Identification and prospective stability of electronic nose (eNose)-derived inflammatory phenotypes in patients with severe asthma

Paul Brinkman, MSc, a Ariane H. Wagener, MD, PhD, a Pieter-Paul Hekking, MD, PhD, a Aruna T. Bansal, PhD, b Anke-Hilse Maitland-van der Zee, PhD, b Yuan Yue Wang, MSc, a Hans Weda, PhD, c Hugo H. Knobel, MSc, d Teunis J. Vink, PhD, s Nicholas J. Rattray, PhD, b Arnaldo D’Amico, PhD, b Giorgio Pennazza, PhD, b Marco Santonico, PhD, g Diane Lefaudeux, MSc, b Bertrand De Meulder, PhD, h Charles Auffray, PhD, h Per S. Bakke, MD, PhD, i Massimo Caruso, MD, PhD, j Pascal Chanez, MD, PhD, k Kian F. Chung, MD, PhD, l Julie Corfield, MSc, m,n Sven-Erik Dahlén, PhD, s Ratko Djukanovic, MD, PhD, p Thomas Geiser, MD, PhD, s Ildiko Horvath, MD, PhD, q Nobert Krug, MD, PhD, r Jacex Musial, MD, PhD, t Kai Sun, PhD, u John H. Riley, PhD, v Dominic E. Shaw, MD, PhD, w Thomas Sandström, MD, PhD, x Ana R. Sousa, PhD, y Paolo Montuschi, MD, PhD, y Stephen J. Fowler, MD, PhD, z,a and Peter J. Sterk, MD, PhD, x on behalf of the U-BIOPRED Study Group*

Amsterdam and Eindhoven, The Netherlands; Cambridge, London, Nottingham, Southampton, Stockley Park, Manchester, and Preston, United Kingdom; New Haven, Conn; Rome and Catania, Italy; Lyon and Marseille, France; Bergen, Norway; Malmö, Stockholm, and Umeå, Sweden; Bern, Switzerland; Budapest, Hungary; Hannover, Germany; and Krakow, Poland

Background: Severe asthma is a heterogeneous condition, as shown by independent cluster analyses based on demographic, clinical, and inflammatory characteristics. A next step is to identify molecularly driven phenotypes using “omics” technologies. Molecular fingerprints of exhaled breath are associated with inflammation and can qualify as noninvasive assessment of severe asthma phenotypes.

Objectives: We aimed (1) to identify severe asthma phenotypes using exhaled metabolomic fingerprints obtained from a composite of electronic noses (eNoses) and (2) to assess the stability of eNose-derived phenotypes in relation to within-patient clinical and inflammatory changes.

Methods: In this longitudinal multicenter study exhaled breath samples were taken from an unselected subset of adults with severe asthma from the U-BIOPRED cohort. Exhaled metabolites were analyzed centrally by using an assembly of eNoses. Unsupervised Ward clustering enhanced by similarity profile analysis together with K-means clustering was performed. For internal validation, partitioning around medoids and topological data analysis were applied. Samples at 12 to 18 months of prospective follow-up were used to assess longitudinal within-patient stability.

Results: Data were available for 78 subjects (age, 55 years [interquartile range, 45-64 years]; 41% male). Three eNose-driven clusters (n = 26/33/19) were revealed, showing differences in circulating eosinophil (P < 0.045) and neutrophil (P < 0.017) percentages and ratios of patients using oral corticosteroids (P < 0.005). Longitudinal within-patient cluster stability was associated with changes in sputum eosinophil percentages (P < 0.045).

Conclusions: We have identified and followed up exhaled molecular phenotypes of severe asthma, which were associated with changing inflammatory profile and oral steroid use. This suggests that breath analysis can contribute to the management of severe asthma. (J Allergy Clin Immunol 2019;nnn:nnn–nnn.)
Key words: Electronic nose technology, exhaled breath, volatile organic compound, follow-up, severe asthma, unbiased clustering, eosinophils, neutrophils, oral corticosteroids

From a clinical point of view, asthma is defined as a disease of episodic and recurrent chest symptoms and variable airflow limitation, including features of airways inflammation and structural changes.\textsuperscript{4} When considering clinical management, international guidelines distinguish different levels of asthma control and severity.\textsuperscript{2,3} Patients with severe asthma comprise nearly 5% of the total asthmatic population.\textsuperscript{6} It is well recognized that severe asthma is a heterogeneous chronic inflammatory disease with several clinical presentations, physiologic characteristics, inflammatory and structural profiles, and outcomes.\textsuperscript{5,6} This heterogeneity of severe asthma shows the need to recognize distinct phenotypes, which might allow more personalized management.

Disease phenotyping is based on classification of patients into subgroups by using clinical and/or biological parameters.\textsuperscript{7,8} Identification of such subgroups can be accomplished through unbiased or hypothesis-driven methods,\textsuperscript{7,8,10} resulting in clusters that might or might not be linked to currently known pathophysiologic pathways. Based on clinical, physiologic, and/or cellular disease features (eg, disease onset, sex, body mass index, lung function, and eosinophilic profile), independent studies have revealed remarkably similar clusters of severe asthma.\textsuperscript{7,12} This supports the clinical potential of improving long-term disease outcomes by tailoring individual treatment.\textsuperscript{1,5,13}

