ORIGINAL ARTICLE

# Gene Expression in Relation to Exhaled Nitric Oxide Identifies Novel Asthma Phenotypes with Unique Biomolecular Pathways

Brian D. Modena[1], John R. Tedrow[1], Jadranka Milosevic[1], Eugene R. Bleecker[2], Deborah A. Meyers[2], Wei Wu[3], Ziv Bar-Joseph[3], Serpil C. Erzurum[4], Benjamin M. Gaston[5], William W. Busse[6], Nizar N. Jarjour[6], Naftali Kaminski[7], and Sally E. Wenzel[1]

[1]Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh School of Medicine, University of Pittsburgh Asthma Institute at UPMC, Pittsburgh, Pennsylvania; [2]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston Salem, North Carolina; [3]Lane Center for Computational Biology School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania; [4]Department of Pathobiology Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio; [5]Division of Pediatric Pulmonary, Allergy, and Immunology, Case Western Reserve University and Rainbow Babies and Children's Hospital, Cleveland, Ohio; [6]Division of Allergy, Pulmonary, and Critical Care Medicine, University of Wisconsin, Madison, Wisconsin; and [7]Pulmonary, Critical Care, and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut

## Abstract

**Rationale:** Although asthma is recognized as a heterogeneous disease associated with clinical phenotypes, the molecular basis of these phenotypes remains poorly understood. Although genomic studies have successfully broadened our understanding in diseases such as cancer, they have not been widely used in asthma studies.

**Objectives:** To link gene expression patterns to clinical asthma phenotypes.

**Methods:** We used a microarray platform to analyze bronchial airway epithelial cell gene expression in relation to the asthma biomarker fractional exhaled nitric oxide ($Fe_{NO}$) in 155 subjects with asthma and healthy control subjects from the Severe Asthma Research Program (SARP).

**Measurements and Main Results:** We first identified a diverse set of 549 genes whose expression correlated with $Fe_{NO}$. We used *k*-means to cluster the patient samples according to the expression of these genes, identifying five asthma clusters/phenotypes with distinct clinical, physiological, cellular, and gene transcription characteristics—termed "subject clusters" (SCs). To then investigate differences in gene expression between SCs, a total of 1,384 genes were identified that highly differentiated the SCs at an unadjusted *P* value $< 10^{-6}$. Hierarchical clustering of these 1,384 genes identified nine gene clusters or "biclusters," whose coexpression suggested biological characteristics unique to each SC. Although genes related to type 2 inflammation were present, novel pathways, including those related to neuronal function, WNT pathways, and actin cytoskeleton, were noted.

**Conclusions:** These findings show that bronchial epithelial cell gene expression, as related to the asthma biomarker $Fe_{NO}$, can identify distinct asthma phenotypes, while also suggesting the presence of underlying novel gene pathways relevant to these phenotypes.

**Keywords:** exhaled nitric oxide; clustering; severe asthma

Severe asthma affects approximately 10% of the total asthma population but accounts for 50% of the disease-associated costs (1). The diagnosis of severe asthma, like asthma itself, is based on a broad definition that sheds little light on the underlying molecular pathobiology or phenotypic differences (2). Prior efforts to classify and understand asthma phenotypes have used

(Received in original form June 16, 2014; accepted in final form October 21, 2014)

Supported by the National Institutes of Health, National Heart, Lung, and Blood Institute grants HL109250 (S.C.E. and B.M.G.), HL103453 (S.C.E. and B.M.G.), HL69174 (S.E.W.), HL069116 (W.W.B. and N.N.J.), HL69167 (E.R.B., D.A.M.), RC2 HL101487 (S.E.W., D.A.M., E.R.B., and N.K.), CTSI UL1 RR024153 (S.E.W.), and UL1 RR025011 (W.W.B. and N.N.J.)

Author Contributions: Acquisition of data: J.M., E.R.B., D.A.M., S.C.E., B.M.G., W.W.B., N.N.J., and S.E.W.; conception and design: S.E.W., N.K., B.D.M., J.R.T., W.W., and Z.B.-J.; analysis and interpretation: B.D.M., J.R.T., J.M., S.E.W., W.W., and N.K.; wrote manuscript: B.D.M. and S.E.W.; approved and edited manuscript: J.R.T., J.M., E.R.B., D.A.M., W.W., Z.B.-J., S.C.E., B.M.G., W.W.B., N.N.J., and N.K.

Correspondence and requests for reprints should be addressed to Sally Wenzel, M.D., Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh School of Medicine, University of Pittsburgh Asthma Institute at UPMC, NW 931 Montefiore, 3459 Fifth Avenue, Pittsburgh, PA 15213. E-mail: wenzelse@upmc.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 190, Iss 12, pp 1363–1372, Dec 15, 2014
Copyright © 2014 by the American Thoracic Society
Originally Published in Press as DOI: 10.1164/rccm.201406-1099OC on October 22, 2014
Internet address: www.atsjournals.org

## At a Glance Commentary

**Scientific Knowledge on the Subject:** Global gene expression in asthma, especially in relation to clinical phenotypes, has not been fully explored.

**What This Study Adds to the Field:** In this study, we analyzed gene expression profiles of bronchial airway cells in relation to exhaled nitric oxide from patients with asthma and healthy control participants. This approach characterized clinically meaningful asthma phenotypes and identified novel molecular pathways for future investigation.

mostly clinical phenotypic data (3–8), with little integration of more complex biologic information.

Gene expression microarrays, a snapshot of genome-scale steady-state messenger RNA (mRNA) expression, have been used to identify novel disease subtypes, predict drug response, and suggest pathological mechanisms in diseases such as cancer and inflammatory disorders (9, 10). In asthma, expression of three bronchial airway epithelial cell (BAEC) genes was used to identify an asthma phenotype (termed "Th2 high") among a group of corticosteroid (CS)-naive patients with mild asthma associated with elevated serum IgE, CS responsiveness, and increased blood and airway eosinophils (11, 12). Molecular phenotyping beyond these three epithelial genes in a broader range of patients with asthma has not yet been performed.

Fractional exhaled nitric oxide ($F_{ENO}$), an asthma biomarker that associates with atopy and eosinophilia (6, 13), is strongly induced by type 2 cytokines *in vitro*. Its inhibition *in vivo* by biologic therapies targeted toward IL-4 and -13 confirms its association with this immune pathway (14–17). However, $F_{ENO}$ elevation is observed across a range of disease states, including a high-risk asthma phenotype, CS refractory asthma, and even atopy without asthma (18, 19). These findings limit its clinical usefulness (20–22) and suggest that multiple biological factors beyond type 2 inflammation may regulate its production and/or downstream consequences.

We hypothesized that clinically distinct asthmatic phenotypes could be identified within a diverse population of subjects by clustering BAEC samples according to the expression of a subset of genes correlated with $F_{ENO}$. We further hypothesized that clustering of genes that are most differentially expressed between these phenotypes would provide biologic information on disease mechanisms. To address this hypothesis, bronchoscopically obtained epithelial brushings from extensively characterized subjects with asthma and healthy subjects in the National Heart Lung and Blood Institute–sponsored Severe Asthma Research Program (SARP) were analyzed by gene expression microarray. Genes strongly correlating with $F_{ENO}$ levels were used to cluster epithelial samples and identify potentially meaningful clinical subject clusters (SCs). Additional genes not correlating with $F_{ENO}$, but still differentiating the SCs, were then clustered to explore molecular and biologic pathways specific to each SC. Some of the results have been previously reported as an abstract (23).

## Methods

*See* online supplement for complete details.

### Study Population

Samples and data were obtained and characterized from subjects in SARP from 2009 to 2011. In this cross-sectional observational study, adherence to prescribed medications was not formally addressed in the period before sampling. However, coordinators and investigators informally addressed compliance to inhaled and systemic CSs over the course of three to four study visits. Subjects believed to have poor adherence to medications were generally excluded.

### Oligonucleotide Microarray Experiments

The complete data set is available online (National Center for Biotechnology Information's Gene Expression Omnibus database; http://www.ncbi.nlm.nih.gov/geo/; accession numbers GSE63142 and GSE43696). Low-variance probes were removed and median values used for probes with identical RefSeq numbers, resulting in a filtered list of 19,567 genes.

### Real-Time Quantitative Reverse Transcription–Polymerase Chain Reaction

Real-time quantitative reverse transcription–polymerase chain reaction (qRT-PCR) confirmed microarray gene expression with RNA extracted from the same bronchial brushings. Genes validated included those identified in the clustering analysis (n = 19) as well as additional genes of potential relevance to asthma (n = 5).

### Subject Sample Clustering by $F_{ENO}$-correlated Genes

$F_{ENO}$ level was not normally distributed and therefore correlated with $F_{ENO}$ using Spearman correlations at a threshold false discovery rate of 5%. $k$ was chosen by a previously described method that minimizes the difference between error sum of squares for the array data and a randomized set of data (24). As shown in Figure E1, the rate of improvement (slope of the curve) tapers as $k$ reaches the range of 4 to 7. The variable $k$ chosen (n = 5) was the largest value that resulted in at least 10 subjects per cluster, a minimum size needed to statistically detect phenotypic differences.

### Comparison of SC Phenotype Data after *k*-Means Clustering

Phenotypic differences between SCs were detected by statistical methods in the online data supplement. All significant intergroup comparisons listed in the tables achieve the appropriate Bonferroni corrected $P$ value of 0.005 or less (for comparisons among five groups).

### Identification of Genes Most Differentially Expressed between SCs

After establishment of SCs, all genes expressed in the epithelial brushings (total = 19,957) were compared across SCs for differential expression using intergroup Student $t$ tests. Genes differentially expressed to a $P$ value $< 10^{-6}$ between at least two of the SCs were used to compose a set of "highly differentiating" genes. Of note, 149 of the original 549 $F_{ENO}$-correlated genes were not included, as they did not meet this criterion. To ensure that the genes in the highly differentiating gene set were not identified by "chance," epithelial samples were randomly reassigned to same-sized SCs with gene expression comparisons repeated for five

iterations. In these randomly reassigned SCs, there were no genes differentially expressed at a $P$ value $< 10^{-6}$ between any two of the five SCs.

### Hierarchical Clustering of Genes to Explore Gene Clusters

As mentioned, a highly differentiating gene list was generated by finding the most differentially expressed genes across the five SCs. Hierarchical clustering using only these highly differentiating genes was then performed to determine genes with similar expression patterns (i.e., "coexpressed" genes). Based on the hierarchical clustering results, these genes were placed into gene clusters (GCs), and analyzed using Ingenuity and PubMed gene annotation.

## Results

### Demographics by Traditional Asthma Severity Criteria

Samples and data from 155 participants were obtained from 2009 to 2011 (Table 1). There were 26 healthy control subjects, and 51 participants met the American Thoracic Society 2000 definition of severe asthma. There were no differences in race and sex across the groups. Subjects with severe asthma (SA) were older than subjects with mild to moderate asthma/no inhaled CSs (ICS) and healthy control subjects (HCs), and had a higher body mass index (BMI) than the HCs. $FEV_1$ decreased with severity. Although blood IgE and $Fe_{NO}$ differed among the groups, there were no intergroup differences. Of the subjects with SA, 63% used oral CSs (OCS) on a regular basis, supporting the severity of disease in these patients.

### Identification of Genes Correlating with $Fe_{NO}$

One hundred twenty-seven of 155 subjects had $Fe_{NO}$ data available. Correlations of $Fe_{NO}$ with all differentially expressed genes (n = 19,567) identified 549 genes (*see* Table E2 in the online supplement) that correlated with $Fe_{NO}$ (false discovery rate $< 0.05$), ranging from 0.3 to 0.72 for positive correlations and $-0.54$ to $-0.3$ for negative correlations. The enzyme most likely responsible for NO production, inducible nitric oxide synthase (NOS2), had the strongest correlation (rho = 0.72, $P$ value $< 10^{-7}$).

### Identification of SCs

Using the subset of genes that correlated with $Fe_{NO}$, samples were clustered by $k$-means with $k = 5$ (Figure 1). This choice was supported by the markedly different characteristics of the resulting five SCs shown in Table 2 and Figure 2. Three SCs had abnormally high $Fe_{NO}$ levels (SC2, 3, and 5), and two had normal $Fe_{NO}$ levels (SC1 and 4). Nearly 90% of all HCs are in SC1 and 4, and 65% of traditionally defined subjects with severe asthma are in SC2 and 3 (Figure 2).

### Description of SCs

*See* Tables E3–E5 for differences in specific characteristics.

***SC1.*** SC1, the largest cluster, contains 73% of all HCs in the study, and more than two-thirds are either HCs or subjects with mild asthma on no or low-dose ICS. Fifty percent of all subjects with mild asthma treated with ICS are in this cluster, suggesting that effective ICS therapy may nearly normalize epithelial gene expression. Despite this, 16% of subjects in this cluster are traditionally classified as having SA. Consistent with the large percentage of HCs in this cluster, lung function is near normal ($FEV_1 = 85 \pm 16\%$), and asthma symptoms are infrequent. There is little evidence for inflammation according to $Fe_{NO}$ (median = 18 ppb), blood, or bronchoalveolar lavage (BAL) cells. Atopy and IgE levels (median = 49 IU/ml) are also relatively low (Table 2).

