Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation

Kei E Fujimura1, Alexandra R Sitarik2, Suzanne Havstad2, Din I Lin1, Sophia Levan1, Douglas Fadrosh1, Ariane R Panzer1, Brandon LaMere1, Elze Rackaityte1, Nicholas W Lukacs3, Ganesa Wegienka4, Homer A Boushey4, Dennis R Ownby5, Edward M Zoratti6, Albert M Levin2, Christine C Johnson2,7 & Susan V Lynch1,7

Gut microbiota bacterial depletions and altered metabolic activity at 3 months are implicated in childhood atopy and asthma1. We hypothesized that compositionally distinct human neonatal gut microbiota (NGM) exist, and are differentially related to relative risk (RR) of childhood atopy and asthma. Using stool samples (n = 298; aged 1–11 months) from a US birth cohort and 16S rRNA sequencing, neonates (median age, 35 d) were divisible into three microbiota composition states (NGM1–3). Each incurred a substantially different RR for multisensitized atopy at age 2 years and doctor-diagnosed asthma at age 4 years. The highest risk group, labeled NGM3, showed lower relative abundance of certain bacteria (for example, Bifidobacterium, Akkermansia and Faecalibacterium), higher relative abundance of particular fungi (Candida and Rhodotorula) and a distinct fecal metabolome enriched for pro-inflammatory metabolites. Ex vivo culture of human adult peripheral T cells with sterile fecal water from NGM3 subjects increased the proportion of CD4+ cells producing interleukin (IL)-4 and reduced the relative abundance of CD4+CD25+FOXP3+ cells. 12,13-DiHOME, enriched in NGM3 versus lower-risk NGM states, recapitulated the effect of NGM3 fecal water on relative CD4+CD25+FOXP3+ cell abundance. These findings suggest that neonatal gut microbiome dysbiosis might promote CD4+ T cell dysfunction associated with childhood atopy.

Atopy, the propensity to produce IgE antibodies in response to allergens, is one of the most common chronic health issues2 and is considered to be a substantial risk factor for childhood asthma development3. Recently, the condition has been linked to bacterial ataxia of the human gut microbiota at 3 months, but not at 12 months, of age1. We therefore hypothesized that compositionally and functionally distinct neonatal (~1 month of age) gut microbiota states exist, and that their associated products idiosyncratically influence CD4+ populations in a manner that relates to the RR of atopy and asthma development in childhood. We studied independent fecal samples collected during a study visit that targeted 1 month olds (median age 35 d; range 16–138 d; n = 130; ‘neonates’) or 6 month olds (median age 201 d; range 170–322 d; n = 168; ‘infants’) from participants in the racially and socioeconomically diverse Wayne County Health, Environment, Allergy and Asthma Longitudinal Study birth cohort4. Predominantly multisensitized (PM) atopy at age 2 years was defined using latent-class analysis, an unsupervised statistical algorithm that clusters subjects according to their pattern of serum specific-IgE (sIgE) responses to a panel of ten food and aeroallergens5 (Supplementary Table 1 and Supplementary Fig. 1).

At the human population level (independent of atopy status), bacterial community α-diversity (number of taxa and their distribution) expanded with increasing age (Pearson’s correlation, r = 0.47, P < 0.001; Supplementary Fig. 2a). In parallel, fungal α-diversity contracted (Pearson’s correlation, r = −0.23, P = 0.0014; Supplementary Fig. 2b), and a reciprocal relationship between these microbial kingdoms existed (Shannon’s index; Pearson’s correlation, r = −0.24, P < 0.001; Fig. 1a). Both bacterial and fungal β-diversity (interpersonal taxonomic composition) were related to participant age (PERMANOVA; R2 = 0.056, P < 0.001; and R2 = 0.034, P < 0.001, respectively; Fig. 1b–c). Neonatal fecal microbiota were typically dominated by Bifidobacteriaceae, Enterobacteriaceae (Fig. 1d), Malasseziales (Malassezia) and Saccharomycetales (Saccharomyces; Fig. 1e). Infant participants exhibited sustained presence, but diminished relative abundance, of Bifidobacteriaceae and Enterobacteriaceae, an expansion of Lachnospiraceae (Blautia and Ruminococcus) and fungal communities characterized by Saccharomycetales (Saccharomyces and Candida; Fig. 1e), the dominant fungal order in healthy adults6. These findings indicate an interkingdom gut microbial co-evolution along an age-associated developmental gradient over the first year of life.

