 min, then 25 cycles of [94°C for 45 s, 55°C for 45 s, 72°C for 5.5 min], then 72°C for 10 min.

The custom primer mix consisted of three forward and three reverse primers, each targeting the sequence immediately flanking the HincII site in the pZE21 MCS1 vector, and staggered by one base pair. The staggered primer mix ensured diverse nucleotide composition during early Illumina sequencing cycles and contained the following primer volumes (from a 10 mM stock) in a single PCR reaction: (primer F1, CCGAATTCATTAAAGAGGAGAAAG, 0.5 µl); (primer F2, CGAATT CATTAAAGAGGAGAAAGG, 0.5 µl); (primer F3, GAATTCATTAAAGAGGAGAAAGGTAC, 0.5 µl); (primer R1, GATATCAAGCTTATCGATACCGTC, 0.21 µl); (primer R2, CGATATCAAGCTTATCGATACCG, 0.43 µl); (primer R3, TCGATATCAAGCTTATCGATACC, 0.86 µl). The amplified metagenomic inserts were then cleaned using the Qiagen QIAquick PCR purification kit and quantified using the Qubit fluorometer HS assay kit (Life Technologies).

For amplified metagenomic inserts from each antibiotic selection, elution buffer was added to PCR template for a final volume of 200 µl and sonicated in a half-skirted 96-well plate on a Covaris E210 sonicator with the following setting: duty cycle, 10%; intensity, 5; cycles per burst, 200; sonication time, 600s. Following sonication, sheared DNA was purified and concentrated using the MinElute PCR Purification kit (Qiagen) and eluted in 20 µl of pre-warmed nuclease-free H₂O. In the first step of library preparation, purified sheared DNA was end-repaired

1. Mix the following for a 25 µl reaction volume: 20 µl of elute, 2.5 µl T4 DNA ligase buffer with 10 mM ATP (10X, New England BioLabs), 1 µl dNTPs (1 mM, New England BioLabs), 0.5 µl T4 polymerase (3 U/µl, New England BioLabs), 0.5 µl T4 PNK (10 U/µl, New England BioLabs), and 0.5 µl Taq Polymerase (5 U/µl, New England BioLabs).
2. Incubate the reaction at 25°C for 30 min followed by 20 min at 75°C.

Next, to each end-repaired sample, 5 µl of 1 µM pre-annealed, barcoded sequencing adapters were added (adapters were thawed on ice). Barcoded adapters consisted of a unique 7-bp oligonucleotide sequence specific to each antibiotic selection, facilitating the de-multiplexing of mixed-sample sequencing runs. Forward and reverse sequencing adapters were stored in TES buffer (10 mM Tris, 1 mM EDTA, 50 mM NaCl, pH 8.0) and annealed by heating the 1 µM mixture to 95°C followed by a slow cool (0.1°C per second) to a final holding temperature of 4°C. After the addition of barcoded adapters, samples were incubated at 16°C for 40 min and then for 10 min at 65°C. Before size-selection, 10 µl each of adapter-ligated samples were combined into pools of 12 and concentrated by elution through a MinElute PCR Purification Kit (Qiagen), eluting in 14 µl of elution buffer (10 mM Tris-Cl, pH 8.5). The pooled, adaptor-ligated, sheared DNA was then size-selected to a target range of 300–400 bp on a 2% agarose gel in 0.5X TBE, stained with GelGreen dye (Biotium) and extracted using a MinElute Gel Extraction Kit (Qiagen). The purified DNA was enriched using the following protocol

1. Mix the following for a 25 μ l reaction volume: 2 μ l of purified DNA, 12.5 μ l 2X Phusion HF Master Mix (New England BioLabs), 1 μ l of 10 mM Illumina PCR Primer Mix (5'-AAT GAT ACG GCG ACC ACC GAG ATC-3' and 5'-CAAGCAGA A GAC GGC ATA CGA GAT-3'), and 9.5 μ l of nuclease-free H₂O.
2. PCR cycle as follows: 98°C for 30 s, then 18 cycles of [98°C for 10s, 65°C for 30 s, 72°C for 30s], then 72°C for 5 min.

Amplified DNA was measured using the Qubit fluorometer HS assay kit (Life Technologies) and 10 nM of each sample were pooled for sequencing. Subsequently, samples were submitted for paired-end 101-bp sequencing using the Illumina Next Seq platform at the DNA Sequencing and Innovation Lab at the Edison Center for Genome Sciences and Systems Biology, Washington University in St Louis, USA). In total, three sequence runs were performed at 10 pM concentration per lane.

Assembly and annotation of functionally-selected fragments

Illumina paired-end sequence reads were binned by barcode (exact match required), such that independent selections were assembled and annotated in parallel. Assembly of the resistance-conferring DNA fragments from each selection was achieved using PARFuMS⁷ (Parallel Annotation and Reassembly of Functional Metagenomic Selections), a tool developed specifically for the high-throughput assembly and annotation of functional metagenomic selections.

Open reading frames (ORFs) were predicted in assembled contigs using MetaGeneMark⁸ and annotated by searching amino acid sequences against Pfam, TIGRfam, and an ARG specific profile hidden Markov model (pHMM) database, Resfams⁹ (<http://www.dantaslab.org/resfams>), with HMMER3¹⁰. MetaGeneMark was run using default gene-finding parameters while *hmmScan* (HMMER3) was run with the option *--cut_ga* as implemented in the script *annotate_functional_selections.py*. Selections were excluded from analysis if (a) more than 200 contigs were assembled or (b) the number of contigs assembled exceeded the number of colonies on the selection plate by a factor of ten. Further, assembled contigs less than 500 bp were discarded. Since many assembled contigs include multiple annotated ORFs, the subset of proteins considered causative resistance determinants for downstream analysis were classified using the following hierarchical scheme. First, if a contig encoded a protein with a 100% amino acid identity hit to the CARD database¹¹, it was considered the causative resistance determinant on that contig. Next, if a contig encoded a protein with a significant hit to a Resfams pHMM using profile specific gathering thresholds, it was considered the causative resistance determinant on that contig. In absence of a high scoring hit to the CARD or Resfams databases, contigs were manually curated to identify plausible resistance determinants on an antibiotic specific basis. The rationale for this hierarchical classification scheme was to first identify perfect matches to known resistance determinants via BLAST to CARD (with a threshold of 100% amino acid identity), and subsequently identify variants of known resistance determinants using Resfams pHMMs. Using these criteria, 1184 of the 5658 unique predicted proteins (20.9%) were classified as resistance determinants.

The percent identity of all resistance determinants were determined using BlastP¹² query against both the NCBI non-redundant protein database (retrieved May 21, 2018) and the CARD¹¹ database (version 1.2.1, retrieved January 24, 2018). Once the top local alignment was identified with BlastP, it was used for a global alignment using the Needleman-Wunsch algorithm as implemented in the *needle* program of EMBOSS¹³ version 6.6.0 as previously described¹⁴.

Putative mobile genetic elements were identified on functionally-selected contigs based on string matches to one of the following keywords in Pfam and TIGRFam annotations: 'transposase', 'transposon', 'conjugative', 'integrase', 'integron', 'recombinase', 'conjugal', 'mobilization', 'recombination', or 'plasmid'.

