

DR. STUART E TURVEY (Orcid ID : 0000-0003-1599-1065)

Article type : Letter to the Editor

LETTER TO THE EDITOR

Mining the infant gut microbiota for therapeutic targets against atopic disease

Boutin, Rozlyn C.T., BScH^{1,2†}, Sbihi, Hind, PhD^{1,3†}, Dsouza, Melissa, PhD^{4†}, Malhotra, Raunaq, PhD⁵, Petersen, Charisse, PhD^{1,3}, Dai, Darlene, MSc^{1,3}, Sears, Malcolm R., MBChB⁶, Moraes, Theo J., MD PhD⁷, Becker, Allan B., MD⁸, Azad, Meghan B., PhD⁸, Mandhane, Piush J., MD PhD⁹, Subbarao, Padmaja, MD^{7‡}, Finlay, B. Brett, PhD^{1,2‡}, Turvey, Stuart E., MBBS DPhil^{1,3‡*}

† Equal contribution as first authors; ‡ Equal contribution as senior authors

¹University of British Columbia, Vancouver, Canada; ²Michael Smith Laboratories, Vancouver, Canada; ³British Columbia Children's Hospital, Vancouver, Canada; ⁴Commense Inc. Current affiliation Vedanta Biosciences Inc., Boston, USA; ⁵Commense Inc. Current affiliation GNS Healthcare Inc., Boston, USA; ⁶McMaster University, Hamilton, Canada; ⁷Hospital for Sick Children, Toronto, Canada; ⁸University of Manitoba, Winnipeg, Canada; ⁹University of Alberta, Edmonton, Canada.

*Corresponding Author:

Stuart E. Turvey, MBBS, DPhil, FRCPC

BC Children's Hospital

950 West 28th Avenue

Vancouver, BC, V5Z 4H4, Canada

Phone: 604 875 2345 ext. 3130

Email: sturvey@cw.bc.ca

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/ALL.14244](#)

This article is protected by copyright. All rights reserved

Capsule Summary: Increased relative abundance of nine bacterial genera within the infant gut microbiota is associated with a reduced risk of developing a spectrum of atopic disease outcomes at age one year.

Key Words: Atopic Dermatitis; Atopy; Wheeze; Food Sensitization; Microbiota; Live Biotherapeutic Product.

List of abbreviations:

LBP: live biotherapeutic product

rDNA: ribosomal deoxyribonucleic acid

SCFA: Short-chain fatty acid

To the Editor:

Forty percent of the world's population is sensitized to at least one environmental allergen and atopic diseases such as asthma, atopic dermatitis, and food allergy are among the most frequent health challenges faced by children today. Atopic conditions often manifest differently in similarly predisposed individuals and frequently co-occur within the same individual over time; a phenomenon best understood in the context of the "atopic march" (1). Yet the underlying causes of many allergic diseases are not fully understood and no effective curative or preventative immunomodulatory therapies exist. Recently, however, multiple converging lines of evidence suggest that the gut microbiota may play a role in the etiology of atopic disease, especially during a critical window early in life when both the microbiota and infant immune system are maturing (2,3). To date, most studies have focused on single atopic disease outcomes. Here we leverage the richness of the outcomes assessed in the CHILD Cohort Study to investigate the relationship between the early life gut microbiota and the development of several early life atopic conditions, with the goal of identifying specific 'missing' microbes involved in the natural history of the atopic march. Deliberate early life supplementation with specific 'missing' immunomodulatory microbes may represent a novel approach to the primary prevention of multiple human atopic diseases and atopic multimorbidity.

We hypothesized that a common set of microbes in the early life gut microbiota of a precisely defined cohort of "healthy" children would be associated with a decreased risk of atopic dermatitis (AD), food and inhalant allergen sensitization, and recurrent wheeze (RW; one of the strongest risk factors for asthma) at age one year. To test this hypothesis, we applied a series of complementary machine learning approaches (Figures E1 & E2) to integrate comprehensive clinical data in the CHILD cohort with 16S rDNA amplicon sequencing data from stool samples collected at three months of age (n=837; see Supplement for full description of Methods and Table E1). Using this approach, we nominated a consortium of microbes in the infant gut as candidates for inclusion in a live biotherapeutic product (LBP) that can be tested for its potential to prevent the onset of a variety of frequently co-occurring atopic disease manifestations (see Supplement reference 43 and relevant patent applications listed in the Supplement describing current trials using a similar approach).

Clinical trials designed to test the efficacy of LBPs for the prevention of atopic disease are constrained by the necessity to evaluate early measures of efficacy, and age one year is thus a desirable end point. The clinical relevance of the one year allergic phenotypes used in this study is emphasized by the fact that infants with a positive atopic composite index (defined as having two or more atopic outcomes (AD, food sensitization (FS), RW, or atopy plus wheeze (AW)) at age one year were 2.1 times more likely than healthy controls to have a positive allergic composite index at age five years (defined as having two or more diagnoses of asthma, allergic rhinitis, AD, or FS) and 2-10 times more likely to have one of asthma, allergic rhinitis, AD, or FS at age five years (Figure E3). Healthy controls were defined as children who were negative for all signs and symptoms of allergic disease (RW, asthma, AD, food allergy, allergic rhinitis, and sensitization to food or inhalant allergens) up to age five years.

Consistent with previous studies (4), we observed differences in global measures of gut microbiota diversity at three months of age between infants who remained healthy and those with an allergic diagnosis at age one year. Specifically, increased bacterial diversity (as quantified by the Shannon α -diversity index) was associated with protection against allergic sensitization (atopy), wheeze, AW, FS, and AD (Figures 1A & E4). We then sought to determine whether specific microbes might be involved in driving these effects as these may represent therapeutic targets. We applied three complementary machine learning techniques (details in Supplement) on 16S rRNA amplicon sequence data taxonomically annotated to the genus level to identify bacterial genera that differed in relative abundance between healthy children and children who developed an allergic disorder by age one year. As anticipated, specific allergic disorders showed significant overlap within single individuals at one year of age (Figure 1B): AD, FS, RW, and AW (Figure 1B).

Bacterial genera identified by at least one machine learning method to be consistently reduced in relative abundance among cases relative to healthy controls and found in at least 15% of healthy control samples were included in the final nomination. Using these methods, we identified nine bacterial genera whose relative abundances were inversely related to the risk of developing at least two allergic disease manifestations at age one year (Table 1; Figure 2).

Although it is now well established that gut microbiota communities of infants who go on to develop allergic manifestations differ from those of healthy children (5), few clinical trials have been completed in an attempt to translate these findings to clinical practice. Moreover, many of the small clinical trials completed to date have shown disappointing results (6,7), likely at least in part because many LBPs used in these trials have focused on only one or a few microbes with little population-based evidence to support their use in allergic disease. *Lactobacillus*, for instance, is commonly tested for its efficacy as a probiotic for allergy prevention, but has been found to have limited long-term efficacy. In light of our findings suggesting that *Lactobacillus* is not deficient in infants who develop atopic disease manifestations by age one year, these studies together suggest that the design of effective microbiota-based therapeutics requires tailoring to a carefully defined target population. Additionally, many functional features of the gut microbiota associated with protection against allergy, such as the production of short-chain fatty acid (SCFA) metabolites, depend on the existence of a complex, dynamic community of microbes. Thus, supplementation with a microbial consortium rather than individual microbes may be a more effective strategy for modulating the development of allergic disease (8,9). Based on our results, we propose the assessment of a nine-component microbial cocktail comprised of bacteria from the following genera for use in a clinical trial as a LBP for the primary prevention of allergic disease and atopic multimorbidity: *Faecalibacterium*, *Lachnospira*, *Coproccoccus* (*Anaerostipes/Eubacterium_E*), *Oscillospira* (*Intestinimonas/Flavonifractor*), *Roseburia*, *Blautia*, *Dorea* (*Tyzzereella*), *Parabacteroides*, and *Ruminococcus* (Figure 2, Table 1, Table E3). Given that the functional capacities of bacterial species and strains within a specific genus can be vastly disparate, the cultivation and functional testing of bacterial strains isolated from healthy children in animal and other models will be critical to the development of such a LBP.

