

## Supplementary Materials for:

### **Asthma severity and corticosteroid response depends on variable type 1 and type 2 inflammation in the airway**

John V. Fahy<sup>1</sup>, Nathan D. Jackson<sup>2</sup>, Satria P. Sajuthi<sup>2</sup>, Elmar Pruesse<sup>2</sup>, Camille M. Moore<sup>2</sup>, Jamie L. Everman<sup>2</sup>, Cydney Rios<sup>2</sup>, Monica Tang<sup>1</sup>, Marc Gauthier<sup>3</sup>, Sally E. Wenzel<sup>3,4</sup>, Eugene R. Bleecker<sup>5</sup>, Mario Castro<sup>6</sup>, Suzy A. Comhair<sup>7</sup>, Serpil C. Erzurum<sup>7</sup>, Annette T. Hastie<sup>8</sup>, Wendy Moore<sup>8</sup>, Elliot Israel<sup>9</sup>, Bruce D. Levy<sup>9</sup>, Loren Denlinger<sup>10</sup>, Nizar N. Jarjour<sup>10</sup>, Mats W. Johansson<sup>10</sup>, David T. Mauger<sup>11</sup>, Brenda R. Phillips<sup>11</sup>, Kaharu Sumino<sup>12</sup>, Prescott G. Woodruff<sup>1</sup>, Michael C. Peters<sup>1</sup>, Max A. Seibold<sup>2, 13, 14</sup>, for the National Heart, Lung, and Blood Institute Severe Asthma Research Program-3

<sup>1</sup>Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA.

<sup>2</sup>Center for Genes, Environment, and Health, National Jewish Health, Denver, CO.

<sup>3</sup>Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA.

<sup>4</sup>Department of Environmental and Occupational Health, Graduate School of Public Health University of Pittsburgh, Pittsburgh, PA.

<sup>5</sup>Division of Genetics, Genomics and Precision Medicine, Department of Medicine, University of Arizona, Tucson, AZ.

<sup>6</sup>Division of Pulmonary, Critical Care and Sleep Medicine, University of Kansas School of Medicine, The University of Kansas Medical Center, Kansas City, KS.

<sup>7</sup>Lerner Research Institute, Cleveland Clinic, Cleveland, OH.

<sup>8</sup>Department of Internal Medicine, School of Medicine, Wake Forest University, Medical Center Blvd, Winston-Salem, NC.

<sup>9</sup>Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

<sup>10</sup>Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI.

<sup>11</sup>Division of Statistics and Bioinformatics, Department of Public Health Sciences, Pennsylvania State University, Hershey, PA.

<sup>12</sup>Division of Pulmonary Critical Care Medicine, Washington University, St. Louis MO.

<sup>13</sup>Department of Pediatrics, National Jewish Health, Denver, CO.

<sup>14</sup>Division of Pulmonary Sciences and Critical Care Medicine, University of Colorado School of Medicine, Aurora, CO.

## Supplementary Materials and Methods

### *Processing of RNA-seq data*

Raw sequence reads were trimmed using skewer v0.2.2 (end-quality 15, mean-quality 25, min 30) (1) and then aligned to the human reference genome (GRCh38) using HISAT2 v2.1.0 (2) with default parameter settings. Gene quantification was performed using htseq-count (mode=intersection-nonempty, stranded=reverse, a=20) (3) with GRCh38 Ensembl v84. Samples were filtered out based on standard QC parameters (number of mapped reads > 15 million, % reads to mitochondrial genes < 40), resulting in RNA-seq data for 782 samples collected from 384 participants (347 with asthma) across the different time points.

