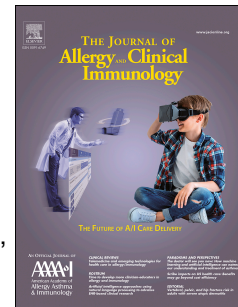


Journal Pre-proof

The nasal microbiome, nasal transcriptome, and pet sensitization

Yoojin Chun, MS, Anh Do, PhD, Galina Grishina, MS, Zoe Ardit, BS, Victoria Ribeiro, BS, Alexander Grishin, PhD, Alfin Vicencio, MD, Supinda Bunyavanich, MD, MPH



PII: S0091-6749(21)00217-7

DOI: <https://doi.org/10.1016/j.jaci.2021.01.031>

Reference: YMAI 14971

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 29 September 2020

Revised Date: 24 December 2020

Accepted Date: 29 January 2021

Please cite this article as: Chun Y, Do A, Grishina G, Ardit Z, Ribeiro V, Grishin A, Vicencio A, Bunyavanich S, The nasal microbiome, nasal transcriptome, and pet sensitization, *Journal of Allergy and Clinical Immunology* (2021), doi: <https://doi.org/10.1016/j.jaci.2021.01.031>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology.

The nasal microbiome, nasal transcriptome, and pet sensitization

Yoojin Chun, MS¹, Anh Do, PhD¹, Galina Grishina, MS², Zoe Ardit, BS¹, Victoria Ribeiro, BS¹, Alexander Grishin, PhD², Alfin Vicencio, MD³, Supinda Bunyavanich, MD, MPH^{1,2*}

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA

²Division of Allergy and Immunology, Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA

³Division of Pulmonary Medicine, Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA

***Corresponding Author:** Supinda Bunyavanich, MD, MPH, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue #1498, New York, NY 10029 USA, Tel. +1.212.659.8262 | Fax +1.212.426.1902, supinda@post.harvard.edu

Funding: US National Institutes of Health R01 AI118833

Declaration of interests: The authors declare that no conflicts of interest exist.

Key Messages

- Cat sensitization is associated with lower nasal microbial diversity and differential nasal bacterial composition.

- Nasal abundances of *Corynebacterium sp.* and *Staphylococcus epidermidis* are associated with absence of cat and/or dog sensitization
- These protective associations are causally mediated by reduced nasal expression of genes related to IgE-hypersensitivity and mast cell function.

Capsule Summary: Nasal microbiota and nasal gene expression vary by pet sensitization status. Higher abundances of nasal *Corynebacterium sp.* and *Staphylococcus epidermidis* are associated with absence of pet sensitization and correlate with lower host expression of inflammatory genes.

Key words: allergen sensitization, cat, dog, pet, nasal microbiome, nasal transcriptome, gene expression, fatty acid, mitochondria, *Corynebacterium sp.*, *Staphylococcus epidermidis*

Abbreviations:

DEG: differentially expressed gene

FDR: false discovery rate

ICS/LABA: inhaled corticosteroid/long-acting beta agonist

LDA: linear discriminant analysis

LEfSe: Linear discriminant effect size

PERMANOVA: permutational multivariate analysis of variance

sIgE: specific IgE

Treg: regulatory T cell

Abstract

Background: Pet allergies are common in children with asthma. Microbiota and host responses may mediate allergen sensitization.

Objective: To uncover host-microbe relationships in pet allergen sensitization via joint examination of the nasal microbiome and nasal transcriptome.

Methods: We collected nasal samples from 132 children with asthma for parallel 16S rRNA and RNA sequencing. Specific IgE levels for cat and dog dander were measured. Analyses of the nasal microbiome, nasal transcriptome, and their correlations were performed with respect to pet sensitization status.

Results: Among the 132 children, 91 (68.9%) were cat sensitized and 96 (72.7%) were dog sensitized. Cat sensitization was associated with lower nasal microbial diversity by Shannon Index ($P=0.021$) and differential nasal bacterial composition by weighted UniFrac distance (PERMANOVA $P=0.035$). *Corynebacterium sp.* and *Staphylococcus epidermidis* were significantly less abundant, and the metabolic process “fatty acid elongation in mitochondria” was lower in pet sensitized vs. unsensitized children. Correlation networks revealed that the nasal expression levels of 47 genes representing inflammatory processes were negatively correlated with the relative abundances of *Corynebacterium sp.* and *Staphylococcus epidermidis*. Thus, these species were not only directly associated with the absence of pet sensitization, but also with the under-expression of host gene expression of inflammatory processes that contribute to allergen sensitization. Causal mediation analyses revealed that the associations between these nasal species and pet sensitization were mediated by nasal gene expression.

71 **Conclusion:** Higher abundances of nasal *Corynebacterium sp.* and *Staphylococcus*
72 *epidermidis* are associated with absence of pet sensitization and correlate with lower
73 expression of inflammatory genes.

Introduction

Allergies to cat and dog dander are common, and exposure to these pet allergens often triggers asthma exacerbations in individuals with asthma.^{1, 2} Pet exposure is associated with risk of allergen sensitization, and this is thought to be partly mediated by microbiota.^{2, 3} While many studies have reported associations between gut microbiota and allergen sensitization^{4, 5}, including to pet allergens^{6, 7}, fewer studies have specifically examined interactions between airway microbiota and allergen sensitization.⁸ In particular, given the nasal passages are the first point of contact between inhaled allergens and the airway, examination of nasal microbiota with respect to allergen sensitization and nasal gene expression could reveal novel insight into host-microbe relationships in allergen sensitization. We hypothesized that multi-omic examination of subjects with and without pet sensitization could reveal novel associations and mediation relationships between nasal microbiota, nasal gene expression, and allergen sensitization to cat and/or dog.

Results and Discussion

Detailed methods are provided in the Online Supplement.

Study population and design

Children with persistent asthma based on National Heart Lung Blood Institute Expert Panel Report 3 guidelines⁹ were recruited from the Mount Sinai Health System, New York, NY, USA. The study was approved by the Mount Sinai Institutional Review Board, and written informed consent was obtained from participants' parents. The

children (median age=12.0 years, interquartile range=5 years) included those with mild/moderate (74%) and severe (26%) persistent asthma (**Table S1**). Asthma control varied among the participants with mean Asthma Control Test (ACT) score 16.6 (SD 4.3). Among the children, asthma medication use included short-acting beta agonist (92%), inhaled corticosteroid (ICS) (23%), combined ICS/long-acting beta agonist (LABA) (21%), leukotriene receptor agonist (21%), and omalizumab (2%). Additional characteristics are shown in **Table S1**.

Allergen sensitization to cat and dog dander was determined by serum specific IgE (sIgE) ≥ 0.10 kU_A/L (ImmunoCAP, ThermoScientific). Among the 132 children, 91 (68.9%) were cat sensitized and 96 (72.7%) were dog sensitized. Compared to their unsensitized counterparts, there were no differences in age, sex, race, asthma severity, ACT score, last upper respiratory infection, last antibiotic use, or asthma medication use except for ICS/LABA in subjects sensitized to cat or dog (**Table S1**). Pet ownership was not associated with sensitization to the respective pet, although dog ownership was associated with cat sensitization.

All subjects underwent nasal swab of one nare and nasal brushing of the contralateral nare. The swab and brush samples respectively underwent DNA and RNA isolation, 16S rRNA and RNA sequencing, sequence data processing, quality control, and analysis using previously described methods.¹⁰ Briefly, nasal microbiome profiling of isolated DNA was performed with dual indices and Illumina sequencing adapters attached to the V3V4 amplicon using Nextera XT Index Kit v2 (Illumina, San Diego, CA), with sequencing of amplified 16S rRNA on the Illumina MiSeq platform with 2X250bp paired-end reads; except where indicated, analyses with amplicon sequence variants (ASVs)¹¹ were performed using Qiime 2¹² with Silva reference data version 138¹³. Nasal

RNA sequencing involved library preparation with the TruSeq RNA Sample Prep Kit v2 protocol (Illumina, San Diego, CA) and sequencing on the Illumina HiSeq 2500 platform with a per-sample target of 40-50 million 100bp paired-end reads. Following data processing and quality control of the above data, there were 19,269 ASVs and 17,337 genes for analysis.

