nature microbiology

SUPPLEMENTARY INFORMATION

ARTICLE NUMBER: 16031 | DOI: 10.1038/NMICROBIOL.2016.31

Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype

Leopoldo N. Segal[1,2], Jose C. Clemente*[3,4], Jun-Chieh J. Tsay*[1,2], Sergei B Koralov*[5], Brian C. Keller*[6], Benjamin G. Wu*[1,2], Yonghua Li*[1,2], Nan Shen*[3], Elodie Ghedin*[7], Alison Morris*[8], Phillip Diaz*[6], Laurence Huang*[9], William R. Wikoff*[10], Carles Ubeda*[11], Alejandro Artacho*[11], William N Rom*[1,2], Daniel H. Sterman*[1,2], Ronald G. Collman*[12], Martin J. Blaser*[2], Michael D. Weiden*[*1,2]*

[1] Division of Pulmonary, Critical Care and Sleep Medicine, New York University School of Medicine, NY
[2] Department of Medicine, New York University School of Medicine, New York, NY
[3] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY.
[4] Immunology Institute, Icahn School of Medicine at Mount Sinai, NY
[5] Department of Pathology, New York University School of Medicine, New York, NY
[6] Division of Pulmonary and Critical Care Medicine, The Ohio State University, Columbus, OH
[7] Department of Biology, Center for Genomics & Systems Biology, & College of Global Public Health, New York University, New York, NY
[8] Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, PA
[9] Department of Medicine, University of California San Francisco, San Francisco, California, CA
[10] Department of Molecular and Cellular Biology & Genome Center, University of California, Davis, CA
[11] Center for Public Health Research, FISABIO, Valencia, Spain
[12] Department of Medicine and Microbiology, University of Pennsylvania School of Medicine, Philadelphia, PA.

Leopoldo N. Segal, MD, MS[1]	Leopoldo.Segal@nyumc.org
Jose C. Clemente, PhD	jose.clemente@mssm.edu
Jun-Chieh Tsay, MD	Jun-Chieh.Tsay@nyumc.org
Sergei B. Koralov, PhD	Sergei.Koralov@nyumc.org
Benjamin G. Wu, MD	Benjamin.Wu@nyumc.org
Yonghua Li, MD, PhD	Yonghua.Li@nyumc.org
Nan Shen, MSc	nan.shen@mssm.edu
Elodie Ghedin, PhD	elodie.ghedin@nyu.edu
Alison M. Morris, MD, MS	morrisa@upmc.edu
Phillip Diaz, MD	Philip.Diaz@osumc.edu
Laurence Huang, MD	Laurence.Huang@ucsf.edu

DOI: 10.1038/NMICROBIOL.2016.31

William R. Wikoff, PhD wrwikoff@ucdavis.edu
Carles Ubeda, PhD ubeda_carmor@gva.es
Alejandro Artacho, PhD artacho_ale@gva.es
Brian C. Keller, MD, PhD bkeller@path.wustl.edu
William N. Rom, MD, MPH William.Rom@nyumc.org
Daniel H. Sterman, MD Daniel.Sterman@nyumc.org
Ronald G. Collman, MD collmanr@mail.med.upenn.edu
Martin J. Blaser, MD Martin.Blaser@nyumc.org
Michael D. Weiden, MD, MPH [1] Michael.Weiden@nyumc.org

*Corresponding Author/ Address for Reprints:
Leopoldo N. Segal, MD, MS Leopoldo.Segal@nyumc.org
Michael D. Weiden, MD, MPH Michael.Weiden@nyumc.org

Research support funding: K23 AI102970 (LNS); R01 HL125816 (SBK); K24 AI080298 (MDW); CTSI Grant #UL1 TR000038; EDRN 5U01CA086137-13; Diane Belfer Program for Human Microbial Ecology; Michael Saperstein Scholarship Fund; R01DK090989; UH2 AR57506; U01AI111598; R01 HL090339; K24 HL123342 (AM); U01 HL098962 (AM, EG); K24HL123342; HL087713 (LH); NIAID and NCI UO1-AI-35042; 5-MO1-RR-00722 (GCRC); UL1TR000124 (UCLA CTRC); UO1-AI-35043; UO1-AI-37984; UO1-AI-35039; UO1-AI-35040; UO1-AI-37613; UO1-AI-35041 (Multicenter AIDS Cohort); NIAID and NICHHD UO1-AI-35004; UO1-AI-31834; UO1-AI-34994; UO1-AI-34989; UO1-AI-34993; UO1-AI-42590; UO1-HD-32632 (WIHS); U01-HL098957 and R01-HL113252 (RGC).

Authors Contributions
1) Conception and design: LNS, JCC, MJB, MDW.
2) Acquisition of data: LNS, JJT, AM, LH, PD, WW.
3) Analysis and interpretation of data: LNS, JCC, JJT, SBK, BGW, YL, NS, WRW, CU, AA, BCK, RGC, MJB, MDW.
4) Drafting or revising of article: LNS, JCC, JJT, SBK, EG, AM, PD, LH, WRW, BCK, WNR, DHS, RGC, MJB, MDW.
5) Final approval of the manuscript: LNS, JCC, JJT, SBK, BGW, YL, NS, EG, AM, PD, LH, WRW, CU, AA, BCK, WNR, DHS, RGC, MJB, MDW.

Financial Disclosure: None

Supplementary Methods

16S rRNA gene sequencing

Amplicon library preparation was performed using an automated platform (Biomek 4000) using a custom liquid handling method. Reagent controls were sequenced and analyzed as quality controls. For each sample, the V4 region of the bacterial 16S rRNA gene was amplified in duplicate reactions, using primer set 515F/806R, which nearly universally amplifies bacterial and archaeal 16S rRNA genes.[1,2] Each unique barcoded amplicon was generated in pairs of 25µl reactions with the following reaction conditions: 11µl PCR-grade H2O, 10µl Hot MasterMix (5 Prime Cat# 2200410), 2µl of forward and reversed barcoded primer (5µM) and 2µl template DNA. Reactions were run on a C1000 Touch Thermal Cycler (Bio-Rad) with the following cycling conditions: initial denaturing at 94°C for 3 min followed by 35 cycles of denaturation at 94°C for 45 seconds, annealing at 58°C for 1 minute, and extension at 72 C for 90 seconds, with a final extension of 10 min at 72°C. Amplicons were quantified using Agilent 2200 TapeStation system and pooled. Purification was then performed using Ampure XT (Beckman Coulter Cat# A63882) as per the manufacturer instructions. Sequencing was then performed in MiSeq (Illumina) to produce 150 base-paired end reads.

The obtained 16S rRNA gene sequences were analyzed using the QIIME package for analysis of community sequence data [3]. Reads were demultiplexed and quality filtered with default parameters. Sequences were then clustered into operational taxonomic units (OTUs) using a 97% similarity threshold with UCLUST [4] and the Greengenes 16S reference dataset and taxonomy[5]. After

DOI: 10.1038/NMICROBIOL.2016.31

curation and removal of sequences potentially derived from reagent controls, the absolute OTU sequence counts were normalized to obtain the relative abundances of the taxa within each sample. These relative abundances at 97% OTU similarity and each of the 5 higher taxonomic levels (phylum, class, order, family, genus) were tested for univariate associations with clinical variables. To decrease the number of features, we only focused on major taxa and OTUs, defined as those having mean relative abundance >1% in at least one sample.

Since the distributions of microbiome data are non-normal, and no distribution-specific tests are available, we used non-parametric tests of association. For association with discrete factors, we used either the Mann-Whitney test (in the case of 2 categories) or the Kruskal Wallis ANOVA (in case of > 2 categories). For tests of association with continuous variables, we used non-parametric Spearman correlation tests. False discovery rate (FDR) was used to control for multiple testing.[6] Weighted UniFrac was used to measure β diversity of bacterial communities and to perform principal coordinate analysis (PCoA).[7] We used the ade4 package in R to PCoA on weighted UniFrac distances.[8] To avoid negative eigenvalues in the analysis, we used the Cailliez method to convert the weighted UniFrac distance matrix into a closest corresponding matrix with Euclidean properties, which was further used for PCoA.[9]

Shotgun sequencing and phageome analysis

DNA was extracted from BAL samples as described for 16S sequencing above. Following fragmentation by ultrasonication (Covaris E210), nucleic acid was prepared into sequencing libraries using the NEBNext Ultra DNA Library Prep Kit

for Illumina (New England Biolabs). Equimolar concentrations of libraries were pooled (20 samples/run) and sequenced on the Illumina Miseq platform (Washington University Center for Genome Sciences, 2 x 250 bp paired-end reads, 6-8 pM loading concentration, 1% PhiX spike-in). Sequences were analyzed using VirusSeeker,[10] a custom bioinformatic pipeline that assigns viral taxonomy based on nucleotide and amino acid sequence homology to reference databases. Briefly, adapter sequences were trimmed with Cutadapt,[11] and overlapping Read1 and Read2 sequences were stitched into a single read using fastq-join (ea-utils package, http://code.google.com/p/ea-utils). Sequence quality control was performed with Prinseq[12] with average PHRED cut-off of 25 and minimum quality cut-off score of 10. Sequences were deduplicated at 95% identity over 95% of sequence length with CD-HIT,[13,14] and low complexity and repeat sequences were masked with Tantan and RepeatMasker (http://www.repeatmasker.org), respectively. Human sequences were removed by mapping against the reference human genome used by the 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/human_g1k_v37.fasta.gz). High-quality, deduplicated sequences were then sequentially queried against custom viral databases using BLASTn (E-value cutoff 1e-5) and BLASTx (E-value cutoff 1e-3).[15] Sequences with homology to bacteriophage were binned and used for downstream analyses. Bacteriophage sequences were converted into a tabular datatable in MEGAN v5.10.5.[16] Differential taxa were identified from

DOI: 10.1038/NMICROBIOL.2016.31

the datatable using Linear Discriminant Analysis (LDA) Effect Size (LEfSe)[17] with alpha value 0.01 and an "all-against-all" multi-class analysis strategy.

Measurement of metabolites in BAL fluid

Briefly, after entering the study design into the MiniX database[18] samples were aliquotted and maintained at −80°C until use, at which point samples were thawed, extracted and derivatized.[19] The sample was split into 15 µl aliquots which were each extracted with 1 ml of degassed acetonitrile:isopropanol:water (3:3:2) at −20°C, centrifuged and decanted with subsequent solvent evaporation to complete dryness. An acetonitrile/water (1:1) clean-up step removed lipids and the supernatant was again dried. Internal FAME standards (C8 through C30) were added and the sample was derivatized with methoxyamine hydrochloride in pyridine and subsequently by MSTFA (Sigma-Aldrich) for trimethylsilylation of acidic protons.

An automatic liner exchange system (ALEX, Gerstel MPS2) was used to eliminate between run cross-contamination from sample matrix. One microliter of sample was injected at 50°C (ramped to 250°C) in splitless mode with a 25 sec splitless time. An Agilent 6890 gas chromatograph (Santa Clara, CA) was used with a 30 m long, 0.25 mm i.d. Rtx5Sil-MS column with 0.25 µm 5% diphenyl film; an additional 10 m integrated guard column was used (Restek, Bellefonte PA).[20-22] Chromatography was performed at a constant flow of 1 ml/min, ramping the oven temperature from 50°C to 330°C over 22 min. Mass spectrometry used a Leco Pegasus IV time-of-flight mass (TOF) spectrometer with 280°C transfer line temperature, electron ionization at −70 V and an ion source temperature of 250°C. Mass spectra were acquired from *m/z* 85–500 at 20 spectra/sec and 1750 V detector voltage. Results[22] were exported to the netCDF format for further data

evaluation with BinBase [18] and filtered by multiple parameters to exclude noisy or inconsistent peaks. Quantification was reported as peak height using the unique ion.[23] All database entries in BinBase were matched against the Fiehn mass spectral library of 1,200 authentic metabolite spectra using retention index and mass spectrum information or the NIST11 commercial library. Identified metabolites were reported if present with at least 50% of the samples per study design group. Quantitative data were normalized to the intensity sum of all known metabolites and used for statistical investigation.

DOI: 10.1038/NMICROBIOL.2016.31

Supplementary Results

Demographics and clinical characteristics

To test the generalizability of our prior observation of two lower airway pneumotypes among healthy subjects, we used lower airways samples obtained from 49 subjects recruited from three separate cohorts. A significant proportion (63%) were smokers (current or former) (**Table 1**). Enrolled subjects had a relatively preserved FEV_1 (median [IQR] = 90 [67-93] % predicted), FVC (92 [71-104] % predicted), and FEV_1/VC (81 [69-85] % predicted). Similarly, for the most part, lung volumes were within normal range (TLC = 99 [69-116], FRC = 85 [60-121] and RV/TLC = 0.34 [0.25-0.37] % predicted). In this group, no major diffusion abnormalities were present (DLCO = 81 [77-100] % predicted).

Bacterial 16S rRNA gene sequencing

All samples yielded 16S rRNA gene V4 sequences with a median depth of sequencing of 12,385 reads per sample (IQR = 10840-16136). Unsupervised hierarchical analysis based on relative abundance of most abundant taxa (> 3% relative abundance at a genus level) in background samples and upper airway samples showed differential clustering for both sample types (**Supplemental Figure 1, Panel A**). We then applied Random Forest to identify taxa most discriminant between background and upper airway samples (**Supplemental Figure 1, Panel B**). Among the most discriminant taxa for upper airways, *Prevotella*, *Veillonella*, and *Fusobacterium* were the most abundant. In contrast, among the most discriminant taxa for background samples, *Pseudomonas* and *Acidocella* were the most abundant taxa.

To evaluate the relative contribution of the background microbiome to the BAL samples, we used SourceTracker.[24] This analysis showed that a large percentage (51[45-54]%) of the taxa found in BAL samples of pneumotype$_{BPT}$ were present in the saline or the bronchoscope. In contrast, the percentage of the background taxa found in BAL samples of pneumotype$_{SPT}$ was significantly smaller (34[23-42]%, p=0.0009). Analysis of α-diversity showed that there were no significant rarefaction differences between the two pneumotypes (**Supplemental Figure 2**).

Since in a subgroup of these samples (n=28) 16S rRNA gene sequences were previously obtained using a 454 platform and targeting V1-V2 variable region,[25] we compared the β diversity distribution of the BAL samples between these two sequencing approaches to evaluate consistency of the data. **Supplemental Figure 3 Panel A** shows that there are significant differences in β diversity between the two sequencing approach. Using LEfSe analysis, we identified multiple significant (LDA>2) taxonomic differences. Samples sequenced with a 454 platform and targeting V1-V2 were enriched with the genera *Corynebacterium*, *Paenibacillus*, *Streptococcus*, *Staphylococcus*, *Propionibacterium*, and *Lactococcus* among others (**Supplemental Figure 3 Panel B**). Samples sequenced with the MiSeq platform and targeting V4 were enriched with the genera *Acidocella*, *Sphingomonas*, *Chryseobacterium*, *Flavobacterium*, *Acinetobacter*, and *Janthinobacterium* among others. To examine whether these samples were still assign to similar "microbiome clusters" (or pneumotype) despite the above-mentioned differences in these two different

DOI: 10.1038/NMICROBIOL.2016.31

approaches we utilized a Procrustes analysis. This analysis compared the fitness of the compositional characteristics of the microbiome in these samples (based on β diversity distribution) when processed by two different platforms (454 and MiSeq) and two different primers (V1-V2 and V4). This analysis showed that there was a significant correlation between the microbial compositions of paired samples sequenced in both platforms, where each sample was consistently assigned to same pneumotype, regardless of the sequence approach used (**Supplemental Figure 3 Panel C**).

To evaluate for differences between BAL samples classified as pneumotype$_{SPT}$ and upper airway samples, we performed LEfSe analysis (**Supplemental Figure 4 Panel A**). Data showed that BAL samples classified as pneumotype$_{SPT}$ were enriched with *Prevotella*, *Selenomonas*, *Veillonella* and *Campylobacter* while upper airway samples were enriched with *Acinetobacter*, *Variovorax* and *Sphingomonas*. We then utilized similar approach to examine for taxonomic differences between BAL samples classified as pneumotype$_{BPT}$ and background samples (**Supplemental Figure 4 Panel B**). Data showed that BAL samples classified as pneumotype$_{BPT}$ were enriched with *Sphingomonas*, *Staphylococcus*, *Acinetobacter* and *Janthinobacterium* while background samples were enriched with *Ralstonia*, *Corynebacterium*, and *Shewanella*.