***SC2.*** SC2 is the second-largest SC, with the highest $Fe_{NO}$ overall (median = 55 ppb).

**Table 1.** Demographics by Traditional Clinical Characteristics

| | HCs | Mild-Mod No ICS | Mild + ICS | Mod + ICS | Severe | *P* Value | Significant Intergroup Comparison* |
|---|---|---|---|---|---|---|---|
| Age, mean ± SD | 32 ± 12 | 29 ± 10 | 37 ± 12 | 36 ± 12 | 45 ± 11 | <0.0001 | SA > HC<br>SA > Mild no ICS |
| Sex, M/F | 12/14 | 13/24 | 6/16 | 8/11 | 16/35 | 0.61 | No intergroup difference |
| Race, W/AA/O | 18/5/3 | 21/10/5 | 14/5/2 | 8/10/1 | 31/14/3 | 0.43 | No intergroup difference |
| BMI, median (25–75th quartile) | 24 (22–30) | 28 (23–32) | 28 (25–34) | 29 (25–32) | 32 (26–37) | 0.006 | SA > HC |
| $FEV_1$% predicted, mean ± SEM | 94 ± 9 | 84 ± 16 | 91 ± 16 | 68 ± 8 | 56 ± 20 | <0.0001 | HC > SA<br>Mild no ICS > SA<br>Mild + ICS > SA<br>HC > Mod + ICS<br>Mild no ICS > Mod + ICS<br>Mild + ICS > Mod + ICS |
| Atopy, yes/no | 16/10 | 31/4 | 15/4 | 16/2 | 39/8 | 0.08 | No intergroup difference |
| IgE, median (25–75th quartile), IU/ml | 59 (11–132) | 113 (54–432) | 139 (56–468) | 142 (55–314) | 104 (26–434) | 0.042 | No intergroup difference |
| Blood eos, median (25–75th quartile), % | 2 (2–3) | 3 (2–5) | 2 (2–5) | 4 (2–7) | 3 (1–5) | 0.14 | No intergroup difference |
| $Fe_{NO}$, median (25–75th quartile), ppb | 20 (13–37) | 37 (19–60) | 21 (13–39) | 39 (22–59) | 37 (21–63) | 0.043 | No intergroup difference |

*Definition of abbreviations*: AA = African American; BMI = body mass index; eos = eosinophils; F = female; $Fe_{NO}$ = fractional exhaled nitric oxide; HC = healthy control subject; ICS = inhaled corticosteroids; M = male; Mod = moderate; O = other; SA = patient with severe asthma; W = white.

*All significant intergroup differences listed achieved a Bonferroni corrected $P$ value of $P \leq 0.005$.

**Table 2.** Demographics by Subject Clusters

| | SC1 | SC2 | SC3 | SC4 | SC5 | *P* Value | Significant Intergroup Comparisons* |
|---|---|---|---|---|---|---|---|
| $F_{E_{NO}}$, median (25–75th quartile), ppb | 18 (12–25) | 55 (38–103) | 39 (18–50) | 19 (14–27) | 38 (29–75) | <0.0001 | SC2 > SC1<br>SC3 > SC1<br>SC5 > SC1<br>SC2 > SC3<br>SC2 > SC4 |
| Age, mean ± SD, yr | 36 ± 12 | 38 ± 13 | 42 ± 14 | 42 ± 12 | 27 ± 9 | 0.0013 | SC1 > SC5<br>SC2 > SC5<br>SC3 > SC5<br>SC4 > SC5 |
| Sex, M/F | 15/41 | 18/29 | 11/12 | 5/7 | 6/11 | 0.44 | No intergroup difference |
| Race, W/AA/O | 36/15/3 | 25/15/6 | 14/4/3 | 9/1/2 | 8/9/0 | 0.15 | No intergroup difference |
| BMI, median (25–75th quartile) | 28 (22–32) | 31 (27–37) | 30 (24–37) | 31 (25–34) | 25 (21–29) | 0.0024 | SC2 > SC1<br>SC2 > SC5 |
| $FEV_1$% predicted, mean ± SEM | 85 ± 19 | 66 ± 21 | 66 ± 22 | 76 ± 26 | 83 ± 16 | <0.0001 | SC1 > SC2<br>SC1 > SC3<br>SC5 > SC2† |
| Change after BD, median (25–75th quartile), % | 7 (4–14) | 19 (10–41) | 15 (11–21) | 10 (5–14) | 11(6–16) | 0.0001 | SC1 > SC2<br>SC1 > SC3 |
| Atopic sensitization, yes/no | 34/18 | 37/7 | 18/2 | 12/0 | 16/1 | 0.0070 | No intergroup difference |
| Number of positive skin tests, mean ± SD | 3 ± 3 | 4 ± 3 | 4 ± 3 | 5 ± 3 | 5 ± 1 | 0.016 | No intergroup difference |
| IgE, median (25–75th quartile), IU/ml | 49 (15–106) | 142 (51–433) | 232 (102–342) | 66 (40–266) | 294 (169–783) | <0.0001 | SC2 > SC1<br>SC3 > SC1<br>SC5 > SC1 |
| Blood eos, median (25–75th quartile), % | 2 (1–3) | 4 (3–6) | 2 (1–5) | 3 (2–3) | 4 (3–8) | <0.0001 | SC2 > SC1<br>SC5 > SC1<br>SC5 > SC4‡ |
| Uses oral CS, % | 4 | 27 | 47 | 20 | 14 | 0.0019 | No intergroup difference |

*Definition of abbreviations*: AA = African American; BD = bronchodilator; BMI = body mass index; CS = corticosteroids; eos = eosinophils; F = female; $F_{E_{NO}}$ = fractional exhaled nitric oxide; M = male; O = other; W = white.

*k*-means clustering of subjects with genes (n = 549) that correlated strongly (positively and negatively) with exhaled nitric oxide ($F_{E_{NO}}$) identified five unique subject clusters (SCs). This table details the demographics of these newly identified SCs.

*All significant intergroup differences listed achieved a Bonferroni corrected *P* value of $P \leq 0.005$.

†$P = 0.0074$.

‡$P = 0.0059$.

Traditionally defined moderate to severe asthma makes up 73% of this SC, with low lung function ($FEV_1 = 66 \pm 21\%$), a high degree of symptoms, the greatest reversibility (19%; 25–75th percentile, 10–41%) and a high percentage on regular OCS (27%). Eosinophils in blood and BAL were higher than those in SC1 and 4. Despite the highest $F_{E_{NO}}$ and high blood eosinophil percentages, IgE level (median = 142 IU/ml) and number of positive skin tests are intermediate among the SCs. Subjects in this cluster generally developed disease in childhood (median, 8.5 yr; 25–75th percentile, 4–18 yr) and are obese.

***SC3.*** The third-largest SC is arguably the most severe. Seventy-four percent have moderate to severe disease by traditional measures, with 56% classified as severe. Similar to SC2 above, SC3 has low lung function ($FEV_1 = 66 \pm 22\%$) and high $F_{E_{NO}}$ (median = 39 ppb) but tends to have a later age at onset (13 yr; 25–75th percentile, 2–31 yr) and shorter duration of disease. There is little history of asthma in a parent (10% of subjects), although IgE levels are high (median = 232 IU/ml). Despite containing the highest percentage of patients receiving regular OCS (47%), BAL cell eosinophilia is similar to the other $F_{E_{NO}}$-high clusters (2%), consistent with a systemic CS effect. SC3 has the highest levels of peripheral blood and BAL neutrophils (62 and 4.9%, respectively), also consistent with high OCS use. It also has the second highest percent of BAL lymphocytes, which suggests a mixed inflammatory milieu in the lungs. SC3 has the highest percent with a history of nasal polyps (26%) and sinus surgery (30%) (Table E5) and is similar to SC2 in BMI.

***SC4.*** SC4 has normal $F_{E_{NO}}$ levels (median = 19 ppb) and is the smallest SC (n = 12). Although 33% are HCs (n = 4), in comparison to cluster SC1, which contains 73% of all HCs, this represents only 15% of all HCs. Unlike SC1, this cluster consists of 50% patients with moderate to severe asthma, with associated high health care use (27% ever intubated) and moderately high systemic CS use (20%). After removing the HCs, the lung function is low ($FEV_1 = 66 \pm 7\%$) and associated with the earliest age at onset (3 yr; 25–75th percentile, 0–6 yr) and the longest duration of disease (median = 34 yr; 25–75th percentile, 33–54 yr). Notwithstanding low $F_{E_{NO}}$, blood and BAL eosinophils, and total IgE (median = 66 IU/ml), 100% are atopic, and the number of positive skin tests does not differ from the high $F_{E_{NO}}$ clusters. Despite the lack of eosinophilic or neutrophilic inflammation, this cluster has the highest BAL lymphocytes (median = 13.2%). The median BMI is the highest of any SC but does not differ from other SCs, likely due to its small sample size.

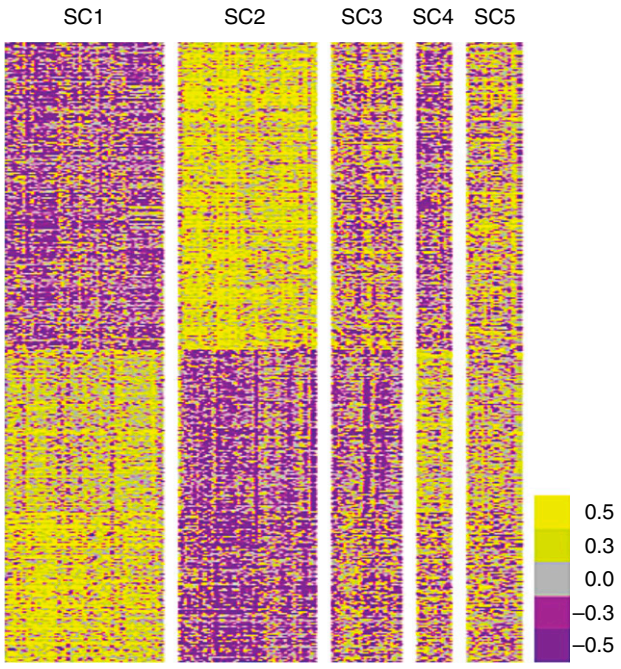



**Figure 1.** Heat map derived by *k*-means clustering of subject samples with genes that correlated with exhaled nitric oxide ($F_{ENO}$). Gene expression was correlated with $F_{ENO}$ using a Spearman correlation method, identifying 549 genes that correlated strongly (positively and negatively) with $F_{ENO}$. *k*-means clustering of subject bronchial epithelial brushing samples with these genes identified five unique SCs shown here (*x axis*, *top*).

***SC5.*** SC5 is the youngest group (mean age = 27 yr), with an elevated $F_{ENO}$ (38 ppb), having an early age at onset (6 yr; 25–75th percentile, 2–11 yr) and shorter duration of disease than in SC4. More than 50% identify as African Americans (highest in any cluster), with the highest percentage of any cluster reporting asthma in a parent (59%). Only 24% have severe disease by traditional measures, and they have normal lung function ($FEV_1 = 83 \pm 16\%$), similar to SC1. Despite this, 35% report an ICU stay or intubation. As would be consistent with their younger age and African racial background, IgE levels are highest in this group (median = 294 IU/ml), and eosinophilic inflammation in both blood and BAL is relatively high and similar to SC2. There is very little OCS use (14%) or history of nasal polyps/sinus disease. Perhaps reflective of their younger age, this SC is the leanest of all SCs (median BMI = 27).

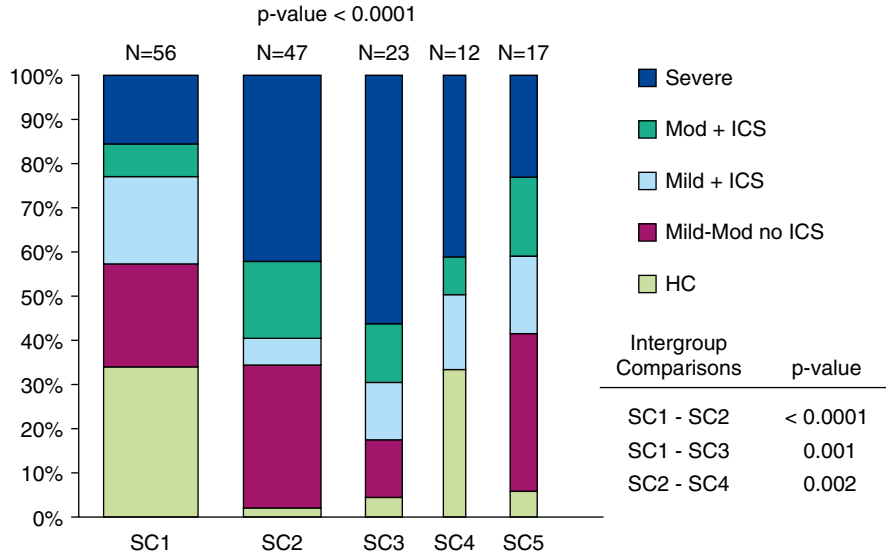



**Figure 2.** Composition of subject clusters (SC1–5) according to traditional asthma severity criteria. Gene expression was correlated with exhaled nitric oxide ($F_{ENO}$) using a Spearman correlation method, identifying 549 genes that correlated strongly (positively and negatively) with $F_{ENO}$. *k*-means clustering of subject bronchial epithelial brushing samples with these genes identified five unique SCs. As shown in the figure, SCs greatly differed according to traditional asthma severity criteria. SC2 and 3 had the highest percentage of patients with severe asthma. HCs = healthy control subjects; ICS = inhaled corticosteroids; Mild-Mod no ICS = mild to moderate asthma not on ICS; Mild + ICS = mild asthma on ICS; Mod + ICS = moderate asthma on ICS; Severe = severe asthma.

## GC Results

Complete linkage hierarchical clustering of the highly differentiating set of genes (n = 1,384) along the *y* axis of the heat map identified nine clusters of coexpressed genes using a cutline at the fourth division of the hierarchical tree (Figure 3). The biologic characteristics of the clusters are summarized below and in Table 3. A complete listing by GC is found in Tables E7.1–7.9. Twenty-one genes were validated by mulitplex qRT-PCR, although not all of them were used in this study. The majority (n = 17) demonstrated a strong correlation (Spearman correlation, $r > 0.5$) (Table E6).

### GC1 (n = 131) (High = SC1; Intermediate = SC3, 4, and 5; Low = SC2)

GC1 includes genes of innate/antimicrobial defense and adaptive immunity (3% combined), including: BPI fold containing family A, member 1 (BPIFA1) (validated by qPCR) (25) and mucin 5B (MUC5B) (26–28). GC1 is also enriched for genes related to cell growth, proliferation, and survival (2%), including the epidermal growth factor receptor gene ERBB2 (validated by qPCR) and sphingosine-1-phosphate receptor (S1PR). The #1 and #2 ranked Ingenuity canonical pathway genes involve immunologic signaling and leukocyte mobility (Table E8).