To address our primary hypothesis, a Dirichlet multinomial mixture (DMM) model was used to group participants on the basis of bacterial-community composition7; three distinct NGM states (NGM1,
NGM-associated RR of PM atopy was supported by the sum of specific parental-reported, doctor-diagnosed asthma at age 4 years (Table 1). No significant difference in RR between NGM groups was observed (Table 1). However, when the asthma-predictive PM atopy definition was used, NGM3 participants incurred a higher RR of atopy at age 2 years, as compared to either NGM1 (RR = 2.94; 95% CI 1.42–6.09, P = 0.004; Table 1) or NGM2 groups (RR = 2.06; 95% CI 1.01–4.19, P = 0.048; Table 1). Even larger effect sizes for NGM3 were observed for the RR of parental-reported, doctor-diagnosed asthma at age 4 years (Table 1). NGM-associated RR of PM atopy was supported by the sum of specific IgE responses at age 2 years (Fig. 2c). IGM participants did not exhibit different RRs for PM atopy (RR = 1.02; 95% CI 0.59–1.75, P = 0.94; Supplementary Table 2) or asthma (RR = 0.51; 95% CI 0.22–1.17, P = 0.11), possibly owing to increased age range and microbial heterogeneity within this group. Using available early-life characteristics, we identified factors, including season of birth, age at sample collection and breastfeeding, to be substantially distinct across IGM states (Supplementary Table 3). Detectable dog allergen (Can f 1) concentrations (P = 0.045) in the home during the neonatal study visit (lowest in the NGM3 group) and baby gender (NGM3 was almost entirely male) significantly differed across NGMs (P = 0.038; Supplementary Table 4). Despite adjustment for these and other early-life factors commonly related to allergic disease, the relationship between NGM and atopy or asthma persisted (Supplementary Table 5). Only one other large pediatric gut-microbiota atopy study exists1, the youngest participants of which were substantially older (~100 d) than the neonates in our cohort (median age, 35 d). The application of our DMM model parameters to this data set identified two compositionally distinct groups (Bifidobacteria-dominated NGM1 and Lachnospiraceae-dominated NGM2; Supplementary Note), indicating that examination of neonatal stool samples is necessary to identify compositionally distinct pioneer microbiota related to differential RR of childhood disease.

NGM3 participants were characteristically depleted of bacterial taxa, including Bifidobacteria (Bifidobacteriaceae), Lactobacillus (Lactobacillaceae), Faecalibacterium (Clostridiaceae) and Akkermansia (Verrucomicrobiaceae), when compared with the NGM1 group (zero-inflated negative binomial regression (ZINB), Benjamini–Hochberg, q < 0.05; Fig. 2d; Supplementary Table 6). These observations were consistent when NGM3 was compared to NGM2 (Supplementary Fig. 4; Supplementary Table 7) and also with previously described atopy-associated early-life gut microbiota taxonomic depletions1.
Mycologically, NGM3 subjects were consistently depleted of multiple Malassezia taxa (Fig. 2e; ZINB; Benjamini–Hochberg, q < 0.20; Supplementary Tables 8 and 9)—striking, given our population-based observation that this genus is characteristically enriched in the neonatal gut microbiota. Fungal taxonomic enrichments in the NGM3 group were also consistent when compared to either of the lower-risk NGM1 and NGM2 groups. (Fig. 2b; Kruskal–Wallis; P = 0.256). Box plots are defined by the 25th and 75th percentiles. Center line represents the median (50th percentile). Whiskers are defined as 1.5 times the interquartile range (IQR, 75th–25th percentile), plus or minus the 75th and 25th percentiles, respectively. (c) The sum of allergen-specific serum IgE concentrations measured at 2 years of age (n = 130) is significantly higher for NGM3 compared with NGM1 participants (Welch’s t test; P = 0.034). Box plots are constructed as defined in b. (d) Bacterial taxonomic comparison of NGM3 with NGM1 subjects; taxa exhibiting significant differences (zero-inflated negative binomial regression (ZINB); Benjamini–Hochberg, q < 0.05) in mean relative abundance (values are natural-log-transformed for purposes of illustration) are shown. Bar height indicates the magnitude of between-group relative abundance difference. (e) Relative abundance of fungal genera differs across NGMs.

![Figure 2](image_url)

**Figure 2** Compositionally distinct, age-independent NGM states exist in neonates, exhibit significant differences in fungal taxonomy and are related to the RR of PM atopy at the age of 2 years. (a) NGM designation significantly explains the observed variation (n = 130; PERMANOVA of unweighted UniFrac distances, R² = 0.09; P < 0.001) in bacterial β-diversity. (b) NGM participants do not differ significantly in age (n = 130; Kruskal–Wallis; P = 0.256). Box plots are defined by the 25th and 75th percentiles. Center line represents the median (50th percentile). Whiskers are defined as 1.5 times the interquartile range (IQR, 75th–25th percentile), plus or minus the 75th and 25th percentiles, respectively. (c) The sum of allergen-specific serum IgE concentrations measured at 2 years of age (n = 130) is significantly higher for NGM3 compared with NGM1 participants (Welch’s t test; P = 0.034). Box plots are constructed as defined in b. (d) Bacterial taxonomic comparison of NGM3 with NGM1 subjects; taxa exhibiting significant differences (zero-inflated negative binomial regression (ZINB); Benjamini–Hochberg, q < 0.05) in mean relative abundance (values are natural-log-transformed for purposes of illustration) are shown. Bar height indicates the magnitude of between-group relative abundance difference. (e) Relative abundance of fungal genera differs across NGMs.