Metagenomic differences between the pneumotypes

To assess the functional metabolic capability of distinct microbial communities we inferred the metagenome using PICRUSt.[26] We estimated the Nearest

Sequenced Taxon Index (NSTI), a proxy for PICRUSt predictive accuracy.[27] In our cohort the NSTI was 0.10 ± 0.03, comparable to values previously reported and indicating an overall accurate prediction.[28] **Figure 2 Panel A** shows a significant difference (Adonis $p<0.001$) in coding potential between $pneumotype_{BPT}$ and $pneumotype_{SPT}$ using a Jensen–Shannon divergence analysis. Since annotation of KEEG ortholog is not limited to functions of bacterial origin, we have limited our analysis to metabolic pathways only in order to remove potential artifacts. **Figure 2 Panel B** shows that when protein-coding genes are assigned to specific metabolic function, $pneumotype_{SPT}$ is enriched with genes allocated to energy metabolism, enzyme families, glycan biosynthesis and metabolism, and metabolism of cofactors and vitamins. In contrast, $pneumotype_{BPT}$ is enriched for genes allocated to lipid metabolism, metabolism of amino acids, metabolism of terpenoids and polyketides, xenobiotics biodegradation and metabolism, and biosynthesis of other secondary metabolites. STAMP was utilized to determine the significance of different metabolic pathways between the pneumotypes. Of the 4,688 KEGG Ontology (KO) assignments inferred in BAL samples, genes with an annotation related to microbial metabolism were summarized into a total of 154 metabolic processes. Among them, 114 metabolic pathways were found differentially represented between $pneumotype_{SPT}$ and $pneumotype_{BPT}$ ($p <0.05$, FDR corrected). Sixty-five metabolic pathways were enriched in $pneumotype_{SPT}$ while 94 were enriched in $pneumotype_{BPT}$ (**Supplemental Table 1**). **Figure 2 Panel C** shows effect size (eta-squared) and relative abundance of the most significant pathways. Among

DOI: 10.1038/NMICROBIOL.2016.31

them, pneumotype$_{SPT}$ was enriched in metabolism pathways for DNA synthesis such as purines, pyrimidines, amino sugars, and nucleotide sugars and one carbon metabolism. Pneumotype$_{SPT}$ was also enriched for metabolism of fructose, mannose, and galactose. Alternately, pneumotype$_{BPT}$ was enriched in metabolism pathways for long and short chain fatty acids, valine/leucine/isoleucine, and phenylalanine.

Metabolomic differences between two pneumotypes

To examine if the differences in the genomic composition of metabolic pathways of these two pneumotypes support a different metabolic environment in the lower airways, metabolites in BAL fluid were assayed for 29 subjects from the NYU cohort by GC-TOF mass spectrometry, yielding 122 identified metabolites. Of these, 83 metabolites had a KEGG annotation and were then correlated with the taxonomic differences identified in pneumotype$_{SPT}$ and pneumotype$_{BPT}$. In BAL samples with pneumotype$_{BPT}$, the intensity of glyceric acid, glycerol alpha phosphate, cellobiose, and isothreonic acid were significantly higher than in BAL samples with pneumotype$_{SPT}$ (**Supplemental Table 2**). **Supplemental Table 3** shows that BAL UniFrac distance to upper airway correlated with bacterially-synthesized carbohydrates fucose-rhamnose and a sugar acid derived from threose and arachidonic acid.

To further evaluate the microbiome-metabolome interaction, we tested whether the genomic potential (metagenome) correlates with levels of the end products (metabolome) in the lung environment by fitting these two datasets using a Procrustes approach. Because pneumotype$_{BPT}$ is similar to the saline

background, we hypothesized that pneumotype$_{BPT}$ would correlate less well to the metabolome than pneumotype$_{SPT}$. Monte Carlo analysis of goodness of fit metagenome/metabolome demonstrated better correlation between metabolic pathways and metabolites for pneumotype$_{SPT}$ than for pneumotype$_{BPT}$ (M^2 0.10 for pneumotype$_{SPT}$ vs. 0.85 for pneumotype$_{BPT}$ $p<0.01$). Correlation matrices were then constructed between metagenome and metabolome for both pneumotypes. A comparison of correlation coefficients between the two data sets shows that pneumotype$_{SPT}$ had significantly higher Spearman Rho than pneumotype$_{BPT}$ (**Supplemental Figure 6**, $p<0.001$). The correlation between the genomic potential of pneumotype$_{SPT}$ and the metabolome in the lower airway samples supports the presence of active microbial metabolism in this environment. For instance, fucose-rhamnose, a bacterially synthesized carbohydrate, is directly correlated with BAL UniFrac distance to upper airway. Rhamnose is a deoxyhexose sugar found widely in bacteria but not in humans. Importantly, rhamnose is a constituent of the cell wall and a substrate for lipopolysaccharides (LPS) of some gram-negative bacteria.[29-31] The decrease in levels of this metabolite in association with enrichment with oral taxa in the lower airways (lower pairwise BAL UniFrac distance to upper airway) may therefore represent its utilization for synthesis of LPS. Alternatively, bacteria present in pneumotype$_{BPT}$ may be metabolically active and able to synthesize this cell wall precursor. However, the poorer correlation between metagenome and metabolome for pneumotype$_{BPT}$ suggests that metabolome found in these

DOI: 10.1038/NMICROBIOL.2016.31

samples is predominantly representative of the endogenous lung mammalian cell metabolism.

BAL cell differential and transcriptome of bronchial epithelial cells

BAL cell differentials were available for subjects from the NYU and LHMP cohorts (n=45). Overall, these subjects had relatively normal BAL cell differentials. However, compared to subjects with pneumotype$_{BPT}$, subjects with pneumotype$_{SPT}$ had significantly higher percentages of lymphocyte (4.5 [3.2-6.7] vs. 7.4 [4.6-11.9] percent of BAL cells, respectively) and a non-significant trend towards lower percentage of macrophages (**Supplemental Table 4**). No differences were noted in neutrophils or eosinophils in BAL. To better understand the lymphocyte dynamics in the pulmonary compartment, we examined Th-17/Treg balance in paired BAL and blood samples of 11 subjects (6 with pneumotype$_{SPT}$ and 5 with pneumotype$_{BPT}$) in whom sufficient number of lymphocytes (>10^6) were obtained from BAL cells. **Supplemental Figure 7 Panel A** shows a representative pair of FACS analysis of blood and BAL lymphocytes. **Supplemental Figure 7 Panel B** shows BAL has a higher Th17/Treg ratio than blood.

Transcriptome of bronchial epithelial cells

To compare the host mucosal immune response to the two pneumotypes in a subset of 12 subjects (5/12 with lower airway microbiome characterized as

pneumotype$_{SPT}$ and 7/12 with lower airway microbiome characterized as pneumotype$_{BPT}$), paired airway brushings were obtained and used to analyze the transcriptome of bronchial epithelial cells by Affymetrix array. This platform yielded a total of 54,675 mRNAs. However, despite the small number of paired samples available, 2,834 mRNAs were statistically significantly different between both pneumotypes ($p<0.05$). Top discriminant mRNAs ($p<0.05$, fold change [FC] >2.0) were used for a hierarchical cluster analysis (total of 62 probes, **Supplemental Figure 8**) of which 53 mRNAs were up-regulated in pneumotype$_{SPT}$ compared with pneumotype$_{BPT;}$ 33/53 have immune phenotypes; 17/33 support the innate or adaptive immunity, (Zfand5, MMP10, SUCNR1, Phactr2, TSLP, CDH11, PTPRC [CD45], EVI2B [CD361], CYYB, PTPRO, MS4A7, TLR8, TREM-1 [CD354], POST, NTS, LPL and ORI)[32-59] while 16/33 have counter-regulatory phenotypes limiting the immune response (PPARG, Tbrg1, LYZ, PTPRG, RUNX1T1, FST, EGFR, MARCH1, LEPR, Snai2, PIK3R1, PLEKHA1, SULT1E1, TNKS/TNKS2, CD52, SVL).[60-83] Eight mRNAs were down-regulated pneumotype$_{SPT}$ including mRNAs associated with mRNA translation (EEF1D, RPL37A, RPL38, and MRPL43)[84-87] and APOBEC3B, which is associated with mRNA editing and innate immune antiviral gene activity.[88] Ingenuity Pathway Analysis (IPA) also identified inflammatory response as the top diseases/biological function ($p<0.01E-04$). Transcriptome analysis of airway tissue showed that UniFrac distance to upper airway was associated with increased expression of STAT-3 (**Figure 4 Panel C** of main document). Top regulator effect networks were the ERK1/2 and PI3K/Akt signaling pathways

DOI: 10.1038/NMICROBIOL.2016.31

(Consistency Score =15.501, **Supplemental Figure 9**), which have previously been shown to be associated with inflammation, and, more specifically, with Treg/Th17 balance[89-91] by increasing Th17 differentiation.[92,93] Interestingly, many of STAT3 downstream molecules (FST, LYZ, HP, SNAI2, and LEPR)[94-99] were significantly higher compared to pneumotype$_{BPT}$ (**Supplemental Figure 8**). The up-regulated leptin receptor (LEPR) gene in epithelial cells associated with pneumotype$_{SPT}$ has been shown to be required for Th17 differentiation through STAT3 interaction.[100] TSLP, an IL-7-like cytokine also up-regulated in bronchial epithelial cells in association with pneumotype$_{SPT}$, can induce the differentiation of Th17 cells though dendritic cell activation.[101] IPA integrated mRNA-metabolome analysis was also performed for subjects where paired airway brushing samples and BAL metabolome was done. This analysis showed 21/83 metabolites (Score=47, #1 network) with KEGG annotations were associated with the ERK 1/2 signaling pathway and 15/83 metabolites (Score =31, #3 network) were associated with PI3k/Akt signaling pathway. These findings raise the hypothesis that with pneumotype$_{SPT}$, the airway mucosa has transcriptional changes relevant to establishing a Th17 phenotype.

Pneumotype and in-vivo lung inflammation

To further characterize lower airway immune phenotype associated with each pneumotype, BAL cytokine levels were measured using a Luminex multiplex platform for the 29 subjects from whom sufficient BAL fluid was available (**Supplemental Table 5**). Compared with pneumotype$_{BPT}$, pneumotype$_{SPT}$ had

significantly higher levels of IL-1α, IL-1 receptor antagonist (IL-1ra), IL-7, IL-8, epidermal growth factor (EGF), transforming growth factor-α (TGF-α), GRO and Fractalkine. Compared to pneumotype$_{SPT}$, no BAL cytokines were significantly elevated in pneumotype$_{BPT}$. Thus, pneumotype$_{SPT}$ and its marker taxa (e.g. *Veillonella, Prevotella*) were associated with increased *in vivo* levels of several cytokines relevant for Th17 differentiation (IL-1β, IL-6) or chemotaxis (Fractalkine). Naïve T cell differentiation into the Th17 subset is initiated by IL-6, while IL-1 cytokines are more important at later stages inducing effector cytokine production in a T cell receptor (TCR)-independent manner.[102] Other cytokines important for Th17 differentiation, such as IL-23 and TGFβ,[103] were not measured. IL-7, important for maturation of both T and B cell lymphoid lineage, was associated with pneumotype$_{SPT}$ and its marker taxa. IL-7 has a substantial role in shaping the pulmonary lymphocyte repertoire by enhancing maturation of T cells, including Th1, Th17, CD8 and γδ T cells. Fractalkine is involved in homing of Th17 cells to mucosa in the gut and the lung.[104] Th17 cells secrete IL-17A, IL-17F and IL-22, which act on resident antigen presenting cells and adjacent epithelial and endothelial cells to elicit inflammatory cytokine and chemokine production, recruiting neutrophils and innate epithelial defenses.

Pneumotype and ex-vivo TLR response

To evaluate the balance of pro-inflammatory and counterregulatory innate immune pathways, we performed *ex-vivo* TLR-4 stimulation with LPS on alveolar

DOI: 10.1038/NMICROBIOL.2016.31

macrophages (since they are the predominant BAL leukocyte) obtained from 18 subjects of the NYU cohort. Of these 18 subjects, 8 were from pneumotype$_{BPT}$ while 10 were from pneumotype$_{SPT}$, and we were thus able to compare cytokine production in response to LPS stimulation in both groups. Compared with BAL macrophages from pneumotype$_{BPT}$, BAL macrophages from pneumotype$_{SPT}$ were less responsive to LPS as shown by a lower fold increase of IL-6, MDC, TNF-α and MIP-1α (**Supplemental Table 6**). Thus, our data indicate close association between pneumotype$_{SPT}$ and pro-inflammatory Th17 airway phenotype, and with a blunted alveolar macrophage TLR4 response. These latter phenomena could stem from counter-regulatory mechanisms triggered by chronic subclinical inflammation or, alternatively, immune tolerance due to direct interactions with particular microbes within the microbiota.

Multivariate analysis to examine the effects of smoking

To examine if smoking confounded the association between pneumotype and any of the biomarkers found associated with pneumotype$_{SPT}$ we used a multivariate logistic regression model. In this model, metabolites, cells, and cytokines were considered outcome (dichotomized as below or above the median), while pneumotype and smoking status (never smoker vs. smoker) were predictors (covariates). This analysis (shown in **Supplemental Figure 10)** demonstrates that among the metabolites found to be decreased in pneumotype$_{SPT}$, the intensity of glyceric acid and cellobiose were associated with pneumotype$_{SPT}$ independently of smoking status. Among the BAL cells found to

be associated with pneumotype$_{SPT}$, lower percentage of macrophages were associated with pneumotype$_{SPT}$ independently of smoking status. After adjusting for smoking, pneumotype$_{SPT}$ predicted elevated levels of IL-1α, IL-7, EGF, TGF-α and Fractalkine. Finally, among the cytokines in which TLR4 induced productions were found to be blunted in pneumotype$_{SPT}$, IL-6 remained statistically significant after adjusting by smoking status. The other biomarkers examined were no longer significantly associated with pneumotype, indicating that in those associations, smoking is likely to be either a confounder or a variable with significant collinearity.

DOI: 10.1038/NMICROBIOL.2016.31

Supplementary References

1 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme J* **6**, 1621-1624, doi:10.1038/ismej.2012.8 (2012).

2 Walters, W. A. *et al.* PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* **27**, 1159-1161, doi:10.1093/bioinformatics/btr087 (2011).

3 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).

4 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).

5 McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *Isme J* **6**, 610-618, doi:10.1038/ismej.2011.139 (2012).

6 Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368-375 (2003).

7 Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**, 8228-8235, doi:71/12/8228 [pii]
10.1128/AEM.71.12.8228-8235.2005 (2005).

8 Dray, S. a. D., A.B. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**, 1-20 (2007).

9 Cailliez, F. The analytical solution of the additive constant problem. *Psychometrika* **48**, 305-310 (1983).

10 Zhao, G. *et al.* Identification of novel viruses using VirusHunter--an automated data analysis pipeline. *PLoS One* **8**, e78470, doi:10.1371/journal.pone.0078470 (2013).

11 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, doi:10.14806/ej.17.1.200
pp. 10-12 (2011).

12 Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863-864, doi:10.1093/bioinformatics/btr026 (2011).

13 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).

14 Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659, doi:10.1093/bioinformatics/btl158 (2006).

15 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).
16 Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome research* **21**, 1552-1560, doi:10.1101/gr.120618.111 (2011).
17 Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome biology* **12**, R60, doi:10.1186/gb-2011-12-6-r60 (2011).
18 Scholz, M. & Fiehn, O. SetupX--a public study design database for metabolomic projects. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 169-180 (2007).
19 Fiehn, O. *et al.* Quality control for plant metabolomics: reporting MSI-compliant studies. *The Plant journal : for cell and molecular biology* **53**, 691-704, doi:10.1111/j.1365-313X.2007.03387.x (2008).
20 Weckwerth, W., Wenzel, K. & Fiehn, O. Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* **4**, 78-83 (2004).
21 Fiehn, O. Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Analyt Chem* **27**, 261-269 (2008).
22 Kind, T., Tolstikov, V., Fiehn, O. & Weiss, R. H. A comprehensive urinary metabolomic approach for identifying kidney cancerr. *Anal Biochem* **363**, 185-195 (2007).
23 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
24 Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nat Methods* **8**, 761-763, doi:10.1038/nmeth.1650 (2011).
25 Segal, L. N. *et al.* Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* **1**, 19, doi:10.1186/2049-2618-1-19 (2013).
26 Huang, Y. J. *et al.* Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. *Journal of clinical microbiology* **52**, 2813-2823, doi:10.1128/JCM.00035-14 (2014).
27 Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* **31**, 814-821, doi:10.1038/nbt.2676 (2013).
28 Metcalf, J. L. *et al.* Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science*, doi:10.1126/science.aad2646 (2015).
29 Yokota, S. *et al.* Characterization of a polysaccharide component of lipopolysaccharide from Pseudomonas aeruginosa IID 1008 (ATCC 27584) as D-rhamnan. *European journal of biochemistry / FEBS* **167**, 203-209 (1987).
30 Senchenkova, S. N., Shashkov, A. S., Knirel, Y. A., McGovern, J. J. & Moran, A. P. The O-specific polysaccharide chain of Campylobacter fetus serotype B

lipopolysaccharide is a D-rhamnan terminated with 3-O-methyl-D-rhamnose (D-acofriose). *European journal of biochemistry / FEBS* **239**, 434-438 (1996).
31 Kocharova, N. A., Knirel, Y. A., Widmalm, G., Jansson, P. E. & Moran, A. P. Structure of an atypical O-antigen polysaccharide of Helicobacter pylori containing a novel monosaccharide 3-C-methyl-D-mannose. *Biochemistry* **39**, 4755-4760 (2000).
32 Magazin, M. *et al.* Sezary syndrome cells unlike normal circulating T lymphocytes fail to migrate following engagement of NT1 receptor. *The Journal of investigative dermatology* **122**, 111-118, doi:10.1046/j.0022-202X.2003.22131.x (2004).
33 Saada, S. *et al.* Differential expression of neurotensin and specific receptors, NTSR1 and NTSR2, in normal and malignant human B lymphocytes. *J Immunol* **189**, 5293-5303, doi:10.4049/jimmunol.1102937 (2012).
34 Masuoka, M. *et al.* Periostin promotes chronic allergic inflammation in response to Th2 cytokines. *J Clin Invest* **122**, 2590-2600, doi:10.1172/JCI58978 (2012).
35 Bouchon, A., Facchetti, F., Weigand, M. A. & Colonna, M. TREM-1 amplifies inflammation and is a crucial mediator of septic shock. *Nature* **410**, 1103-1107, doi:10.1038/35074114 (2001).
36 Gibot, S. *et al.* Soluble triggering receptor expressed on myeloid cells and the diagnosis of pneumonia. *N Engl J Med* **350**, 451-458, doi:10.1056/NEJMoa031544 (2004).
37 Heil, F. *et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526-1529, doi:10.1126/science.1093620 (2004).
38 Zuccolo, J. *et al.* Expression of MS4A and TMEM176 Genes in Human B Lymphocytes. *Frontiers in immunology* **4**, 195, doi:10.3389/fimmu.2013.00195 (2013).
39 Juszczynski, P. *et al.* BCL6 modulates tonic BCR signaling in diffuse large B-cell lymphomas by repressing the SYK phosphatase, PTPROt. *Blood* **114**, 5315-5321, doi:10.1182/blood-2009-02-204362 (2009).
40 Porter, C. D. *et al.* p22-phox-deficient chronic granulomatous disease: reconstitution by retrovirus-mediated expression and identification of a biosynthetic intermediate of gp91-phox. *Blood* **84**, 2767-2775 (1994).
41 Stasia, M. J. *et al.* Characterization of six novel mutations in the CYBB gene leading to different sub-types of X-linked chronic granulomatous disease. *Human genetics* **116**, 72-82, doi:10.1007/s00439-004-1208-5 (2005).
42 Matesanz-Isabel, J. *et al.* New B-cell CD molecules. *Immunology letters* **134**, 104-112, doi:10.1016/j.imlet.2010.09.019 (2011).
43 Hermiston, M. L., Xu, Z. & Weiss, A. CD45: a critical regulator of signaling thresholds in immune cells. *Annual review of immunology* **21**, 107-137, doi:10.1146/annurev.immunol.21.120601.140946 (2003).
44 Lee, D. M. *et al.* Cadherin-11 in synovial lining formation and pathology in arthritis. *Science* **315**, 1006-1010, doi:10.1126/science.1137306 (2007).