A next step toward tailoring asthma management is to phenotype patients based on multimolecular profiles. The availability of high-throughput large-scale analytic methods and complex statistical and computational procedures is making the search for new biomarkers at a high-dimensional “omics” level a tangible option.\textsuperscript{7,16} Omics concerns the acquisition and analysis of such large-scale biological data sets with the aim to discover and identify biomarkers of diseases and new pathophysiologic mechanisms. This approach purposely avoids \textit{a priori} assumptions about markers that might be associated with a particular disease or phenotype.\textsuperscript{8} The omics field has been demonstrated to be valuable in asthma research. A transcriptomics study on airway epithelial brushings identified 2 asthma subgroups defined by the degree of T\textsubscript{H}2 inflammation,\textsuperscript{17} whereas gene expression profiling of induced sputum samples from patients with severe asthma showed not only 3 distinct phenotypes\textsuperscript{8} but also its capability to distinguish T\textsubscript{H}2-high and T\textsubscript{H}2-low subtypes of asthma.\textsuperscript{18} Furthermore, clustering of sputum cytokine-high profiles revealed 5 unique asthma molecular phenotypes.\textsuperscript{19} Although these data demonstrate the merit of molecular profiling in asthmatic patients, the technologies are analytically demanding and far from being applicable in daily medical care. Therefore there is a need to bridge the gap between omics technologies and clinical diagnostics and monitoring.

Metabolomics of exhaled air (breathomics) can serve the purpose of molecular profiling in asthma at point of care. Exhaled breath contains a complex gas mixture of volatile organic compounds (VOCs), which can be measured noninvasively. These composite VOC samples can be analyzed by detecting individual molecular compounds by using gas chromatography–mass spectrometry (GC-MS) or through pattern recognition of exhaled VOCs using cross-reactive sensor arrays with electronic nose (eNose) technology.\textsuperscript{20} Indeed, GC-MS has been applied in the discovery of biomarkers for the prediction of eosinophilic versus noneosinophilic asthma,\textsuperscript{21} the discrimination between clinically stable and unstable episodes of asthma,\textsuperscript{22} and the inflammatory phenotyping of patients with chronic obstructive pulmonary disease (COPD).\textsuperscript{23} Notably, when using the faster eNose technology,\textsuperscript{20} based on a combination of cross-reactive sensors and pattern recognition algorithms, exhaled breath from patients with asthma, patients with COPD, and healthy control subjects could be discriminated in several independent studies.\textsuperscript{24-26} Moreover, cluster analysis with an eNose among a study population of both asthmatic patients and patients with COPD resulted in clusters that are not determined based on diagnosis but rather based on clinical and inflammatory characteristics.\textsuperscript{27} This raises the questions of whether phenotyping of patients with severe asthma based on solely exhaled breath profiles is possible and what the within-patient stability of those phenotypes is over time.

We hypothesized that eNose technology is suitable for noninvasive identification of severe asthma phenotypes. Therefore our aims were (1) to identify severe asthma phenotypes using unbiased and benchmarked cluster analysis based on metabolomic fingerprints from exhaled breath by a composite of different brands of eNoses (U-BIOPRED eNose platform) and (2) to assess within-patient stability in eNose-derived clusters in relation to changes in clinical and inflammatory characteristics after 12 to 18 months of prospective follow-up.

METHODS

Subjects

In this study an unselected subset of adult subjects (aged ≥18 years) from the pan-European U-BIOPRED cohort study was included.\textsuperscript{7} All participants were given a diagnosis of severe asthma according to Innovative Medicines Initiative criteria.\textsuperscript{7} In short, these patients had a prescription for high-dose inhaled corticosteroids (≥1000 μg fluticasone propionate or equivalent) plus at least 1 other controller medication and symptoms defined as uncontrolled according to Global Initiative for Asthma guidelines and/or 2 or more severe exacerbations per year and/or required prescription of oral corticosteroids (OCSs) minimally daily to achieve asthma control. Subjects were excluded if they changed asthma medication or required high-dose OCSs or a doubling of maintenance OCS for at least 3 days because of a severe asthma exacerbation in the month before study visits. The study was approved by all local medical ethics committees, and all patients provided written informed consent. The study was registered at ClinicalTrials.gov under the identifier NCT01976767.

Design

The U-BIOPRED study in adults consisted of 3 visits in the severe asthma cohorts.\textsuperscript{28} At the first visit, participants were screened for eligibility to...
participate according to the inclusion and exclusion criteria. For the purpose of the present study, several measurements were performed. At the baseline visit and at the 12- to 18-month prospective follow-up visit, fraction of exhaled nitric oxide (FNOx) values were measured, followed by exhaled breath collection. Subsequently, prebronchodilator and postbronchodilator spirometry and sputum induction were performed. Finally, allergy tests, blood draws, and questionnaires were performed, as outlined previously.23

Measurements

Exhaled breath collection. Exhaled breath of patients with severe asthma was collected at 7 participating sites, as previously described.24,25,29 After refraining from eating, drinking, and smoking for at least 2 hours, patients breathed for 5 minutes at tidal volume through a 3-way nonrebreathing valve and an inspiratory carbon VOC filter (A2; North Safety, Middelburg, The Netherlands). Next, the subjects exhaled a single vital capacity volume into a 10-L Tedlar bag (SKC, Eighty Four, Pa). The VOCs in the Tedlar bag were trapped on thermal desorption tubes containing Tenax (80 mg) by drawing air through the Tenax tube with a peristaltic pump at a flow rate of 250 mL/min. Such storage of VOCs preserves the exhaled marker signal.30

Tubes were sent to the Academic Medical Centre Amsterdam for central analysis. After desorption of VOCs with a thermal desorption oven (TDS 3; Gerstel, Mülheim an der Ruhr, Germany), the stored VOCs were transferred into a Tedlar bag with nitrogen as a carrier gas. Subsequent analysis was carried out with the composite U-BIOPRED eNose platform. This eNose platform consists of an assembly of 4 eNoses, all from different developers, using distinct sensor technologies: (1) Cyanosense C320 using carbon black-polymer sensors (32 sensors); (2) Tor Vergata eNose using quartz crystal microbalances (QMBs) covered with metalloporphyrins (8 sensors); (3) Common Invent eNose using metal oxide semiconductor sensors (8 sensors),31 and (4) Owlsense Lonestar based on field asymmetric ion mobility spectrometry (110 data points).32 For all samples analyzed with the U-BIOPRED eNose platform, the structure of the output is similar: a 158-data-point counting profile based on responses of all 4 eNoses.