Mycologically, NGM3 subjects were consistently depleted of multiple Malassezia taxa (Fig. 2e; ZINB; Benjamini–Hochberg, q < 0.20; Supplementary Tables 8 and 9)—striking, given our population-based observation that this genus is characteristically enriched in the neonatal gut microbiota. Fungal taxonomic enrichments in the NGM3 group were also consistent when compared to either of the lower-risk NGM1 and NGM2 groups.

| NGMs exhibit significantly different RRs of PM atopy development at age 2 years and of parental report of doctor-diagnosed asthma at age 4 years. Significance of risk ratios between microbiota states was calculated on the basis of log-binomial regression. |
|---|---|---|---|---|---|
| **DMM community states** | **RR (95% CI)** | **Overall P value** |
| Atopy (PM) | 13 (18.6%) | 13 (26.5%) | 6 (54.5%) | NGM2 versus NGM1: 1.43 (0.73–2.81) P = 0.30 | 0.034 |
| Parental report of doctor-diagnosed asthma | 8 (13.6%) | 5 (11.9%) | 4 (40.0%) | NGM3 versus NGM1: 0.87 (0.31–2.50) P = 0.81 | 0.13 |
| Atopy (IgE > 0.35 IU ml⁻¹) | 29 (41.4%) | 25 (51%) | 7 (63.6%) | NGM3 versus NGM1: 1.23 (0.83–1.82) P = 0.30 | 0.30 |

*P = 19 missing information regarding doctor diagnosis of asthma; average age at interview, 4.1 years (s.d. of 0.84).
groups, and included Rhodotorula and Candida (Supplementary Tables 8 and 9). Hence, neonatal interkingdom microbiota dysbiosis is characteristic of PM atopy and asthma development in childhood.

NGM3-associated bacterial taxonomic alterations were predicted to result in a deficiency in amino acid, lipid and xenobiotic metabolism pathways (Supplementary Fig. 5). Untargeted liquid chromatography mass spectrometry-identified fecal metabolites present in a subset (n = 28) of the representative subjects from each NGM (those with the highest posterior probability of NGM membership). Substantial correlations existed between the 16S rRNA profile, predicted metagenome and the metabolome of NGMs (Procrustes; Supplementary Table 10), indicating a deterministic relationship between bacterial community composition and the metabolic microenvironment of the neonatal gut. Between-group comparisons identified specific metabolites enriched in each NGM (Welch’s t test; P < 0.05; Supplementary Fig. 6; Supplementary Tables 11 and 12). As previously reported from analysis of the urine of subjects with atopy1, NGM3 participants exhibited fecal enrichment of primary and secondary bile metabolites. However, more expansive metabolic dysfunction, involving lipid, amino acid, carbohydrate, peptide, xenobiotic, nucleotide, vitamin and energy metabolism pathways—essentially, the bacterial pathways predicted to be deficient in NGM3—was evident. Although the NGM1 and NGM2 groups exhibited distinct metabolic programs, a common subset of metabolites differentiated them from NGM3. These included anti-inflammatory polyunsaturated fatty acids, docosapentaenoate (n3 DPA; 22 n5) and dihomo-γ-linolenate9,10 (DGLA; 20:3n3 or n6), succinate and the breast-milk oligosaccharides, 3-fucosyllactose and lacto-N-fucopentaose II, which are known to influence gut epithelial colonization11,12 (Supplementary Fig. 6, Supplementary Tables 11 and 12). By contrast, NGM3 participants were consistently enriched for 12,13-DiHOME, stigma- and sitosterols, 8-hydroxyoctanoate, α-CEHC and γ-tocopherol.

Sterile fecal water from NGM3 participants (as compared to that from NGM1), decreased the ratio of CD4+IFNγ−:CD4+IL-4+ cells (linear mixed-effects model (LME), P = 0.095; Supplementary Fig. 7), increased the proportion of CD4+IL-4+ cells (LME, P < 0.001; Fig. 3a) and the concentration of IL-4 released (LME, P = 0.045; Fig. 3b) and reduced the percentage of CD4+CD25+FOXP3+ cells (compared with control). When the same ex vivo assay that was performed in Figure 3a–c was used, 12,13-DiHOME significantly reduced the proportion of CD4+CD25+FOXP3+ cells at three different concentrations compared to vehicle control (LME; P = 0.04, P < 0.001, P = 0.001 for concentrations of 75, 130 and 200 µM, respectively; center line represents mean proportion of cells).