45 Kou, X. X. *et al.* Estradiol promotes M1-like macrophage activation through cadherin-11 to aggravate temporomandibular joint inflammation in rats. *J Immunol* **194**, 2810-2818, doi:10.4049/jimmunol.1303188 (2015).

46 Wu, M., Xu, T., Zhou, Y., Lu, H. & Gu, Z. Pressure and inflammatory stimulation induced increase of cadherin-11 is mediated by PI3K/Akt pathway in synovial fibroblasts from temporomandibular joint. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* **21**, 1605-1612, doi:10.1016/j.joca.2013.07.015 (2013).

47 Allakhverdi, Z. *et al.* Thymic stromal lymphopoietin is released by human epithelial cells in response to microbes, trauma, or inflammation and potently activates mast cells. *J Exp Med* **204**, 253-258, doi:10.1084/jem.20062211 (2007).

48 Ito, T. *et al.* TSLP-activated dendritic cells induce an inflammatory T helper type 2 cell response through OX40 ligand. *J Exp Med* **202**, 1213-1223, doi:10.1084/jem.20051135 (2005).

49 Leikauf, G. D. *et al.* Functional genomic assessment of phosgene-induced acute lung injury in mice. *Am J Respir Cell Mol Biol* **49**, 368-383, doi:10.1165/rcmb.2012-0337OC (2013).

50 Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124, doi:10.1038/nature11582 (2012).

51 Hakak, Y. *et al.* The role of the GPR91 ligand succinate in hematopoiesis. *Journal of leukocyte biology* **85**, 837-843, doi:10.1189/jlb.1008618 (2009).

52 Rubic, T. *et al.* Triggering the succinate receptor GPR91 on dendritic cells enhances immunity. *Nat Immunol* **9**, 1261-1269, doi:10.1038/ni.1657 (2008).

53 Gosselink, J. V. *et al.* Differential expression of tissue repair genes in the pathogenesis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **181**, 1329-1335, doi:10.1164/rccm.200812-1902OC (2010).

54 Dower, K., Ellis, D. K., Saraf, K., Jelinsky, S. A. & Lin, L. L. Innate immune responses to TREM-1 activation: overlap, divergence, and positive and negative cross-talk with bacterial lipopolysaccharide. *J Immunol* **180**, 3520-3534 (2008).

55 He, G., Sun, D., Ou, Z. & Ding, A. The protein Zfand5 binds and stabilizes mRNAs with AU-rich elements in their 3'-untranslated regions. *J Biol Chem* **287**, 24967-24977, doi:10.1074/jbc.M112.362020 (2012).

56 Klingenberg, R. *et al.* Depletion of FOXP3+ regulatory T cells promotes hypercholesterolemia and atherosclerosis. *J Clin Invest* **123**, 1323-1334, doi:10.1172/JCI63891 (2013).

57 Joo, H. *et al.* C-type lectin-like receptor LOX-1 promotes dendritic cell-mediated class-switched B cell responses. *Immunity* **41**, 592-604, doi:10.1016/j.immuni.2014.09.009 (2014).

58 Wang, L. *et al.* Triglyceride-rich lipoprotein lipolysis releases neutral and oxidized FFAs that induce endothelial cell inflammation. *Journal of lipid research* **50**, 204-213, doi:10.1194/jlr.M700505-JLR200 (2009).

59 Michaud, S. E. & Renier, G. Direct regulatory effect of fatty acids on macrophage lipoprotein lipase: potential role of PPARs. *Diabetes* **50**, 660-666 (2001).

60 Liu, H. P. *et al.* Association of supervillin with KIR2DL1 regulates the inhibitory signaling of natural killer cells. *Cellular signalling* **23**, 487-496, doi:10.1016/j.cellsig.2010.11.001 (2011).

61 Watanabe, T. *et al.* CD52 is a novel costimulatory molecule for induction of CD4+ regulatory T cells. *Clinical immunology* **120**, 247-259, doi:10.1016/j.clim.2006.05.006 (2006).

62 Bandala-Sanchez, E. *et al.* T cell regulation mediated by interaction of soluble CD52 with the inhibitory receptor Siglec-10. *Nat Immunol* **14**, 741-748, doi:10.1038/ni.2610 (2013).

63 Levaot, N. *et al.* Loss of Tankyrase-mediated destruction of 3BP2 is the underlying pathogenic mechanism of cherubism. *Cell* **147**, 1324-1339, doi:10.1016/j.cell.2011.10.045 (2011).

64 Xu, Y. *et al.* Estrogen sulfotransferase (SULT1E1) regulates inflammatory response and lipid metabolism of human endothelial cells via PPARgamma. *Mol Cell Endocrinol* **369**, 140-149, doi:10.1016/j.mce.2013.01.020 (2013).

65 Carissimi, C. *et al.* miR-21 is a negative modulator of T-cell activation. *Biochimie* **107 Pt B**, 319-326, doi:10.1016/j.biochi.2014.09.021 (2014).

66 Landego, I. *et al.* Interaction of TAPP adapter proteins with phosphatidylinositol (3,4)-bisphosphate regulates B-cell activation and autoantibody production. *European journal of immunology* **42**, 2760-2770, doi:10.1002/eji.201242371 (2012).

67 Deau, M. C. *et al.* A human immunodeficiency caused by mutations in the PIK3R1 gene. *J Clin Invest* **124**, 3923-3928, doi:10.1172/JCI75746 (2014).

68 Lucas, C. L. *et al.* Heterozygous splice mutation in PIK3R1 causes human immunodeficiency with lymphoproliferation due to dominant activation of PI3K. *J Exp Med* **211**, 2537-2547, doi:10.1084/jem.20141759 (2014).

69 Pioli, P. D., Chen, X., Weis, J. J. & Weis, J. H. Fatal autoimmunity results from the conditional deletion of Snai2 and Snai3. *Cell Immunol* **295**, 1-18, doi:10.1016/j.cellimm.2015.02.009 (2015).

70 Busso, N. *et al.* Leptin signaling deficiency impairs humoral and cellular immune responses and attenuates experimental arthritis. *J Immunol* **168**, 875-882 (2002).

71 Chattopadhyay, G. & Shevach, E. M. Antigen-specific induced T regulatory cells impair dendritic cell function via an IL-10/MARCH1-dependent mechanism. *J Immunol* **191**, 5875-5884, doi:10.4049/jimmunol.1301693 (2013).

72 Tze, L. E. *et al.* CD83 increases MHC II and CD86 on dendritic cells by opposing IL-10-driven MARCH1-mediated ubiquitination and degradation. *J Exp Med* **208**, 149-165, doi:10.1084/jem.20092203 (2011).

73 Shim, J. J. *et al.* IL-13 induces mucin production by stimulating epidermal growth factor receptors and by activating neutrophils. *Am J Physiol Lung Cell Mol Physiol* **280**, L134-140 (2001).

74 Mascia, F., Mariani, V., Girolomoni, G. & Pastore, S. Blockade of the EGF receptor induces a deranged chemokine expression in keratinocytes leading to enhanced skin inflammation. *Am J Pathol* **163**, 303-312, doi:10.1016/S0002-9440(10)63654-1 (2003).

75 Pastore, S. *et al.* ERK1/2 regulates epidermal chemokine expression and skin inflammation. *J Immunol* **174**, 5047-5056 (2005).

76 Hardy, C. L. *et al.* The activin A antagonist follistatin inhibits cystic fibrosis-like lung inflammation and pathology. *Immunology and cell biology* **93**, 567-574, doi:10.1038/icb.2015.7 (2015).

77 Aoki, F., Kurabayashi, M., Hasegawa, Y. & Kojima, I. Attenuation of bleomycin-induced pulmonary fibrosis by follistatin. *Am J Respir Crit Care Med* **172**, 713-720, doi:10.1164/rccm.200412-1620OC (2005).

78 Guo, C., Hu, Q., Yan, C. & Zhang, J. Multivalent binding of the ETO corepressor to E proteins facilitates dual repression controls targeting chromatin and the basal transcription machinery. *Molecular and cellular biology* **29**, 2644-2657, doi:10.1128/MCB.00073-09 (2009).

79 Yan, M. *et al.* A previously unidentified alternatively spliced isoform of t(8;21) transcript promotes leukemogenesis. *Nat Med* **12**, 945-949, doi:10.1038/nm1443 (2006).

80 Mirenda, M. *et al.* Protein tyrosine phosphatase receptor type gamma is a JAK phosphatase and negatively regulates leukocyte integrin activation. *J Immunol* **194**, 2168-2179, doi:10.4049/jimmunol.1401841 (2015).

81 Lee, W., Ku, S. K., Na, D. H. & Bae, J. S. Anti-Inflammatory Effects of Lysozyme Against HMGB1 in Human Endothelial Cells and in Mice. *Inflammation*, doi:10.1007/s10753-015-0171-8 (2015).

82 Reed, S. M. *et al.* NIAM-deficient mice are predisposed to the development of proliferative lesions including B-cell lymphomas. *PLoS One* **9**, e112126, doi:10.1371/journal.pone.0112126 (2014).

83 Shan, M. *et al.* Agonistic induction of PPARgamma reverses cigarette smoke-induced emphysema. *J Clin Invest* **124**, 1371-1381, doi:10.1172/JCI70587 (2014).

84 Gazda, H. T. *et al.* Defective ribosomal protein gene expression alters transcription, translation, apoptosis, and oncogenic pathways in Diamond-Blackfan anemia. *Stem Cells* **24**, 2034-2044, doi:10.1634/stemcells.2005-0554 (2006).

85 Fumagalli, S. *et al.* Absence of nucleolar disruption after impairment of 40S ribosome biogenesis reveals an rpL11-translation-dependent mechanism of p53 induction. *Nat Cell Biol* **11**, 501-508, doi:10.1038/ncb1858 (2009).

86 Kondrashov, N. *et al.* Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383-397, doi:10.1016/j.cell.2011.03.028 (2011).

87 Antonicka, H. & Shoubridge, E. A. Mitochondrial RNA Granules Are Centers for Posttranscriptional RNA Processing and Ribosome Biogenesis. *Cell reports*, doi:10.1016/j.celrep.2015.01.030 (2015).

88 Mangeat, B. *et al.* Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103, doi:10.1038/nature01709 (2003).

89 Barbi, J., Pardoll, D. & Pan, F. Metabolic control of the Treg/Th17 axis. *Immunological reviews* **252**, 52-77, doi:10.1111/imr.12029 (2013).

90 Okkenhaug, K. *et al.* The p110delta isoform of phosphoinositide 3-kinase controls clonal expansion and differentiation of Th cells. *Journal of immunology (Baltimore, Md. : 1950)* **177**, 5122-5128 (2006).

91 Sauer, S. *et al.* T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7797-7802, doi:10.1073/pnas.0800928105 (2008).

92 Kurebayashi, Y. *et al.* PI3K-Akt-mTORC1-S6K1/2 axis controls Th17 differentiation by regulating Gfi1 expression and nuclear translocation of RORgamma. *Cell reports* **1**, 360-373, doi:10.1016/j.celrep.2012.02.007 (2012).

93 Liu, H. *et al.* ERK differentially regulates Th17- and Treg-cell development and contributes to the pathogenesis of colitis. *European journal of immunology* **43**, 1716-1726, doi:10.1002/eji.201242889 (2013).

94 Lee, J. H. *et al.* Signal transducer and activator of transcription-3 (Stat3) plays a critical role in implantation via progesterone receptor in uterus. *FASEB J* **27**, 2553-2563, doi:10.1096/fj.12-225664 (2013).

95 Kim, H. & Baumann, H. The carboxyl-terminal region of STAT3 controls gene induction by the mouse haptoglobin promoter. *J Biol Chem* **272**, 14571-14579 (1997).

96 Zauberman, A., Lapter, S. & Zipori, D. Smad proteins suppress CCAAT/enhancer-binding protein (C/EBP) beta- and STAT3-mediated transcriptional activation of the haptoglobin promoter. *J Biol Chem* **276**, 24719-24725, doi:10.1074/jbc.M005813200 (2001).

97 Heinrich, P. C. *et al.* Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *The Biochemical journal* **374**, 1-20, doi:10.1042/BJ20030407 (2003).

98 Yang, P., Li, Z., Fu, R., Wu, H. & Li, Z. Pyruvate kinase M2 facilitates colon cancer cell migration via the modulation of STAT3 signalling. *Cellular signalling* **26**, 1853-1862, doi:10.1016/j.cellsig.2014.03.020 (2014).

99 Pathak, R. R. *et al.* Loss of phosphatase and tensin homolog (PTEN) induces leptin-mediated leptin gene expression: feed-forward loop operating in the lung. *J Biol Chem* **288**, 29821-29835, doi:10.1074/jbc.M113.481523 (2013).

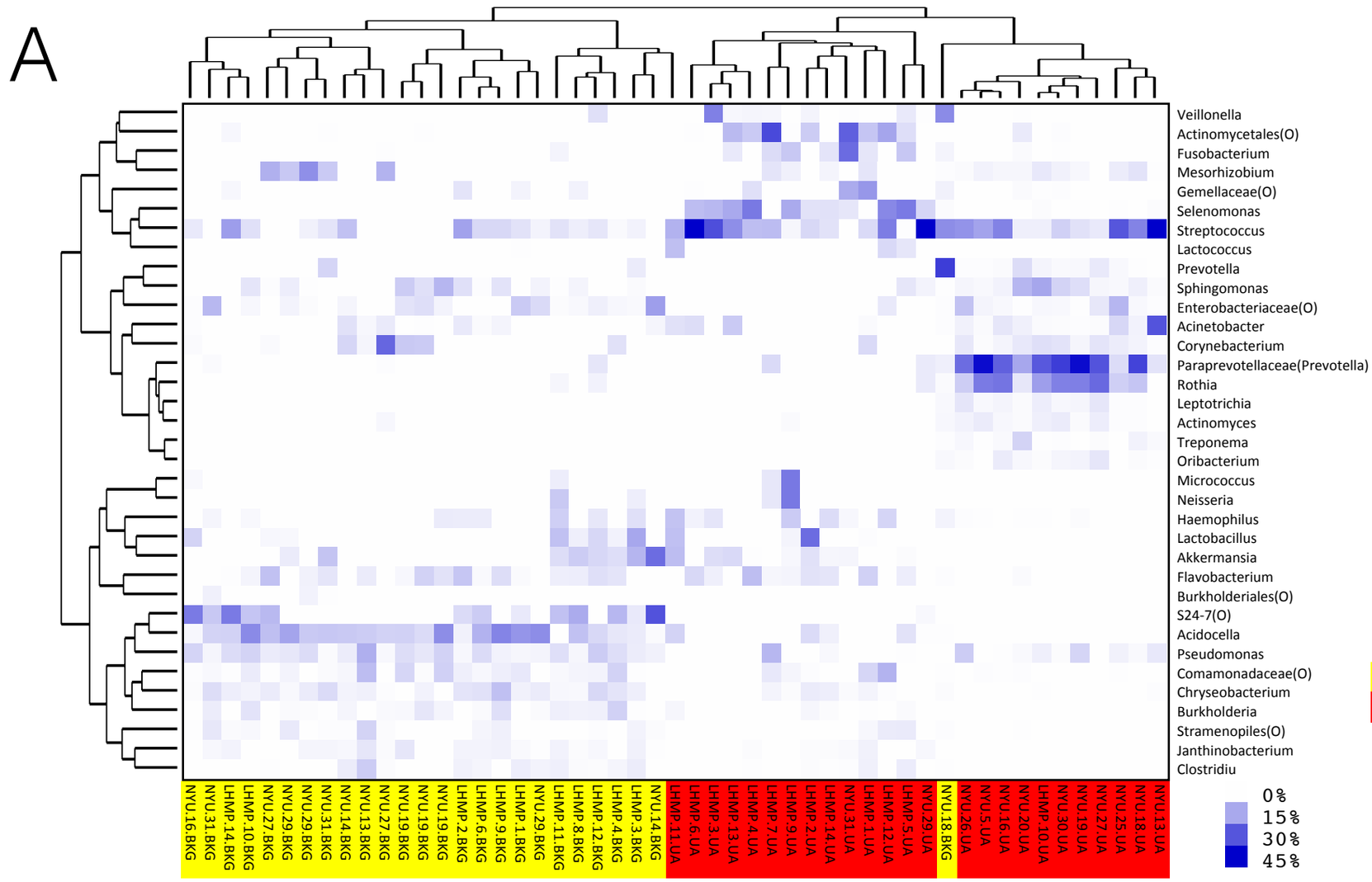
100 Reis, B. S. *et al.* Leptin receptor signaling in T cells is required for Th17 differentiation. **194**, 5253-5260, doi:10.4049/jimmunol.1402996 (2015).

