E-Cigarette Use, Cigarette Use, and Sex Modify the Nasal Microbiome and Nasal Host-Microbiota Interactions

Elise Hickman
UNC Chapel Hill
Andrew Hinton
UNC Chapel Hill
Bryan Zorn
UNC Chapel Hill
Meghan Rebuli
UNC Chapel Hill https://orcid.org/0000-0003-1918-2257
Carole Robinette
UNC Chapel Hill
Matthew Wolfgang
University of North Carolina https://orcid.org/0000-0003-4534-6470
Peter Mucha
Dartmouth University
Ilona Jaspers (ilona_jaspers@med.unc.edu)
UNC Chapel Hill

Article

Keywords: E-Cigarettes, nasal microbiome, rRNA

Posted Date: July 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-725763/v1

Title: E-Cigarette Use, Cigarette Use, and Sex Modify the Nasal Microbiome and Nasal Host-Microbiota Interactions

Authors: Elise Hickman[1,2]*, Andrew Hinton[3,4,]*, Bryan Zorn[5], Meghan E. Rebuli[1,2], Carole Robinette[1], Matthew Wolfgang[5, 6], Peter J. Mucha[3,7], Ilona Jaspers[1,2#]
*These authors contributed equally to the published work.
[#] Corresponding author

Affiliations:
[1] Center for Environmental Medicine, Asthma, and Lung Biology
[2] Curriculum in Toxicology & Environmental Medicine
[3] Curriculum in Bioinformatics and Computational Biology
[4] UNC Food Allergy Initiative, School of Medicine
[5] Marsico Lung Institute
[6] Department of Microbiology and Immunology
[7] Department of Mathematics and Department of Applied Physical Sciences
University of North Carolina, Chapel Hill, NC, United States

Abstract (150-word limit)

E-cigarettes are often perceived as safer than cigarettes, but previous research suggests that e-cigarettes can alter respiratory innate immune function. The respiratory microbiome plays a key role in respiratory host defense, but the effect of e-cigarettes on the respiratory microbiome has not been studied. Using 16S rRNA gene sequencing on nasal epithelial lining fluid samples from adult e-cigarette users, smokers, and nonsmokers, followed by novel computational analysis of pairwise log ratios, we determined that e-cigarette use and smoking causes differential respiratory microbiome dysbiosis, which was further affected by sex. We also collected nasal lavage fluid for analysis of immune mediators associated with host-microbiota interactions. Our analysis identified disruption of the relationships between host-microbiota mediators in the nose of e-cigarette users and smokers, which is indicative of disrupted respiratory mucosal immune responses. Our approach provides a novel platform that robustly identifies host immune dysfunction caused by e-cigarette use or smoking.

Introduction

Approximately 7 million adults and more than 3.5 million youth are current electronic cigarette (e-cigarette) users.[1-3] E-cigarettes heat and aerosolize e-liquids containing nicotine and flavorings dissolved in humectants propylene glycol and glycerin. E-cigarette use has been steadily increasing over the past decade, especially among teenagers and young adults, reversing the previous decline in youth tobacco use. [3,4] Public health crises, such as the outbreak of e-cigarette and vaping-associated lung injury in 2019-2020 and the ongoing SARS-CoV-2 global pandemic, highlight the importance of research examining the effects of e-cigarettes on respiratory immune function.[5,6]

There is emerging evidence that e-cigarettes disrupt respiratory innate immunity. Previous work has demonstrated the potential for e-cigarette toxicity and impairment of respiratory immune defense using *in vitro* and *in vivo* models as well as in samples from human

subjects.[7-12] For example, e-cigarette users have altered markers of innate immune responses in induced sputum and bronchoalveolar lavage fluid in comparison with smokers and nonsmokers,[8,12] and chronic e-cigarette exposure in mice can dysregulate endogenous lung lipid homeostasis and innate immunity.[11,13] *In vitro* studies have demonstrated that e-liquids, e-cigarette aerosols, and their components can impair the function of ciliated airway cells and respiratory immune cells.[9,14-18] Furthermore, e-cigarette exposure has been shown to enhance bacterial virulence and adhesion to airway cells,[19,20] suggesting that e-cigarette exposure may impact the respiratory microbiome. However, the effects of e-cigarette use on the respiratory microbiome in humans have not been evaluated.

The respiratory microbiome includes distinct communities of microbiota along the length of the respiratory tract.[21] Similar to microbial communities at other body sites, respiratory microbiota interface with the host immune system, and dysbiosis of the respiratory tract microbiome has been associated with diseases, including cystic fibrosis, chronic obstructive pulmonary disease, asthma, and chronic rhinosinusitis, as well as with disease exacerbations and smoking cigarettes.[21-24] Sampling the nasal microbiome is straightforward in contrast to the lower airway microbiome, which is easily contaminated with oral microbiota during specimen collection.[25] In addition, the nose is an important gatekeeper in the respiratory tract, as potential pathogens must often colonize this region before progressing to the lower respiratory tract.[21] This role has become even more clear and relevant with the emergence of SARS-CoV-2, with recent studies showing associations between the nasal microbiome and SARS-CoV-2 infection. [26,27] Of note is that dysbiosis of the nasal microbiome specifically has been associated with smoking cigarettes,[23] and gene expression and histopathological changes due to smoking are similar in the nasal and lower airway epithelium,[7] supporting the use of the nasal microbiome for studying the effects of environmental exposures on the respiratory microbiome.

Mechanistic study of the human microbiota is an important focus when studying the human microbiome, where identifying microbes associated with disease is paramount.[28] To uncover complex interactions in microbiome association studies changes to classical statistical methods are required.[29] In addition, computational methods that robustly integrate disparate data types with 16S microbiome data for association testing have been limited.[30] In particular, microbiome datasets have interspecies interactions, small sample sizes, high dimensionality (where the number of features greatly exceed the number of samples), are sparse (where the data matrix contains many zeroes), and when converted to relative abundance are compositional, meaning the total number of reads is not informative.[31] Combined, these challenges significantly confound the multivariate integrative analysis required to improve our understanding of host-microbiome interactions. Thus, novel analytical tools are necessary to uncover true signals hidden within small sample size microbiome data.

In this study, we sampled the nasal microbiomes of smokers, nonsmokers, and e-cigarette users using a non-invasive absorptive strip to collect nasal epithelial lining fluid. We then used high-throughput sequencing of the bacterial 16S rRNA gene from the strips to identify bacteria present and analyze the bacterial composition of the nasal microbiome in our subjects. Because these microbial communities are composed of highly interdependent taxa that have complex interaction patterns, multivariate data analysis is critical to extract biologically relevant information.

Here, we leverage Selection Energy Permutation [32], a novel multivariate association test that simultaneously tests associations while identifying robust subsets of pairwise log ratios in the setting of high-dimensional, low sample size data. These reduced subsets are then used to

integratively analyze nasal microbiome and matched cell-free nasal lavage fluid mediator data to determine: 1) whether there were significant compositional differences in the nasal microbiomes of E-cigarette users, smokers, and nonsmokers, 2) whether levels of nasal lavage fluid (NLF) mediators are significantly different in e-cigarette users and smokers in comparison with nonsmokers, and 3) whether changes in levels of these mediators correlate with nasal microbiome dysbiosis. Our data demonstrate nasal microbiome dysbiosis and unique networks of host-microbiota mediators in e-cigarette users and smokers in comparison with nonsmokers. This is indicative of disrupted respiratory mucosal immune responses in these groups and potentially increased susceptibility to infection by specific bacterial taxa. We also observed significant sex differences in the nasal microbiome, highlighting the importance of including sex as a biological variable in nasal microbiome studies.

Methods

Subject recruitment. Nasal epithelial lining fluid (NELF) strips, nasal lavage fluid (NLF), and venous blood were obtained from healthy adult human e-cigarette users, smokers, and nonsmokers as described previously (**Table 1**),[33] forming our exposure groups. Inclusion criteria were healthy adults age 18-50 years who are either nonsmokers not routinely exposed to environmental tobacco smoke, active regular cigarette smokers, or active e-cigarette users. Active cigarette smoking and e-cigarette use were determined as described previously.[7] Exclusion criteria were current symptoms of allergic rhinitis (deferred until symptoms resolve), asthma, FEV_1 less than 75% predicted at screen, bleeding disorders, recent nasal surgery, immunodeficiency, current pregnancy, chronic obstructive pulmonary disease, cardiac disease, or any chronic cardiorespiratory condition. After the consent process was completed, a medical history and substance use questionnaire was obtained, and subjects were issued a diary to document smoking/vaping for up to 4 weeks, after which they returned for sample collection. E-cigarette users averaged less than 1.5 cigarettes/day in their smoking/vaping diaries, while cigarette users ranged from 4.93-20 cigarettes per day in their diaries. To compare demographic characteristics between subjects in the different exposure groups, age, BMI, and serum cotinine levels were tested for normality using the Shapiro-Wilk test, and groups were compared using the Kruskal-Wallis test followed by the Steel-Dwass method for non-parametric multiple comparisons (analogous to a one-way ANOVA with Tukey's HSD for parametric data).

Serum Cotinine Measurement. Venous blood was collected in BD Vacutainer serum-separating tubes (Fisher Scientific, Waltham, MA) and allowed to clot for a minimum of 15 minutes at room temperature. The blood was then centrifuged at 1200 x g for 10 minutes, and the serum layer was transferred to a fresh tube and stored at -80°C until samples were collected from all subjects. Serum was assayed for cotinine, a metabolite of nicotine that can be measured as a biomarker of nicotine consumption, using a commercially available ELISA kit (Calbiotech, Mannheim, Germany) per manufacturer's instructions. Absorbance was read on a CLARIOstar plate reader (BMG Labtech, Ortenberg, Germany). The limit of quantification for serum cotinine was 5 ng/mL. For samples below the limit of detection, a value of zero was assigned. Serum was not available for one subject in the cohort.

NELF Strip Metagenomic Sequencing. DNA was extracted from whole NELF strips using Powersoil DNA Isolation Kit (MoBio Laboratories). Sequencing libraries were prepared as

previously described.[34] Samples were sequenced on an Illumina MiSeq kit version V3 2x300 paired end over the V3-V4 bacterial 16s gene. Raw sequencing data were demultiplexed and processed to generate a table of operational taxonomic units (OTUs). Specific primer schema, qPCR data, and the OTU table (having at least 10 sequences per OTU across all samples) are provided in the supplement. Raw sequence data have been uploaded under the BioProject accession number PRJNA746950 within the Sequence Read Archive.

NLF Processing and Soluble Mediator Measurement. Cell-free nasal lavage fluid was obtained via processing of raw nasal lavage fluid as described previously.[35] Briefly, raw nasal lavage fluid from each nostril was pooled and centrifuged at 500x g through a 40 μm strainer for 10 minutes. Supernatant (cell-free NLF) was collected and stored at -80°C until samples were collected from all subjects. Cell-free NLF was assayed for mediators of host-microbiota interaction (neutrophil elastase, immunoglobulin A (IgA), lactoferrin, lysozyme, interleukin 8 (IL-8), alpha-defensin 1, beta-defensin 1, beta-defensin 2, cathelicidin (LL-37)) using commercially available ELISA kits per manufacturer's instructions as described in Supplementary Table 1. Absorbance was read on a CLARIOstar plate reader. For samples below the limit of detection, a value of ½ the lowest standard was assigned. Cell-free nasal lavage fluid was not available for one subject in the cohort (Figure S1).