Cat sensitization is associated with lower nasal microbial diversity and differential nasal bacterial composition

Alpha diversity measured by Shannon index, a measure of the richness and evenness of species at a site, was significantly lower in children sensitized (n=91) vs. unsensitized (n=41) to cat (P=0.021, **Figure 1A**). Although children sensitized to dog (n=96) showed a similar trend of lower nasal bacterial diversity relative to those without dog sensitization (n=36), this difference was not significant (**Figure 1B**). Comparison of nasal microbial composition showed compositional differences associated with cat sensitization (weighted UniFrac PERMANOVA P=0.035, pseudo-F = 2.8; Jaccard PERMANOVA P=0.016, pseudo-F=1.3) but not dog sensitization (weighted UniFrac PERMANOVA P=0.23, pseudo-F=1.3; Jaccard PERMANOVA P=0.46, pseudo-F=0.98) (**Figure 1C, 1D, Figure S1**).

Linear discriminant effect size (LEfSe) analysis¹⁴, a method for biomarker discovery, was then used to identify species with relative abundance $\geq 1\%$ that best characterize cat and dog sensitization. LEfSe revealed that an unnamed *Corynebacterium* sp. was significantly associated with absence of sensitization to cat (linear discriminant analysis (LDA) score=5.4, P=0.011) and dog (LDA score=5.0, P=0.043). *Staphylococcus epidermidis* was also associated with absence of

sensitization to cat (LDA score=5.1, $P=0.0045$). No other species showed significant results. *Corynebacterium spp.* have been previously associated with reduced risk of airway diseases.^{10, 15} *Staphylococcus epidermidis* is a commensal bacterium associated with healthy maturation of the nasal microbiome and the production of beneficial antimicrobial peptides.¹⁶

Sensitivity analyses with LEfSE adjusted for age, sex, ICS/LABA use, and cat or dog ownership revealed no significant changes to these findings (Online Supplement). To test the specificity of our findings to pet sensitization, we used the same metrics to compare children in the cohort with and without sensitization to pollen (tree, grass, or weed), finding no significant differences by Shannon Index, weighted UniFrac, or LEfSe.

Potential functional implications of nasal microbiota associated with pet sensitization

To explore the functional implications of our findings, we pursued two paths. In the first, we applied metagenomic inference using PICRUST¹⁷ and STAMP¹⁸ to infer microbial functions associated with the compositional differences in nasal bacteria associated with pet sensitization. Metagenomic inference showed that compared to their unsensitized counterparts, the metabolic process “fatty acid elongation in mitochondria” was lower in children sensitized to cat (FDR = $7.5E-03$) (**Figure 2A**) and dog (FDR = $1.6E-02$) (**Figure 2B**). Others have shown that altered lipid metabolism in mitochondria modulates regulatory T cell (Treg) suppressive capacity.¹⁹ Experimental blockade of fatty acid binding proteins leads to decreased fatty acid elongation in mitochondria and perturbation of Treg function.¹⁹ Given the role of Tregs in mediating tolerance against airborne antigens²⁰, our finding of reduced Treg-associated mitochondrial fatty acid processes in pet sensitized children raises the possibility that nasal microbiota could

potentiate allergen sensitization via downregulation of this metabolic process. These metagenomic inference findings merit direct validation.

The second path of our functional investigation aimed to explore the potential effects of *Corynebacterium sp.* and *Staphylococcus epidermidis* on host from the lens of host gene expression and as related to pet sensitization. We first identified nasal genes differentially expressed based on cat sensitization, finding 58 differentially expressed genes (DEGs) at $FDR \leq 0.05$ and $|\log_2 \text{fold change}| > 1$ (**Figure 3A**); DEG analysis was performed using R packages *voom*²¹ and *limma*¹⁰. We then analogously identified nasal genes differentially expressed based on dog sensitization, finding 20 dog DEGs (**Figure 3B**). Weighted gene coexpression network analysis²² showed that most of these DEGs belonged to two gene coexpression modules representing inflammatory response and allergy (**Figure 3A, 3B**) (Online supplement). Fourteen of the cat and dog DEGs overlapped with one another (**Figure 3C**). There was no significant change in DEGs when analyses were adjusted for age, sex, ICS/LABA use, and cat or dog ownership (Online Supplement).

Of the 64 genes associated with cat and/or dog sensitization, the expression levels of 47 DEGs were each significantly correlated (Spearman correlation $FDR \leq 0.05$) with the relative abundance of *Corynebacterium sp.* and/or *Staphylococcus epidermidis* (**Figure 4**). Each of these DEGs has been previously associated with at least one outcome related to allergen sensitization, including IgE and IgE-mediated hypersensitivity (*MS4A2*, *GATA2*, *CEBPE*, *PCSK6*, *GCNT4*), Type 2 inflammation (*POSTN*, *B3GNT6*, *GCSAML*), mast cell function (*ADCYAP1*, *CPA3*, *TPSAB1*, *TPSB2*, *TPSD1*, *SIGLEC6*, *SIGLEC8*, *HDC*), eosinophil activity (*CAPN14*, *GATA1*, *GATA2*, *CCL24*, *CLC*), lung function (*SNTG2*, *DLC1*, *DPP4*), and mucosal inflammation

(*SLC9A3*, *CDH26*, *SCGB3A1*, *ALOX15*)²³⁻²⁵. Except for *SCGB3A1*, *Corynebacterium sp.* and *Staphylococcus epidermidis* abundances were negatively correlated ($FDR \leq 0.05$) with the expression of each of these DEGs associated with pet sensitization (**Figure 4**); that is, higher abundances of these two species were associated with lower levels of pet sensitization gene expression.

As LEfSe analysis had shown that *Corynebacterium sp.* and *Staphylococcus epidermidis* were directly associated with protection from sensitization to pet allergens, we found reassuring consistency from the correlation network that these nasal species also had near uniform negative correlation with the nasal expression of the sensitization-associated DEGs. In line with prior observations that *Corynebacterium spp.* are associated with reduced risk of airway disease^{10, 15} and *Staphylococcus epidermidis* promotes a healthy nasal microbiome,¹⁶ we found that both nasal species were not only associated with the absence of pet sensitization in children with asthma, but also with the under-expression of genes associated with inflammatory processes that contribute to allergen sensitization.

Nasal gene expression mediates associations between nasal microbiota and pet sensitization

To further examine relationships between *Corynebacterium sp.*, *Staphylococcus epidermidis*, and DEGs associated with pet sensitization, we performed causal mediation analysis. The abundances of *Corynebacterium sp.* and *Staphylococcus epidermidis* were inversely associated with cat and/or dog sensitization, and these associations were causally mediated by reduced expression of 7 genes related to IgE-mediated hypersensitivity and mast cell function ($FDR \leq 0.05$, **Figure 5**). Interestingly, we

218 did not find that associations between DEGs and pet sensitization were mediated by
219 *Corynebacterium sp.* or *Staphylococcus epidermidis*.

221 Conclusions

222 To our knowledge, this is the first study to jointly examine the nasal microbiome,
223 nasal transcriptome, and their relationship to pet allergen sensitization in children with
224 asthma. The limitations of this study include its observational, cross-sectional design
225 and the possibility of residual confounding. We found that the nasal microbiome and
226 nasal transcriptome of children with asthma vary by pet sensitization. Higher abundance
227 of nasal *Corynebacterium sp.* and *Staphylococcus epidermidis* in particular were
228 associated with the absence of pet sensitization and correlated with lower host
229 expression of inflammatory genes, supporting that these nasal species are associated
230 with allergy health. Causal mediation analysis identified several DEGs that mediate
231 associations between *Corynebacterium sp.*, *S. epidermidis*, and pet sensitization. Our
232 study provides novel insight into airway host-microbe relationships in pet allergen
233 sensitization.