Eight of the nine nominated bacterial genera belong to the class Clostridia, seven are known producers of SCFAs (Table 1; Table E4), none are known pathogens or pathobionts, and many of the nominated genera have been identified in other studies to be associated with atopic outcomes and/or have beneficial immunomodulatory effects (see Supplement for additional references). Consistent with our results, levels of fecal SCFAs including acetate, butyrate, and propionate are often found to be reduced in infants at risk of allergic diseases and data from animal models indicates that SCFA-producing bacteria in the gut protect against allergic diseases through the promotion of anti-

Accepted Article

inflammatory and tolerogenic immune responses. Thus, our results suggest that supplementation with a LBP comprised of multiple bacterial genera may prevent the onset or progression of a wide array of allergic disease manifestations. Future work will validate these findings, verify the functional SCFA-producing capacities of the LBP through metagenomic sequencing and/or experimental approaches, and determine whether administration of the proposed LBP in infancy is safe and can protect against the onset and/or reduce the severity of allergic disease manifestations in mice and humans. We anticipate that these findings will be useful in guiding the development of future rigorous clinical trials.

References

1. Zheng T, Yu J, Oh MH, Zhu Z. The atopic march: Progression from atopic dermatitis to allergic rhinitis and asthma. *Allergy, Asthma Immunol Res* 2011;**3**:67–73.
2. Stiemsma LT, Turvey SE. Asthma and the microbiome: Defining the critical window in early life. *Allergy, Asthma Clin Immunol* 2017;**13**:1–9.
3. Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* 2015;**7**:307ra152.
4. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 2016;**8**:343–382.
5. Boutin RCT, Finlay BB. Microbiota-mediated immunomodulation and asthma: Current and future perspectives. *Curr Treat Options Allergy* 2016;**3**:292–309.
6. Tan-Lim, CSC & Esteban-Ipac, NAR. Probiotics as treatment for food allergies among pediatric patients: a meta-analysis. *World Allergy Organ. J.* **11**, 1–13 (2018)
7. Cabana MD, McKean M, Caughey AB, Fong L, Lynch S, Wong A et al. Early Probiotic Supplementation for Eczema and Asthma Prevention: A Randomized Controlled Trial. *Pediatrics* 2018;**142**:S213.1-S213.
8. Stein RR, Tanoue T, Szabady RL, Bhattarai SK, Olle B, Norman JM et al. Computer-guided design of optimal microbial consortia for immune system modulation. *Elife* 2018;**7**:1–17.
9. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N et al. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* 2016;**535**:94–103.

Boutin, Rozlyn C.T., BScH^{1,2†}, Sbihi, Hind, PhD^{1,3†}, Dsouza, Melissa, PhD^{4†}, Malhotra, Raunaq, PhD⁵, Petersen, Charisse, PhD^{1,3}, Dai, Darlene, MSc^{1,3}, Sears, Malcolm R., MBChB⁶, Moraes, Theo J., MD PhD⁷, Becker, Allan B., MD⁸, Azad, Meghan B., PhD⁸, Mandhane, Piush J., MD PhD⁹, Subbarao, Padmaja, MD^{7‡}, Finlay, B. Brett, PhD^{1,2‡}, Turvey, Stuart E., MBBS DPhil^{1,3‡*}

† Equal contribution as first authors; ‡ Equal contribution as senior authors

¹University of British Columbia, Vancouver, Canada; ²Michael Smith Laboratories, Vancouver, Canada; ³British Columbia Children's Hospital, Vancouver, Canada; ⁴Commense Inc. Current affiliation Vedanta Biosciences Inc., Boston, USA; ⁵Commense Inc. Current affiliation GNS Healthcare Inc., Boston, USA; ⁶McMaster University, Hamilton, Canada; ⁷Hospital for Sick Children, Toronto, Canada; ⁸University of Manitoba, Winnipeg, Canada; ⁹University of Alberta, Edmonton, Canada.

***Corresponding Author:**

Stuart E. Turvey, MBBS, DPhil, FRCPC
BC Children's Hospital
950 West 28th Avenue
Vancouver, BC, V5Z 4H4, Canada
Phone: 604 875 2345 ext. 3130
Email: sturvey@cw.bc.ca

Funding: Core funding for the CHILD Cohort Study was provided by the Canadian Institutes of Health Research (CIHR; AEC-85761, PJT-148484, FDN-159935), the Allergy, Genes and Environment Network of Centres of Excellence (AllerGen NCE) (12CHILD) and Genome Canada/Genome BC (274CHI). R.C.T.B. is supported by a Vanier Canada Graduate Scholarship and a Vancouver Coastal Health-Canadian Institutes of Health Research (CIHR)-UBC MD/PhD Award. HS is supported by Michael Smith Foundation for Health Research Postdoctoral Trainee Fellowship and CIHR postdoctoral Fellowship. SET holds the Aubrey J Tingle Professorship in Pediatric Immunology and Canada Research Chair in Pediatric Precision Health.

Conflict of interest statement: MD and RM were employees of Commense Inc. when this work was completed and BBF is a Scientific Advisory Board Member of Commense Inc. A provisional patent related to this work entitled 'Microbiome compositions and methods of treatment of atopic diseases' has been filed. All other authors have no conflicts of interest in relation to this manuscript.

Figure legends:

Figure 1. (A) Risk of developing an allergic disease manifestation (RW: recurrent wheeze; ARW: Atopy+recurrent wheeze; AW: atopy+wheeze; Food sens: food sensitization; AD: atopic dermatitis) by age one year associated with one interquartile range (IQR) increase in the alpha diversity (Shannon index) of the three-month gut microbiota determined by univariate logistic regression in the CHILD cohort study sample included in this study. (B) Number of children with one or more of: atopy plus wheeze, food sensitization, recurrent wheeze, and atopic dermatitis at age one year.

Figure 2. Overlap of bacterial genera nominated as potentially protective against clinical indications of atopic dermatitis, food sensitization, recurrent wheeze, and atopy plus wheeze diagnosed one year after birth in a subset of infants from the CHILD Cohort Study based on 16S rRNA gene sequencing.

Table 1. Bacterial genera nominated as potentially protective against clinical indications of atopic dermatitis, food sensitization, recurrent wheeze, and atopy plus wheeze diagnosed one year after birth. Check marks indicate nominees for each clinical indication.