### GC2 (n = 172) (High = SC4; Intermediate = SC1 and 5; Low = SC2 and 3)

GC2 is highly enriched for cilia-related genes (12%), including dynein motor proteins and subunits of the intraflagellar transport complex, a complex responsible

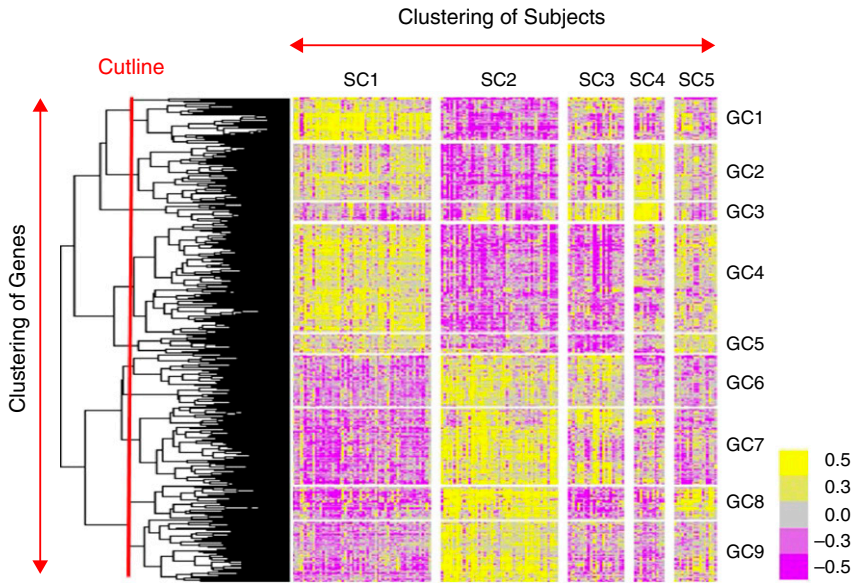



**Figure 3.** Complete linkage hierarchical clustering of genes that most strongly differentiate the subject clusters (SCs). After establishment of SCs, all genes expressed in the epithelial brushings (total = 19,957) were compared across SCs for differential expression using intergroup Student *t* tests. Genes differentially expressed to a *P* value $< 10^{-6}$ between at least two of the SCs were used to compose a set of "highly differentiating" genes (n = 1,384) shown in the *heat map* here. Complete linkage hierarchical clustering of this highly differentiating gene set along the *y axis* of the *heat map* identified nine clusters of coexpressed genes using a cutline at the fourth division of the hierarchical tree (*y axis*, *left*). SCs remain on the *x axis*. GC = gene cluster.

for the genesis, resorption, and signaling of primary cilia.

**GC3 (n = 52) (High = SC4 and 3; Intermediate = SC2; Low = SC1 and 5)**
This small GC is highly enriched for muscle-related genes (12%) and tumor necrosis factor (TNF)-related signaling (6%).

**GC4 (n = 325) (High = SC1; Intermediate = SC4 and 5; Low = SC2 and 3)**
This large GC4 is enriched with diverse functions and pathways. Surprisingly, 8% of genes associate with neuronal development and function. GC4 also includes NOTCH (4%) and WNT developmental pathway genes. The #5 ranked Ingenuity upstream analysis is WNT signaling protein (WNT3A, $P = 4.89 \times 10^{-4}$; Table E8). Finally, GC4 contains several reducing enzymes, and the #1 ranked Ingenuity pathway is glutathione-mediated detoxification ($P = 7.8 \times 10^{-4}$), including Glutathione S-transferase α 2 (GSTA2), validated by qRT-PCR.

**GC5 (n = 50) (High = SC1 and 5; Intermediate = SC2 and 4; Low = SC3)**
Although some well-described genes are in this subset, including Bruton's tyrosine kinase (BTK), there is not enrichment for a known molecular or cellular process. Likewise, there was no significant pathway identified in Ingenuity.

**GC6 (n = 150) (High = SC2 and 3; Intermediate = SC4 and 5; Low = 1)**
GC6 is enriched with mitochondria- (3%), microtubules- (2%), and actin- (2%) related genes. Affirmatively, the #3, #4, and #6 ranked Ingenuity pathways include oxidative phosphorylation ($4.2 \times 10^{-2}$), mitochondrial dysfunction ($2.8 \times 10^{-2}$), and an actin-based pathway ($4.5 \times 10^{-2}$), respectively. This cluster also contains neuronal genes (2%), including brain-derived neurotrophic factor (BDNF). The #1 ranked Ingenuity upstream analysis is neuronal development gene atrophin 1 (ATN1) ($2.9 \times 10^{-5}$).

**GC7 (n = 228) (High = SC2 and 3; Intermediate = SC4; Low = SC1 and 5)**
GC7 includes inducible NOS2, the gene correlating most strongly with $F_{ENO}$. Affirmatively, the #3 Ingenuity pathway is iNOS signaling ($P = 1.12 \times 10^{-3}$). GC7 is also enriched for type 2 and 3 interferon genes (2%), including IL-18RA and DUOX2 (dual oxidase-2) genes, validated by qRT-PCR. Ingenuity upstream analysis supports a prominent effect from TNF ($P = 9.62 \times 10^{-6}$), likely contributing to the enhancement of genes associated with apoptosis, including TRADD (TNFRSF1A-associated via death domain) and CARD14 (Caspase Recruitment Domain Family, Member 14). At least 3% of genes in GC7 are associated with cell–matrix interactions.

**GC8 (n = 96) (High = SC2 and 5; Low = SC1, 3, and 4)**
GC8 is enriched for mast cell–associated genes (11%), including TPSAB1 (tryptase) (tryptase genes validated by qRT-PCR), MS4A2 (β chain of high-affinity IgE receptor), and HPGDS (hematopoietic prostaglandin D2 [PGD2] synthase), with their tight clustering demonstrating strong evidence for coexpression. CRTH2, one of the receptors for PGD2, is also found here. Other genes include mucin-related (including MUC5AC, validated by qRT-PCR) (5%), cysteine metabolism (4%), vasoconstriction, and glycolipid antigen presentation (CD1 genes). The #1 Ingenuity upstream analysis is cytokine IL-13 ($P = 3.35 \times 10^{-7}$).

**GC9 (n = 180) (High = SC2; Intermediate = SC3, 4, and 5; Low = SC1)**
GC9 contains the "Th2-like" genes periostin (POSTN), chemokine ligand 26 (CCL26), and 15 lipoxygenase 1 (ALOX15A). Both POSTN and CCL26 were validated by qRT-PCR. GC9 is also enriched for genes related to intracellular trafficking (5%), mitochondrial function (3%) and O-linked glycosylation (2%). The #1 ranked Ingenuity upstream analysis protein is transcription factor XBP1 ($5.4 \times 10^{-11}$).

## Discussion

This report details the first large-scale study to combine exhaled nitric oxide and bronchial epithelial cell gene expression to objectively cluster a broad range of patients with asthma into distinct phenotypes. Detailed clinical, physiological, cellular, and gene transcription characteristics were combined to fully describe these newly defined clusters. This study is also unique in the use of gene expression patterns to uncover known and novel molecular pathways relevant to these asthma phenotypes.

**Table 3.** Gene Cluster Molecular and Biological Enrichment Characteristics

| GC (n) | High (SCs) | Int (SCs) | Low (SCs) | Functional Pathway/Cellular Association | Genes |
|---|---|---|---|---|---|
| 1 (131) | 1 | 3, 4, 5 | 2 | Innate immunity/antibacterial function | BPIFA1, MUC5B, PGLYPR4, NOX1 |
| | | | | Cell proliferation/apoptosis | ERBB2, S1PR, IL36RN |
| | | | | Lymphocyte activation/migration | LAG3, CCL15, SLAMF7, SLAMF9, MALL, TNFSF4 |
| | | | | Other | HCRT, FGF family |
| 2 (172) | 4 | 1, 5 | 2, 3 | Cilia structure/function | DNAH1, DNAH3, IFT122, DNAI1, FBF1, TMEM231, BBS1, ARL3, IFT140, HYDIN, KIF19, DNAH10 |
| | | | | Other | CRY2, INSR, SCGB1A1 and 3A1, SMAD3, DTX3, TMEM8B, ATXN2 |
| 3 (52) | 4, 3 | 2 | 1, 5 | TNF-α | SPPL2B, MAPK4K4, IKBKB |
| | | | | Muscle | DMPK. POPDC2, MUSTN1, TNNI3, FERL5, MYO15B |
| 4 (325) | 1 | 4, 5 | 2, 3 | Notch signaling/cell adhesion | WNK4, JAG2, EPDR1, CD302, CDON, GXYLT2, HEY2, THBS4, RBPJ, CLDN16, OSBPL3, GJB7, CDH1 |
| | | | | Neuronal function | SHANK2, LRP8, KCNN3, MAOB, SEMA3E, LMO3, EFHD1, NDN, CHN1, NPAS3, NDNF, NAV3, SLITRK5, SLC6A16, SNCAIP, TUB, INA, MAPT, ORS1E1, SCSK1, EFNB2, PMMA3, CLU, STXBP1, ENO2, OPRL1, DDC |
| | | | | Dystrophin family | SNTB1, SGCE, SSPN, CNTD1 |
| | | | | WNT family | WIF1, WNT5B, RHOU, ASCDD1, DKK3 |
| | | | | Ion channels | KCNA1, KCNC1, CACNB1, CACNA202, SCN3A, NEDD4L |
| | | | | Other | BPHLHE31, PER2, IL5RA, GSTA1, 2, 3, 5, HNMT, GAB2, PIAS1, FOXA2, LTF |
| 5 (50) | 1, 5 | 2, 4 | 3 | No obvious functional/cellular associations "Interesting genes" | PFDN5, KCNMB2, BTK, TNFSF12, GPAN |
| 6 (150) | 2, 3 | 4,5 | 1 | Microtubules (2%) | TUBA4A, ACOT13, CKAP2 |
| | | | | Mitochondrial (3%) | ATP5H, MTERF, ETFA, NNT, CPT1A. CYCS |
| | | | | Actin related (2%) | LIMA1, PDF1B, ACTR3, TMP4 |
| | | | | Neuronal (2%) | BDNF, GRIK2, NAV1, RIT1 |
| | | | | Other | IRAK4, GART, GMP5, SMAP1, ADAM9, IBTK, ITGB6, CXCL14 |
| 7 (228) | 2, 3 | 4 | 1, 5 | Interferons (type 1 and 2) (∼2%) | DUOX2, MX2, IL18R1, NOS2 |
| | | | | Apoptosis (∼3%) | CARD14, TRADD, CASP6, TXNDC5, AK2, TNFRFS18, MAGED1 |
| | | | | P38 related | MAPK13, MAP2K6 |
| | | | | Keratins | KRT23, KRT6A, KRT18 |
| | | | | Sialyl Lewis antigen | FUT3, 5, 6 |
| | | | | Cell matrix interactions (∼3.5%) | DIAPH3, LGALS7, LGALS7B, FERMT1, LOXL4, RHOC, LOX, TNC |
| | | | | Other | IRAK3, TEP and TIMELESS, PTK6, MUC4, DEFB1, PLA2G4A and ABCC1, LRRC8A, ICA1, POP1 |
| 8 (96) | 2, 5 | | 1, 3, 4 | Cysteine metabolism (4%) | CCBL1, CST1, CST2, FETUB |
| | | | | Mucins (5%) | MUC5AC, TFF3, TFF2, FOXA3, GANLT12 |
| | | | | Mast cells (11%) | HPGDS, PTGS1, P2RY14, CTSG, HDC, IL1RL1, KIT, CPA3, TPSAB1, MS4A2, TPSD1, CD22/SIGLEC2, SIGLEC17P, SLC18A2, DPP4 |
| | | | | Vasoconstrictors (possibly MC) (3%) | UTS2, P3RX1 |
| | | | | Glycolipid antigen presentation | CD1C, CD1E |
| | | | | Other | PTGDR2, MYO1C, CNRIP1, PRB1, 2 and 4 |
| 9 (180) | 2 | 3, 4, 5 | 1 | Mitochondria (3%) | MTHFD2, MRPL15, COX17, ATP5O, PYCR1, HK2 |
| | | | | Intracellular trafficking (5%) | STKBP5, RAB27B, TBC1D15, S100A6. ARL1, KIF16B, SYTL4, SEC24D, SEC22A |
| | | | | O-linked glycosylation (2%) | GALNT1, 4, 7, 10 |
| | | | | N-linked glycosylation (2%) | MAGT1, FUT8, ALG5, ERLEC1 |
| | | | | "Type 2 genes" (2%) | ALOX15A, CCL26, SERPINB2, POSTN |
| | | | | Other | BCL2L14, RIPK2, CORO1C, ARNTL2, MUC2, MGST2, NTRK2, CD44 |

*Definition of abbreviations*: GC = gene cluster; Int = intermediate; SCs = subject clusters; TNF = tumor necrosis factor.
Genes differentially expressed to a $P$ value of $P < 10^{-6}$ between at least two of the five SCs were added to the $F_{E_{NO}}$-correlated genes to form a highly differentiating set of genes. Complete linkage hierarchical clustering of this gene set formed nine GCs. This table details important genes with associated functional pathways and/or cellular associations in each of these GCs. The columns "High," "Int," and "Low" indicate comparative expression levels.

$Fe_{NO}$ has emerged as a viable asthma biomarker. NOS2 is the gene most strongly correlated with $Fe_{NO}$ (rho = 0.72, $P$ value $< 10^{-7}$), supporting its known link to BAEC expression and the biologic plausibility of its correlation with BAEC gene expression profiles (Table E2) (29–31). Similarly, the "type 2" immunity–associated genes POSTN, CLCA1, and SERPINB2 also strongly correlate with $Fe_{NO}$ (rho values from 0.49–0.60, $P$ values all $< 10^{-7}$), whereas the mucin gene MUC5B, which is strongly repressed by IL-13 (12), has the strongest negative correlation (rho = 0.54, $P$ value $< 10^{-7}$). This study confirms an association of $Fe_{NO}$ with type 2 immune processes. However, it adds many genes negatively and positively correlated with $Fe_{NO}$, providing a diverse set of genes to identify SCs in subsequent cluster analysis.

