Epigenetic landscape links upper airway microbiota in infancy with allergic rhinitis at 6 years of age

Andréanne Morin, PhD,[a] Chris G. McKennan, PhD,[b]‡ Casper-Emil T. Pedersen, PhD,[c] Jakob Stokholm, MD, PhD,[c] Bo L. Chawes, MD, PhD, DMSc,[c] Ann-Marie Malby Schoos, MD, PhD,[c] Katherine A. Naughton, MS,[a] Jonathan Thorsen, MD, PhD,[c] Martin S. Mortensen, PhD,[d] Donata Vercelli, MD,[e,f] Urvish Trivedi, PhD,[d] Søren J. Sørensen, PhD,[d] Hans Bisgaard, MD, DMSc,[c] Dan L. Nicolae, PhD,[a,b]* Klaus Bønnelykke, MD, PhD,[c]* and Carole Ober, PhD[a]* *Chicago, Ill, Copenhagen, Denmark, and Tucson, Ariz*

Background: The upper airways present a barrier to inhaled allergens and microbes, which alter immune responses and subsequent risk for diseases, such as allergic rhinitis (AR).
Objective: We tested the hypothesis that early-life microbial exposures leave a lasting signature in DNA methylation that ultimately influences the development of AR in children.
Methods: We studied upper airway microbiota at 1 week, 1 month, and 3 months of life, and measured DNA methylation and gene expression profiles in upper airway mucosal cells and assessed AR at age 6 years in children in the Copenhagen Prospective Studies on Asthma in Childhood birth cohort.
Results: We identified 956 AR-associated differentially methylated CpGs in upper airway mucosal cells at age 6 years, 792 of which formed 3 modules of correlated differentially methylated CpGs. The eigenvector of 1 module was correlated with the expression of genes enriched for lysosome and bacterial invasion of epithelial cell pathways. Early-life microbial diversity was lower at 1 week (richness $P = .0079$) in children with AR at age 6 years, and reduced diversity at 1 week was also correlated with the same module's eigenvector ($\rho = -0.25$; $P = 3.3 \times 10^{-5}$). We show that the effect of microbiota richness at 1 week on risk for AR at age 6 years was mediated in part by the epigenetic signature of this module.
Conclusions: Our results suggest that upper airway microbial composition in infancy contributes to the development of AR during childhood, and this trajectory is mediated, at least in part, through altered DNA methylation patterns in upper airway mucosal cells. (J Allergy Clin Immunol 2020;■■■:■■■-■■■.)

Key words: *Allergic rhinitis, microbiota, DNA methylation, gene expression, early life, upper airways*

Allergic rhinitis (AR) is an inflammatory disease of the nasal mucosa that affects more than 400 million people worldwide.[1] In contrast to other forms of rhinitis, AR is also associated with allergic sensitization (AS), mainly to inhaled allergens. The disease itself is complex, with important contributions from both genetic and environmental factors. For example, a recent genome-wide association study of AR in more than 200,000 individuals reported 24 independent loci contributing to risk,[2] whereas many nongenetic factors, such as geography, season of birth, presence of siblings, early-life aeroallergen exposure, immune response patterns, and exposures to infections, have also been associated with risk of or protection from AR.[1,3,4] The epigenome bridges these 2 features by mediating the effects of the host response to environmental exposures on gene expression, thereby modifying disease risk profiles. DNA methylation (DNAm) is a widely studied epigenetic modification, and methylation patterns in airway cells have been associated with asthma and asthma-related phenotypes,[5-8] but only 1 previous study has explored the relationship between airway cell DNAm and asthma-related phenotypes using the most comprehensive array currently available.[5]

Early-life microbiota can shape host immune trajectories and influence subsequent development of disease.[9] During the first year of life, the microbiota is highly dynamic, ultimately "seeding" the microbial communities that become established at different body sites.[10,11] We previously demonstrated that the microbial composition of the upper airways at age 1 month was associated with subsequent development of asthma, using both culturing[12] and sequencing[13] techniques. Similar associations with early-life microbiota have been reported by others for respiratory tract infections[14,15] and rhinitis.[16]

Because the upper airways are the first site exposed to inhaled particles and provide signals that ultimately modulate disease susceptibility, we hypothesized that early-life exposures,

From the Departments of [a]Human Genetics and [b]Statistics, The University of Chicago, Chicago; [c]COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, Copenhagen; [d]the Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen; and [e]the Department of Cellular and Molecular Medicine and [f]the Asthma and Airway Disease Research Center, University of Arizona Health Sciences, Tucson.

*These authors contributed equally to this work.

‡Chris G. McKennan is currently at the Department of Statistics, University of Pittsburgh, Pittsburgh, Pa.

This work was supported by the National Institutes of Health (NIH) (grant no. HL129735 to H.B. and C.O.). Copenhagen Prospective Studies on Asthma in Childhood is funded by private and public research funds (complete list on www.copsac.com), including the Lundbeck Foundation (grant no. R16-A1694), the Danish Ministry of Health (grant no. 903516), the Danish Council for Strategic Research (grant no. 0603-00280B), the Danish Council for Independent Research (grant nos. 10-082884 and 271-08-0815), and the Capital Region Research Foundation. A.M. is supported by Bourse de formation postdoctorale – Fonds de recherche du Québec - Santé. D.V. was supported in part by the NIH (grant nos. AI133765 and AI144722). U.T. and S.J.S. are supported by the Novo Nordisk Fonden.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication May 6, 2020; revised June 19, 2020; accepted for publication July 2, 2020.

Corresponding author: Carole Ober, PhD, Department of Human Genetics, 920 E 58th St, Rm 507C, Cummings Life Science Center, University of Chicago, Chicago, IL 60615. E-mail: c-ober@genetics.uchicago.edu.

0091-6749

https://doi.org/10.1016/j.jaci.2020.07.005

Abbreviations used
AR: Allergic rhinitis
AS: Allergic sensitization
COPSAC: Copenhagen Prospective Studies on Asthma in Childhood
DMC: Differentially methylated CpG
DNAm: DNA methylation
FDR: False-discovery rate
QC: Quality control
RA: Relative abundance
WGCNA: Weighted gene coexpression network analysis

reflected in upper airway microbiota composition during the first months of life, leave lasting signatures in airway DNAm patterns that ultimately influence the development of AR in childhood. Our results revealed that upper airway microbiota diversity at age 1 week was associated with both DNAm patterns and AR at age 6 years. Although the effect of microbial richness at 1 week on AR risk at age 6 years was statistically significant, the conditional effect (controlling for the epigenetic signature) was 61% smaller and not statistically significant, suggesting that a DNAm signature mediates the effect of microbial richness on the development of AR.

