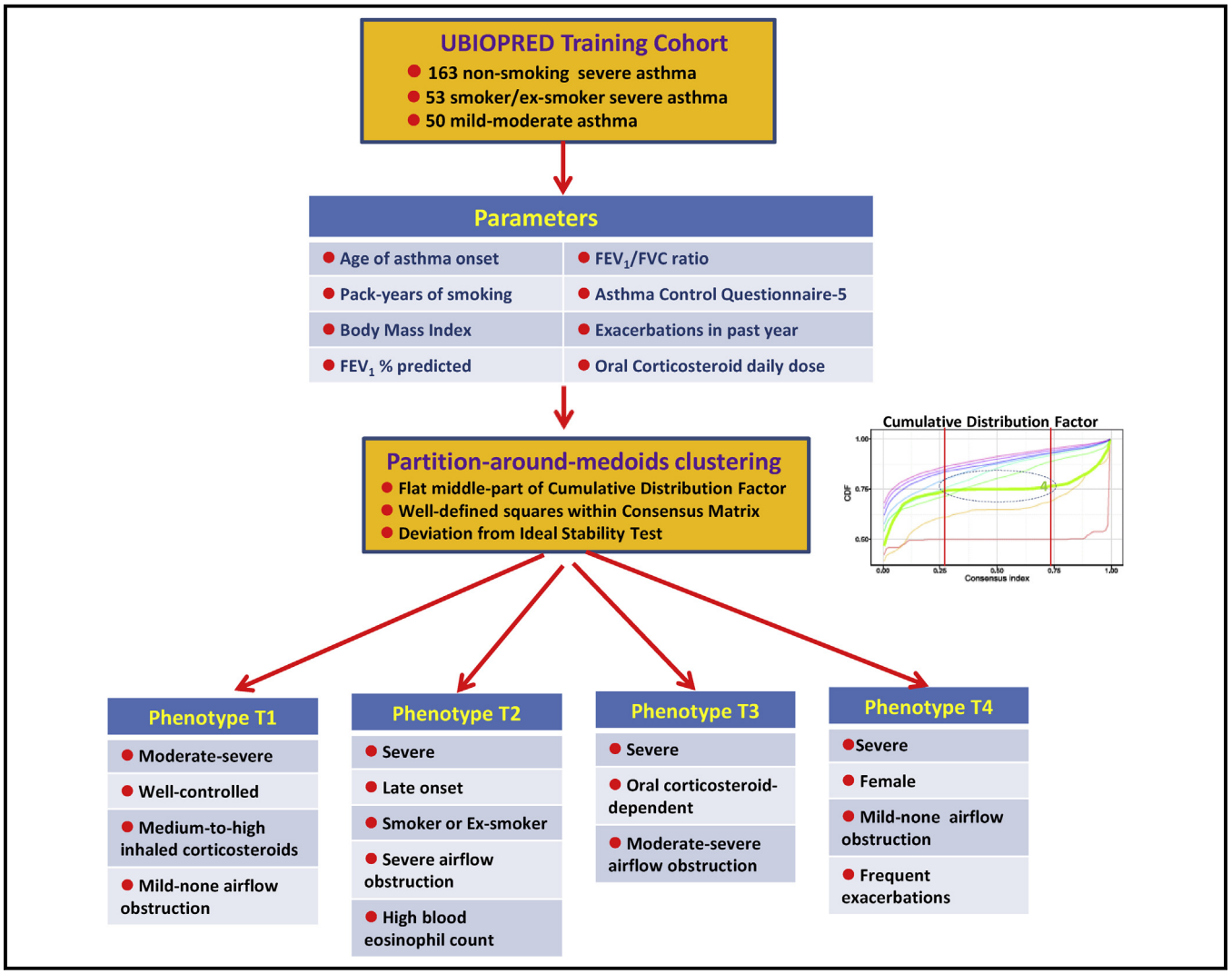
U-BIOPRED clinical adult asthma clusters linked to a subset of sputum omics

Diane Lefaudeux, MSc,[a]* Bertrand De Meulder, PhD,[a]* Matthew J. Loza, PhD,[b] Nancy Peffer, BS,[b] Anthony Rowe, PhD,[c] Frédéric Baribaud, PhD,[b] Aruna T. Bansal, PhD,[d] Rene Lutter, PhD,[e] Ana R. Sousa, PhD,[f] Julie Corfield, MSc,[g] Ioannis Pandis, PhD,[h] Per S. Bakke, MD,[i] Massimo Caruso, MD,[j] Pascal Chanez, MD,[k] Sven-Erik Dahlén, MD,[l] Louise J. Fleming, MD,[m] Stephen J. Fowler, MD,[n] Ildiko Horvath, MD,[o] Norbert Krug, MD,[p] Paolo Montuschi, MD,[q] Marek Sanak, MD,[r] Thomas Sandstrom, MD,[s] Dominic E. Shaw, MD,[t] Florian Singer, MD,[u] Peter J. Sterk, MD, PhD,[v] Graham Roberts, MD,[w] Ian M. Adcock, PhD,[m] Ratko Djukanovic, MD,[w] Charles Auffray, PhD,[a] Kian Fan Chung, MD,[m] and the U-BIOPRED Study Group‡

Lyon and Marseille, France; Spring House, Pa; High Wycombe, Cambridge, Stockley Park, Nottingham, London, Manchester, and Southampton, United Kingdom; Amsterdam, The Netherlands; Bergen, Norway; Catania and Rome, Italy; Molndal, Stockholm, and Umeå, Sweden; Budapest, Hungary; Hannover, Germany; Krakow, Poland; and Bern, Switzerland

GRAPHICAL ABSTRACT



Background: Asthma is a heterogeneous disease in which there is a differential response to asthma treatments. This heterogeneity needs to be evaluated so that a personalized management approach can be provided.
Objectives: We stratified patients with moderate-to-severe asthma based on clinicophysiologic parameters and performed an omics analysis of sputum.
Methods: Partition-around-medoids clustering was applied to a training set of 266 asthmatic participants from the European Unbiased Biomarkers for the Prediction of Respiratory Diseases Outcomes (U-BIOPRED) adult cohort using 8 prespecified clinic-physiologic variables. This was repeated in a separate validation set of 152 asthmatic patients. The clusters were compared based on sputum proteomics and transcriptomics data.
Results: Four reproducible and stable clusters of asthmatic patients were identified. The training set cluster T1 consists of patients with well-controlled moderate-to-severe asthma, whereas cluster T2 is a group of patients with late-onset severe asthma with a history of smoking and chronic airflow obstruction. Cluster T3 is similar to cluster T2 in terms of chronic airflow obstruction but is composed of nonsmokers. Cluster T4 is predominantly composed of obese female patients with uncontrolled severe asthma with increased exacerbations but with normal lung function. The validation set exhibited similar clusters, demonstrating reproducibility of the classification. There were significant differences in sputum proteomics and transcriptomics between the clusters. The severe asthma clusters (T2, T3, and T4) had higher sputum eosinophilia than cluster T1, with no differences in sputum neutrophil counts and exhaled nitric oxide and serum IgE levels.
Conclusion: Clustering based on clinicophysiologic parameters yielded 4 stable and reproducible clusters that associate with different pathobiological pathways. (J Allergy Clin Immunol 2017;139:1797-807.)

Key words: *Severe asthma, clustering, sputum eosinophilia, partition-around-medoids algorithm*

Abbreviations used
BMI: Body mass index
CDF: Cumulative distribution function
COPD: Chronic obstructive pulmonary disease
OCS: Oral corticosteroids
SARP: Severe Asthma Research Program
U-BIOPRED: Unbiased Biomarkers for the Prediction of Respiratory Diseases Outcomes

Although clinicians have been focusing on the definition and classification of asthma severity and disease risk for the past decade, there is now a consensus that a deeper understanding of the basis of the heterogeneity of asthma is necessary to find targeted treatments for specific asthma phenotypes.[1] This is imperative for patients with severe asthma because this group of patients does not fully respond to currently available asthma medications[1] and is likely to constitute a number of different asthma phenotypes.[2] Therefore there is a need to improve the identification and definition of these asthma phenotypes.