*k*-means clustering of the 549 $Fe_{NO}$-correlated genes identified five molecularly defined and clinically distinct SCs, three with higher $Fe_{NO}$ (SC2, 3, and 5) and the remaining two with low/normal $Fe_{NO}$ levels (SC1 and 4) (Table 2; Tables E2–E5). The majority of HCs, including all of those without atopy, clustered into SC1. It is therefore considered to be a "healthy" cluster. The patients with asthma who clustered into SC1 primarily had mild or well-controlled disease. Traditionally defined severe asthma was most prevalent in SC2 and 3, identifying these two clusters as the most "severe" SCs. Both of these clusters have high $Fe_{NO}$ levels, high inhaled and systemic CS use, BAL eosinophilia, and low lung function. However, when compared with SC2, SC3 has a later age at onset and more nasal polyposis and neutrophilic inflammation. SC4 is 100% atopic, has the earliest age of asthma onset, longest disease duration, and severe disease, but normal $Fe_{NO}$ and a high percentage of BAL lymphocytes. SC5, the youngest cluster, has high $Fe_{NO}$, a high percentage of African Americans (>50%), and, perhaps not surprisingly, the highest IgE levels (32–34). Although eosinophil levels are high, lung function does not differ from the "healthy" SC1.

Hierarchical clustering with the highly differentiating gene set of 1,384 genes (Figure 3) identified nine GCs that helped deepen our molecular understanding of each SC. Exploration of GC composition helped identify transcriptional patterns in severe asthma SCs that indicated both ongoing epithelial injury as well as inflammatory processes. GC1, 2, and 4 are all expressed at lower levels in "severe" SC2 and 3 compared with SC1. GC1 contains critical host defense genes, such as BPIFA1 (25) and the mucin gene MUC5B (26–28), which likely translates to an increased susceptibility to infection among patients with more severe asthma. MUC5B deficiency is associated with chronic infections in mice (28), with low levels previously described in asthmatic airways in relation to type 2 inflammation (12). Similarly, suppression of cilia-associated genes (i.e., dynein motor proteins of GC2), cell adhesion, and wound repair–related genes (GC4) suggest abnormal airway epithelial defense/structure among those with more severe disease. Fascinatingly, 8% of GC4 genes relate to neuronal function, suggesting neural features are an integral part of a healthy epithelial layer, whereas loss of expression could contribute to disease pathogenesis.

GC transcription patterns also revealed inflammatory conditions unique to each SC. Phenotypically, there are more type 2 inflammatory characteristics (i.e., elevated blood eosinophils and $Fe_{NO}$ [Table 2] and BAL eosinophils [Table E4]) in SC2 and 5 than in the "healthy" SC1. Supportively, they have increased expression of the mast cell–related GC8, and the "Th2-high" asthma phenotype genes found in GC9. Notwithstanding the fact that SC3 has an elevated $Fe_{NO}$ and SC4 has high atopy (100%), both have relatively lower expression of the mast cell–associated GC8. This suggests that a mast cell signature is quite specific for atopic/allergic and inflamed asthma.

Genes in GC7 are more highly expressed in "severe" SC2 and 3 compared with "healthy" SC1. GC7 contains iNOS and related pathway genes. These include IFN-γ signaling, supporting prior reports that indicate higher levels of IFN-γ are found in "severe" asthma and suggesting an involvement of type 1 pathways in iNOS expression (35–37). GC7 also associates with TNF signaling. The enhancement of TNF-α pathways and apoptosis-related genes in GC7 support overall derangements of cell death–related properties.

Finally, GC3 is higher in "severe" SC3 and "moderate-to-severe" SC4 compared with "healthy" SC1. GC3 contains muscle-related genes and those related to TNF signaling. SC4 also has intermediate expression of GC7, which, as mentioned above, also contains genes related to TNF signaling. This supports the role of TNF-induced inflammation in SC3 and suggests it may be a predominant or novel disease mechanism in SC4. Whether these patients might selectively benefit from anti-TNF therapeutic approaches remains to be determined.

Limitations to the study include the cross-sectional nature and the lack of studies that associate this mRNA expression to protein expression. In addition, $Fe_{NO}$ levels are influenced by post-transcriptional/translational modifications of iNOS and other relevant proteins, as well as by substrate levels. However, iNOS (NOS2) mRNA correlates to $Fe_{NO}$ with a rho = 0.72, suggesting ~50% of the $Fe_{NO}$ level is explained transcriptionally. Cells were not sorted before RNA extraction, making it unclear whether mRNA expression was derived from epithelial cells or other cell types. However, in many cases, mRNA specific for certain cell types, particularly mast cells, goblet cells, and ciliated epithelial cells, clustered together, supporting their expression from a common cell type. Finally, both use of and nonadherence to CS therapy could impact the relevance of the transcriptional data reported here. Notwithstanding, despite having the highest reported CS use and other indications of systemic CS adherence (e.g., blood and BAL neutrophilia in SC3), SC2 and 3 demonstrated strong inflammatory signals, including high gene expression of iNOS and associated genes in GC7. Also, OCS (or ICS) use was not specific to any SC. Although the study is likely biologically pertinent, its clinical importance must be tested by the ability of the gene and SCs to predict response to therapy.

In conclusion, this was the first large-scale study to combine exhaled nitric oxide and bronchial epithelial cell gene expression to objectively identify novel asthma phenotypes while also uncovering unique and complex relationships between gene expression and $Fe_{NO}$. This analysis may help to (*1*) improve current understanding of underlying pathobiology among asthma phenotypes, and (*2*) lead to the development of biomarkers that enhance the usefulness of $Fe_{NO}$ measurements. This hypothesis-generating study opens the door to future studies to confirm and/or refine these asthma phenotypes as well as identify novel molecular pathways in disease. ■

**Author disclosures** are available with the text of this article at www.atsjournals.org.

## References


1. Sullivan SD, Rasouliyan L, Russo PA, Kamath T, Chipps BE; TENOR Study Group. Extent, patterns, and burden of uncontrolled disease in severe or difficult-to-treat asthma. *Allergy* 2007;62:126–133.
2. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, Adcock IM, Bateman ED, Bel EH, Bleecker ER, *et al*. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* 2014;43:343–373.
3. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, Wenzel SE, Aujla S, Castro M, Bacharier LB, *et al*; National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. Heterogeneity of severe asthma in childhood: Confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *J Allergy Clin Immunol* 2011;127:382–389.e1–13.
4. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med* 2008;178:218–224.
5. Judie AH, Brooke S, Anne F, Scott TW, Benjamin R. Using spectral clustering of phenotypes to identify novel asthma subtypes [abstract]. *Am J Respir Crit Care Med* 2010;181:A1877.
6. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R Jr, Castro M, Curran-Everett D, Fitzpatrick AM, *et al*.; National Heart, Lung, and Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med* 2010; 181:315–323.
7. Siroux V, Basagaña X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, Slama R, Jarvis D, Anto JM, Kauffmann F, *et al*. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J* 2011; 38:310–317.
8. Wu T, Boezen HM, Postma DS, Los H, Postmus PE, Snieder H, Boomsma DI. Genetic and environmental influences on objective intermediate asthma phenotypes in Dutch twins. *Eur Respir J* 2010;36:261–268.
9. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–537.
10. Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002;32:502–508.
11. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C, *et al*. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci USA* 2007;104:15858–15863.
12. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180:388–395.
13. Yamamoto M, Tochino Y, Chibana K, Trudeau JB, Holguin F, Wenzel SE. Nitric oxide and related enzymes in asthma: relation to severity, enzyme function and inflammation. *Clin Exp Allergy* 2012;42: 760–768.
14. Chibana K, Trudeau JB, Mustovich AT, Hu H, Zhao J, Balzar S, Chu HW, Wenzel SE. IL-13 induced increases in nitrite levels are primarily driven by increases in inducible nitric oxide synthase as compared with effects on arginases in human primary bronchial epithelial cells. *Clin Exp Allergy* 2008;38:936–946.
15. Corren J, Lemanske RF, Hanania NA, Korenblat PE, Parsey MV, Arron JR, Harris JM, Scheerens H, Wu LC, Su Z, *et al*. Lebrikizumab treatment in adults with asthma. *N Engl J Med* 2011;365:1088–1098.
16. Wenzel S, Ford L, Pearlman D, Spector S, Sher L, Skobieranda F, Wang L, Kirkesseli S, Rocklin R, Bock B, *et al*. Dupilumab in persistent asthma with elevated eosinophil levels. *N Engl J Med* 2013;368: 2455–2466.
17. Wenzel S, Wilbraham D, Fuller R, Getz EB, Longphre M. Effect of an interleukin-4 variant on late phase asthmatic response to allergen challenge in asthmatic patients: results of two phase 2a studies. *Lancet* 2007;370:1422–1431.
18. Dweik RA, Sorkness RL, Wenzel S, Hammel J, Curran-Everett D, Comhair SA, Bleecker E, Busse W, Calhoun WJ, Castro M, *et al*.; National Heart, Lung, and Blood Institute Severe Asthma Research Program. Use of exhaled nitric oxide measurement to identify a reactive, at-risk phenotype among patients with asthma. *Am J Respir Crit Care Med* 2010;181:1033–1041.
19. Fitzpatrick AM, Gaston BM, Erzurum SC, Teague WG; National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. Features of severe asthma in school-age children: atopy and increased exhaled nitric oxide. *J Allergy Clin Immunol* 2006;118:1218–1225.
20. Dweik RA, Boggs PB, Erzurum SC, Irvin CG, Leigh MW, Lundberg JO, Olin AC, Plummer AL, Taylor DR; American Thoracic Society Committee on Interpretation of Exhaled Nitric Oxide Levels (FENO) for Clinical Applications. An official ATS clinical practice guideline: interpretation of exhaled nitric oxide levels (FENO) for clinical applications. *Am J Respir Crit Care Med* 2011;184: 602–615.
21. Shaw DE, Berry MA, Thomas M, Green RH, Brightling CE, Wardlaw AJ, Pavord ID. The use of exhaled nitric oxide to guide asthma management: a randomized controlled trial. *Am J Respir Crit Care Med* 2007;176:231–237.
22. Wysocki K, Park SY, Bleecker E, Busse W, Castro M, Chung KF, Gaston B, Erzurum S, Israel E, Teague WG, *et al*. Characterization of factors associated with systemic corticosteroid use in severe asthma: data from the Severe Asthma Research Program. *J Allergy Clin Immunol* 2014;133:915–918.
23. Modena BD, Tedrow J, Milosevic J, Kaminski N, Wenzel SE. Bronchial epithelial cell gene expression in relation to exhaled nitric oxide identifies new molecular asthma phenotypes [abstract]. *J Allergy Clin Immunol* 2014;133:AB176.
24. Ketchen DJ, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. *Strateg Manage J* 1996;17:441–458.
25. Schleimer RP, Kato A, Kern R, Kuperman D, Avila PC. Epithelium: at the interface of innate and adaptive immune responses. *J Allergy Clin Immunol* 2007;120:1279–1284.
26. Kirkham S, Sheehan JK, Knight D, Richardson PS, Thornton DJ. Heterogeneity of airways mucus: variations in the amounts and glycoforms of the major oligomeric mucins MUC5AC and MUC5B. *Biochem J* 2002;361:537–546.
27. Ordoñez CL, Khashayar R, Wong HH, Ferrando R, Wu R, Hyde DM, Hotchkiss JA, Zhang Y, Novikov A, Dolganov G, *et al*. Mild and moderate asthma is associated with airway goblet cell hyperplasia and abnormalities in mucin gene expression. *Am J Respir Crit Care Med* 2001;163:517–523.
28. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, *et al*. Muc5b is required for airway defence. *Nature* 2014;505: 412–416.
29. Dweik RA, Laskowski D, Abu-Soud HM, Kaneko F, Hutte R, Stuehr DJ, Erzurum SC. Nitric oxide synthesis in the lung. Regulation by oxygen through a kinetic mechanism. *J Clin Invest* 1998;101: 660–666.
30. Guo FH, Comhair SA, Zheng S, Dweik RA, Eissa NT, Thomassen MJ, Calhoun W, Erzurum SC. Molecular mechanisms of increased nitric oxide (NO) in asthma: evidence for transcriptional and post-translational regulation of NO synthesis. *J Immunol* 2000;164: 5970–5980.
31. Guo FH, De Raeve HR, Rice TW, Stuehr DJ, Thunnissen FB, Erzurum SC. Continuous nitric oxide synthesis by inducible nitric oxide synthase in normal human airway epithelium in vivo. *Proc Natl Acad Sci USA* 1995;92:7809–7813.
32. Burrows B, Martinez FD, Halonen M, Barbee RA, Cline MG. Association of asthma with serum IgE levels and skin-test reactivity to allergens. *N Engl J Med* 1989;320:271–277.
33. Naqvi M, Choudhry S, Tsai H-J, Thyne S, Navarro D, Nazario S, Rodriguez-Santana JR, Casal J, Torres A, Chapela R, *et al*. Association between IgE levels and asthma severity among African American, Mexican, and Puerto Rican patients with asthma. *J Allergy Clin Immunol* 2007;120: 137–143.

34. Vergara C, Murray T, Rafaels N, Lewis R, Campbell M, Foster C, Gao L, Faruque M, Oliveira RR, Carvalho E, *et al*. African ancestry is a risk factor for asthma and high total IgE levels in African admixed populations. *Genet Epidemiol* 2013;37:393–401.
35. Jeon SG, Oh SY, Park HK, Kim YS, Shim EJ, Lee HS, Oh MH, Bang B, Chun EY, Kim SH, *et al*. TH2 and TH1 lung inflammation induced by airway allergen sensitization with low and high doses of double-stranded RNA. *J Allergy Clin Immunol* 2007;120:803–812.
36. Kim HY, DeKruyff RH, Umetsu DT. The many paths to asthma: phenotype shaped by innate and adaptive immunity. *Nat Immunol* 2010;11:577–584.
37. Voraphani N, Gladwin MT, Contreras AU, Kaminski N, Tedrow JR, Milosevic J, Bleecker ER, Meyers DA, Ray A, Ray P, *et al*. An airway epithelial iNOS-DUOX2-thyroid peroxidase metabolome drives Th1/Th2 nitrative stress in human severe asthma. *Mucosal Immunol* 2014;7:1175–1185.