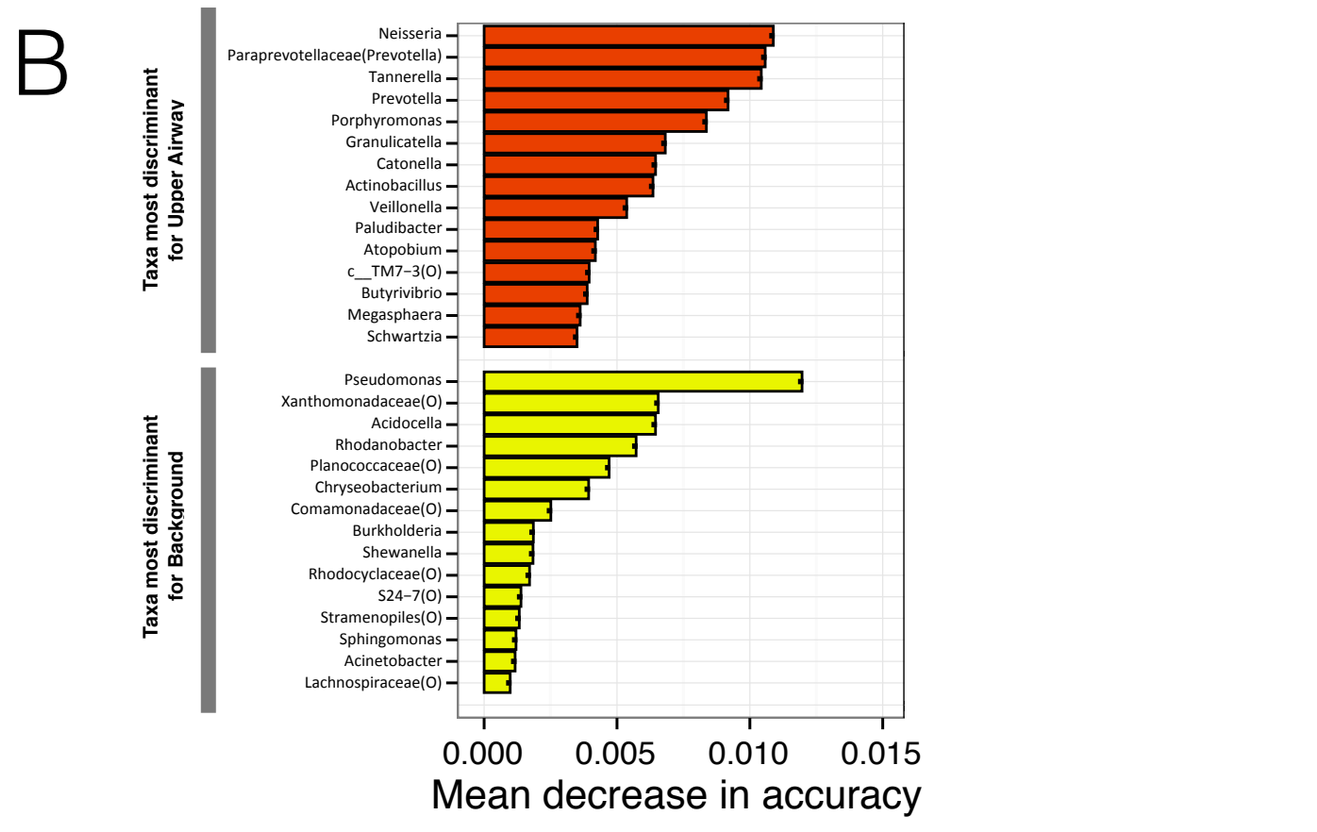
101 Tanaka, J. *et al.* Human TSLP and TLR3 ligands promote differentiation of Th17 cells with a central memory phenotype under Th2-polarizing conditions. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **39**, 89-100, doi:10.1111/j.1365-2222.2008.03151.x (2009).

102 Guo, L. *et al.* IL-1 family members and STAT activators induce cytokine production by Th2, Th17, and Th1 cells. *Proc Natl Acad Sci U S A* **106**, 13463-13468, doi:10.1073/pnas.0906988106 (2009).
103 Yang, L. *et al.* IL-21 and TGF-beta are required for differentiation of human T(H)17 cells. *Nature* **454**, 350-352, doi:10.1038/nature07021 (2008).
104 Medina-Contreras, O. *et al.* CX3CR1 regulates intestinal macrophage homeostasis, bacterial translocation, and colitogenic Th17 responses in mice. *J Clin Invest* **121**, 4787-4795, doi:10.1172/JCI59150 (2011).

Supplementary Figures

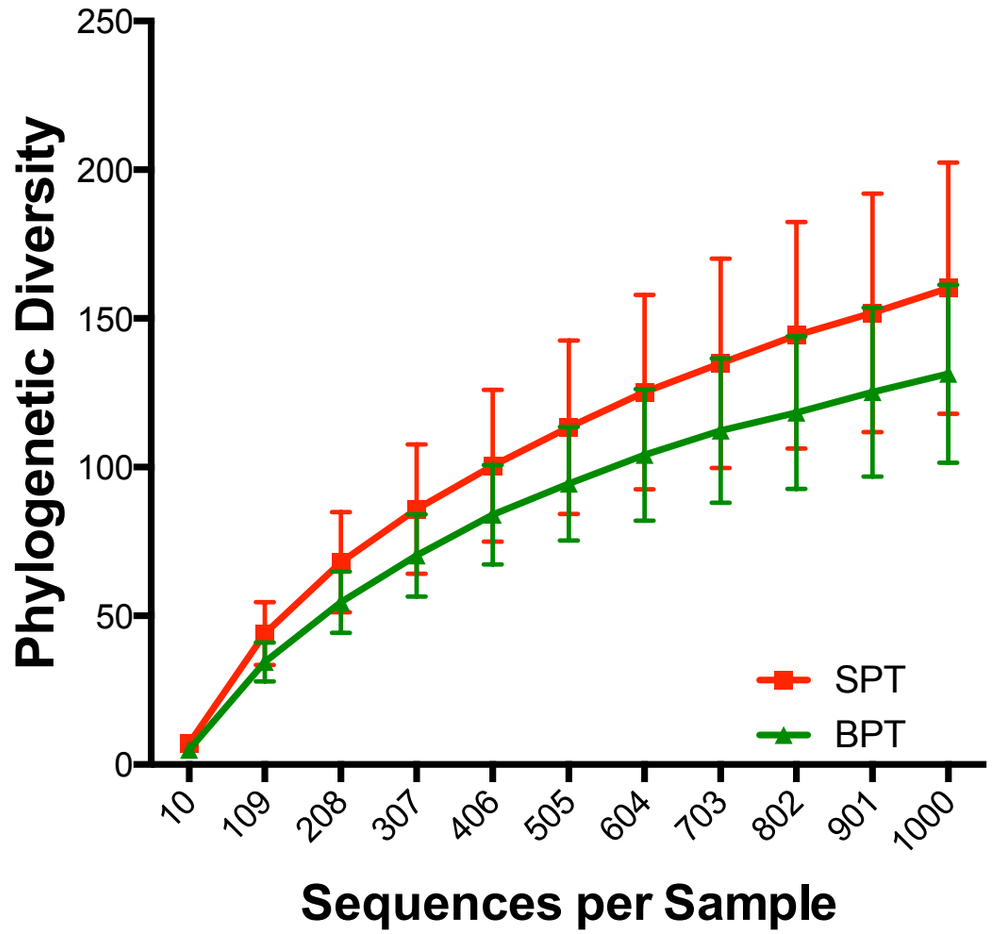
Figure S1


A
Veillonella
Actinomycetales(O)
Fusobacterium
Mesorhizobium
Gemellaceae(O)
Selenomonas
Streptococcus
Lactococcus
Prevotella
Sphingomonas
Enterobacteriaceae(O)
Acinetobacter
Corynebacterium
Paraprevotellaceae(Prevotella)
Rothia
Leptotrichia
Actinomyces
Treponema
Oribacterium
Micrococcus
Neisseria
Haemophilus
Lactobacillus
Akkermansia
Flavobacterium
Burkholderiales(O)
S24-7(O)
Acidocella
Pseudomonas
Comamonadaceae(O)
Chryseobacterium
Burkholderia
Stramenopiles(O)
Janthinobacterium
Clostridiu
Background
Upper Airway
0%
15%
30%
45%
NYU.16.BKG
NYU.31.BKG
LHMP.14.BKG
LHMP.10.BKG
NYU.27.BKG
NYU.29.BKG
NYU.29.BKG
NYU.31.BKG
NYU.14.BKG
NYU.13.BKG
NYU.27.BKG
NYU.19.BKG
NYU.19.BKG
NYU.19.BKG
LHMP.2.BKG
LHMP.6.BKG
LHMP.9.BKG
LHMP.1.BKG
NYU.29.BKG
LHMP.11.BKG
LHMP.8.BKG
LHMP.12.BKG
LHMP.4.BKG
LHMP.3.BKG
NYU.14.BKG
LHMP.11.UA
LHMP.6.UA
LHMP.3.UA
LHMP.13.UA
LHMP.4.UA
LHMP.7.UA
LHMP.9.UA
LHMP.2.UA
LHMP.14.UA
NYU.31.UA
LHMP.1.UA
LHMP.12.UA
LHMP.5.UA
NYU.29.UA
NYU.18.BKG
NYU.26.UA
NYU.5.UA
NYU.16.UA
NYU.20.UA
LHMP.10.UA
NYU.30.UA
NYU.19.UA
NYU.27.UA
NYU.25.UA
NYU.18.UA
NYU.13.UA


B
Taxa most discriminant for Upper Airway
Neisseria
Paraprevotellaceae(Prevotella)
Tannerella
Prevotella
Porphyromonas
Granulicatella
Catonella
Actinobacillus
Veillonella
Paludibacter
Atopobium
c__TM7-3(O)
Butyrivibrio
Megasphaera
Schwartzia
Taxa most discriminant for Background
Pseudomonas
Xanthomonadaceae(O)
Acidocella
Rhodanobacter
Planococcaceae(O)
Chryseobacterium
Comamonadaceae(O)
Burkholderia
Shewanella
Rhodocyclaceae(O)
S24-7(O)
Stramenopiles(O)
Sphingomonas
Acinetobacter
Lachnospiraceae(O)
0.000
0.005
0.010
0.015
Mean decrease in accuracy

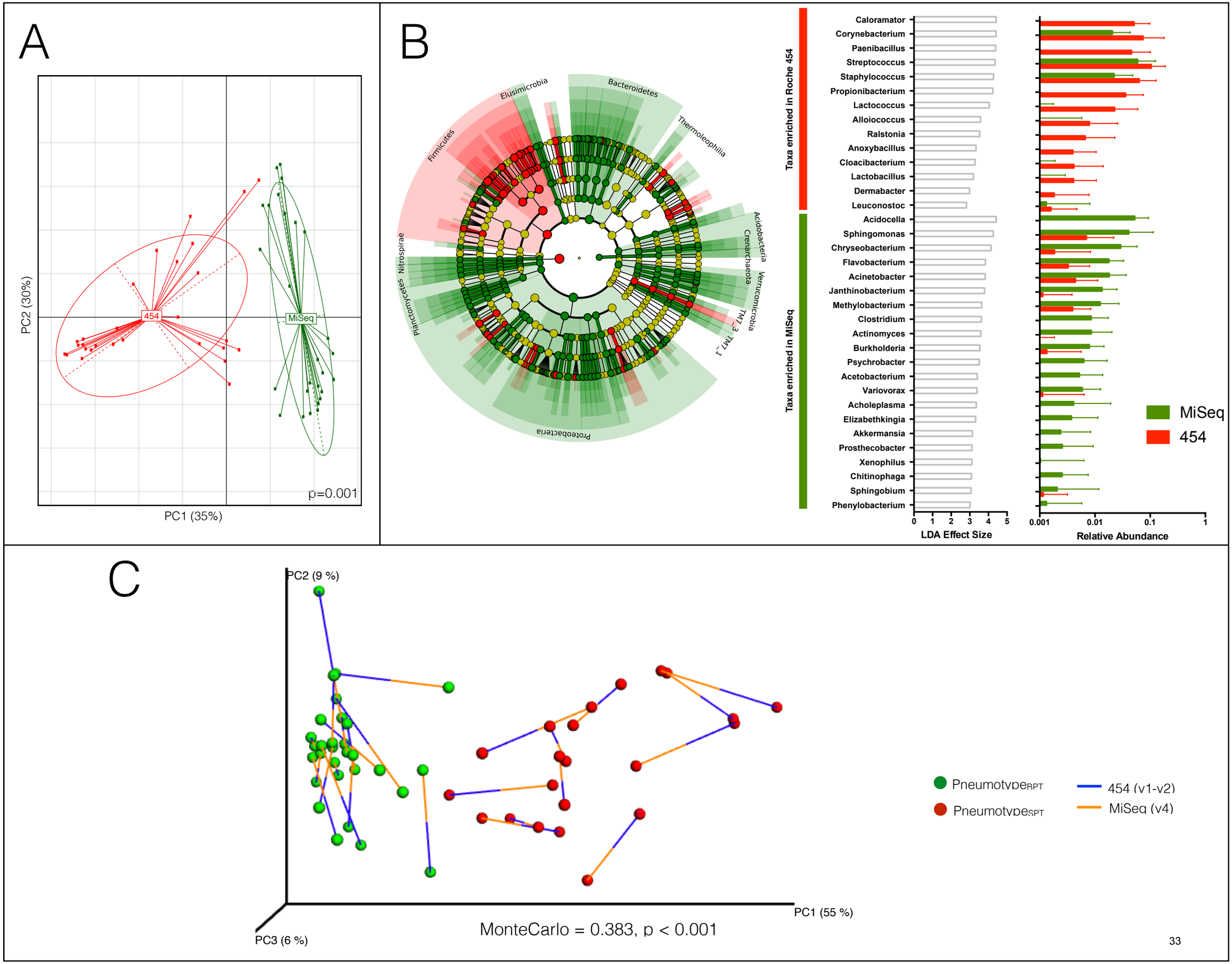
Supplemental Figure 1: Taxonomic differences in background and upper airway samples. Panel A. Unsupervised hierarchical clustering of most abundant taxa (relative abundance ≥ 3% in any sample) for upper airway samples obtained by oral wash or separate scope and background samples obtained by flushing saline through bronchoscope prior to bronchoscopy. Dendrogram shows deep cleft that separates most of upper airway samples from most of background samples. Panel B. Machine learning approach based on Random Forest to identify taxa most predictive for upper airway vs. background samples.