References

1. Gergen PJ, Mitchell HE, Calatroni A, Sever ML, Cohn RD, Salo PM, et al. Sensitization and Exposure to Pets: The Effect on Asthma Morbidity in the US Population. *J Allergy Clin Immunol Pract* 2018; 6:101-7 e2.
2. Mendy A, Wilkerson J, Salo PM, Cohn RD, Zeldin DC, Thorne PS. Exposure and Sensitization to Pets Modify Endotoxin Association with Asthma and Wheeze. *J Allergy Clin Immunol Pract* 2018; 6:2006-13 e4.
3. Kim H, Sitarik AR, Woodcroft K, Johnson CC, Zoratti E. Birth Mode, Breastfeeding, Pet Exposure, and Antibiotic Use: Associations With the Gut Microbiome and Sensitization in Children. *Curr Allergy Asthma Rep* 2019; 19:22.
4. Zimmermann P, Messina N, Mohn WW, Finlay BB, Curtis N. Association between the intestinal microbiota and allergic sensitization, eczema, and asthma: A systematic review. *J Allergy Clin Immunol* 2019; 143:467-85.
5. Cait A, Cardenas E, Dimitriu PA, Amenyogbe N, Dai D, Cait J, et al. Reduced genetic potential for butyrate fermentation in the gut microbiome of infants who develop allergic sensitization. *J Allergy Clin Immunol* 2019; 144:1638-47 e3.
6. Sbihi H, Boutin RC, Cutler C, Suen M, Finlay BB, Turvey SE. Thinking bigger: How early-life environmental exposures shape the gut microbiome and influence the development of asthma and allergic disease. *Allergy* 2019; 74:2103-15.
7. Tun HM, Konya T, Takaro TK, Brook JR, Chari R, Field CJ, et al. Exposure to household furry pets influences the gut microbiota of infant at 3-4 months following various birth scenarios. *Microbiome* 2017; 5:40.
8. Anderson HM, Jackson DJ. Microbes, allergic sensitization, and the natural history of asthma. *Curr Opin Allergy Clin Immunol* 2017; 17:116-22.
9. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington DC: National Heart Lung and Blood Institute and National Asthma Education and Prevention Program; 2007.
10. Chun Y, Do A, Grishina G, Grishin A, Fang G, Rose S, et al. Integrative study of the upper and lower airway microbiome and transcriptome in asthma. *JCI Insight* 2020; 5:e133707.
11. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; 13:581-3.
12. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; 37:852-7.
13. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; 41:D590-6.
14. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011; 12:R60.
15. Man WH, de Steenhuijsen P, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 2017; 15:259-70.
16. Liu Q, Liu Q, Meng H, Lv H, Liu Y, Liu J, et al. *Staphylococcus epidermidis* Contributes to Healthy Maturation of the Nasal Microbiome by Stimulating Antimicrobial Peptide Production. *Cell Host Microbe* 2020; 27:68-78 e5.
17. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013; 31:814-21.

- 283 18. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of
284 taxonomic and functional profiles. *Bioinformatics* 2014; 30:3123-4.
- 285 19. Field CS, Baixauli F, Kyle RL, Puleston DJ, Cameron AM, Sanin DE, et al. Mitochondrial
286 Integrity Regulated by Lipid Metabolism Is a Cell-Intrinsic Checkpoint for Treg
287 Suppressive Function. *Cell Metab* 2020; 31:422-37 e5.
- 288 20. Bacher P, Scheffold A. The effect of regulatory T cells on tolerance to airborne allergens
289 and allergen immunotherapy. *J Allergy Clin Immunol* 2018; 142:1697-709.
- 290 21. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model
291 analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15:R29.
- 292 22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
293 analysis. *BMC Bioinformatics* 2008; 9:559.
- 294 23. Gene, National Center for Biotechnology Information (NCBI). National Library of
295 Medicine 2020:Available from <https://www.ncbi.nlm.nih.gov/gene/>, accessed 10
296 December 2020.
- 297 24. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
298 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted
299 arrays and summary statistics 2019. *Nucleic Acids Res* 2019; 47:D1005-D12.
- 300 25. PubMed, National Center for Biotechnology Information (NCBI). National Library of
301 Medicine 2020:Available from <https://pubmed.ncbi.nlm.nih.gov.eresources.mssm.edu/>,
302 accessed 10 December 2020.

303

Figure Legends

Figure 1: Alpha diversity and relative abundances of nasal bacterial species in

children sensitized and unsensitized to cat and dog. Alpha diversity of nasal bacterial species was measured by Shannon index at a rarefaction depth of 4,980.

Alpha diversity comparison by Kruskal-Wallis test between **(A)** children sensitized (Cat+, n=91) and unsensitized (Cat-, n=41) to cat, and **(B)** children sensitized (Dog+, n=96) and unsensitized (Dog-, n=36) to dog. Relative abundance of nasal bacterial species in **(C)** children sensitized (Cat+, n=91) and unsensitized (Cat-, n=41) to cat, and **(D)** children sensitized (Dog+, n=96) and unsensitized (Dog-, n=36) to dog. Unnamed species are indicated with “sp.”. Species with relative abundance less than 1% were combined as “Other”.

Figure 2: Fatty acid elongation in mitochondria of children sensitized and

unsensitized to cat and dog. Comparison of fatty acid elongation in mitochondria, a metagenomically inferred metabolic process, between **(A)** children sensitized (Cat+, n=91) and unsensitized (Cat-, n=41) to cat, and **(B)** children sensitized (Dog+, n=96) and unsensitized (Dog-, n=36) to dog. FDR values from White’s non-parametric t-tests are shown.

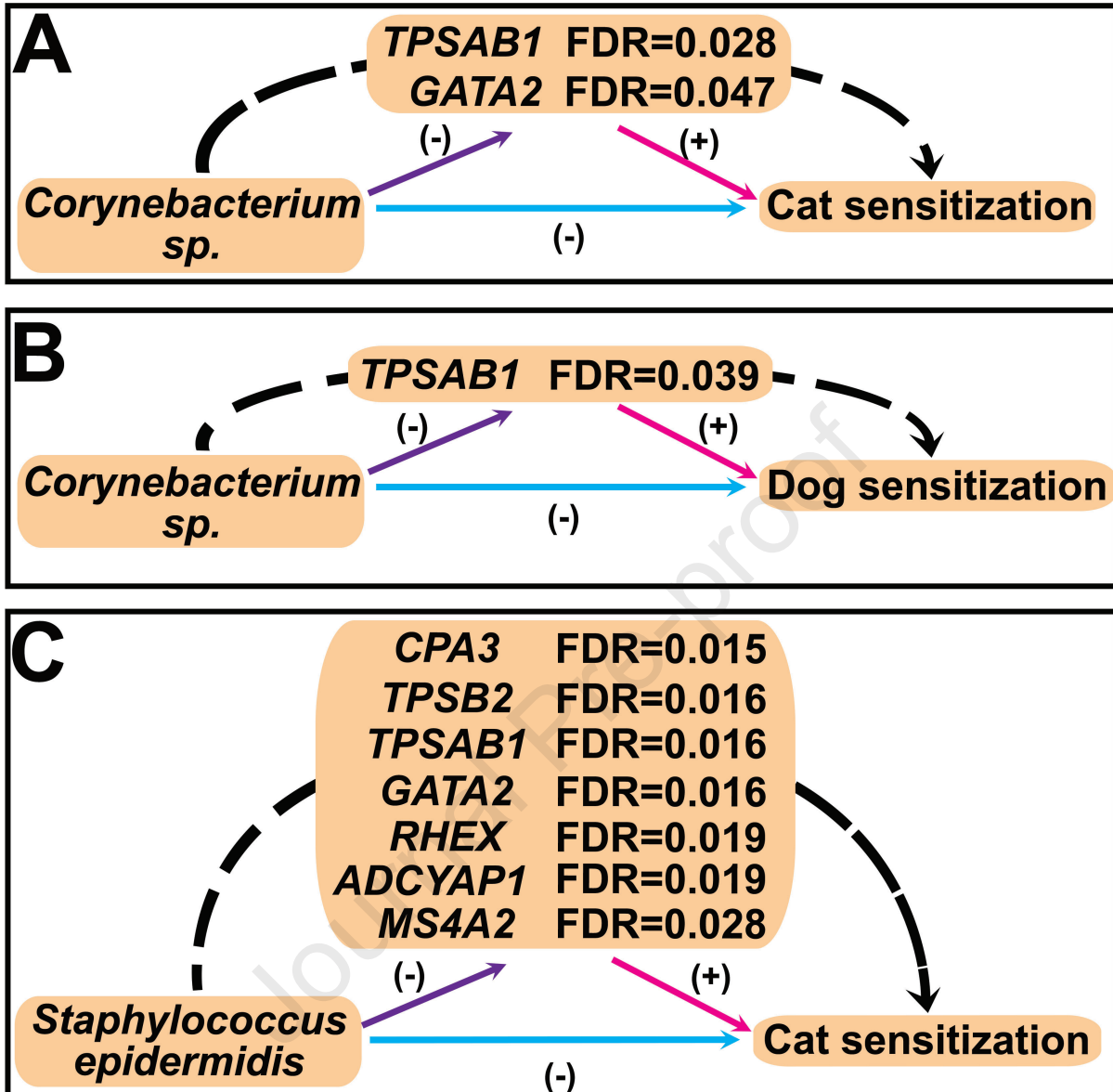
Figure 3: Differentially expressed nasal genes of children sensitized and