Cluster analysis with unsupervised statistical approaches has already led to the definition of clusters on the basis of similarities in clinical and inflammatory biomarkers.[3,4] However, these studies have used relatively homogeneous populations and therefore might not reflect the real-life situation. One example is the exclusion of current or previous smokers with asthma, a group that might have asthma–chronic obstructive pulmonary disease (COPD) overlap syndrome.[5] In addition, previously derived clusters have not been linked to underlying biological profiles apart from the use of blood or sputum eosinophil counts. Other approaches have been to use unsupervised gene and protein omics data to cluster patients with asthma.[6-8]

From [a]the European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon; [b]Janssen Research and Development LLC, Spring House; [c]Janssen Research and Development Ltd, High Wycombe; [d]Acclarogen, St John's Innovation Centre, Cambridge; [e]the Academic Medical Centre, University of Amsterdam; [f]the Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park; [g]AstraZeneca R&D Molndal, and Areteva R&D, Nottingham; [h]Data Science Institute, Imperial College London; [i]the Department of Clinical Science, University of Bergen; [j]the Department of Clinical and Experimental Medicine, University of Catania; [k]Département des Maladies Respiratoires, Aix Marseille Université Marseille; [l]the Centre for Allergy Research, Karolinska Institutet, Stockholm; [m]National Heart and Lung Institute, Imperial College & Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London; [n]the Centre for Respiratory Medicine and Allergy, University of Manchester; [o]the Department of Pulmonology, Semmelweis University, Budapest; [p]Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover; [q]the Faculty of Medicine, Catholic University of the Sacred Heart, Rome; [r]the Department of Medicine, Jagiellonian University Medical School, Krakow; [s]the Department of Public Health and Clinical Medicine, Medicine, Umeå university; [t]the Respiratory Research Unit, University of Nottingham; [u]University Children's Hospital Bern; [v]NIHR Respiratory Biomedical Research Unit, Clinical and Experimental Sciences, Southampton; [w]the Faculty of Medicine, University of Southampton.
*These authors contributed equally to this work.
‡Other Consortium Study Group members are shown in the acknowledgements section.
U-BIOPRED is supported through an Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115010, resources of which are composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution (www.imi.europa.eu). We would also like to acknowledge help from the IMI-funded eTRIKS project (EU Grant Code no. 115446). This study is registered on ClinicalTrials.gov (identifier: NCT01982162).
Disclosure of potential conflict of interest: D. Lefaudeux receives grant support from the European Union's Innovative Medicines Initiative (EU IMI). B. De Meulder receives grant support from the EU IMI. N. Peffer is an employee and holds stock options for Janssen. A. Rowe is an employee and holds stock options with Janssen Research. F. Baribaud is an employee and holds stock options for Janssen. R. Lutter receives grant support from the EU IMI. A. R. Sousa is an employee of GlaxoSmithKline. I. Pandis receives grant support from the EU IMI. M. Sanak receives grant support from the EU IMI. L. J. Fleming receives grant support from the EU IMI, serves as a consultant for Vectura, and receives payment for lectures form Novartis. S. J. Fowler receives grant support from the EU IMI. D. E. Shaw serves as a consultant for TEVA and GlaxoSmithKline; receives research support from GlaxoSmithKline; and receives speaker fees from Boehringer Ingelheim and Novartis. P. J. Sterk receives grant support from the EU IMI. G. Roberts receives grant support from the EU IMI. R. Djukanovic receives grant support from the EU IMI and Novartis; serves on the board for TEVA and Novartis; serves as a consultant for Synairgen; receives payment for lectures from TEVA and Novartis; receives payment for educational presentations from TEVA; and receives stock options with Synairgen. C. Auffray receives grant support from the EU IMI. K. F. Chung receives grant support from Pfizer, GlaxoSmithKline, MRC, EU IMI, and the NIH; serves as a board member for GlaxoSmithKline, AstraZeneca, Novartis, TEVA, Boehringer Ingelheim, and J&J; and receives payment for lectures from AstraZeneca, Novartis, and Merck. The rest of the authors declare that they have no relevant conflicts of interest.
Received for publication December 20, 2015; revised July 23, 2016; accepted for publication August 8, 2016.
Available online October 20, 2016.
Corresponding author: Kian F. Chung, MD, National Heart and Lung Institute, Imperial College London, Dovehouse St, London SW3 6LY, United Kingdom. E-mail: f.chung@imperial.ac.uk.
The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections
0091-6749/$36.00

http://dx.doi.org/10.1016/j.jaci.2016.08.048

In this study we used a robust clustering approach with clinical and physiologic parameters that are available to the asthma physician in a broad range of participants with mild/moderate to severe asthma, including smokers and ex-smokers recruited in the Unbiased Biomarkers for the Prediction of Respiratory Diseases Outcomes (U-BIOPRED) project.[9] The second step was to explore the underlying pathobiological pathways of these clusters by examining the differential expression of the transcriptome of sputum cells and the proteome of sputum supernatants that exist between the clusters generated to determine whether they exhibit any differences in specific pathobiological pathways.

METHODS

U-BIOPRED cohorts

We used a subset of the U-BIOPRED adult baseline data. The U-BIOPRED cohort comprises 509 patients with asthma, both mild-to-moderate and severe, and includes nonsmokers, ex-smokers, current smokers, and 101 nonasthmatic control subjects. These subjects had undergone detailed phenotypic characterization by using established standard operating procedures, as described previously.[9] The study participants were split randomly into training and validation data sets in a 2:1 ratio, with the 2 groups being balanced in terms of asthma severity, age, and sex. The validation group was used for internal replication. All participants provided written informed consent to participate in the study, which was approved by national ethics committees.

Clinical variables

The cluster analysis was focused on key variables that are readily accessible to the general practitioner representing important historical, clinical, and physiologic parameters underlying each participant with asthma. These variables were age of onset of asthma symptoms, pack years of cigarette smoking, body mass index (BMI), FEV_1 as a percentage of predicted value (FEV_1 percent predicted), FEV_1/forced vital capacity ratio, the average score of the 5 first questions of the Asthma Control Questionnaire, self-reported numbers of exacerbations in the previous year, and daily dose of oral prednisolone or equivalent.

Data preprocessing and cluster analysis

Box-Cox power transformation[10] was used to approximate the data to a normal distribution by using the powerTransform function from the R package *car*,[11] which uses maximum likelihood to determine the best λ value. Data were then center-scaled to ensure similar ranges for all the parameters and reduced by using principal component analysis, ensuring that there was no correlation between the composite variables, thereby avoiding skewing of the analysis.

Clustering schemes are descriptive methods that group participants with similar characteristics. The Euclidean distance (which actually measures dissimilarity by using the ordinary straightline distance between 2 points) was used to determine similarity between participants. Clustering was performed by using the partition-around-medoids algorithm, a more robust generalization of the k-means method.[12] Bootstrapping (also known as consensus clustering) was performed by randomly removing 10% of the data and repeating the clustering a total of 1000 times to assess cluster stability.[13] Cluster stability was assessed by studying the cumulative distribution function (CDF), which, as represented in Fig 1, *A*, describes the proportion of pairs of participants (on the y-axis) clustered together in at most x percentage of bootstrap iterations (on the x-axis). Thus, if the curve is flat, such as between x values of 0.2 and 0.8, this means that there are no pairs of participants who are clustered together between 20% and 80% of the iterations (ie, they are almost never clustered together [<20%] or almost always clustered together [>80%]). Clustering results were considered stable when the middle part of the CDF was flat.

To further define the stability of the clusters, we used an in-house objective called "deviation from ideal stability" (see Fig E1 in this article's Online Repository at www.jacionline.org), which is at its best when it is close to zero. Additionally, to gain confidence in the existence of these clusters,[14] internal validity was checked by using the Calinski and Harabasz index,[15] which measures the ratio of between-cluster variance to within-cluster variance such that clusters are better defined at higher values.

Sputum induction, transcriptomics, and protein analytes

Sputum induction was performed after inhalation of hypertonic (0.9% to 4.5%) saline with a DeVilbiss 2000 Ultrasonic nebulizer (DeVilbiss, Somerset, Pa), according to a standardized protocol.[16] Sputum plugs were selected and liquefied with dithioerythritol. Differential cell counts were determined by means of assessment of a maximum of 500 to 1000 inflammatory cells on Diff-Quick–stained cytospin preparations. Cytospin assessments were performed centrally with the outcome of the cytospin analysis, determining the suitability of the sample for analysis by accepting only those samples with a cell viability of 50% or greater and squamous cell counts of 40% or less.