Sequencing Data Processing and Filtering. Five samples were removed from the dataset due to a low number of reads (Figure S1). A spiked pseudomonas positive control was identified correctly as pseudomonas. To control for potential contamination on the NELF strips, the decontam R package was used to remove contaminants.[36] This package uses an algorithm that takes into account the relative abundance of OTUs in samples and controls to remove the most likely contaminants and has been shown useful for respiratory samples.[37] This reduced the number of OTUs from 5346 to 4677. Alpha diversity measures (Observed, Chao1, ACE, Shannon, Simpson, Fisher) were calculated using the phyloseq R library before trimming OTU counts less than 5 for downstream analysis. This brought the number of OTUs to 3059 for downstream analysis.

Alpha diversity. Shannon and Simpson diversity indices were computed for each sample. Diversity indices were tested for normality using the Shapiro-Wilk test and further statistical tests to compare groups were carried out using the appropriate parametric (two-tailed t-test, ANOVA) or non-parametric (Kruskal-Wallis, Steel Dwass) tests. These analyses were performed using JMP Pro 14 and GraphPad Prism 8.

Nasal Microbiome Compositional Data Analysis. To limit spurious findings and because absolute sequencing counts are uninformative,[31,38,39] compositional data analysis (CoDA)[40] was carried out on the OTU count table after aggregating OTUs (O = 3059) by family (min. level assigned) and genera (max level assigned) and removing taxa not present in at least 20% of samples. The 20% sparsity threshold was selected to maximize class-specific information (Sex, Exposure group) while ensuring the microbial signatures were robust and contained minimal noise due to excessive sparsity. After aggregating OTUs, we define the taxa count matrix, $\mathbf{X} \in \mathbb{R}^{n \times p}$, with n = 62 samples and p = 143 taxa. The closure operator, $C[\cdot]$, was then used to map the count data of each element x_{ij} of $\mathbf{X}$ onto its corresponding coordinate on the unit-sum simplex, defining $\mathbf{X}' = C[\mathbf{X}]$ in terms of matrix elements as

$$x'_{ij} = (C[\boldsymbol{X}])_{ij} = \frac{x_{ij}}{\sum_{k=1}^{P} x_{ik}}$$

Because the presence of zeros is a major limitation of the log ratio transformation essential to CoDA, all zeroes must be robustly imputed to non-zero values. To overcome this we use the ratio-preserving multiplicative replacement strategy which has been shown to have several theoretical advantages over simple additive replacement:[41] we set the δ imputed values to a single constant equal to the smallest nonzero value encountered in $\mathbf{X}'$. From this, we impute zeros and replace $\mathbf{X}'$ with $\mathbf{Z}$ defined in matrix elements as:

$$z_{ij} = \begin{cases} \delta & , \quad x'_{ij} = 0 \\ \left(1 - \sum_{k|x_{ik}=0} \delta\right) x'_{ij} & , \quad x'_{ij} > 0 \end{cases}$$

Partial redundancy analysis to remove variation due to Sex. To remove the significant effect of Sex (which otherwise obscures the exposure group effect) on $\mathbf{Z}$, partial Redundancy Analysis (pRDA)[42] was used. Here we encode the Sex variable into the design matrix **S.** Additionally, to ensure multiple regression computations used in pRDA are performed on symmetric vectors in real space that preserves the inter-sample Euclidean distances, a center log ratio (clr) transformation was applied[40] to $\mathbf{Z}$, defining the clr values $\mathbf{C}$ for each sample as $\boldsymbol{c}_i = [c_1, \ldots, c_p]$ such that:

$$\mathrm{c_{ij}} = \log\left(\frac{\mathrm{z_j^i}}{\mathrm{G_i}}\right) \text{ where } G_i = \left(\prod_j z_{ij}\right)^{\frac{1}{p}}$$

With $\mathbf{C}$ defined, pRDA was carried out in the vegan R package.[43] Multivariate linear regression of $\mathbf{C}$ on $\mathbf{S}$ (i.e. computed as a series of multiple linear regression on individual features) was used to produce the fitted values $\widehat{\mathbf{C}}$. To remove the Sex effect as in pRDA, the adjusted values of $\mathbf{C}$ were computed by $\mathbf{P} = \mathbf{C} - \widehat{\mathbf{C}}$ where $\widehat{\mathbf{C}}$ contains all variation attributable to Sex. With $\mathbf{P}$ defined in Euclidean coordinates which are not suitable for downstream pairwise log ratio transformations, an inverse clr transformation was applied to map the adjusted coordinates back to the unit-sum simplex. The Sex adjusted relative abundance matrix $\mathbf{M}$ with elements m_{ij} is computed as:

$$m_{ij} = \frac{\exp(p_{ij})}{\sum_{k=1}^{p} \exp(p_{ik})}$$

Nasal Microbial Signature identification using Selection Energy Permuation. To identify microbial log ratio signatures in the setting of high-dimensional low sample size data we utilized the recently developed Selection Energy Permutation (SelEnergyPerm) method, which has been shown to have increased statistical power over several existing multivariate hypothesis testing

methods under hypothesis testing settings like this.[32] The SelEnergyPerm method simultaneously selects a reduced subset of log ratios while maximizing the association between groups. Let the group distributions be defined as $X \in \mathbb{R}^{n\,x\,f}$ and $Y \in \mathbb{R}^{m\,x\,f}$. In this work, we use SelEnergyPerm with the energy statistic (E-statistic)[44] defined by

$$\mathcal{E}_{n,m}(X,Y) = 2A - B - C,$$

where A, B, and C are specified, in terms of the vectors of $\mathbb{R}^f$ indexed by sample, by

$$A = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\|\boldsymbol{x}_i - \boldsymbol{y}_j\|, \qquad B = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|, \qquad C = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|\boldsymbol{y}_i - \boldsymbol{y}_j\|$$

From this, the pooled multi-class (#classes (c) $\geq$ 2) E-statistic becomes

$$S = \sum\nolimits_{1\leq j<k\leq k}\left(\frac{n_j + n_k}{2N}\right)\left[\frac{n_j n_k}{n_j + n_k}\,\mathcal{E}_{\mathrm{n_j},n_k}(X_j, X_k)\right]$$

The pooled E-statistic is then maximized using forward selection on a subset selected from the full set of pairwise log ratios to explain maximal variation when compared to the full set of pairwise log ratios. Similar to the approach in Greenacre et al.,[45] the reduced subset of log ratios are selected from the $\frac{p(p-1)}{}$ 2-dimensional feature space (all pairs). However, there are p^{p-2} possible ways to select a subset of log ratios that explain the total log ratio variance. To overcome this, SelEnergyPerm scores each log ratio using the differential compositional variation scoring method and then iteratively computes acyclic subsets of log ratios, [32] with permutation testing via Monte Carlo sampling[46] to assess the significance and prevent overfitting of the log ratio signature. Specifically, given a log ratio signature discovered with true labels, SelEnergyPerm tests if the observed pooled E-statistic (S^*) is more extreme than E-statistics sampled from the permutation distribution of log ratio signatures selected under random labels (S_i, indexing different random-label samples). With γ such E-statistics randomly sampled from the permutation distribution the one-sided p-value becomes

$$\hat{p} = \frac{1 + \sum_{i=1}^{\gamma} I(S_i > S^*)}{\gamma + 1}$$

As expected, we find that removing large numbers of uninformative features increases statistical power in the high-dimensional low-sample-size setting. To identify the Sex nasal microbial signatures in this study we utilized $\mathbf{Z}$ with labels $= Sex$ and for the Exposure group microbial signature we utilized $\mathbf{M}$ with labels $= Subject\ Group$. Using these data, we applied the SelEnergyPerm method with default settings using 200 permutations. Additionally, to reduce noise from sparse features, we further reduced taxa included in the analysis by first identifying the number of taxa to include in the microbial signature. We tested the following subset sizes: [5,10,20,40,60,80,100]. Applying the SelEnergyPerm method on each subset and normalizing the energy statistic[44] we selected the subset that maximized the normalized pooled energy statistic (**Figure S2**) and then tested if the observed S^* was more extreme than random. In

this way, we increase the statistical power of our analysis and reduce the chance of overfitting. While this is good for identifying associations, it can come at the expense of reduced overall discriminatory potential.

Network Visualization of Microbial signature. To visualize the microbial log ratio signatures, we constructed undirected graphs connecting the key taxa (vertices/nodes) by edges representing the formation of a ratio between two taxa with edge weight corresponding to the between-group Kruskal-Wallis H-statistic. While the full log ratio structure is directed in distinguishing numerators from denominators, directedness in the visualizations used here does not fundamentally change our interpretation. Graphs were visualized using Gephi[47] and R-igraph.[48]

Multivariate statistical test for microbial signals. To confirm associations between microbial log ratio signatures and Sex/Exposure group multivariate hypothesis testing was done using permutational multivariate analysis of variation[49] and implemented using the R vegan package.[43] Unsupervised lower-dimensional projections of samples and group centroids were done using principal coordinate analysis (PCoA) and were implemented using the R stats package.

Partial Least Squares Discriminate Analysis. We utilized partial least squares discriminate analysis (PLS-DA),[50,51] a versatile multivariate statistical regression technique, to model and understand the relationship between Sex/Exposure group to their microbial signatures. Shown to have reliable performance on compositional and genomic datasets,[52,53] PLS-DA models perform classification, inference, and are inherently linear thus offering improved model interpretability. We specified a priori the number of PLS-DA components (ncomp) as follows: for the between Sex nasal microbial signature, ncomp = 1; for the between Exposure group nasal microbial signature, ncomp = 2. Model fitting was done using the R caret [54] *plsda* function, with latent space projections and loadings extracted from the final models fit using all samples using R caret.[54] PLS-DA biplots were created by scaling and superimposing the loading vectors onto the score coordinates extracted from the final fitted model. PLS-DA biplots were visualized using the R ggplot2 package.[55]

Receiver operating characteristic curve analysis and PLS-DA performance metric. To understand how well the binary PLS-DA models discriminate between Sex using the nasal microbiome signature, we utilized the area under the receiver operating characteristic metric, AUC, which represents the probability that a randomly selected instance of class 1 will be ranked higher than a randomly selected instance of class 2.[56] Additionally, to understand the discriminatory potential of the ternary PLS-DA Exposure group models, the multi-class AUC metric was used. The multi-class AUC generalizes binary AUC through pairwise class AUC averaging and has the useful property of being independent of cost and priors as in AUC while having a similar interpretation to misclassification rate.[57] AUC metrics were estimated using repeated k-fold cross-validation.[58] The R pROC package[59] was used to compute all AUC metrics. ROC curves, which graph the false positive and true positive rate of a classifier over a range of thresholds, were computed using the R pROC package[59] and visualized using the R ggplot2 package.[55]

NLF mediator and microbiome data integration. We define the nasal lavage data matrix, $\mathbf{L} \in \mathbb{R}^{n \times f}$, where n = 66 samples and f = 7 mediators. Treating the data as relative such that sample-

wise absolute concentrations in our study are considered unimportant (**Figure S3A**), zeroes were imputed after applying the closure operator to **L** as described in our compositional data analysis methods. From this, we define $\mathbf{L}' \in \mathbb{R}^{n \times k}$, with k = 21, to include all pairwise log ratios from **L**. To remove uninformative NLF mediators, we computed the differential compositional variation (DCV) score [32] and assigned each NLF mediator log ratio a score by averaging the within-fold DCV score using 20 repeats of 10-fold cross-validation. NLF log ratios with a DCV score < 0 were considered uninformative and were removed (**Figure S3B**). From this ' was reduced to $\hat{\mathrm{L}} \in \mathbb{R}^{n \times k}$ where k=4 (**Figure S3C**) log ratios. To test for univariate associations between NLF mediator log ratios and Exposure group the Kruskal-Wallis test was applied followed by pairwise Wilcoxon rank-sum testing if $\alpha < 0.05$. The nasal microbiome signal was obtained by applying the SelEnergyPerm method to **M** to get $\widehat{\mathbf{M}} \in \mathbb{R}^{n \times r}$ where n = 62 and r = 9 log ratios.