METHODS

Study population

The Copenhagen Prospective Studies on Asthma in Childhood (COPSAC)$_{2010}$ is an ongoing prospective population-based cohort of 700 Danish mother-child pairs recruited at 2 sites near Copenhagen. The inclusion and exclusion criteria and baseline characteristics are detailed elsewhere.[17-19] At the children's sixth-year visit, rhinitis symptoms were assessed in 657 children and sensitization to 10 aeroallergens was measured in 538 of these children. Among the latter, children who presented with rhinitis symptoms and sensitization to at least 1 aeroallergen were classified as having AR (n = 40); children with neither AR nor sensitization were classified as controls (n = 428) (see Table E1 in this article's Online Repository at www.jacionline.org). Of the 468 included children, 236 were males and 232 were females. Children who were sensitized but without concomitant AR symptoms (n = 70) were included in the RNA and DNAm normalization steps, but excluded from all primary analyses.

AS was determined from IgE measurements to 10 aeroallergens (*Dermatophagoides pteronyssinus* [d1], cat [e1], horse [e3], dog [e5], grass [g6], birch [t3], mugwort [w6], *Cladosporium herbarum* [m2], *Aspergillus fumigatus* [m3], and *Alternaria tenuis* [m6]) and/or skin prick test to 10 common aeroallergens (*Alternaria spp*, birch, cat, *Cladosporium spp*, *Dermatophagoides farinae*, *D pteronyssinus*, dog, grass, horse, and mugwort) in 538 children, assessed during the same visit at which the nasal brush samples were obtained, as described.[20] Children were considered sensitized if they were positive for at least 1 specific IgE (≥0.35 kU_A/L) or 1 skin prick test (wheal ≥3 mm larger than that with negative control, n = 109). Rhinitis cases were defined as reoccurring sneezing and blocked or runny nose that severely affected the well-being of the child in the previous 12 months during periods without an accompanying common cold or flu, alone for nonallergic rhinitis (n = 34) or with congruence between symptoms, relevant allergen exposure, and sensitization for AR (n = 40). Atopic dermatitis (n = 60) was defined according to the criteria of Hanifin and Rajka and more details can be found in Thorsteinsdottir et al.[21] Food sensitization (n = 71) was defined the same way as AS, but for egg whites (f1), cow milk (f2), wheat (f4), and peanut (f13). None of the cases was treated with nasal steroids at the time of sampling.

The following 17 variables were also assessed in the children: clinical diagnosis of maternal or paternal rhinitis, mode of delivery, season of birth, duration of breast-feeding, older siblings, and exposure or treatment during pregnancy or in the first few years of life with fishoil and/or vitamin D supplementation, antibiotics, pets (cat or dog), smoking in household, and lower respiratory tract infection. Details on their measurements and definitions can be found in Table E2 in this article's Online Repository at www.jacionline.org.

At the sixth-year visit, DNA and RNA were collected from inferior turbinate epithelial cell scrapings from 562 children. Samples were obtained using a Rhino-probe nasal curette, stored in RNAprotect Cell Reagent (Qiagen, Germantown, Md), and then cryopreserved at −80°C until nucleotide extractions. After removing samples that did not pass quality control (QC) or that did not have genotyping data available (n = 60), 454 and 381 samples were retained for DNAm and gene expression studies, respectively (see Fig E1 in this article's Online Repository at www.jacionline.org), of which 348 and 288 were evaluated for AR (Table E1). The 60 samples without genotypes were excluded so that we could perform QC for sample swaps and correct for ancestry principal components (PCs) in the analyses. The overlapping samples between time points and measurements are described in Fig E2 in this article's Online Repository at www.jacionline.org. The microbiota was studied in hypopharyngeal samples collected from children at age 1 week, 1 month, and 3 months in 549, 647, and 657 children, respectively. Details on sampling, processing, and inclusion criteria for the microbiota studies were previously described.[22] Of these children, 361 (1 week), 441 (1 month), and 445 (3 months) were evaluated for AR at age 6 years (Table E1).

The study was conducted in accordance with the guiding principles of the Declaration of Helsinki and was approved by the Local Ethics Committee (H-B-2008-093) and the Danish Data Protection Agency (2015-41-3696). Both parents gave written informed consent before enrollment.

DNAm studies

Methylation profiles were assessed in DNA from upper airway mucosal cells, using the Illumina 850k EPIC array (Illumina, San Diego, Calif). QC and filtering were performed using the R package *minfi* (version 1.30).[23]

Of 866,836 probes on the array, we removed 19,040 with detection *P* value of more than .01 in 90% of the samples, 19,479 located on sex chromosome, 87,148 that overlapped a known common single nucleotide polymorphism (minor allele frequency > 0.05), and 34,057 that mapped to multiple locations on the bisulfite-converted genome (cross-hybridizing).[24] The remaining 707,112 CpGs that passed QC were used in further analyses. Normalization was performed using the SWAN algorithm[25] from the R package *minfi* and quantile normalization from the R package *lumi* (version 2.36).[26] We used M-values in analyses. Principal-component analysis was used to identify the effect of confounding variables on DNAm: DNA concentration, array, and site of sampling significantly correlated with at least 1 of the first 10 components.

To adjust for potential batch effects, sampling site and methylation array were regressed out of the DNAm data using ComBat (R package *sva* version 3.32.1).[27] Latent factors were defined after protecting the AR phenotype using the CorrConf method[28] and included as covariates to correct for hidden unwanted variation. Sex and the first 2 ancestry PCs were also included as covariates in DNAm and gene expression analyses.

To assess methylation differences between children with and without AR, we used the M-values in a linear model (*limma*[29], version 3.40.6), with sex, DNA concentration, ancestry PCs, and latent factors included as covariates. Differentially methylated CpGs (DMCs) were assessed using the Benjamini-Hochberg procedure (false-discovery rate [FDR] < 5%).

Gene expression studies

RNA extraction was performed using SMART-seq V4 Ultra Low input RNA kit, and cDNA libraries were prepared with the Illumina Nextera XT kit, according to the manufacturer's instructions. Concentration and quality were assessed using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, Calif). Sequencing was performed on the Illumina HiSeq 2500 platform using 6 pools of 91 to 95 samples in 5 flow cells each. Read quality was assessed using FastQC[30] and MultiQC,[31] for which all samples passed QC. FastQ files were then combined and RNA-seq reads were mapped to the genome (hg19) using STAR[32] (v2.5.1). All QC filtering of samples can be found in Fig E1.