Quantification of antibiotic resistance genes in metagenomes

Relative abundance of antibiotic resistance genes was calculated using ShortBRED³¹. Causative resistance determinants, as identified using the hierarchical annotation scheme described above, were used as proteins of interest for identification of marker families using *shortbred_identify.py*. These proteins included all antibiotic resistance genes in CARD (version 1.2.1, retrieved January 24, 2018)⁵⁴ and antibiotic resistance proteins identified using functional metagenomic selections performed in the current study. Thus, the custom ShortBRED database included markers to canonical antibiotic resistance determinants as well as resistance determinants functionally identified in this study which are most relevant to the infant gut microbiota. In order to calculate relative abundance of resistance genes in metagenomes, *shortbred_quantify.py* was used.

Bacterial isolation from infant stools

Approximately 50 mg of frozen stool was resuspended in 1 mL Tryptic Soy Broth (TSB) and incubated with shaking at 37°C for four hours. 50 µL of culture was streaked for isolation using the four quadrant method on each of the following agars: Bile Esculin Agar, ESBL Agar, MacConkey Agar, MacConkey Agar+cefotaxime, MacConkey Agar +ciprofloxacin, and Blood Agar (Hardy Diagnostics catalog numbers G12, G321, G35, G121, G258, A10, respectively). Plates were incubated for 18–24 hours at 37°C. Four colonies of each distinct morphology on each plate were sub-streaked onto blood agar and incubated for 18–24 hours at 37°C. Following confirmation of morphology, a 1 mL TSB was inoculated with a single colony and grown overnight at 37°C with shaking. Overnight cultures were frozen in 15% glycerol in TSB.

Genomic DNA isolation

1.5 mL TSB was inoculated from isolate glycerol stocks and grown overnight at 37°C with shaking. DNA was extracted using the BiOstic Bacteremia DNA Isolation Kit (MoBio Laboratories) following manufacturer's protocols. Genomic DNA was quantified using a Qubit fluorometer dsDNA BR Assay (Invitrogen) and stored at –20°C.

Isolate sequencing library preparation

Isolate sequencing libraries were prepared in the same manner as described for metagenomic sequencing libraries, following the protocol described in Baym *et al.*, 2015. For each

sequencing lane, 10 nM of approximately 300 samples were pooled three independent times. These pools were quantified using the Qubit dsDNA BR Assay and combined in an equimolar fashion. Samples were submitted for 2×150 bp paired-end sequencing on an Illumina NextSeq High-Output platform at the Center for Genome Sciences and Systems Biology at Washington University in St. Louis with a target sequencing depth of 1 million paired end reads per sample.

Assembly of isolate genomes

Prior to all downstream analysis, Illumina paired end reads were binned by index sequence. Adapter and index sequences were trimmed using Trimmomatic v0.36⁵ using the following parameters: *java -Xms2048m -Xmx2048m -jar trimmomatic-0.33.jar PE -phred33 ILLUMINACLIP: NexteraPE-PE.fa:2:30:10:1:true*. Contaminating human reads were removed using DeconSeq¹⁵ and unpaired reads were discarded. Reads were assembled using SPAdes¹⁶ with the following parameters: *spades.py -k 21,33,55,77 -careful*. Contigs less than 500 bp were excluded from further analysis. Assembly quality was assessed using QUAST¹⁷. Average coverage across the assembly was calculated by mapping raw reads to contigs using bbmap (<https://jgi.doe.gov/data-and-tools/bbtools/>).

Isolate genomic analysis

A total of 406 assemblies had an N50 greater than 50,000 and fewer than 500 total contigs longer than 1000 bp and were included in further analysis. Genomes were annotated using Prokka¹⁸ with default parameters. Multilocus sequence types were determined using *in silico* MLST (<https://github.com/tseemann/mlst>). Species assignments were determined by querying assemblies against a RefSeq sketch using Mash identifying RefSeq hit with the minimum Mash distance¹⁹. Assemblies were binned by species according to Mash designation. For each of the seven most commonly occurring species, pangenome analysis was performed using Roary, with core genome alignments created with PRANK²⁰. An outgroup assembly of the same genus but different species was downloaded from NCBI and included in each pangenome analysis (Supplementary Table 2). Maximum likelihood core genome phylogenies were constructed using RAXML under the GTRGAMMA model with 1000 bootstraps and maximum likelihood optimization initialized from a random starting tree. Average nucleotide identities were computed using pyani (<https://github.com/widdowquinn/pyani>). Pairwise single nucleotide polymorphism distances were calculated from core genome alignments generated by Roary using snp-dists (<https://github.com/tseemann/snp-dists>). Resistance genes were annotated via nucleotide blast to the resfinder database (<https://bitbucket.org/genomicepidemiology/resfinder/src/master/README.md>).

Enterococcus biofilm formation assay

Mid-log phase cultures in freshly-prepared tryptic soy broth containing 0.5% glucose (TSBG) were diluted to OD 0.1. 200ul of the diluted culture was added in quadruplicate to 96 well polystyrene plates and incubated at room temperature or 37°C without shaking. After 24 hours of growth, wells were decanted, washed three times with sterile PBS, and fixed for 30 minutes with 200 µl Bouin's solution. Fixative was removed by washing three times with sterile PBS, then wells were stained with 0.1% crystal violet for 30 minutes. Excess stain was removed by washing three times with sterile PBS, the stain was solubilized

in 200ul ethanol, and absorbance read at 590nm. *E. faecalis* strains TX5682 (biofilm negative) and TX82 (biofilm positive) were used as controls²¹.

***Enterococcus* vancomycin susceptibility testing**

Isolates identified as *Enterococcus* were phenotyped for vancomycin resistance using microbroth dilution according to the CLSI guidelines. ATCC29212 (vancomycin susceptible) and ATCC51299 (vancomycin-resistant) were included in all assays as controls. Isolates were grown to mid-log phase, diluted in culture media to 1×10^6 CFU/ml, and used to inoculate plates containing vancomycin ranging from 128–2ug/ml. After 24 hours of static growth at 37°C, optical density was read at 600 nm and MIC was determined by scoring by eye for turbidity.

Vancomycin resistance of biofilms was assayed after establishing biofilms as above. After 24 hours of static growth at 37°C, planktonic cells were removed by washing three times with sterile water, and then 200 µl of TSBG containing 5 mg/ml, 5 µg/ml, or no vancomycin, and the plates incubated at 37°C for an additional 24 hours. After washing planktonic cells three times with sterile water, 200 µl sterile water was added to each well and the viability of the cells in the biofilm was assessed using an XTT Cell Viability Kit (Cell Signaling Technology, #9095) according to manufacturer's protocols, reading absorbance at 450nm 60 minutes after addition of reagents.

Generalized linear mixed model of microbiota diversity

To model the effect of clinical variables on microbiota diversity, a generalized linear mixed model was fit by maximum likelihood using the lme4 package in R. All variables in Supplementary Table 1 were included in initial modeling, and a final model was fit via backwards elimination of variables. Pseudo-R² was determined using r.squaredGLMM function in the MuMin package. P values were corrected for multiple hypotheses using the glht function in the multcomp (lincfit = mcp(tension = 'Tukey')).