Lung function. Spirometry was performed before and 10 minutes after inhalation of 400 µg of salbutamol through a spacer according to European Respiratory Society (ERS) recommendations using daily calibrated equipment.33

Sputum induction. Sputum was induced by means of inhalation of hypertonic saline, according to the ERS recommendations.34 Selected samples were processed with 0.1% dithioerythritol, and differential cell counts were measured with a portable analyzer (NIOX Mino System; Aerocrine, Solna, Sweden) at a constant flow rate of 50 mL/s, according to American Thoracic Society/ERS recommendations.35

Blood. Blood eosinophil and neutrophil percentages were obtained from standard complete blood counts.

Allergic status. Allergy testing was performed with total and specific serum IgE measurements and skin prick test (SPTs) to a panel of common aeroallergens. Atopy was defined as the presence of sensitization on SPTs (wheal ≥3 mm) or serum specific IgE (≥0.35 kU/L).

FENO. Fraction of exhaled lower respiratory nitric oxide values were measured with a portable analyzer (NIOX Mino System; Aerocrine, Solna, Sweden) at a constant flow rate of 30 mL/s, according to American Thoracic Society/ERS recommendations.

Questionnaires. Asthma control was assessed by using the Juniper Asthma Control Questionnaire, which is a validated 5-item questionnaire, whereas the 20-item Sino-Nasal Outcome Test questionnaire was used as a measure for rhinosinusitis status.

Statistical analysis

Repeatability testing. For the purpose of repeatability testing, means ± SDs of the within-subject coefficient of variation among all 158 sensors was calculated based on available duplicate samples from the present study.

Data preprocessing. After ComBat selects the optimal number of variables35 and pairwise p values between baseline and follow-up visits (ie, changes over time) regarding clinical, physiologic, and inflammatory variables were evaluated by using the Mann–Whitney–Wilcoxon test for continuous data or the Pearson χ² test categorical data. Clinical variables were considered statistically significant at a P value of .05 or less. Furthermore, the distribution of patients from individual clinical sites among the revealed clusters was evaluated by using Pearson χ² test.

Between-cluster comparisons. Between-cluster comparisons and pairwise post hoc analyses of clinical, physiologic, and inflammatory variables at baseline were performed by using the Kruskal-Wallis test for continuous data and the Pearson χ² test for categorical data. Clinical variables were considered statistically significant at a P value of .05 or less. Furthermore, the distribution of patients from individual clinical sites among the revealed clusters was evaluated by using Pearson χ² test.

Cluster benchmarking. Clustering results of the baseline data set were benchmarked by using 2 more clustering techniques, partitioning around medoids (PAM) and topological data analysis (TDA), with the Ayasdi Workbench (version 7.5.0; Ayasdi, Menlo Park, Calif).10,16 Concordance between K-means and TDA was qualitatively assessed by combining results in a graphic display. Similarity of outcomes between Ward, K-means, and PAM clustering was quantified by means of Rand indexing, which results in a value between 0 (complete disagreement) and 1 (complete agreement).

Follow-up. Data workup to appraise the longitudinal behavior of the clusters consisted of 3 steps: (1) data preprocessing similar to baseline, (2) calculation of PCs based on loading factors of the baseline data set, and (3) cluster membership prediction for each patient by using the former K-means clustering results for modeling.

Baseline and longitudinal clustering outcomes were cross-tabulated. Absolute Δ values between baseline and follow-up visits (ie, changes over time) regarding clinical, physiologic, and inflammatory variables were evaluated by using the Mann–Whitney–Wilcoxon test for continuous data or the Pearson χ² test categorical variables to compare the characteristics between “cluster-stable” patients versus “cluster-migrating” patients.

Exploratory analyses. A series of exploratory analyses was performed through partial least-squares discriminant analysis (PLS-DA) and cross-validated partial least-squares regression (PLS regression): 1. analysis of correlation between the 158 eNose sensors using PLS-DA at the eNose sensor level versus obtained clusters; 2. examination of the relationship between significantly different inflammatory parameters and VOC patterns through PLS regression; and 3. testing of association between the U-BIOPRED sputum transcriptomic clusters (TACs; TAC1, TAC2, and TAC3) delineated by Kuo et al32 and VOC patterns reported here through PLS-DA.

Apart from TDA analysis, all analyses were performed in R studio (version 1.1.383) with R software (version 3.3.3) as the engine and supported by R packages: caret, clue, clustseg, ConsensusClusterPlus, data.table, fossil, Hmisc, MASS, stringr, sva, tableone, mixOmics, plsdepot and boot.