![Figure 3](https://example.com/figure3.png) Sterile fecal water from NGM3 participants induces CD4+ cell population dysfunction associated with atopic asthma. Dendritic cells and autologously purified naïve CD4+ cells from the serum of two healthy adult donors (biological replicates) were incubated with sterile fecal water from NGM1 (n = 7; three biological replicates per sample) or NGM3 (n = 5; three biological replicates per sample) participants. (a,b) Fecal water from NGM3 participants induced significantly increased proportions of CD4+IL-4+ cells (LME, P < 0.001; center line represents mean) (a) and expression of IL-4 (LME; P = 0.045) (b). (c) Fecal water from both NGM1 and NGM3 participants induced significantly decreased proportions of CD4+CD25+FOXP3+ cells (LME; P < 0.001 for NGM1 and P = 0.017 for NGM3), compared with control. (d) Weighted correlation network analysis identified a metabolic module that differentiates NGM3 from NGM2 and NGM1 participants (n = 28; ANOVA; P = 0.038). Box plots define the 25th and 75th percentiles; the median is represented by the center line. IQR (75th–25th percentile) is represented by whiskers. (e) Scatterplot of metabolite significance (transformed P values) versus module membership (MM) of the 12 metabolites in the NGM3-discriminating metabolic module. Metabolites with a higher metabolite significance value discriminate NGM3 from other NGMs. Metabolites plotted above the dashed line (representing the overall P-value for between-NGM differences) are significantly associated with NGM differentiation (P < 0.05), and were detected in higher concentrations in NGM3 compared to the other NGMs. MM values indicate the degree of interconnectedness of a specific metabolite to other metabolites in the module (higher MM value indicates greater interconnectedness). (f) When the same ex vivo assay that was performed in Figure 3a–c was used, 12,13-DiHOME significantly reduced the proportion of CD4+CD25+FOXP3+ cells at three different concentrations compared to vehicle control (LME; P = 0.04, P < 0.001, P = 0.001 for concentrations of 75, 130 and 200 µM, respectively; center line represents mean proportion of cells).
control; LME, \( P < 0.017 \) Fig. 3c) ex vivo, indicating that the NGM3 gut microenvironment promotes adaptive immune dysfunction associated with established atopic asthma. Weighted correlation network analysis identified 32 metabolic modules (Supplementary Fig. 8), one of which discriminated the three NGMs (ANOVA; \( P = 0.038 \); Fig. 3d). Notably, this module contained 12,13-DiHOME, which was identified both as a hub metabolite (highest module membership (MM) value = 0.91; Fig. 3e) and was NGM-discriminatory (\( P < 0.05 \); Fig. 3e). In addition, NGM3 subjects were enriched for 12,13-DiHOME compared to NGM1 and NGM2 (\( P < 0.05 \) for both; Supplementary Figs. 6 and 9).

**Supplementary Tables 11 and 12.** All concentrations of 12,13-DiHOME examined reduced the proportion of CD4+CD25+FOXP3+ cells, compared with vehicle treatment (LME, \( P = 0.04, P < 0.001, P = 0.001 \) respectively; Fig. 3f).

These findings indicate that neonatal gut microbiota influences susceptibility to childhood allergic asthma, potentially via alterations in the gut microenvironment that influence CD4+ T cell populations and function. This suggests that very early-life interventions to manipulate the composition and function of the gut microbiome might offer a viable strategy for disease prevention.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Accession codes.** All sequence data related to this study are available from the European Nucleotide Archive (ENA) under accession number PRJEB13896.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

This study was supported by the US National Institutes of Health and National Institute of Allergy and Infectious Diseases P01 AI089473-01 (C.C.J.), D.R.O., H.A.B., N.W.L., S.V.L., G.W., and E.M.Z.) and the Alfred P. Sloan Foundation 2013-6-03 (S.V.L.). We thank C. Arrieta and B. Finlay for graciously sharing sequencing data from the CHILD study.

**AUTHOR CONTRIBUTIONS**

S.V.L., C.C.J., D.R.O., H.A.B., N.W.L., G.W., and E.M.Z., designed research; C.C.J., D.R.O., K.E.F., D.E., B.L., D.L.L., S.L., A.R.P., E.R., and G.W., performed research; A.R.S., S.H., and A.M.L. contributed new analytic tools; K.E.F., A.R.S., S.H., S.L., A.M.L., and S.V.L. analyzed data; and K.E.F. and S.V.L. wrote the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Arrieta, M.C. et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. 7, 307ra152 (2015).
2. Asher, M.I. et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood. Lancet 368, 733–743 (2006).
3. Simpson, A. et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. Am. J. Respir. Crit. Care Med. 181, 1200–1206 (2010).
4. Aichbhaumik, N. et al. Prenatal exposure to household pets influences fetal immunoglobulin E production. Clin. Exp. Allergy 38, 1787–1794 (2008).
5. Havstad, S. et al. Atopic phenotypes identified with latent class analyses at age 2 years. J. Allergy Clin. Immunol. 134, 722–727.e2 (2014).
6. Hoffmann, C. et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One 8, e66019 (2013).
7. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One 7, e30126 (2012).
8. Langille, M.G.I. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814–821 (2013).
9. Morin, C., Blier, P.U. & Fortin, S. Eicosapentaenoic acid and docosapentaenoic acid monoglycerides are more potent than docosahexaenoic acid monoglyceride to resolve inflammation in a rheumatoid arthritis model. Arthritis Res. Ther. 14, 122 (2012).
10. Amagai, Y. et al. Dihomo-γ-linolenic acid prevents the development of atopic dermatitis through prostaglandin D1 production in NC/Tnd mice. J. Dermatol. Sci. 79, 30–37 (2015).
11. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22, 1147–1162 (2012).
12. Weichert, S. et al. Bioengineered 2′-fucosyllactose and 3-fucosyllactose inhibit the adhesion of Pseudomonas aeruginosa and enteric pathogens to human intestinal and respiratory cell lines. Nutr. Res. 33, 831–838 (2013).
ONLINE METHODS

Study population. Pregnant women (n = 1,258) between the ages of 21 and 49 were recruited from August 2003–November 2007 as part of the Wayne County Health, Environment, Allergy and Asthma Longitudinal Study (WHEALS). WHEALS is a prospective birth cohort from southeastern Michigan designed to investigate early-life risk factors for allergic diseases, as previously described. Briefly, women were considered eligible if they lived in a predefined cluster of contiguous zip codes in and surrounding Detroit, Michigan, had no intention of moving out of the area and provided informed written consent. Five follow-up interviews were conducted at 1, 6, 12, 24 and 48 months after the birth of their child, with the 24-month appointment being at a standardized study clinic so that the child could be evaluated by a board-certified allergist. Stool samples were collected from the child at the 1- and 6-month home visits. All aspects of this research were approved by the Henry Ford Hospital Institutional Review Board.