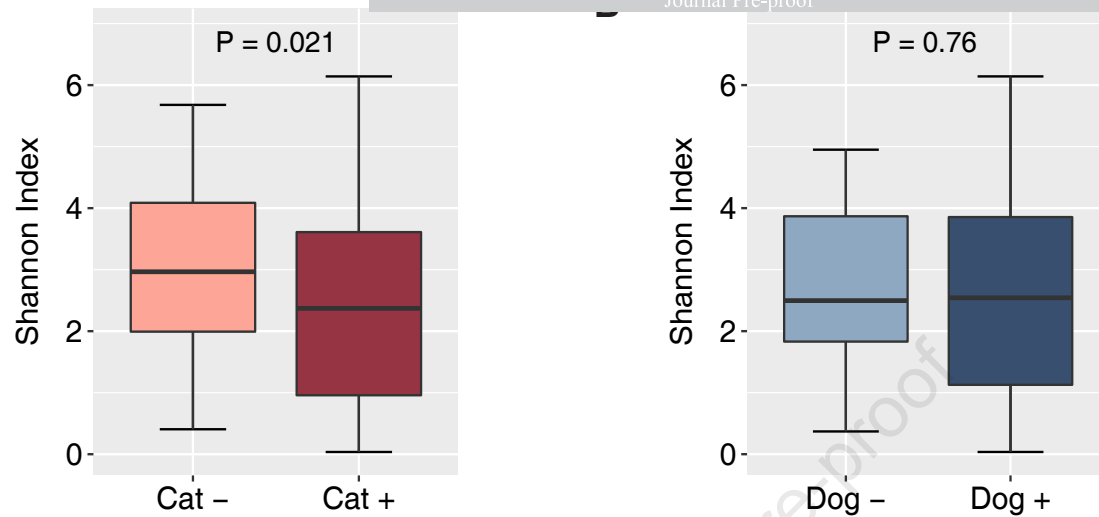
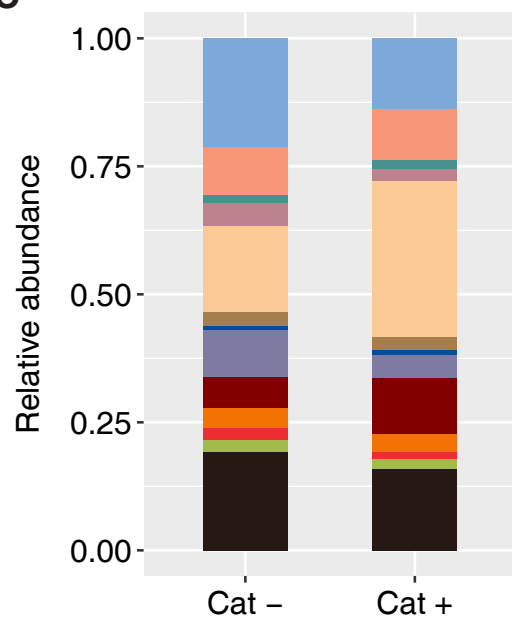
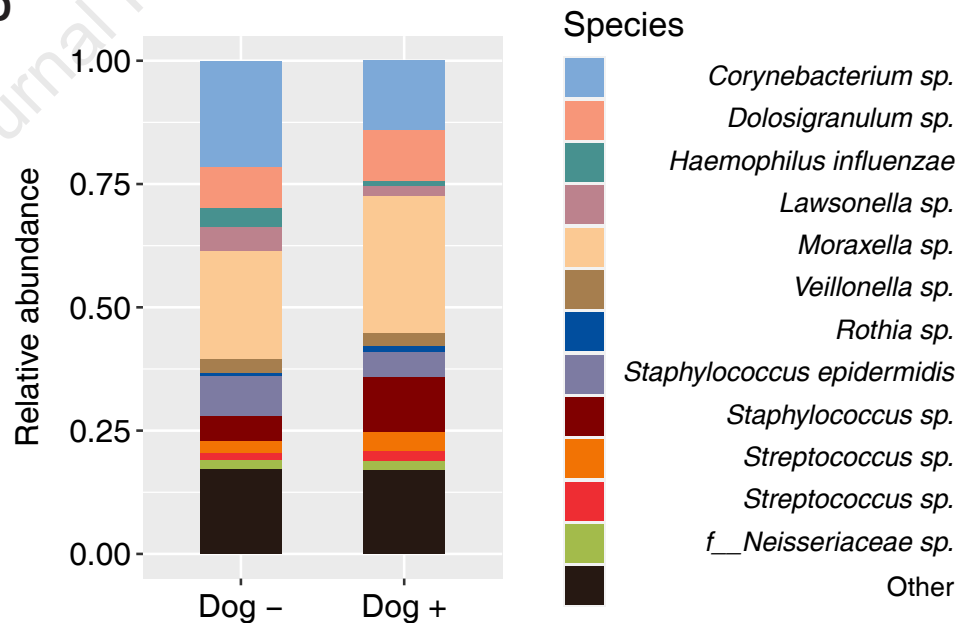
unsensitized to cat and dog. Differentially expressed nasal genes (rows) associated with **(A)** cat sensitization and **(B)** dog sensitization at $FDR \leq 0.05$ and $|\log_2 \text{fold change}| > 1$ across subjects (columns). Green = higher expression in sensitized subjects. Cyan =