RESULTS

With a success rate of 100%, baseline breath samples of 80 patients with severe asthma recruited from 7 different sites across 5 countries in Europe were available. Two samples were lost because of wet-laboratory technicalities, resulting in complete data for 78 patients. Baseline characteristics are summarized in Table I.
TABLE I. Demographic data and baseline and longitudinal characteristics of study population

<table>
<thead>
<tr>
<th>Subjects (no.)</th>
<th>Baseline</th>
<th>Longitudinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), median (IQR)</td>
<td>55.5 (45.0-64.0)</td>
<td>57.0 (49.5-64.5)</td>
</tr>
<tr>
<td>Male sex (%)</td>
<td>41.0</td>
<td>43.1</td>
</tr>
<tr>
<td>Body mass index (kg/m²), median (IQR)</td>
<td>27.8 (24.4-32.6)</td>
<td>27.8 (25.1-32.3)</td>
</tr>
<tr>
<td>Age of onset (y), median (IQR)</td>
<td>27.5 (7.0-40.5)</td>
<td>30.0 (10.5-41.5)</td>
</tr>
<tr>
<td>Smoking status, current smoker/ ex-smoker/non-smoker (%)</td>
<td>12.8/33.3/53.8</td>
<td>15.7/35.3/49.0</td>
</tr>
<tr>
<td>Pack years, median (IQR)</td>
<td>14.5 (7.8-21.3)</td>
<td>13.9 (4.6-22.3)</td>
</tr>
<tr>
<td>OCS use (%)</td>
<td>42.3</td>
<td>36.7</td>
</tr>
<tr>
<td>Total daily OCS dose (median [IQR])</td>
<td>10.0 (10.0-12.5)</td>
<td>10.0 (10.0-12.5)</td>
</tr>
<tr>
<td>Exacerbations/year, median (IQR)</td>
<td>1.0 (0.0-2.8)</td>
<td>1.0 (0.0-2.0)</td>
</tr>
<tr>
<td>Atopy, positive (%)</td>
<td>67.9</td>
<td>62.7</td>
</tr>
<tr>
<td>ACQ-5 score, median (IQR)</td>
<td>2.5 (1.4-3.4)</td>
<td>1.6 (0.7-3.0)</td>
</tr>
<tr>
<td>PbFEV₁ (% predicted), median (IQR)</td>
<td>76.5 (60.9-88.5)</td>
<td>71.8 (60.6-86.2)</td>
</tr>
<tr>
<td>PhFEV₁/FVC actual ratio, median (IQR)</td>
<td>80.3 (69.7-89.8)</td>
<td>77.8 (70.6-87.5)</td>
</tr>
<tr>
<td>Sino-Nasal Outcome Test Questionnaire; Total daily OCS dose, daily OCS dose normalized to prednisolone among OCS users.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unbiased cluster analysis of exhaled breath profiles

From the original data set based on 158 eNose sensors (mean ± SD within-subject coefficient of variation, 6.28% ± 1.70%; n = 11), 34 principal components with an eigenvalue of greater than 1 were derived. By applying K-means clustering with a predefined number of clusters based on Ward clustering, 3 clusters of patients with severe asthma were delineated solely based on their exhaled breath profiles (Fig 1). Cluster 1 consists of 26 patients, cluster 2 consists of 33 patients, and cluster 3 consists of 19 patients. The first group includes 33% of the patients (n = 26) and is the middlemost cluster. With blood eosinophil percentages of 4.1% (interquartile range [IQR], 1.8% to 6.9%), blood neutrophil percentages of 57.8% (IQR, 54.0% to 67.1%), and an OCS user percentage of 39%, it represents the second highest/lowest levels at all 3 discriminating variables among the clusters.

Cluster 2. The second cluster includes 42% (n = 33) of the study population and can be described as neutrophilic inflammation predominant, with the highest blood neutrophil percentage of the 3 clusters being 65.6% (IQR, 60.0% to 76.2%). With an OCS use of 58%, this is the cluster with the greatest number of patients using OCS maintenance therapy.

Cluster 3. The final cluster comprises 24% of the patients (n = 19). This phenotype differentiates from the other 2 because only 4 (21%) of 19 patients are chronically using OCSs, and it shows the highest percentages of peripheral blood eosinophilia 4.7% (IQR, 2.64% to 8.00%) among the 3 groups.

Benchmarking

TDA demonstrated largely similar findings compared with K-means clustering. Data points in Fig 2 are color coded based on the 3 significant clusters revealed by K-means clustering to demonstrate coherence between both methods. Quantitative assessment of the similarity between K-means versus Ward and PAM clustering resulted in Rand index scores of 0.82 and 0.74, respectively.

Follow-up

Fifty-one of the 78 patients included in the baseline visit provided a breath sample at the 12- to 18-month follow-up visit (Table I). When comparing baseline clustering results with the longitudinal allocation, 21 (41%) patients were cluster stable, whereas 30 patients migrated to another cluster (Table III). There was a significant difference between these 2 groups in their change (absolute Δ) of sputum eosinophil percentages (Table IV).

Exploratory analyses

The Partial Least Square Discriminant Analysis (PLS-DA) correlation plot (Fig E2 in this article’s Online Repository at www.jacionline.org) indicates how the 158 sensors correlate with each other. Four of 8 sensors from the Tor Vergata eNose have a high correlation with a large number of sensors from the Cyanose C320.

Based on clustering outcomes (significantly different inflammatory parameters between the eNose clusters), PLS regression of blood neutrophils and eosinophils (as percentages) at baseline versus VOC patterns and absolute Δ values of sputum eosinophils (as a percentage) versus absolute Δ values of VOC patterns was performed, resulting in $R^2$ values of 0.63 (95% CI, 0.52-0.74), 0.41 (95% CI, 0.13-0.75), and 0.87 (95% CI, 0.76-0.99), respectively (see Fig E3-E5 in this article’s Online Repository at www.jacionline.org).

From a total of 28 patients, complete data concerning sputum transcriptomic profile and exhaled breath pattern were available. Based on VOC patterns, the 3 TAC clusters could be discriminated with an accuracy of 0.93 (see Fig E6 in this article’s Online Repository at www.jacionline.org).