Transcriptomic analysis was performed with the Affymetrix HT HG-U133 + PM GeneChip (Affymetrix, Santa Clara, Calif) on extracted RNA from sputum cells derived from cell pellets with a specific cutoff of 30% or fewer squamous cells. Technical and biological quality checks were performed according to Affymetrix recommendations with only RNA samples of high purity (RNA integrity number > 6.5) used for amplification; raw data were preprocessed by using the robust microarray analysis method from the *affy* R package[17] to derive the expression matrix.

From each of the frozen aliquots of sputum supernatant, 1129 analytes were quantified by using the SomaScan v3 platform (SomaLogic, Boulder, Colo; www.somalogic.com) with SOMAmer (Slow Off-rate Modified Aptamer) protein-binding reagents. These assays combine the best properties of antibodies and traditional aptamers, which are highly specific for the corresponding cognate proteins.[18] Analyte levels were reported as relative fluorescence units, cross-plate calibrated, and median normalized.

Statistical analysis

All analyses were undertaken with R software for statistical computing (version 3.1.2). Clinical variables were compared between clusters by using ANOVA for multiple-group comparison of normally distributed variables. The Kruskal-Wallis test was used for multiple-group comparison of ordered categorical or nonnormally distributed variables, and the χ^2 test was used for qualitative variables. ANOVA was performed on the data (transformed with base 2 logarithm), adjusting for age and sex, followed by a Tukey *post hoc* pairwise comparison test to compare protein abundance or transcript expression. Protein analytes or probe sets were defined to be consistently differentially abundant or expressed when their respective *P* values were less than .05 in both the training and validation sets analyzed separately and when these sets were analyzed together (preventing the inclusion of features that would have a different direction of change in the training and the validation sets). This allows for a reproducible and relatively stringent feature-selection process and decreases the false-positive rate despite not correcting the *P* value for multiple testing. Both proteomics and transcriptomics data sets have been checked for any batch or site effect and corrected accordingly by using the ComBat method.[19]

Pathway enrichment analysis was performed by using the results of the statistical analysis described above. The lists of contrast-specific features consistently found in both the training and validation sets were submitted to the *g:Profiler* Web tool for enrichment analysis.[20] The *P* values for enrichment analysis were corrected for false discovery rate by using the Benjamini-Hochberg method.[21] Feature lists for each comparison were tested for enrichment against the KEGG[22] and Reactome[23] databases.

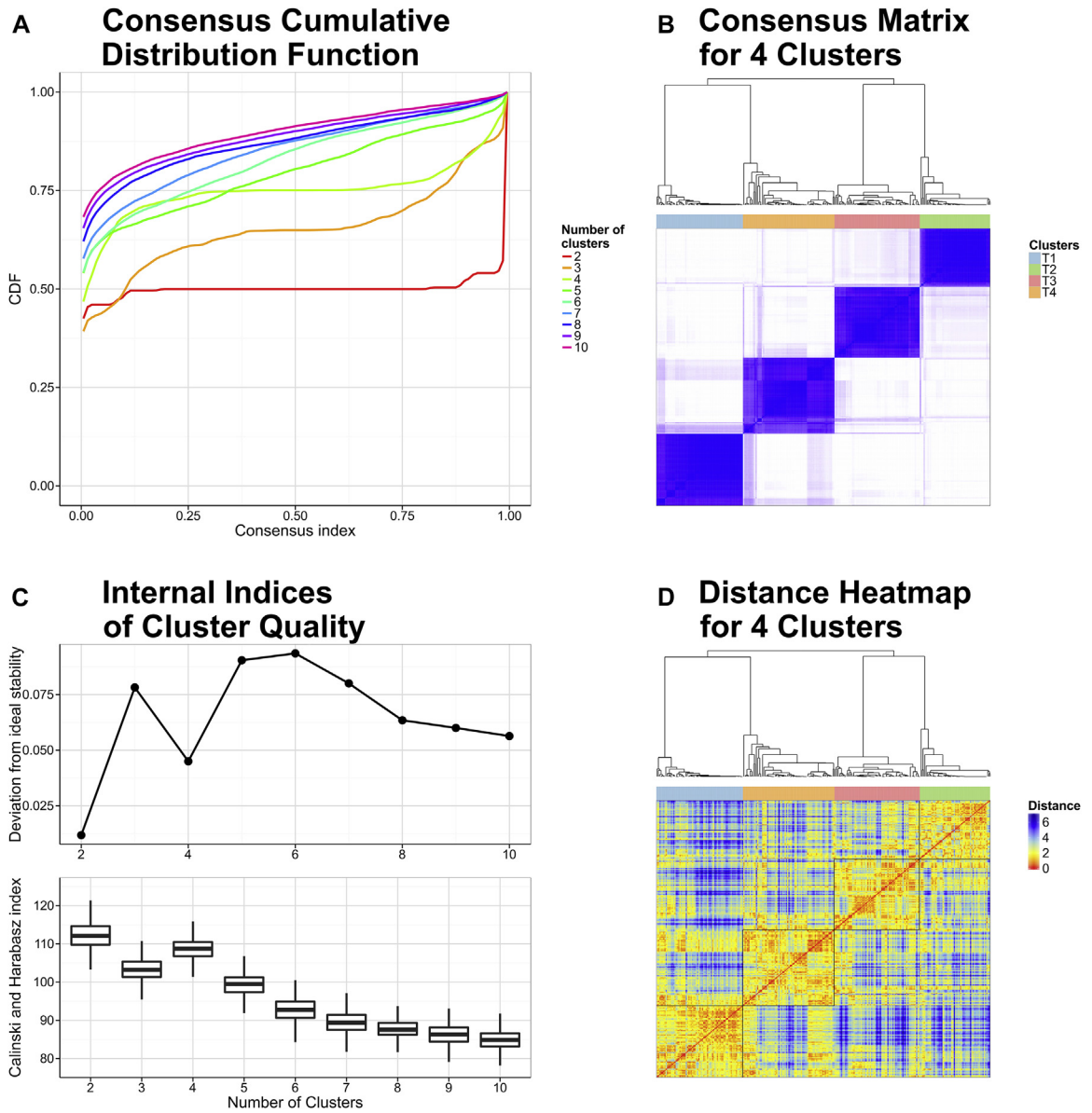


FIG 1. Clustering using the partition-around-medoids *(PAM)* algorithm on the training set. **A,** Consensus CDF of the consensus matrix shown in **B**. **C,** Two internal quality indexes: deviation from ideal stability and the Calinski and Harabasz index. **D,** Heat map of pairwise distances between participants.

RESULTS

Participants

A total of 418 of 509 asthmatic patients with a complete set of data for the 8 variables were available for analysis and were split into training (n = 266) and validation (n = 152) sets. The distribution of asthma severity, age, and sex and all 8 variables included in the clustering for the training and validation sets was not statistically different between the 2 sets, although FEV_1 (percent predicted) and daily dose of oral corticosteroids (OCS) were incompletely balanced ($P = .07$ and .06, respectively; see Table E1 in this article's Online Repository at www.jacionline.org).

Training set clusters

Consensus clustering on the training set was run to assess stability for a number of potential cluster numbers varying from 2 to 10. This resulted in the separation of 2 or 4 stable groups after resampling, as defined by a flat middle part of the consensus CDF (Fig 1, *A*),[13] well-defined squares within the consensus matrix (Fig 1, *B*), and minimal values for deviation from the ideal stability index (Fig 1, *C*, top). These cluster numbers were also associated with the 2 highest Calinski and Harabasz indices (Fig 1, *C*, bottom), indicating that the clusters were more compact than the overall data. Although separating into 2 and 4 clusters resulted in almost similar quality, 4 clusters were chosen for

further analysis (denoted T1 to T4) to allow for a more precise subphenotype definition. Indeed, the 2-cluster allocation mainly regroups T1 with T4 and T2 with T3. Finally, Fig 1, *D*, represents a heat map of distances between the participants in the 4 clusters.