# Gene Expression in Relation to exhaled NO Identifies Novel Asthma Phenotypes with Unique Biomolecular Pathways

Brian D. Modena, John R. Tedrow, Jadranka Milosevic, Eugene R. Bleecker, Deborah A. Meyers, Wei Wu, Ziv Bar-Joseph, Serpil C. Erzurum, Benjamin M. Gaston, William W. Busse, Nizar N. Jarjour, Naftali Kaminski, Sally E. Wenzel

**ONLINE DATA SUPPLEMENT**

**Study population**

All samples and data were obtained from subjects in the Severe Asthma Research Program (SARP) from 2009-2011, at which time the network consisted of Cleveland Clinics/University of Virginia, the University of Pittsburgh, the University of Wisconsin and Wake Forest University. The enrollment and phenotypic characterization for this cohort, enriched for severe asthma, has been previously described [6]. Subjects signed informed consent and the individual Institutional Review Boards of the 4 participating SARP sites approved the study. All subjects were non-smokers (less than 5 pack-years) and bronchoscopies were performed 4 weeks after any exacerbations or respiratory tract infections. Severe asthmatic subjects (SA) met the American Thoracic Society (ATS) 2000 definition of severe asthma. Non-severe asthmatic subjects were classified initially as: Mild-moderate asthma/No inhaled corticosteroids (ICS), with an FEV1 >60% predicted on no inhaled corticosteroids (ICS), Mild asthma/+ICS (FEV1>80% predicted on low to moderate dose ICS) and Moderate asthma/+ICS (FEV1 60-80% predicted on low-moderate dose ICS). Healthy controls were recruited and tested in a similar manner. Demographics, asthma symptoms, medication use, medical/family history and health care utilization were derived from questionnaires. Pre- and post-bronchodilator spirometry with allergy skin prick testing (SPT), serum IgE and eosinophils were obtained within 1 month of bronchoscopy while FeNO was obtained the day of bronchoscopy. All subjects underwent bronchoscopy with bronchial brushings and lavage as previously described per SARP protocol [6, 37, 38]. Bronchoalveolar Lavage was performed using two 50 ml aliquots of saline. Cell counts and differentials (cytospins stained with Diff-Quik, City, State) were performed on cell pellets from the returned fluid.

**Oligonucleotide Microarray Experiments**

Total RNA was extracted from the fresh bronchial epithelial brushings suspended in Trizol solution using the QIACube system (Qiagen, Valencia, CA). RNA quality was determined using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA), and only samples with an RIN higher than 7 used for microarray experiments. Labeling reactions were performed by Quick Amp labeling kit (Agilent, Santa Clara, CA). RNA was hybridized to Agilent Human GE 4 × 44 Kv2 whole human genome microarrays, and scanned with an Agilent scanner (Agilent Technologies, Santa Clara, CA) according to manufacturer's protocol. The data was extracted using the Agilent Feature Extraction software 10.7.3.1 (Agilent Technologies). The data discussed in this manuscript have been deposited in NCBI's Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) and are accessible through Gene Expression Omnibus Series accession number GSE43696.

**Microarray data processing**

Microarray images were processed according to Agilent Feature Extraction protocol (http://www.genomics.agilent.com/article.jsp?pageId=2037&_requestid =767154). The arrays were performed in 3 batches and normalized by cyclic-LOESS with use of Bioconductor package [39], and floor value was set as the mean signal across all negative control probes across all samples. There were no samples excluded due to poor normalization. Redundant probe IDs were excluded after normalization, by choosing the probe with the highest mean signal across all samples, which reduced the probe set of 30,889 probes to a gene set of 19,567 genes.

**Real-time PCR analysis TaqMan Low Density Arrays**

To confirm expression of genes identified with microarrays, RNA was extracted from the same bronchial brushing cells for real-time RT-PCR (RT-PCR). The genes were selected by overall relevance to this and other analyses, as well as by known clinical interest (see Table E1). Specifically, 50 ng of total RNA from 155 subjects (26 healthy control subjects, 78 not severe and 51 severe asthmatics) were reverse transcribed to single-stranded cDNA using the High Capacity cDNA Reverse Transcription kit (Life Technology, Foster City, CA). Real-time PCR analysis was performed using TaqMan Low Density Arrays (Life Technology).

TaqMan Low-density array (TLDA) cards are 384-well microfluidic cards with eight ports, each containing 48 connected wells. The primers and probe for each assay are preloaded and dried onto the designated duplicate wells. All probes used on the card are conjugated to 6-carboxyfluorescein (FAM). The cards contained two control assays, the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and beta-D-glucuronidase (BGUS) genes and 22 assays of interest. Real-time RT-PCR was performed using an ABI Prism 7900HT according manufacturer protocol. The data was extracted and analyzed using ExpressionSuite Software v1.0.3

**Subject sample clustering by FeNO-correlated genes**

FeNO level was not normally distributed and therefore correlated with FeNO using Spearman correlations. Genes chosen for subsequent K-means clustering correlated with FeNO at a threshold false discovery rate (FDR) of 5%, determined by BRB ArrayTools statistical package. K was chosen by a previously described method that minimizes the difference between error sum of squares for the array data and a randomized set of data (24). As shown in Supplemental Figure E1, the rate of improvement (slope of the curve) tapers as k reaches the range of 4-7. The variable k

chosen (n=5) was the largest value that resulted in at least ten subjects per cluster, a minimum size needed to statistically detect phenotypic differences. Gene expression analysis was performed using BRB-Array Tools stable release 4.3.0, developed by Dr. Richard Simon and the BRB Array-Tools Development Team. Clustering was done using Cluster 3.0 and visualized with Java TreeView 1.1.6r3.

**Comparison of subject cluster (SC) phenotype data after K-means clustering**

Phenotypic information was compared across SCs using JMP software (SAS, Cary, NC). For continuous characteristics (e.g. BMI), Shapiro-Wilk test for normality was performed. As the majority of the data did not fit a normal distribution, nonparametric comparisons were performed by the Kruskal-Wallis variation of the Wilcoxon test to obtain an overall p-value, with Wilcoxon testing for intergroup comparisons, using a Bonferroni correction factor for multiple comparisons. For categorical data, a Pearson chi-squared calculation tested for associations in a multi-group contingency table. When a contingency was present in the multi-group analysis, intergroup comparisons were evaluated by a Pearson chi-squared test of the 2x2 contingency tables, correcting for multiple comparisons with a Bonferroni correction factor. All significant intergroup comparisons listed in the tables achieve the appropriate Bonferroni corrected p-value of 0.005 or less (for comparisons among 5 groups). Of the phenotypic features in this study, only 5 had greater than 5% missing data, including FeNO (18%), IgE (16%), Asthma Quality of Life Questionnaire (AQLQ) (10.9%), number of positive allergen skin prick tests (SPTs) (7%) and history of any positive allergen SPT (6%). No imputation for missing data was performed.

**Identification of genes most differentially expressed between subject clusters (SCs)**

Following establishment of SCs, all genes expressed in the epithelial brushings (total=19,957) were compared across SCs for differential expression using intergroup student t-tests. Genes differentially expressed to a p-value < $10^{-6}$ between at least 2 of the SCs were used to comprise a set of 'highly-differentiating' genes. Of note, 149 of the original 549 FeNO-correlated genes were not included as they did not meet this criteria. In order to ensure that the genes in the 'highly-differentiating' gene set were not identified 'by chance', epithelial samples were randomly reassigned to same-sized subject clusters with gene expression comparisons repeated for 5 iterations. In these randomly-reassigned subject clusters, there were no genes differentially expressed at a p-value < $10^{-6}$ between any 2 of the 5 SCs. At an p-value<10-3, there were only an average of 23 genes differentially expressed between these randomized clusters.

**Hierarchical clustering of genes to explore gene clusters (GCs)**

As mentioned, a 'highly-differentiating' gene list was generated by finding the most differentially expressed genes across the five SCs. Hierarchical clustering using only these 'highly-differentiating' genes was then performed in order to determine genes with similar expression patterns, i.e. 'co-expressed' genes. Based on the hierarchical clustering results, these genes were placed into Gene Clusters (GCs), and analyzed using Ingenuity and PubMed gene annotation. Specifically, co-expressed genes from each bicluster were analyzed by both Ingenuity pathway analysis (IPA) software application (http://www.ingenuity.com/products/ipa), and annotation information taken from PubMed online gene search (http://www.ncbi.nlm.nih.gov/pubmed). In IPA, global canonical pathway (GCP) analysis and upstream regulator analysis were considered using threshold cutoff of p-value < $10^{-2}$. GCP determines genes associate together by chance, and assigns a likelihood p-value. Upstream regulator analysis attempts to

discover upstream transcriptional regulators (e.g. cytokines, transcriptional factors, drugs, etc.) that could explain changes in gene expression.

**Table E1**. List of genes chosen for qSCR validation**.** To confirm expression of differentiating genes identified on the microarrays, RNA was extracted from the same bronchial brushing cells for real-time quantitative reverse transcription-polymerase chain reaction (qRT-SCR). The genes were selected by overall relevance to this and other analyses, as well as by known clinical interest. Table S6 presents the results of the correlation.

| | **Name (SYMBOL)** |
|---|---|
| 1 | chloride channel accessory 1 (CLCA1) |
| 2 | cystatin 1 (CST1) |
| 3 | chemokine (C-C motif) CCL26 |
| 4 | acetylcholinesterase (ACHE) |
| 5 | tryptase alpha-1 beta-1 (TPSAB1) |
| 6 | proline-rich protein subfamily 2 (PRB2) |
| 7 | periostin (POSTN) |
| 8 | inducible nitric oxide synthase (NOS2) |
| 9 | mucin 5AC (MUC5AC) |
| 10 | trefoil factor 3 (TFF3) |
| 11 | prostaglandin synthase 1 (PTGS1) |
| 12 | dual oxidase 2 (DUOX |
| 13 | receptor tyrosine-protein kinase erbB-2 (ERBB2/HER2) |
| 14 | lactotransferrin (LTF) |
| 15 | secretoglobin 3A1 (SCGB3A1) |
| 16 | jun B proto-oncogene (JUNB) |
| 17 | palate, lung, and nasal epithelium clone protein (PLUNC) |
| 18 | cluster of differentiation 86 (CD86) |
| 19 | interleukin-18 receptor 1 (IL18R1) |
| 20 | interleukin-1 receptor-like 1 (IL1RL1/ST2) |
| 21 | mucin 5B (MUC5B) |
| 22 | FBJ murine osteosarcoma viral oncogene homolog B (FOSB) |
| 23 | early growth response 2 (EGFR2) |
| 24 | glucuronidase, beta (GUSB) |

**Table E2**. Top 50 genes with positive or negative correlation with FeNO. Gene correlation performed with (1) all participants and (2) excluding those with moderate asthma on inhaled corticosteroid treatment and severe disease (total 86 participants). Genes highlighted in green signify genes no longer found to be correlated with FeNO once 69 patients with moderate or severe asthma are removed from the data set.

Gene expression was correlated with FeNO using a Spearman correlation method, chosen because of the non-linear distribution of FeNO values. Genes listed correlated with FeNO with a 90% probability of containing no more than 5% of false discoveries (FDR). Genes are ranked according to absolute value of r-value.

| **Top 50 Genes that correlate with FeNO using all participants (n=155)** | | | **Top 50 Genes that correlate with FeNO excluding Mod-ICS and Severe phenotypes (n=86)** | | |
|---|---|---|---|---|---|
| **Gene Symbol** | **rho value** | **p-value** | **Symbol** | **rho value** | **p-value** |
| NOS2 | 0.719 | < 1e-07 | NOS2 | 0.721 | < 1e-07 |
| CDH26 | 0.642 | < 1e-07 | CDH26 | 0.714 | < 1e-07 |
| CST1 | 0.612 | < 1e-07 | CST1 | 0.65 | < 1e-07 |
| CCL26 | 0.601 | < 1e-07 | CST2 | 0.631 | 1.00E-07 |
| CST2 | 0.597 | < 1e-07 | SERPINB10 | 0.624 | 1.00E-07 |
| POSTN | 0.595 | < 1e-07 | CCL26 | 0.619 | 1.00E-07 |
| CLCA1 | 0.584 | < 1e-07 | POSTN | 0.608 | 2.00E-07 |
| HPDL | 0.582 | < 1e-07 | CCL26 | 0.606 | 2.00E-07 |
| ZMAT4 | 0.568 | < 1e-07 | NRAP | 0.586 | 6.00E-07 |
| CD44 | 0.563 | < 1e-07 | CPA3 | 0.582 | 8.00E-07 |
| GSN | 0.558 | < 1e-07 | HPDL | 0.582 | 8.00E-07 |
| FETUB | 0.556 | < 1e-07 | CLCA1 | 0.58 | 8.00E-07 |
| IL1RL1 | 0.543 | < 1e-07 | SLC16A9 | 0.572 | 1.30E-06 |
| CCBL1 | 0.538 | < 1e-07 | PDIA5 | 0.57 | 1.50E-06 |
| MUC5B | -0.535 | < 1e-07 | ZMAT4 | 0.569 | 1.50E-06 |
| DHX35 | 0.53 | < 1e-07 | TPSAB1 | 0.569 | 1.50E-06 |
| SLC39A8 | 0.52 | < 1e-07 | DHX35 | 0.569 | 1.50E-06 |
| CD36 | 0.517 | < 1e-07 | RUNX2 | 0.553 | 3.30E-06 |
| NRAP | 0.505 | < 1e-07 | KCNE3 | 0.55 | 3.80E-06 |
| CISH | 0.503 | < 1e-07 | C6orf164 | -0.549 | 3.90E-06 |
| FOLR1 | -0.496 | < 1e-07 | MS4A2 | 0.547 | 4.30E-06 |
| KCNE3 | 0.494 | < 1e-07 | CHDH | 0.546 | 4.50E-06 |
| SERPINB2 | 0.492 | < 1e-07 | SIGLECP3 | 0.545 | 4.70E-06 |
| PDIA5 | 0.491 | < 1e-07 | TPSD1 | 0.542 | 5.50E-06 |
| RBPMS | -0.491 | < 1e-07 | SLC39A8 | 0.541 | 5.60E-06 |
| SLC24A3 | 0.49 | < 1e-07 | KAZALD1 | 0.54 | 5.80E-06 |
| SIGLECP3 | 0.482 | < 1e-07 | GSN | 0.538 | 6.40E-06 |
| SHISA2 | -0.481 | < 1e-07 | SERPINB2 | 0.537 | 6.80E-06 |
| MS4A2 | 0.478 | < 1e-07 | P2RY14 | 0.534 | 7.80E-06 |
| HDC | 0.478 | < 1e-07 | HDC | 0.533 | 8.10E-06 |
| CREB3L4 | 0.475 | < 1e-07 | CCBL1 | 0.533 | 8.10E-06 |
| KAZALD1 | 0.474 | < 1e-07 | SERPINB2 | 0.531 | 9.00E-06 |
| CRABP2 | -0.469 | 1.00E-07 | TMEM71 | 0.529 | 9.80E-06 |