### *Defining T1 and T2 endotypes based on sputum co-expression networks*

To cluster genes into co-expressing groups that may correspond to discrete biological function, we applied weighted gene co-expression network analysis (WGCNA) to VST-normalized counts (DESeq2; 4) for the top 10,000 most variant genes (selected after filtering out ribosomal and mitochondrial genes, unannotated ENSG genes, and genes without 10 or more counts observed in at least 15% of samples). Note that in addition to all 782 samples in the current study, this analysis also included 48 non-SARP asthma samples that are not part of the current study. The analysis was run using soft threshold=16, type=signed, minClusterSize=10, deep split=3, pamStage=FALSE, and pamRespectsDendro=TRUE. This analysis identified 42 networks, which were assigned biological functions based on pathway enrichment analysis using Enrichr (5). Among these networks was one strongly enriched in markers of T1 inflammation (e.g., *STAT1*, *IFNG*, *CD8A*, *CD8B*; 65 genes total) and another strongly enriched in markers of T2 inflammation (e.g., *IL4*, *IL13*, *IL1RL1*, *CCL26*; 48 genes total). We categorized the 782 samples into T1- and T2-“high” and “low” groups by hierarchically clustering samples (ward.D2 clustering of a Euclidean distance matrix using *hclust* in R) based on all network genes, for T1 and T2 gene sets in turn, whose expression exhibited a Pearson correlation  $\geq 0.7$  with the network overall (i.e., network eigengenes) (36 T1 genes and 35 T2 genes). The first split in the dendrograms were used to distinguish high and low-expressers of each endotype.

## References for Supplementary Methods

1. H. S. Jiang, R. Lei, S. W. Ding, S. F. Zhu, Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, (2014).
2. D. Kim, B. Landmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357-U121 (2015).
3. S. Anders, P. T. Pyl, W. Huber, HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
4. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).
5. E. Y. Chen, C. M. Tan, Y. Lou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, M. a. A., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **128**, (2013).