Four-cluster analysis (T1-T4)

The 4 clusters are described in Table I and Table E2 in this article's Online Repository at www.jacionline.org. Briefly, cluster T1 is composed of patients with moderate-to-severe well-controlled asthma with normal FEV_1, low sputum eosinophilia, almost no OCS use (6%), and a high proportion of atopic participants (84.1%). Cluster T2 is mainly composed of overweight to obese (79% with BMI ≥25 kg/m^2 and 41% with BMI ≥30 kg/m^2) patients with late-onset severe asthma who smoked and who had relatively poor control, severe airflow obstruction (mean FEV_1, 58.9% of predicted value), and the highest sputum and blood eosinophilia, with a lower proportion of atopic participants than in the other 3 clusters (55.6%). Cluster T3 is similar to cluster T2, except that the asthmatic patients were nonsmokers, were less overweight, had poorer lung function, and had a higher proportion of atopic participants (70.6%). Cluster T4 is mostly composed of obese female asthmatic patients (83% female, 88% with BMI ≥25 kg/m^2, and 56% with BMI ≥30 kg/m^2) experiencing frequent exacerbations with poor asthma quality of life despite near-normal lung function and 73.6% of positive atopy status. Fraction of exhaled nitric oxide and serum IgE levels were not differentially distributed among the 4 clusters.

Validation set clusters

The same analysis was done on the validation set. It yielded 5 relatively stable clusters after resampling (denoted V1, V2, V3, V4a, and V4b to align with the training set), as shown by a flat CDF and a low deviation from ideal stability (see Fig E2 and Table E3 in this article's Online Repository at www.jacionline.org). The Calinsky and Harasbaz index was slightly better for 4 clusters. The difference in the number of clusters compared with the training set might be due to the fact that the validation set was smaller. When comparing the training and validation clusters using the least statistical differences of clinical variables, cluster V1 was found to be similar to cluster T1, cluster V2 was found to be similar to cluster T2, and cluster V3 was found to be similar to cluster T3, whereas cluster V4a combined with cluster V4b was found to be similar to cluster T4 (see Table E4 in this article's Online Repository at www.jacionline.org). For ease of recall, clusters T1 and V1 will be referred to as phenotype 1, clusters T2 and V2 will be referred to as phenotype 2, clusters T3 and V3 will be referred to as phenotype 3, and clusters T4, V4a, and V4b will be referred to as phenotype 4.

The distributions of the main clinical characteristics of the training and validation clusters were similar (Fig 2), with the exception of the V4a and V4b clusters covering 2 different parts of T4. Cluster V4a consists of less obese asthmatic patients associated with later onset of disease, lower OCS use, and better asthma control when compared with cluster V4b.

Algorithm to predict clinical phenotype

The support vector machine algorithm with a Gaussian radial basis kernel[24] was used to predict phenotypes from the 8 clinical parameters. The model was trained on the training set only by using a 10-fold cross-validation method to prevent overfitting with *caret*[25] and *kernlab*[24] R packages. The prediction model yielded an almost perfect accuracy of 97% on the training set. It predicted phenotype assignment on the validation set and achieved a very good accuracy rate of 86%. An xlsm file has been developed that can be used to predict the clinical phenotype (see this article's Prediction algorithm Excel file in this article's Online Repository at www.jacionline.org).

Biological characterization in a subset

The results of proteomics and transcriptomics profiling in sputum samples were compared between the phenotypes to determine whether they could represent a useful categorization of asthma. Because not all patients were able to produce any sputum or good-quality sputum for analysis and because of technical quality control, the number of participants used in the proteomics and transcriptomics analyses were 86 and 94, respectively. The clinical profiles of these participants who provided these samples was not different from those of the whole cohort, as shown in Table E5 in this article's Online Repository at www.jacionline.org. Protein data were available for 86 participants (56 in the training set and 30 in the validation set). Ten of the 1129 proteins measured were identified as being consistently differentially abundant between phenotypes (Table II). This number of hits was too small to allow for any meaningful pathway enrichment analysis. Sputum transcript expression data were available for 94 participants (56 in the training set and 38 in the validation set). A total of 345 transcripts (291 annotated) were found to be consistently significantly differentially expressed in at least 1 of the pairwise comparisons between the phenotypes (see Table E6 in this article's Online Repository at www.jacionline.org). Pathway enrichment results are shown in Table III.[20]

Differential protein abundance in sputum supernatants

Both the comparison of phenotype 2 ([ex-]smokers with severe asthma) and phenotype 3 (nonsmokers with severe asthma) with phenotype 1 (well-controlled asthma) highlighted levels of IL-16, a natural ligand of CD4 and CD9 that induces preferential migration of human regulatory T cells,[26] as being increased in the more severe phenotypes. Additionally, compared with phenotype 2, there was greater abundance in phenotype 1 of (1) connective tissue-activating peptide III (CTAP-III; or chemokine [C-X-C] ligand 7), a potent chemoattractant and activator of neutrophils; (2) granulocyte-macrophage colony stimulating factor (GM-CSF), which controls the production, differentiation, and function of granulocytes and macrophages; and (3) trypsin 2, which degrades the extracellular matrix. On the contrary, hyaluronan and proteoglycan link protein 1 (HAPLN1), which is involved in cell adhesion, was less abundant.

Phenotype 3 was associated with reduced levels of cathepsin G involved in connective tissue remodeling at the site of inflammation compared with phenotypes 1 and 4. Moreover, phenotype 3 also exhibited increased levels of arylsulfatase B precursor (ARSB), an arylsulfatase involved in cell adhesion and migration regulation, and proteasome subunit α2 (PSA2), a member of the peptidase T1A family, when compared with phenotype 1. Lastly,

TABLE I. Characteristics of the 4 asthma clusters from the training set*

Variables	Missing or uncertain, T1/T2/T3/T4	Cluster T1 (n = 69 [moderate to severe, well controlled])	Cluster T2 (n = 56 [severe late-onset asthma with airway obstruction, high BMI, smoking, and OCS use])	Cluster T3 (n = 68 [severe asthma with airway obstruction and OCS use but no smoking history])	Cluster T4 (n = 73 [severe asthma with female predominance, high BMI, frequent exacerbations, and OCS use but no history of smoking or airway obstruction])	*P* value
Age (y)		42.9 ± 15.6	57.4 ± 10.1	52.5 ± 15	47.5 ± 13.6	<.001‡
Female sex (%)		55.07	57.14	47.06	83.56	<.001‖
Total daily OCS dose (normalized to mg of prednisolone)†		0 (0-0)	0 (0-10)	4 (0-10)	0 (0-10)	<.001§
Asthma onset (y)†		17 (5-30)	42.5 (30.8-52)	17.5 (5.75-37)	20 (7-37)	<.001§
FEV_1 (% predicted)†		88.5 ± 16.9	58.9 ± 15.6	48.5 ± 13.8	79.2 ± 15.4	<.001‡
FEV_1/FVC ratio†		0.737 ± 0.0836	0.557 ± 0.0978	0.505 ± 0.0787	0.741 ± 0.0832	<.001‡
ACQ-5 score†		0.8 (0.25-1.6)	2 (1.2-2.8)	2.4 (1.6-3.6)	2.6 (1.8-3.2)	<.001§
No. of exacerbations in past year†		0 (0-1)	1 (0.75-3)	2 (1-3.25)	3 (2-4)	<.001§
BMI (kg/m^2)†		25.1 (21.8-28.4)	29 (26-33.3)	25.4 (23.4-28.5)	31.6 (26.8-35.8)	<.001§
Pack years†		0 (0-0)	16 (4.94-25.9)	0 (0-0)	0 (0-0)	<.001§
High ICS dose (%)	2/5/2/5	37.3	98.0	96.9	95.5	<.001‖
Blood eosinophils ($\times 10^3/\mu L$)	0/1/1/4	0.199 (0.1-0.3)	0.301 (0.119-0.56)	0.22 (0.1-0.494)	0.2 (0.0997-0.385)	.0273§
Blood neutrophils ($\times 10^3/\mu L$)	0/1/1/4	3.98 (3.10-4.87)	5.09 (3.97-7.15) NA: 1	5.13 (3.74-7.61) NA: 1	4.73 (3.62-6.70) NA: 4	<.001§
Sputum eosinophils (%)	39/20/44/45	0.78 (0.252-5.88)	4.88 (1.42-20) NA: 20	3.67 (1.01-29.6) NA: 44	2.42 (0.288-7.06) NA: 45	.0147§
Sputum neutrophils (%)	39/20/44/45	54.4 (27.4-64.1)	59.7 (44.0-70.5)	63.4 (40.2-86.3)	49.4 (31.5-70.0)	.268§
FENO (ppb)	1/1/0/10	24 (14-42.9)	29 (15.5-53)	33.8 (21.2-56.5)	26.5 (15.5-48)	.332§
Total IgE (IU/mL)	0/1/2/3	131 (54-316)	140 (47-310)	113 (41.5-352)	124 (41.2-319)	.995§
Atopy (positive) SPT response or specific IgE level (%)	0/2/0/1	84.1	55.6	70.6	73.6	.006‖