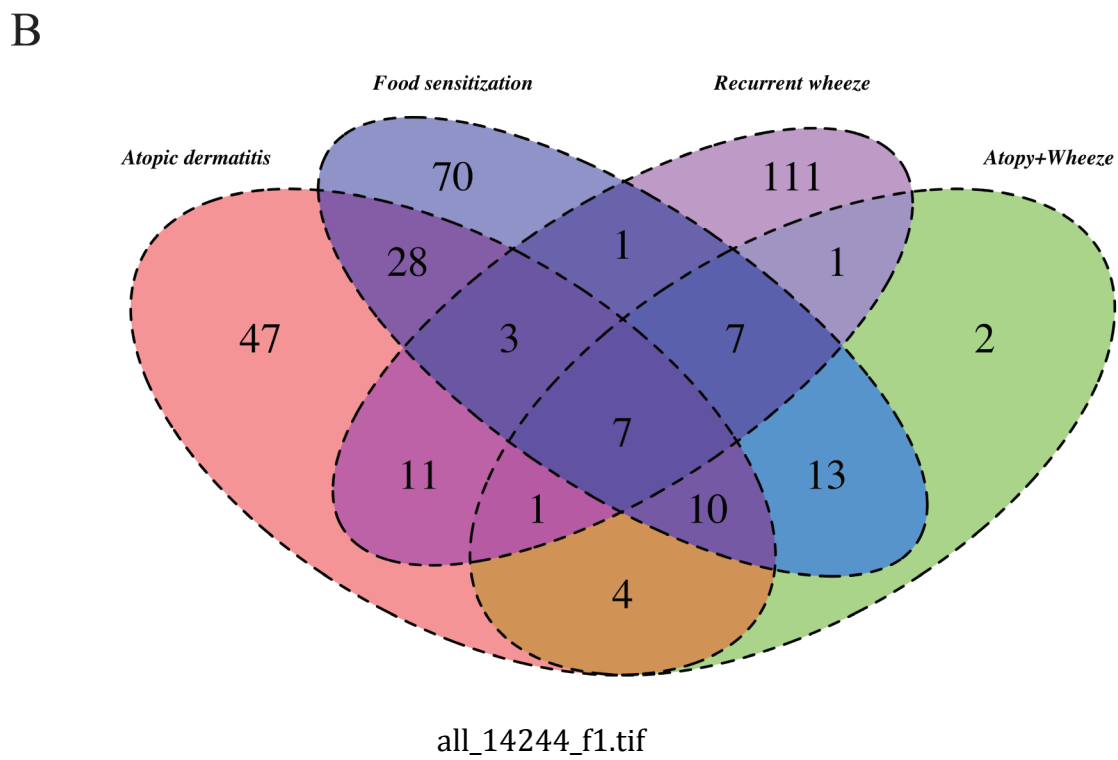
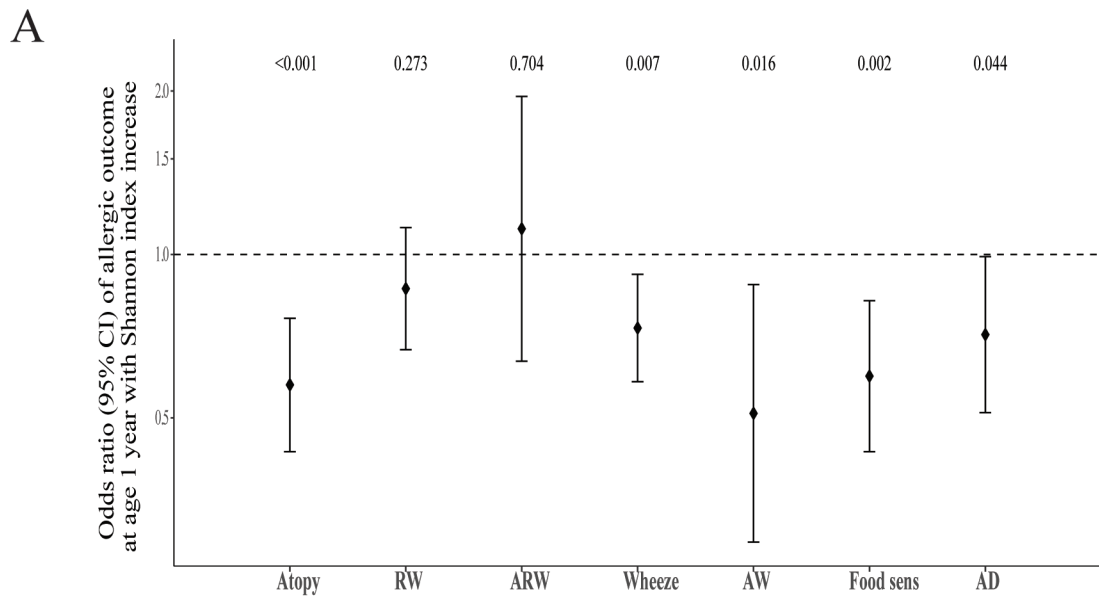
Nominated genera	Atopic dermatitis	Food sensitization	Recurrent Wheeze	Atopy + Wheeze
<i>Faecalibacterium</i> *	√	√	√	√
<i>Lachnospira</i> †		√	√	√
<i>Coprococcus</i> †	√	√	√	√
<i>Oscillospira</i> *	√	√	√	
<i>Roseburia</i> *◇	√	√	√	√
<i>Blautia</i> *	√	√		√
<i>Dorea</i> **	√	√		
<i>Parabacteroides</i>	√			√
<i>Ruminococcus</i> *	√	√		√

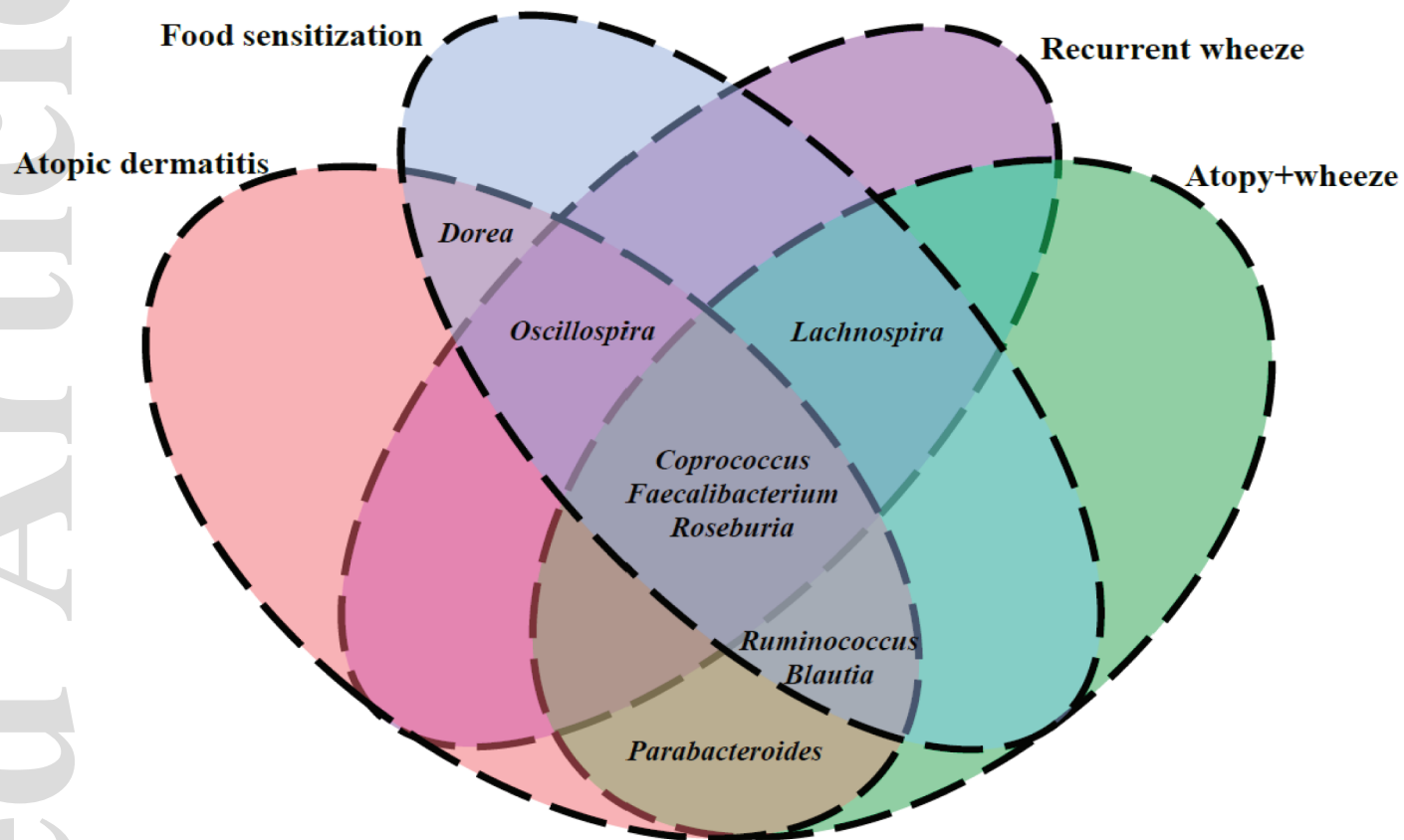
*=known butyrate producer

**=possible butyrate producer

†=known acetate producer

◇=known propionate producer





all_14244_f2.tif

SUPPLEMENTARY APPENDIX

Mining the infant gut microbiota for therapeutic targets against atopic disease

Boutin, Rozlyn C.T., BScH^{1,2†}, Sbihi, Hind, PhD^{1,3†}, Dsouza, Melissa, PhD^{4†}, Malhotra, Raunaq, PhD⁵, Petersen, Charisse, PhD^{1,3}, Dai, Darlene, MSc^{1,3}, Sears, Malcolm R., MBChB⁶, Moraes, Theo J., MD PhD⁷, Becker, Allan B., MD⁸, Azad, Meghan B., PhD⁸, Mandhane, Piush J., MD PhD⁹, Subbarao, Padmaja, MD^{7‡}, Finlay, B. Brett, PhD^{1,2‡}, Turvey, Stuart E., MBBS DPhil^{1,3‡*}
† Equal contribution as first authors; ‡ Equal contribution as senior authors

¹University of British Columbia, Vancouver, Canada; ²Michael Smith Laboratories, Vancouver, Canada; ³British Columbia Children's Hospital, Vancouver, Canada; ⁴Commense Inc. Current affiliation Vedanta Biosciences Inc., Boston, USA; ⁵Commense Inc. Current affiliation GNS Healthcare Inc., Boston, USA; ⁶McMaster University, Hamilton, Canada; ⁷Hospital for Sick Children, Toronto, Canada; ⁸University of Manitoba, Winnipeg, Canada; ⁹University of Alberta, Edmonton, Canada.