Concatenating these data, we define the integrated NLF mediator and nasal microbiome matrix as $\mathbf{D} \in \mathbb{R}^{n \times f}$ where n = 61 (6 samples were removed due to either missing nasal microbiome or NLF data) and f = 13 (4-nasal lavage and 9 microbiome log-ratio features). Exposure group discrimination was estimated separately for each of $\hat{\mathrm{L}}$, $\widehat{\mathbf{M}}$, and **D** using multi-class AUC from 50 repeats of 10-fold cross-validation using 2-component PLS-DA models. Multi-class AUC estimates using $\hat{\mathrm{L}}$, $\widehat{\mathbf{M}}$, and **D** were compared between groups using the non-parametric Wilcoxon rank-sum test.

Nasal NLF mediator and microbiome association analysis

A final 2-component PLS-DA model to discriminate between exposure groups was fit to $\widehat{\mathbf{M}}$. Using dimensionality reduction inherent to PLS-DA, the first PLS-DA component (explaining the most variation) was extracted as a latent variable for further analysis. Pearson's correlation coefficients (PCC) and subsequent p-values were computed between the first PLS-DA component and **L''** represent the reduced nasal microbiome exposure group signature. PCC p-values, adjusted for multiple comparisons (q-value) using the Benjamini-Hochberg (BH) correction,[60] were considered significant if $q \leq 0.10$. These analyses were carried out using the R stats and caret packages.

Between Exposure group Correlation analysis. Partitioning the samples of **D** into 3 matrices based on exposure group (nonsmokers, e-cig users, or smokers), we calculate all pairwise PCC and p-values between features for each group. We also report q-values after adjusting for multiple comparisons within each group using the BH method. Correlations were considered significant if $q \leq 0.10$. Significant PCC within each subject were then aggregated across all exposure groups and visualized as a graph using the R igraph package.[48]

Confidence Intervals and univariate statistical test for log ratios. Log ratio 95% confidence interval estimates were calculated by

$$CI_i = \bar{x}_i \pm 1.96 \frac{s_i}{\sqrt{n}}$$

where for the *ith* log ratio, $\bar{x}_i$ = sample mean, s_i = sample standard deviation and n =number samples. Log ratios with confidence intervals bounds that do not include 0 are interpreted as enriched on average for the numerator if $\bar{x} > 0$ or denominator if $\bar{x} < 0$. The Kruskal-Wallis

and Wilcoxon rank-sum test were used for univariate comparisons of log ratios between Sex or Exposure groups. Moreover, p-values were adjusted for multiple comparisons using the BH correction using the R stats library and are reported as q-values.

Results

Subject Demographics

Demographic, questionnaire, and smoking/vaping diary data are summarized in **Table 1**. The study cohort was comprised of 30% nonsmokers (n = 20), 42% e-cigarette users (n = 28), and 28% smokers (n = 19) with at least n = 8 per sex within each exposure group. E-cigarette users were significantly younger (26.39 ± 1.44) than nonsmokers (30.75 ± 1.32) and smokers (31.89 ± 1.91) ($p < 0.05$). BMI did not differ significantly between the exposure groups. Questionnaires and smoking/vaping diaries were completed for 95% (19/20) of nonsmokers and 100% of e-cigarette users and smokers. However, there was variability in the completeness of diaries filled out by e-cigarette users, particularly for the e-cigarette use parameters (mL/day, puffs/day, nicotine concentration, flavor, device). Cigarette users smoked an average of 12.68 ± 0.96 cigarettes per day, whereas 25% (7/28) of e-cigarette users smoked a cigarette during the diary period with an average of 0.14 ± 0.07 cigarettes per day, while 13 e-cigarette users reported puffs per day and 16 reported mL e-liquid/day and e-liquid nicotine concentration in mg/mL. These e-cigarette users averaged 53.90 ± 16.54 puffs/day, 3.60 ± 0.70 mL of e-liquid, and 19.43 ± 4.92 mg/mL nicotine in e-liquids. One smoker reported vaping on one day of the diary, which is the reason for the non-zero values for e-cigarette use parameters in the smoker category. Nonsmokers did not report previous cigarette smoking or marijuana use, whereas 79% (22/28) of e-cigarette users were former cigarette smokers, while 14% (4/28) of e-cigarette users and 21% (4/19) of smokers reported marijuana use in their diaries. Cotinine, a metabolite of nicotine, was not detectable in the serum of nonsmokers and was significantly elevated in the serum of e-cigarette users (127.99 ± 15.42) and smokers (170.16 ± 21.41) in comparison with nonsmokers ($p < 0.0001$), as expected.

Nasal Microbiome Characteristics

The 4677 OTUs included in the dataset represented OTUs from 19 unique phyla and 225 unique genera. The top four most abundant phyla by average relative abundance across all samples were *Actinobacteria* (50.2%), *Firmicutes* (36%), *Proteobacteria* (12.0%), and *Bacteroidetes* (1.6%). The top six most abundant genera by average relative abundance across all samples were *Corynebacterium* (40.7%), *Staphylococcus* (19.9%), *Propionibacterium* (11.8%), *Alliococcus* (8.5%), *Moraxella* (5.3%), and *Streptococcus* (4.2%). This microbial composition is similar to previously reported studies of the nasal microbiome.[61,62] These data are summarized in **Figure 1**, where relative abundances by exposure group and sex are plotted for the most highly abundant phyla and genera.

Alpha Diversity

To determine whether there are differences in alpha diversity between the nasal microbiomes of smokers, nonsmokers, and e-cigarette users, we calculated alpha diversity indices (Observed,

Chao1, ACE, Shannon, Simpson, Fisher) using phyloseq.[63] We did not find any statistically significant differences between the exposure groups for any measure of alpha diversity; however, we did observe a non-significant trend of increased alpha diversity in smokers (**Figures 2A and 2B**). Because our group and others have previously observed sex differences in respiratory mucosal immune responses[64,65] we also tested whether alpha diversity was significantly different between male and female subjects. We found that both the Shannon and Simpson indices were significantly higher in males than females ($p = 0.021$ and $p = 0.0078$, respectively) (**Figures 2C and 2D**). We then tested for the interaction between sex and exposure group and found that sex was a significant source of observed variation ($p = 0.0286$ for Shannon; $p = 0.0102$ for Simpson), while exposure group was not. When the data were stratified by exposure group, the only male-female comparison that remained significant was in the e-cigarette user group ($p = 0.0361$ for Shannon; $p = 0.0124$ for Simpson) (**Figures 2E and 2F**). These results suggest that sex is an important biological variable to consider in studies of the nasal microbiome.

Compositional Difference of the Nasal Microbiome by Sex

Because we observed distinctions in alpha diversity between sexes, we next tested whether there were significant compositional differences between the sexes and to identify specific genera capable of explaining these dissimilarities. Given challenges presented by sparse, compositional 16S rRNA sequencing data combined with high-dimensionality (genera = 255) and small sample size (n=62), we leveraged the SelEnergyPerm [32] method to identify a robust signature of nasal microbiome taxa (among sparse noisy data) capable of explaining compositional differences between sexes.

By applying this method, we discovered (beyond random noise) a subset of genera ($g = 6$) capable of maximizing the energy distance between male and female samples ($p = 0.0123$, **Figure S1A**). This microbial signature was comprised of four log ratios between *Rhodococcus*, *Finegoldia*, *Sneathia*, *Abiotrophia*, *Tannerella*, and *Yaniella* genera (**Figure 3A**). Using the identified log ratio signature, PERMANOVA analysis (pseudo-F = 16.586, $p = 0.0002$, **Figure 3B**) also confirmed the existence of differences in the nasal microbiome composition between sex. Analysis of individual taxa log ratios between sexes demonstrated important nasal microbiome compositional differences (**Figure 3C**). In female samples, *Yaniella* was more abundant on average than *Rhodoccous* and *Tannerella,* while the reverse was true for males. In male samples, *Abiotrophia* was more abundant on average than *Sneathia,* while the opposite was true for females. Finally, in both males and females, *Finegoldia* was observed to be more abundant than *Yaniella,* however, *Finegoldia* was significantly more enriched relative to *Yaniella* in males compared to females.

Next, we analyzed the microbial signature as a whole using Partial Least Squares Discriminate Analysis (PLS-DA) with a single component to predict sex. Using 20 repeats of 10-fold cross-validation, the average area under the receiver operating characteristic curve (AUC) for predicting sex given the reduced microbial signature was 0.862 (95% CI 0.842 – 0.883, **Figure 3D**). With strong cross-validated predictive performance, a final PLS-DA model was trained on all samples (n=62). Scores from the single PLS-DA component indicated strong separation between sexes (**Figure 3E**). The PLS-DA loading plot (**Figure 3F**), which shows how each log ratio contributes to the final score, demonstrates key relationships between taxa log ratios.

Increased abundance of *Abiotrophia and Finegoldia* (in log ratios where they appear) were characteristic of males, and increased abundance of *Yaniella* was associated with females. Overall, these findings indicate there exists a compositionally distinct taxa subset that differs strongly in the nasal microbiomes of males and females. Therefore, controlling for sex differences present in the nasal microbiome is important in further analysis.

Compositional Difference of the Nasal Microbiome by Exposure group

We next examined whether there were distinct nasal microbiome compositions between exposure groups (e-cigarette users: n = 24; smokers: n=19; nonsmokers: n=19; See Methods and **Table 1)**. Taking into account nasal microbiome sex differences and applying SelEnergyPerm, we identified a subset of genera (g = 12) important for explaining key nasal microbiome alterations between exposure groups (p = 0.032, **Figure S1B**). This microbial signature comprised nine log ratios (edges) between 12 key genera (nodes) (**Figure 4A**). PERMANOVA analysis (pseudo-F = 8.4889, p =0.0002, **Figure 4B**) confirmed differences in nasal microbiome composition between exposure groups given the microbial signature of 9 log ratios.

Individual analyses of log ratios elucidated specific compositional differences between exposure groups (**Figure 4C**). In e-cigarette users, *Lactobacillus* taxa were significantly more abundant relative to *Bacillus* taxa, while in smokers and nonsmokers, these taxa presented in similar proportions, suggesting an enrichment of *Lactobacillus* among e-cigarette users. E-cigarette users' nasal microbiomes also contained significantly more *Staphylococcus* relative to *Bacillus* than what was observed in nasal microbiomes of both smokers (q = 0.0097) and nonsmokers (q = 0.0031). In smokers, *Maccrocus* genera were significantly more abundant on average relative to *Hymenobacter*, *Mycobacterium*, *Varibaculum*, and *Rhodococcus,* suggesting that smoking may enrich *Macrococcus* taxa populations in the nasal passage. Additionally, smoker nasal microbiomes contained more *Hymenobacter* relative to *Moryella,* whereas the opposite was true for nonsmokers, both in contrast to e-cigarette users, which maintained on average equal amounts of both genera. In nonsmokers, *Lautropia* taxa were significantly more abundant relative to *Bulleidia,* but this was not observed in smokers and e-cigarette users.