DISCUSSION

This study shows that unbiased clustering of exhaled breath profiles captured by using eNose technology identifies 3 phenotypes of severe asthma. These clusters significantly differed with respect to systemic inflammatory markers and use of anti-inflammatory medication. These findings were benchmarked...
by using different clustering techniques. Follow-up at 12 to 18 months showed a significant difference between cluster-stable and cluster-migrating patients with regard to their longitudinal changes in sputum eosinophil values. Our results support the concept of using exhaled breath analysis with the eNose for quick and noninvasive inflammatory phenotyping of patients with severe asthma, which can be of clinical value with respect to personalized management or monitoring of these difficult-to-treat patients.

To our knowledge, this is the first study that establishes clusters of patients with severe asthma based on exhaled breath analysis by using eNose technology. The second novelty of this study is the assessment of stability of patients in those clusters during longitudinal follow-up. Previously, clustering techniques have been used to describe and monitor phenotypes of asthma based on clinical characteristics and for eNose technology focused on phenotyping broad cohorts of obstructive pulmonary diseases. However, the present study merged both approaches by using unbiased molecular profiling as a starting point for the follow-up of severe asthma phenotypes.

At a cross-sectional level, this concept is similar to the sputum gene expression profiling by Baines et al., Kuo et al., and Seys...
et al., as well as the serum cytokine profile clustering by Liang et al or imaging-based cluster analysis by Choi et al. These approaches cluster patients solely based on manifest molecular mechanisms and pathophysiology, thereby allowing the biology to drive the classification. In this way omics technologies are used to generate clusters based on both known and unknown pathophysiologic pathways rather than examining a priori–defined molecular mechanisms, such as T H2o rT H17 pathways. The present eNose clusters show that exhaled metabolites yield differential signals mainly related to inflammatory profiles, which extends similar observations based on profiling at RNA levels in sputum. Therefore our data suggest that breathomics can bridge the gap between clinical and laboratory assessments in the phenotyping of severe asthma.

Our study has a number of strengths. First, exhaled breath samples were analyzed on the U-BIOPRED eNose platform according to the most recent standards for quality control, an assembly of eNoses from 4 different brands, all driven by different sensor techniques. Using this composite system, we expect to have approximated an optimal data set from VOC mixtures based on available eNose technology.

Second, the participants were carefully characterized and recruited from 7 different sites across 5 European countries in this study. This international and multicenter character strengthens the general validity of the observed eNose phenotypes of severe asthma.

Finally, we used distinct clustering techniques to test the stability of clusters and added similarity profile analysis for determining the number of significant clusters with the assumption of no a priori groups. The latter strengthens the validity of our findings.

Our study also has limitations. First, an external validation in a new and independent cohort was lacking. This would have required us to repeat the U-BIOPRED study, which understandably is a future goal. The second best option would have been a split into a training and validation set, but this resulted in sample sizes too small for adequate cluster analysis, thereby promoting inevitable overfitting. However, we did validate our findings by benchmarking 3 specifically different clustering techniques: hierarchical methods versus partitioning versus clumping (continuous) variation combined with similarity testing by means of Rand indexing. The consistency of these results adds to the plausibility of our findings.

Second, eNose technology seems to be suitable for the noninvasive identification of severe asthma phenotypes but is principally unable to identify the individual VOCs driving the distinction between the subgroups. Specific analysis with GC-MS will then be necessary to unravel the identity of the combination of VOCs. Recent GC-MS studies have shown that asthma in particular features exhaled VOCs that are associated with lipid peroxidation and inflammatory parameters, the distribution of which is likely to drive the distinction among the 3 clusters.

Finally, given the nature of the project, there is an inevitable influence of medication and smoking in addition to endogenous biological mechanisms. Inhaled and systemic corticosteroids, as well as short-acting β2-agonists, are likely to affect metabolic fingerprints and thereby patients’ “breathprints.” Notably, the present analysis picked this up by showing that systemic steroid use was one of the distinguishing features. Therefore we believe that including real-life patients with varying levels of treatment is not a major limitation of the present study.

How can we interpret these results? It appears that the eNose platform predominantly captured features of inflammation and anti-inflammatory treatment, such as eosinophil and neutrophil percentages and oral steroid use. Similar to several previous studies on exhaled VOC analysis in patients with obstructive