TABLE OF CONTENTS

Supplementary Methods	S3
Figure E1. An overview of the methods used to analyze amplicon-based microbiota data to nominate taxa enriched in healthy individuals versus individuals with an atopic disease manifestation at age one year.....	S9
Figure E2. Relative abundance of genera within the three-month-old gut microbiota reported as potential nomination for inclusion in a live biotherapeutic product by modeling methods in healthy subjects (blue box plots) and subjects with: (A) atopic dermatitis (orange box plots); (B) food sensitization (green box plots); (C) recurrent wheeze (red box plots); and (D) atopy plus wheeze (purple box plots) at one year of age. Taxa included in the final nomination (according to the Greengenes database) for each clinical measurement assessed at one year are indicated with an arrow.	S10
Figure E3. Odds ratios and 95% confidence intervals for children with a positive atopic composite score (having at least two of atopic dermatitis, food sensitization, recurrent wheeze, atopy+wheeze) at age one year of being diagnosed with food allergy, atopic dermatitis, allergic rhinitis, asthma, or two or more of these conditions (allergic composite score) at age five years.	S11
Figure E4. Risk of developing an allergic disease manifestation by age one year associated with an interquartile range (IQR) increase in alpha diversity (Shannon index) if the three-month gut microbiota stratified by potentially modifiable risk factors ((A) mode of delivery and (B) prenatal smoking exposure) and non-modifiable risk factors ((C) sex and (D) parental asthma status). RW: recurrent wheeze; ARW: atopy plus recurrent wheeze; AW: atopy plus wheeze; AD: atopic dermatitis.....	S11
Table E1. Derivation of sample sizes used in each analysis according to 16S data processing.	S13
Table E2. Nominated genera from the three-month-old for each clinical measurement assessed at age one year.	S13
Table E3. Nominated genera correspondence across two taxonomic classification method.....	S13
Table E4. Short-chain fatty acid-producing capacity of bacterial genera nominated for inclusion in a live biotherapeutic product designed for allergic disease prevention and references indicating a potentially therapeutic immunomodulatory effect of each genus.	S14
Supplemental References	S16

SUPPLEMENTARY METHODS

Ethics Approval

The human clinical research ethics boards at all universities and institutions directly involved with the CHILD cohort (McMaster University, University of British Columbia, the Hospital for Sick Children, University of Manitoba, and University of Alberta) approved this study and a parent or legal guardian gave signed informed consent to participate in the study.

Sample collection and sequencing data acquisition

DNA from fecal samples collected at three months of age from CHILD cohort participants was extracted using the MO BIO PowerMag Soil DNA Isolation Kit. The V4 hypervariable region of the 16S rRNA gene was amplified by PCR using universal bacterial primers (V4-515f: V4-806r) and pooled PCR amplicons were subjected to paired-end sequencing on the Illumina MiSeq platform.

CHILD cohort description and clinical definitions

CHILD cohort recruited a total of 3,621 pregnant women aged 18 years or older (19 or older in Vancouver) from four cities across Canada (see (1,2) for a detailed description). Infants born after 34 weeks of gestation with no congenital abnormalities were eligible and 3,455 began the study at birth. The current analysis was based on a subset of 659 children for whom fecal microbiota data at three months of age had been analyzed and 1, 3, and 5 year clinical outcomes data were available.

The CHILD cohort has prospectively followed children prenatally to age 5 years and collected a comprehensive array of clinical objective measurements and assessments, allowing the diagnosis of atopic diseases including asthma, allergic sensitization, wheeze, allergic rhinitis, and atopic dermatitis (or eczema). The health status of the CHILD study subjects was recorded using a combination of questionnaires and structured in-person clinical assessments at 11 time points over a period of five years after birth.

Children were diagnosed with IgE-mediated allergic sensitization (also referred to as atopy) based on skin prick testing to multiple common food and environmental inhalant allergens, using ≥ 2 mm average wheal size as indicating a positive test relative to the negative control. Allergens tested at the one year visit were German cockroach, *Alternaria alternata*, house dust mites (*Dermatophagoides pteronyssinus* and *Dermatophagoides farinae*), cat hair, dog epithelium, cow's milk, peanut, egg white and soybean. Glycerin and histamine served as the negative and positive controls, respectively. Additional allergens tested at three and five year visits were *Cladosporium*, tree mix, grass mix, *Aspergillus fumigatus*, weed mix, ragweed mix and penicillium mix. Food sensitization was defined at ages one, three, and five years after birth based on skin prick test measurements to the following foods: eggs, peanut, milk, and soy bean.

Atopic dermatitis was diagnosed at ages one, three, and five years after birth based on a clinical assessment by a CHILD healthcare professional using the UK Working Party criteria (3).

Recurrent wheeze was diagnosed at ages one, three, and five years after birth if the subject was reported to have two or more episodes of wheeze within the preceding year as indicated by the

response to the Child Health Questionnaire or clinical assessment. Asthma was diagnosed at ages 3 and 5 years based on a study subspecialist pediatrician response to the question: “In your opinion, does this child have asthma? (Yes / Possible / No)”. The pediatric asthma specialist or the highly trained health care professional working under their supervision conducted a structured interview with the accompanying parent or guardian identifying symptoms consistent with asthma, namely recurrent wheeze and coughing without a cold, and noted any physical findings. Asthma was considered definite if the parent reported physician-diagnosed asthma, or use of a bronchodilator prescribed by a physician for episodes of coughing or wheezing, or use of a prescribed daily controller medication, or frequent wheezing (3 or more distinct episodes over the previous year) with no alternative diagnosis. Atopy was not essential for this diagnosis but, together with parental history, weighted the likelihood of diagnosing recurrent wheezing as definite asthma. Possible asthma was recorded if there were less frequent episodes of wheeze or coughing without colds, and no report of medication use. Only definite cases of asthma (“Yes” to the initial interview question) were considered as cases for the purposes of the present study.

Healthy subjects were defined as CHILD study subjects who did not have asthma, recurrent wheeze, atopic dermatitis, atopy, or allergic sensitization (see table E1 for full description of sample sizes).

Sequence processing and statistical analysis

Demultiplexed fastq-files were processed in Qiime2 (4) to generate sub-Operational Taxonomic Units (sOTUs) as recommended by the deblur workflow (5) (<https://github.com/biocore/deblur>). The taxonomic assignment of the sOTUs was performed using the Greengenes database (6)

(version 13_8) (http://greengenes.secondgenome.com/downloads/database/13_8). All downstream microbiota data analyses were performed in the statistical software package R version 3.3.0, using packages including phyloseq (7) (<https://github.com/joey711/phyloseq>), randomForest (8) (<https://www.stat.berkeley.edu/~breiman/RandomForests/>), and mixOmicsCaret (9,10) (<https://github.com/jonathanh/mixOmicsCaret>). The overall data analysis workflow is described in Figure E1 and was designed to respect the compositional nature of amplicon sequencing data. Odds ratios (ORs) were calculated in R using logistic regression.