Figure S2


Phylogenetic Diversity
250
200
150
100
50
0
10
109
208
307
406
505
604
703
802
901
1000
SPT
BPT
Sequences per Sample

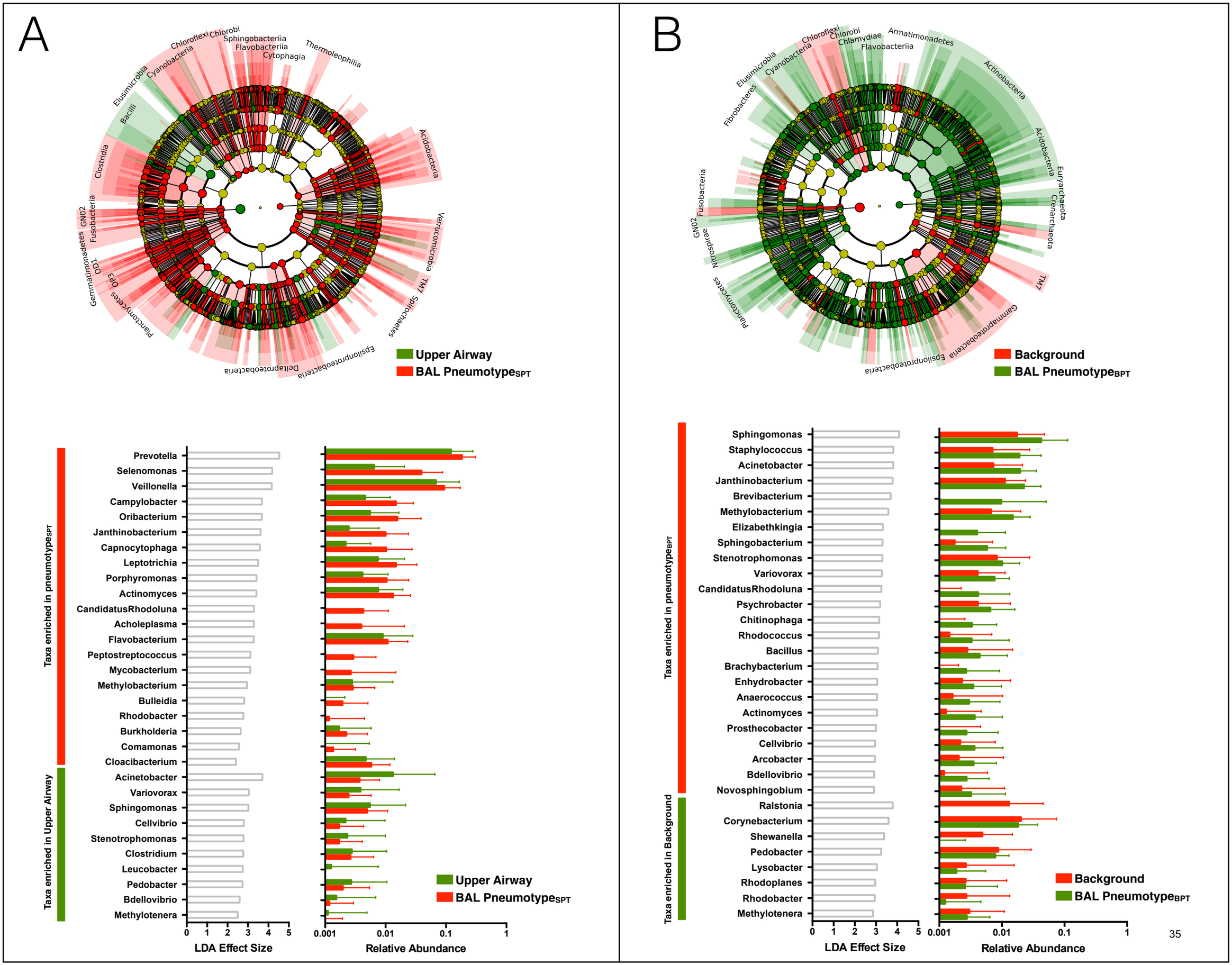
Supplemental Figure 2: Differences in α diversity between two pneumotypes. Alpha rarefaction curve of observed species as different sequence depth shows no significant differences in α diversity between $pneumotype_{BPT}$ and $pneumotype_{SPT}$.

Figure S3


A
PC2 (30%)
PC1 (35%)
454
MiSeq
p=0.001
B
Elusimicrobia
Bacteroidetes
Thermoleophilia
Firmicutes
Acidobacteria
Crenarchaeota
Verrucomicrobia
TM7_3
TM7_1
Nitrospirae
Planctomycetes
Proteobacteria
Taxa enriched in Roche 454
Taxa enriched in MiSeq
Caloramator
Corynebacterium
Paenibacillus
Streptococcus
Staphylococcus
Propionibacterium
Lactococcus
Alloiococcus
Ralstonia
Anoxybacillus
Cloacibacterium
Lactobacillus
Dermabacter
Leuconostoc
Acidocella
Sphingomonas
Chryseobacterium
Flavobacterium
Acinetobacter
Janthinobacterium
Methylobacterium
Clostridium
Actinomyces
Burkholderia
Psychrobacter
Acetobacterium
Variovorax
Acholeplasma
Elizabethkingia
Akkermansia
Prosthecobacter
Xenophilus
Chitinophaga
Sphingobium
Phenylobacterium
0 1 2 3 4 5
LDA Effect Size
0.001 0.01 0.1 1
Relative Abundance
MiSeq
454
C
PC2 (9 %)
PC1 (55 %)
PC3 (6 %)
MonteCarlo = 0.383, p < 0.001
PneumotypeBPT
PneumotypeSPT
454 (v1-v2)
MiSeq (v4)

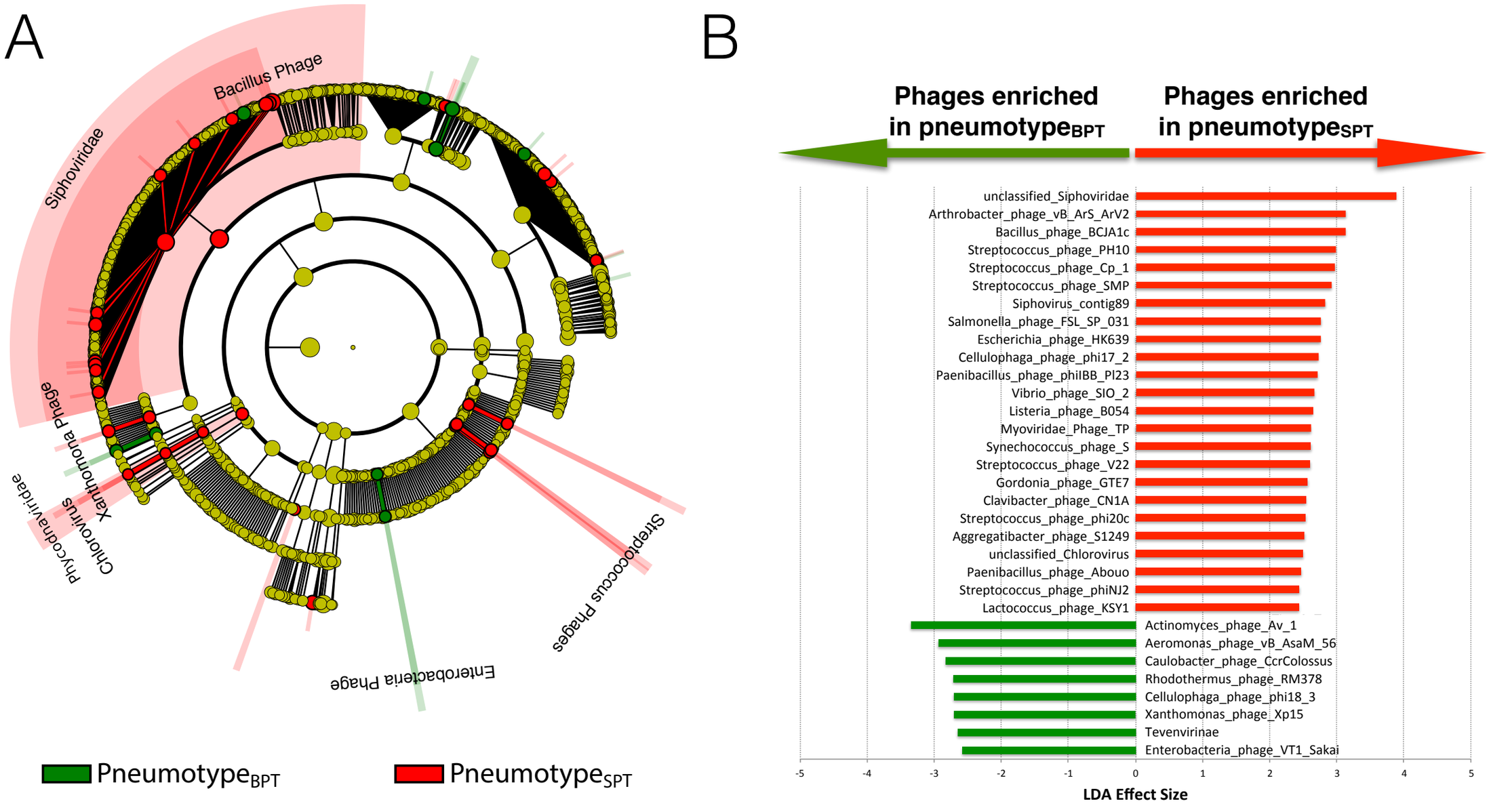
Supplemental Figure 3: Pneumotype assignment is consistent despite the use of two sequencing techniques. Panel A. PCoA based on weighted UniFrac Distance shows significant differences (Adonis p=0.001) in the microbiome beta diversity of 28 BAL samples sequenced by two different approaches: a 454 sequence targeting V1-V2 (data previously reported, reference number 3 of manuscript) and a MiSeq sequence targeting V4. Panel B. LEfSe analysis was performed to examine taxonomic differences between the two sequencing approaches and showed multiple taxonomic differences (LDA>2). Panel C. Procrustes analysis for β diversity distribution of these 28 BAL samples in which the two sequencing approaches were used. Pairs are connected and show similar β diversity distribution for the two clusters defined as pneumotypes (in terms of pneumotype_{SPT} being distinct from pneumotype_{BPT}) with both sequencing approaches.

Figure S4


A
Chloroflexi
Chlorobi
Sphingobacteriia
Flavobacteriia
Cytophagia
Thermoleophilia
Cyanobacteria
Elusimicrobia
Bacilli
Clostridia
Acidobacteria
Fusobacteria
GN02
Verrucomicrobia
Gemmatimonadetes
TM7
OD1
Spirochaetes
OP3
Planctomycetes
Deltaproteobacteria
Epsilonproteobacteria
Upper Airway
BAL Pneumotype SPT
B
Chloroflexi
Chlorobi
Chlamydiae
Armatimonadetes
Flavobacteriia
Elusimicrobia
Cyanobacteria
Actinobacteria
Fibrobacteres
Acidobacteria
Euryarchaeota
Fusobacteria
Crenarchaeota
GN02
Nitrospirae
TM7
Planctomycetes
Gammaproteobacteria
Epsilonproteobacteria
Background
BAL Pneumotype BPT
Taxa enriched in pneumotype SPT
Taxa enriched in Upper Airway
Prevotella
Selenomonas
Veillonella
Campylobacter
Oribacterium
Janthinobacterium
Capnocytophaga
Leptotrichia
Porphyromonas
Actinomyces
CandidatusRhodoluna
Acholeplasma
Flavobacterium
Peptostreptococcus
Mycobacterium
Methylobacterium
Bulleidia
Rhodobacter
Burkholderia
Comamonas
Cloacibacterium
Acinetobacter
Variovorax
Sphingomonas
Cellvibrio
Stenotrophomonas
Clostridium
Leucobacter
Pedobacter
Bdellovibrio
Methylotenera
0 1 2 3 4 5
LDA Effect Size
0.001 0.01 0.1 1
Relative Abundance
Upper Airway
BAL Pneumotype SPT
Taxa enriched in pneumotype BPT
Taxa enriched in Background
Sphingomonas
Staphylococcus
Acinetobacter
Janthinobacterium
Brevibacterium
Methylobacterium
Elizabethkingia
Sphingobacterium
Stenotrophomonas
Variovorax
CandidatusRhodoluna
Psychrobacter
Chitinophaga
Rhodococcus
Bacillus
Brachybacterium
Enhydrobacter
Anaerococcus
Actinomyces
Prosthecobacter
Cellvibrio
Arcobacter
Bdellovibrio
Novosphingobium
Ralstonia
Corynebacterium
Shewanella
Pedobacter
Lysobacter
Rhodoplanes
Rhodobacter
Methylotenera
0 1 2 3 4 5
LDA Effect Size
0.001 0.01 0.1 1
Relative Abundance
Background
BAL Pneumotype BPT

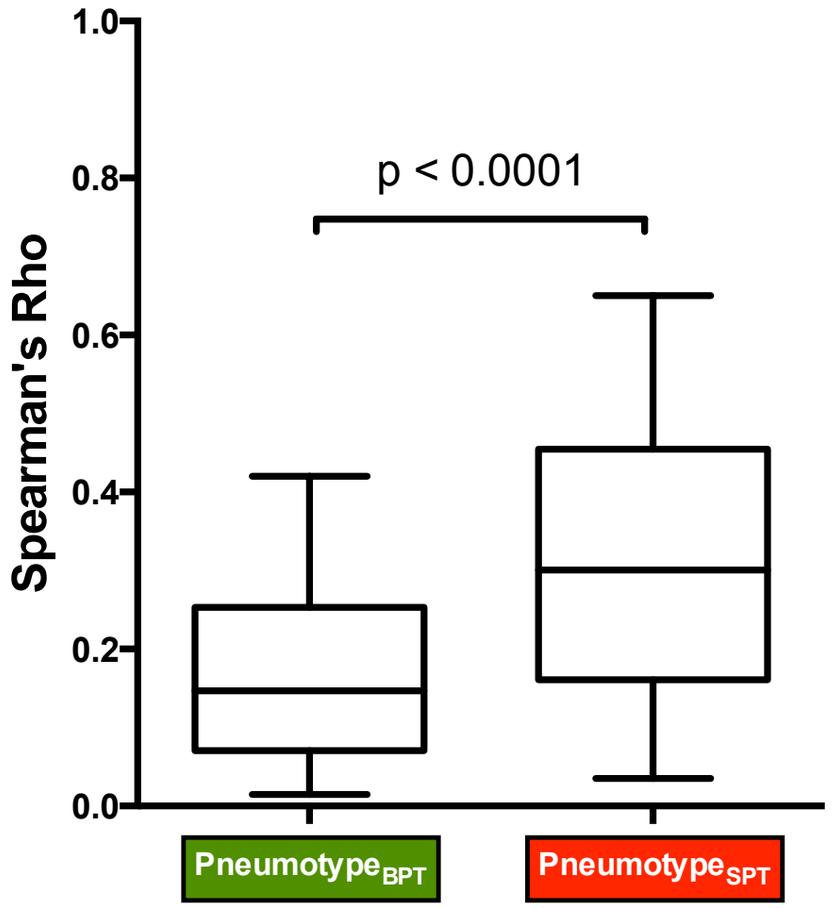
Supplemental Figure 4: Taxonomic differences between groups of samples. Panel A. Linear discriminant analysis (LDA) Effect Size (LEfSe) comparing taxonomic composition of BAL samples from pneumotype$_{SPT}$ vs. upper airway samples. Cladogram shows multiple differences at various taxonomic levels while histogram below represents LDA effect size and relative abundance for most discriminant genus. Panel B. Similar analysis was performed comparing taxonomic composition of BAL samples from pneumotype$_{BPT}$ vs. background samples.

Figure S5


A
Bacillus Phage
Siphoviridae
Phycodnaviridae
Chlorovirus
Xanthomona Phage
Enterobacteria Phage
Streptococcus Phages
Pneumotype$_{BPT}$
Pneumotype$_{SPT}$
B
Phages enriched in pneumotype$_{BPT}$
Phages enriched in pneumotype$_{SPT}$
unclassified_Siphoviridae
Arthrobacter_phage_vB_ArS_ArV2
Bacillus_phage_BCJA1c
Streptococcus_phage_PH10
Streptococcus_phage_Cp_1
Streptococcus_phage_SMP
Siphovirus_contig89
Salmonella_phage_FSL_SP_031
Escherichia_phage_HK639
Cellulophaga_phage_phi17_2
Paenibacillus_phage_phiIBB_Pl23
Vibrio_phage_SIO_2
Listeria_phage_B054
Myoviridae_Phage_TP
Synechococcus_phage_S
Streptococcus_phage_V22
Gordonia_phage_GTE7
Clavibacter_phage_CN1A
Streptococcus_phage_phi20c
Aggregatibacter_phage_S1249
unclassified_Chlorovirus
Paenibacillus_phage_Abouo
Streptococcus_phage_phiNJ2
Lactococcus_phage_KSY1
Actinomyces_phage_Av_1
Aeromonas_phage_vB_AsaM_56
Caulobacter_phage_CcrColossus
Rhodothermus_phage_RM378
Cellulophaga_phage_phi18_3
Xanthomonas_phage_Xp15
Tevenvirinae
Enterobacteria_phage_VT1_Sakai
-5 -4 -3 -2 -1 0 1 2 3 4 5
LDA Effect Size

Supplemental Figure 5: Differences in phageome between pneumotypes. Panel A. Cladogram represents results from calculated linear discriminant analysis (LDA) Effect Size (LEfSe) comparing phage composition of BAL samples from pneumotype$_{SPT}$ vs. pneumotype$_{BPT}$. Panel B. Multiple significant phage differences (LDA>2) were observed between pneumotypes.