### Table II. Characteristics of baseline clusters

<table>
<thead>
<tr>
<th>Subjects (no.)</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y), median IQR</td>
<td>59.0 (50.3-63.8)</td>
<td>55.0 (43.0-62.0)</td>
<td>56.0 (44.5-68.5)</td>
<td>.405</td>
</tr>
<tr>
<td>Male sex (%)</td>
<td>42</td>
<td>46</td>
<td>32</td>
<td>.640</td>
</tr>
<tr>
<td>Body mass index (kg/m²), median IQR</td>
<td>28.1 (25.2-32.3)</td>
<td>28.2 (24.9-33.4)</td>
<td>25.8 (23.6-33.0)</td>
<td>.618</td>
</tr>
<tr>
<td>Age of onset (y), median IQR</td>
<td>30.0 (15.0-44.0)</td>
<td>18.0 (6.0-39.0)</td>
<td>27.0 (17.5-34.5)</td>
<td>.702</td>
</tr>
<tr>
<td>Smoking status, current smoker/ex-smoker/nonsmoker (%)</td>
<td>11.5/46.2/42.3</td>
<td>9.1/33.3/57.6</td>
<td>21.1/15.8/63.2</td>
<td>.233</td>
</tr>
<tr>
<td>Total daily OCS dose, median (IQR)</td>
<td>10.0 (5.8-20.5)</td>
<td>13.8 (4.1-23.8)</td>
<td>20.5 (11.4-31.3)</td>
<td>.479</td>
</tr>
<tr>
<td>Atopy, positive (%)</td>
<td>38.5</td>
<td>57.6</td>
<td>21.1</td>
<td>.035</td>
</tr>
<tr>
<td>Total daily OCS dose, median (IQR)</td>
<td>3.0 (1.0-11.0)</td>
<td>4.0 (2.0-14.5)</td>
<td>10.0 (4.5-19.5)</td>
<td>.428</td>
</tr>
<tr>
<td>SNOT-20 score, median (IQR)</td>
<td>1.7 (1.1-2.0)</td>
<td>1.5 (1.1-2.2)</td>
<td>1.9 (1.2-2.5)</td>
<td>.393</td>
</tr>
<tr>
<td>OCS use (%)</td>
<td>57.7</td>
<td>72.7</td>
<td>73.7</td>
<td>.958</td>
</tr>
<tr>
<td>ACQ-5, median (IQR)</td>
<td>2.5 (1.4-3.4)</td>
<td>2.2 (1.4-3.4)</td>
<td>2.8 (1.1-3.6)</td>
<td>.531</td>
</tr>
<tr>
<td>PFBEV1 (%) predicted, median (IQR)</td>
<td>74.7 (52.0-84.2)</td>
<td>76.6 (62.5-92.1)</td>
<td>75.4 (66.9-87.7)</td>
<td>.146</td>
</tr>
<tr>
<td>PbFEV1/FVC actual ratio, median (IQR)</td>
<td>75.8 (66.9-85.9)</td>
<td>81.5 (70.1-90.4)</td>
<td>82.7 (77.7-94.9)</td>
<td>.840</td>
</tr>
<tr>
<td>SNOT20 score, median (IQR)</td>
<td>25.5 (14.0-53.0)</td>
<td>33.0 (16.9-46.5)</td>
<td>28.0 (14.5-49.5)</td>
<td>.550</td>
</tr>
<tr>
<td>Sputum eosinophils (%), median (IQR)</td>
<td>4.8 (1.6-15.9)</td>
<td>5.3 (2.1-16.5)</td>
<td>2.2 (1.0-29.5)</td>
<td>.915</td>
</tr>
<tr>
<td>Sputum neutrophils (%), median (IQR)</td>
<td>52.2 (39.0-65.4)</td>
<td>52.8 (45.0-65.8)</td>
<td>44.2 (9.4-85.8)</td>
<td>.733</td>
</tr>
<tr>
<td>Blood eosinophils (%), median (IQR)</td>
<td>4.1 (1.8-6.9)</td>
<td>2.3 (1.0-4.4)</td>
<td>4.7 (2.6-8.0)</td>
<td>.045</td>
</tr>
<tr>
<td>Blood neutrophils (%), median (IQR)</td>
<td>57.8 (54.0-67.1)</td>
<td>65.6 (60.0-76.2)</td>
<td>55.5 (52.3-64.0)</td>
<td>.017</td>
</tr>
</tbody>
</table>

Differences between clusters were tested by using Kruskal-Wallis testing for continuous data and Pearson χ² tests for categorical data.

Statistically significant values are in boldface.

**ACQ-5:** Juniper 5-item Asthma Control Questionnaire; **Atopy:** presence of sensitization on SPT (wheat ≥3 mm) or serum specific IgE (≥0.35 kU/L); **FENO:** fraction of exhaled nitric oxide; **FVC:** forced vital capacity; **OCS use:** regular daily use of oral corticosteroids; **PFBEV1:** postbronchodilator FEV1; **SNOT-20:** 20-item Sino-Nasal Outcome Test Questionnaire; **Total daily OCS dose:** daily OCS dose normalized to prednisolone among OCS users.

---

et al., as well as the serum cytokine profile clustering by Liang et al or imaging-based cluster analysis by Choi et al. These approaches cluster patients solely based on manifest molecular mechanisms and pathophysiology, thereby allowing the biology to drive the classification. In this way omics technologies are used to generate clusters based on both known and unknown pathophysiologic pathways rather than examining a priori–defined molecular mechanisms, such as T H2o rT H17 pathways. The present eNose clusters show that exhaled metabolites yield differential signals mainly related to inflammatory profiles, which extends similar observations based on profiling at RNA levels in sputum. Therefore our data suggest that breathomics can bridge the gap between clinical and laboratory assessments in the phenotyping of severe asthma.

Our study has a number of strengths. First, exhaled breath samples were analyzed on the U-BIOPRED eNose platform according to the most recent standards for quality control, an assembly of eNoses from 4 different brands, all driven by different sensor techniques. Using this composite system, we expect to have approximated an optimal data set from VOC mixtures based on available eNose technology.

Second, the participants were carefully characterized and recruited from 7 different sites across 5 European countries in this study. This international and multicenter character strengthens the general validity of the observed eNose phenotypes of severe asthma.

Finally, we used distinct clustering techniques to test the stability of clusters and added similarity profile analysis for determining the number of significant clusters with the assumption of no a priori groups. The latter strengthens the validity of our findings.

Our study also has limitations. First, an external validation in a new and independent cohort was lacking. This would have required us to repeat the U-BIOPRED study, which understandably is a future goal. The second best option would have been a split into a training and validation set, but this resulted in sample sizes too small for adequate cluster analysis, thereby promoting inevitable overfitting. However, we did validate our findings by benchmarking 3 specifically different clustering techniques: hierarchical methods versus partitioning versus clumping (continuous) variation combined with similarity testing by means of Rand indexing. The consistency of these results adds to the plausibility of our findings.

Second, eNose technology seems to be suitable for the noninvasive identification of severe asthma phenotypes but is principally unable to identify the individual VOCs driving the distinction between the subgroups. Specific analysis with GC-MS will then be necessary to unravel the identity of the combination of VOCs. Recent GC-MS studies have shown that asthma in particular features exhaled VOCs that are associated with lipid peroxidation and inflammatory parameters, the distribution of which is likely to drive the distinction among the 3 clusters.