The Data QC step defines the healthy cohort for all clinical measurements, removes rare sOTUs and low-quality reads. The microbiota data were examined at aggregated genus level given the taxonomic resolution afforded by 16S sequencing. At this order, the sequencing resulted in a mean of 11215 reads per sample ranging from 2 to 49915. In order to correct for sequencing variation across the samples and sequencing batch effects, the microbiota data were rarefied (5000 reads) and percentile normalized, respectively. Additionally, for some methods which can incorporate sequencing depth and batch effects as parameters in the modeling, raw microbiota data was used along with sequencing batch and depth information as additional parameters.

Three distinct differential abundance methods were used to nominate taxa that were significantly and differentially abundant between individuals with a specific clinical phenotype and healthy individuals. Differences in the relative abundances at the genus level were analyzed using a number of differential abundance tests present in the R package DAtest (11) (<https://github.com/Russel88/DAtest>) including t-test, Wilcoxon rank-sum test, linear regressions, limma, SAMseq, edgeR, DESeq2, metagenomeSeq, and ANCOM. The DAtest

package allows comparison of different methods for differential abundance testing in order to select methods most appropriate for the specific microbiota dataset and assumes that most features are not associated with the outcome of interest. The importance of differentially abundant taxa (at genus level) for differentiating individuals with and without allergic disease was also determined using the sparse Partial Least Square – Discriminant Analysis (sPLS-DA) (10) and random forest classification (8). After log ratio transformation and total sum scaling of microbiota sequencing data to account for the compositionality of the dataset, sPLS-DA methods select the most predictive features in a dataset that help classify samples according to the outcome of interest. Random forest classification, in turn, is an accurate non-linear model for determining the most important features for predicting sample classification by training an outcome classification algorithm based on microbiota composition using a subset of samples and testing the accuracy of the model on the remaining samples. An optimized number of input variables were selected using repeated 10-fold cross-validation of the Area Under the Curve (AUC) statistic to avoid overfitting.

Subsequently nominated taxa for each condition were further filtered for a prevalence greater than 15% in the healthy children and relative abundances greater than 0.2% (Figure E2). A statistical significance level of 0.05 and an AUC cut-off of greater than 0.6 were used in all analyses.

Nomination for clinical measurements

Using the above-mentioned methods, the microbiota data collected at three months of age was used to nominate taxa that were associated with protection against clinical measures/phenotypes

at age one year. The clinical measurements for which the method was applied were: atopic dermatitis, food sensitization, recurrent wheeze, and atopy plus wheeze. The nominated genera for each clinical measure are listed in Table E2. As anticipated, there was a significant overlap in the nominated genera for the individual clinical measurements. These genera were combined into a single nomination for a composite atopic disease which is defined as having two or more of the clinical measurements: atopic dermatitis, food sensitization, recurrent wheeze, and atopy plus wheeze. Taxa included in the final nomination were further refined by including only those present at greater relative abundance in controls relative to cases for at least three of the four clinical phenotypes assessed, regardless of whether this difference was statistically significant. No bacterial genera were identified to be consistently increased in cases relative to controls.

The sequences of the nominated genera were also mapped to the Genomic Taxonomy Database (version 86) (<http://gtdb.ecogenomic.org/>) (12), and their corresponding genera are listed in Table E3.

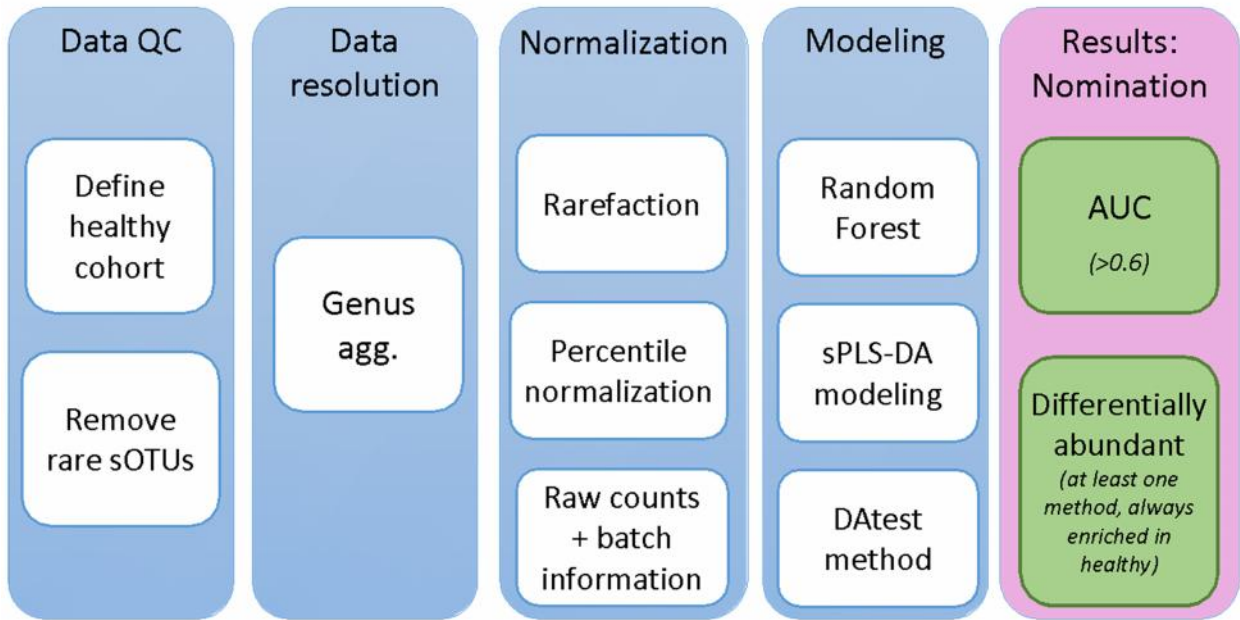


Figure E1. An overview of the methods used to analyze amplicon-based microbiota data to nominate taxa enriched in healthy individuals versus individuals with an atopic disease manifestation at age one year.