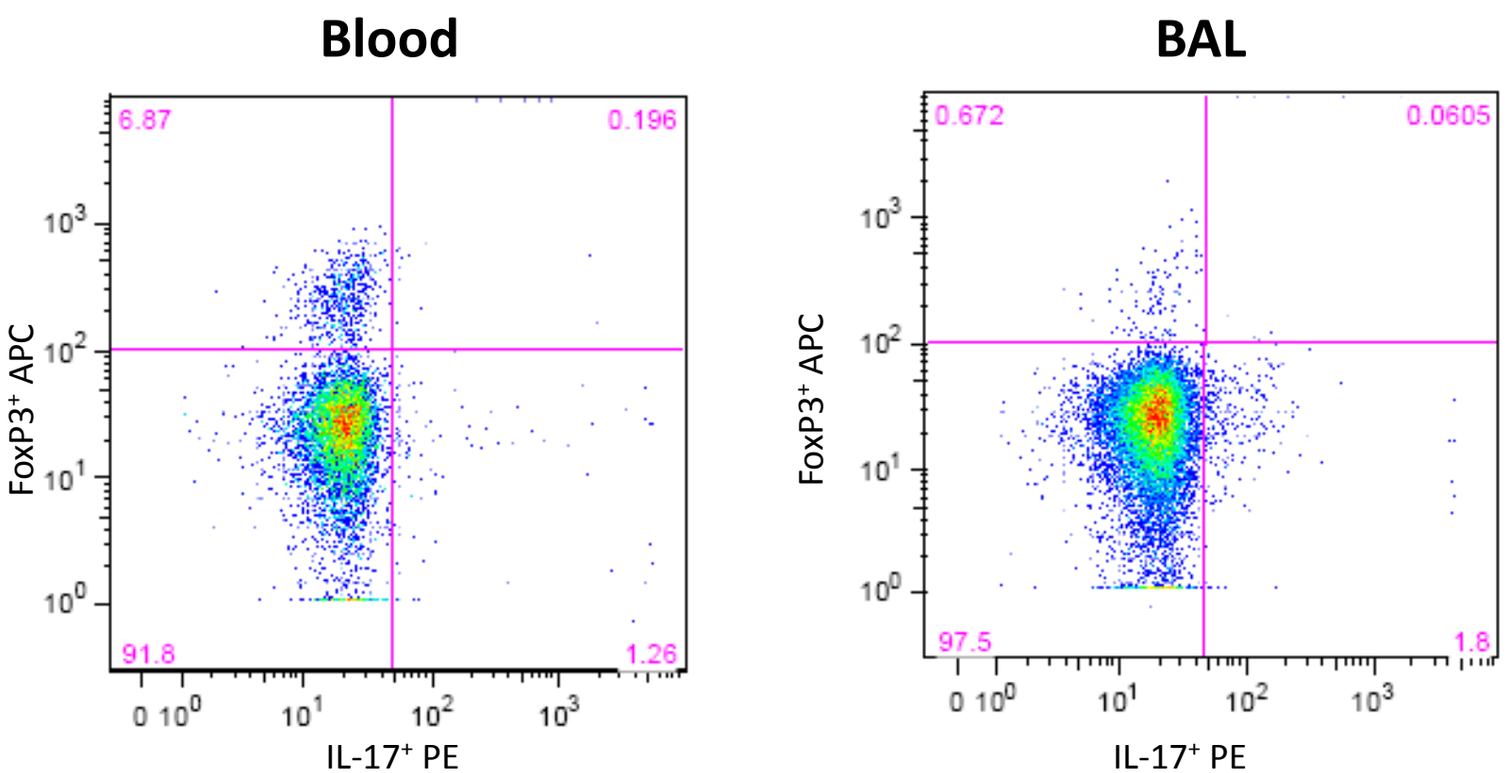
Figure S6


1.0
0.8
0.6
0.4
0.2
0.0
Spearman's Rho
p < 0.0001
Pneumotype BPT
Pneumotype SPT

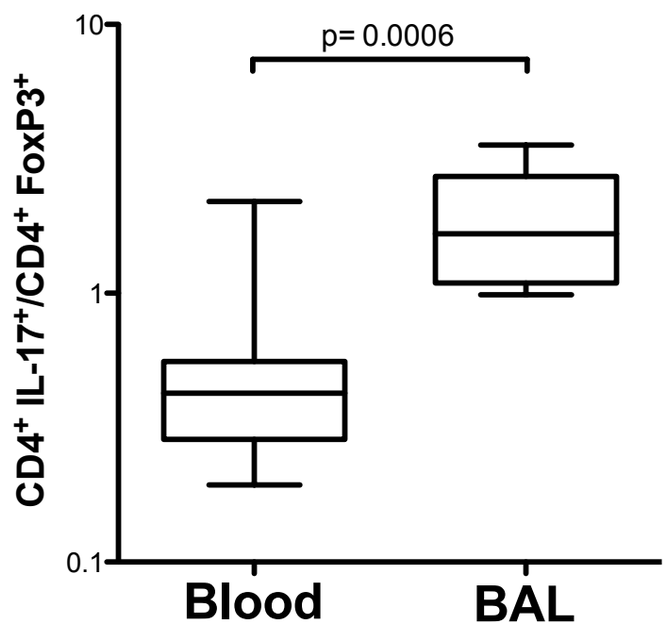
Supplemental Figure 6: Predicted metagenome correlates more tightly to the metabolome in pneumotype$_{SPT}$ than in pneumotype$_{BPT}$. Comparison of non-parametric correlation coefficients for matrix generated between the PICRUSt inferred metagenome and the metabolome data (11,952 paired metagenome/metabolome correlations, data represented as median[IQR], p-value based on Mann-Whitney U).

Figure S7

A

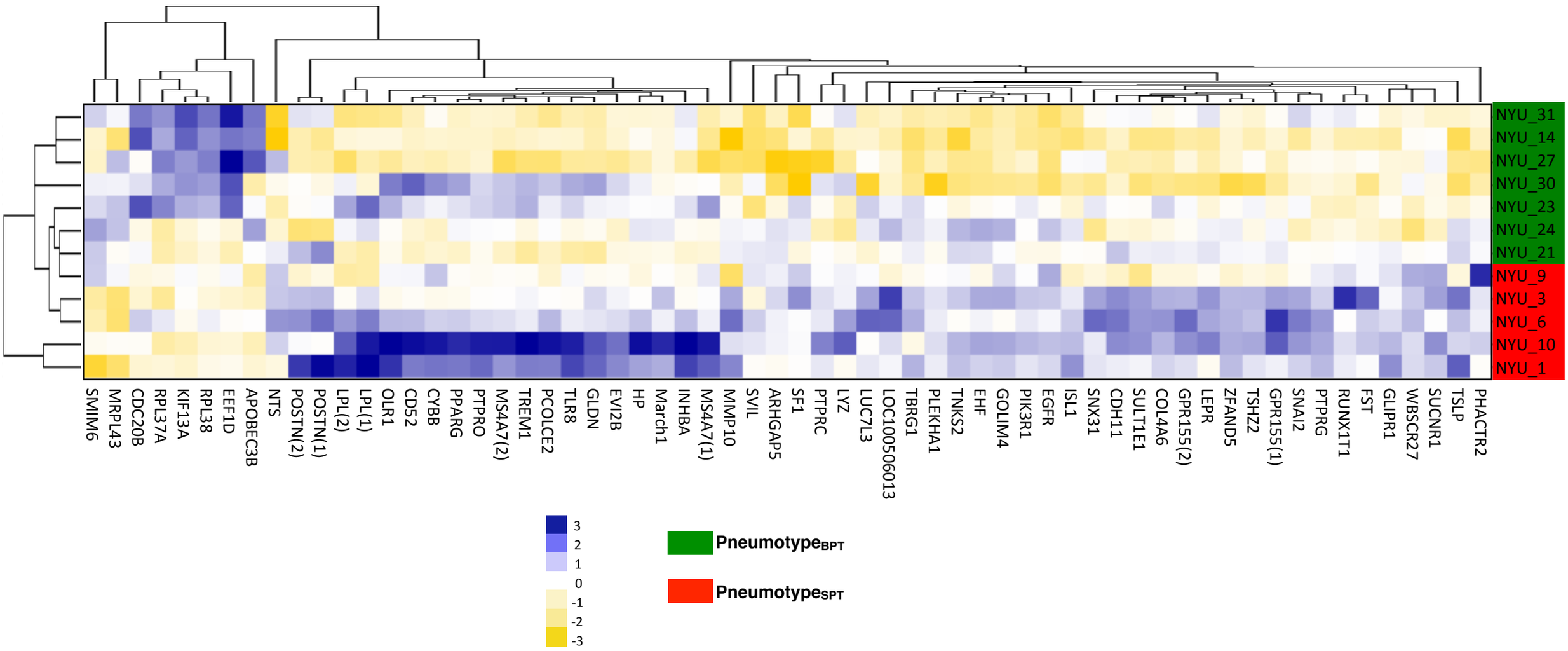

Blood
BAL
6.87
0.196
91.8
1.26
0.672
0.0605
97.5
1.8
FoxP3⁺ APC
IL-17⁺ PE
0
10^0
10^1
10^2
10^3

B


p= 0.0006
CD4⁺ IL-17⁺/CD4⁺ FoxP3⁺
10
1
0.1
Blood
BAL

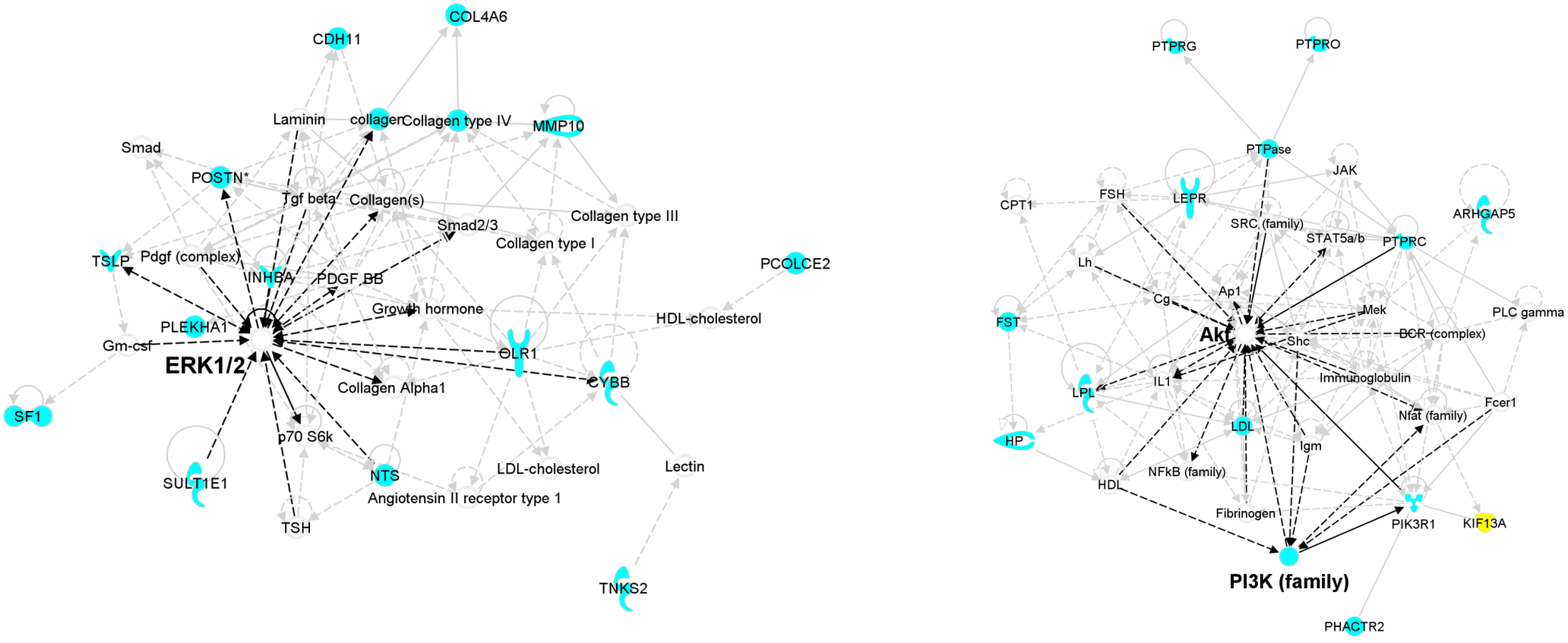
Supplemental Figure 7: Greater Th17 cell in the bronchoalveolar compartment. Panel A. Representative flow cytometry of $CD4^+$ cells intracellularly stained for IL-17 or FoxP3. Panel B. $CD4^+$ cells in BAL fluid have higher IL-17/FoxP3 ratio than $CD4^+$ cells in peripheral blood supporting a favored Th17 phenotype in the lung mucosa (based on 11 BAL samples, 6 from $pneumotype_{SPT}$ and 5 from $pneumotype_{BPT}$, data represented as median[IQR], p-value based on Mann-Whitney U).

Figure S8


NYU_31
NYU_14
NYU_27
NYU_30
NYU_23
NYU_24
NYU_21
NYU_9
NYU_3
NYU_6
NYU_10
NYU_1
SMIM6
MRPL43
CDC20B
RPL37A
KIF13A
RPL38
EEF1D
APOBEC3B
NTS
POSTN(2)
POSTN(1)
LPL(2)
LPL(1)
OLR1
CD52
CYBB
PPARG
PTPRO
MS4A7(2)
TREM1
PCOLCE2
TLR8
GLDN
EVI2B
HP
March1
INHBA
MS4A7(1)
MMP10
SVIL
ARHGAP5
SF1
PTPRC
LYZ
LUC7L3
LOC100506013
TBRG1
PLEKHA1
TNKS2
EHF
GOLIM4
PIK3R1
EGFR
ISL1
SNX31
CDH11
SULT1E1
COL4A6
GPR155(2)
LEPR
ZFAND5
TSHZ2
GPR155(1)
SNAI2
PTPRG
RUNX1T1
FST
GLIPR1
WBSCR27
SUCNR1
TSLP
PHACTR2
3
2
1
0
-1
-2
-3
Pneumotype BPT
Pneumotype SPT

Supplemental Figure 8: Differences in transcriptome of bronchial epithelial cells of pneumotype$_{SPT}$ vs. pneumotype$_{BPT}$. Unsupervised hierarchical clustering analysis of mRNA obtained from bronchial epithelial cells and found to be significantly differently expressed between the two pneumotypes.

Figure S9

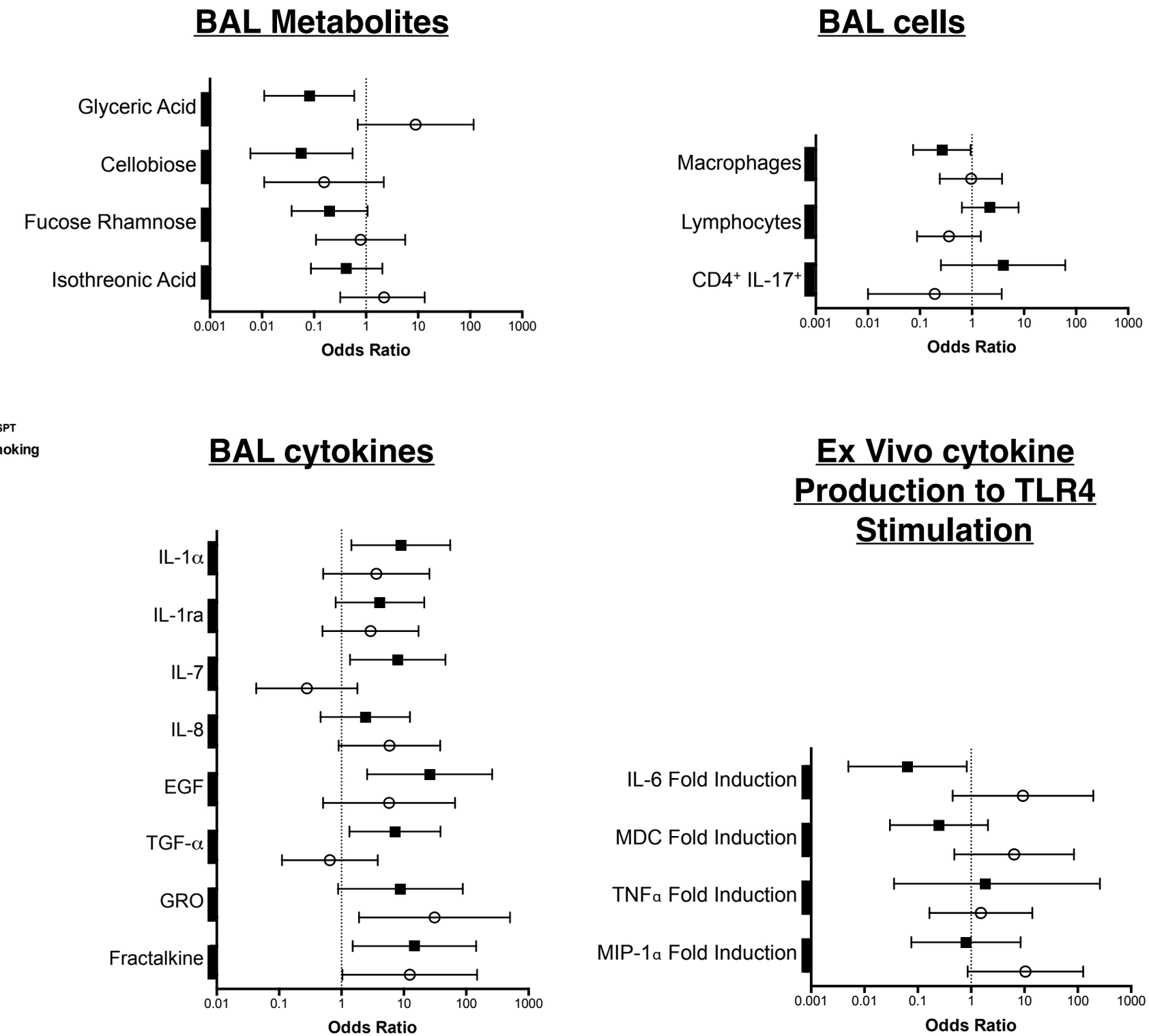


Up-regulated in Pneumotype$_{SPT}$

Up-regulated in Pneumotype$_{BPT}$

Supplemental Figure 9: Up-regulation of ERK1/2 and PIK3/AKT signaling pathway in bronchial epithelial cells associated with pneumotype$_{SPT}$. Ingenuity Pathway Analysis network generation algorithm identified ERK 1/2 signaling pathway as the top network (#1 IPA network, score=30, Focus Molecules= 14) and PI3K/Akt signaling pathway (#4 IPA network, score=20) from top discriminated mRNAs ($p<0.05$, FC>2) between pneumotype$_{SPT}$ and pneumotype$_{BPT}$. Solid lines denote positively correlated proteins while dotted lines denote negatively correlated proteins.