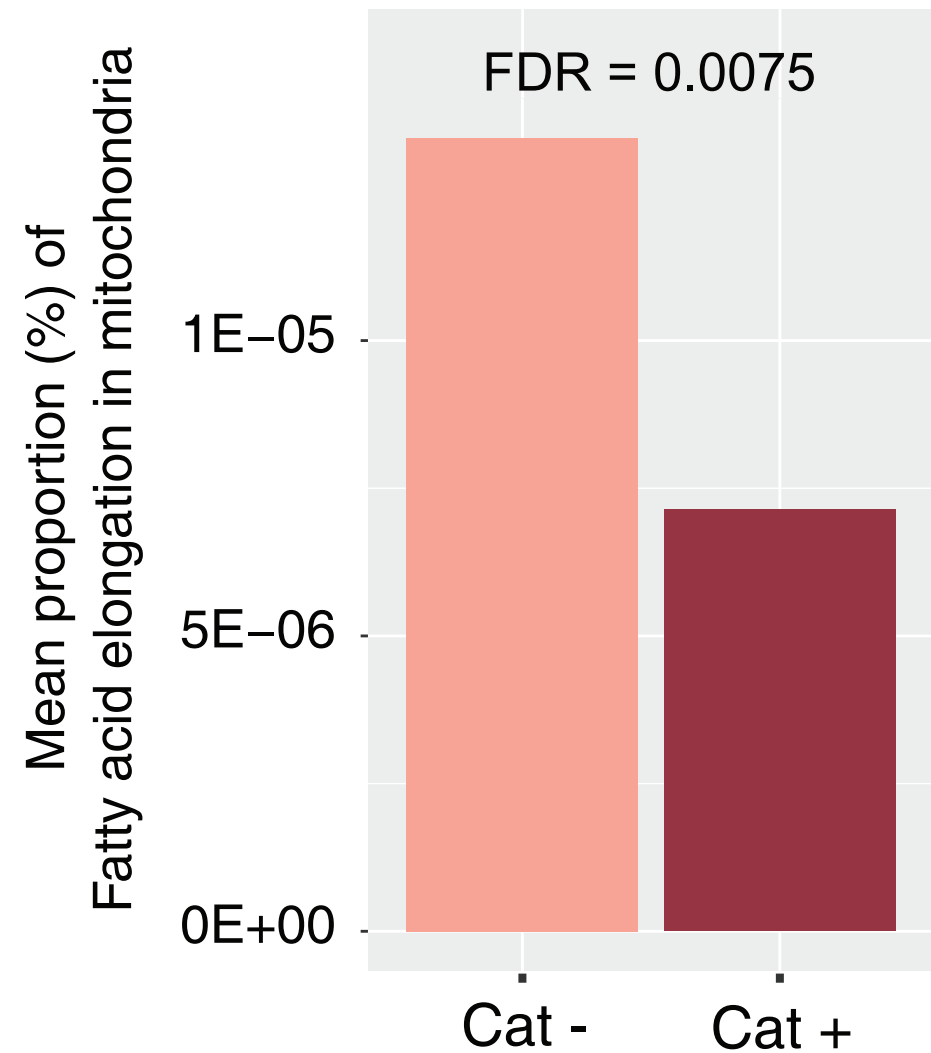
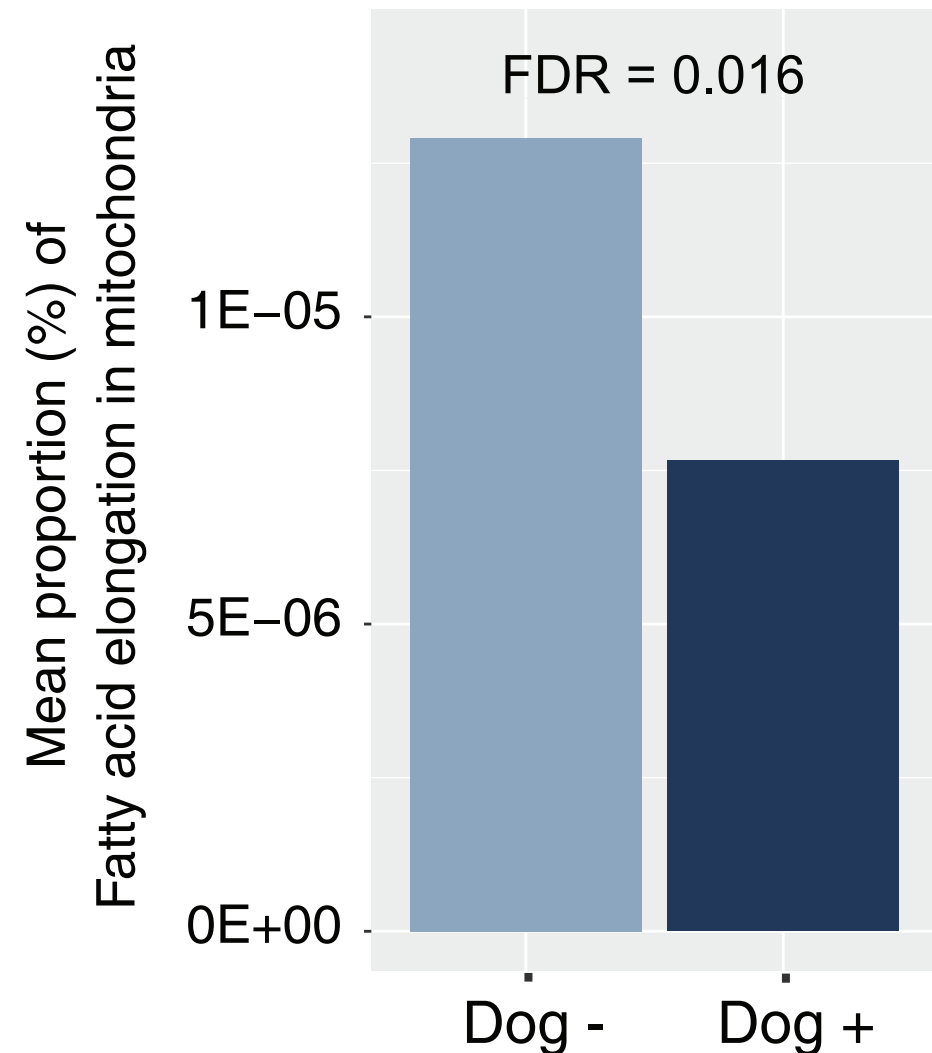
higher expression in unsensitized subjects. Gene expression values were centered and scaled by gene. Hierarchical clustering was performed based on complete linkage. Genes in magenta and purple fonts are members of the inflammatory response and allergy coexpression modules, respectively, from weighted gene coexpression network analysis. **(C)** Venn diagram for differentially expressed genes associated with cat sensitization and dog sensitization.