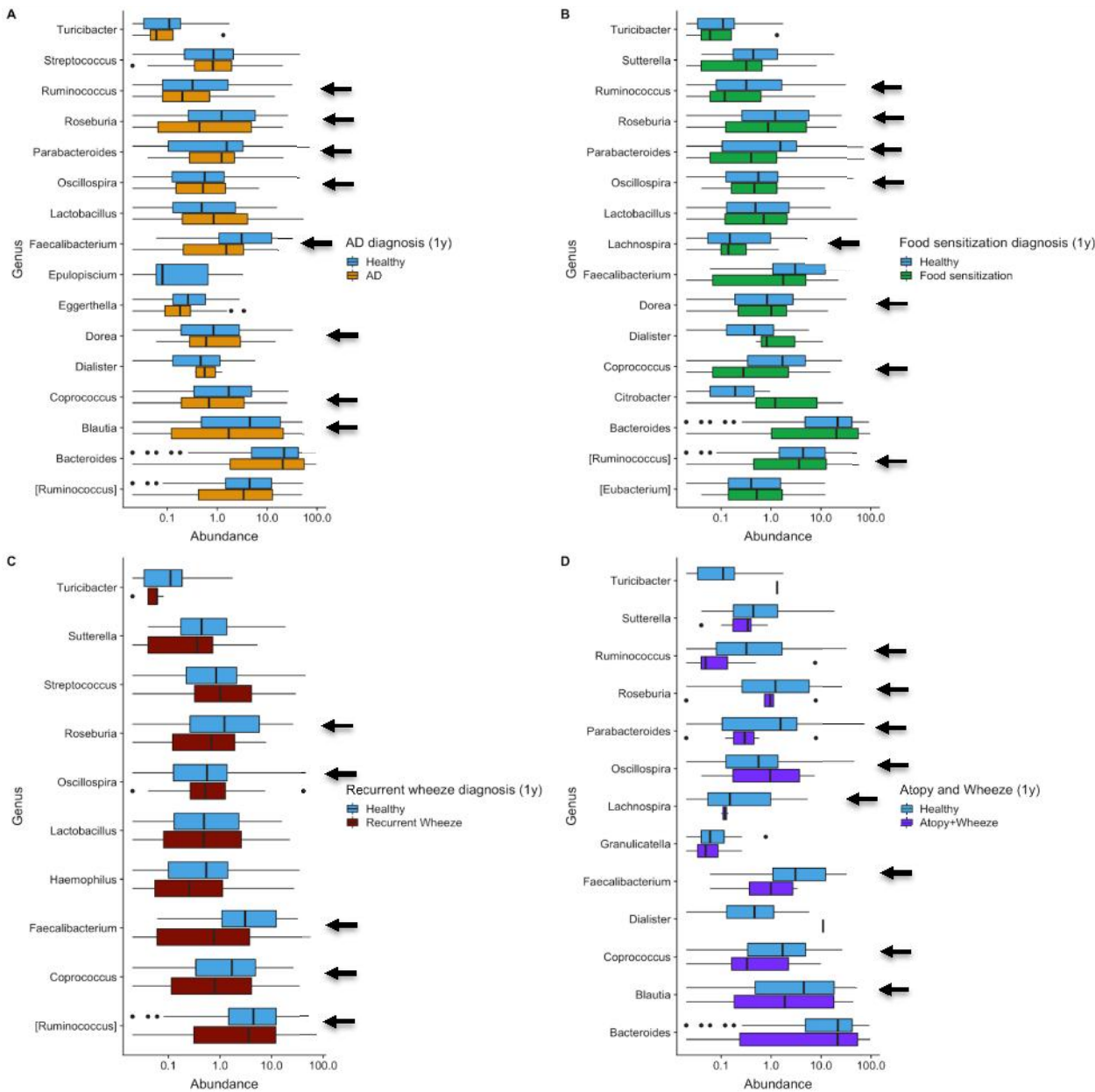


Figure E2. Relative abundance of genera within the three-month-old gut microbiota reported as potential nomination for inclusion in a live biotherapeutic product by modeling methods in healthy subjects (blue box plots) and subjects with: (A) atopic dermatitis (orange box plots); (B) food sensitization (green box plots); (C) recurrent wheeze (red box plots); and (D) atopy plus wheeze (purple box plots) at one year of age. Taxa included in the final nomination (according to the Greengenes database) for each clinical measurement assessed at one year are indicated with an arrow.

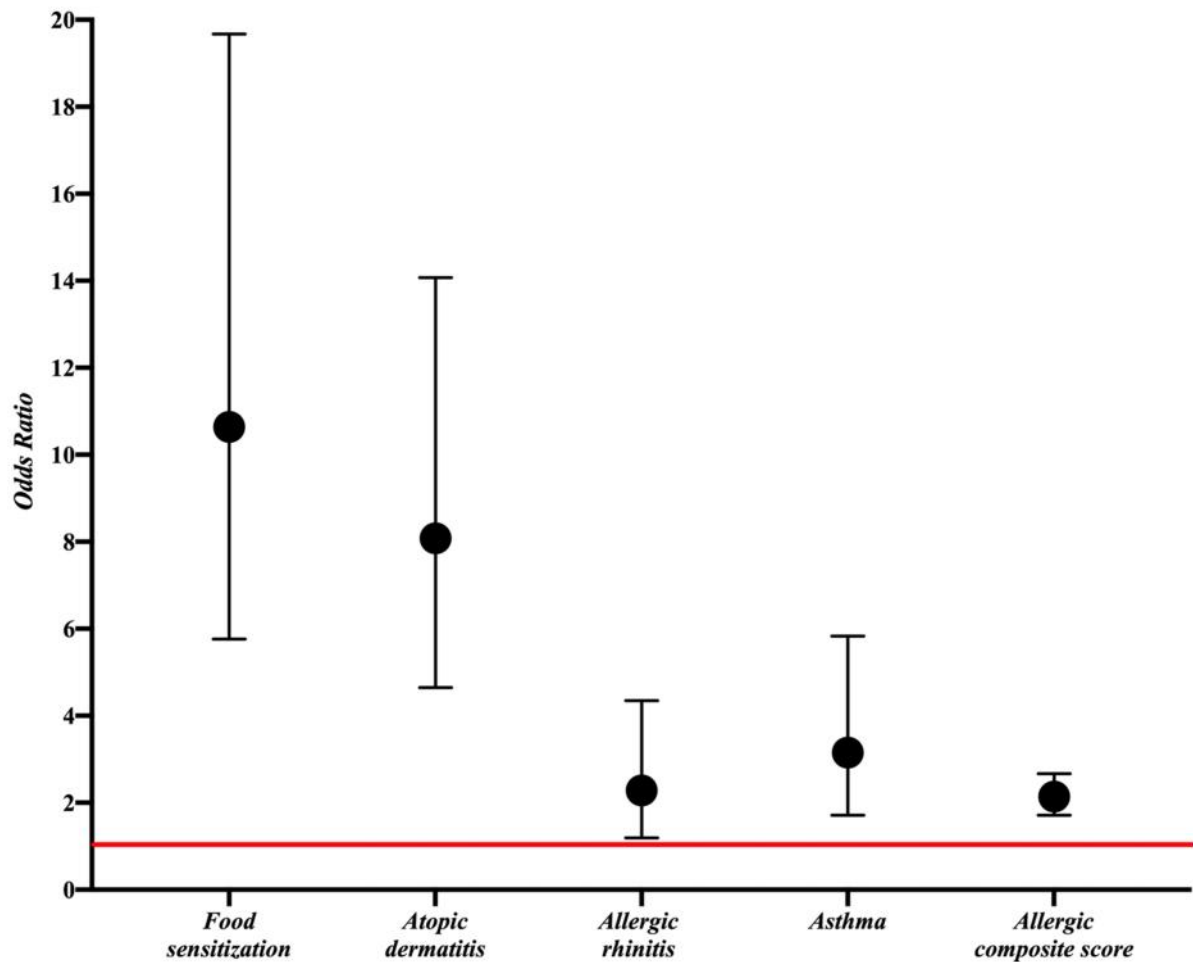


Figure E3. Odds ratios and 95% confidence intervals for children with a positive atopic composite score (having at least two atopic disease manifestations (atopic dermatitis, food sensitization, recurrent wheeze, atopy+wheeze)) at age one year of being diagnosed with food sensitization, atopic dermatitis, allergic rhinitis, asthma, or two or more of these conditions (allergic composite score) at age five years.