To understand how taxa log ratios work together to discriminate between exposure groups, PLS-DA was used with 20 repeats of 10-fold cross-validation (**Figure 4D**). The estimated multi-classification AUC was 0.851 (95% CI 0.835 – 0.866) suggesting excellent exposure group discrimination. Pairwise examination of exposure group classifications shows strong differences between the nasal microbiomes of nonsmokers/e-cigarette users (AUC = 0.895: 95% CI 0.874 – 0.915) and smokers/e-cigarette users (AUC = 0.893: 95% CI 0.873 – 0.913), with weaker yet distinct differences between smokers/nonsmokers (AUC = 0.803: 95% CI 0.773 – 0.833) (**Figure 4D**). The relative importance of taxa log ratios for discriminating between exposure groups was computed using a final PLS-DA model fit using all samples (n=62). The log ratio between *Macrococcus* relative to *Hymenobacter* was found to be most important for classifying samples as smoker (least important for e-cigarette user classification), and the log ratio between *Bacillus* taxa relative to taxa from the *Micrococcaceae* family was most important for samples to be classified as e-cigarette users (least important to be classified as smokers). (**Figure 4E**). Interestingly, inspection of relative log ratio importance data failed to uncover log ratios disproportionately important for nonsmokers. This observation suggests smoking and e-cigarette

use recognizably alter the nasal microbiome in otherwise healthy adults. Overall, analysis of the taxa log ratios signature suggests alterations in *Macrococcus* and *Bacillus* genera are important for distinguishing between these exposure groups.

Differences in NLF mediator Expression Patterns Between Exposure groups

Because smoking and e-cigarette use were associated with distinct changes in the nasal microbiome, we next explored if there was altered expression of innate immune response mediators in the exposure groups. Accounting for differences in absolute concentration (**Figure S3A**) and subsequently applying differential compositional variation scoring [32] (See Methods, **Figure S3B**), we identified four log ratios among NLF mediators that showed strong intergroup variability (**Figure S3C**). These ratios comprised the following NLF mediators: IL-8, DEFB4A-2, neutrophil elastase, IgA, and lactoferrin. Kruskal-Wallis one-way testing (**Figure S3D**) of each log ratio suggest there exist intergroup differences in NLF mediator expression formed between the concentrations of neutrophil elastase relative to IL-8 ($H = 6.4417$; $p = 0.0399$; $q = 0.0798$) and lactoferrin relative to IL-8 ($H = 8.2080$; $p = 0.0165$; FDR = 0.0660). There were no significant differences between exposure groups among log ratios formed by IgA relative to IL-8 or DEFB4A-2 relative to neutrophil elastase. However, multivariate analysis with PERMANOVA (pseudo-F = 3.7678, $p = 0.0030$) using the four key log ratios confirmed there were differences in NLF mediator expression patterns between exposure groups when considered together. To better understand which groups were different, we applied PLS-DA. Training a PLS-DA model with the NLF mediator expression patterns revealed the strongest between-subject-group discrimination to be among Smokers and Nonsmokers (AUROC = 0.8230, 95%CI 0.7920-0.8530, **Figure S3E**). Notably, e-cigarette users' NLF mediators were weakly distinguishable from nonsmokers (AUROC = 0.6720, 95%CI 0.6350-0.7100, **Figure S3E**) but more discernible from smokers (AUROC = 0.7480, 95%CI 0.7130-0.7820, **Figure S3E**). Together, these results suggest that the expression of NLF mediators in smokers was distinct from that of e-cigarette users and healthy adults.

Integration of NLF mediators and nasal microbiome composition improves exposure group discrimination

Finally, we aimed to understand if alterations in NLF mediator expression are associated with nasal microbiome dysbiosis resulting from smoking or e-cigarette use. To this end, we first estimated the discriminatory AUROC of a 2-component PLS-DA model fit on log ratios from NLF mediators (**Figures S2C - S3**), nasal microbiome (**Figures 4A – S4A**), or both nasal microbiome and NLF mediators (**Figure 5A**). When compared to individual signatures, improved discriminatory AUROC (**Figure 5B**) was observed when PLS-DA models were fit using the combined nasal microbiome and NLF mediator signatures. Therefore, with established synergy between mediator expression and nasal microbiome composition in discriminating between exposure groups, we next examined if correlations were present between the two.

Association between altered NLF mediator expression and nasal microbiome dysbiosis

Using the first PLS-DA component of the nasal microbiome signature, we found significant correlations with NLF mediator expression, showing an association between the nasal

microbiome composition and NLF mediator expression (**Figure 5C**). Examination of the location of samples by exposure group projected along the first PLS-DA component show important projective distinctions between smokers (on average negative projections) and both e-cigarette users and nonsmokers (on average positive projections) (**Figure S4B**). Given this, these correlations suggest nasal microbiome dysbiosis caused by cigarette smoke exposure is associated with increased expression of IL-8 relative to neutrophil elastase, Total IgA, and lactoferrin (**Figure 5C**). Moreover, the loadings along the first PLS-DA component (**Figure S4C**) show log ratios with higher abundance of *Maccroccous* as being the most important contributor to negative projections. Combined, these data propound an important link between dysbiosis in *Macrococcus* communities within the nasal microbiome and NLF IL-8 expression.

Microbial functional and mediator expression differences between exposure groups

Correlation analysis of the combined NLF mediator expression and nasal microbiome signature reveal distinct correlation patterns within exposure groups suggesting distinct functional differences (**Figures 5D**). Most notably, a significant negative correlation between log ratios formed by *Hymenobacter/Moryella* and *Macrococcus/Hymenobacter* was observed only in the nonsmoker group. This negative correlation highlights a possible role of *Hymenobacter*, in that it appears to be important for maintaining a healthy balance of *Maccroocous* and *Moryella.* In the e-cigarette and smoking groups, we observed a significant positive correlation between the log ratios formed by IgA/IL-8 *and* Lactoferrin/IL-8. Analysis of this correlation pattern reveals that increased expression of IL-8 in these groups may come at the expense of decreased expression of IgA and lactoferrin or vice versa. We also observed a significant negative correlation between the log ratios formed by Neutrophil Elastase/IL-8 and DEFB4A-2/Neutrophil Elastase in the e-cigarette and smoking groups. These strong negative correlations show that increased expression of IL-8 and DEFB4A-2 subsequently results in decreased expression of neutrophil elastase. The final significant correlation pattern observed was in smokers only and consisted of four positively correlated log ratios formed by *Macrococcus* relative to *Hymenobacter*, *Mycobacterium*, *Varibaculum*, and *Rhodococcus* (**Figure 5D**). Relatively interpreting these correlations between log ratios suggests that as *Macrococcus* becomes more abundant (among these ratios) the abundance of *Hymenobacter*, *Mycobacterium*, *Varibaculum*, and *Rhodococcus* decreases. This suggests that cigarette smoke exposure may produce favorable colonization conditions for *Maccroccous* genera which subsequently reduces the abundance of Hymenobacter, *Mycobacterium*, *Varibaculum*, and *Rhodococcus*.

From these analyses, our results demonstrate there exists a strong association between altered NLF mediator expression and nasal microbiome dysbiosis. Our findings indicate nasal microbiome dysbiosis from smoking results in the simultaneous increase in IL-8 expression and *Maccroccous* abundance. Additionally, variations in the correlation networks among e-cigarette users and smokers, while similar, were distinct from nonsmokers, suggesting functional differences at the microbial and mediator levels between exposure groups.

Discussion

Despite the growing body of research showing that e-cigarette use can disrupt the respiratory immune system, no studies to date have assessed the effects of e-cigarettes on the

579 respiratory microbiome and host-microbiota interactions. In this study, after adjusting for sex
580 differences, we found that e-cigarette users, smokers, and nonsmokers have unique nasal
581 microbiomes, with differences driven by the relationships between a subset of key taxa. We also
582 found a subset of immune mediators that had distinct relationships between each other in the
583 different exposure groups. Importantly, we found a link between nasal microbiome dysbiosis and
584 soluble immune mediator networks.
585 A fundamental feature of our study is that we detected microbial signatures from the
586 nasal microbiome that explained differences between sex and exposure groups using the novel
587 SelEnergyPerm computational method. This method directly accounts for the sparse, high-
588 dimensional and compositional nature of the 16S relative abundance data. Additionally,
589 SelEnergyPerm identifies subsets of robust log ratios between taxa, as opposed to analyzing taxa
590 relative abundance alone, yielding higher statistical power in the sparse association setting with
591 low-sample-size compositional data. [32] Most importantly, traditional statistical techniques such
592 as PERMANOVA, ANOSIM, and ANCOM alone were unable to detect these sparse
593 associations within the high-dimensional nasal microbiome feature space. Further, our
594 parsimonious yet statistically significant signatures were then integrated with NLF mediators
595 where we were then able to uncover novel interactions between a taxa subset within the nasal
596 microbiome and the NLF mediators in response to exposure to cigarette or e-cigarette aerosol.
597 We observed that there were relationships between a subset of taxa that were important in
598 separating the microbial communities of smokers, nonsmokers, and e-cigarette users (**Figure 4**).
599 Only a few studies have previously compared the nasal microbiome of smokers and
600 nonsmokers.[23,66] Charlson et al. found specific bacteria genera that were differentially abundant
601 in smokers and that some genera belonging to the phylum *Firmicutes* were important in
602 distinguishing smokers from nonsmokers.[23] Other studies did not find any significant differences
603 in diversity measures or relative taxa abundance between smokers and nonsmokers.[66] In our
604 study, which focused on the composition of the nasal microbiome and ratios between taxa rather
605 than relative abundance of individual taxa, we found that alterations in *Macrococcus* and
606 *Bacillus* genera are important for distinguishing between exposure groups. Our data also suggest
607 an enrichment of *Lactobacillus* and *Staphylococcus* relative to *Bacillus* in e-cigarette users and
608 enrichment of *Macrococcus* relative to *Hymenobacter*, *Mycobacterium*, *Varibaculum*, and
609 *Rhodococcus* in smokers. A shift from *Lactobacillus* to *Bacillus* in the lung microbiome has been
610 previously demonstrated in response to influenza A infection and increases in anaerobic bacteria,
611 such as *Lactobacillus,* have been associated with chronic rhinosinusitis.[61] Furthermore, *Bacillus*
612 have been shown to produce antimicrobials against *S. aureus*,[67] indicating that the patterns we
613 have observed may be directly linked to specific interactions between taxa. An increase in
614 *Staphylococcus* relative to *Bacillus* in e-cigarette users is also notable due to the role of species
615 such as *Staphylococcus aureus*, which is carried normally by about 30% of people and is also
616 considered to be a potential pathogen of the skin and mucosal surfaces.[68,69] Our data provide
617 evidence that e-cigarette and smoker nasal microbiomes are distinctly shifted from nonsmokers.
618 Importantly, we also observed that different subsets of taxa were important in separating e-
619 cigarette users and smokers, rather than effects on a continuum from nonsmokers to e-cigarette
620 users to smokers, highlighting the concept that the effects of e-cigarettes are likely unique from
621 those of smokers, even though they are commonly directly compared.
622 We also measured concentrations of mediators of host-microbiota interactions in nasal
623 lavage fluid to determine whether the changes in the nasal microbiome in different exposure
624 groups are potentially caused by direct effects on the microbiome, mediated by changes in the

host immune system, or both. Our data indicate that the expression of immune mediators in nasal lavage fluid samples differed among exposure groups and was driven by shifts in neutrophil elastase and lactoferrin relative to IL-8. Neutrophil elastase and IL-8 are associated with inflammation and neutrophil recruitment, while lactoferrin is an antimicrobial protein primarily produced by epithelial cells and has a wide array of functions, including antioxidant and immune-modulating properties.[70] Our results suggest that e-cigarette users and smokers may have altered immune mediator milieu, indicating a shift away from immune homeostasis and towards increased inflammation and neutrophil recruitment. This shift could be partially driving observed differences in the nasal microbiome.

