

## **Supplemental Methods**

### **Clinical methods and selection criteria**

Control samples and subjects for the cross-sectional cohort were selected to minimize population level differences by matched group mean and standard deviation in age at the time of sampling, gestational age at birth, and mode of delivery. Samples for the longitudinal cohort were collected from the RSV group at approximately one month of age, during acute RSV infection, and approximately one month after illness, and at corresponding timepoints from the healthy controls. Control subjects were selected to match on an individual basis by sex, mode of delivery, and gestational age at birth, and samples were selected to match by age. For both cohorts, subjects were eligible as controls if they had no respiratory illness between birth and at least ten days after the last sample and had not been exposed to antibiotics within 30 days of enrollment. Patients and controls were not exposed to antibiotics within 30 days of the time points studied.

### **Genomic DNA extraction**

Total genomic DNA was extracted from the nasal samples using a modification of the ZymoBIOMICS™ DNA Miniprep Kit (Zymo Research, Irvine, CA) and FastPrep mechanical lysis (MPBio, Solon, OH). 16S ribosomal DNA (rRNA) was amplified with Phusion High-Fidelity polymerase (Thermo Scientific, Waltham, MA) and dual indexed primers specific to the V1-V3 (8F: 5' AGAGTTTGATCCTGGCTCAG 3'; 534R: 3' ATTACCGCGGCTGCTGG 5') hypervariable regions [1]. Amplicons were pooled and paired-end sequenced on an Illumina MiSeq (Illumina, San Diego, CA) in the University of Rochester Genomics Research Center. Each sequencing run included: (1) positive controls consisting of a 1:5 mixture of *Staphylococcus aureus*, *Lactococcus lactis*, *Porphyromonas gingivalis*, *Streptococcus mutans*, and *Escherichia coli*; and (2) negative controls consisting of sterile saline.

### **Microbiota background control**

The background microbiota was monitored at multiple stages of sample collection and processing. All sterile saline, buffers, reagents, plasticware and flocked nylon swabs used for sample collection, extraction and amplification of nucleic acid were UV irradiated to eliminate possible DNA background contamination. Elimination of potential background from the irradiated buffers, reagents, plasticware and swabs was confirmed by 16S rRNA amplification. After sample collection, multiple aliquots of sterile saline with swabs used for sample collection were carried through our entire sequencing protocol as individual samples, including DNA extraction, 16S rRNA amplification, library construction and sequencing to monitor potential background microbiome [2]. Data from these background control samples is deposited in SRA along with positive controls.

### **Bioinformatics analysis**

Raw data from the Illumina MiSeq was first converted into FASTQ format 2x312 paired end sequence files using the bcl2fastq program, version 1.8.4, provided by Illumina. Format conversion was performed without de-multiplexing and the EAMMS algorithm was disabled. All other settings were default. Reads were multiplexed using a configuration described previously [1]. Briefly, for both reads in a pair, the first 12 bases were a barcode, which was followed by a primer, then a heterogeneity spacer, and then the target 16S rRNA sequence. QIIME 1.9.1 [3] was used to extract the barcodes into a separate file for importing into QIIME 2 [4], which was used to perform all subsequent processing. Reads were demultiplexed requiring exact barcode matches, and 16S primers were removed allowing 20% mismatches and requiring at least 18 bases. Cleaning, joining, and denoising were performed using DADA2 [5]: forward reads were truncated to 275 bps and reverse reads to 260 bps, error profiles were learned with a sample of one million reads, and a maximum expected error of two was allowed. Taxonomic classification was performed with a custom naïve Bayesian classifier trained on the August, 2013 release of

GreenGenes [6, 7]. Sequence variants that could not be classified at least at the phylum level were discarded. Sequencing variants observed fewer than ten times total, or in only one sample, were discarded. Samples with fewer than 900 reads were discarded.

Phylogenetic trees were constructed for each cohort using MAFFT for sequence alignment and FastTree for tree construction [8, 9]. Prior to diversity analyses, samples were rarefied to a depth of 900 reads. Faith's PD and the Shannon index were used to measure alpha diversity, and Kruskal-Wallis to test for differences. Weighted and Unweighted Unifrac distances were used to measure beta diversity [10] and pairwise PERMANOVA to test for differences.

Infected vs. healthy and healthy vs. mild vs. severe classification, and severity score regression, were performed using the Sample Classifier plugin [11] in QIIME 2, using the Gradient Tree Boosting Classifier/Regressor, five-fold cross-validation, 20% data hold-out for testing, 5,000 estimators, parameter tuning, and feature selection. Both exact sequence variant abundances and abundances of taxa summarized at species level were tried as inputs, and whichever performed better was used and reported.

### **Associations of taxon abundance with RSV infection and disease severity**

Univariate tests for differential taxon abundance between groups was performed using both ANCOM [12] and LefSe [13]. Multivariate regression models using gneiss [14] and MaAsLin [15] were employed to assess associations of taxon abundance with RSV infection and disease severity while controlling for the potentially confounding covariates sex, mode of delivery, age at sampling, reads per sample, and antibiotic usage. The cross-sectional and longitudinal cohorts were analyzed independently. All reported results were significant by at least two tests.

## **References**

1. Fadrosh DW, Ma B, Gajer P, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2014**; 2:6.
2. Grier A, McDavid A, Wang B, et al. Neonatal gut and respiratory microbiota: coordinated development through time and space. *Microbiome* **2018**; 6:193.
3. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **2010**; 7:335-6.
4. Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **2019**; 37:852-7.
5. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* **2016**; 13:581-3.
6. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* **2006**; 72:5069-72.
7. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **2012**; 6:610-8.
8. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**; 30:772-80.
9. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **2010**; 5:e9490.
10. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *The ISME journal* **2011**; 5:169-72.
11. Bokulich NA, Dillon MR, Zhang Y, et al. q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. *mSystems* **2018**; 3.

12. Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* **2015**; 26:27663.
13. Segata N, Izard J, Waldron L, et al. Metagenomic biomarker discovery and explanation. *Genome biology* **2011**; 12:R60.
14. Morton JT, Sanders J, Quinn RA, et al. Balance Trees Reveal Microbial Niche Differentiation. *mSystems* **2017**; 2.
15. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology* **2012**; 13:R79.