Samples with no genotype available (n = 60) or less than 8M exon mapped reads were removed (n = 98). We used VerifyBamID[33] to detect sample swaps or sample contamination, and subsequently removed another 23 samples. Outlier samples that had more than 3-fold absolute deviation from the median count per million were removed (n = 17). A total of 381 samples were used for the normalization steps. Genes with less than 1 count per million in more than 85% of the samples (n = 43,983) or located on X and Y chromosomes were removed (n = 906). Median number of mapped reads was 10,936,189 (range, 8,044,014-86,750,179). Normalization of the 15,363 remaining genes using log-transformed count per million was performed using the trimmed mean of M-value[34] method and variance modeling (voom).[35] Batch effect and covariates were identified using principal-component analysis: RNA concentration, RNA integrity number, site as well as sex for the same reason mentioned above.

Assessing cell-type–specific gene expression profiles

To estimate the cell-type composition of the upper airway mucosal samples, we used gene expression profiles in these samples and cell-type–specific signatures described in a previous single-cell RNA sequencing study. Those signatures were derived from nonsurgical nasal scrapings from 3 healthy subjects and 6 subjects with inflamed airways (see Supplementary Table 3 extended Fig 6, *D*, in Ordovas-Montanes et al[36]).

Using the lists of predictor genes for basal cells, ciliated cells, secretory cells, eosinophils, macrophages, mast cells, neutrophils, and T cells,[36] we evaluated both whether the latent factors accounted for cell-type composition across samples and differences in cell composition between the AR cases and controls. For these analyses, we extracted expression data for the gene predictors of each cell type from normalized gene expression after the covariates sex, RNA concentration, RNA integrity number, and clinic site were regressed out. We used the first principal component of the cell-type signature genes for each cell type and first correlated their PC1 eigenvectors with the latent factors from the gene expression and DNAm data using Spearman correlation, and then tested for association between each cell-type signature and AR.

Co-DNAm networks using weighted gene coexpression network analysis

Methylation M-values of DMCs between AR cases and controls were used to create comethylation modules using the R package weighted gene coexpression network analysis (weighted gene coexpression network analysis [WGCNA]; version 1.68).[37] We used a soft thresholding power of 4 and required modules to include at least 30 DMCs. Correlations between the eigenvectors for each module with microbial diversity, relative abundance (RA) measures, and normalized gene expression were performed using Spearman correlation.

Pathway and functional features enrichment analyses

Gene-related pathways were assessed using iPathwayGuide (Advaita Corporation 2019) (pathways >1 genes and FDR 10%).

Enrichment permutation

To determine whether CpGs were enriched in specific analyses, we performed 10,000 random resampling of the number of CpGs observed. Empiric *P* values were determined on the basis of the number of time the permuted number was equal to or greater than the observed. If the observed value was never observed, *P* was considered as less than 1×10^{-4}.

Collection and analysis of microbiota samples

Fluid was aspirated with a soft catheter passed through the nose to the hypopharyngeal region. Catheters were immediately flushed with 1 mL of sterile 0.9% NaCl solution and stored at −80°C. Sampling happened at the COPSAC clinic during visits.

DNA was extracted from the cells using the Mobio Powersoil kit (Qiagen, Germany) on the epMotion 5075 robotic platform (Eppendorf, Germany), and amplified using a 2-step PCR reaction with the primers 515F[38] and 806R[39] targeting the hypervariable region V4 of the 16S ribosomal RNA gene. Sequencing was performed on the Miseq platform (Illumina) using the v2 kit (paired-end 250bp reads). A full description of the laboratory workflow has been described previously.[40] Sequencing adapters were removed using Cutadapt v1.15.[41] Reads were analyzed using QIIME 2 v2018.2.0[42] and denoized using DADA2.[43] Resulting amplicon sequence variants were compared with the 99% identity clustered SILVA database v132[44] using a naive Bayes classifier[45] trained on the amplified region. Diversity was quantified as richness (number of amplicon sequence variants per sample) and by the Shannon diversity index. To assess RA, amplicon sequence variants were filtered at the genus level, keeping those present in more than 10% of the children within each of the 3 time points and with a median RA greater than 0.01% (42 taxa). We used the analysis of composition of microbiome algorithm to assess the differences in RA between AR cases and controls.[46] Rarefaction of the richness value was performed using the *vegan* R package.[47] For more robust data, samples with more than 12,000 reads were considered.

Mediation analyses

To assess whether the association between early-life airways microbiota diversity and AR at age 6 years was mediated through DNAm patterns, we used logistic regression using the equations shown below, which first exclude (1) and then include (2) the eigenvectors from the WGCNA module of DMCs in the model. The eigenvector +1 was used to compensate for values less than 0, and the diversity measures were log-transformed to use in logistic regression.

$$Y_{AR} = \beta_0 + \beta_{richness1} + Sex + Read\ counts \quad (1)$$

$$Y_{AR} = \beta_0 + \beta_{richness2} + \beta_{blue\ module2} + Sex + Read\ counts \quad (2)$$

We calculated the proportion of risk mediated by DNAm using Equation 3.

$$\%\ explained = \frac{(\beta_{blue\ module2} \times \beta_{blue\ module4})}{((\beta_{blue\ module2} \times \beta_{blue\ module4}) + \beta_{richness2})} \quad (3)$$

$$Y_{richness} = \beta_0 + \beta_{blue\ module4} \quad (4)$$

Equation 4 is a linear regression of richness at 1 week and the eigenvectors from the WGCNA blue module, for which the estimate is used in Equation 3.

We also used mediation analyses to assess whether the association between early-life airways microbiota diversity and AR at age 6 years could be explained by other measured exposures (see Table E3 in this article's Online Repository at www.jacionline.org).

RESULTS

DNAm profiles in nasal mucosal cells differ between children with and without AR

We first examined DNAm profiles in nasal mucosal cells at age 6 years to identify patterns that differed between AR cases and controls using 707,112 CpGs on the array that passed QC. We identified 956 AR DMCs (FDR < 5%), of which 741 (78%) were less methylated in cells from children with AR (see Fig E3 and Table E4 in this article's Online Repository at www.jacionline.org). To assess the specificity of the DMCs for AR, we repeated the analysis considering AS, atopic dermatitis, sensitization to food allergens, and rhinitis alone at age 6 years. We detected 228 DMCs associated with AS at an FDR of 5%, of which 183 (80%) overlapped with AR DMCs. Only 1 DMC remained after excluding AR from AS cases (45 cases and 331 controls). We observed only 4