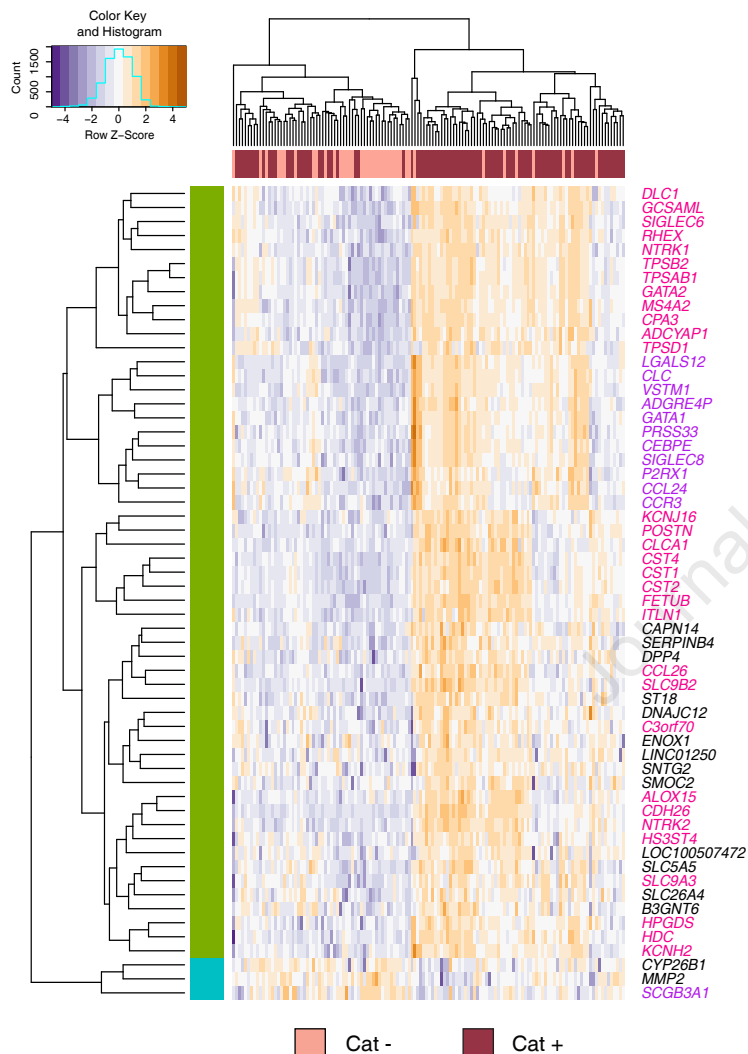
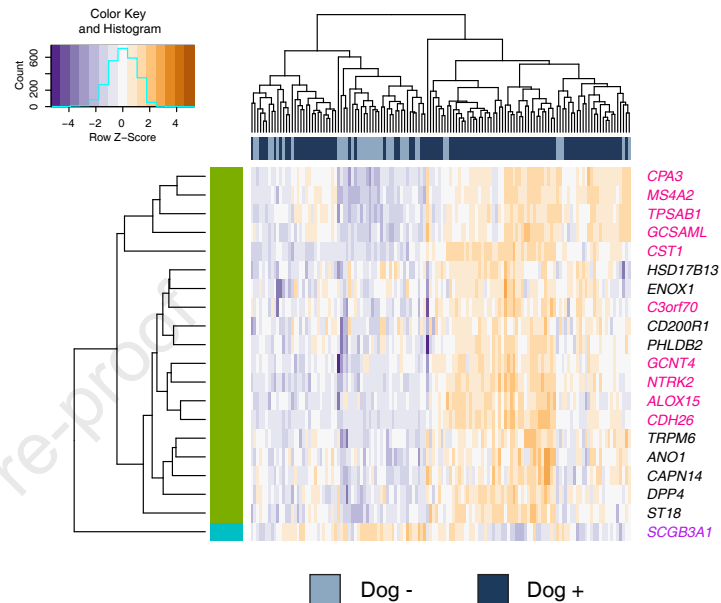
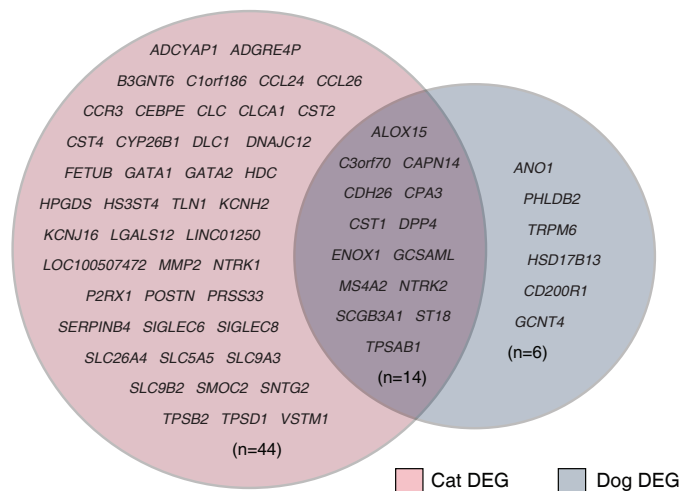
Figure 4: Correlation network between nasal *Corynebacterium* sp., nasal *Staphylococcus epidermidis*, and nasal transcripts associated with cat and/or dog sensitization. Circular and rectangular nodes represent differentially expressed genes (DEGs) associated with cat and/or dog sensitization at $FDR \leq 0.05$ and $|\log_2 \text{fold change}| > 1$. Edges show significant Spearman correlations ($FDR \leq 0.05$) between the relative abundance of nasal species and gene expression level. Circled genes are DEGs identified by causal mediation analysis to mediate associations between *Corynebacterium* sp., *Staphylococcus epidermidis*, and pet sensitization.

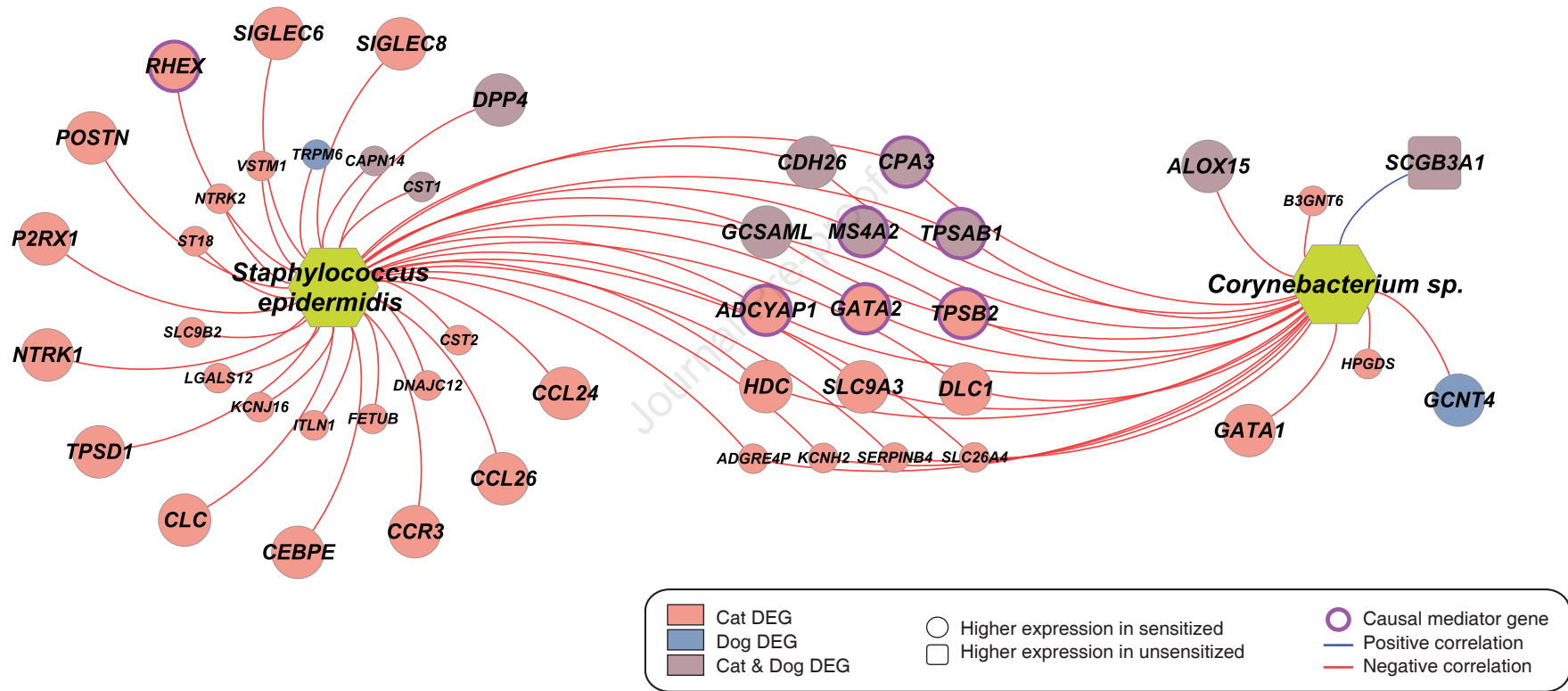
Figure 5: Causal mediation analysis for *Corynebacterium* sp., *Staphylococcus epidermidis*, DEGs, and pet sensitization. Associations between: (1) nasal species and pet sensitization are shown in blue; (2) nasal species and DEGs are shown in purple; and (3) DEGs and pet sensitization are shown in pink. Positive (+) and negative (-) associations are as indicated. Dashed arrows indicate mediation of the association between nasal species and pet sensitization by gene expression. FDR values from the causal mediation analysis are shown. Results are shown for *Corynebacterium* sp. (panels A,B) and *Staphylococcus epidermidis* (C).