Our data indicate that both e-cigarette users and smokers have altered nasal microbial communities and relationships between markers of innate immune response, which could imply that they are at increased susceptibility to respiratory infections and/or that they exist in a state of inflammation and altered immune response. We also uncovered interactions of key immune mediators with the host and microbiota, such as IL-8, neutrophil elastase, and lactoferrin, that are also disrupted by e-cigarette and cigarette use. The microbial shifts we observed in association with e-cigarette and cigarette use could be driven by changes in the microenvironment, such as temperature, pH, free radical formation, and availability of metabolic substrates (e.g. sugars) that could then alter the fitness of different bacteria in the nasal microbial community. The shifts we observed could also be mediated through direct effects on respiratory host defense function, inflammation, and/or specific microbes. Multiple processes are likely at play, but our novel findings on the effects of e-cigarettes on the nasal microbiome add to the growing body of literature demonstrating that e-cigarettes are not without health effects and that they should be more thoroughly investigated for inhalational toxicity.

Because sex differences in the human immune system and its response to respiratory disease and toxicant exposure have been observed previously,[64,71] we also investigated whether there were sex differences in the nasal microbiomes of our subjects. We observed that the relationships between six genera were important in separating the nasal microbiomes of males and females (Figure 4A). Increased abundance of *Abiotrophia* and *Finegoldia* (in log ratios where they appear) were characteristic of males, and increased abundance of *Yaniella* was associated with Females. Many of these genera have been detected in previous studies of skin, oral, and/or respiratory microbiomes,[23,61,72-76] but detailed information on the functions of these bacteria as part of the microbial community, as well as their impact on host health, are not available for all taxa. Although some of these genera, such as *Abiotrophia* and *Finegoldia* have been associated with disease- and exposure-driven alterations in the respiratory microbiome,[23,61,72,73] we hypothesize that the observed sex difference is neither good nor bad; rather, it is reflective of a different baseline composition in males and females or altered microenvironments in males and females due to differences in toxicant metabolism rates or mechanisms of immune regulation.[77,78] In other body sites, such as the gut, sex differences have been detected and have been attributed to a variety of factors, including sex hormone levels, pharmaceutical use, and diet.[79,80] In mice, sex-related differences in gut microbiota were shown to impact pulmonary responses to ozone.[65] However, few studies have explored sex differences in the respiratory microbiome.[81] In the studies that have analyzed data by sex, detection of sex differences is not consistent between studies and is typically not explored in-depth.[62,68,82] Importantly for the data presented here, compositional differences in the nasal microbiomes of e-cigarette users, smokers, and nonsmokers were not apparent until sex was properly adjusted for,

further underscoring the importance of considering sex as a biological variable which significantly modifies exposure effects and can substantially affect data interpretation.

Though our study reveals important community shifts in nasal microbiota and immune mediators associated with e-cigarette and cigarette use as well as with sex, there are limitations to our study. Our novel analysis approach, while properly accounting for the compositional nature of the data, limits us in comparing our work to previous studies, which have been more focused on specific taxa rather than ratios across the microbial community as a whole. As with any study of human subjects, there is also inherent inter-subject variability that can interfere with detection of differences between groups. In our e-cigarette user group, there was considerable variability in factors that could impact the exposure subjects are receiving, including e-liquid flavor, device, nicotine content, and frequency of use. The e-cigarette user group also includes previous smokers and some marijuana use was reported in both smoker and e-cigarette user questionnaires. These factors were included in our analysis and did not show a significant impact on our overall findings due to the nature of the computational models we used. In future studies, larger cohort sizes coupled with more extensive questionnaires could improve the ability to detect which, if any, of these factors may be driving changes in microbiota composition and would also increase power to detect overall changes and shifts in the nasal microbiomes of such subjects given the compositional and sparse nature of 16S sequencing data.

As a whole, our results support and expand on the previously published notion that exposure to inhaled toxicants, including tobacco products, can influence the respiratory microbiome.[23,83,84] The novel, robust computational approach in terms of pairwise log ratios that we applied allowed us to uncover both exposure- and sex-dependent effects on nasal mucosal host defense responses using straightforward, non-invasive sampling of the upper respiratory tract of human subjects. Importantly, we were able to integrate 16S sequencing data with expression of soluble immune mediators to understand interactions between the nasal microbiome and host milieu by appropriately handling the sparse, compositional data generated by 16S sequencing, accounting for inter-individual variability between subjects' mediator levels, and selecting for features that were most important for separating classes, resulting in interpretable, biologically meaningful results. Conventional analysis pipelines would have limited our ability to integrate these two types of data and detect the exposure and sex-dependent effects we observed, highlighting the importance of applying innovative computational methods to address specific research questions and integrating multiple factors in understanding biological outcomes of exposure and disease.

Acknowledgments: This research was funded by the National Institutes of Health (NIH) Grant Nos. R01 HL139369, T32 ES007126, P50 HL120100, and F31 HL154758, by an HHMI Gilliam Award (GT11504), and by a JSMF Complex Systems Scholar Award (#220020315). The content is solely the responsibility of the authors and does not necessarily represent the official views of any agency funding this research.

Author Contributions: I.J. and E.H. conceived the study and were in charge of overall direction and planning. M.E.R., C.R., and I.J. contributed to clinical research operations, including the collection and processing of samples and subject demographic data. B.Z. and M.W. performed metagenomic sequencing, quality control, and analysis through OTU assignment. E.H. performed experiments to measure proteins in nasal lavage fluid. A.H. led the remaining data

analysis with help from E.H. and feedback from P.J.M. and I.J.. E.H. and A.H. took the lead in writing the manuscript. All authors provided feedback on and helped shape the final manuscript.

Competing Interests Statement: The authors report no competing interests.

Data Availability: Raw sequencing data is available under the SRA BioProject accession number PRJNA746950. Processed OTU and NLF tables by exposure group and sex have been deposited in the github repository: https://github.com/andrew84830813/nasalMicrobiome_EcigSmoking/Data/

Code accessibility

All nasal microbiome analyses were done using version 4.0.0 of the R statistical programming language. All input data, R script, and functions used in the analysis presented here can be retrieved from the github repository: https://github.com/andrew84830813/nasalMicrobiome_EcigSmoking.git

Materials & Correspondence: Ilona Jaspers, Ilona_jaspers@med.unc.edu

734

735 **Table 1.** Subject demographics. Reported values are mean ± standard error. Groups were
736 compared using the Steel Dwass method for non-parametric multiple comparisons. AA = African
737 American. # $p < 0.05$ in comparison with nonsmokers and smokers. **** $p < 0.0001$ in
738 comparison with nonsmokers.

	Nonsmokers	**E-Cigarette Users**	**Smokers**
n	20	28	19
Sex (Male/Female)	8/12	19/9	10/9
Race (White/AA/Asian/Other)	16/1/2/1	18/4/5/1	10/8/0/1
Age	30.75 ± 1.32	26.39 ± 1.44#	31.89 ± 1.91
BMI	27.11 ± 1.31	30.07 ± 1.51	27.65 ± 1.43
Cigarettes/Day	0 ± 0	0.14 ± 0.07	12.68 ± 0.96
mL E-Liquid/Day	0 ± 0	3.60 ± 0.70	0.015 ± 0.015
E-Cigarette Puffs/Day	0 ± 0	53.90 ± 16.54	0.466 ± 0.414
E-Liquid Nicotine (mg/mL)	0 ± 0	19.43 ± 4.92	0.158 ± 0.158
Former Cigarette Smoker (Yes/No)	0/20	22/6	19/0
Marijuana Use (Yes/No)	0/20	4/24	4/15
Serum Cotinine (ng/mL)	0 ± 0	127.99 ± 15.42****	170.16 ± 21.41****

739
740
741

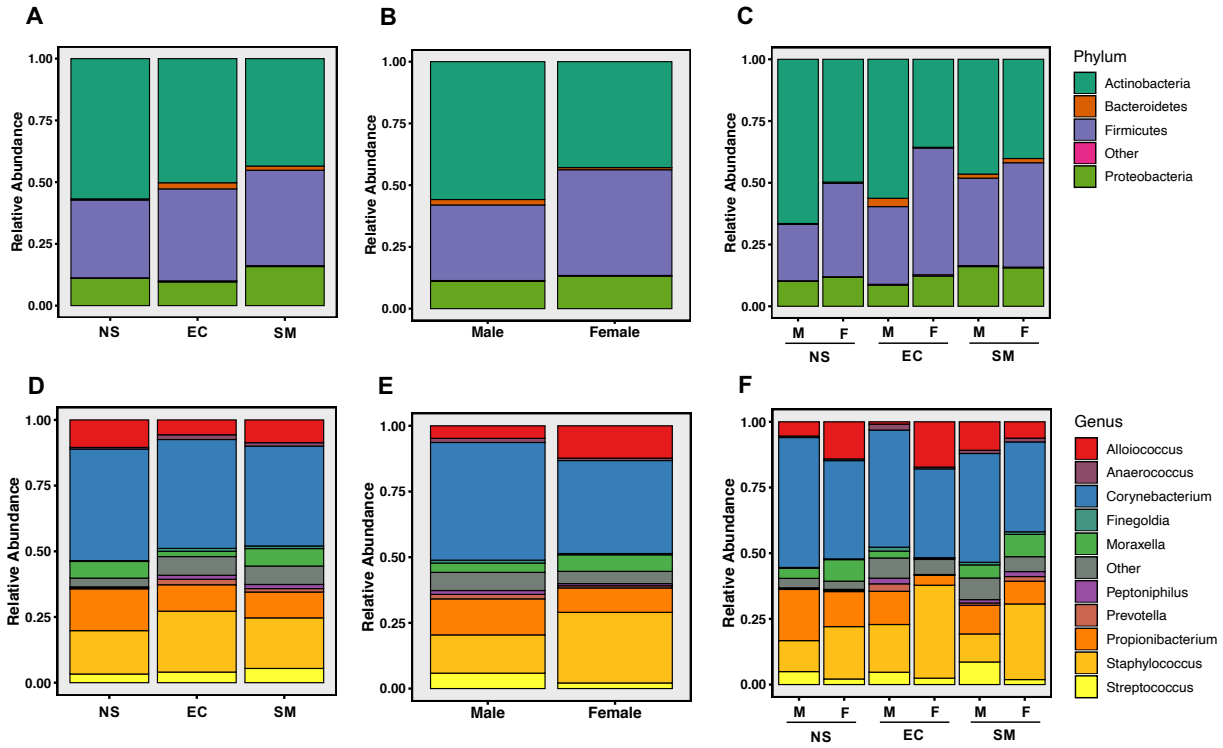

Figure 1. Average relative abundances of the top 4 phyla (A-C) and top 10 genera (D-E) plotted by exposure group (A, D), sex (B, E), and sex within exposure groups (C, F). NS = nonsmoker, EC = e-cigarette user, SM = smoker, M = male, F = female.