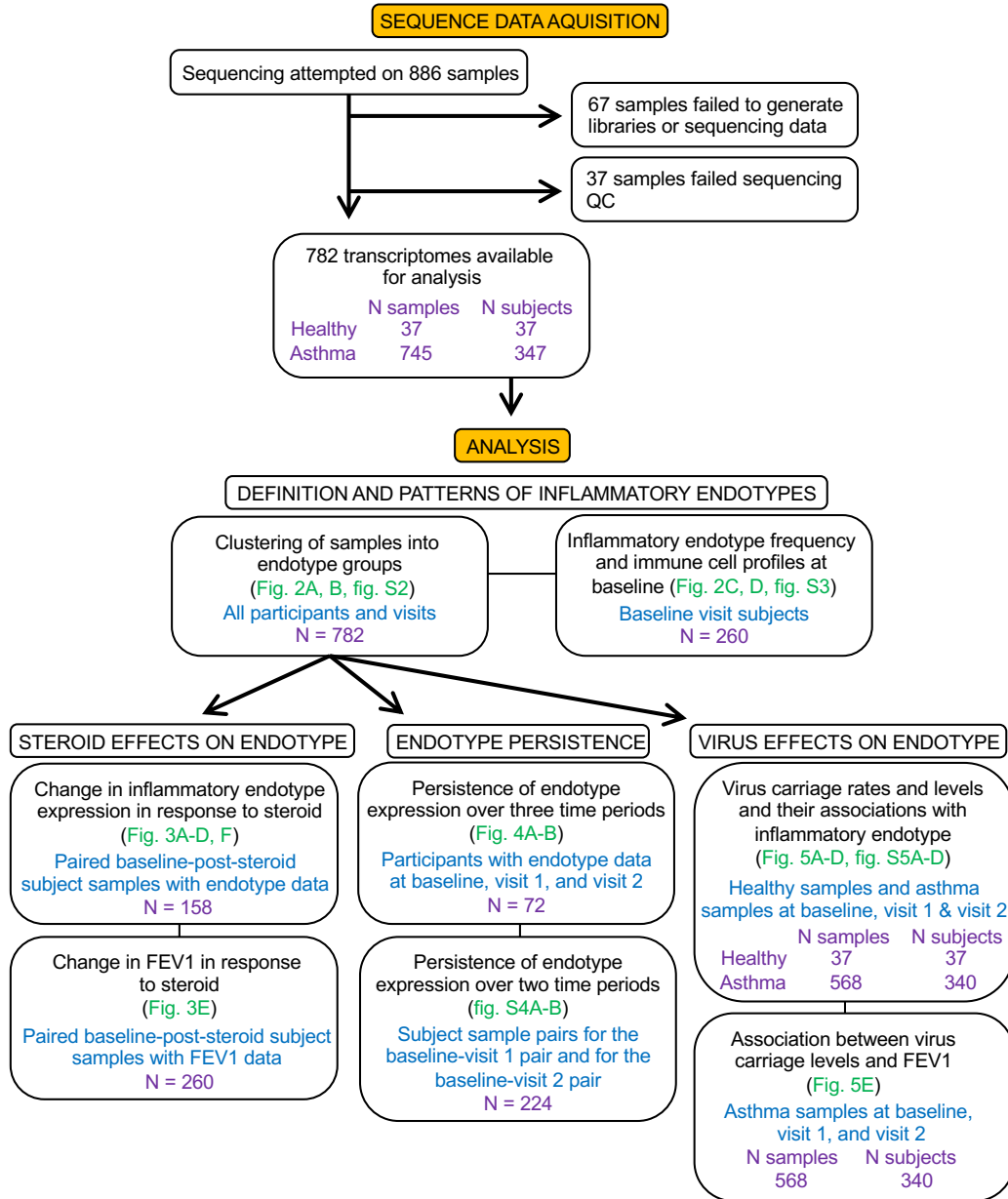
## Supplementary Tables

**Supplementary Table 1.** Description of SARP participants in this study.

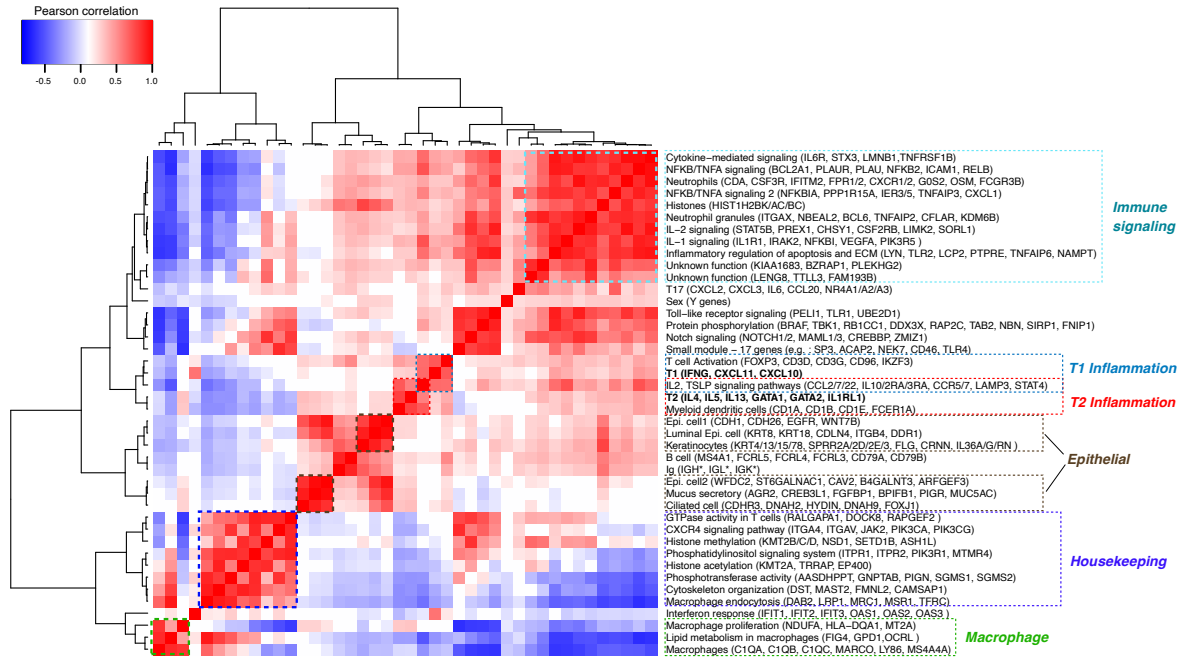
	Healthy (n=37)	Asthma SARP (n=347)
Age (years), mean $\pm$ sd	38.9 $\pm$ 13.3	47.7 $\pm$ 13.8
Male, n (%)	8 (25.0)	109 (31.4)
BMI, mean $\pm$ sd	25.9 $\pm$ 5.6	32.4 $\pm$ 8.6
Race, n (%)		
White	17 (53.1)	230 (66.3)
Black	4 (12.5)	79 (22.8)
Asian	6 (18.8)	13 (3.7)
Other/Multiple	4 (12.5)	17 (6.6)
Hispanic, n (%)	1 (3.1)	12 (3.5)
Pre-BD FEV1 (% pred), mean $\pm$ sd	96.9 $\pm$ 9.7	73.2 $\pm$ 18.2
Pre-BD FVC (% pred), mean $\pm$ sd	98.0 $\pm$ 12.2	85.3 $\pm$ 16.4