Subsample criteria of WHEALS subjects for stool microbiome analyses. For this study, we selected children who had completed their 24-month clinic visit, which included a blood draw for IgE measurements, and had had dust samples collected from their homes at the same time as their stool-sample collection (n = 308). Stool samples from children ranging from age 1–11 months were collected from field staff during home visits and stored at −80 °C. Samples were randomized before being shipped to the University of California, San Francisco (UCSF), on dry ice, where they were also stored at −80 °C until processed.

PM atopy and asthma definition. Blood drawn at the 2-year clinic visit was used to determine participants’ levels of total and ten allergen-specific IgEs (sIgE): Alternaria (Alternaria alternata), German cockroach (Blattella germanica Bla g 2), dog (Canis lupus familiaris Can f 1), house dust mites (Dermatophagoides farinae Der f 1), hen’s egg (egg), cat (Felis domesticus Fel d 1), cow’s milk (milk), peanut (Arachis hypogaea), common ragweed (Ambrosia artemisiifolia) and Timothy grass (Phleum pratense). Specific IgEs were measured using the Pharmacia UniCap system (ThermoFisher Scientific, Waltham, MA, USA). Latent class analysis was used to group participants into four discrete atopic classes according to sensitization patterns of the ten allergen sIgEs, as with the entire WHEALS cohort. Our subset was assigned to one of four latent classes: (i) Low or no sensitization (n = 226); (ii) highly sensitized (both food and inhalant allergens; n = 9); (iii) milk- and egg-dominated (n = 50) sensitization or (iv) peanut- and inhalant(s)-dominated (n = 13) sensitization. Because of the sample size, latent classes ii–iv were collapsed and considered to be “predominately multisensitized” (PM atopy; n = 72); remaining subjects represented the “low or no sensitization” group. The conventional definition of atopy (at least one positive test (sIgE ≥ 0.35 IU ml−1) to any of the ten allergens) was also used for comparative purposes. Children were defined as having asthma according to parental-reported doctor diagnosis of asthma at the 4-year interview.

Bacterial- and fungal-community profiling, PICRUSt and metabolic analyses. DNA extraction. Stool samples from 308 subjects were extracted using a modified cetyltrimethylammonium bromide (CTAB)-buffer-based protocol. Briefly, 0.5 ml of modified CTAB extraction buffer was added to 25 mg of stool in a 2-ml Lysing Matrix E tube (MP Biomedicals, Santa Ana, CA) and then incubated (65 °C, 15 min). Samples were bead-beaten (5.5 m s−1, 30 s) in a Fastprep-24 (MP Biomedicals, Santa Ana, CA), which was followed by the addition of 0.5 ml of phenol:chloroform:isoamyl alcohol (25:24:1). After centrifugation (14,000 rpm, 5 min), the supernatant was added to a heavy phase–light phase separation and then incubated (65 °C, 15 min). Samples were centrifuged (14,000 rpm, 5 min), and the resulting supernatants were added to fresh tubes, which was followed by the addition of 1 μl of linear acrylamide before PEG- NaCl (2×v:v). Samples were incubated (21 °C, 2 h), washed with 70% ETOH and resuspended in 10 mM Tris-Cl, pH 8.5.

Sequencing preparation. The V4 region of the 16S rRNA gene was amplified, as designed by Caporaso et al. PCR reactions were performed in 25-μl reactions using 0.025 U Takara Hot Start ExTaq (Takara Mirus Bio Inc., Madison, WI), 1X Takara buffer with MgCl2, 0.4 pmol/μl of F515 and R806 primers, 0.56 mg/ml of bovine serum albumin (BSA; Roche Applied Science, Indianapolis, IN), 200 μM of dNTPs and 10 ng of gDNA. Reactions were performed in triplicate with the following: initial denaturation (98 °C, 2 min), 30 cycles of 98 °C (20 s), annealing at 50 °C (30 s), extension at 72 °C (45 s) and final extension at 72 °C (10 min). Amplicons were pooled and verified using a 2% TBE agarose gel (Life Technologies, Grand Island, NY), before undergoing purification using AMPure SPRI beads (Beckman Coulter, Brea, CA), being quality checked with the Bioanalyzer DNA 1000 Kit (Agilent, Santa Clara, CA) and being quantified using the Qubit 2.0 Fluorometer and the dsDNA HS Assay Kit (Life Technologies, Grand Island, NY). Samples were pooled and sequenced on the Illumina MiSeq platform, as previously described.