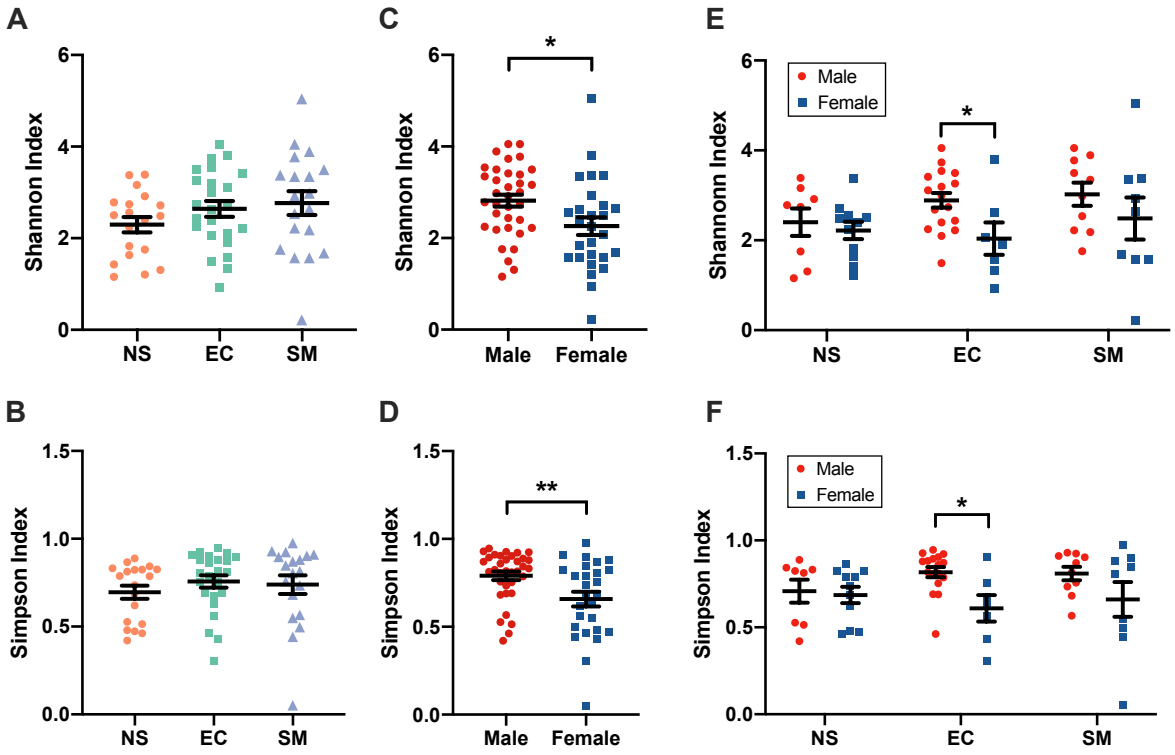

Figure 2. Shannon and Simpson indices of alpha diversity are significantly different between sexes, and this difference is most pronounced in e-cigarette users. The Shannon and Simpson indices for alpha diversity were calculated and plotted by exposure group (A, B), sex (C, D), and sex within exposure groups (E, F). NS = nonsmoker, EC = e-cigarette user, SM = smoker. Data are presented as mean ± standard error. * $p < 0.05$, ** $p < 0.01$ by t-test (C), Kruskal-Wallis test (D), or two-way ANOVA with Fisher's LSD (E, F).

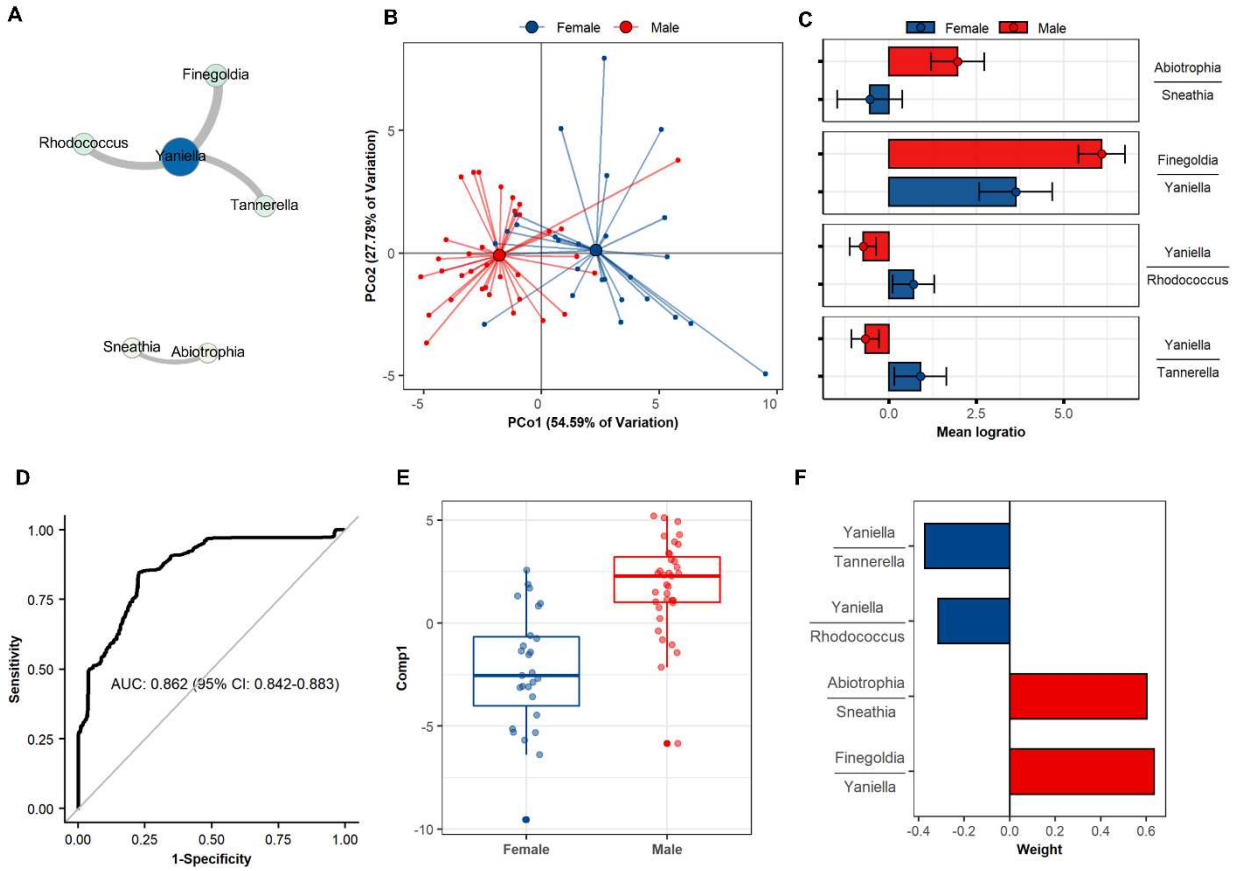

Figure 3 | Nasal microbiome differences between sexes (Males: n=35; Females: n=27). (A) Network representation of SelEnergyPerm (p=0.0123) derived genus aggregated taxa log ratio signature of nasal microbiome differences between sexes (Node = genera; edge = log ratio between taxa, Edge-weight = Kruskal Wallis H-statistic between sexes, Size/Color = node strength). (B) Principal coordinate analysis plot of nasal microbiome log ratio signature between sex explaining 82.37% of the total variation. (C) Univariate analysis of log ratio signature showing average depletion or enrichment of specific taxa log ratios between sexes. Error bars reflect 95% confidence intervals of the mean log-ratio value for males and females. (D) Receiver operating characteristics (ROC) curve displaying the area under the curve (AUC) predictive performance (20x10-fold cross-validation) of 1-component partial least squares discriminant analysis (PLS-DA) models trained on nasal microbiome signature between sexes. (E) PLS-DA scores plot of single discriminating component between sexes. Final PLS-DA model fit using all samples (n=62). (F) PLS-DA loadings plot showing contributions of each log ratio to final scores.

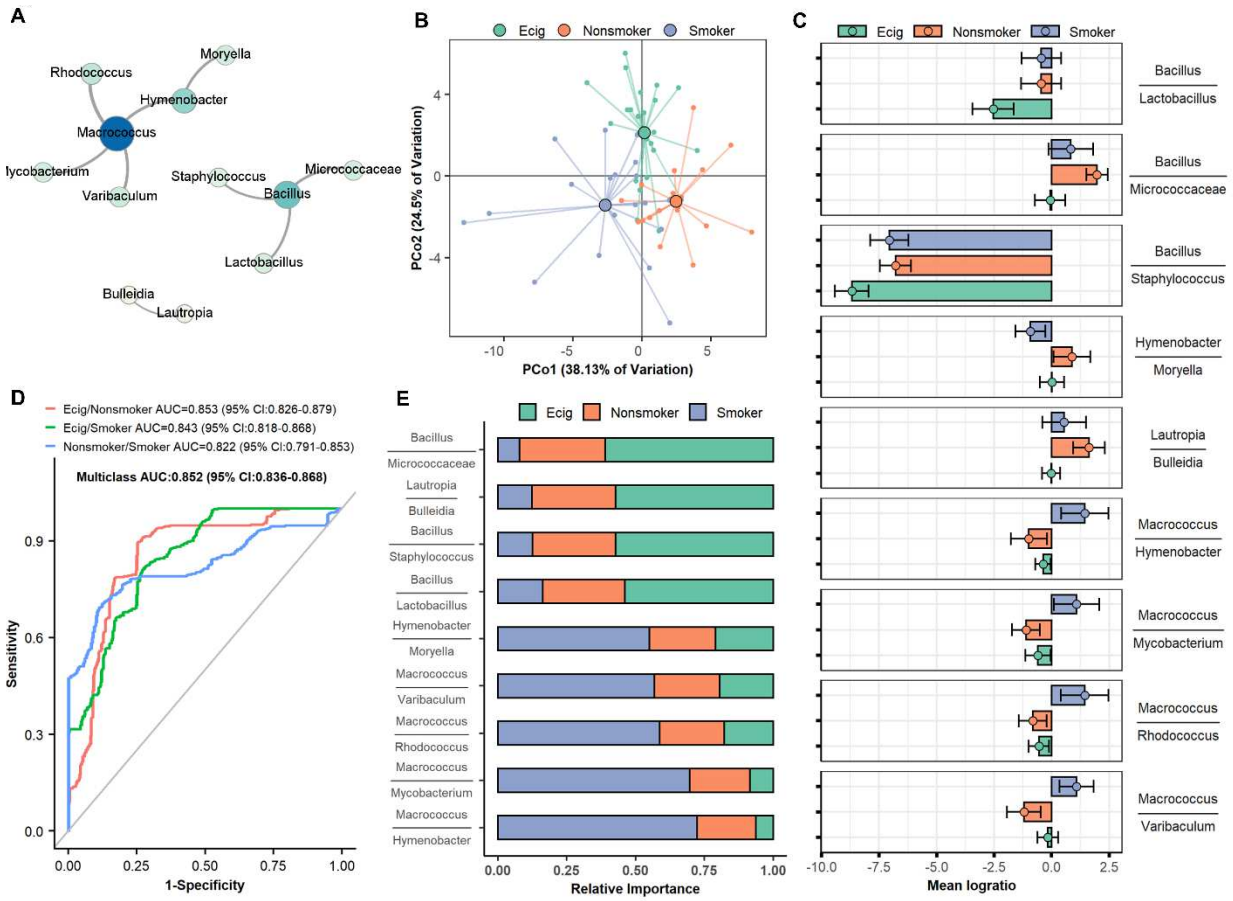

Figure 4 | Nasal microbiome differences between exposure groups (Ecig-users: n=24; Nonsmokers: n=19; and Smokers: n=19) adjusted for sex. (A) Network representation of SelEnergyPerm (p=0.032) derived genus aggregated taxa log ratio signature of nasal microbiome differences between exposure groups (Node = genera; edge = log ratio between taxa, Edge-weight = Kruskal Wallis H-statistic between sex, Size/Color = node strength). (B) Principal coordinate analysis plot of nasal microbiome log ratio signature between exposure groups explaining 62.63% of the total variation. (C) Univariate analysis of log ratio signature showing average depletion or enrichment of specific taxa log ratios between exposure groups. Error bars reflect 95% confidence intervals of the mean log-ratio value for each exposure group. (D) ROC curve displaying the multi-classification AUC for predicting exposure group (20x10-fold cross-validation) of 2-component PLS-DA models trained on nasal microbiome signature between exposure groups. (E) Relative importance of log ratios for distinguishing between exposure groups in PLS-DA model trained on all samples (n=62).