Microbiota age regression using Random Forests

Random Forests was used to regress the relative abundances of all species predicted by MetaPhlan2 in infant stool samples against their chronological age using the R package "randomForest" as previously described²². The default parameters were used with the following exceptions: *ntree*=10,000, *importance*=TRUE. Fivefold cross-validation was performed using the *rfcv* function over 100 iterations to estimate the minimum number of features needed to accurately predict microbiota age. The features most important for prediction were identified over 100 iterations of the *importance* function, and a sparse model consisting of the 50 most important features was constructed and trained on a set of nine antibiotic-naïve near-term infants randomly selected from the larger near-term infant set. This model was validated in the remaining eight antibiotic-naïve near-term infants, and then applied preterm infants to predict microbiota age. Microbiota for age Z-score was computed as previously described²². This allowed for comparisons of microbiota maturity between age bins as the metric accounts for differing variance in predicted microbiota age throughout infant development.

Classification of post-discharge samples

A single sample from each individual was selected (the final post-discharge sample collected from each preterm infant and a roughly age matched sample from each near-term infant). All metagenomic data (species and ARG abundances, centered and scaled) were initially used as input for logistic regression, k-nearest neighbor, support vector machine, naïve bayes, and random forests classifiers. Ultimately, a support vector machine as implemented in the R package e1071 was selected as it was both the highest performing and most parsimonious classifier. Feature importance was determined by computing the elementwise absolute value of the matrix of weights by the matrix of support vectors. A sparse model was subsequently constructed consisting of only the fifteen most important features.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported in part by awards to G.D through National Institute of General Medical Sciences of the National Institutes of Health (R01 GM099538), National Institute of Allergy and Infectious Diseases of the National Institutes of Health (R01 AI123394), the US Centers for Disease Control and Prevention (200–2016-91955), to P.I.T. through National Institutes of Health (5P30 DK052574 [Biobank, DDRCC]), to G.D., P.I.T., and B.B.W. through the Eunice Kennedy Shriver National Institute Of Child Health and Human Development of the National Institutes of Health (R01 HD092414), and to P.I.T and B.B.W. through the Children’s Discovery Institute at St. Louis Children’s Hospital and Washington University School of Medicine, A.J.G. received support from a NIGMS training grant through award number T32 GM007067 (Jim Skeath, Principal Investigator) and from the NIDDK Pediatric Gastroenterology Research Training Program under award number T32 DK077653 (P.I.T., Principal Investigator). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors thank members of the Dantas lab for helpful discussion of the manuscript, and the Edison Family Center for Genome Sciences & Systems Biology staff, Eric Martin, Brian Koebe, and Jessica Hoisington-López for technical support and sequencing expertise.

References

1. Sommer F & Bäckhed F The gut microbiota — masters of host development and physiology. *Nat. Rev. Microbiol* 11, 227–238 (2013). [PubMed: 23435359]
2. Pantoja-Feliciano IG et al. Biphase assembly of the murine intestinal microbiota during early development. *ISME J.* 7, 1112–1115 (2013). [PubMed: 23535917]
3. Yatsunenko T et al. Human gut microbiome viewed across age and geography. *Nature* 486, 222–227 (2012). [PubMed: 22699611]
4. Cox LM et al. Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. *Cell* 158, 705–721 (2014). [PubMed: 25126780]
5. Abrahamsson TR et al. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin. Exp. Allergy* 44, 842–850 (2014). [PubMed: 24330256]
6. Livanos AE et al. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat. Microbiol* 1, 16140 (2016). [PubMed: 27782139]
7. Cho I et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* 488, 621–626 (2012). [PubMed: 22914093]
8. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK & Knight R Diversity, stability and resilience of the human gut microbiota. *Nature* 489, 220–230 (2012). [PubMed: 22972295]
9. Trasande L et al. Infant antibiotic exposures and early-life body mass. *Int. J. Obes* 37, 16–23 (2013).
10. Hviid A, Svanstrom H & Frisch M Antibiotic use and inflammatory bowel diseases in childhood. *Gut* 60, 49–54 (2011). [PubMed: 20966024]

11. Penders J et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. *Gut* 56, 661–667 (2007). [PubMed: 17047098]
12. Arrieta M-C et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med* 7, 307ra152–307ra152 (2015).
13. Ahmadizar F et al. Early life antibiotic use and the risk of asthma and asthma exacerbations in children. *Pediatr. Allergy Immunol* 28, 430–437 (2017). [PubMed: 28423467]
14. Stokholm J et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun* 9, 141 (2018). [PubMed: 29321519]
15. Missaghi B, Barkema H, Madsen K & Ghosh S Perturbation of the Human Microbiome as a Contributor to Inflammatory Bowel Disease. *Pathogens* 3, 510–527 (2014). [PubMed: 25438009]
16. Tremlett H et al. Gut microbiota in early pediatric multiple sclerosis: a case–control study. *Eur. J. Neurol* 23, 1308–1321 (2016). [PubMed: 27176462]
17. Russell SL et al. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep.* 13, 440–447 (2012). [PubMed: 22422004]
18. Zanvit P et al. Antibiotics in neonatal life increase murine susceptibility to experimental psoriasis. *Nat. Commun* 6, 8424 (2015). [PubMed: 26416167]
19. Azad MB, Bridgman SL, Becker AB & Kozyrskyj AL Infant antibiotic exposure and the development of childhood overweight and central adiposity. *Int. J. Obes* 38, 1290–1298 (2014).
20. Boursi B, Mamtani R, Haynes K & Yang Y-X The effect of past antibiotic exposure on diabetes risk. *Eur. J. Endocrinol* 172, 639–648 (2015). [PubMed: 25805893]
21. Shaw SY, Blanchard JF & Bernstein CN Association Between the Use of Antibiotics in the First Year of Life and Pediatric Inflammatory Bowel Disease: *Am. J. Gastroenterol* 105, 2687–2692 (2010). [PubMed: 20940708]
22. Ungaro R et al. Antibiotics Associated With Increased Risk of New-Onset Crohn’s Disease But Not Ulcerative Colitis: A Meta-Analysis. *Am. J. Gastroenterol* 109, 1728–1738 (2014). [PubMed: 25223575]
23. Kronman MP, Zaoutis TE, Haynes K, Feng R & Coffin SE Antibiotic Exposure and IBD Development Among Children: A Population-Based Cohort Study. *PEDIATRICS* 130, e794–e803 (2012). [PubMed: 23008454]
24. Lexmond WS et al. Involvement of the iNKT Cell Pathway Is Associated With Early-Onset Eosinophilic Esophagitis and Response to Allergen Avoidance Therapy. *Am. J. Gastroenterol* 109, 646–657 (2014). [PubMed: 24513807]
25. Blencowe H et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet* 379, 2162–2172 (2012).
26. Liu L et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet* 388, 3027–3035 (2016).
27. Stoll BJ Neurodevelopmental and Growth Impairment Among Extremely Low-Birth-Weight Infants With Neonatal Infection. *JAMA* 292, 2357 (2004). [PubMed: 15547163]
28. Hsieh E et al. Medication Use in the Neonatal Intensive Care Unit. *Am. J. Perinatol* 31, 811–822 (2013). [PubMed: 24347262]
29. Flannery DD et al. Temporal Trends and Center Variation in Early Antibiotic Use Among Premature Infants. *JAMA Netw. Open* 1, e180164 (2018). [PubMed: 30646054]
30. Rose G et al. Antibiotic resistance potential of the healthy preterm infant gut microbiome. *PeerJ* 5, e2928 (2017). [PubMed: 28149696]
31. Rahman SF, Olm MR, Morowitz MJ & Banfield JF Machine Learning Leveraging Genomes from Metagenomes Identifies Influential Antibiotic Resistance Genes in the Infant Gut Microbiome. *mSystems* 3, (2018).
32. Pärnänen K et al. Maternal gut and breast milk microbiota affect infant gut antibiotic resistome and mobile genetic elements. *Nat. Commun* 9, 3891 (2018). [PubMed: 30250208]