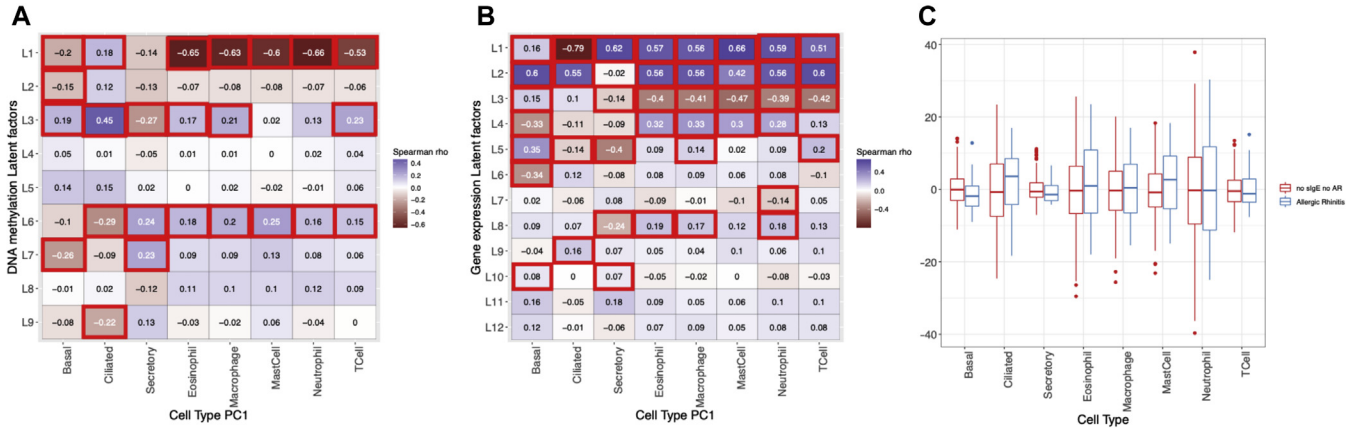

FIG 1. Latent factors capture cell-type composition in upper airway mucosal cells at age 6 years. Spearman correlations between cell-type signatures and (**A**) DNAm latent factors (n = 266) and (**B**) gene expression latent factors (n = 288). Cells outlined by red boxes are the Spearman correlation *P* values that remain significant after applying Benjamini-Hochberg procedure ($P < .05$). **C,** No differences between AR cases and controls were observed for any of cell-type signatures (Wilcoxon rank-sum test). Latent factors were not regressed out for this analysis. *sIgE,* Specific IgE.

TABLE I. Descriptions of WGCNA comethylation modules

Modules	Blue	Brown	Turquoise
N DMCs	260	161	371
More methylated in AR, n (%)	121 (46.5)	34 (21.1)	10 (2.7)
N-associated DMCs in a previous study of upper airway mucosal cells (5)	233 ($P < 1 \times 10^{-4}$)	128 ($P = .11$)	314 ($P < 1 \times 10^{-4}$)
Asthma (n = 84; 8.7%)	22 ($P = .628$)	4 ($P = 4 \times 10^{-4}$)*	56 ($P < 1 \times 10^{-4}$)
FENO (n = 717; 74.3%)	232 ($P < 1 \times 10^{-4}$)	128 ($P = .086$)	312 ($P < 1 \times 10^{-4}$)
Allergic asthma (n = 277; 28.7%)	107 ($P < 1 \times 10^{-4}$)	25 ($P < 1 \times 10^{-4}$)*	134 ($P = 2 \times 10^{-4}$)
N-correlated genes ($\rho >$ |0.15|)†	228	248	126
Top enriched KEGG pathway	Lysosome	Ribosome	Inflammatory mediator regulation of transient receptor potential channels

FENO, Fractional exhaled nitric oxide; *KEGG*, Kyoto Encyclopedia of Genes and Genomes
Uncorrelated DMCs (gray module) are not shown.
Permutations *P* values are described in the Methods section.
*Significant depletion of overlapping CpGs.
†Correlated genes have a $\rho >$ |0.15| using all children first (n = 266), but remain correlated when testing only in control children (n = 250). All pathways are listed in Table E8.

DMCs associated with rhinitis, 3 of which overlapped with AR DMCs. None remained after excluding AR from rhinitis cases (23 cases and 387 controls). No DMCs were identified for atopic dermatitis or sensitization to food allergen. These results highlight the specificity of these associations with the AR phenotype.

To assess the robustness of our findings, we compared the DMCs identified in our study with those reported in Cardenas et al,[5] who explored the relationships between methylation and asthma-related traits in DNA from nasal swab cells from 547 multiethnic teenage children, using the same DNAm array as in our study. Overall, 722 of the DMCs in our study (76%) were associated with at least 1 phenotype in their study[5] (Table E4). Of the 8777 DMCs reported in the Cardenas et al study for any phenotype, 4501 (51%) were associated with AR in our study at $P < .05$ (see Table E5 in this article's Online Repository at www.jacionline.org). This indicates that DNAm results are both consistent and stable prepuberty and postpuberty and across ethnicities, and that many DMCs are shared across allergy- and asthma-associated phenotypes.

Because mucosal cells are a mix of cell types, we used upper airway cell–specific gene signatures to explore the possibility that DNAm differences between AR cases and controls were due to cell-type heterogeneity between AR cases and controls[36] (Fig 1). We first tested for correlation between each cell-type–specific gene expression signatures and each latent factor[28] for DNAm and for gene expression (see Methods). There were strong correlations between cell-type signatures and latent factors 1, 3, 6, and