Figure S10


BAL Metabolites
Glyceric Acid
Cellobiose
Fucose Rhamnose
Isothreonic Acid
0.001
0.01
0.1
1
10
100
1000
Odds Ratio
BAL cells
Macrophages
Lymphocytes
CD4+ IL-17+
0.001
0.01
0.1
1
10
100
1000
Odds Ratio
Pneumotype SPT
History of Smoking
BAL cytokines
IL-1α
IL-1ra
IL-7
IL-8
EGF
TGF-α
GRO
Fractalkine
0.01
0.1
1
10
100
1000
Odds Ratio
Ex Vivo cytokine Production to TLR4 Stimulation
IL-6 Fold Induction
MDC Fold Induction
TNFα Fold Induction
MIP-1α Fold Induction
0.001
0.01
0.1
1
10
100
1000
Odds Ratio

Supplemental Figure 10: Multivariate logistic regression analysis to evaluate the contribution of smoking to differences noted between the two pneumotypes in BAL metabolites, cells, cytokines, and ex vivo cytokine production to TLR4 stimulation. For this analysis, levels of metabolites, cells, and cytokines (above or below the median) found to be significantly different between $pneumotype_{SPT}$ and $pneumotype_{BPT}$ were used as outcome and predictors in the model included $pneumotype_{SPT}$ and smoking status. Age and gender were also forced into the model. This analysis showed that the association between $pneumotype_{SPT}$ with lower levels of glyceric acid, cellobiose, macrophages, and with higher levels of IL-1α, IL-7, EGF, TGF-α, and Fractalkine was independent of smoking status. The association between $pneumotype_{SPT}$ with lower ex-vivo production of IL-6 was also independent of smoking status. Smoking was associated with increased levels of GRO and Fractalkine independently of pneumotype. The association of $pneumotype_{SPT}$ and the other biomarkers tested and represented in the figure were not statistically significant when adjusted by these covariates. Data presented as odds ratio±95%CI.

Supplemental Table 1: Metabolic pathways found to be differentially enriched between pneumotype$_{SPT}$ vs. pneumotype$_{BPT}$

	Relative Abundance		Effect size	FDR
	BPT	SPT		
Pathways enriched in pneumotype$_{SPT}$				
Carbohydrate Metabolism; Amino sugar and nucleotide sugar metabolism	1±0.048	1.246±0.131	0.624	4.95E-09
Carbohydrate Metabolism; Fructose and mannose metabolism	0.585±0.037	0.71±0.06	0.621	5.81E-09
Carbohydrate Metabolism; Galactose metabolism	0.373±0.036	0.515±0.081	0.576	8.81E-08
Nucleotide Metabolism; Pyrimidine metabolism	1.413±0.113	1.845±0.248	0.570	1.19E-07
Metabolism of Cofactors and Vitamins; One carbon pool by folate	0.494±0.039	0.634±0.082	0.559	2.25E-07
Glycan Biosynthesis and Metabolism; Peptidoglycan biosynthesis	0.66±0.057	0.834±0.099	0.549	3.72E-07
Amino Acid Metabolism; Amino acid related enzymes	1.306±0.071	1.508±0.114	0.541	5.63E-07
Metabolism of Terpenoids and Polyketides; Zeatin biosynthesis	0.037±0.005	0.058±0.013	0.529	1.06E-06
Nucleotide metabolism; Nucleotide metabolism_Unclassified	0.028±0.006	0.047±0.012	0.526	1.23E-06
Metabolism of Cofactors and Vitamins; Thiamine metabolism	0.383±0.015	0.438±0.035	0.525	1.33E-06
Nucleotide Metabolism; Purine metabolism	2.031±0.136	2.39±0.207	0.521	1.58E-06
Metabolism of Other Amino Acids; D-Glutamine and D-glutamate metabolism	0.128±0.007	0.155±0.017	0.519	1.74E-06
Energy Metabolism; Carbon fixation in photosynthetic organisms	0.513±0.025	0.6±0.056	0.513	2.32E-06
Glycan Biosynthesis and Metabolism; Glycosyltransferases	0.357±0.019	0.422±0.043	0.505	3.42E-06
Xenobiotics Biodegradation and Metabolism; Drug metabolism - other enzymes	0.223±0.023	0.295±0.048	0.495	5.73E-06
Amino Acid Metabolism; Cysteine and methionine metabolism	0.823±0.028	0.909±0.057	0.494	5.92E-06
Amino Acid Metabolism; Lysine biosynthesis	0.606±0.019	0.663±0.037	0.490	7.05E-06
Metabolism of Other Amino Acids; D-Alanine metabolism	0.095±0.013	0.121±0.013	0.482	1.05E-05
Metabolism of Terpenoids and Polyketides; Biosynthesis of ansamycins	0.065±0.005	0.079±0.01	0.469	1.91E-05
Enzyme Families; Peptidases	1.559±0.061	1.767±0.153	0.461	2.74E-05
Metabolism of Terpenoids and Polyketides; Terpenoid backbone biosynthesis	0.527±0.049	0.62±0.058	0.431	9.83E-05
Carbohydrate Metabolism; Starch and sucrose metabolism	0.601±0.045	0.693±0.065	0.409	0.00025
Carbohydrate Metabolism; Pentose phosphate pathway	0.706±0.025	0.754±0.032	0.403	0.00031
Glycan Biosynthesis and Metabolism; Lipopolysaccharide biosynthesis proteins	0.425±0.047	0.55±0.102	0.396	0.00042
Glycan Biosynthesis and Metabolism; Lipopolysaccharide biosynthesis	0.335±0.041	0.447±0.093	0.395	0.00044
Amino Acid Metabolism; Phenylalanine, tyrosine and tryptophan biosynthesis	0.714±0.026	0.772±0.045	0.389	0.00055
Glycan Biosynthesis and Metabolism; Glycosaminoglycan degradation	0.038±0.011	0.069±0.027	0.384	0.00068
Glycan Biosynthesis and Metabolism; Other glycan degradation	0.107±0.028	0.168±0.049	0.376	0.00094
Glycan Biosynthesis and Metabolism; Glycosphingolipid biosynthesis - globo series	0.042±0.01	0.075±0.029	0.370	0.00116
Glycan Biosynthesis and Metabolism; Glycosphingolipid biosynthesis - ganglio series	0.024±0.007	0.046±0.021	0.357	0.00194

Lipid Metabolism; Sphingolipid metabolism	0.087±0.015	0.121±0.03	0.348	0.00271
Metabolism of Cofactors and Vitamins; Folate biosynthesis	0.451±0.023	0.493±0.037	0.326	0.00597
Metabolism of Other Amino Acids; Selenocompound metabolism	0.364±0.014	0.386±0.019	0.312	0.00985
Pathways enriched in pneumotype$_{BPT}$				
Metabolism of Terpenoids and Polyketides; Carotenoid biosynthesis	0.06±0.012	0.025±0.013	0.677	1.32E-10
Xenobiotics Biodegradation and Metabolism; Metabolism of xenobiotics by cytochrome P450	0.203±0.022	0.102±0.05	0.643	1.39E-09
Xenobiotics Biodegradation and Metabolism; Drug metabolism - cytochrome P450	0.212±0.023	0.107±0.053	0.640	1.76E-09
Xenobiotics Biodegradation and Metabolism; Chlorocyclohexane and chlorobenzene degradation	0.098±0.016	0.045±0.025	0.617	7.69E-09
Xenobiotics Biodegradation and Metabolism; Fluorobenzoate degradation	0.056±0.01	0.025±0.015	0.615	8.93E-09
Metabolism of Other Amino Acids; Glutathione metabolism	0.482±0.029	0.359±0.066	0.611	1.09E-08
Amino Acid Metabolism; Valine, leucine and isoleucine degradation	0.846±0.092	0.513±0.17	0.609	1.28E-08
Amino Acid Metabolism; Phenylalanine metabolism	0.35±0.032	0.237±0.058	0.606	1.54E-08
Lipid Metabolism; Synthesis and degradation of ketone bodies	0.16±0.024	0.086±0.035	0.600	2.13E-08
Amino Acid Metabolism; Tryptophan metabolism	0.579±0.066	0.339±0.126	0.598	2.45E-08
Xenobiotics Biodegradation and Metabolism; Styrene degradation	0.094±0.015	0.051±0.021	0.594	3.11E-08
Xenobiotics Biodegradation and Metabolism; Chloroalkane and chloroalkene degradation	0.282±0.022	0.18±0.058	0.589	4.16E-08
Lipid Metabolism; Fatty acid metabolism	0.683±0.087	0.412±0.141	0.581	6.40E-08
Xenobiotics Biodegradation and Metabolism; Ethylbenzene degradation	0.08±0.007	0.057±0.012	0.581	6.61E-08
Lipid Metabolism; Ether lipid metabolism	0.022±0.004	0.011±0.006	0.580	6.82E-08
Amino Acid Metabolism; Lysine degradation	0.462±0.056	0.28±0.097	0.579	7.41E-08
Carbohydrate Metabolism; Propanoate metabolism	0.906±0.078	0.668±0.126	0.571	1.12E-07
Carbohydrate Metabolism; Butanoate metabolism	1.015±0.066	0.771±0.139	0.570	1.20E-07
Lipid Metabolism; alpha-Linolenic acid metabolism	0.036±0.006	0.019±0.009	0.563	1.80E-07
Enzyme Families; Protein kinases	0.41±0.035	0.292±0.067	0.560	2.05E-07
Carbohydrate Metabolism; Inositol phosphate metabolism	0.209±0.018	0.143±0.039	0.557	2.52E-07
Metabolism of Cofactors and Vitamins; Retinol metabolism	0.084±0.007	0.058±0.016	0.552	3.17E-07
Metabolism of Other Amino Acids; beta-Alanine metabolism	0.438±0.042	0.306±0.075	0.552	3.23E-07
Carbohydrate Metabolism; Glyoxylate and dicarboxylate metabolism	0.753±0.054	0.574±0.105	0.549	3.70E-07
Xenobiotics Biodegradation and Metabolism; Benzoate degradation	0.526±0.071	0.34±0.098	0.549	3.77E-07
Lipid Metabolism; Biosynthesis of unsaturated fatty acids	0.261±0.021	0.178±0.052	0.541	5.69E-07
Metabolism of Terpenoids and Polyketides; Geraniol degradation	0.319±0.051	0.172±0.083	0.539	6.43E-07
Xenobiotics Biodegradation and Metabolism; Aminobenzoate degradation	0.414±0.053	0.265±0.084	0.538	6.80E-07
Metabolism of Terpenoids and Polyketides; Limonene and pinene degradation	0.348±0.049	0.204±0.083	0.534	8.05E-07
Biosynthesis of Other Secondary Metabolites; beta-Lactam resistance	0.045±0.005	0.031±0.008	0.531	9.40E-07
Amino Acid Metabolism; Tyrosine metabolism	0.468±0.023	0.39±0.048	0.531	9.47E-07

Xenobiotics Biodegradation and Metabolism; Caprolactam degradation	0.22±0.037	0.114±0.062	0.530	1.01E-06
Energy Metabolism; Oxidative phosphorylation	1.445±0.061	1.282±0.093	0.524	1.34E-06
Xenobiotics Biodegradation and Metabolism; Naphthalene degradation	0.242±0.021	0.177±0.04	0.516	2.02E-06
Biosynthesis of Other Secondary Metabolites; Penicillin and cephalosporin biosynthesis	0.064±0.008	0.043±0.012	0.493	6.17E-06
Lipid Metabolism; Lipid biosynthesis proteins	0.781±0.032	0.697±0.054	0.483	9.75E-06
Xenobiotics Biodegradation and Metabolism; Toluene degradation	0.19±0.016	0.159±0.016	0.477	1.31E-05
Amino Acid Metabolism; Arginine and proline metabolism	1.17±0.063	1.025±0.095	0.454	3.68E-05
Carbohydrate Metabolism; Pyruvate metabolism	1.113±0.026	1.018±0.073	0.449	4.49E-05
Xenobiotics Biodegradation and Metabolism; Nitrotoluene degradation	0.086±0.011	0.063±0.013	0.447	4.89E-05
Carbohydrate Metabolism; Ascorbate and aldarate metabolism	0.178±0.01	0.149±0.022	0.437	7.61E-05
Amino Acid Metabolism; Histidine metabolism	0.581±0.028	0.517±0.045	0.425	0.00013
Xenobiotics Biodegradation and Metabolism; Atrazine degradation	0.061±0.012	0.033±0.02	0.422	0.00015
Xenobiotics Biodegradation and Metabolism; Bisphenol degradation	0.123±0.015	0.085±0.031	0.404	0.00031
Metabolism of Cofactors and Vitamins; Lipoic acid metabolism	0.075±0.007	0.056±0.015	0.395	0.00044
Lipid Metabolism; Linoleic acid metabolism	0.074±0.01	0.053±0.016	0.387	0.00061
Others; Others_Unclassified	1.041±0.04	0.953±0.071	0.377	0.00090
Biosynthesis of Other Secondary Metabolites; Phenylpropanoid biosynthesis	0.1±0.015	0.078±0.013	0.377	0.00090
Amino acid metabolism; Amino acid metabolism_Unclassified	0.217±0.012	0.194±0.018	0.373	0.00105
Metabolism of Terpenoids and Polyketides; Tetracycline biosynthesis	0.146±0.012	0.118±0.026	0.344	0.00309
Lipid Metabolism; Fatty acid biosynthesis	0.551±0.016	0.516±0.032	0.340	0.00365
Lipid Metabolism; Steroid biosynthesis	0.022±0.005	0.012±0.009	0.340	0.00367
Metabolism of cofactors and vitamins; Metabolism of cofactors and vitamins_Unclassified	0.189±0.016	0.163±0.021	0.331	0.00493
Biosynthesis and biodegradation of secondary metabolites; Biosynthesis and biodegradation of secondary metabolites	0.075±0.009	0.057±0.017	0.322	0.00706
Biosynthesis of Other Secondary Metabolites; Indole alkaloid biosynthesis	0.002±0.001	0.001±0.001	0.314	0.00921
Energy Metabolism; Photosynthesis - antenna proteins	0.028±0.016	0.01±0.009	0.306	0.01230
Energy Metabolism; Sulfur metabolism	0.324±0.024	0.29±0.028	0.296	0.01737
Metabolism of Terpenoids and Polyketides; Biosynthesis of vancomycin group antibiotics	0.051±0.004	0.059±0.008	0.294	0.01866
Amino Acid Metabolism; Glycine, serine and threonine metabolism	0.935±0.026	0.899±0.033	0.274	0.03739

Data presented as Mean±SD Relative abundance (% of total reads annotated to metabolism based on PICRUSt); n=49
FDR = false discovery rate

Supplemental Table 2: Metabolites in BALF

	Pneumotype$_{BPT}$	**Pneumotype$_{SPT}$**	p-value
Glyceric Acid	212680[182960-451002]	150105[70339-175507]	*0.007*
Glycerol Alpha Phosphate	151720[62243-210562]	75165[32864-109428]	*0.016*
Cellobiose	111036[57497-163914]	52584[32471-102231]	*0.026*
Isothreonic Acid	22376[11126-40002]	13431[8516-18725]	*0.037*
Fucose Rhamnose	161467[130191-277157]	113822[67636-163330]	0.051
Erythritol	47258[17249-120023]	26961[12076-42609]	0.057
Threitol	16781[6138-42439]	8312[3641-14970]	0.086
Cysteine	51372[9524-155553]	14965[4575-56604]	0.095
Cholesterol	13742965[2205911-22856976]	6842542[2223233-12216410]	0.114
Oxalic Acid	672362[92496-2408361]	288001[3259-620238]	0.114
Erythronic Acid Lactone	148039[17929-271405]	67230[1347-105879]	0.114
Proline	132421[86471-406760]	190107[156948-278999]	0.126
Benzoic Acid	487173[134230-1132627]	311850[64597-522344]	0.137
Gluconic Acid	12747[8247-24994]	10364[7813-14400]	0.150
Tocopherol Alpha	97401[17840-224454]	50645[13034-101682]	0.164
Beta Alanine	10220[5892-17274]	5441[3910-14206]	0.194
Azelaic Acid	35073[4519-91690]	9490[4321-34581]	0.194
Fructose	35026[19597-51108]	26656[15364-39141]	0.210
Malic Acid	16280[11478-40053]	16267[9919-20798]	0.227
Threonic Acid	15864[6621-72716]	9879[7204-16256]	0.227
2-Hydroxybutanoic Acid	127018[70967-246284]	93151[66003-129266]	0.227
Ribose	12166[7916-69686]	23534[12015-42419]	0.246
Uridine	18238[9733-56162]	27388[19680-71487]	0.246
Ribitol	65569[14631-104406]	26263[11142-80501]	0.246
Pseudo Uridine	36712[16070-86369]	23760[13901-40510]	0.246
Hydroxylamine	444094[63043-1089163]	317357[11729-490341]	0.265
Pelargonic Acid	266077[97265-715234]	125932[87552-277154]	0.265
Indole-3-Lactate	31799[5498-83398]	12957[6431-35434]	0.265
Behenic Acid	25095[12941-54184]	19376[13028-25954]	0.265
Nicotinamide	104807[21909-200725]	61636[19791-112554]	0.286
Glutaric Acid	28546[6367-81993]	15216[5705-36561]	0.286
Capric Acid	121363[19036-275183]	51501[8926-163267]	0.286
1-Monostearin	36782[13927-72079]	26636[15006-41456]	0.286
Beta Sitosterol	17019[14617-41574]	17831[9841-25089]	0.307
Glycerol	637219[424517-1272975]	982170[580988-1691424]	0.330
Glycolic Acid	62701[46700-92040]	52102[35202-71640]	0.330
Succinic Acid	76907[24596-168372]	50545[30887-71237]	0.403
Oxoproline	465225[111364-802549]	200449[141322-546374]	0.430
1-Monopalmitin	58897[14215-121755]	45555[9998-66993]	0.458
Lauric Acid	399741[105863-1010952]	274397[79921-633459]	0.458
Aspartic Acid	61624[31442-98186]	41974[19268-67072]	0.486