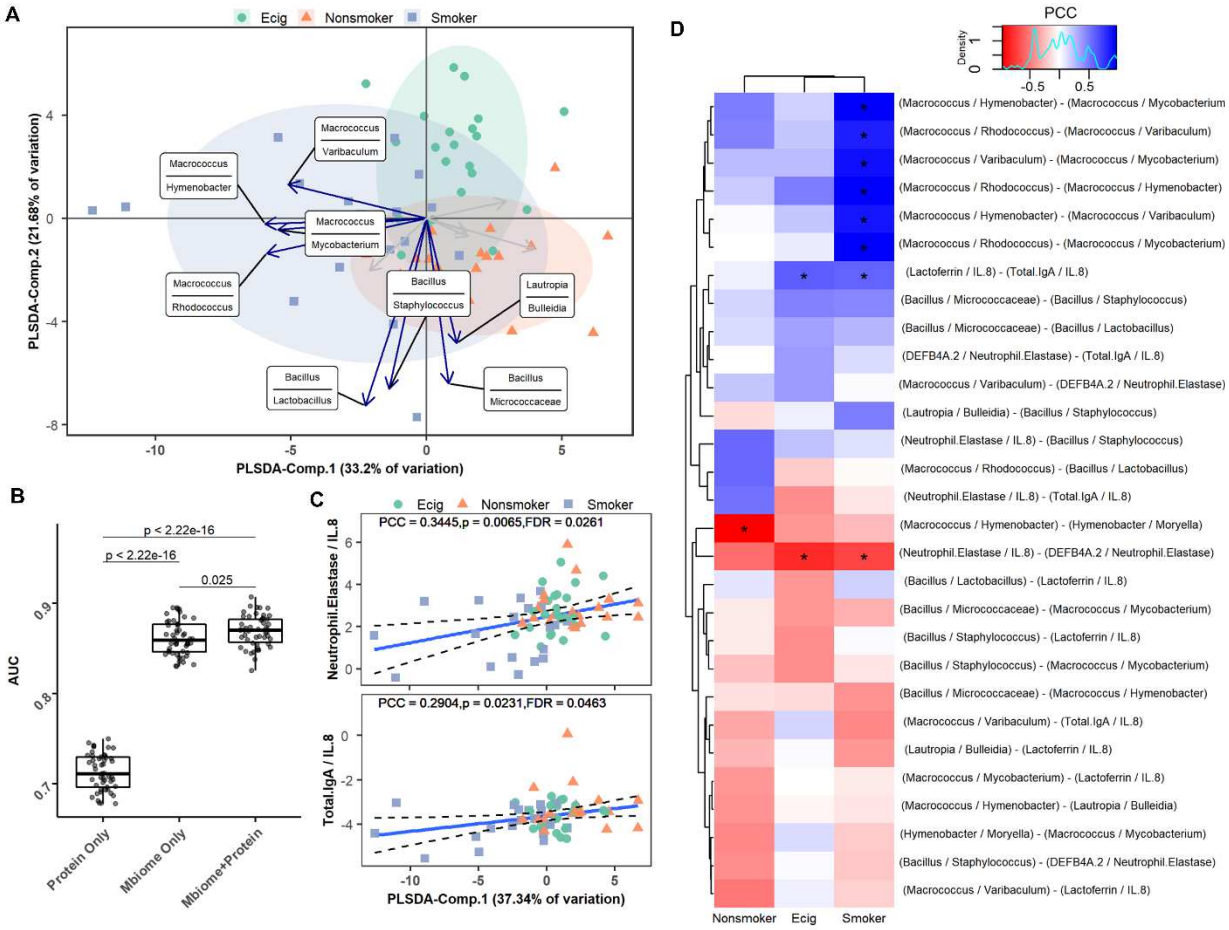

Figure 5 | Integrating data uncovers association between NLF mediators and nasal microbiome along with identifying distinct correlation patterns between exposure groups (Ecig-users: n=23; Nonsmokers: n=19; and Smokers: n=19). (A) PLS-DA biplot of integrated NLF mediators and nasal microbiome (B) Box and whisker's plot comparing area under the receiver operating characteristic curve performance of 2-component PLS-DA model (50x10-fold cross-validation) using each data type alone or integrated. (C) Scatter plot showing correlations between log ratios formed between concentrations$(\mu g/mL)$ of Lactoferrin, Neutrophil Elastase relative to IL-8 and the first PLS-DA component of the nasal microbiome. (D) Correlation heatmap showing Pearson's correlation coefficients (PCC) between and within the microbiome and protein log ratio signatures. ($*$ indicates within group $q \leq 0.10$)

References

1 Wang, T., Asman, K., Gentzke, A. & al., e. Tobacco Product Use Among Adults — United States, 2017. *MMWR Morb Mortal Wkly Rep* **67**, 1225–1232, doi: http://dx.doi.org/10.15585/mmwr.mm6744a2 (2017).

2 Leventhal, A. M. *et al.* Flavors of e-Cigarettes Used by Youths in the United States. *Jama*, doi:10.1001/jama.2019.17968 (2019).

3 Wang, T. W. *et al.* E-cigarette Use Among Middle and High School Students - United States, 2020. *MMWR Morb Mortal Wkly Rep* **69**, 1310-1312, doi:10.15585/mmwr.mm6937e1 (2020).

4 Cullen, K. A. *et al.* e-Cigarette Use Among Youth in the United States, 2019. *Jama*, doi:10.1001/jama.2019.18387 (2019).

5 Kiernan, E. *et al.* A Brief Overview of the National Outbreak of e-Cigarette, or Vaping, Product Use-Associated Lung Injury and the Primary Causes. *Chest* **159**, 426-431, doi:10.1016/j.chest.2020.07.068 (2021).

6 McAlinden, K. D. *et al.* COVID-19 and vaping: risk for increased susceptibility to SARS-CoV-2 infection? *Eur Respir J* **56**, doi:10.1183/13993003.01645-2020 (2020).

7 Martin, E. M. *et al.* E-cigarette use results in suppression of immune and inflammatory-response genes in nasal epithelial cells similar to cigarette smoke. *Am J Physiol Lung Cell Mol Physiol* **311**, L135-144, doi:10.1152/ajplung.00170.2016 (2016).

8 Reidel, B. *et al.* E-Cigarette Use Causes a Unique Innate Immune Response in the Lung, Involving Increased Neutrophilic Activation and Altered Mucin Secretion. *Am J Respir Crit Care Med* **197**, 492-501, doi:10.1164/rccm.201708-1590OC (2018).

9 Clapp, P. W. *et al.* Flavored e-cigarette liquids and cinnamaldehyde impair respiratory innate immune cell function. *Am J Physiol Lung Cell Mol Physiol* **313**, L278-L292, doi:10.1152/ajplung.00452.2016 (2017).

10 Clapp, P. W. *et al.* Cinnamaldehyde in flavored e-cigarette liquids temporarily suppresses bronchial epithelial cell ciliary motility by dysregulation of mitochondrial function. *Am J Physiol Lung Cell Mol Physiol* **316**, L470-l486, doi:10.1152/ajplung.00304.2018 (2019).

11 Madison, M. C. *et al.* Electronic cigarettes disrupt lung lipid homeostasis and innate immunity independent of nicotine. *J Clin Invest* **129**, 4290-4304, doi:10.1172/jci128531 (2019).

12 Ghosh, A. *et al.* Chronic E-Cigarette Use Increases Neutrophil Elastase and Matrix Metalloprotease Levels in the Lung. *Am J Respir Crit Care Med*, doi:10.1164/rccm.201903-0615OC (2019).

13 Sussan, T. E. *et al.* Exposure to electronic cigarettes impairs pulmonary anti-bacterial and anti-viral defenses in a mouse model. *PLoS One* **10**, e0116861, doi:10.1371/journal.pone.0116861 (2015).

14 Gerloff, J. *et al.* Inflammatory Response and Barrier Dysfunction by Different e-Cigarette Flavoring Chemicals Identified by Gas Chromatography-Mass Spectrometry in e-Liquids and e-Vapors on Human Lung Epithelial Cells and Fibroblasts. *Appl In Vitro Toxicol* **3**, 28-40, doi:10.1089/aivt.2016.0030 (2017).

15 Muthumalage, T. *et al.* Inflammatory and Oxidative Responses Induced by Exposure to Commonly Used e-Cigarette Flavoring Chemicals and Flavored e-Liquids without Nicotine. *Front Physiol* **8**, 1130, doi:10.3389/fphys.2017.01130 (2017).
16 Behar, R. Z., Luo, W., McWhirter, K. J., Pankow, J. F. & Talbot, P. Analytical and toxicological evaluation of flavor chemicals in electronic cigarette refill fluids. *Sci Rep* **8**, 8288, doi:10.1038/s41598-018-25575-6 (2018).
17 Clapp, P. W. *et al.* Cinnamaldehyde in Flavored E-Cigarette Liquids Temporarily Suppresses Bronchial Epithelial Cell Ciliary Motility by Dysregulation of Mitochondrial Function. *Am J Physiol Lung Cell Mol Physiol*, doi:10.1152/ajplung.00304.2018 (2019).
18 Hickman, E., Herrera, C. A. & Jaspers, I. Common E-Cigarette Flavoring Chemicals Impair Neutrophil Phagocytosis and Oxidative Burst. *Chemical research in toxicology* **32**, 982-985, doi:10.1021/acs.chemrestox.9b00171 (2019).
19 Hwang, J. H. *et al.* Electronic cigarette inhalation alters innate immunity and airway cytokines while increasing the virulence of colonizing bacteria. *J Mol Med (Berl)* **94**, 667-679, doi:10.1007/s00109-016-1378-3 (2016).
20 Miyashita, L. *et al.* E-cigarette vapour enhances pneumococcal adherence to airway epithelial cells. *Eur Respir J* **51**, doi:10.1183/13993003.01592-2017 (2018).
21 Man, W. H., de Steenhuijsen Piters, W. A. & Bogaert, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* **15**, 259-270, doi:10.1038/nrmicro.2017.14 (2017).
22 Ubags, N. D. J. & Marsland, B. J. Mechanistic insight into the function of the microbiome in lung diseases. *Eur Respir J* **50**, doi:10.1183/13993003.02467-2016 (2017).
23 Charlson, E. S. *et al.* Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One* **5**, e15216, doi:10.1371/journal.pone.0015216 (2010).
24 Ramakrishnan, V. R., Hauser, L. J. & Frank, D. N. The sinonasal bacterial microbiome in health and disease. *Curr Opin Otolaryngol Head Neck Surg* **24**, 20-25, doi:10.1097/moo.0000000000000221 (2016).
25 Grønseth, R. *et al.* Protected sampling is preferable in bronchoscopic studies of the airway microbiome. *ERJ Open Research* **3**, 00019-02017, doi:10.1183/23120541.00019-2017 (2017).
26 Di Stadio, A. *et al.* The Microbiota/Host Immune System Interaction in the Nose to Protect from COVID-19. *Life (Basel)* **10**, 345, doi:10.3390/life10120345 (2020).
27 Rosas-Salazar, C. *et al.* SARS-CoV-2 Infection and Viral Load are Associated with the Upper Respiratory Tract Microbiome. *J Allergy Clin Immunol*, doi:10.1016/j.jaci.2021.02.001 (2021).
28 Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nature Medicine* **24**, 392-400, doi:10.1038/nm.4517 (2018).
29 Li, H. in *Handbook of Statistical Genomics* 977-550 (2019).
30 Jiang, D. *et al.* Microbiome Multi-Omics Network Analysis: Statistical Considerations, Limitations, and Opportunities. *Frontiers in Genetics* **10**, doi:10.3389/fgene.2019.00995 (2019).
31 Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology* **8**, doi:10.3389/fmicb.2017.02224 (2017).