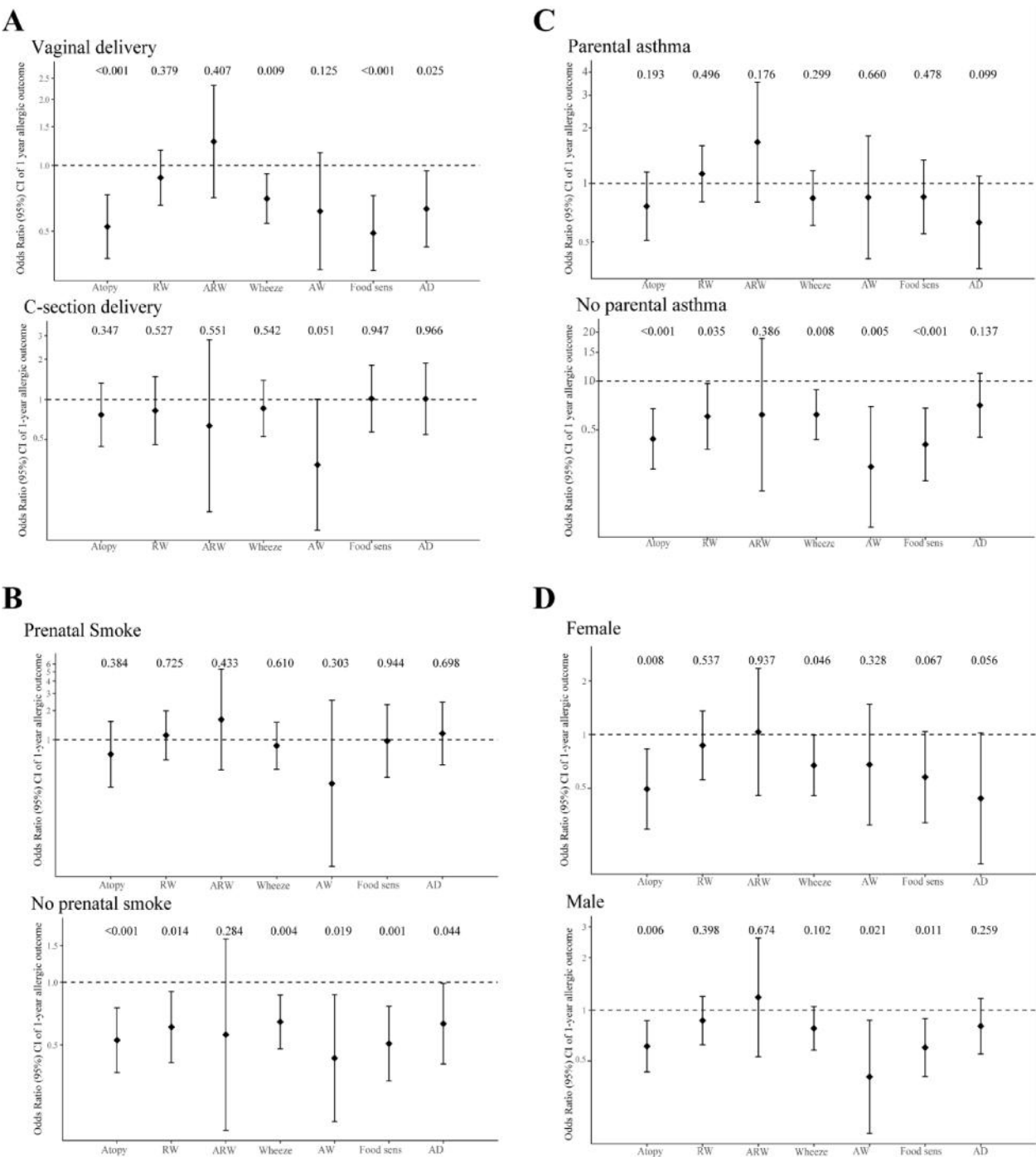


Figure E4. Risk of developing an allergic disease manifestation by age one year associated with an interquartile range (IQR) increase in alpha diversity (Shannon index) of the three-month gut microbiota stratified by potentially modifiable risk factors ((A) mode of delivery and (B) prenatal smoking exposure) and non-modifiable risk factors ((C) sex and (D) parental asthma status). RW: recurrent wheeze; ARW: atopy plus recurrent wheeze; AW: atopy plus wheeze; AD: atopic dermatitis.

Table E1. Derivation of sample sizes used in each analysis according to 16S data processing.

Description	Total sample size	Phenotype sample sizes
Total cohort with 16S data available	837	Healthy controls: 140 At least one allergic disorder: 316
Cohort with 16S data available after data processing and rarefaction	659	Healthy controls: 112 At least one allergic disorder: 244

Table E2. Nominated genera from the three-month-old for each clinical measurement assessed at age one year.

Atopic Dermatitis	Food Sensitization	Recurrent Wheeze	Atopy+Wheeze
<i>Blautia</i>	<i>Blautia</i>		<i>Blautia</i>
<i>Coprococcus</i>	<i>Coprococcus</i>	<i>Coprococcus</i>	<i>Coprococcus</i>
<i>Dorea</i>	<i>Dorea</i>		
<i>Faecalibacterium</i>	<i>Faecalibacterium</i>	<i>Faecalibacterium</i>	<i>Faecalibacterium</i>
<i>Oscillospira</i>	<i>Oscillospira</i>	<i>Oscillospira</i>	
	<i>Lachnospira</i>	<i>Lachnospira</i>	<i>Lachnospira</i>
<i>Parabacteroides</i>			<i>Parabacteroides</i>
<i>Roseburia</i>	<i>Roseburia</i>	<i>Roseburia</i>	<i>Roseburia</i>
<i>Ruminococcus</i>	<i>Ruminococcus</i>		<i>Ruminococcus</i>

Table E3. Nominated genera correspondence across two taxonomic classification methods.

Genus (Greengenes database v13_8)	Genus (GTDB database v86)
<i>Faecalibacterium</i>	<i>Faecalibacterium</i>
<i>Lachnospira</i>	<i>Lachnospira</i>
<i>Coprococcus</i>	<i>Anaerostipes</i>
<i>Coprococcus</i>	<i>Eubacterium_E</i>
<i>Oscillospira</i>	<i>Intestinimonas</i>
<i>Oscillospira</i>	<i>Flavonifractor</i>
<i>Roseburia</i>	<i>Roseburia</i>
[<i>Ruminococcus</i>]	<i>Ruminococcus_B</i>
<i>Ruminococcus</i>	<i>Ruminococcus_E</i>
<i>Dorea</i>	<i>Tyzzereella</i>
<i>Blautia</i>	<i>Blautia</i>
<i>Blautia</i>	<i>Blautia_A</i>
<i>Parabacteroides</i>	<i>Parabacteroides</i>

Table E4. Short-chain fatty acid-producing capacity of bacterial general nominated for inclusion in a live biotherapeutic product designed for allergic disease prevention (13–50).

Nominated genera	Short-chain fatty acids produced
<i>Faecalibacterium</i>	Butyrate
<i>Lachnospira</i>	Acetate
<i>Oscillospira</i>	Butyrate
<i>Roseburia</i>	Butyrate, Propionate
<i>Blautia</i>	Acetate, Butyrate

Acknowledgments: We are grateful to all the families who took part in the CHILD Study, and the entire CHILD team, which includes interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists.

Author contributions: SET, BBF, PS, RCTB, HS, and MD conceptualized the study. MS, RM, HS, and RCTB designed the study. MD, RM, DD, CP, RCTB, HS formalized the pipelines and analyzed the data. RCTB, HS, and MD interpreted the data and wrote the first draft of the manuscript. ABB, PS, PJM, SET, MBA, TJM and MRS directed data collection for the CHILD cohort. All authors contributed to critical revision of the manuscript and approved the final version.

Data and materials availability: The microbiota and clinical data related to this study (which have been stripped of any personally identifiable information) will be deposited in CHILDDb: a database for the CHILD cohort study (www.childdb.ca).

Supplemental References

1. Subbarao P, Anand SS, Becker AB, Befus AD, Brauer M, Brook JR et al. The Canadian Healthy Infant Longitudinal Development (CHILD) study: Examining developmental origins of allergy and asthma. *Thorax* 2015;**70**:998–1000.
2. Moraes TJ, Lefebvre DL, Chooniedass R, Becker AB, Brook JR, Denburg J et al. The canadian healthy infant longitudinal development birth cohort study: Biological samples and biobanking. *Paediatr Perinat Epidemiol* 2015;**29**:84–92.
3. Dharma C, Lefebvre DL, Tran MM, Lu Z, Lou WYW, Subbarao P et al. Diagnosing atopic dermatitis in infancy: Questionnaire reports vs criteria-based assessment. *Paediatr Perinat Epidemiol* 2018;**32**:556–567.
4. Bolyen E, Dillon M, Bokulich N, Abnet C, Al-Ghalith G, Alexander H et al. Reproducible, interactive, scalable, and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–857.
5. Amir A, McDonald D, Navas-Molina JA, C Evguenia Kopylova A, James T. Morton, a Zhenjiang Zech Xu, a Eric P. Kightley B, Luke R. Thompson A et al. Deblur Rapidly Resolves Single-. *Am Soc Microbiol* 2017;**2**:e00191-16.
6. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–5072.
7. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013;**8**.
8. Breiman L. Random Forests. *Mach Learn* 2001;**45**:5–32.
9. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;**13**:1–14.
10. Chung D, Keles S. Sparse partial least squares classification for high dimensional data. *Stat Appl Genet Mol Biol* 2010;**9**. doi:10.2202/1544-6115.1492
11. Russel J, Thorsen J, Brejnrod AD, Bisgaard H, Sørensen SJ, Burmølle M. DAtest: a framework for choosing differential abundance or expression method. *bioRxiv* 2018;;1–7.
12. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;**36**:996–1004.
13. Belkaid Y, Bouladoux N, Hand TW. Effector and memory T cell responses to commensal bacteria. *Trends Immunol* 2013;**34**:299–306.
14. Planer JD, Peng Y, Kau AL, Blanton L V., Ndao IM, Tarr PI et al. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* 2016;**534**:263–266.
15. Singh N, Gurav A, Sivaprakasam S, Brady E, Padia R, Shi H et al. Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* 2014;**40**:128–139.