33. Hourigan SK et al. Comparison of Infant Gut and Skin Microbiota, Resistome and Virulome Between Neonatal Intensive Care Unit (NICU) Environments. *Front. Microbiol* 9, 1361 (2018). [PubMed: 29988506]
34. Gibson MK et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol* 1, 16024 (2016). [PubMed: 27572443]
35. Fouhy F et al. High-Throughput Sequencing Reveals the Incomplete, Short-Term Recovery of Infant Gut Microbiota following Parenteral Antibiotic Treatment with Ampicillin and Gentamicin. *Antimicrob. Agents Chemother* 56, 5811–5820 (2012). [PubMed: 22948872]
36. Greenwood C et al. Early Empiric Antibiotic Use in Preterm Infants Is Associated with Lower Bacterial Diversity and Higher Relative Abundance of Enterobacter. *J. Pediatr* 165, 23–29 (2014). [PubMed: 24529620]
37. Stewart CJ et al. Preterm gut microbiota and metabolome following discharge from intensive care. *Sci. Rep* 5, 17141 (2015). [PubMed: 26598071]
38. Zwitink RD et al. Association between duration of intravenous antibiotic administration and early-life microbiota development in late-preterm infants. *Eur. J. Clin. Microbiol. Infect. Dis* 37, 475–483 (2018). [PubMed: 29368074]
39. Moles L et al. Preterm infant gut colonization in the neonatal ICU and complete restoration 2 years later. *Clin. Microbiol. Infect* 21, 936.e1–936.e10 (2015).
40. Committee Opinion No 579: Definition of Term Pregnancy. *Obstet. Gynecol* 122, 1139–1140 (2013). [PubMed: 24150030]
41. Raju TNK, Higgins RD, Stark AR & Leveno KJ Optimizing Care and Outcome for Late-Preterm (Near-Term) Infants: A Summary of the Workshop Sponsored by the National Institute of Child Health and Human Development. *PEDIATRICS* 118, 1207–1214 (2006). [PubMed: 16951017]
42. Truong DT et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903 (2015). [PubMed: 26418763]
43. Franzosa EA et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968 (2018). [PubMed: 30377376]
44. Bradley PH & Pollard KS Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome* 5, 36 (2017). [PubMed: 28330508]
45. Lindberg TP et al. Preterm infant gut microbial patterns related to the development of necrotizing enterocolitis. *J. Matern. Fetal Neonatal Med* 1–10 (2018). doi:10.1080/14767058.2018.1490719
46. Warner BB et al. Gut bacteria dysbiosis and necrotising enterocolitis in very low birthweight infants: a prospective case-control study. *The Lancet* 387, 1928–1936 (2016).
47. Abrahamsson TR et al. Low diversity of the gut microbiota in infants with atopic eczema. *J. Allergy Clin. Immunol* 129, 434–440.e2 (2012). [PubMed: 22153774]
48. Kriss M, Hazleton KZ, Nusbacher NM, Martin CG & Lozupone CA Low diversity gut microbiota dysbiosis: drivers, functional implications and recovery. *Curr. Opin. Microbiol* 44, 34–40 (2018). [PubMed: 30036705]
49. Zou Z-H et al. Prenatal and postnatal antibiotic exposure influences the gut microbiota of preterm infants in neonatal intensive care units. *Ann. Clin. Microbiol. Antimicrob* 17, 9 (2018). [PubMed: 29554907]
50. Zhu D et al. Effects of One-Week Empirical Antibiotic Therapy on the Early Development of Gut Microbiota and Metabolites in Preterm Infants. *Sci. Rep* 7, 8025 (2017). [PubMed: 28808302]
51. Subramanian S et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 510, 417–421 (2014). [PubMed: 24896187]
52. Breiman L Random Forests. *Mach. Learn* 45, 5–32 (2001).
53. Forsberg KJ et al. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* 337, 1107–1111 (2012). [PubMed: 22936781]
54. Jia B et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573 (2017). [PubMed: 27789705]
55. Wyres KL & Holt KE *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol* 45, 131–139 (2018). [PubMed: 29723841]

56. Navon-Venezia S, Kondratyeva K & Carattoli A Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol. Rev* 41, 252–275 (2017). [PubMed: 28521338]
57. Goldstone RJ & Smith DGE A population genomics approach to exploiting the accessory ‘resistome’ of Escherichia coli. *Microb. Genomics* 3, (2017).
58. Crofts TS, Gasparrini AJ & Dantas G Next-generation approaches to understand and combat the antibiotic resistome. *Nat. Rev. Microbiol* 15, 422–434 (2017). [PubMed: 28392565]
59. Kaminski J et al. High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED. *PLOS Comput. Biol* 11, e1004557 (2015). [PubMed: 26682918]
60. Carl MA et al. Sepsis From the Gut: The Enteric Habitat of Bacteria That Cause Late-Onset Neonatal Bloodstream Infections. *Clin. Infect. Dis* 58, 1211–1218 (2014). [PubMed: 24647013]
61. Zhang X, Feng Y, Zhou W, McNally A & Zong Z Cryptic transmission of ST405 Escherichia coli carrying bla NDM-4 in hospital. *Sci. Rep* 8, 390 (2018). [PubMed: 29321680]
62. Izdebski R et al. MLST reveals potentially high-risk international clones of Enterobacter cloacae*. *J. Antimicrob. Chemother* 70, 48–56 (2015). [PubMed: 25216820]
63. Gurnee EA et al. Gut Colonization of Healthy Children and Their Mothers With Pathogenic Ciprofloxacin-Resistant Escherichia coli. *J. Infect. Dis* 212, 1862–1868 (2015). [PubMed: 25969564]
64. Kim HB et al. oqxAB Encoding a Multidrug Efflux Pump in Human Clinical Isolates of Enterobacteriaceae. *Antimicrob. Agents Chemother* 53, 3582–3584 (2009). [PubMed: 19528276]
65. Fevre C, Passet V, Weill F-X, Grimont PAD & Brisse S Variants of the Klebsiella pneumoniae OKP Chromosomal Beta-Lactamase Are Divided into Two Main Groups, OKP-A and OKP-B. *Antimicrob. Agents Chemother* 49, 5149–5152 (2005). [PubMed: 16304190]
66. Gasparrini AJ et al. Antibiotic perturbation of the preterm infant gut microbiome and resistome. *Gut Microbes* 7, 443–449 (2016). [PubMed: 27472377]
67. Furtado I et al. Enterococcus faecium and Enterococcus faecalis in blood of newborns with suspected nosocomial infection. *Rev. Inst. Med. Trop. São Paulo* 56, 77–80 (2014). [PubMed: 24553613]
68. Akturk H et al. Vancomycin Resistant Enterococci Colonization in a Neonatal Intensive Care Unit: Who will be infected? *J. Matern. Fetal Neonatal Med* 1–22 (2016). doi: 10.3109/14767058.2015.1132693
69. Brooks B et al. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome* 2, 1 (2014). [PubMed: 24468033]
70. Fernández-Canigia L, Cejas D, Gutkind G & Radice M Detection and genetic characterization of β -lactamases in Prevotella intermedia and Prevotella nigrescens isolated from oral cavity infections and peritonsillar abscesses. *Anaerobe* 33, 8–13 (2015). [PubMed: 25623818]
71. Singh B et al. Probiotics for preterm infants: A National Retrospective Cohort Study. *J. Perinatol. Off. J. Calif. Perinat. Assoc* 39, 533–539 (2019).
72. Kerr-Wilson CO, Mackay DF, Smith GCS & Pell JP Meta-analysis of the association between preterm delivery and intelligence. *J. Public Health* 34, 209–216 (2012).
73. Johnson S et al. Academic attainment and special educational needs in extremely preterm children at 11 years of age: the EPICure study. *Arch. Dis. Child. - Fetal Neonatal Ed* 94, F283–F289 (2009). [PubMed: 19282336]
74. Bhutta AT, Cleves MA, Casey PH, Cradock MM & Anand KJS Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 288, 728–737 (2002). [PubMed: 12169077]
75. Tinnion R, Gillone J, Cheetham T & Embleton N Preterm birth and subsequent insulin sensitivity: a systematic review. *Arch. Dis. Child* 99, 362–368 (2014). [PubMed: 24363362]
76. Parkinson JRC, Hyde MJ, Gale C, Santhakumaran S & Modi N Preterm Birth and the Metabolic Syndrome in Adult Life: A Systematic Review and Meta-analysis. *PEDIATRICS* 131, e1240–e1263 (2013). [PubMed: 23509172]
77. Crump C, Winkleby MA, Sundquist K & Sundquist J Risk of Hypertension Among Young Adults Who Were Born Preterm: A Swedish National Study of 636,000 Births. *Am. J. Epidemiol* 173, 797–803 (2011). [PubMed: 21320866]