## Supplementary Figures

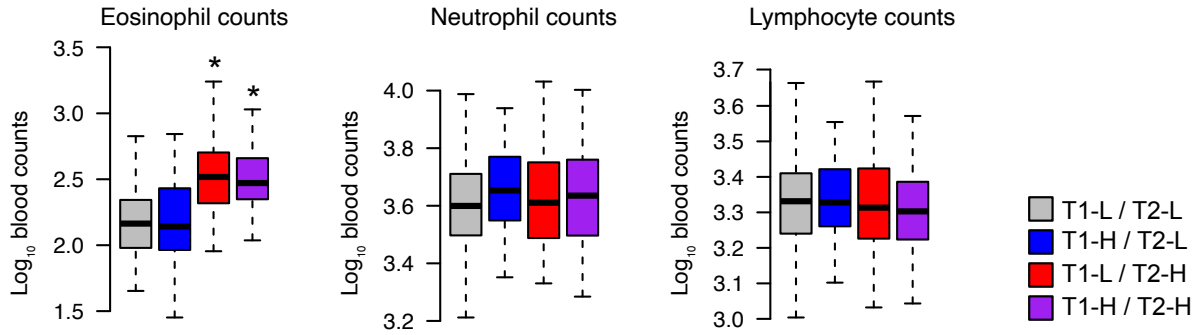
**Fig. S1.** Consort-style diagram summarizing samples and subjects sequenced and analyzed in the study.