| | | | | | |
|---|---|---|---|---|---|
| CPA3 | 0.468 | 1.00E-07 | FETUB | 0.528 | 1.00E-05 |
| HOMER2 | 0.468 | 1.00E-07 | IL1RL1 | 0.527 | 1.04E-05 |
| TMEM45A | -0.468 | 1.00E-07 | CHDH | 0.525 | 1.17E-05 |
| RUNX2 | 0.465 | 1.00E-07 | HPGDS | 0.52 | 1.42E-05 |
| COMTD1 | 0.461 | 1.00E-07 | CHDH | 0.517 | 1.62E-05 |
| C1orf186 | 0.46 | 1.00E-07 | CPEB1 | -0.509 | 2.25E-05 |
| ADCY4 | 0.46 | 1.00E-07 | ST6GAL1 | 0.508 | 2.41E-05 |
| CCL15 | -0.457 | 1.00E-07 | SLC18A2 | 0.505 | 2.68E-05 |
| TMEM111 | 0.456 | 1.00E-07 | LOC100506125 | -0.502 | 3.03E-05 |
| STRA13 | 0.456 | 1.00E-07 | C6orf164 | -0.501 | 3.15E-05 |
| PEG10 | -0.455 | 2.00E-07 | LOC440356 | -0.501 | 3.23E-05 |
| SREK1IP1 | 0.453 | 2.00E-07 | CSGALNACT1 | -0.498 | 3.54E-05 |
| CHDH | 0.452 | 2.00E-07 | CWH43 | -0.498 | 3.60E-05 |
| UGT2A1 | 0.45 | 2.00E-07 | CEP72 | 0.495 | 4.11E-05 |
| LRRC8D | 0.45 | 2.00E-07 | CHST15 | -0.494 | 4.17E-05 |
| FAM110C | 0.45 | 2.00E-07 | ST6GAL1 | 0.493 | 4.35E-05 |
| TUBB2B | -0.45 | 2.00E-07 | ELOVL5 | 0.492 | 4.61E-05 |

**Table E3.1** Comparison of physiologic and clinical characteristics by subject clusters (SCs). K-means clustering of the 549 FeNO-correlated genes identified 5 molecularly defined and clinically distinct subject clusters (SCs). This table compares extensive list of physiological and clinical features. Significant intergroup comparisons are these features are presented in Table 3.2 below.

| **Feature** | **SC1** | **SC2** | **SC3** | **SC4** | **SC5** | **p-value** |
|---|---|---|---|---|---|---|
| FEV1% pred mean±SEM | 85±19 | 66±21 | 66±22 | 76±26 | 83±16 | <0.0001 |
| % change post BD median($25^{th}$-$75^{th}$ percentile | 7 (4-14) | 19 (10-41) | 15 (11-21) | 10 (5-14) | 11 (6-16) | 0.0001 |
| Post BD FEV1% pred mean±SEM | 94±13 | 77±21 | 75±26 | 82±24 | 87±12 | 0.0035 |
| AQLQ (total) Mean±SEM | 6±1 | 4±1 | 4±2 | 5±1 | 5±2 | <0.0001 |
| Wheezing frequency per week median $25^{th}$-$75^{th}$ quartile | 1 (1-3) | 4 (3-5) | 4 (2-5) | 3 (1-4) | 2 (1.5-3) | <0.0001 |
| Shortness of breath frequency per week median $25^{th}$-$75^{th}$ quartile | 2 (1-3.5) | 4 (3-5) | 4 (3-6) | 3 (1-5) | 3 (2-3.5) | <0.0001 |
| Nocturnal awakening frequency per week median $25^{th}$-$75^{th}$ quartile | 1 (1-2) | 3 (1-5) | 2 (1-4) | 1 (1-4) | 2 (1-2.5) | 0.0002 |
| Chest tightness frequency per week median $25^{th}$-$75^{th}$ quartile | 2 (1-3) | 3 (2-5) | 3 (2-5) | 2 (1-5) | 2 (2-3) | 0.0006 |
| Number positive skin tests median $25^{th}$-$75^{th}$ quartile | 2 | 4 | 3 | 3.5 | 4 | 0.016 |
| Asthma caused by pets (%) | 42 | 70 | 55 | 27 | 71 | 0.016 |
| Asthma caused by exercise (%) | 45 | 84 | 80 | 45 | 79 | 0.0003 |
| Asthma caused by infection (%) | 57 | 91 | 86 | 73 | 86 | 0.0017 |
| Asthma caused by routine activity (%) | 30 | 69 | 64 | 45 | 71 | 0.0011 |
| Any inhaled CS use (%) | 40 | 66 | 79 | 60 | 71 | 0.021 |
| 3 oral CS use in last year (%) | 11 | 24 | 53 | 20 | 21 | 0.011 |
| Uses oral CS (%) | 4 | 27 | 47 | 20 | 14 | 0.0019 |
| Uses anti-IgE therapy (%) | 2 | 2 | 21 | 10 | 7 | 0.046 |
| Hospitalized last 12 mos for asthma (%) | 4 | 17 | 22 | 27 | 24 | 0.08 |
| Intubated or ICU admission ever (%) | 6 | 29 | 22 | 27 | 35 | 0.02 |

**Table E3.2** Intergroup comparisons for significant features in Table E3.1 above.

| Feature | p-value | Significantly Different Clusters* |
|---|---|---|
| FEV1% pred mean±SEM | <0.0001 | SC1>SC2, SC1>SC3, SC5>2 |
| % change post BD median ($25^{th}$-$75^{th}$ percentile | 0.0001 | SC1 > SC2, SC1>SC3 |
| Post BD FEV1% pred mean±SEM | 0.0035 | SC2>SC1, SC3>SC1 |
| AQLQ (total) Mean±SEM | <0.0001 | SC1>SC2, SC1>3 |
| Wheezing frequency median $25^{th}$-$75^{th}$ quartile | <0.0001 | SC2>SC1, SC3>SC1, SC2>SC5 |
| Shortness of breath frequency median $25^{th}$-$75^{th}$ percentile | <0.0001 | SC2>SC1, SC3>SC1, SC2>SC5, SC3>SC5 |
| Nocturnal awakening frequency median $25^{th}$-$75^{th}$percentile | 0.0002 | SC2>SC1, SC3>SC1 |
| Chest tightness frequency median $25^{th}$-$75^{th}$ percentile | 0.0006 | SC2>SC1, SC3>SC1 |
| Number positive skin tests median $25^{th}$-$75^{th}$ percentile | 0.016 | SC5>SC1 (p=0.0052**) |
| Asthma caused by pets (%) | 0.016 | NA† |
| Asthma caused by exercise (%) | 0.0003 | NA† |
| Asthma caused by infection (%) | 0.0017 | NA† |
| Asthma caused by routine activity (%) | 0.0011 | NA† |
| Any inhaled CS use (%) | 0.021 | NA† |
| 3 oral CS use in last year (%) | 0.011 | NA† |
| Uses oral CS (%) | 0.0019 | NA† |
| Uses anti-IgE therapy (%) | 0.046 | NA† |
| Hospitalized last 12 mos for asthma (%) | 0.08 | NA† |
| Intubated or ICU admission ever (%) | 0.02 | NA† |

* All significant intergroup differences listed achieved a Bonferroni corrected p value of $p \leq 0.005$.
**Marginally greater than strict Bonferroni corrected p-value (p=0.005) for multiple comparisons.
†NA: not applicable. There was no significant difference in intergroup comparisons using a strict Bonferroni corrected p-value (p=0.005).

**Table E4.1**. Comparison of blood and lung inflammatory cell characteristics by subject clusters (SCs). K-means clustering of the 549 FeNO-correlated genes identified 5 molecularly defined and clinically distinct subject clusters (SCs). This table compares extensive list of blood and lung inflammatory cell characteristics. Significant intergroup comparisons are these features are presented in Table 4.2 below.

| | **SC1** | **SC2** | **SC3** | **SC4** | **SC5** | **p-value*** |
|---|---|---|---|---|---|---|
| Blood eosinophils (%) | 2 (1-3) | 4 (3-6) | 2 (1-5) | 3 (2-3) | 4 (3-8) | <0.0001 |
| BAL eosinophils (%) | 0.3 (0-0.6) | 0.6 (0.2-1.3) | 0.6 (0.1-2.2) | 0.3 (0-0.8) | 0.9 (0.2-2.8) | 0.0045 |
| Blood neutrophils (%) | 56 (50-62) | 58 (47-67) | 62 (54-73) | 61 (53-65) | 55 (46-60) | 0.05 |
| BAL neutrophils (%) | 1.7 (0.6-3.0) | 1.5 (0.8-3.3) | 4.9 (1.8-9.5) | 2.9 (2.1-3.9) | 1.7 (0.5-3.2) | 0.0107 |
| Blood lymphocytes (%) | 33 (29-40) | 29 (24-38) | 26 (14-29) | 27 (25-34) | 34 (29-39) | 0.0053 |
| BAL lymphocytes (%) | 7.7 (3.6-12.2) | 5.9 (3.3-11.5) | 12.9 (5.8-19.5) | 13.2 (7.3-20.8) | 7.8 (5.3-11.7) | 0.0156 |

**Table E4.2** Intergroup comparisons for significant features in Table E4.1 above.

| Feature | p-value | Significantly Different Clusters* |
|---|---|---|
| Blood eosinophils (%) | <0.0001 | SC2>SC1, SC5>SC1, SC5>SC4 |
| BAL eosinophils (%) | 0.0045 | SC2>SC1, SC5>SC1 (p=0.0075**) |
| Blood neutrophils (%) | 0.0500 | SC3>SC5 (p=0.0056*) |
| BAL neutrophils (%) | 0.0107 | SC3>SC1, SC3>SC2 (p=0.0080**), SC3> SC5 |
| Blood lymphocytes (%) | 0.0053 | SC1>SC3, SC5>SC3 |
| BAL lymphocytes (%) | 0.0156 | SC3>SC2 (p=0.0080**). |

* All significant intergroup differences listed achieved a Bonferroni corrected p value of $p \leq 0.005$.
**Marginally greater than strict Bonferroni corrected p-value (p=0.005) for multiple comparisons.

**Table E5**. Comparison of nasal and/or sinus disease by subject clusters (SCs). K-means clustering of the 549 FeNO-correlated genes identified 5 molecularly defined and clinically distinct subject clusters (SCs). This table compares nasal and sinus disease.

| | **SC1** | **SC2** | **SC3** | **SC4** | **SC5** | **p-value*** |
|---|---|---|---|---|---|---|
| Ever have nasal polyp (%) | 4 | 7 | 26 | 9 | 0 | 0.013 |
| Ever have nasal polyp removed (%) | 2 | 7 | 17 | 18 | 0 | 0.047 |
| Ever have sinus surgery (%) | 8 | 6 | 32 | 18 | 6 | 0.019 |

* All significant intergroup differences listed achieved a Bonferroni corrected p value of $p \leq 0.005$.

**Table E6.** Results of correlation between microarray vs qRT-SCR. To confirm expression of differentiating genes identified on the microarrays, RNA was extracted from the same bronchial brushing cells for real-time quantitative reverse transcription-polymerase chain reaction (qRT-SCR). The genes were selected by overall relevance to this and other analyses, as well as by known clinical interest.

| Entrez Gene Symbol | Gene cluster (GC) | Spearman (r) correlation | P-value |
|---|---|---|---|
| TFF3 | 8 | 0.9317 | 0.001 |
| BPIFA1 | 1 | 0.93 | 0.001 |
| POSTN | 9 | 0.9088 | 0.001 |
| NOS2 | 7 | 0.9 | 0.001 |
| LTF | 4 | 0.89 | 0.001 |
| CLCA1 | 8 | 0.8827 | 0.001 |
| TPSB2 | 8 | 0.872 | 0.001 |
| EGR2 | NA* | 0.87 | 0.001 |
| SCGB3A1 | 2 | 0.864 | 0.001 |
| GSTA2 | 4 | 0.855 | 0.001 |
| DUOX2 | 7 | 0.824 | 0.001 |
| FOSB | NA* | 0.8 | 0.001 |
| PTGS1 | 8 | 0.7843 | 0.001 |
| MUC5AC | 8 | 0.737 | 0.001 |
| CCL26 | 9 | 0.72 | 0.001 |
| IL18R1 | 8 | 0.6879 | 0.001 |
| IL1RL1 | 8 | 0.5136 | 0.001 |
| JUNB | NA* | 0.418 | 0.001 |
| ACHE | NA* | 0.312 | 0.002 |
| CD86 | NA* | 0.3 | 0.0025 |
| ERBB2 | 1 | 0.256 | 0.012 |

*NA: not applicable. These genes were not among those (n=1384) that most highly differentiated the SCs. Therefore; it was not clustered into a gene cluster.