The internal transcribed spacer region 2 (ITS2) of the rDNA gene was amplified using the primer pair ITS5′ (5′-GTGARTCAATCGAATTCTTG-3′) and ITS4′ (5′-TCCCTCGCGTATTAGTGATGCG-3′). Primers were designed for the Illumina MiSeq platform, as described above. PCR reactions were performed in triplicate in a 25-μl reaction with 1X Takara buffer (Takara Mirus Bio), 200 nM of each primer, 200 μM dNTPs, 2.75 mM of MgCl2, 0.56 mg/ml of BSA (Roche Applied Science, Indianapolis, IN), 0.025 U Takara Hot Start ExTaq and 50 ng of gDNA. Reactions were conducted under the following conditions: initial denaturation (94 °C, 5 min), 30 cycles of 94 °C (30 s), annealing at 54 °C (30 s), extension at 72 °C (30 s) and a final extension at 72 °C (7 min). PCR verification and purification were performed as described above. Samples were quantified using KAPA SYBR (KAPA Biosystems, Wilmington, MA) qPCR, following the manufacturer’s protocol. Samples were pooled in equal moles (50 ng), and prepped and denatured libraries with PhiX spike-in control, as described above, were sequenced using an Illumina MiSeq platform.

Sequencing-data processing and quality control. For bacterial sequences, paired-end reads were assembled using FLASH v1.2.7, de-multiplexed by barcode, and low-quality reads (Q score, <30) were discarded in QUIME. If three consecutive bases were <Q30, then the read was truncated and the resulting read retained in the data set only if it was at least 75% of the original length. Sequences were checked for chimeras using UCHIME and filtered from the data set before operational taxonomic unit (OTU) picking at 97% sequence identification using UCLUST against the Greengenes database version 13.5. Sequencing reads that failed to cluster with a reference sequence were clustered de novo. Sequences were aligned using PyNAST, and taxonomy was assigned using the RDP classifier and Greengenes reference database version 13.5. To de-noise the OTU table, taxa with fewer than five total sequences across all samples were removed. A bacterial phylogenetic tree was built using FastTree with Taxa with fewer than five total sequences across all samples were removed. A bacterial phylogenetic tree was built using FastTree. Fungal sequences were quality trimmed (Q score, <25) and adaptor sequences removed using cutadapt, then paired-end reads were assembled with FLASH. Sequences were demultiplexed by barcode and truncated to 150 bp before undergoing clustering using USEARCH v7. pipeline, specifically the UPARSE function, and being chimera-checked using UCHIME. Taxonomy was assigned using UNITE vers. 6.

To normalize variation in read depth across samples, data were rarefied to the minimum read depth of 202,367 sequences per sample for bacteria (n = 298) and 30,590 for fungi (n = 188). To ensure that a truly representative community was used for analysis for each sample, sequence subsampling at these defined depths was rarefied 100 times. The representative community composition for each sample was defined as that which exhibited the minimum average Euclidean distance to all other OTU vectors generated from all subsamplings for that particular sample (Supplementary Note). Investigators at UCSF were blinded to sample identity until microbiota data sets underwent the aforementioned processing and were ready for statistical analyses.

Phylogenetic reconstruction of unobserved states (PICRUSt). PICRUSt was used to predict the pathways of those taxa significantly enriched in each NGM state, according to zero-inflated negative binomial regression and corrected for multiple testing using the Benjamini–Hochberg false-discovery rate. These taxa were used to generate a new OTU table normalized in PICRUSt, and discriminatory pathways were illustrated in a heat map constructed in R.

Metabolic profiling. Stool samples (200 mg) from each of the three microbiota states, eight NGM3 subjects, and ten from each of NGM1 and NGM2 groups were provided to Metabolon (Durham, NC) for ultra-high performance liquid chromatography–tandem mass spectrometry (UPLC–MS/MS) and gas chromatography–mass spectrometry (GC–MS) using their standard protocol.
Ex vivo dendritic cell challenge and T cell co-culture. Fecal samples from five of ten NGM1 and seven of eight NGM3 neonates that had undergone metabolic profiling were used (biological replicates). Excluded samples from these groups had insufficient volume for analyses. Fecal samples were homogenized 1 g ml\(^{-1}\) (w/v) in pre-warmed phosphate-buffered saline (PBS) containing 20% FBS (FBS). Samples were vortexed, incubated (37 °C, 10 min) and centrifuged (14,000 rpm, 30 min). Supernatant was filter-sterilized through a 0.2-μm filter before being used in the dendritic cell (DC) T cell assay described below. PBS was used as the negative control. Treatment conditions used for the DiHOME experiment included: 75 μM, 130 μM and 200 μM 12,13-DiHOME (Cayman Chemical, Ann Arbor, MI) solubilized in 0.4%, 0.15% and 0.05% DMSO, respectively. DiHOME solutions were added to R10 media (Roswell Park Memorial Institute media 1640 with 10% heat-inactivated FBS (antigen activator) and 2 mM l-glutamine and 100 U ml\(^{-1}\) penicillin-streptomycin added; Life Technologies) and exposed to DCs within 1 h of preparation. Controls included PBS and DMSO (delivered in R10 media) at corresponding percentages used to dissolve the different concentrations of DiHOME. Treatment group size was determined on the basis of preliminary assays that demonstrated the effect size for the suppression of CD4\(^+\)CD25\(^+\)FOXP3\(^+\) using 130 μM of 12,13-DiHOME was approximately seven, indicating that at least two samples per group were required to achieve a power of >0.80.