78. Kowalski RR et al. Elevated Blood Pressure with Reduced Left Ventricular and Aortic Dimensions in Adolescents Born Extremely Preterm. *J. Pediatr* 172, 75–80.e2 (2016). [PubMed: 26873655]
79. Crump C, Winkleby MA, Sundquist J & Sundquist K Risk of Asthma in Young Adults Who Were Born Preterm: A Swedish National Cohort Study. *PEDIATRICS* 127, e913–e920 (2011). [PubMed: 21422091]
80. Lum S et al. Nature and severity of lung function abnormalities in extremely pre-term children at 11 years of age. *Eur. Respir. J* 37, 1199–1207 (2011). [PubMed: 20947682]

Methods References

1. La Rosa PS et al. Patterned progression of bacterial populations in the premature infant gut. *Proc. Natl. Acad. Sci* 111, 12522–12527 (2014). [PubMed: 25114261]
2. Planer JD et al. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* 534, 263–266 (2016). [PubMed: 27279225]
3. Baym M et al. Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLOS ONE* 10, e0128036 (2015). [PubMed: 26000737]
4. Truong DT et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903 (2015). [PubMed: 26418763]
5. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014). [PubMed: 24695404]
6. Franzosa EA et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968 (2018). [PubMed: 30377376]
7. Forsberg KJ et al. The Shared Antibiotic Resistome of Soil Bacteria and Human Pathogens. *Science* 337, 1107–1111 (2012). [PubMed: 22936781]
8. Zhu W, Lomsadze A & Borodovsky M Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38, e132–e132 (2010). [PubMed: 20403810]
9. Gibson MK, Forsberg KJ & Dantas G Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.* 9, 207–216 (2015). [PubMed: 25003965]
10. Finn RD, Clements J & Eddy SR HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37 (2011). [PubMed: 21593126]
11. Jia B et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573 (2017). [PubMed: 27789705]
12. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ Basic local alignment search tool. *J. Mol. Biol* 215, 403–410 (1990). [PubMed: 2231712]
13. Rice P, Longden I & Bleasby A EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG* 16, 276–277 (2000). [PubMed: 10827456]
14. Gibson MK et al. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol* 1, 16024 (2016). [PubMed: 27572443]
15. Schmieder R & Edwards R Fast Identification and Removal of Sequence Contamination from Genomic and Metagenomic Datasets. *PLoS ONE* 6, e17288 (2011). [PubMed: 21408061]
16. Bankevich A et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol* 19, 455–477 (2012). [PubMed: 22506599]
17. Gurevich A, Saveliev V, Vyahhi N & Tesler G QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075 (2013). [PubMed: 23422339]
18. Seemann T Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069 (2014). [PubMed: 24642063]
19. Ondov BD et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132 (2016). [PubMed: 27323842]
20. Page AJ et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693 (2015). [PubMed: 26198102]
21. Mohamed JA, Huang W, Nallapareddy SR, Teng F & Murray BE Influence of Origin of Isolates, Especially Endocarditis Isolates, and Various Genes on Biofilm Formation by *Enterococcus faecalis*. *Infect. Immun* 72, 3658–3663 (2004). [PubMed: 15155680]

22. Subramanian S et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 510, 417–421 (2014). [PubMed: 24896187]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