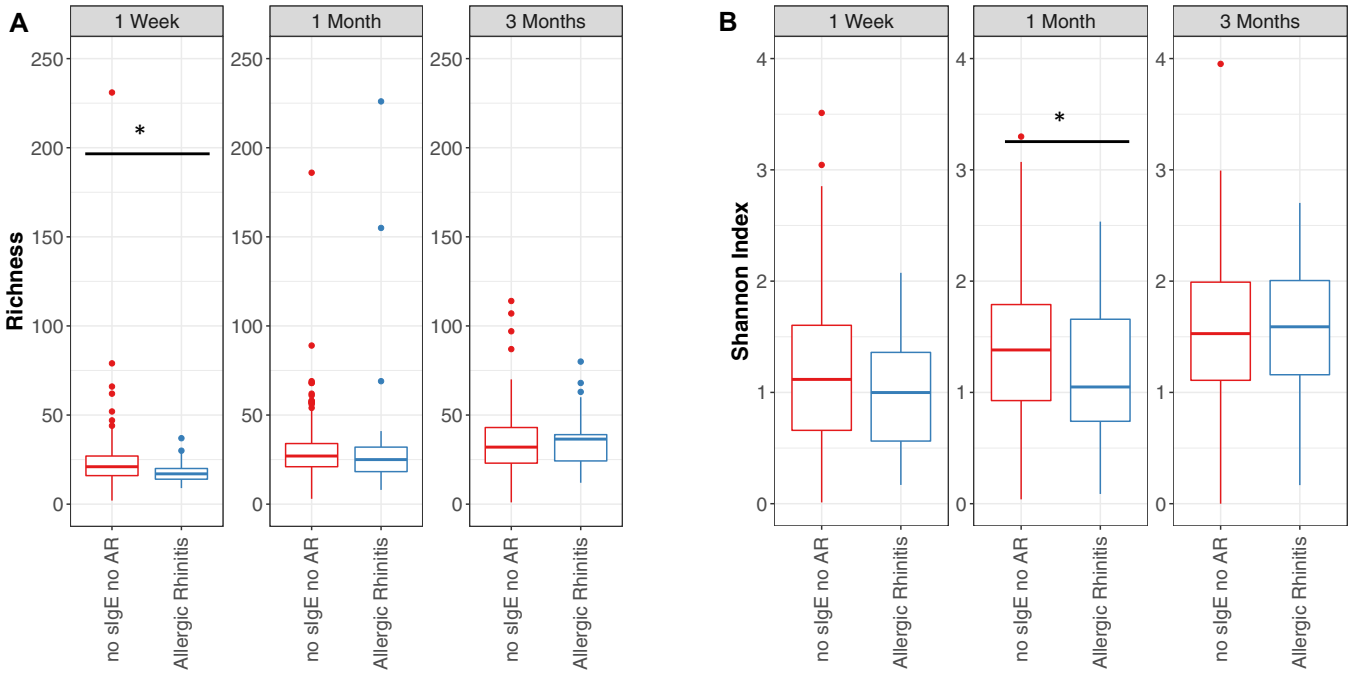

FIG 2. Early-life microbiota composition and the development of AR at age 6 years. Differences between children with AR compared with children without AR for (**A**) richness and (**B**) Shannon index at 1 week (n = 361), 1 month (n = 441), and 3 months (n = 445) of life. *sIgE*, Specific IgE. Significant differences ($P < .05$; Wilcoxon rank sum test) are indicated by asterisk. Richness at 1 week $P = .0079$, median AR =17 and controls = 21; 1 month $P = .34$, median AR = 25 and controls = 27; 3 months $P = .40$, median AR = 36.5 and controls = 32. Shannon diversity index at 1 week $P = .30$, median AR = 1 and controls = 1.11; 1 month $P = .045$, median AR = 1.05 and controls = 1.38; 3 months $P = .76$, median AR = 1.59 and controls = 1.53.

7 for DNAm (Fig 1, *A*) and latent factors 1 through 4 for gene expression (Fig 1, *B*). Importantly, none of the cell-type signatures differed between AR cases and controls (Fig 1, *C*). These observations indicated that the latent factors effectively accounted for variation in cell composition and that differences in cell composition do not underlie the observed methylation differences between the AR cases and controls.

Comethylation networks of DMCs reveal gene expression signatures and pathway enrichments

We used the R package WGCNA to further characterize the DMCs in nasal mucosa cells and assess their correlation structure.[37] Of the 965 DMCs, 792 (82%) formed 3 modules of correlated (comethylated) DMCs (Table I). Unsurprisingly, all 3 modules were correlated with AR (Wilcoxon rank-sum test–adjusted $P < 10^{-8}$; see Table E6 in this article's Online Repository at www.jacionline.org). Of all the exposure variables measured in these children, only a diagnosis of rhinitis in the mother was modestly associated with the blue module (adjusted $P = .041$; Table E6).

To draw additional biological inferences from these data, we examined gene expression profiles from RNA-seq studies in the same cells used for DNAm studies. First, we tested for correlations between the eigenvectors of each comethylation module and global gene expression, using all 15,363 genes detected as expressed in nasal mucosa cells and a liberal threshold for correlation of Spearman rho more than $|0.15|$ and P less than or equal to .015. We considered only those genes that were correlated in the combined samples and in only the control subjects to avoid spurious correlations. This analysis revealed 228 genes whose expression was correlated with the eigenvector for the blue module, 248 with the eigenvector for the brown module, and 126 with the eigenvector for the turquoise module (see Table E7 in this article's Online Repository at www.jacionline.org). The genes correlated with the eigenvector for the blue module were enriched in pathways associated with "lysosome" and "bacterial invasion of epithelial cells"; those with the eigenvector for the brown module were enriched in pathways associated with "ribosome", "cytokine-cytokine receptor interaction" and "oxidative phosphorylation"; and those with the eigenvector for the turquoise module were enriched in pathways associated with "inflammatory mediator regulation of transient receptor potential channels" and "Influenza A susceptibility" (see Table E8 in this article's Online Repository at www.jacionline.org). These combined data indicate that the DMCs in the different modules reflect different biological processes and potentially distinct facets of the AR phenotype.

Children with AR at age 6 years had less diverse airways microbiota in early life

The blue comethylation module was correlated with the expression of genes enriched in "bacterial invasion of epithelial cells", suggesting a connection to microbial exposures. Therefore, we next explored relationships between upper airway microbiota in early life and both AR and DNAm patterns at age 6 years. We first examined the bacterial composition in the