- 294 16. Chang P V., Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate
295 regulates intestinal macrophage function via histone deacetylase inhibition. *Proc Natl*
296 *Acad Sci* 2014;**111**:2247–2252.
- 297 17. Wang HB, Wang PY, Wang X, Wan YL, Liu YC. Butyrate enhances intestinal epithelial
298 barrier function via up-regulation of tight junction protein claudin-1 transcription. *Dig Dis*
299 *Sci* 2012;**57**:3126–3135.
- 300 18. Arpaia N, Campbell C, Fan X, Dikiy S, van der Veecken J, DeRoos P et al. Metabolites
301 produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*
302 2013;**504**:451–455.
- 303 19. Thio CLP, Chi PY, Lai ACY, Chang YJ. Regulation of type 2 innate lymphoid cell–
304 dependent airway hyperreactivity by butyrate. *J Allergy Clin Immunol* 2018;**142**:1867–
305 1883.
- 306 20. Wu W, Sun M, Chen F, Cao AT, Liu H, Zhao Y et al. Microbiota metabolite short-chain
307 fatty acid acetate promotes intestinal IgA response to microbiota which is mediated by
308 GPR43. *Mucosal Immunol* 2017;**10**:946–956.
- 309 21. Zhang M, Qiu X, Zhang H, Yang X, Hong N, Yang Y et al. Faecalibacterium prausnitzii
310 inhibits interleukin-17 to ameliorate colorectal colitis in rats. *PLoS One* 2014;**9**:6–15.
- 311 22. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humará n LG, Gratadoux J-J et
312 al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified
313 by gut microbiota analysis of Crohn disease patients. *PNAS* 2008;**105**:16731–16736.
- 314 23. Quévrain E, Maubert MA, Michon C, Chain F, Marquant R, Tailhades J et al.
315 Identification of an anti-inflammatory protein from Faecalibacterium prausnitzii, a
316 commensal bacterium deficient in Crohn’s disease. *Gut* 2016;**65**:415–425.
- 317 24. Uematsu S, Fujimoto K, Jang MH, Yang BG, Jung YJ, Nishiyama M et al. Regulation of
318 humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like
319 receptor 5. *Nat Immunol* 2008;**9**:769–776.
- 320 25. Rescigno M, Rotta G, Valzasina B, Ricciardi-Castagnoli P. Dendritic cells shuttle
321 microbes across gut epithelial monolayers. *Immunobiology* 2001;**204**:572–581.
- 322 26. Del Chierico F, Nobili V, Vernocchi P, Russo A, Stefanis C De, Gnani D et al. Gut
323 microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients
324 unveiled by an integrated meta-omics-based approach. *Hepatology* 2017;**65**:451–464.
- 325 27. Kaakoush NO, Day AS, Huinao KD, Leach ST, Lemberg DA, Dowd SE et al. Microbial
326 dysbiosis in pediatric patients with Crohn’s disease. *J Clin Microbiol* 2012;**50**:3258–3266.
- 327 28. Haro C, Montes-Borrego M, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gamez-Delgado F,
328 Pérez-Martínez P et al. Two healthy diets modulate gut microbial community improving
329 insulin sensitivity in a human obese population. *J Clin Endocrinol Metab* 2016;**101**:233–
330 242.
- 331 29. Tims S, Derom C, Jonkers DM, Vlietinck R, Saris WH, Kleerebezem M et al. Microbiota
332 conservation and BMI signatures in adult monozygotic twins. *ISME J* 2013;**7**:707–717.
- 333 30. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R et al. Human

- genetics shape the gut microbiome. *Cell* 2014;**159**:789–799.
31. Escobar JS, Klotz B, Valdes BE, Agudelo GM. The gut microbiota of Colombians differs from that of Americans, Europeans and Asians. *BMC Microbiol* 2015;**14**:1–14.
 32. Machiels K, Joossens M, Sabino J, De Preter V, Arijis I, Eeckhaut V et al. A decrease of the butyrate-producing species *roseburia hominis* and *faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 2014;**63**:1275–1283.
 33. Patterson AM, Mulder IE, Travis AJ, Lan A, Cerf-Bensussan N, Gaboriau-Routhiau V et al. Human gut symbiont *Roseburia hominis* promotes and regulates innate immunity. *Front Immunol* 2017;**8**:1–14.
 34. Van Den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, Thas O et al. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J* 2013;**7**:949–961.
 35. Takahashi K, Nishida A, Fujimoto T, Fujii M, Shioya M, Imaeda H et al. Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion* 2016;**93**:59–65.
 36. Zohreh T-S, Imen S, Latifa B, Olivier L, Vincent M, Shao Bing F et al. *Roseburia* spp.: A marker of health? *Future Microbiol* 2017;**12**:157–170.
 37. Jenq RR, Taur Y, Devlin SM, Ponce DM, Goldberg JD, Ahr KF et al. Intestinal *Blautia* Is Associated with Reduced Death from Graft-versus-Host Disease. *Biol Blood Marrow Transplant* 2015;**21**:1373–1383.
 38. Paul B, Royston KJ, Li Y, Stoll ML, Skibola CF, Wilson LS et al. Impact of genistein on the gut microbiome of humanized mice and its role in breast tumor inhibition. *PLoS One* 2017;**12**:1–20.
 39. Hoen AG, Li J, Moulton LA, O'Toole GA, Housman ML, Koestler DC et al. Associations between gut microbial colonization in early life and respiratory outcomes in cystic fibrosis. *J Pediatr* 2015;**167**:138–147.
 40. Nycz BT, Dominguez SR, Friedman D, Hilden JM, Ir D, Robertson CE et al. Evaluation of bloodstream infections, *clostridium difficile* infections, and gut microbiota in pediatric oncology patients. *PLoS One* 2018;**13**:1–16.
 41. Chen SY, Tsai CN, Lee YS, Lin CY, Huang KY, Chao HC et al. Corrigendum: Intestinal microbiome in children with severe and complicated acute viral gastroenteritis. *Sci Rep* 2017;**7**:46820.
 42. Kverka M, Zakostelska Z, Klimesova K, Sokol D, Hudcovic T, Hrnčir T et al. Oral administration of *Parabacteroides distasonis* antigens attenuates experimental murine colitis through modulation of immunity and microbiota composition. *Clin Exp Immunol* 2011;**163**:250–259.
 43. Plummer EL, Lozinsky AC, Tobin JM, Uebergang JB, Axelrad C, Garland SM et al. Postnatal probiotics and allergic disease in very preterm infants: sub-study to the ProPrems randomized trial. *Allergy* 2019;*In press*.
 44. Cait A, Hughes MR, Antignano F, Cait J, Dimitriu PA, Maas KR et al. Microbiome-

- driven allergic lung inflammation is ameliorated by short-chain fatty acids. *Mucosal Immunol* 2018;**11**:785–795.
45. Cait A, Cardenas E, Dimitriu PA, Amenyogbe N, Dai D, Cait J et al. Reduced genetic potential for butyrate fermentation in the gut microbiome of infants who develop allergic sensitization. *J Allergy Clin Immunol* 2019;**144**:1638-1647.
46. Arrieta MC, Arévalo A, Stiemsma L, Dimitriu P, Chico ME, Loo S et al. Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. *J Allergy Clin Immunol* 2018;**142**:424-434.
47. Trompette A, Gollwitzer ES, Yadava K, Sichelstiel AK, Sprenger N, Ngom-Bru C et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med* 2014;**20**:159–166.
48. Thorburn AN, McKenzie CL, Shen S, Stanley D, Macia L, Mason LJ et al. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat Commun* 2015;**6**:7320.
49. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* 2013;**341**:569–573.
50. Roduit C, Frei R, Ferstl R, Loeliger S, Westermann P, Rhyner C et al. High levels of butyrate and propionate in early life are associated with protection against atopy. *Allergy Eur J Allergy Clin Immunol* 2019;**74**:799–809.

Relevant Patent applications: US9603876B2, US20180264053A1, US9750902B2, US20180050070A1, US20180333440, US20190134107, US10391131, US20180117097, US20190142877