ACQ-5, Five first questions of the Asthma Control Questionnaire; *FENO*, fraction of exhaled nitric oxide; *FVC*, forced vital capacity; *ICS*, inhaled corticosteroid; *NA*, not available.
*Values are presented as means ± SDs, medians (first-third quartiles), or percentages.
†Variables included in the clustering.
‡ANOVA.
§Kruskal-Wallis test.
‖χ^2 Test.

tyrosine-protein kinase LYN and fucosyl transferase 5 (FUT5) were found to be decreased in phenotype 3 when compared with phenotype 2.

Differential transcript expression in sputum cells

Comparing phenotype 2 with phenotype 1 yielded 8 differentially expressed genes, 2 of which are linked to the hematopoietic cell lineage pathway (*CSF1* and *CD1B*), both being more expressed in phenotype 2. The comparison of phenotype 3 with phenotype 1 highlighted 147 genes, 5 of them encoding the proteins CTSB, PDIA3, CD4, CD74, and CALR, which are linked to antigen processing and presentation pathway. Pathway enrichment analysis of phenotype 3 compared with phenotype 2 revealed pathways related to the regulation of the actin cytoskeleton (ITGB1, ITGB8, FN1, DIAPH2, F2R, and ACTN2), and to fibronectin matrix formation (ITGB1 and FN1), which are potentially linked to the effect of smoking in patients with severe asthma. Three probe sets, one of which annotated to a known gene, *DAGLB*, which encodes for the enzyme diacylglycerol lipase, were highlighted in the comparison of phenotype 4 with phenotype 2, and therefore no pathway enrichment analysis was done. The comparison of phenotype 4 with phenotype 3 revealed 14 differentially expressed genes, including those encoding proteins related to the cell cycle and growth factor–regulating pathways (*MAPK1*, *E2F1*, and *SPRY2*) and to the modulation of immune system responses, particularly the interferon signaling pathway (*OASL*, *OAS3*, and *TRIM14*).

DISCUSSION

Using the partition-around-medoids clustering algorithm and a bootstrapping method on the large U-BIOPRED cohort of participants with moderate-to-severe asthma, we have identified 4 clusters of asthma: one composed of patients with well-controlled asthma with almost normal lung function but receiving low-to-high doses of inhaled corticosteroid and the other 3 composed of patients with severe asthma. Two of the clusters relate to chronic airflow obstruction, with 1 cluster associated with smokers and ex-smokers with late-onset asthma who had the highest blood and sputum eosinophil counts, whereas the third cluster is associated with nonsmokers receiving OCS therapy. Finally, the fourth cluster of severe asthma relates to obese female asthmatic patients with recurrent exacerbations and near-normal lung function. Because of a bias in the patient recruitment process, there are significantly more patients from one site with a history of smoking and associated with cluster T2, so much so

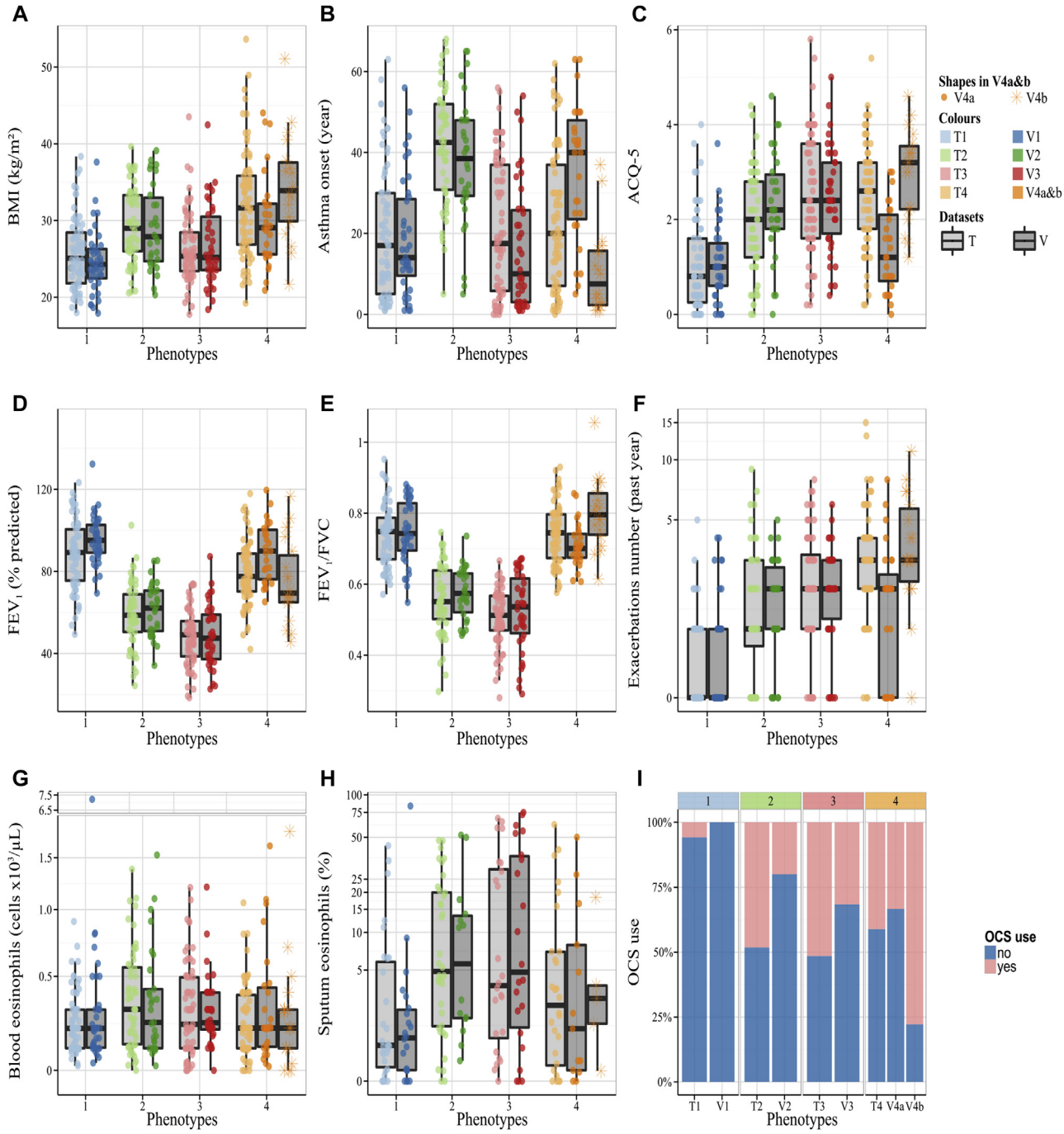


FIG 2. Box plots of main variables. Box plots of some variables used for clustering **(A-F)** together with OCS use as a binary variable **(I)**, blood eosinophil counts **(G)**, and sputum eosinophil percentages **(H)**. Clusters found in the training and validation sets exhibit similar distributions (light and dark hue of the same color), the same color). The V4a and V4b clusters relate to 2 subsets of cluster T4.

that the clinical variables for center and smoking status were confounded. We chose not to adjust the *P* values for the center effect because it would remove part or all of the variability associated with the smoking status.