Peripheral blood mononuclear cells (PBMCs) were purified from plasma obtained from healthy, de-identified human donors (Blood Centers of the Pacific, San Francisco, CA) through the cell-sourcing program that ensures donor confidentiality. Donors signed an agreement acknowledging that their blood may be used for research. PBMCs were isolated using Ficoll–Hypaque gradient centrifugation, washed twice with R10 media and incubated for 18 h. Dendritic cells (DCs) were isolated from PBMCs using the EasySep Human Pan-DC Pre-Enrichment Kit (STEMCELL Technologies, Vancouver, BC). DCs (0.5 × 10\(^6\) cells ml\(^{-1}\)) from two donors (biological replication) were treated in triplicate (treatment replicate) with either cell-free fecal water (0.22 μM filtered) or varying concentrations of DiHOME, and cultured in R10 media supplemented with 10 ng ml\(^{-1}\) GM-CSF and 20 ng ml\(^{-1}\) IL-4 at 37 °C for 2 d, for the fecal-water assay, or 5 d, for the DiHOME experiment. For the DiHOME experiment, freshly prepared media containing DiHOME or control exposures was replaced every 48 h. For the fecal-water experiment, the assay was repeated twice on one donor (technical replicates) and once on donor B owing to insufficient numbers of cells recovered from the latter donor. Treatment replicates were also considered biological replicates because the human donor cells are not clonal.

24 h before co-culture with CD4\(^+\) T cells, DC maturation was stimulated by using DC growth mediators (10 ng ml\(^{-1}\) TNF-α, 10 ng ml\(^{-1}\) IL-1β, 10 ng ml\(^{-1}\) IL-6 and 1 mM prostaglandin E2 (PGE2)). In preparation for co-culture, DCs were washed in fresh R10 media, counted via flow cytometry and plated in TexMACs Medium (Miltenyi Biotec, San Diego, CA) at 0.5 × 10\(^6\) live DC4\(^+\) cells per well.

Autologous T lymphocytes were purified from the PBMCs using a naive CD4\(^+\) T cell isolation kit (Miltenyi Biotec). After purification, naive autologous CD4\(^+\) T cells were suspended in the TexMACS Medium (Miltenyi Biotec) and added to the treated DCs at a ratio of 10:1 in the presence of soluble anti-CD28 and anti-CD49d (1 μg ml\(^{-1}\)). T and DC cells were co-cultured for 5 d at 37 °C and replenished with fresh TexMACS media every 48 h. To assess cytokine production, the co-cultures were mixed with Phorbol Myristate Acetate-Ionomycin (SIGSa, St Louis, MO) and GolgiPlug (Gplug; BD Biosciences, San Jose, CA) for 16 h before flow cytometry. Cell-free media from the co-cultures was collected at 48 h and 5 d, before PMA–Gplug addition, to assess cytokine secretion. Cytokine secretion was evaluated by cytometric bead array, following the manufacturer’s protocol (BD Biosciences).

For flow cytometry, single-cell suspensions were stained using a panel of antibodies, including anti-CD3 (SP34-2, 1:100), anti-CD4 (L200, 1:100), anti-CD25 (M-A251, 1:25), anti-IFN-γ (R727, 1:200; BD Biosciences); anti-CD69 (RPA-T8, 1:100; BioLegend, San Diego, CA); anti-IL-4 (7A3-3, 1:20; Miltenyi Biotec); anti-IL-17A (64DEC17, 1:20) and anti-FOXPI (PCH101, 1:20; Affymetrix eBioscience, Santa Clara, CA). Validation for each primary antibody is provided on the manufacturers’ websites. Dead cells were stained positive with LIVE–DEAD Aqua Dead Cell Stain (Life Technologies). Permeabilization buffer (Affymetrix eBioscience) was used to permeabilize cells before staining for the intracellular markers IFN-γ, IL-4, IL-17A and FoxP3. For flow analysis, live T cells were gated as CD3\(^+\)CD4\(^+\) cells, wells containing <50% live cells were excluded from analyses. Among CD4\(^+\) T cell subpopulations, T helper 1 (TH1) cells were IFN-γ\(^+\), T helper 2 (TH2) cells were IL-4\(^+\); T helper 17 (TH17) cells were IL-17A\(^+\), and T regulatory (Treg) cells were both CD25\(^+\) and FOXP3\(^+\). Stained cells were assayed via a flow cytometer on a BD LSR II (BD Biosciences).

Statistical analysis. Shannon’s diversity index was calculated using QIIME. Pearson’s correlation was used to test for a relationship between bacterial and fungal Shannon’s diversity. Distance matrices (unweighted UniFrac and Bray–Curtis) were calculated in QIIME to assess compositional dissimilarity between samples, and visualized using PCoA plots constructed in Emperor. Permutational multivariate analysis of variance (PERMANOVA) was performed using Adonis in the R environment to determine factors that significantly (P < 0.05) explained variation in microbiota β-diversity.