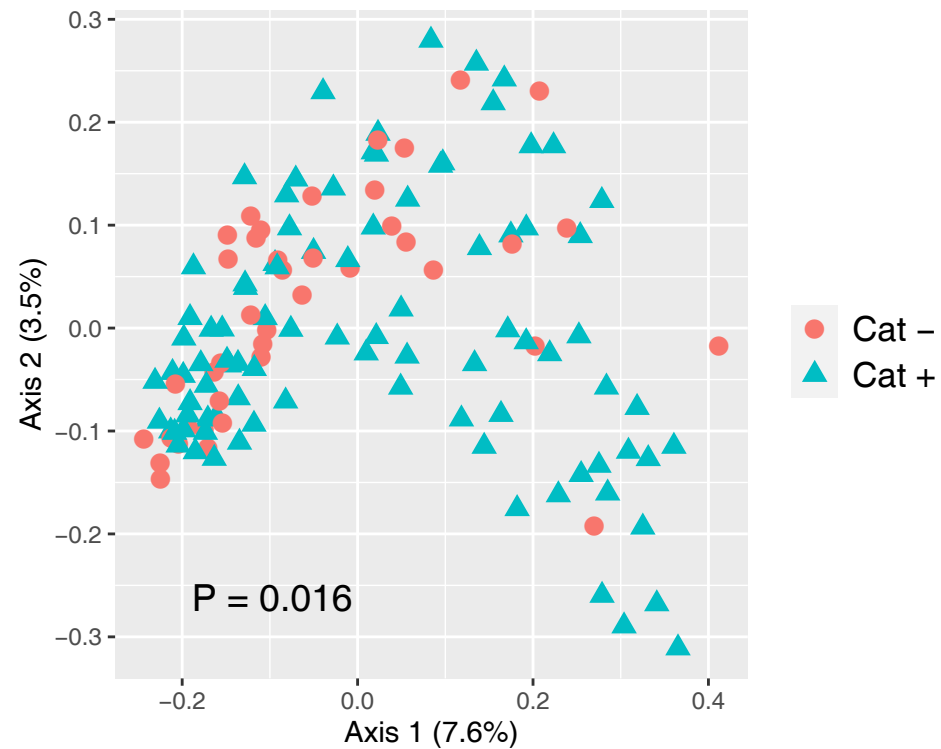
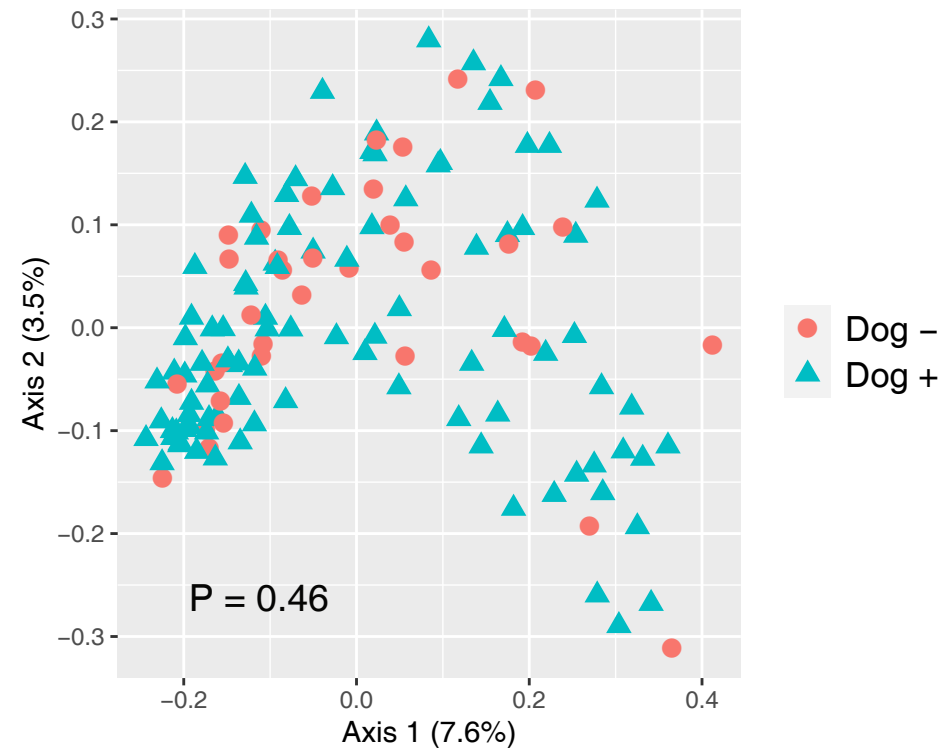


A**C****D**

A**B**

A**B****C**



A**B**

Online Supplement

The nasal microbiome, nasal transcriptome, and pet sensitization

Yoojin Chun, MS¹, Anh Do, PhD¹, Galina Grishina, MS², Zoe Arditi, BS¹, Victoria Ribeiro, BS¹,
Alexander Grishin, PhD², Alfin Vicencio, MD³, Supinda Bunyavanich, MD, MPH^{1,2*}

Methods

Study population and design

Children with physician-diagnosed persistent asthma based on National Asthma Education and Prevention Program (NAEPP): Third Expert Panel on the Diagnosis and Management of Asthma (EPR3) guidelines¹ were recruited from general pediatric, pulmonology, and allergy/immunology clinics within the Mount Sinai Health System, New York, NY, USA throughout all seasons from 2015 to 2018. Inclusion criteria included physician-diagnosed persistent asthma based on NAEPP EPR3 guidelines¹, demonstration of airway hyperresponsiveness by pre/post bronchodilator spirometry or methacholine challenge, and age 4-21 years. Exclusion criteria included history of non-asthma pulmonary disease, sinus disease, bleeding disorder, or anticoagulant or aspirin use. Recruitment occurred at clinical encounters. Questionnaires addressing demographics, pet ownership, asthma medication use, last upper respiratory infection, and last oral antibiotic use were administered to each subject, along with the Asthma Control Test (ACT).² Allergen sensitization to cat dander, dog dander, tree pollen mix, grass pollen mix, and weed pollen mix was evaluated using serum specific IgE measurements by ImmunoCAP (ThermoFisher-PHADIA, Uppsala, Sweden). Allergen sensitization was defined as serum specific IgE (sIgE) level ≥ 0.10 kUA/L.

26 *Microbiome profiling and analysis*

27 Nasal samples were obtained from children with asthma using a sterile cotton swab. DNA
28 extraction was performed using Qiagen DNeasy Mini Kit (Valencia, CA). Sample DNA yield and
29 quality was accessed using Nanodrop (Thermo Fisher Scientific). 16S rRNA library preparation
30 was done accordingly to the protocol described in (16S Metagenomic Sequencing Library
31 Preparation by Illumina Inc.). We also used both regular agarose DNA gel electrophoresis and
32 Bioanalyzer for analysis and quantification of the final PCR product. Dual indices and Illumina
33 sequencing adapters were attached to the V3V4 amplicon using the Nextera XT Index Kit v2
34 (set A and B, Illumina Inc.). DNA isolated from blank swabs and PCR blanks were included with
35 the samples for sequencing as negative controls. DNA sequencing was performed on amplified
36 16S rRNA using the Illumina MiSeq platform with 2x250bp paired-end reads. Reads were
37 quality controlled and denoised using the DADA2³ plugin in Qiime2⁴. The sequences were
38 summarized to the amplicon sequence variant (ASV) table via DADA2³, and the non-zero
39 counts detected from the blank swabs and PCR blanks were subtracted from all other samples.
40 The Silva 138 database⁵ was used to generate a phylogenetic tree and classify taxonomy of the
41 reads. All samples met the minimum sequences per sample threshold of 2,000.

42 Alpha and beta diversities and compositional analyses were performed using Qiime2⁴.
43 Alpha diversity was estimated by Shannon Index at sampling depth of 4,980, which was the
44 lowest sequencing depth. Alpha diversity was measured using ASV counts, and differences
45 between groups were compared with the Kruskal-Wallis test. To calculate significance of
46 differences in beta diversity between groups, weighted UniFrac and Jaccard distances were
47 measured using ASV counts, and PERMANOVA tests with 1,000 permutations were performed.
48 PICRUST⁶ and STAMP⁷ were used to analyze predicted metagenomic profiles between
49 sensitized and unsensitized children. The input OTU table for the PICRUST analysis was
50 generated via Qiime 1.9.1⁸. P-value was calculated using STAMP¹⁶ with White's non-parametric
51 t-test followed by FDR correction. Linear discriminant effect size (LEfSe)⁹ was used to calculate

differentially abundant species between groups with $P < 0.05$ and $LDA \geq 2.0$. The species analyzed in the analysis were pre-filtered to only include species with relative abundance $\geq 1\%$.