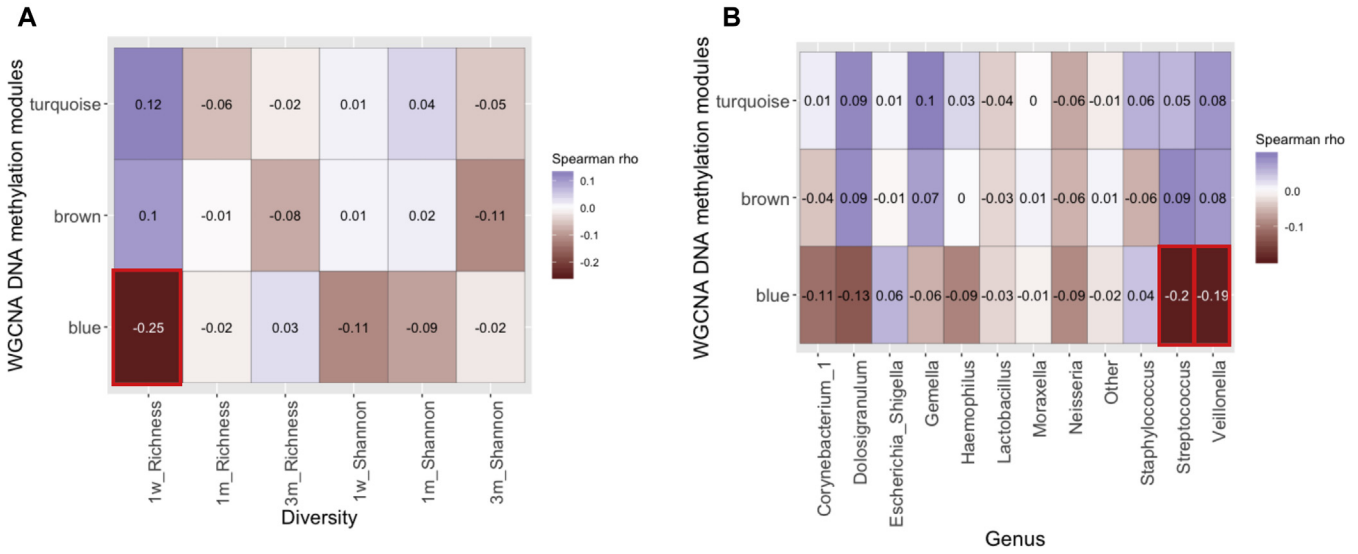

FIG 3. Microbial diversity and RAs of microbial genera are correlated with the blue WGCNA comethylation module. Spearman correlation between (**A**) microbial diversity measures at 1 week (n = 268), 1 month (n = 326), and 3 months (n = 332) and (**B**) RA of genera with the RA of all taxa occurring more than 1% at 1 week and WGCNA comethylation modules. *sIgE*, Specific IgE. Red outlines highlight Spearman correlation *P* values reaching significance after applying Benjamini-Hochberg procedure.

airways at age 1 week, 1 month, and 3 months and AR status at age 6 years.[22] This revealed significantly less richness at age 1 week in children with AR at age 6 years (median richness = 17 and 21, respectively, Wilcoxon rank-sum test $P = .0079$; Fig 2, *A*) and a lower Shannon diversity index at 1 month (median Shannon diversity index = 1.05 and 1.38, respectively; Wilcoxon rank-sum test $P = .045$; Fig 2, *B*) compared with control children. No differences were observed at 3 months (Fig 2). Read counts did not differ between the AR cases and controls at any of the 3 time points (see Fig E4 in this article's Online Repository at www.jacionline.org). Richness and Shannon index were correlated with each other at each time point, but neither was correlated between time points (see Fig E5 in this article's Online Repository at www.jacionline.org), reflecting the dynamic changes that occur in the first 3 months of life. To confirm the robustness of these results, we performed a rarefied richness analysis, in which similar results were observed (median rarefied richness = 17.2 and 20.4, respectively, Wilcoxon rank-sum test $P = .015$; see Fig E6, *A*, in this article's Online Repository at www.jacionline.org), as well as repeated the analysis with 1 outlier removed in the controls at 1 week (median richness = 17 and 21, respectively, Wilcoxon rank-sum test $P = .0083$; see Fig E7 in this article's Online Repository at www.jacionline.org). Furthermore, richness and Shannon index were not correlated with any of the 17 exposures measured in these children (Table E2), and these exposures did not significantly impact the association between richness at 1 week and AR, after adjusting for multiple testing (see Tables E3 and E9 in this article's Online Repository at www.jacionline.org), although some exposures reduced the effect of Shannon index at 1 month on AR (Table E3).

To assess the specificity of these findings, we tested for associations between microbiota diversity and 3 other allergy-related traits (AS to aeroallergens, AS to food allergens, and atopic dermatitis) and 1 nonallergic trait (non-AR) at age 6 years (see Table E10 in this article's Online Repository at www.jacionline.org). Only children with AS had significantly lower richness at 1 week ($P = .0092$, median 19 in cases and 21 in controls). However, the difference was not significant after removing children with AR from the analysis ($P = .14$; see Table E11 in this article's Online Repository at www.jacionline.org), suggesting that the difference was driven by AR and not AS.

Upper airway microbiota diversity is correlated with the blue comethylation module

To determine whether microbiota diversity in the first few months of life is correlated with DNAm patterns at age 6 years, we first confirmed that significant differences at 1 week were observed when including only those children who also had DNAm measured at age 6 years (Wilcoxon rank-sum test $P = .0033$; n = 268; see Fig E8 in this article's Online Repository at www.jacionline.org), and then tested for correlation between the 2 diversity measures at each age with each of the 3 DNAm modules. Richness at age 1 week was associated only with the blue modules eigenvector (Spearman $\rho = -0.25$, $P = 3.27 \times 10^{-5}$; $P_{adj} = 5.9 \times 10^{-4}$; n = 268; Fig 3, *A*; see Fig E9 in this article's Online Repository at www.jacionline.org), highlighting a potential signature of the early-life microbial composition on DNAm patterning that is specifically captured by this module. Richness at other ages and Shannon index at any age were not correlated with any of the module eigenvectors (Fig 3, *A*). To exclude the possibility that the correlation between richness at 1 week and the blue module eigenvector was because both the diversity index and the DMCs in the modules were associated with AR, we tested for correlation only in the controls (n = 251). After removing the AR cases, the correlation between richness at 1 week and the blue module eigenvector remained significant (Spearman $\rho = -0.21$, $P = 6.1 \times 10^{-4}$, $P_{adj} = .01$), indicating a robust correlation between these 2 metrics even in the absence of AR.

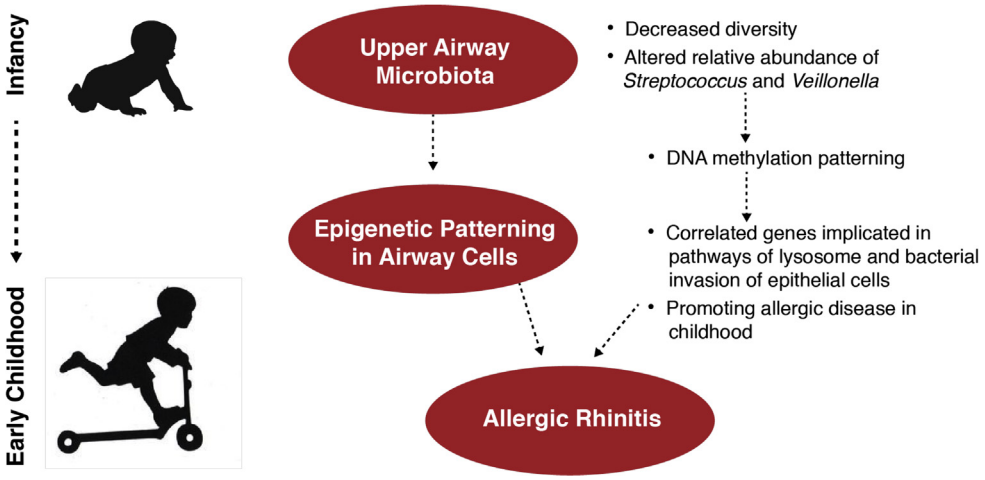

FIG 4. Early-life microbial environment contributes to the development of AR though modification of the epigenetic landscape.