To identify clusters of subjects on the basis of bacterial-taxonomy, DMM models were used, which implement an unsupervised Bayesian approach that is based on a Dirichlet prior. The best-fitting DMM model was determined using the Laplace approximation to the negative-log model evidence, testing up to ten underlying microbiota states. Each sample was assigned to a particular neonatal gut microbiota (NGM) state on the basis of the maximum posterior probability of NGM membership. Kruskal–Wallis test was used to test whether age differentiated the microbiota states. Relative risk (RR) ratios and corresponding 95% confidence intervals were calculated using PROC GENMOD in SAS version 9.4 (Cary, NC). Unadjusted and adjusted RRs were calculated on the basis of log-binomial regression using maximum likelihood estimation or robust Poisson regression, when prevalence ratios were near one or when the log-binomial model did not converge. Two-tailed Welch’s t test was used to test whether slgE concentrations (log-transformed) were significantly different between the three NGM states.

To determine which OTUs differed in relative abundance between NGM groups, zero-inflated negative binomial regression (pscl package) was used as a primary modeling strategy, appropriate for sequence-count data. In cases in which OTU distributions were not zero-inflated and the model failed to converge, standard negative binomial was used as a secondary modeling strategy. These were corrected for multiple testing using the false-discovery rate (q < 0.05 for bacteria; q < 0.20 for fungi). Results were natural-log transformed for illustration on phylogenetic trees using iTOL version 3.0.6. When examining the association between early-life factors and NGMs, P values were calculated on the basis of covariate distribution by ANOVA (numerical, normally distributed), Kruskal–Wallis (numerical, skewed), chi-square (categorical) or Fisher’s exact (sparse categorical). Log-binomial-regression model was used to test for confounding factors when assessing the RR of individuals with different microbiota states developing atopy or asthma (PROC GENMOD in SAS version 9.4).

Metabolites exhibiting significantly (P < 0.05) different concentrations (log-transformed) between lower-risk NGM states and NGM3 were identified using two-tailed Welch’s t test. Shared and distinct super- and sub-pathway products among NGMs were illustrated using Cytoscape, ver. 3.2.1 (ref. 31). Co-occurrence networks of metabolites were constructed using weighted correlation network analysis (WGCNA) with the R package WGCNA to find modules of highly interconnected, mutually exclusive metabolites. Pearson correlations were used to determine intermetabolite relationships, wherein modules are composed of positively correlated metabolites. To avoid spurious modules, the minimum module size was set to five. Module ‘eigenmetabolites’ (referred to as eigengenes) were defined as the first principal component of a given module and considered as a representative measure of the joint metabolic profile of that module. Each ‘eigenmetabolite’ was used to test (ANOVA) the association between its respective module and NGM, module membership was used to determine the interconnectedness of each metabolite to its assigned module and to identify ‘hub’

DOI: 10.1038/nm.4176
metabolites: this was defined as the correlation between each metabolite and the
eigenmetabolite (strong positive values indicate high interconnectedness).

Procrustes was used to test for concurrence between communities described
by 16S phylogeny, PICRUSt and metabolomics data sets.

To test for T cell and cytokine differences, a linear mixed-effects model
(LME) was used (R package lmerTest) and adjusted for donors. Except
where indicated, all analyses were conducted in the R statistical program-
ning environment.

13. DeAngelis, K.M. et al. Selective progressive response of soil microbial community
to wild oat roots. *ISME J.* 3, 168–178 (2009).
14. Caporaso, J.G. et al. Global patterns of 16S rRNA diversity at a depth of millions
of sequences per sample. *Proc. Natl. Acad. Sci. USA* 108 (Suppl.), 4516–4522
(2011).
15. Caporaso, J.G. et al. Ultra-high-throughput microbial community analysis on the
Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624 (2012).
16. Magoc, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve
geneome assemblies. *Bioinformatics* 27, 2997–2993 (2011).
17. Caporaso, J.G. et al. QIIME allows analysis of high-throughput community
sequencing data. *Nat. Methods* 7, 335–336 (2010).
18. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. UCHIME improves
sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200
(2011).
19. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST.
*Bioinformatics* 26, 2460–2461 (2010).
20. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for
ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618
(2012).
21. Caporaso, J.G. et al. PyNAST: a flexible tool for aligning sequences to a template
alignment. *Bioinformatics* 26, 266–267 (2010).
22. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree 2—approximately maximum-
likelihood trees for large alignments. *PLoS One* 5, e9490 (2010).
23. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
reads. *EMBnet.journal* 17, 10–12 (2011).
24. Edgar, R.C. UPARSE: highly accurate OTU sequences from microbial amplicon
reads. *Nat. Methods* 10, 996–998 (2013).
25. Abarenkov, K. et al. The UNITE database for molecular identification of fungi—recent
updates and future perspectives. *New Phytol.* 186, 281–285 (2010).
26. Benjamin, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and
powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57, 289–300
(1995).
27. Obermaier, B. et al. Development of a new protocol for 2-day generation of
mature dendritic cells from human monocytes. *Biol. Proced. Online* 5, 197–203
(2003).
28. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing
microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235 (2005).
29. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPeror: a tool for
visualizing high-throughput microbial community data. *GigaScience* 2, 16
(2013).
30. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of
phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–W478 (2011).
31. Shannon, P. et al. Cytoscape: A software environment for integrated models of
biomolecular interaction networks cytoscape. *Genome Res.* 13, 2498–2504
(2003).