Our phenotypes are quite distinguishable in terms of controlled versus uncontrolled asthma (phenotype 1 vs phenotypes 2, 3, and 4), airflow obstruction versus normal lung function (phenotypes 2 and 3 vs phenotypes 1 and 4), and infrequent exacerbations versus frequent exacerbations (phenotypes 1, 2, and 3 vs phenotype 4). Thus our robust approach to clustering on the basis of clinicophysiologic parameters has yielded phenotypes characterized based on asthma control, airflow obstruction, recurrent exacerbations, and OCS dependence, which are all well-known features of severe asthma. The 3 clusters of predominantly severe asthma (clusters 2, 3, and 4) had the highest incidence of nasal polyps and exacerbations, including admission to the intensive care unit in the past year, compared with cluster 1. In addition, these 3 clusters also had the greatest use of rescue inhalers and

TABLE II. Differentially abundant proteins in sputum supernatants between clusters in a subset of patients

		Phenotype 2 vs 1 (n = 28 vs 22)		Phenotype 3 vs 1 (n = 20 vs 22)		Phenotype 3 vs 2 (n = 20 vs 28)		Phenotype 4 vs 3 (n = 16 vs 20)	
Target name	**Gene symbol**	**log_2 (FC)‡**	***P* value*†**	**log_2 (FC)‡**	***P* value*†**	**log_2 (FC)‡**	***P* value*†**	**log_2 (FC)‡**	***P* value*†**
IL-16	*IL16*	0.968	**9.87e-6 (4.18e-3)**	1.034	**2.04e-5 (.023)**				
CTAP-III	*PPBP*	1.096	**4.61e-4 (9.13e-3)**						
Trypsin 2	*PRSS2*	0.930	**1.85e-4 (6.72e-3)**						
GM-CSF	*CSF2*	0.898	**6.69e-4 (0.0111)**						
HPLN1	*HAPLN1*	−0.571	**6.68e-5 (5.39e-3)**						
Cathepsin G	*CTSG*			−2.088	**3.91e-5 (.0221)**			1.556	2.29e-3 (.590)
ARSB	*ARSB*			0.966	9.53e-4 (.131)				
PSA2	*PSMA2*			0.955	7.82e-4 (.126)				
LYN	*LYN*					−2.376	**4.32e-5 (.0488)**		
FUT5	*FUT5*					−2.104	6.96e-4 (.0982)		

*Tukey *P* value adjusted for age and sex in the training and validation set pooled together of the proteins that were consistently found both in the training and validation sets for that specific contrast.
†*P* values are shown as nominal (false discovery rate corrected). Boldface indicates that the difference is still significant, even when correcting for false discovery rate by using the Benjamini-Hochberg method.
‡Fold change (FC) is given in base 2 logarithm. If it is positive in X versus Y, this means that the analyte is more abundant in X than Y; negative values mean it is less abundant.

TABLE III. Enriched pathways* with differentially expressed genes between clusters in a subset of patients

Term	**Source**	**Size**	**Phenotype 2 vs 1†‡ (n = 24 vs 31)**	**Phenotype 3 vs 1†‡ (n = 20 vs 31)**	**Phenotype 3 vs 2†‡ (n = 20 vs 24)**	**Phenotype 4 vs 3†‡ (n = 19 vs 20)**
Hematopoietic cell lineage	KEGG	84	0.05 (2)			
Antigen processing and presentation	KEGG	69		.05 (5)		
Regulation of actin cytoskeleton	KEGG	216			.05 (6)	
Arrhythmogenic right ventricular cardiomyopathy	KEGG	74			.032 (4)	
Melanoma	KEGG	72				.032 (2)
Glioma	KEGG	66				.027 (2)
Non–small cell lung cancer	KEGG	57				.020 (2)
MicroRNAs in cancer	KEGG	258				.024 (3)
Prostate cancer	KEGG	90				.05 (2)
Pancreatic cancer	KEGG	67				.028 (2)
Bladder cancer	KEGG	38				.009 (2)
Chronic myeloid leukemia	KEGG	74				.034 (2)
N-glycan trimming in the endoplasmic reticulum and calnexin/calreticulin cycle	REAC	13		.05 (3)		
Fibronectin matrix formation	REAC	74			.05 (2)	
Negative regulation of fibroblast growth factor receptor signaling	REAC	16				.002 (2)
Spry regulation of fibroblast growth factor signaling	REAC	16				.002 (2)
Oncogene-induced senescence	REAC	31				.010 (2)
Pre-NOTCH expression and processing	REAC	57				.035 (2)
Cytokine signaling in immune system	REAC	309				.005 (4)
Interferon signaling	REAC	194				.022 (3)
IFN-γ signaling	REAC	91				.002 (3)
IFN-α/β signaling	REAC	68				.05 (2)

*Pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases.
†*P* values are calculated by using hypergeometric test with false discovery rate correction for multiple testing, as described in Reimand et al.[20]
‡Number of genes from the list found in the pathway (hits) is indicated in brackets. Only pathways with 2 hits or more are shown.

OCSs and, despite this, also had higher scores on the 5 first questions of the Asthma Control Questionnaire. The differential expression of proteins and genes measured in sputum has also provided some insight into the potential pathophysiologic pathways that might govern these phenotypes, particularly those related to the characteristics of severe asthma, namely chronic airflow obstruction and frequent exacerbations.

In contrast to the training set, clustering of the validation set resulted in 1 additional cluster, even though the results were not as stable to resampling as in the training set, as shown by the CDF and deviation from the ideal stability index. A potential reason for this might relate to the fact that we had fewer participants in the validation set compared with the training set, thus increasing the difficulty of finding stable (to resampling) clusters. Furthermore, FEV_1 (percent predicted) and OCS doses, 2 of the variables

included in the clustering, were slightly different between the training and validation sets. Nevertheless, the training and validation set clusters shared similarities, with cluster 4 being divided in the validation set into 2 clusters. This provides an internal replication of the clusters in our cohort.

Our clusters exhibit similarities with some of the clusters previously reported in 2 similar cohorts with mild/moderate and severe asthma, namely the Severe Asthma Research Program (SARP)[4] and the Leicester cohorts,[3] even though they used different clustering algorithms (Ward hierarchical clustering and k-means, respectively). In relation to the SARP cohorts, phenotype 1 relates to SARP cluster 2, phenotype 3 relates to SARP clusters 4 and 5, and phenotype 4 relates to SARP cluster 3. One rather unique feature of our study is the inclusion of a smoking or ex-smoking cohort of patients with severe asthma who were grouped principally in a late-onset, severe airflow obstruction cluster with high blood and eosinophil counts and 55% of the group with evidence of atopy. These patients represent a group of asthmatic patients with features of COPD, namely chronic airflow obstruction, fulfilling the criteria of the asthma-COPD overlap syndrome.[5] A similar cluster has been previously reported,[27,28] although in one cluster of smoking asthmatic patients, the degree of airflow obstruction was minimal, but the cohort that was studied was not one of severe asthma.[29]

In the Leicester study[3] sputum eosinophil counts were also used in clustering, generating a late-onset, obese female severe asthmatic cluster with low sputum eosinophil counts, which is similar to phenotype 4. In our clusters blood and sputum eosinophil counts varied within each of these clusters, with the highest found in the smoking and ex-smoking patients in phenotype 2, supporting further the concept of the asthma-COPD overlap syndrome.[5] However, the clinical clusters did not segregate according to levels of serum IgE or fraction of exhaled nitric oxide. Furthermore, levels of sputum periostin, which were used as a biomarker of T_H2-associated protein, did not differ among the 4 groups.