Transcriptome profiling and analysis

Nasal brushing was performed with a standard cytology brush for all subjects. Brushes were immediately placed in RNALater (Thermo Fisher Scientific, Fair Lawn, NJ). RNA extraction was performed with the Qiagen RNeasy Mini Kit (Valencia, CA). Samples were assessed for yield and quality using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and Qubit fluorometry (Thermo Fisher Scientific, Grand Island, NY). The RNA sequencing library was prepared with the standard TruSeq RNA Sample Prep Kit v2 protocol (Illumina). The mRNA libraries were sequenced on the Illumina HiSeq 2500 platform with a per-sample target of 40-50 million 100 bp paired-end reads. The data were put through Mount Sinai's standard mapping pipeline¹⁰ (using STAR¹¹ Picard¹² and Subread¹³). Mapped data were subjected to quality control with RNA-SeQC¹⁴. R package edgeR¹⁵ was used to normalize the dataset after filtering out lowly expressed genes by keeping genes where counts per million (CPM) is greater than 1 in at least one quarter of the total number of samples.

Differential gene expression analysis was performed using R packages voom¹⁶ and limma¹⁷, and p-values were adjusted for inflation/deflation via a Bayesian method based on estimation of the empirical null distribution (R package Bacon¹⁸). Differentially expressed genes (DEGs) were identified with $FDR \leq 0.05$ and $|\log_2 \text{fold change}| > 1$. Heatmaps for the DEGs were plotted using R. The gene expression values of the DEGs were centered and scaled by gene for heatmap visualization, and hierarchical clustering dendrograms for the DEGs were generated using the complete linkage method. A weighted gene coexpression network was built using all genes and WGCNA¹⁹.

Correlation network between microbiome and host transcriptome

Spearman correlations were calculated between the relative abundances of *Corynebacterium* sp. and *Staphylococcus epidermidis*, and gene expression levels for each of the DEGs for cat and/or dog sensitization. $FDR \leq 0.05$ was used as the threshold for inclusion in the network.

Cytoscape²⁰ was used for the network visualization.

Causal mediation analysis

Causal mediation analysis was performed with nasal species associated with cat and/or sensitization by the LEfSe analyses (*Corynebacterium* sp. and *S. epidermidis*) and cat and/or dog sensitization-associated DEGs correlated with these two nasal species. The robust structural modeling equation (SEM), implemented by the sem function in the lavaan R package²¹, was used to test two different mediation effects (1) whether the effects of nasal species on pet sensitization is mediated by difference in gene expression levels of a gene (1st mediation analysis); (2) whether the effects of gene expression on pet sensitization is mediated by difference in abundance of a nasal species (2nd mediation analysis). Mediation analyses were performed in two steps. First, three different regression models were generated. The first logistic regression model was regressed on cat and/or dog sensitization with either each nasal species (for the 1st mediation analysis) or each gene (for the 2nd mediation analysis) as an independent variable to estimate the direct effect of nasal species or gene on pet sensitization. The second model was regressed on gene expression as an independent variable to estimate effects of nasal species on gene expression (for the 1st mediation analysis) or reverse (regressing nasal species as an independent variable on gene expression for the 2nd mediation analysis). The third model estimates the effect of mediator (either gene expression in the 1st mediation analysis or nasal species in the 2nd mediation analysis) on pet sensitization. The second step used outputs from previous regression models to estimate mediation effects by

computing a robust standard error by sandwich-type covariance matrices and the Santorra-Bentler statistic²¹. FDR $\leq$ 0.05 was used as our threshold for significance.

Supplementary results

Consideration of potential confounders

We performed LEfSe analyses adjusted for potential confounders (age, sex, ICS/LABA use, cat or dog ownership). We used MaAsLin2²² to adjust the ASV data for age, sex, ICS/LABA, and cat or dog ownership. We then converted the residual from this model into relative abundances for input into LEfSe analysis. As was the case for the unadjusted ASV-based analyses, adjusted LEfSe still showed *Corynebacterium* sp. (LDA = 3.99) and *S. epidermis* (LDA = 3.95) as the top two species associated with lack of cat sensitization, and *Corynebacterium* sp. (LDA = 4.20) as the top species associated with lack of dog sensitization.

We also performed differential gene expression analysis adjusted for age, sex, ICS/LABA, cat or dog ownership. Compared to the DEGs identified from unadjusted analyses, the results were highly and significantly overlapping: for cat sensitization DEGs, fold enrichment = 254.2, hypergeometric test FDR = 4.26E-137; for dog sensitization DEGs, fold enrichment = 808.1, FDR = 1.52E-47. These results support minimal effect from these potential confounders.

WGCNA

Twenty four modules were identified by WGCNA¹⁹. Enrichment of each module for DEGs correlated with *Corynebacterium* sp. and/or *Staphylococcus epidermidis* was calculated and tested with the hypergeometric test. Two modules (magenta and purple) were significantly enriched for such DEGs (magenta: fold enrichment = 244.6, Hypergeometric test FDR = 3.39E-66; purple: fold enrichment = 2.7, FDR = 4.35E-03). Gene ontology (GO) analysis^{23, 24} revealed

that the purple module was enriched for the biological process “inflammatory response” (fold-enrichment=5.3, FDR = 4.5E-56). GO analysis for the magenta module was precluded by its small sample size (n=39); however, it contained many genes previously associated with allergy outcomes as discussed in the main text.

Supplementary references

1. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. Washington DC: National Heart Lung and Blood Institute and National Asthma Education and Prevention Program; 2007.
2. Schatz M, Kosinski M, Yarlas AS, Hanlon J, Watson ME, Jhingran P. The minimally important difference of the Asthma Control Test. *J Allergy Clin Immunol* 2009; 124:719-23 e1.
3. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; 13:581-3.
4. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; 37:852-7.
5. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013; 41:D590-6.
6. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* 2013; 31:814-21.
7. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 2014; 30:3123-4.

- 153 8. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al.
154 QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*
155 2010; 7:335-6.
- 156 9. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic
157 biomarker discovery and explanation. *Genome Biol* 2011; 12:R60.
- 158 10. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al.
159 Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat*
160 *Neurosci* 2016; 19:1442-53.
- 161 11. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
162 universal RNA-seq aligner. *Bioinformatics* 2013; 29:15-21.
- 163 12. Broad Institute. Picard Toolkit. Broad Institute, GitHub repository 2019. Available at
164 <https://broadinstitute.github.io/picard/>.
- 165 13. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
166 assigning sequence reads to genomic features. *Bioinformatics* 2014; 30:923-30.
- 167 14. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-
168 SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*
169 2012; 28:1530-2.
- 170 15. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
171 expression analysis of digital gene expression data. *Bioinformatics* 2010; 26:139-40.
- 172 16. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model
173 analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15:R29.
- 174 17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
175 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*
176 2015; 43:e47.

- 177 18. van Iterson M, van Zwet EW, Consortium B, Heijmans BT. Controlling bias and inflation
178 in epigenome- and transcriptome-wide association studies using the empirical null
179 distribution. *Genome Biol* 2017; 18:19.
- 180 19. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network
181 analysis. *BMC Bioinformatics* 2008; 9:559.
- 182 20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a
183 software environment for integrated models of biomolecular interaction networks.
184 *Genome Res* 2003; 13:2498-504.
- 185 21. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical*
186 *Software* 2012; 48:36.
- 187 22. Multivariable Association in Population-scale Meta-omics Studies. Available from
188 <https://huttenhower.sph.harvard.edu/maaslin/>.
- 189 23. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large
190 gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4:44-57.
- 191 24. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward
192 the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37:1-
193 13.

194

195 **Figure S1: Principal Coordinate Analysis (PCoA) plots show separation of nasal bacterial**
196 **composition based on cat and dog sensitization.** (A) cat sensitized subjects (n=91) and cat
197 unsensitized subjects (n=41) were compared by PERMANOVA with 1,000 permutations
198 (P=0.016, pseudo-F=1.3). (B) Dog sensitized subjects (n=96) and dog unsensitized subjects
199 (n=36) were compared by PERMANOVA with 1,000 permutations (P=0.46, pseudo-F=0.98).
200 Plots based on Jaccard distances, with each point corresponding to a subject and colored by
201 sensitization status.