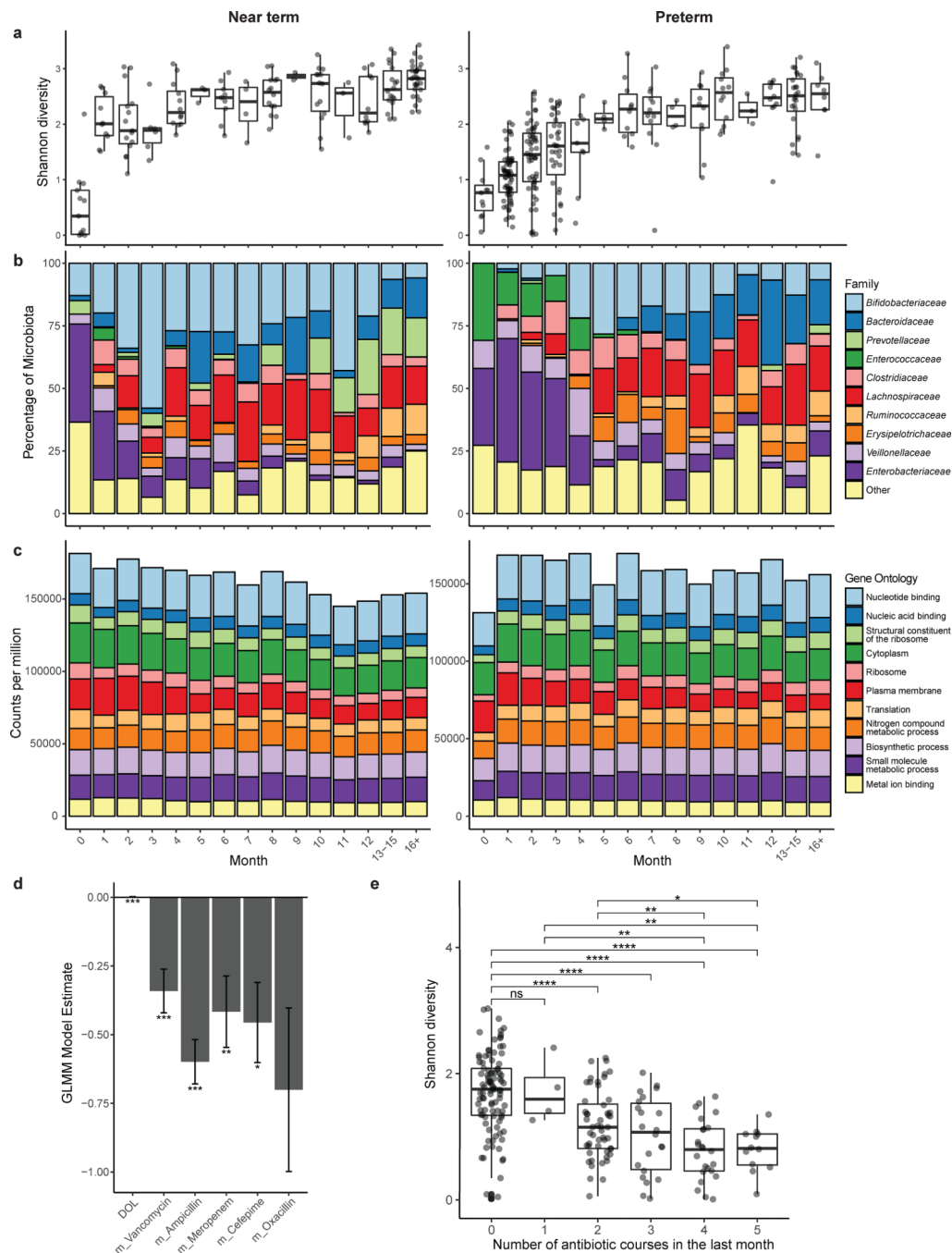


Figure 1 | Clinical variables predict microbiota diversity and composition.

a, Shannon diversity of all near-term ($n=17$) and preterm ($n=41$) infants in this study, by month of life. Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within $1.5\times$ the interquartile range. **b**, Microbiota species and functional compositions inferred by MetaPhlAn2 of all near-term ($n=17$) and preterm ($n=41$) infants in this study. **c**, Microbiota species and functional compositions inferred by HUMAnN2 of all near-term ($n=17$) and preterm ($n=41$) infants in this study. **d**, Day of life is significantly associated with an increase in microbiota Shannon

diversity, while vancomycin, ampicillin, meropenem, or cefepime treatment within the month prior to sampling is associated with significantly decreased species richness (***, $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; generalized linear mixed model with subject as random effect using 437 infant gut metagenomes.) Oxacillin was included in the model but was not significant after correction for multiple comparisons. Error bars indicate s.e. e, Shannon diversity is significantly lower in infants who have received >1 course of antibiotic treatment in the past month compared to infants who had not received antibiotic treatment during that time span (**** $p < 0.0001$, ***, $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; two-sided Wilcoxon with Benjamini-Hochberg correction, $n = 212$ samples). Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within $1.5 \times$ the interquartile range.

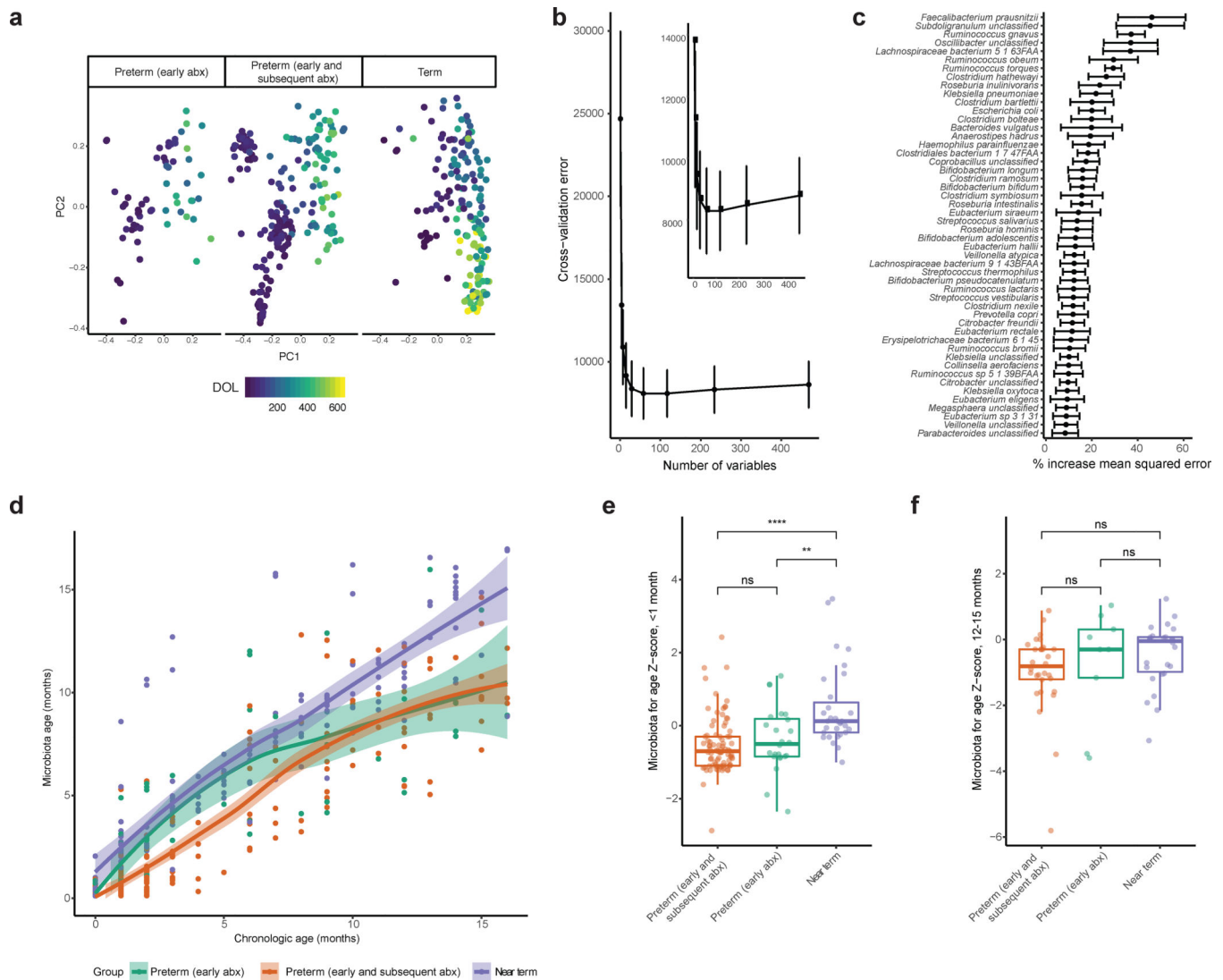


Figure 2 |. Partial architectural recovery of preterm infant gut microbiota following discharge from NICU.

a, Microbiota composition is distinct between near-term infants, preterm infants with early only antibiotic treatment, and preterm infants with early and subsequent antibiotic treatment (Bray-Curtis, $p < 0.001$, Adonis, $n = 437$ samples), but chronological day of life (DOL) is a major driver of microbiota composition. **b**, Fivefold cross-validation indicates that 50 variables are sufficient for random forests prediction of chronological age of near-term infants based on microbiota composition. Inset details vertex. Points indicate mean and error bars indicate s.e. computed over 100 iterations. **c**, The 50 most informative predictors to the random forests model. These species were included in a sparse model. Points indicate mean and error bars indicate s.e. computed over 100 iterations. **d**, The sparse random forests model accurately predicts near-term infant chronological age, but preterm infant chronological age is predicted to be less than actual age across numerous stages of development. Curves are loess regression fit to each group and shading depicts 95% confidence interval, $n = 437$ samples. **e**, Preterm infant microbiota for age Z-score (MAZ) is

significantly lower than that of near-term infants in the first month of life, indicating early microbiota immaturity (** $p < 0.01$, **** $p < 0.0001$, two-sided Wilcoxon, $n = 140$ samples). **f**, Preterm infant MAZ is statistically indistinguishable from near-term infant by 12–15 months of life, indicating resolution of microbiota immaturity by this time point ($p > 0.05$, two-sided Wilcoxon, $n = 65$ samples). Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within $1.5 \times$ the interquartile range.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