Fumaric Acid	44459[8538-77739]	32121[6966-55250]	0.486
Glucose	4353984[1133389-21323749]	3817115[870729-7731547]	0.486
Palmitic Acid	3297592[521203-8147453]	3319768[265295-5071008]	0.486
Stearic Acid	15312207[2544560-32484628]	13329871[1502088-17975676]	0.486
Dihydroabietic Acid	125522[14194-445874]	160779[9875-232057]	0.486
Serine	45290[28366-90497]	43646[26502-68958]	0.516
Thymine	14485[4498-35415]	15014[2581-29882]	0.516
1,5-Anhydroglucitol	381599[228945-1398827]	374671[229249-593452]	0.516
Octadecanol	40500[13534-80378]	31824[7117-62647]	0.516
Threonine	131247[51557-217972]	85735[40917-150272]	0.546
Propane-1,3-Diol NIST	70644[31882-229295]	58737[34908-124804]	0.546
Phenylethylamine	13052[6293-142769]	10998[5538-26479]	0.546
Leucine	199672[121346-375419]	228400[161198-351023]	0.577
4-Hydroxybenzoate	2797951[184084-20025680]	4229243[1024802-9623629]	0.577
Citric Acid	43252935[18383937-68717502]	37422499[18335754-56317964]	0.577
Urea	9124327[7371423-22200543]	9728264[6651056-15557519]	0.610
Sucrose	8922[4486-21715]	7148[3977-15058]	0.610
Maltose	40794[28466-67371]	48400[27987-85786]	0.610
Pentadecanoic Acid	328709[89109-786462]	367548[55421-495714]	0.610
Glycine	1143022[154321-2303517]	964216[99069-1668517]	0.642
Alanine	229697[145034-579922]	212302[195955-273800]	0.642
Lactic Acid	491789[217544-803442]	344632[214217-697484]	0.642
Dodecanol	87778[12597-352366]	66973[12058-177847]	0.676
Glutamine	245804[18012-661134]	142515[13039-378225]	0.710
Arachidonic Acid	154377[16862-494916]	193590[28290-411165]	0.745
Inosine	469025[101943-1443518]	325835[141305-877117]	0.745
Guanosine	29224[10260-107891]	27134[18158-61130]	0.745
Isoleucine	105582[63084-161202]	98659[79154-194451]	0.745
Glutamic Acid	117470[53561-191841]	82905[61354-189085]	0.781
Creatinine	130430[36048-368282]	79835[47795-248165]	0.781
Valine	246249[154347-450670]	262706[219969-357951]	0.816
Salicylaldehyde	43136[16583-162771]	46311[16349-131921]	0.816
Myristic Acid	308199[160572-574972]	360195[181571-453475]	0.853
2-Ketoisocaproic Acid	26995[10804-41559]	25987[16070-42757]	0.889
Methionine	18169[11601-35797]	15382[12878-26057]	0.926
Taurine	74471[31203-376954]	95937[39545-294675]	0.926
Linoleic Acid	45203[6696-109302]	47990[11604-90814]	0.926
Palmitoleic Acid	126514[22753-279426]	112400[15481-275263]	0.926
Oleic Acid	113128[23518-302528]	165242[16132-296372]	0.963
Mannose	81838[31145-188229]	73709[55204-109288]	1.000
Ethanolamine	1422425[643823-3822229]	1397693[991858-2925678]	1.000
Aconitic Acid Delta	339556[167099-862089]	479463[125861-1000540]	1.000

Data presented as Median (IQR); n=28

Supplemental Table 3: Correlation of UniFrac distance to upper airway with levels of metabolites in BAL

	Spearman's Rho	p-value
BAL Metabolytes		
Cellobiose	0.500	***0.007***
Fucose Rhamnose	0.473	***0.011***
Isothreonic Acid	0.446	***0.017***
Glyceric Acid	0.423	***0.025***
Arachidonic Acid	-0.472	***0.011***
Threonic Acid	0.381	***0.045***
Alanine	0.372	0.051
Oleic Acid	-0.369	0.053
4-Hydroxybenzoate	-0.330	0.086
Linoleic Acid	-0.322	0.095
Myristic Acid	-0.303	0.117
Palmitoleic Acid	-0.288	0.137
Citric Acid	0.282	0.146
Lactic Acid	0.280	0.149
Uridine	-0.275	0.156
Guanosine	-0.263	0.177
Pentadecanoic Acid	-0.263	0.177
Ethanolamine	0.242	0.214
Palmitic Acid	-0.239	0.221
Dihydroabietic Acid	-0.235	0.228
Dodecanol	-0.231	0.237
Stearic Acid	-0.230	0.239
Glycolic Acid	0.215	0.272
Gluconic Acid	0.211	0.282
Malic Acid	0.210	0.284
Sucrose	-0.204	0.297
Octadecanol	-0.201	0.304
Valine	0.200	0.307
Glutamine	-0.192	0.327
Glutamic Acid	0.191	0.330
1-Monopalmitin	-0.190	0.332
Cysteine	0.190	0.333
2-Hydroxybutanoic Acid	0.167	0.394
Erythritol	0.160	0.415
Fructose	0.151	0.445
Succinic Acid	-0.142	0.470
Glycerol Alpha Phosphate	0.142	0.472
Glycerol	-0.141	0.475
Threitol	0.135	0.495
1-Monostearin	-0.134	0.498

Lauric Acid	-0.133	0.500
Inosine	-0.126	0.521
Fumaric Acid	-0.120	0.543
Propane-1,3-Diol NIST	-0.118	0.549
Behenic Acid	-0.117	0.555
Hydroxylamine	-0.116	0.557
Indole-3-Lactate	-0.115	0.558
Isoleucine	0.114	0.562
Thymine	-0.113	0.568
Leucine	0.111	0.575
Oxoproline	0.106	0.591
Proline	-0.103	0.600
Glutaric Acid	-0.099	0.616
Aconitic Acid Delta	0.088	0.656
Oxalic Acid	0.088	0.656
Creatinine	-0.088	0.658
Beta Sitosterol	0.087	0.660
Salicylaldehyde	-0.083	0.676
Aspartic Acid	0.078	0.692
Pelargonic Acid	0.077	0.696
Serine	0.076	0.700
Nicotinamide	0.074	0.707
Cholesterol	-0.071	0.721
Beta Alanine	0.069	0.727
Maltose	0.061	0.757
Mannose	0.059	0.765
Methionine	-0.056	0.778
Glycine	-0.048	0.808
Taurine	-0.047	0.812
Erythronic Acid Lactone	0.046	0.816
Benzoic Acid	0.043	0.827
2-Ketoisocaproic Acid	0.042	0.834
Azelaic Acid	-0.041	0.836
Tocopherol Alpha	-0.039	0.844
Ribose	-0.037	0.853
Pseudo Uridine	0.029	0.883
Ribitol	-0.017	0.932
Glucose	-0.015	0.941
Phenylethylamine	-0.014	0.945
Capric Acid	-0.011	0.956
Threonine	-0.009	0.965
Urea	-0.007	0.971
1,5-Anhydroglucitol	-0.005	0.978

N=28

Supplemental Table 4: BAL cells and cytokines

	Pneumotype$_{BPT}$	Pneumotype$_{SPT}$	*p-value*
BAL Cell Differential (%)*			
Macrophages	92.8 [90.0-95.5]	90.6 [86.0-92.3]	***0.07***
Lymphocytes	4.5 [3.2-6.7]	7.4 [4.6-11.9]	***0.035***
Neutrophils	1.3 [1.0-2.0]	1.8 [0.9-2.6]	*0.365*
Eosinophils	0.1 [0.0-0.4]	0.0 [0.0-0.4]	*0.32*
Cytokines (pg/mL)**			
EGF	3.22[2.87-5.40]	16.04[10.60-56.64]	***0.006***
IL-1α	29.44[16.72-36.24]	48.3[28.74-90.3]	***0.007***
Fractalkine	32.70[11.43-50.87]	63.46[41.03-99.1]	***0.008***
IL-8	317.74[176.06-577.84]	725.93[422.54-1395.93]	***0.009***
TGF-α	15.62[4.00-21.72]	30.07[19.86-39.65]	***0.009***
IL-1ra	17.18[6.92-49.28]	91.65[21.73-134.52]	***0.023***
GRO	3486.09[2671.76-5205.75]	13101.47[3369.98-38653.56]	***0.027***
IL-7	10.43[5.79-11.98]	13.20[9.86-18.96]	***0.042***
IFN-γ	0.78[0.78-1.17]	0.78[0.78-0.78]	0.087
IL-1β	2.94[2.50-3.36]	3.42[2.63-4.98]	0.088
Eotaxin	13.8[10.35-19.88]	19.26[13.18-30.82]	0.092
IP-10	3113.67[1678.12-11942.6]	8849.09[4635.97-14606.33]	0.101
Fit-3 Ligand	29.48[13.91-66.25]	60.94[20.74-138.24]	0.111
IL-6	23.97[14.47-30.21]	30.24[23.72-48.84]	0.127
sCD40L	35.26[4.08-62.4]	57.92[28.33-122.3]	0.15
G-CSF	807.02[481.61-1338.22]	1074.08[676.43-2348.09]	0.17
FGF-2	7.10[6.51-9.72]	8.42[6.94-11.14]	0.199
IL-17	1.51[1.44-2.03]	1.59[1.48-2.54]	0.207
MIP-1α	16.82[12.62-42.69]	28.79[20.18-45.69]	0.215
GM-CSF	14.74[13.19-22.64]	14.07[10.43-16.48]	0.25
sIL-2Rα	7.56[5.23-11.01]	6.09[4.53-8.30]	0.425
VEGF	1033.33[205.65-1638.29]	885.63[500.64-3034.01]	0.425
IL-5	4.33[1.64-11.72]	2.19[1.36-7.44]	0.438
MCP-3	17.93[4.80-42.44]	15.03[8.36-22.66]	0.491
MIP-1β	33.31[22.72-44.33]	38.06[27.57-91.87]	0.535
TNF-α	6.54[5.06-12.23]	9.93[5.06-18.12]	0.58
MDC	53.75[35.52-116.3]	45.01[29.26-105.98]	0.658
IL-12 p70	4.54[3.59-5.29]	4.38[3.66-6.11]	0.74
MCP-1	2531.3[1511.65-5130.14]	2500.64[1569.05-3936.27]	0.757
IL-15	62.91[44.54-79.33]	58.43[30.87-83.25]	0.79
IL-12 p40	5.40[3.86-8.10]	5.85[3.78-7.65]	0.877

Data presented as Median (IQR)
*Available for the NYU and LHMP cohorts (n=45)
**Available for the NYU cohort (n=31)

Supplemental Table 5: Correlation of UniFrac distance to upper airway with levels of cells and cytokines in BAL

	Spearman's Rho	p-value
BAL Cells*		
Macrophages	0.416	***0.005***
Lymphocytes	-0.404	***0.007***
Neutrophils	-0.232	0.130
Eosinophils	0.070	0.654
BAL Cytokines **		
Fractalkine	-0.552	***0.002***
IL-8	-0.550	***0.002***
GRO	-0.536	***0.003***
EGF	-0.513	***0.004***
IL-1ra	-0.514	***0.004***
MIP-1α	-0.457	***0.013***
IL-1α	-0.451	***0.014***
TGF-α	-0.448	***0.015***
Eotaxin	-0.431	***0.020***
IL-7	-0.390	***0.036***
G-CSF	-0.383	***0.040***
IL-17	-0.373	***0.047***
FGF-2	-0.368	***0.049***
IL-1β	-0.367	0.050
IL-6	-0.360	0.055
sCD40L	-0.316	0.095
IP-10	-0.311	0.100
VEGF	-0.297	0.118
MIP-1β	-0.291	0.125
MDC	0.285	0.134
Fit-3 Ligand	-0.260	0.173
GM-CSF	0.217	0.259
IL-12 p70	-0.203	0.291
IL-12 p40	-0.133	0.490
TNF-α	-0.100	0.607
IL-5	-0.091	0.640
MCP-3	0.076	0.694
IL-15	0.050	0.796
IFN-γ	0.046	0.813
sIL-2Rα	0.041	0.833
MCP-1	-0.031	0.875

*Available for the NYU and LHMP cohorts (n=45)

**Available for the NYU cohort (n=31)

Supplemental Table 6: Ex vivo cytokine production in response to TLR4 stimulation

	Pneumotype$_{BPT}$	**Pneumotype$_{SPT}$**	*p-value*
IL-6	36.88[25.36-181.59]	2.91[1.18-13.26]	***0.0010***
MDC	1.95[1.51-3.95]	1.22[0.99-1.69]	***0.0160***
TNF-α	16.24[5.6-19.95]	2.43[1.04-7.7]	***0.0160***
MIP-1α	2.24[0.91-10.60]	0.92[0.67-1.15]	***0.0430***
G-CSF	945.87[103.92-1929.23]	13.5[1.58-305.85]	0.0550
MIP-1β	6.81[2.77-20.42]	2.44[1.1-4.56]	0.0550
GM-CSF	153.51[53.73-533.57]	8.92[1.94-98.33]	0.0680
IL-10	173.14[39.62-321.38]	20.19[1.34-49.61]	0.0830
GRO	1.94[1.13-6.63]	1.2[0.97-3.15]	0.1010
Fractalkine	1.64[1.04-2.25]	1.15[1.01-1.36]	0.1730
Flt-3Ligand	1.49[0.92-1.73]	1.13[0.67-1.3]	0.2030
MCP-1	1.35[1-4.92]	1.09[0.59-2.2]	0.2740
MCP-3	0.67[0.57-1.03]	0.87[0.64-1.75]	0.2740
IL1-ra	1.32[0.95-2.09]	0.99[0.88-1.08]	0.3150
TGF-α	1[1-5.06]	1.64[1-2.62]	0.3600
IL-1α	18.89[1.44-48.34]	3.4[1.09-26]	0.4080
IL-8	1.03[0.29-1.16]	0.62[0.27-1.22]	0.4080
FGF-2	1[1-1.18]	1[1-1.04]	0.4600
IL-12-p40	62[1-337.51]	7.36[1-172]	0.4600
IL-1β	296.22[1-706.91]	8.8[3.12-70.14]	0.4600
IL-9	1.01[1-1.04]	1.01[0.99-1.07]	0.5730
VEGF	1[1-1]	1[1-1.02]	0.6330
EGF	1[1-1.76]	1[1-1.01]	0.6960
Eotaxin	1.01[0.99-1.13]	1[0.99-1.01]	0.6960
IL-17	1[1-1.16]	1[1-1.04]	0.6960
IL-2	1.09[1.05-1.14]	1.07[0.99-1.7]	0.6960
IP-10	0.91[0.52-2.06]	1.31[0.51-2.29]	0.6960
sCD40L	1[1-1]	1[0.85-1.66]	0.6960
siL2-Rα	2.75[1.11-6.69]	1.24[1.06-2.66]	0.6960
IFNα	1[1-1.38]	1[1-1.02]	0.7620
IL-13	1.39[1.05-1.45]	1.15[0.99-5.05]	0.7620
IL-7	1[1-1.72]	1[1-1.01]	0.7620
IL-4	1[1-1]	1[1-1]	0.8290
IFNγ	1.46[1.26-2.03]	1.3[1.14-5.31]	0.9650
IL-12-p70	1.16[1.01-1.74]	1.08[1.02-1.29]	0.9650
IL-15	1.03[1-1.51]	1.05[0.98-1.12]	1.0000
IL-3	1[1-1.13]	1[1-1.02]	1.0000
TNF-β	1.25[0.97-5.13]	1.15[1.04-3.91]	1.0000

Data presented as Median (IQR) fold induction; (n=31)

Supplemental Table 7: Correlation of UniFrac distance to upper airway with ex-vivo cytokine production during TLR4 stimulation of alveolar macrophages

	Spearman's Rho	p-value
Cytokine fold induction		
MDC	0.566	**0.014**
IL-6	0.551	**0.018**
GM	0.470	**0.049**
MCP	0.425	0.079
IL-3	-0.414	0.088
IL-13	-0.346	0.160
Flt	0.338	0.170
IL-1β	0.333	0.177
IL-9	-0.318	0.199
IFNγ	-0.315	0.203
IL-2	-0.313	0.206
IL-10	0.310	0.211
IL-17	-0.310	0.211
EGF	0.309	0.213
Eotaxin	-0.307	0.215
IP	-0.288	0.247
IL-4	-0.276	0.268
FGF	-0.267	0.284
sCD40L	-0.265	0.289
IFNα	0.264	0.290
TNF	-0.260	0.297
Fractalkine	0.252	0.314
IL-7	0.225	0.369
IL-12-p70	-0.218	0.385
G	-0.217	0.404
IL-12-p40	0.195	0.438
VEGF	0.193	0.442
IL-15	0.190	0.449
IL-8	-0.130	0.608
IL1-ra	0.116	0.647
siL2	-0.109	0.667
MIP-1α	0.091	0.721
IL-1α	-0.085	0.739
MIP-1β	0.063	0.803
TGF	0.056	0.824
TNF	-0.049	0.846
GRO	0.042	0.867
MCP	0.005	0.985

n=31