32 Hinton, A. L. & Mucha, P. J. A simultaneous feature selection and compositional association test for detecting sparse associations in high-dimensional metagenomic data (PREPRINT - Version 1). *PREPRINT Version 1 available at Research Square*, doi:https://doi.org/10.21203/rs.3.rs-703177/v1 (2021).
33 Rebuli, M. E., Speen, A. M., Clapp, P. W. & Jaspers, I. Novel applications for a noninvasive sampling method of the nasal mucosa. *Am J Physiol Lung Cell Mol Physiol* **312**, L288-l296, doi:10.1152/ajplung.00476.2016 (2017).
34 Muhlebach, M. S. *et al.* Initial acquisition and succession of the cystic fibrosis lung microbiome is associated with disease progression in infants and preschool children. *PLoS Pathog* **14**, e1006798, doi:10.1371/journal.ppat.1006798 (2018).
35 Horvath, K. M. *et al.* Nasal lavage natural killer cell function is suppressed in smokers after live attenuated influenza virus. *Respir Res* **12**, 102, doi:10.1186/1465-9921-12-102 (2011).
36 Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226, doi:10.1186/s40168-018-0605-2 (2018).
37 Drengenes, C. *et al.* Laboratory contamination in airway microbiome studies. *BMC Microbiol* **19**, 187, doi:10.1186/s12866-019-1560-1 (2019).
38 Quinn, T. P. *et al.* A field guide for the compositional analysis of any-omics data. *GigaScience* **8**, doi:10.1093/gigascience/giz107 (2019).
39 Tsilimigras, M. C. B. & Fodor, A. A. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology* **26**, 330-335, doi:10.1016/j.annepidem.2016.03.002 (2016).
40 Aitchison, J. The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)* **44**, 139-177 (1982).
41 Martın-Fernandez, J. A., Barcelo-Vidal, C. & Pawlowsky-Glahn, V. Dealing with Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation. *Mathematical Geology*, 26 (2003).
42 Legendre, P. & Legendre, L. in *Developments in Environmental Modelling* Vol. 24 625-710 (Elsevier, 2012).
43 Oksanen, J. *et al.* Vegan: Community Ecology Package. *R Package Version 2.2-1* **2**, 1-2 (2015).
44 Rizzo, M. L. & Székely, G. J. Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics* **8**, 27-38, doi:10.1002/wics.1375 (2016).
45 Greenacre, M. Variable Selection in Compositional Data Analysis Using Pairwise Logratios. *Mathematical Geosciences* **51**, 649-682, doi:10.1007/s11004-018-9754-x (2019).
46 Ernst, M. D. Permutation Methods: A Basis for Exact Inference. *Statistical Science* **19**, 676-685, doi:10.1214/088342304000000396 (2004).
47 Bastian, M., Heymann, S. & Jacomy, M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. (2009).
48 Csardi, G. & Nepusz, T. The Igraph Software Package for Complex Network Research. *InterJournal* **Complex Systems**, 1695 (2005).

49 Anderson, M. J. in *Wiley StatsRef: Statistics Reference Online* 1-15 (American Cancer Society, 2017).

50 Barker, M. & Rayens, W. Partial least squares for discrimination. *Journal of Chemometrics* **17**, 166-173, doi:https://doi.org/10.1002/cem.785 (2003).

51 Brereton, R. G. & Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics* **28**, 213-225, doi:https://doi.org/10.1002/cem.2609 (2014).

52 Kalivodová, A. *et al.* PLS-DA for compositional data with application to metabolomics. *Journal of Chemometrics* **29**, 21-28, doi:https://doi.org/10.1002/cem.2657 (2015).

53 Boulesteix, A.-L. & Strimmer, K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics* **8**, 32-44, doi:10.1093/bib/bbl016 (2006).

54 Kuhn, M. Building Predictive Models in *R* Using the **caret** Package. *Journal of Statistical Software* **28**, doi:10.18637/jss.v028.i05 (2008).

55 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 2 edn, (Springer International Publishing, 2016).

56 Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861-874, doi:10.1016/j.patrec.2005.10.010 (2006).

57 Hand, D. J. & Till, R. J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning* **45**, 171-186, doi:10.1023/A:1010920819831 (2001).

58 Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**, 111-133, doi:https://doi.org/10.1111/j.2517-6161.1974.tb00994.x (1974).

59 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).

60 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).

61 Kumpitsch, C., Koskinen, K., Schöpf, V. & Moissl-Eichinger, C. The microbiome of the upper respiratory tract in health and disease. *BMC Biol* **17**, 87, doi:10.1186/s12915-019-0703-z (2019).

62 De Boeck, I. *et al.* Comparing the Healthy Nose and Nasopharynx Microbiota Reveals Continuity As Well As Niche-Specificity. *Frontiers in microbiology* **8**, 2372-2372, doi:10.3389/fmicb.2017.02372 (2017).

63 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217, doi:10.1371/journal.pone.0061217 (2013).

64 Rebuli, M. E. *et al.* Wood Smoke Exposure Alters Human Inflammatory Responses to Viral Infection in a Sex-Specific Manner: A Randomized, Placebo-Controlled Study. *Am J Respir Crit Care Med*, doi:10.1164/rccm.201807-1287OC (2018).

65 Cho, Y. *et al.* Sex Differences in Pulmonary Responses to Ozone in Mice. Role of the Microbiome. *Am J Respir Cell Mol Biol* **60**, 198-208, doi:10.1165/rcmb.2018-0099OC (2019).

66 Yu, G. *et al.* The effect of cigarette smoking on the oral and nasal microbiota. *Microbiome* **5**, 3, doi:10.1186/s40168-016-0226-6 (2017).
67 Piewngam, P. *et al.* Pathogen elimination by probiotic Bacillus via signalling interference. *Nature* **562**, 532-537, doi:10.1038/s41586-018-0616-y (2018).
68 Liu, C. M. *et al.* Staphylococcus aureus and the ecology of the nasal microbiome. *Science Advances* **1**, e1400216, doi:10.1126/sciadv.1400216 (2015).
69 Sakr, A., Brégeon, F., Mège, J.-L., Rolain, J.-M. & Blin, O. Staphylococcus aureus Nasal Colonization: An Update on Mechanisms, Epidemiology, Risk Factors, and Subsequent Infections. *Frontiers in microbiology* **9**, 2419-2419, doi:10.3389/fmicb.2018.02419 (2018).
70 Actor, J. K., Hwang, S. A. & Kruzel, M. L. Lactoferrin as a natural immune modulator. *Curr Pharm Des* **15**, 1956-1973, doi:10.2174/138161209788453202 (2009).
71 Casimir, G. J., Lefevre, N., Corazza, F. & Duchateau, J. Sex and inflammation in respiratory diseases: a clinical viewpoint. *Biology of sex differences* **4**, 16, doi:10.1186/2042-6410-4-16 (2013).
72 Man, W. H. *et al.* Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study. *Lancet Respir Med* **7**, 417-426, doi:10.1016/s2213-2600(18)30449-1 (2019).
73 Neumann, A., Björck, L. & Frick, I.-M. Finegoldia magna, an Anaerobic Gram-Positive Bacterium of the Normal Human Microbiota, Induces Inflammation by Activating Neutrophils. *Frontiers in Microbiology* **11**, doi:10.3389/fmicb.2020.00065 (2020).
74 Hoggard, M. *et al.* Chronic Rhinosinusitis and the Evolving Understanding of Microbial Ecology in Chronic Inflammatory Mucosal Disease. *Clinical Microbiology Reviews* **30**, 321, doi:10.1128/CMR.00060-16 (2017).
75 Chiu, C.-Y. *et al.* Cross-talk between airway and gut microbiome links to IgE responses to house dust mites in childhood airway allergies. *Scientific Reports* **10**, 13449, doi:10.1038/s41598-020-70528-7 (2020).
76 Bacci, G. *et al.* Pyrosequencing Unveils Cystic Fibrosis Lung Microbiome Differences Associated with a Severe Lung Function Decline. *PLoS One* **11**, e0156807, doi:10.1371/journal.pone.0156807 (2016).
77 Zanger, U. M. & Schwab, M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics* **138**, 103-141, doi:https://doi.org/10.1016/j.pharmthera.2012.12.007 (2013).
78 Vemuri, R. *et al.* The microgenderome revealed: sex differences in bidirectional interactions between the microbiota, hormones, immunity and disease susceptibility. *Semin Immunopathol* **41**, 265-275, doi:10.1007/s00281-018-0716-7 (2019).
79 Kim, Y. S., Unno, T., Kim, B. Y. & Park, M. S. Sex Differences in Gut Microbiota. *World J Mens Health* **38**, 48-60, doi:10.5534/wjmh.190009 (2020).
80 Shin, J. H. *et al.* Serum level of sex steroid hormone is associated with diversity and profiles of human gut microbiome. *Res Microbiol* **170**, 192-201, doi:10.1016/j.resmic.2019.03.003 (2019).
81 Han, M. K. *et al.* Female Sex and Gender in Lung/Sleep Health and Disease. Increased Understanding of Basic Biological, Pathophysiological, and Behavioral Mechanisms

Leading to Better Health for Female Patients with Lung Disease. *Am J Respir Crit Care Med* **198**, 850-858, doi:10.1164/rccm.201801-0168WS (2018).

82 De Boeck, I. *et al.* Anterior Nares Diversity and Pathobionts Represent Sinus Microbiome in Chronic Rhinosinusitis. *mSphere* **4**, e00532-00519, doi:10.1128/mSphere.00532-19 (2019).

83 Chen, Y. W. *et al.* Fine Particulate Matter Exposure Alters Pulmonary Microbiota Composition and Aggravates Pneumococcus-Induced Lung Pathogenesis. *Front Cell Dev Biol* **8**, 570484, doi:10.3389/fcell.2020.570484 (2020).

84 Mariani, J. *et al.* Short-term particulate matter exposure influences nasal microbiota in a population of healthy subjects. *Environmental research* **162**, 119-126, doi:10.1016/j.envres.2017.12.016 (2018).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 20210716MicrobiomePaperFinalSupplement.pdf