We next assessed whether the association between microbiota richness at 1 week and AR at age 6 years was mediated by the DNAm variation captured by the blue module. We first used logistic regression, including sex and read counts as covariates, to confirm the relationship between richness at 1 week and AR at age 6 years using a parametric model ($P = .041$; $\beta = -1.12$; 95% CI, −2.28 to −0.24). We then included the blue module eigenvector as a covariate in the logistic regression. In the latter model, the estimate of the effect of richness at 1 week on AR at age 6 years was reduced by 61% and the correlation was no longer significant ($P = .53$; $\beta = -0.58$; 95% CI, −1.70 to 0.77). This indicated that the nasal mucosal cell DNAm patterns captured by the blue module accounted for more than half of the effect of microbiota composition in early life on the development of AR in childhood. The association between richness at 1 week and AR at age 6 years remained significant when using rarefied richness ($P = .023$; $\beta = -1.69$; 95% CI, −3.11 to −0.31), and the estimate of rarefied richness on AR was also reduced by 58% when the blue module eigenvector was included in the model ($P = .49$; $\beta = -0.74$; 95% CI, −2.4 to 0.86).

Finally, we assessed correlations between the RA of bacteria at 1 week (genus level) with the modules. Although there were no differences in the RAs of any of the genera between AR cases and controls (see Fig E10 in this article's Online Repository at www.jacionline.org), the RAs of 2 genera were specifically correlated with the blue module eigenvector (*Streptococcus* [$\rho = -0.20$; $P = .0013$; mean RA = 17.1%]; *Veillonella* [$\rho = -0.19$; $P = .0013$; mean RA = 2.3%]; $P_{adj} = .026$; n = 268) (Fig 3, *B*). AR was positively correlated with the blue module eigenvector, whereas the abundances of these genera were negatively associated, suggesting that lower RA of *Streptococcus* and *Veillonella* at 1 week were associated with risk for AR.

DISCUSSION

A cardinal feature of AR is inflammation of the upper airway nasal mucosal tissue, which forms the first line of defense to inhaled airborne particles and orchestrates the downstream host responses. By focusing our studies on this tissue in children participating in a prospective birth cohort, we were able to explore mechanistic links of upper airway microbiota at 1 week with epigenetic modifications and AR at age 6 years. This led to several novel observations.

First, our results support the hypothesis that colonization of bacteria in the upper airways in very early life plays an important role in shaping epigenetics profiles in the nasal epithelium that persist at least to later childhood, and contribute to the development of AR. To our knowledge, this study is the first to show a longitudinal connection between microbiota and DNAm profiles in the upper airways, indicating lasting microbial effects on epigenetic patterns in the development of AR. These results suggest both that DNAm changes may be a persistent global marker of early-life exposures and that effects of methylation changes on gene expression are present in later childhood. Taken together these observations may reflect coordinated effects of DNAm patterns on gene expression that are associated with the development of AR.

Second, combining our results with those from another study demonstrated that the methylation findings in nasal mucosal cells are overall reproducible. More than 70% of the DMCs identified in our study were also differentially methylated in a study of asthma-related phenotypes in children that also used nasal brushings and the same DNAm array as in our study.[5] We further showed the specificity of our findings for AR within our study cohort. Despite the larger sample sizes, remarkably few DMCs were associated with rhinitis (sans sensitization) or sensitization (sans rhinitis) and those few were nearly completely overlapping with the AR DMCs. Most DMCs were less methylated in children with AR compared with controls, which is consistent with results observed in other studies of related traits.[5] Although we did not have direct estimates of cell-type proportions, analyses of cell-specific gene expression signatures confirmed that the latent factors we used to correct for unwanted variation in the DNAm and gene expression data captured differences due to cell heterogeneity. This allowed us to conclude that the observed differences between AR cases were not due to differences in cell composition between these children.

Third, we showed that each of the 3 comethylation modules were all associated with AR but correlated with genes enriched in

different pathways, thus capturing different features of AR. The blue module eigenvector was correlated with bacterial richness at age 1 week, with the RA of *Streptococcus* and *Veillonella*, and with genes enriched in bacterial invasion of epithelial cells and lysosome function pathways. Because the studies of microbiota were performed on samples collected during infancy, before the onset of AR, these data raise the possibility that the DNAm patterning in the blue module preceded and contributed to the development of AR, a causal trajectory supported by a mediation analysis (Fig 4). Remarkably, we observed a relationship between DNAm and microbial diversity only at 1 week, but not at the other time points or with other diversity metrics. We interpret this to indicate a narrow developmental window during which reduced microbiota richness can lead to the development of AR. Furthermore, the RAs of 2 genera at 1 week were correlated with the blue module eigenvector. Even though the RAs of the 2 genera were not significantly different between AR cases and controls at 1 week, each was significantly negatively correlated with the blue module eigenvector. These 2 genera have been shown to co-occur and act in a synergistic manner: *Streptococcus* catabolizes carbohydrates to lactic acid (among others), and lactic acid can be used by *Veillonella* as a source of carbon to metabolize lactate into short-chain fatty acids (acetate and propionate). The latter has been shown previously to modify host chromatin states, another epigenetic mark.[48] This could explain the link between the 2 genera and the epigenetic landscape reflected in the blue module. Although we cannot rule out the possibility that DNAm at age 6 years was patterned by an unmeasured factor that also preceded and led to the observed microbiota diversity at age 1 week, we were able to exclude many exposures measured in these children because neither microbial diversity (Table E9) nor the DNAm module eigenvectors (Table E6) were correlated with the measured exposures, ruling out at least some potentially relevant exposures. The DNAm patterns in the brown and turquoise models were correlated with AR, but not with any of the microbiota indices or measured exposures. These epigenetic patterns may therefore be a result of the disease process itself or other exposures not measured in this study.

The population-based ascertainment of infants enrolled in the COPSAC, without regard to disease risk, is a strength of our study because it reduces confounding due to inherent differences between children at high risk for disease development and those who are not, and makes our results generalizable. However, this is also a limitation because relatively few children had AR at 6 years, which impacted the power to detect differences. Therefore, we may have underestimated the number of AR-associated DMCs or missed associations between AR and microbial diversity or abundance in early life. A second limitation is that we did not have DNAm profiles in upper airway cells before the age of 6 years, and cannot observe the longitudinal effects of microbial diversity in infancy on epigenetic modifications or of epigenetic modifications on the development of AR. As a result, we are unable to determine whether the composition of early-life upper airway microbiota influences the epigenetic patterns in upper airway cells, as we propose here, or whether the DNAm patterns in upper airway cells preceded and influenced microbial diversity in early life, or that they were both influenced by an unmeasured exposure. Furthermore, the DNAm profiles and microbiota composition were measured in different upper airway niches (inferior turbinate vs hypopharynx, respectively). Finally, this study was conducted in children of European ancestry. However, the Cardenas study,[5] which included ethnically diverse children, showed high concordance of methylation results with ours, suggesting that the DNAm results are robust to ancestry, although future studies in non-European populations are needed to affirm this assumption and to discover ancestry-specific epigenetic effects.