**Tables E7.1-7.9** Lists of the 1384 genes differentially expressed between at least 2 SCs with high significance ($p<10^{-6}$) on intergroup t-test comparison. Gene expression of all differentially expressed genes was compared between SCs by intergroup student t-tests. Only those genes which met a p-value of $p<10^{-6}$ between at least 2 of the 5 SCs were then chosen to be included in the enhanced geneset. Biclustering with the 'enhanced' gene set (Fig. 3) identified 9 gene clusters (GCs). The 1384 genes are presented below according to GC.

**Table E7.1** Gene Cluster 1 (GC1) genes (n=131). This table lists co-expressed genes that cluster together in GC1 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns.

| Genes by Entrez Gene Symbol | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| BPIFA1 | FAM117A | SOBP | AP3B2 | ALOXE3 | C8orf85 | GHRHR | LOC100132287 | PRRT1 | MALL | TNFSF4 |
| CHST13 | DMRTC1 | NSUN7 | ZNF551 | ERBB2 | PRLHR | HHIPL1 | PTGIR | AMOTL2 | RHOV | |
| CHI3L1 | C19orf42 | AFAP1L1 | PATZ1 | HPN | DPPA5 | ACHE | PLAC4 | FGFR2 | DDIT4L | |
| CRABP2 | MGC3771 | CEP68 | TSHZ1 | KMO | SPATA3 | MINK1 | BHLHE22 | ZMIZ1 | RPRM | |
| C3 | EPB41L1 | C11orf95 | FAM124A | ADARB2-AS1 | CYP2A13 | EPHA10 | HCRT | CYP24A1 | ACOX2 | |
| LAG3 | GTF2IRD1 | SPRY1 | MAMLD1 | C20orf46 | CPXM1 | THEG | SLAMF9 | CXCL6 | TMEM25 | |
| LYPD2 | GTPBP1 | SCD5 | ABCC3 | TGM7 | PPAN-P2RY11 | CD300LB | NCKAP5L | SCNN1G | JAM3 | |
| C2orf65 | TTC22 | EPAS1 | ZNF185 | MANBA | SCDHB7 | GPR144 | MT1G | MUC5B | VASN | |
| CLCA4 | GRAMD2 | HLF | PLA1A | VGLL2 | KRTAP1-1 | OTUD7A | KCND3 | PGLYRP4 | ATP8A2 | |
| CES2 | PFKM | MKL2 | SLC28A3 | WHSC2 | NAT8B | ATCAY | MYOG | FAM181A | RGN | |
| PACS2 | GLI3 | NEURL3 | C8orf68 | CCDC169 | SH3BP1 | PPP1R3F | NELF | NOX1 | KCNMA1 | |
| TSCN1 | GPR125 | FHDC1 | CCL15 | MMP16 | S1PR2 | GPR156 | SPACA1 | SLAMF7 | FRMD4A | |
| SETDB1 | PIP4K2B | OLFM1 | DKFZP434F142 | IL36RN | TCF23 | C7orf26 | PNMA5 | LOC100132330 | THSD7A | |

**Table E7.2** Gene Cluster 2 (GC2) genes (n=171). This table lists co-expressed genes that cluster together in GC2 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns.

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| BZRAP1 | C3orf25 | TMEM231 | HYI | BCAS3 | KNDC1 | SPEN | TP53I11 | KIAA1377 | EFCAB6 |
| CCDC108 | LRRC71 | MAFIP | C17orf50 | C20orf195 | PRKCDBP | SIN3A | FGFRL1 | SYNE1 | DNAH10 |
| CCDC17 | SFI1 | GSTM2 | TRIM39 | PDLIM4 | HNF1A | SMAD3 | CALCOCO1 | KSR1 | HYDIN |
| ARHGAP39 | FARP2 | UBXN11 | ARL3 | PRDX5 | HNF1A-AS1 | HDAC4 | USP21 | ZXDC | MICAL3 |
| C9orf174 | C14orf79 | C2orf81 | SCYT1B | CNNM2 | DHRS12 | NCAPD3 | RBPMS | ADD1 | WDR16 |
| DNAH3 | GAS2L2 | ST5 | SFXN3 | C10orf95 | FBRSL1 | LOC283174 | CCDC88C | ILF3 | ZSCAN18 |
| SPATA6 | ZBTB48 | SCARF2 | GRAMD1A | IFT140 | SEC14L4 | IPW | LOC100507904 | TBX1 | SUGP2 |
| C1orf228 | ANKRD54 | SUV420H2 | PPP1R26 | PPP1R32 | CORO2B | RALGPS1 | LOC729737 | PEAK1 | EZH1 |
| MAPK10 | BBS1 | ASPG | PRPF8 | PTOV1 | PTPRT | DHCR24 | FAM86EP | KLC1 | TMEM8B |
| CCDC135 | KIAA0556 | C2orf50 | KDM4D | USP11 | INSR | RBMX | SSBP4 | THBS3 | PLIN5 |
| IFT122 | INTS3 | GYG2 | USP2 | WFS1 | CSH1 | C1orf93 | DTX3 | CECR7 | KIF19 |
| CCDC151 | C12orf51 | ARX | SLC25A42 | BIN3 | SCGB1A1 | TBL1X | LOC100132247 | LOC100506990 | TCTE1 |
| DNAI1 | STRADA | CRY2 | CES4A | PMM1 | SCGB3A1 | FXYD1 | INPP5E | DZANK1 | CHST15 |
| P4HTM | C16orf48 | JMJD5 | LAMC2 | SRGAP3 | DNAJC17 | SLC29A1 | POFUT2 | ZNF337 | MUM1 |
| CCDC42B | LRRC46 | FTO | EXOC7 | TCEA2 | BAHCC1 | ATP6V1B1 | LOC727849 | MIR600HG | FAM193A |
| DNAH1 | ATXN2 | ARHGEF17 | DALRD3 | LMAN2L | NTN1 | CYP2F1 | DZIP1L | SNED1 | POMT1 |
| FBF1 | RABL2A | ST6GALNAC2 | TTC16 | KIF17 | MECP2 | SUSD4 | WDR65 | PRSS12 | ZNF212 |
| ZNF554 | | | | | | | | | |

**Table E7.3** Gene Cluster 3 (GC3) genes (n=52). This table lists co-expressed genes that cluster together in GC3 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | |
|---|---|---|---|---|
| C16orf7 | CEP164 | MUSTN1 | CYTH1 | DUOXA1 |
| CABIN1 | RFX1 | TNNI3 | IKBKB | LINC00479 |
| DMPK | MAP4K4 | IDS | KIAA0895L | ANKRD18A |
| CCDC57 | MBL1P | FER1L5 | CAPN3 | GOLGA6L9 |
| HGS | POPDC2 | OVGP1 | PDXDC2P | FHAD1 |
| SPPL2B | CAMSAP1 | FAM193B | FAM95B1 | ZFHX2 |
| FLJ20021 | FBXO31 | MYO15B | RANBP10 | NOL3 |
| ARHGAP33 | TMC5 | NSUN5 | NXF5 | PMS2L2 |
| CACNB1 | C16orf93 | RAD9A | STK36 | LINC00260 |
| SLC25A35 | CCDC24 | TMEM234 | WDR90 | PAQR6 |
| TRPV1 | TMEM44 | | | |

**Table E7.4** Gene Cluster 4 (GC4) genes (n=325). This table lists co-expressed genes that cluster together in GC1 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C10orf114 | WDR19 | MGP | RGAG4 | PIAS1 | C18orf1 | POMGNT1 | C2orf72 | BEX1 | ZFYVE28 |
| BICC1 | PTPRM | FOLR1 | JAG2 | KLF12 | LY75 | BCL2 | FXYD6 | DDC | ALDH3B1 |
| C8orf34 | TTLL6 | LMO3 | ZNF134 | TMEM133 | ASCDD1 | SMO | SPHKAP | SCGN | CCDC69 |
| CXorf57 | C3orf18 | GATA6 | EPDR1 | PTPN14 | FOXA2 | KANK1 | CDH18 | SV2B | C17orf79 |
| C20orf26 | MAOB | NDNF | ZNF497 | RHOU | RASSF10 | BEX2 | RNF182 | CHGB | CCDC102A |
| CPEB1 | ASTN2 | EFHD1 | AKAP7 | FMN2 | KCTD12 | SMOC2 | GNG7 | SYT1 | CSNK2A2 |
| HSPA12A | CCDC8 | GPRASP2 | CD302 | BMI1 | OSBPL6 | DKK3 | CACNA2D2 | SCN3A | MORN4 |
| RIBC1 | TMEM121 | B3GNT1 | HNMT | HEY2 | SLC22A4 | ISLR | SFTA1P | NOL4 | PDE6B |
| C11orf88 | FABP6 | NDN | SLC30A1 | PARD6B | SLC6A16 | NRG1 | SUSD2 | SCSK1 | CLU |
| KATNAL1 | PKD2 | ZNF671 | ANKRD46 | ESRRG | FAM167A | ANKH | ATP8B2 | SYT4 | FLNB |
| SHANK2 | BHLHE41 | CHN1 | GCOM1 | TMEM27 | KIAA1244 | FBXL16 | DUSP16 | LINC00261 | STXBP1 |
| FLJ37644 | CBX7 | PEG10 | PQLC3 | APPBP2 | LRBA | THSD4 | PTPRA | SYT13 | TTC5 |
| LRP8 | LRIG1 | SGCE | SMARCA1 | PSIP1 | OSBPL3 | SCN2B | TERF2IP | C2CD4B | ENO2 |
| RBKS | SESN1 | PNMAL1 | KALRN | EFEMP1 | ZNF709 | BCL11A | PNMA2 | PER2 | ZNF853 |
| C6 | HHLA2 | C14orf132 | NAV3 | GPRASP1 | SLIT1 | PRKCA | PDLIM3 | EFNB2 | TGM3 |
| KCNN3 | SEC14L3 | DSCAML1 | PIEZO2 | SSBP2 | ZBTB46 | AR | RCAN2 | IRS2 | FAM50B |
| RAI2 | GSTA2 | FHOD3 | ADARB1 | TMEM237 | CNTN3 | SETBP1 | ANXA6 | CA8 | NEURL1B |
| MSRB3 | GSTA1 | PROS1 | CRB1 | ZAK | C2orf67 | IRX2 | CLDN9 | PRKAR2B | LRRC27 |
| NEGR1 | GSTA3 | ARHGEF26 | MAGI2-AS3 | THBS4 | SNCAIP | PPP1R9A | KRTCAP3 | CNBP | NEDD4L |
| SAV1 | GSTA5 | SIAH3 | CDON | TOX | CPNE8 | SPOCK3 | NACAD | GAB2 | MYO10 |
| CDKL1 | SEMA3E | ATP2A2 | UBE2I | FLJ42709 | RAB3IP | GJB7 | INA | EIF1B | PSMF1 |
| COLQ | ACADM | NPAS3 | RTN4RL1 | SLITRK5 | GLS | SCDH17 | DUSP26 | SERTAD2 | WWC3 |
| C8orf42 | KCNA1 | IGFBP5 | WNT5B | PTGFR | CDC42BPA | ST8SIA4 | GRP | TMCC1 | CSGALNACT1 |
| PRUNE2 | VGLL3 | PLD6 | USP27X | RBPJ | MYO9A | AP3M2 | KCNC1 | HSD17B2 | ERP27 |
| ARHGAP44 | C18orf45 | DDAH1 | CLDN8 | SERP1 | ST13 | PGM5 | INSM1 | RPRML | B3GAT1 |
| DNAH9 | EPB41L2 | SCRN1 | ECI2 | NPNT | PHF2 | RAB6B | SCG5 | CDK14 | TWIST2 |
| LOC158960 | CNTD1 | SSPN | GXYLT2 | TSPAN12 | TUB | GLIS3-AS1 | HOXB5 | STEAP4 | LTF |
| IL5RA | WNK4 | NPR3 | KLF15 | SERPINB7 | USP30 | SHISA2 | SUCNR1 | LOC100288911 | TMEM45A |
| RNF32 | SNTB1 | PPP4R4 | GAS2 | CLDN16 | TGFBR3 | LINC00478 | OR51E1 | TNS3 | OPRL1 |
| EML1 | LINC00341 | WIF1 | PRKCE | CYP4X1 | ALDH3A2 | SMARCD3 | FEZF2 | PNMA3 | ASXL1 |
| SPAG6 | TRMT5 | SLC16A12 | NFIB | NFATC2 | LINC00086 | TBL1Y | MAPT | SLC41A1 | KDM2B |
| SCDP1 | C16orf89 | CCND1 | CYP2J2 | PRICKLE2 | LINC00087 | ZNF469 | KCNA4 | PRDX6 | WNK2 |
| PIGG | POLA2 | LOC646976 | ZNF425 | TRIM68 | | | | | |

**Table E7.5** Gene Cluster (GC5) 5 genes (n=50). This table lists co-expressed genes that cluster together in GC5 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | |
|---|---|---|---|---|
| BTD | FAM174B | LOC100506495 | EFNB3 | TNFSF12 |
| C17orf58 | SPR | TBC1D8 | TMEM146 | PROCR |
| FAM46C | SEMA4F | CAMK2D | SCN4B | HES5 |
| ST3GAL6 | C16orf55 | PFDN5 | SNURF | GMNN |
| ATOH8 | CCDC56 | PLCL2 | METRN | PART1 |
| CLDN22 | LOC400236 | HMGN2 | LRTOMT | GPANK1 |
| GPR116 | ABCA4 | HMGN2P46 | FGF14 | PPP1R3B |
| SLC16A11 | TOMM34 | KCNMB2 | UBXN10 | RASSF3 |
| TMEM98 | SHF | RPS27L | GPR34 | RPL21 |
| ZNF704 | LOC729970 | C1orf87 | BTK | RPL34 |