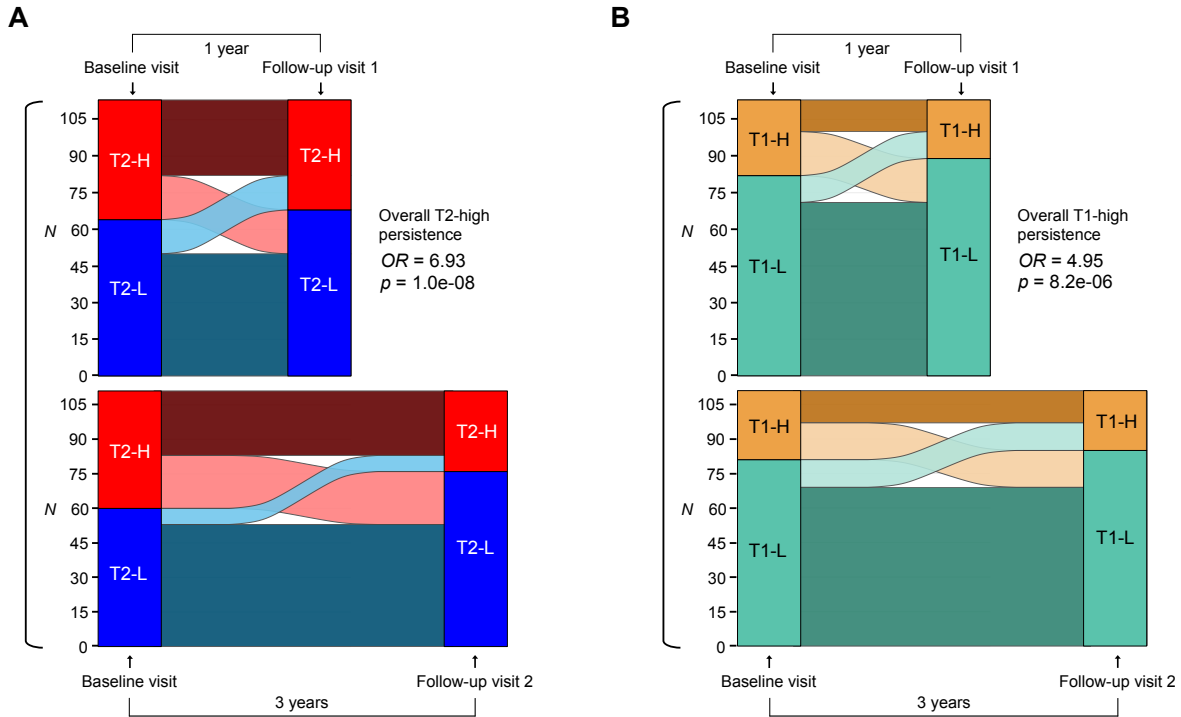
**Fig. S2.** Heat map of pairwise Pearson correlation coefficients among eigenenes for 42 WGCNA gene networks identified from RNA-seq data on 782 SARP samples, enabling visualization of groups of highly correlated networks in the dataset. Identifiable groups, including T1 and T2 inflammation, immune signaling, and epithelial expression, are highlighted. Functional descriptions and key genes for each network are given.



**Fig. S3.** Box plots depicting immune cell counts ( $\log_{10}$ ) in the blood of individuals sampled at baseline ( $N = 260$ ) who also exhibit T1 and/or T2 inflammatory endotypes (from left to right within a plot,  $N=116, 30, 77,$  and  $37$ ). Asterisks indicate when  $\log_{10}$  immune cell counts significantly differ from the dual T1 and T2-low group ( $p < 0.05$ ) based on an ANOVA that accounts for gender, age, and ethnicity ( $p$ -values, from left to right:  $4.19 \times 10^{-08}, 1.73 \times 10^{-04}$ ). Data beyond the end of whiskers are not shown.



**Fig. S4. (A)** An alluvial diagram depicting persistence of T2 status between baseline and 1-year follow-up visits (top; N=113) and between baseline and 3-year follow-up visits (bottom; N=111). The odds ratio (*OR*) and p-value, describing the excess chance that a T2-high individual at follow-up was also T2-high at baseline compared to having switched from T2-low, were estimated using GEE logistic regression, where the two sets of visit pairs were analyzed together. **(B)** The same as panel A, except related to persistence of T1 status.



**Fig. S5.** (A) A pie chart describing the proportion of each human rhinovirus species identified using our metagenomics pipeline (N=68). (B) similar to panel A but for different coronavirus species (N=58). (C) Box plots visualizing no difference in T2 network expression among high virus (N=39), low virus (N=107), and no virus (N=459) samples. Effects of virus group were carried out using a mixed model predicting T2 network expression as a function of virus status, with participant as a random effect. Overlaid points are colored by number of  $\log_{10}$  virus reads in a sample. (D) A bar plot depicting the proportion of samples that carry no, low, or high virus within T2-low (N=359) and T2-high (N=246) samples (proportions are listed on each bar). Odds ratios (*OR*) and p-values, describing the excess chance of carrying low virus (compared to no virus; top) or of carrying high virus (compared to no/low virus; bottom) when T2-high compared to T2-low, were calculated using GEE logistic regression.