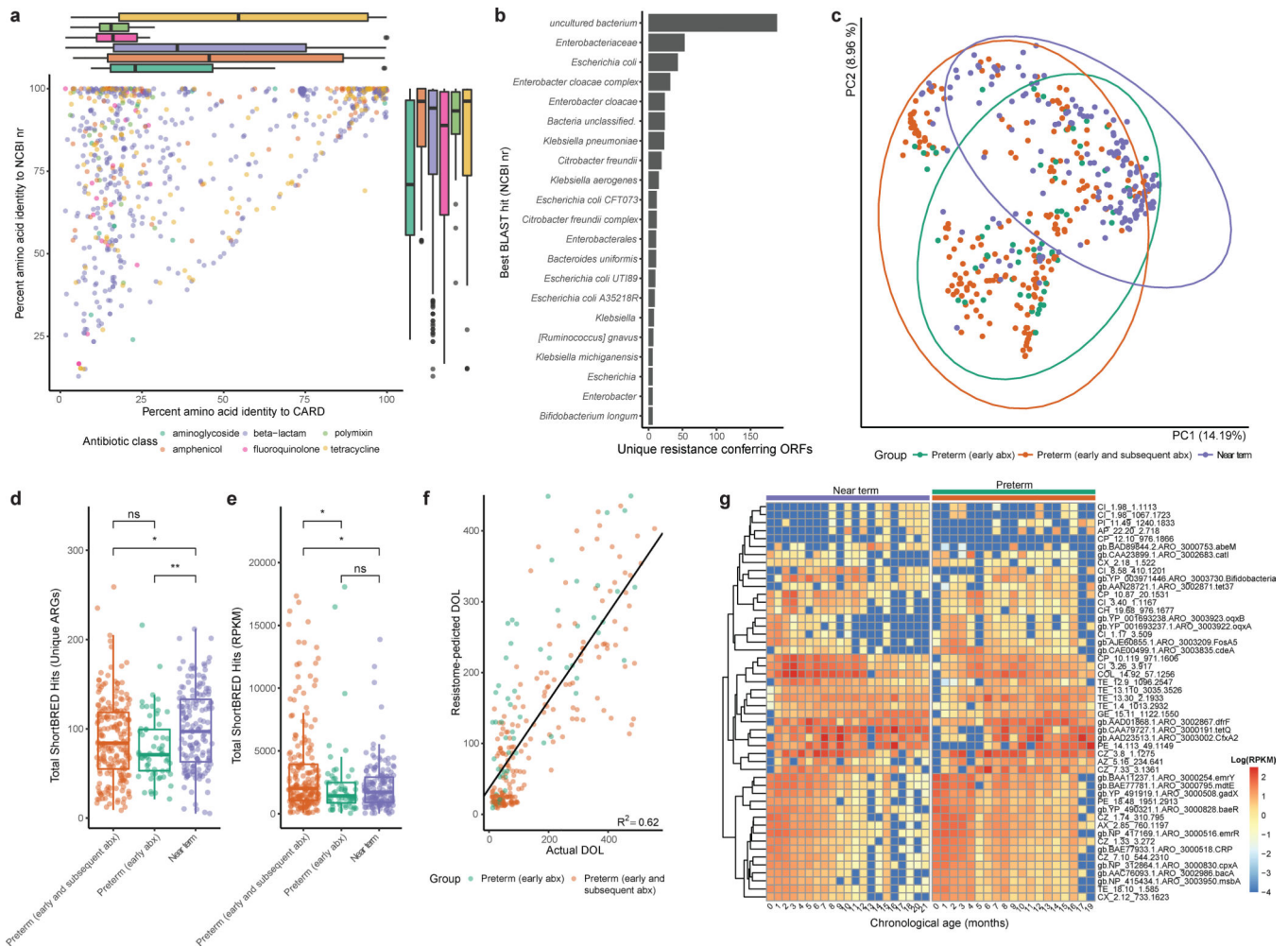


Figure 3 | Preterm infants harbor an enriched gut resistome.

a, Amino acid identity between all functionally-selected ARGs and their top hit in CARD vs their top hit in the NCBI nr protein database, colored by class of antibiotic used for selection. Notably, ARGs recovered by fluoroquinolone and polymyxin selection have very low median identity to CARD. Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within 1.5× the interquartile range, n=879 ARGs. **b**, The most commonly predicted hosts of functionally-selected ARGs based on highest identity BLAST hit in the NCBI nr protein database. **c**, Gut resistome composition is distinct between near-term infants, preterm infants with early only antibiotic treatment, and preterm infants with early and subsequent antibiotic treatment (Bray-Curtis, $p < 0.001$, Adonis, n=437 samples). **d**, Preterm infants had fewer unique ARGs encoded in their gut metagenomes than near-term infants. (* $p < 0.05$, ** $p < 0.01$ two-sided Wilcoxon, n=437 samples). Box plots represent the first quartile, median, and third quartile of the data with whiskers extending to the last data point within 1.5× the interquartile range. **e**, The cumulative resistome relative abundance was significantly higher in the gut microbiota of preterm infants with early and subsequent antibiotic treatment compared to both preterm infants with only early antibiotic treatment and near-term infants (* $p < 0.05$, two-sided Wilcoxon, n=437 samples). Box plots represent the first quartile, median, and

third quartile of the data with whiskers extending to the last data point within $1.5\times$ the interquartile range. **f**, A random forests model trained on preterm infant gut resistome poorly predicts chronological age of near-term infants. Black line is linear regression line of day of life as predicted by the random forests model against the actual day of life ($R^2=0.62$, $n=437$ samples) **g**, Relative abundance of 50 most informative resistance genes over the first months of life in near-term (left heatmap) and preterm (right heatmap) infants. Resistance genes are hierarchically clustered and listed to the right of the heatmaps. Colored bars above heatmaps correspond to colors in panels c-e. For canonical resistance genes, the CARD accession is displayed and for resistance genes functionally-selected in this study, the relevant selection information is listed.