**Table E7.6** Gene Cluster 6 (GC6) genes (n=150). This table lists co-expressed genes that cluster together in GC6 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ARID5B | ATP5H | CYP2C9 | CKAP2 | C8orf76 | GMPS | NIPAL2 | FAM177A1 | ENTPD7 | IDE |
| GJB2 | OCIAD2 | CTTNBP2NL | OAT | GART | BORA | SLC35G1 | POC1B | SLC30A7 | PLA2G12A |
| LAMB1 | MED1 | ITPRIPL2 | SRPK1 | HIGD1A | RTKN2 | ENTPD4 | IMPA1 | CNTLN | CCDC72 |
| ALDH4A1 | IFI16 | ANKRD36 | SYNCRIP | RIT1 | CHMP3 | TBC1D8B | NNT | CYCS | C12orf57 |
| LAD1 | MIS18BP1 | ANKRD36BP2 | ATP1B3 | TIMM17A | SMAP1 | PHLDB2 | TMEM87B | OCLN | UQCRQ |
| CXCL14 | OSTF1 | LOC100506123 | PRCP | UCHL3 | ARSC3 | PRTFDC1 | DENND4A | UGGT1 | NDUFA1 |
| MCAM | NDUFB9 | SERPINB13 | TACC1 | CCK | WSB2 | ACTR3 | CMYA5 | IBTK | KIAA0020 |
| FHL2 | CCDC90B | CEP152 | BDH2 | CYP27B1 | BDNF | BZW1 | RUNX2 | PARD3B | TAOK3 |
| KIAA1199 | MTERF | IRAK4 | IGF2BP3 | PTHLH | VPS13D | TPM3 | PGRMC2 | ATP10D | SLC44A1 |
| SCP4L1 | LIMA1 | UGGT2 | BMS1 | GAD1 | DNAJC10 | ADAM9 | SLC10A7 | ITGB6 | RARS |
| TUBA4A | POF1B | NET1 | LOC96610 | CD109 | KRAS | LRRC8D | C16orf54 | ERO1L | SH3GLB1 |
| ACOT13 | PSMC1 | SLK | ETFA | PXDN | PHKB | AGR2 | CPT1A | BCAP29 | RBM28 |
| C1orf31 | SLC25A46 | SYTL2 | FAM3B | DPYSL3 | SSFA2 | PDIA3 | NAV1 | POMP | OBFC2A |
| FAM162A | TIGD2 | C3orf14 | GPR87 | GRK5 | PPAPDC1B | RPN1 | SCYT1A | CPNE3 | TPM4 |
| NAA38 | CYP2C18 | C6orf115 | RPL22L1 | TMEM200A | CYP2R1 | CCPG1 | GRIK2 | GFPT1 | STK38L |

**Table E7.7** Gene Cluster 7 (GC7) genes (n=228). This table lists co-expressed genes that cluster together in GC7 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns.

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| DENND2C | TIMELESS | KPNA2 | CALML5 | PLEKHJ1 | C11orf24 | PMM2 | SOX21 | AHCYL2 | TMEM165 |
| GPSM2 | HENMT1 | TROAP | HPDL | SEMA3B | ICA1 | S100A16 | INPP4B | RAB27A | ABCC1 |
| KDM4C | SNTB2 | GPR107 | NOS2 | SIN3B | B4GALT2 | NUDT8 | USP28 | PARM1 | DGKA |
| PPIL2 | ARHGEF16 | MFSD2A | C6orf168 | DNAJC22 | GNPNAT1 | TSTA3 | ADM | SLC5A8 | TNC |
| CEACAM1 | MST1R | CHST6 | COPZ2 | BACE2 | GJB1 | TIMP1 | ANGPTL4 | OSTalpha | TMEM171 |
| IRAK3 | TRABD | TRADD | FAM110C | B3GNT6 | LOC100129034 | GALNT5 | INF2 | TMPRSS4 | TNFRSF18 |
| ATP10B | FAM171A1 | ELL3 | MDFI | CASP6 | NEK6 | CEACAM5 | HN1 | AKAP12 | CKMT1A |
| SAT1 | FGFR1 | FHOD1 | RHOC | SLC39A11 | FKBP11 | S100A14 | TUBA1C | IL18R1 | GAPDH |
| DUOX2 | LAPTM4B | STX18 | ANKRD9 | FAM83D | MAGED1 | S100P | PNP | AK2 | TAGLN2 |
| KLHDC7B | FBXL7 | SPAG5 | C9orf142 | HOMER2 | TXNDC5 | FAM114A1 | HIST1H1A | MANF | MYO5B |
| MX2 | TPRXL | COMTD1 | CPSF4 | KRT23 | NPDC1 | RAB11A | HIST1H3B | SYNGR2 | SRPX2 |
| CARD14 | ESRRA | DNAJC1 | PSMD11 | FAM150A | SHROOM1 | UBE2J1 | ASPM | CHPF2 | TXN |
| HRASLS | HOMER3 | PM20D2 | SPHK1 | ANXA3 | TRPM4 | MAPK13 | CASC5 | TEP1 | KRT6A |
| DQX1 | PTK6 | LGALS7 | ULBP2 | KCNE3 | CCDC167 | GALE | NMU | TBC1D7 | FUT3 |
| RCC2 | TP53I3 | LGALS7B | CSTA | MREG | BOLA2B | TMED3 | PTTG1 | WHSC1 | PTPRH |
| LOC100292909 | AGPAT4-IT1 | DEFB1 | LOC100505633 | FUT5 | EIF4E2 | UROS | PTTG2 | ATP1B1 | SDCBP2 |
| SNTG2 | LOC100506029 | S100A10 | DISP1 | FUT6 | CYB561D2 | SMPDL3B | KIF2C | SLC26A2 | TCN1 |
| LOXL1 | MUC4 | FZD5 | ASRGL1 | KCNK6 | GPR172A | ABHD2 | PBK | GPT2 | UPK1B |
| DIAPH3 | SLC9A8 | LRRC8A | TMEM117 | TBC1D16 | ECE2 | CTNNA1 | TYMS | TLE1 | KRT18 |
| NCAPG | ACADVL | FERMT1 | LOC728855 | HIST1H2BK | MTFP1 | CIB1 | C1orf122 | FKBP14 | SERPINB8 |
| NDC80 | SGSM3 | MORC4 | SLC7A4 | PRSS16 | ARMCX6 | TSPAN1 | RRP36 | POP1 | TRIM7 |
| PIF1 | KIAA1875 | LOXL4 | BICD1 | CREB3L4 | GMDS | KBTBD12 | CATSPERB | LOX | TMEM165 |
| KIAA0101 | EZH2 | ADCY4 | SH3RF2 | STRA13 | HM13 | KLF4 | MAP2K6 | PLA2G4A | ABCC1 |

**Table E7.8** Gene Cluster 8 genes (n=96). This table lists co-expressed genes that cluster together in GC8 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C4orf52 | CPNE4 | PDIA5 | CGREF1 | KAZALD1 | PTGS1 | CNRIP1 | CD1E | PRELID2 | MYO1C |
| EEF1B2 | MZB1 | MUC5AC | TFF2 | ZMAT4 | P2RY14 | C1orf186 | DPP4 | MFSD4 | TPM1 |
| ATM | CCBL1 | LOC401127 | SCGB1D2 | PRB1 | CTSG | KIT | ADAMTSL3 | WWP1 | VTI1B |
| SERINC5 | MAGED4B | TFF3 | SDC2 | PRB2 | HDC | CPA3 | ELOVL5 | ALPL | PTGDR2 |
| NEDD9 | CLCA1 | RABAC1 | LIPH | PRR4 | UTS2 | TPSAB1 | TSPAN13 | CTSC | TNFSF11 |
| SLAIN2 | CST1 | HMGCS2 | EGFL6 | MPPED2 | CATSPER1 | MS4A2 | TSPAN3 | FOXA3 | SLC9B2 |
| BMP8A | CST2 | DHX35 | TTPA | PTPRN2 | P2RX1 | SIGLECP3 | SLC7A1 | GALNT12 | MYO1C |
| ANO7 | FETUB | NRAP | FGF13 | KRT1 | GATA2 | SLC18A2 | GNE | CHDH | TPM1 |
| ADRA2A | SLC24A3 | ITLN1 | FMOD | CD22 | IL1RL1 | TPSD1 | ST6GAL1 | P2RX4 | VTI1B |
| C1orf64 | GSN | GUCA2B | PON3 | HPGDS | RGS13 | CD1C | PRR15 | HSCAL1 | PTGDR2 |

**Table E7.9** Gene Cluster 9 (GC9) genes (n=180). This table lists co-expressed genes that cluster together in GC9 by complete linkage hierarchical clustering The order matches the order of the heat map in Fig. 3, starting in the top left corner and going from top to bottom along rows, then left to right along columns

| Genes by Entrez Gene Symbol | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| DHRS7 | OSTC | CD44 | PELO | MYL12B | C4orf34 | SLC39A8 | SEC24D | ENTPD3 | SPEG |
| DHX32 | SEC11C | DOK1 | MRPL15 | TMEM50A | SLC17A5 | TSPAN8 | DNAJC12 | LAMP2 | PHYH |
| EEF1G | MPZL1 | CORO1C | UQCRH | COX17 | C14orf109 | AADAC | SLC39A14 | TPD52 | PRKD1 |
| ACOT9 | CNPY2 | MLKL | TNNC1 | LEO1 | ALG5 | FKBP7 | PGM3 | SLC31A1 | SEC14L1 |
| ATP5G3 | SDF2L1 | RIPK2 | ATL3 | ATP5O | PSCS | PRDX4 | WBP5 | TCEA3 | SERPINA11 |
| SLC25A5 | TMEM125 | GALNT1 | REEP5 | TMEM111 | FBXO25 | TFF1 | FOLH1 | ALOX15 | MGST2 |
| FGFBP2 | ANG | ISOC1 | NME1 | RBM18 | TCEAL8 | ERLEC1 | FOLH1B | GALNT4 | SREK1IP1 |
| HDAC9 | RNASE4 | ERP44 | PPIB | SPECC1 | TMEM9B | KDELR2 | LOC100289255 | AHSG | SEC22A |
| BCL2L14 | SRD5A2 | SFT2D1 | SFN | POLR2K | TFB2M | MAPRE2 | CLDN10 | CCL26 | LRRC17 |
| CCDC109B | C14orf28 | ARNTL2 | GALNT10 | TMEM167A | TXNDC15 | KIF16B | YIPF1 | SERPINB10 | UGT2A1 |
| MTHFD2 | STXBP5 | RAB27B | JTB | NOP10 | CPE | MUM1L1 | LRRC31 | CDH26 | UGT2B10 |
| RHOH | MCTP2 | TBC1D15 | FAM109B | SSCS3 | HEPACAM2 | SYTL4 | VSIG2 | SERPINB2 | MUC2 |
| AGPS | MLLT3 | ECT2 | TNFRSF11B | TXNDC17 | EAF2 | FAM108C1 | PDGFRL | POSTN | ZNF467 |
| MAGT1 | DNAJC3 | TMEM183B | SLC50A1 | ARL1 | ARMCX3 | GALNT7 | KDELR3 | CEP72 | SLC12A2 |
| ERGIC1 | SLC2A13 | ARSC2 | SPINT2 | GGH | SSR3 | MIA3 | PYCR1 | PP7080 | SLC6A14 |
| SDR16C5 | METTL9 | NME2 | TMBIM6 | PTS | FUT8 | PPAPDC2 | WFDC1 | HS3ST4 | SLC16A9 |
| LMAN1 | PRRC1 | MYL6 | CD63 | GPR160 | C12orf23 | GORASP2 | HK2 | NTRK2 | TMEM71 |
| CISD2 | CD36 | S100A6 | MYL12A | SLC35A1 | EDEM3 | LRRC59 | SLC45A3 | UTP20 | VTCN1 |

**Table E8**. Summary of Ingenuity findings believed to be of biological relevance and/or support the gene cluster analysis in the paper's Results section. In IPA, global canonical pathway (GCP) analysis and upstream regulator analysis were considered using threshold cutoff of p-value < $10^{-2}$. Briefly, GCP determines genes associate together by chance, and assigns a likelihood p-value. Upstream regulator analysis attempts to discover upstream transcriptional regulators (e.g. cytokines, transcriptional factors, drugs, etc.) that could explain changes in gene expression

| GC | Type | Rank | Pathway or Upstream analysis | p-value |
|---|---|---|---|---|
| 1 | Canonical pathway | 1 | LPS/IL-1 Mediated Inhibition of RXR Function | 2.09E-03 |
| | Canonical pathway | 2 | Granulocyte Adhesion and Diapedesis | 3.86E-03 |
| 4 | Canonical pathway | 1 | Glutathione-mediated Detoxification | 7.8E-04 |
| | Upstream analysis | 5 | WNT3A | 4.89E-04 |
| 5 | Canonical pathway | 1 | PI3K Signaling in B lymphocytes | 2.4E-03 |
| 6 | Canonical pathway | 3 | Oxidative Phosphorylation | 4.17E-02 |
| | Canonical pathway | 4 | Mitochondrial Dysfunction | 2.79E-02 |
| | Canonical pathway | 6 | Actin nucleation by ARP-WASP | 4.48E-02 |
| | Canonical pathway | 8 | Remodeling of Epithelial Adherens Junctions | 4.29E-02 |
| | Upstream analysis | 1 | Atrophin 1 | 2.92E-05 |
| 7 | Canonical pathway | 3 | iNOS Signaling | 1.13E-3 |
| | Upstream analysis | 1 | TNF | 9.62E-06 |
| 8 | Upstream analysis | 1 | IL13 | 3.35E-07 |
| 9 | Upstream analysis | 1 | XBP1 | 5.40E-11 |

**Figure E1.** Gene expression was correlated with exhaled nitric oxide (FeNO) using a Spearman correlation method, identifying 549 genes that correlated strongly with FeNO. The variable k selected for k-means clustering was the largest value which produced a unique solution and more than ten subjects in the smallest cluster. As shown in this figure, the difference between error sum of squares for the array data and a randomized set of data was not improved with k greater than 4 [36], which confirmed the plausibility of the selection.