These 4 distinct phenotypes of asthma that would be recognizable by the clinician experienced in seeing patients with severe asthma would allow patients to be segregated into these clinical characteristics associated with severe asthma. These clusters were also characterized by different pathobiological pathways. Using very strict criteria for defining the differentially abundant proteins by examining for consistency of their expression in both the training and validation sets analyzed separately, we found that IL-16 (lymphocyte chemoattractant factor) was the only protein to be detected as differentially abundant when comparing both severe asthma clusters with airflow obstruction (present in phenotypes 2 and 3) with patients with well-controlled asthma (phenotype 1). IL-16 has been previously associated with asthma and shown to be expressed in abundance in epithelial cells after histamine challenge, in bronchoalveolar lavage fluid after an allergen challenge, and in airway epithelium and $CD4^+$ T cells of airway biopsy specimens.[30-32] Phenotype 2 showed higher levels than phenotype 1 of CTAP-III (CXCL7) and GM-CSF, which might be a reflection of the effect of smoking exposure because CXCL7 has been used as a biomarker for the risk of lung cancer and because GM-CSF mediates cigarette smoke–induced lung neutrophilia.[33,34] In addition, higher levels of CXCL7 and GM-CSF have been shown in patients with COPD secondary to cigarette smoking.[35,36] On the other hand, phenotype 3 showed decreased sputum levels of cathepsin G compared with phenotypes 1 and 4, in which increased systemic levels have been linked to neutrophilic asthma.[6]

LYN kinase was found to be reduced in phenotype 3 (nonsmoking patients with severe obstructed asthma) when compared with phenotype 2 (smoking or ex-smoking asthmatic patients). LYN kinase is an SRC kinase that controls GATA-3 and induces T_H2 cell differentiation,[37] as well as the susceptibility of epithelial cells to their response to cigarette smoke extracts.[38] It has also been implicated in increasing asthma severity in mouse asthma models.[39]

Comparing gene expression between phenotypes 2 and 3 revealed pathways related to regulation of the actin cytoskeleton and to fibronectin matrix formation. The comparison of phenotype 4 (obese and exacerbation-prone asthmatic patients) with phenotype 3 (airflow-obstructed asthmatic patients), by contrast, yielded differential gene pathways related to immune cytokine signaling, particularly interferon signaling and regulation of fibroblast growth factor and signaling of fibroblast growth factor receptor. These specific pathways might be involved in important pathophysiologic aspects underlying the clinical phenotypes identified through this clustering approach based initially on clinicophysiologic features.

Some of the limitations and biases within this analysis need to be highlighted. First, as in any clinical study, the cohort is biased by its inclusion and exclusion criteria (as discussed in detail by Yan et al[8]), but we have been as inclusive as possible. Second, cluster analysis is a descriptive method, and groups can be defined even when there is no underlying structure in the data; this limitation was addressed by assessing the stability, separation, and reproducibility of the clusters. Moreover, the choice of clinical variables might condition the type of clusters found, but the choice of variables we used can be justified by their relevance to day-to-day clinical practice. The proof that the choice was reasonable is in the description of clinical cohorts that makes sense to the clinician. Finally, unsupervised clustering on the basis of the transcriptomics and proteomics data remains another powerful approach toward molecular phenotyping, work that is currently being performed in U-BIOPRED.

In conclusion, the 4 phenotypes of asthma that we describe from the U-BIOPRED cohort have distinct clinical and molecular characteristics that should prove useful to the clinician in directing management of the particularly severe asthma phenotypes. One phenotype is associated with smoking, emphasizing its influence on asthma. The differential molecular characteristics of the 4 phenotypes are not only potentially useful biomarkers of asthma severity but also represent a starting point for drug discovery efforts and the development of better treatments. This will pave the way toward a more personalized approach to asthma management.

We thank all the members of each recruiting center for the recruitment and assessment of the participants.

U-BIOPRED consortium study group members: Nora Adriaens,[a] Hassan Ahmed,[b] Antonios Aliprantis,[c] Kjell Alving,[d] Philipp Badorek,[e] David Balgoma,[f] Clair Barber,[g] An Bautmans,[h] Annelie F. Behndig,[i] Elisabeth Bel,[a] Jorge Beleta,[j] Ann Berglind,[f,k] Alix Berton,[l] Jeanette Bigler,[m] Hans Bisgaard,[n] Grazyna Bochenek,[o] Michael J. Boedigheimer,[m] Klaus Bønnelykke,[n] Joost Brandsma,[p] Armin Braun,[e] Paul Brinkman,[a] Dominic Burg,[q] Davide Campagna,[r] Leon Carayannopoulos,[s] João P. Carvalho da Purfição Rocha,[t] Amphun Chaiboonchoe,[b] Romanas Chaleckis,[f] Courtney Coleman,[u] Chris Compton,[v] Arnaldo D'Amico,[w] Barbro Dahlén,[f,x] Jorge De Alba,[j] Pim de

Boer,[y] Inge De Lepeleire,[h] Tamara Dekker,[a] Ingrid Delin,[f] Patrick Dennison,[f,z] Annemiek Dijkhuis,[a] Aleksandra Draper,[aa] Jessica Edwards,[u] Rosalia Emma,[r] Magnus Ericsson,[x] Veit Erpenbeck,[bb] Damijan Erzen,[cc] Cornelia Faulenbach,[e] Klaus Fichtner,[cc] Neil Fitch,[aa] Breda Flood,[u] Urs Frey,[dd] Martina Gahlemann,[ee] Gabriella Galffy,[ff] Hector Gallart,[f] Trevor Garret,[aa] Thomas Geiser,[gg] Jilaiha Gent,[t] Maria Gerhardsson de Verdier,[l] David Gibeon,[hh] Cristina Gomez,[f] Kerry Gove,[f] Neil Gozzard,[ii] Yi-Ke Guo,[jj] Simone Hashimoto,[a] John Haughney,[kk] Gunilla Hedlin,[f,k] Pieter-Paul Hekking,[a] Elisabeth Henriksson,[x] Lorraine Hewitt,[g] Tim Higgenbottam,[ll] Uruj Hoda,[t] Jans Hohlfeld,[e] Cecile Holweg,[mm] Peter Howarth,[g] Richard Hu,[m] Sile Hu,[hh] Xugang Hu,[m] Val Hudson,[u] Anna J. James,[f] Juliette Kamphuis,[y] Erika J. Kennington,[u] Dyson Kerry,[nn] Matthias Klüglich,[cc] Hugo Knobel,[oo] Richard Knowles,[pp] Alan Knox,[qq] Johan Kolmert,[f] Jon Konradsen,[f,k] Maxim Kots,[rr] Linn Krueger,[dd] Scott Kuo,[hh] Maciej Kupczyk,[f] Bart Lambrecht,[ss] Ann-Sofie Lantz,[f,k] Lars Larsson,[l] Nikos Lazarinis,[x] Saeeda Lone-Satif,[a] Lisa Marouzet,[g] Jane Martin,[g] Sarah Masefield,[tt] Caroline Mathon,[f] John G. Matthews,[mm] Alexander Mazein,[b] Sally Meah,[hh] Andrea Maiser,[hh] Andrew Menzies-Gow,[t] Leanne Metcalf,[u] Roelinde Middelveld,[f] Maria Mikus,[uu] Montse Miralpeix,[j] Philips Monk,[vv] Nadia Mores,[ww] Clare S. Murray,[xx,yy] Jacek Musial,[o] David Myles,[v] Shama Naz,[f] Katja Nething,[cc] Ben Nicholas,[zz] Ulf Nihlen,[l] Peter Nilsson,[uu] Björn Nordlund,[d,f] Jörgen Östling,[l] Antonio Pacino,[aaa] Laurie Pahus,[bbb] Susanna Palkonnen,[ccc] Stelios Pavlidis,[hh] Giorgio Pennazza,[w] Anne Petrén,[f] Sandy Pink,[g] Anthony Postle,[zz] Pippa Powel,[tt] Malayka Rahman-Amin,[u] Navin Rao,[ddd] Lara Ravanetti,[a] Emma Ray,[g] Stacey Reinke,[f] Leanne Reynolds,[u] Kathrin Riemann,[cc] John Riley,[v] Martine Robberechts,[h] Amanda Roberts,[u] Christos Rossios,[hh] Kirsty Russell,[hh] Michael Rutgers,[y] Giuseppe Santini,[ww] Marco Sentoninco,[z] Corinna Schoelch,[cc] James P. R. Schofield,[a,g] Wolfgang Seibold,[cc] Ralf Sigmund,[cc] Marcus Sjödin,[f] Paul J. Skipp,[q] Barbara Smids,[a] Caroline Smith,[g] Jessica Smith,[u] Katherine M. Smith,[qq] Päivi Söderman,[k] Adesimbo Sogbesan,[t] Doroteya Staykova,[eee] Karin Strandberg,[x] Kai Sun,[hh] David Supple,[u] Marton Szentkereszty,[ff] Lilla Tamasi,[ff] Kamran Tariq,[g,z] John-Olof Thörngren,[x] Bob Thornton,[s] Jonathan Thorsen,[n] Salvatore Valente,[w] Wim van Aalderenm,[a] Marianne van de Pol,[a] Kees van Drunen,[a] Marleen van Geest,[l] Jenny Versnel,[u] Jorgen Vestbo,[xx,yy] Anton Vink,[oo] Nadja Vissing,[n] Christophe von Garnier,[gg] Arianne Wagerner,[a] Scott Wagers,[aa] Frans Wald,[cc] Samantha Walker,[u] Jonathan Ward,[fff] Zsoka Weiszhart,[ff] Kristiane Wetzel,[cc] Craig E. Wheelock,[f] Coen Wiegman,[hh] Siân Williams,[kk] Susan J. Wilson,[fff] Ashley Woodcock,[xx,yy] Xian Yang,[hh] Elizabeth Yeyashingham,[ggg] Wen Yu,[m] Wilhelm Zetterquist,[d,f] and Koos Zwinderman[a]