Finally, given the nature of the project, there is an inevitable influence of medication and smoking in addition to endogenous biological mechanisms. Inhaled and systemic corticosteroids, as well as short-acting β2-agonists, are likely to affect metabolic fingerprints and thereby patients’ “breathprints.” Notably, the present analysis picked this up by showing that systemic steroid use was one of the distinguishing features. Therefore we believe that including real-life patients with varying levels of treatment is not a major limitation of the present study.

How can we interpret these results? It appears that the eNose platform predominantly captured features of inflammation and anti-inflammatory treatment, such as eosinophil and neutrophil percentages and oral steroid use. Similar to several previous studies on exhaled VOC analysis in patients with obstructive

---

**ACQ-5:** Juniper 5-item Asthma Control Questionnaire; **Atopy:** presence of sensitization on SPT (wheat ≥3 mm) or serum specific IgE (≥0.35 kU/L); **FENO:** fraction of exhaled nitric oxide; **FVC:** forced vital capacity; **OCS use:** regular daily use of oral corticosteroids; **PFBEV1:** postbronchodilator FEV1; **SNOT-20:** 20-item Sino-Nasal Outcome Test Questionnaire; **Total daily OCS dose:** daily OCS dose normalized to prednisolone among OCS users.
pulmonary diseases, our data confirm that exhaled VOCs appear to be associated with both systemic and local airway eosinophilic inflammation. Apparently, both types of inflammation express themselves through a different set of exhaled VOCs because there is little to no agreement in outcomes between the 2 inflammatory features among our reported clusters. This might be caused by the distinct molecular pathways underlying eosinophilic inflammation. Furthermore, differences in metabolic profiles between clusters 2 and 3 might particularly be linked to a combination of blood eosinophil percentages and OCS use (see Table II), whereas cluster 1 does not show specific distinctive characteristics, as measured based on the available clinical and inflammatory features. Nevertheless, the latter group differs with regard to its exhaled VOC-driven profile, which points toward a complementary value of molecular assessment on top of what we know from our patients by commonly used assessment. Considering the aims of the U-BIOPRED consortium, such a discovery of previously unknown molecular phenotypes is key to increased understanding of severe asthma. Additional explorative analysis concerning identification of sputum transcriptomic phenotypes (TAC1, TAC2, and TAC3) based on exhaled breath patterns resulted in an accuracy of 0.93. This identification of sputum transcriptics clusters is driven by different PCs than those most relevant for the primary clusters reported in this article. We believe this underlines the wealth of information available in exhaled breath.

What are the clinical implications of our data? When eNose technology is capable of identifying clinically revealed and unrevealed asthma phenotypes, it can serve a transitional role between omics technologies and clinical medicine.

**TABLE III.** Cross-table of clustering results among 51 patients with both a baseline and follow-up visit based on exhaled breath profiles: 21 patients are cluster stable (in boldface), and 30 patients migrate between clusters

<table>
<thead>
<tr>
<th>Baseline clusters</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal clusters</td>
<td>1</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>22</td>
<td>19</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*Baseline clusters* Cluster allocation of patients at baseline visits; *Longitudinal clusters*, cluster allocation of patients at the 12- to 18-month follow-up visit.
**TABLE IV.** Cluster-stable versus cluster-migrated patients: Δ (abs[Baseline–Longitudinal])

<table>
<thead>
<tr>
<th>Subjects (no.)</th>
<th>Cluster stable</th>
<th>Cluster migrated</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCS use, yes/no (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/no</td>
<td>10</td>
<td>17</td>
<td>.477</td>
</tr>
<tr>
<td>Yes/No</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Yes/yes</td>
<td>7</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Total daily OCS dose, median (IQR)</td>
<td>1.0 (1.0-1.8)</td>
<td>1.0 (0.0-1.0)</td>
<td>.283</td>
</tr>
<tr>
<td>Exacerbations/year, median (IQR)</td>
<td>1.0 (0.0-2.0)</td>
<td>1.0 (0.0-2.0)</td>
<td>.418</td>
</tr>
<tr>
<td>ACQ-5 score, median (IQR)</td>
<td>0.8 (0.4-1.2)</td>
<td>0.6 (0.3-1.0)</td>
<td>.430</td>
</tr>
<tr>
<td>PbPEFV1i (% predicted), median (IQR)</td>
<td>9.0 (4.3-11.9)</td>
<td>6.0 (2.3-12.8)</td>
<td>.752</td>
</tr>
<tr>
<td>PbPEFV1/FVC actual ratio, median (IQR)</td>
<td>5.4 (2.0-8.7)</td>
<td>5.4 (2.3-9.6)</td>
<td>.863</td>
</tr>
<tr>
<td>FENO (ppb), median (IQR)</td>
<td>10.6 (8.8-18.5)</td>
<td>7.8 (4.0-14.4)</td>
<td>.181</td>
</tr>
<tr>
<td>SNOT-20 score, median (IQR)</td>
<td>0.49 (0.19-0.81)</td>
<td>0.30 (0.13-0.50)</td>
<td>.204</td>
</tr>
<tr>
<td>Sputum eosinophils (%), median (IQR)</td>
<td>2.2 (1.0-3.8)</td>
<td>7.2 (3.3-12.3)</td>
<td>.046</td>
</tr>
<tr>
<td>Sputum neutrophils (%), median (IQR)</td>
<td>16.5 (12.4-26.1)</td>
<td>12.3 (4.0-20.0)</td>
<td>.166</td>
</tr>
<tr>
<td>Blood eosinophils (%), median (IQR)</td>
<td>1.5 (0.8-4.5)</td>
<td>1.2 (0.7-2.5)</td>
<td>.803</td>
</tr>
<tr>
<td>Blood neutrophils (%), median (IQR)</td>
<td>4.4 (1.9-11.1)</td>
<td>7.9 (2.8-13.7)</td>
<td>.216</td>
</tr>
</tbody>
</table>

ACQ-5, Juniper 5-item Asthma Control Questionnaire; Feno, fraction of exhaled nitric oxide; FVC, forced vital capacity; OCS use, regular daily use of oral corticosteroids at baseline/follow-up; PbPEFV1i, postbronchodilator FEV1i; SNOT-20, 20-item Sino-Nasal Outcome Test Questionnaire; Total daily OCS dose, daily OCS dose normalized to prednisolone among OCS users.