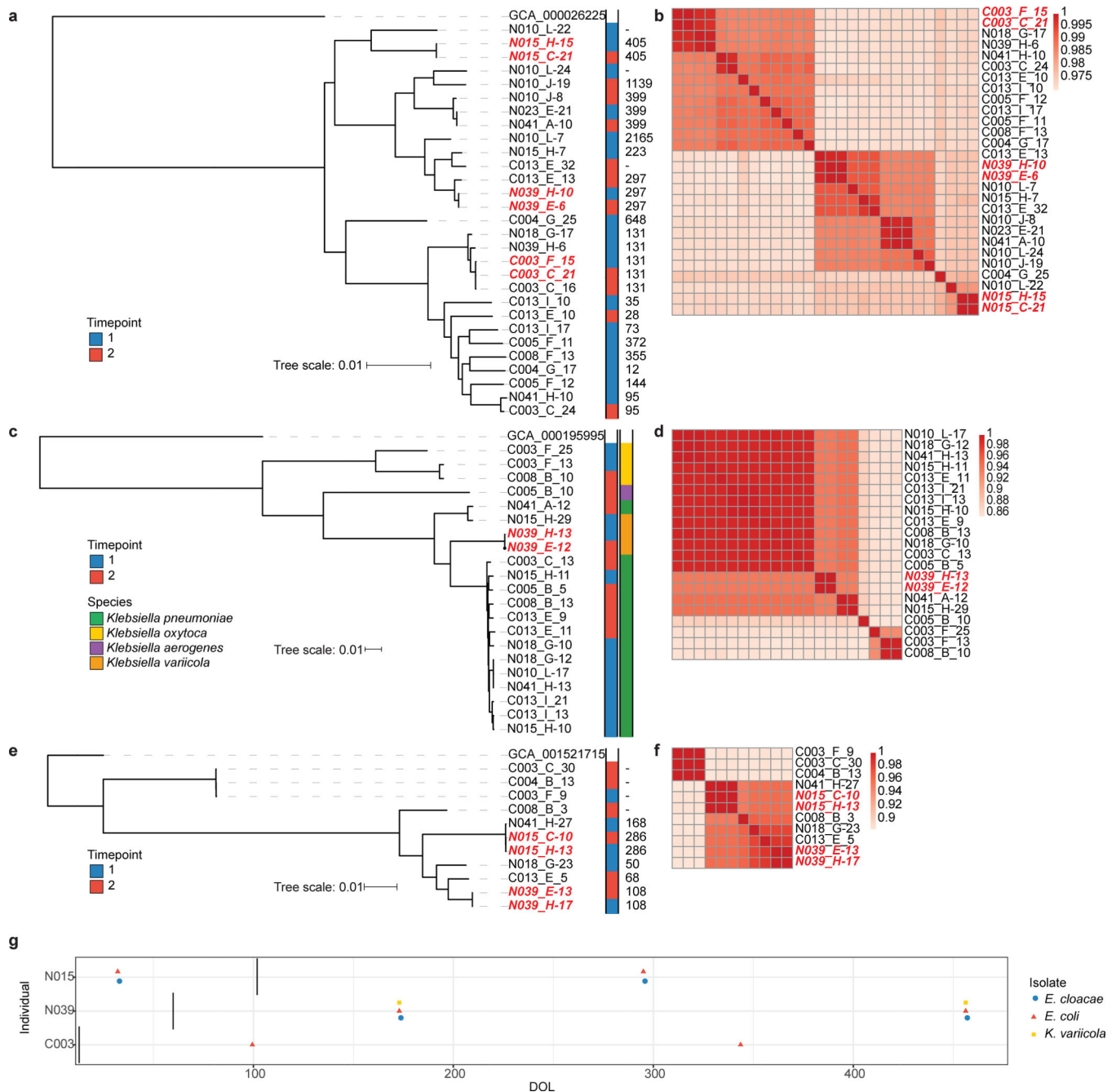


Figure 4 | Multidrug resistant Enterobacteriaceae lineages persist in infant gut microbiota. Maximum likelihood core genome phylogenies of *E. coli* (a), *Klebsiella* spp. (c), and *E. cloacae* (e) isolated from infant stool. Annotations to the right of metadata indicate timepoints of isolation and sequence type (determined by *in silico* MLST) for *E. coli* and *E. cloacae* or species for *Klebsiella*. Persistent isolates are highlighted in red. Average nucleotide identity heat maps for *E. coli* (b), *Klebsiella* spp. (d), and *E. cloacae* (f) indicate that persistent isolates are isogenic, i.e., they share >99.997% nucleotide identity. Persistent isolate pairs are highlighted in red. g, Timeline of isolation of persistent *Enterobacteriaceae* from infant stool. Vertical bar indicates the day at which infants were discharged from the

hospital. The first two infants displayed are preterm infants, while the third is a near-term infant.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

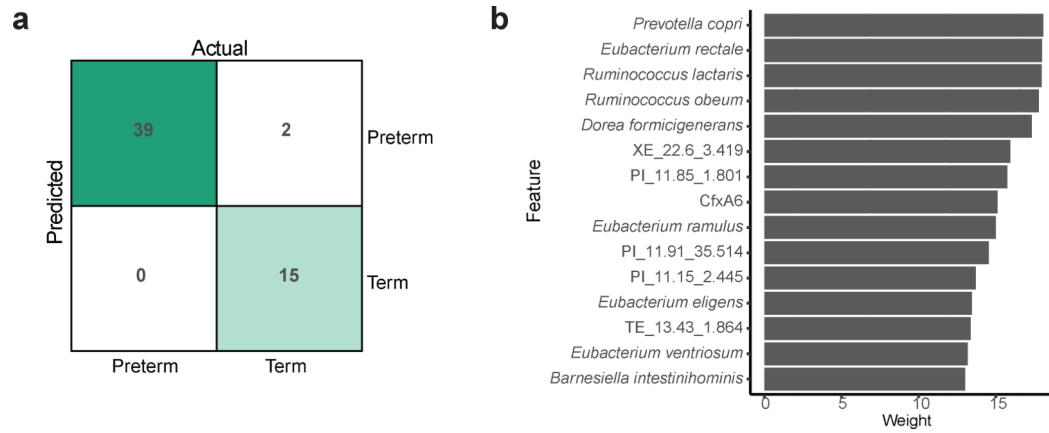


Figure 5 |. Enduring damage to the preterm infant gut microbiota.

a, Support vector machine confusion matrix for classification of gestational age based on the species and ARGs present in the microbiota following discharge from the NICU or at matched timepoints in unhospitalized near-term infants. **b**, Twenty predictors most important to classification. A sparse model trained using only these twenty predictors was highly accurate (96.4% classification accuracy), a persistent metagenomic signature of preterm birth and associated hospitalization and antibiotic treatment.

Table 1 |

Clinical characteristics of infant cohorts analyzed in this study

	Preterm Early Antibiotic Exposure Only (N=9)	Preterm Early + Subsequent Antibiotic Exposure (N=32)	Term Antibiotic-Naive Infants (N=17)
Birth weight, g, median (IQR)	1080 (880, 1270)	830 (698.75, 897.5)	2529 (2359.5, 2966.5)
Gestational age at birth, weeks, median (IQR)	27 (26, 27)	25 (24, 26)	36 (36,37)
Gender, M/F	4/5	15/17	4/13
Route of delivery, C-section/vaginal	6/3	25/7	15/2
Antibiotic exposure, n courses			
Gentamicin	9	74	none
Ampicillin	9	37	none
Vancomycin	none	67	none
Clindamycin	none	16	none
Meropenem	none	14	none
Cefepime	none	11	none
Cefotaxime	none	10	none
Mupirocin	none	7	none
Trimethoprim-sulfamethoxazole	none	4	none
Ticarcillin-clavulanate	none	3	none
Oxacillin	none	3	none
Cefoxitin	none	3	none
Cefazolin	none	2	none
Amoxicillin	none	2	none
Metronidazole	none	1	none
Penicillin G	none	1	none
Bacterial culture positive, n			
Blood	0	22	0
Tracheal	0	30	0
Urine	0	17	0