Conclusions

We provide data in support of the hypothesis that the early-life microbial environment contributes to the risk of developing AR in childhood by shaping the epigenetic landscape in upper airway mucosal cells. These epigenetic patterns lead to perturbations of genes in lysosome and bacterial invasion of epithelial cells pathways in upper airway mucosal cells that persist into later childhood. Taken together, our study supports the view that interventions for modulating microbial composition, as early as the first week of life, could have a significant impact on both the quality of life for people and the economic burdens of nations associated with allergic diseases throughout the world.[49]

We thank the children and families of the $COPSAC_{2010}$ cohort for their participation and commitment. We also thank Andres Cardenas for helpful discussion and advice on data processing.

Key messages

- **Early-life upper airway microbial diversity is lower in children who develop AR by age 6 years.**
- **Epigenetic patterning in upper airway mucosal cells mediates the effects of early-life microbial diversity on the development of AR in later childhood.**

REFERENCES

1. Greiner AN, Hellings PW, Rotiroti G, Scadding GK. Allergic rhinitis. Lancet 2011;378:2112-22.
2. Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nat Genet 2018;50:1072-80.
3. Schoos AM, Chawes BL, Jelding-Dannemand E, Elfman LB, Bisgaard H. Early indoor aeroallergen exposure is not associated with development of sensitization or allergic rhinitis in high-risk children. Allergy 2016;71:684-91.
4. Wang N, Schoos AM, Larsen JM, Brix S, Thysen AH, Rasmussen MA, et al. Reduced IL-2 response from peripheral blood mononuclear cells exposed to bacteria at 6 months of age is associated with elevated total-IgE and allergic rhinitis during the first 7 years of life. EBioMedicine 2019;43:587-93.
5. Cardenas A, Sordillo JE, Rifas-Shiman SL, Chung W, Liang L, Coull BA, et al. The nasal methylome as a biomarker of asthma and airway inflammation in children. Nat Commun 2019;10:3095.
6. Forno E, Wang T, Qi C, Yan Q, Xu CJ, Boutaoui N, et al. DNA methylation in nasal epithelium, atopy, and atopic asthma in children: a genome-wide study. Lancet Respir Med 2019;7:336-46.
7. Yang IV, Pedersen BS, Liu AH, O'Connor GT, Pillai D, Kattan M, et al. The nasal methylome and childhood atopic asthma. J Allergy Clin Immunol 2017;139:1478-88.
8. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight 2016;1:e90151.
9. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. Science 2016;352:539-44.
10. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486:207-14.
11. Man WH, Clerc M, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Kool J, et al. Loss of microbial topography between oral and nasopharyngeal microbiota and development of respiratory infections early in life. Am J Respir Crit Care Med 2019;200:760-70.

12. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bonnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med 2007;357:1487-95.
13. Thorsen J, Rasmussen MA, Waage J, Mortensen M, Brejnrod A, Bonnelykke K, et al. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. Nat Commun 2019;10:5001.
14. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe 2015;17:704-15.
15. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, et al. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences: a prospective cohort study. Am J Respir Crit Care Med 2017;196:1582-90.
16. Ta LDH, Yap GC, Tay CJX, Lim ASM, Huang CH, Chu CW, et al. Establishment of the nasal microbiota in the first 18 months of life: correlation with early-onset rhinitis and wheezing. J Allergy Clin Immunol 2018;142:86-95.
17. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Folsgaard NV, Kreiner-Moller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. Clin Exp Allergy 2013;43:1384-94.
18. Bisgaard H, Stokholm J, Chawes BL, Vissing NH, Bjarnadottir E, Schoos AM, et al. Fish oil-derived fatty acids in pregnancy and wheeze and asthma in offspring. N Engl J Med 2016;375:2530-9.
19. Chawes BL, Bonnelykke K, Stokholm J, Vissing NH, Bjarnadottir E, Schoos AM, et al. Effect of vitamin D3 supplementation during pregnancy on risk of persistent wheeze in the offspring: a randomized clinical trial. JAMA 2016; 315:353-61.
20. Schoos AM, Jelding-Dannemand E, Stokholm J, Bonnelykke K, Bisgaard H, Chawes BL. Single and multiple time-point allergic sensitization during childhood and risk of asthma by age 13. Pediatr Allergy Immunol 2019;30:716-23.
21. Thorsteinsdottir S, Thyssen JP, Stokholm J, Vissing NH, Waage J, Bisgaard H. Domestic dog exposure at birth reduces the incidence of atopic dermatitis. Allergy 2016;71:1736-44.
22. Mortensen MS, Brejnrod AD, Roggenbuck M, Abu Al-Soud W, Balle C, Krogfelt KA, et al. The developing hypopharyngeal microbiota in early life. Microbiome 2016;4:70.
23. Fortin JP, Triche TJ Jr, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics 2017;33: 558-60.
24. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17:208.
25. Maksimovic J, Gordon L, Oshlack A. SWAN: subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol 2012;13:R44.
26. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547-8.
27. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
28. McKennan C, Nicolae D. Accounting for unobserved covariates with varying degrees of estimability in high-dimensional biological data. Biometrika 2019;106: 823-40.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
30. Andrews S. FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed July 15, 2019.
31. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047-8.
32. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
33. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 2012;91:839-48.
34. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;11:R25.
35. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 2014;15:R29.
36. Ordovas-Montanes J, Dwyer DF, Nyquist SK, Buchheit KM, Vukovic M, Deb C, et al. Allergic inflammatory memory in human respiratory epithelial progenitor cells. Nature 2018;560:649-54.
37. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform 2008;9:559.
38. Turner S, Pryer KM, Miao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol 1999;46:327-38.
39. Takai K, Horikoshi K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl Environ Microbiol 2000;66:5066-72.
40. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. Nat Commun 2018;9:141.
41. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 2011;17:10-2.
42. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852-7.
43. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 2016; 13:581-3.
44. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 2013;41:D590-6.
45. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 2018;6:90.
46. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis 2015;26:27663.
47. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-6. 2019. https://CRAN.R-project.org/package=vegan.
48. Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME, et al. Diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. Mol Cell 2016;64:982-92.
49. Meltzer EO, Blaiss MS, Derebery MJ, Mahr TA, Gordon BR, Sheth KK, et al. Burden of allergic rhinitis: results from the Pediatric Allergies in America survey. J Allergy Clin Immunol 2009;124:S43-70.