Statistically significant values are in boldface.

Transcriptomics, proteomics, and metabolomics in tissue, sputum, or blood have demonstrated their value through identification of distinct phenotypes in asthmatic patients. Because pathophysiology mechanisms have been linked to these phenotypes, new specifically targeted treatments are being developed. This leads to an increasing need for point-of-care biomarkers to predict and monitor the responsiveness of patients to particular interventions. Although most omics technologies are difficult to implement in daily medical care, gas sensor–based exhaled breath analysis is developing toward clinical implication, thereby having the potential to fulfill this increasing clinical need. In addition, even though biomarkers for capturing the eosinophilic phenotype and predicting responsiveness to common therapies are available, these are either not widely used because of laborious procedures (sputum eosinophilia) or exhibit insufficient accuracy (exhaled nitric oxide). We believe that the present study adds to strengthening the clinical utility of eNose measurements by showing that migration of patients between the clusters is associated with changes in sputum eosinophilia, whereas we observed relatively high correlations between the exhaled breath profiles and eosinophilic and neutrophilic inflammation. This fits in with the predictive capacity of eNose assessment for steroid responsiveness and its potential to discriminate between clinically stable and unstable episodes of asthma. It will require longer-term follow-up studies to examine the clinical course of the cluster-stable and cluster-unstable patients, as identified by using the eNose.

In conclusion, this study reveals unbiased clusters of patients with severe asthma based on exhaled breath profiles captured by the U-BIOPRED eNose platform. Three significant clusters with differences regarding eosinophilic/neutrophilic inflammation and systemic steroid use were delineated. Notably, after 12 to 18 months of follow-up, cluster-stable and cluster-migrating patients differed with regard to their longitudinal changes in sputum eosinophil values. These results warrant prospective studies on the potential of exhaled breath fingerprinting by using eNose technology as point-of-care procedure for (therapeutic) management of patients with severe asthma.

**Key messages**

- Unbiased clustering of exhaled breath profiles captured by using eNose technology revealed 3 phenotypes of severe asthma, which significantly differ with respect to systemic inflammatory markers and use of anti-inflammatory medication.
- Follow-up at 12 to 18 months showed a significant difference between cluster-stable and cluster-migrating patients with regard to longitudinal changes in sputum eosinophilic.

**REFERENCES**


FIG E1. Left panel, PLS-DA results based on 78 exhaled breath profiles of patients with severe asthma by using the U-BIOPRED eNose platform and the 3 revealed clusters. Blue circles, Cluster 1; orange triangles, cluster 2; gray plus signs, cluster 3. x-axis, PLS-DA component I; y-axis, PLS-DA component II. Right panel, Correlation plot based on components I (x-axis) and II (y-axis) of the U-BIOPRED eNose platforms outcomes by using PLS-DA. Strongly associated (or correlated) variables are projected in the same direction from the origin. The greater distance from the origin, the stronger the association. Green, Cyranose 320; black, Tor Vergata eNose; red, Comon Invent eNose; blue, Owlstone Lonestar.
FIG E2. Box plots of all variables listed in Table II of the main document. Between-cluster comparison of clinical, physiologic, and inflammatory variables at baseline was performed by using the Kruskal-Wallis test for continuous data and the Pearson χ² test for categorical data.
FIG E3. PLS-DA results based on 28 exhaled breath profiles of patients with severe asthma by using the U-BIOPRED eNose platform at the PC level and their sputum transcriptomic phenotypes by Kuo et al. Blue circles, TAC1; orange triangles, TAC2; gray plus signs, TAC3. x-axis, PLS-DA component I; y-axis, PLS-DA component II. Right panel, Correlation plot based on components I (x-axis) and II (y-axis) of the U-BIOPRED eNose platforms outcomes by using PLS-DA. Strongly associated (or correlated) PCs are projected in the same direction from the origin. The greater the distance from the origin, the stronger the association. Note: colorization of PC labels 1 to 34 (red, green, blue, and black) serves no other goal than improvement of readability.
FIG E4. Associations between exhaled markers by using the U-BIOPRED eNose platform versus blood neutrophil percentages with PLS regression analysis: $R^2 = 0.63$ (95% CI, 0.52-0.74). Gray area, 95% CI; dashed lines, 95% prediction interval.
FIG E5. Associations between exhaled markers by using the U-BIOPRED eNose platform versus blood eosinophil percentages by using PLS regression analysis: $R^2 = 0.41$ (95% CI, 0.13-0.75). Gray area, 95% CI; dashed lines, 95% prediction interval.
FIG E6. Associations between absolute Δ values of exhaled breath profiles by using the U-BIOPRED eNose platform versus absolute Δ values of sputum eosinophil percentages by using PLS regression analysis: $R^2 = 0.87$ (95% CI, 0.76-0.99). Gray area, 95% CI; dashed lines, 95% prediction interval.
TABLE E1. Distribution of patients from the same centers among the 3 clusters

<table>
<thead>
<tr>
<th></th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>Site 5</th>
<th>Site 6</th>
<th>Site 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

A χ² analysis resulted in a P value of .079.