From [a]Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; [b]the European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon, France; [c]Merck Research Laboratories, Boston, Mass; [d]the Department of Women's & Children's Health, Uppsala University, Uppsala, Sweden; [e]Fraunhover Institute for Toxicology and Experimental Medicine, Hannover, Germany; [f]the Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; [g]NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, United Kingdom; [h]MSD, Brussels, Belgium; [i]the Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; [j]Almirall S.A., Barcelona, Spain; [k]the Department of Women's & Childern's Health, Karolinska Institutet, Stockholm, Sweden; [l]AstraZeneca, Mölndal, Sweden; [m]Amgen, Seattle, Wash; [n]Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Genofte Hospital, University of Copenhagen, Copenhagen, Denmark; [o]the Department of Internal Medicine, Jagiellonian University Medical College, Krakow, Poland; [p]the Faculty of Medicine, Southampton University, Southampton, United Kingdom; [q]the Centre for Proteomics Research, Institute of Life Sciences, University of Southampton, Southampton, United Kingdom; [r]the Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy; [s]MSD, Kenilworth, NJ; [t]Royal Brompton and Harefield NHS Fondation Trust, London, United Kingdom; [u]Asthma UK, London, United Kingdom; [v]the Respiratory Therapeutic Unit, GlaxoSmothKline, London, United Kingdom; [w]University of Rome Tor Vergata, Rome, Italy; [x]University Hospital, Karolinska Institutet, Stockholm, Sweden; [y]Longfonds, Amersfoort, The Netherlands; [z]NIHR-Wellcome Trust Clinical Research Facility, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; [aa]BioSci Consulting, Maasmechelen, Belgium; [bb]Translational Medicine, Respiratory Profiling, Novartis Institute for Biomedical Research, Basel, Switzerland; [cc]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany; [dd]University Children's Hospital, Basel, Switzerland; [ee]Boehringer Ingelheim (Schweiz) GmbH, Basel, Switzerland; [ff]Semmelweis Universty, Budapest, Hungary; [gg]the Department of Respiratory Medicine, University Hospital Bern, Bern, Switzerland; [hh]the National Heart and Lung Institute, Imperial College, London, United Kingdom; [ii]UCB, Slough, United Kingdom; [jj]the Data Science Institute, Imperial College, London, United Kingdom; [kk]the International Primary Care Respiratory Group, Aberdeen, Scotland; [ll]Allergy Therapeutics, West Sussex, United Kingdom; [mm]Respiratory and Allergy Diseases, Genentech, San Francisco, Calif; [nn]CromSource, Stirling, United Kingdom; [oo]Philips Research Laboratories, Eindhoven, The Netherlands; [pp]Arachos Pharma, Stevenage, United Kingdom; [qq]the Respiratory Research Unit, University of Nottingham, Nottingham, United Kingdom; [rr]Chiesi Pharmaceuticals, SPA, Parma, Italy; [ss]University of Gent, Gent, Belgium; [tt]the European Lung Foundation, Sheffield, United Kingdom; [uu]Science for Life Laboratory and The Royal Institute of Technology, Stockholm, Sweden; [vv]Synairgen Research, Southampton, United Kingdom; [ww]Università Cattolica del Sacro Cuore, Rome, Italy; [xx]the Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester, Manchester, United Kingdom; [yy]University Hospital of South Manchester, Manchester Academic Health Sciences Centre, Manchester, United Kingdom; [zz]the Faculty of Health Science, Southampton University, Southampton, United Kingdom; [aaa]Lega Italiana Anti Fumo, Catania, Italy; [bbb]Assistance Publique des Hôpitaux de Marseille, Clinique des bronches, allergies et sommeil, Espace EthiqueMéditerranéen, Aix-Marseille Université, Marseille, France; [ccc]the European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium; [ddd]Janssen Research & Development, San Diego, Calif; [eee]the Centre for Biological Sciences, University of Southampton, Southampton, United Kingdom; [fff]the Histochemistry Research Unit, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; and [ggg]UK Clinical Operations, GlaxoSmithKline, Stockley Park, United Kingdom.

Clinical implications: The definition of 4 distinct clusters of asthma linked to different pathobiological pathways provides a better template for the phenotyping and personalized treatment of severe asthma, where high unmet needs remain.

REFERENCES

1. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J 2014;43:343-73.
2. Wenzel SE. Complex phenotypes in asthma: current definitions. Pulm Pharmacol Ther 2013;26:710-5.
3. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med 2008;178:218-24.
4. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010;181:315-23.
5. Bateman ED, Reddel HK, van Zyl-Smit RN, Agusti A. The asthma-COPD overlap syndrome: towards a revised taxonomy of chronic airways diseases? Lancet Respir Med 2015;3:719-28.
6. Baines KJ, Simpson JL, Wood LG, Scott RJ, Gibson PG. Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. J Allergy Clin Immunol 2011;127:153-60.e1-9.
7. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, et al. T-helper type 2-driven inflammation defines major subphenotypes of asthma. Am J Respir Crit Care Med 2009;180:388-95.
8. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. Am J Respir Crit Care Med 2015;191:1116-25.

9. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. Eur Respir J 2015;46:1308-21.
10. Box GEP, Cox DR. An analysis of transformations (with discussion). J R Stat Soc B 1964;26:211-52.
11. Fox J, Weisberg S. An R companion to applied regression. 2nd ed. Thousand Oaks (CA): SAGE Publications; 2011.
12. Kaufman L, Rousseeuw P. Clustering by means of medoids. In: Reports of the faculty of Mathematics and Informatics. Delft, The Netherlands: Delft University; 1987.
13. Monti S, Tamayo P, Mesirov J, Golub T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Machine Learning 2003;52:91-118.
14. Senbabaoglu Y, Michailidis G, Li JZ. Critical limitations of consensus clustering in class discovery. Sci Rep 2014;4:6207.
15. Calinski T. A dendrite method for cluster analysis. Biometrics 1968;24:207.
16. Pavord ID, Pizzichini MMM, Pizzichini E, Hargreave FE. The use of induced sputum to investigate airway inflammation. Thorax 1997;52:498-501.
17. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.
18. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, et al. Nucleic Acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. Mol Ther Nucleic Acids 2014;3:e201.
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8:118-27.
20. Reimand J, Arak T, Vilo J. g:Profiler—a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res 2011;39:W307-15.
21. Benjamini Y, Hochberg Y. Controlling the False discovery rate—a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289-300.
22. Kanehisa M, Goto SKEGG. Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
23. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014;42:D472-7.
24. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab—An S4 Package for Kernel Methods in R. J Stat Software 2004;11:20.
25. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A. Caret: classification and regression training. 2012. Available at: http://CRAN.R-project.org/package=caret.
26. McFadden C, Morgan R, Rahangdale S, Green D, Yamasaki H, Center D, et al. Preferential migration of T regulatory cells induced by IL-16. J Immunol 2007;179:6439-45.
27. Konno S, Taniguchi N, Makita H, Nakamaru Y, Shimizu K, Shijubo N, et al. Distinct phenotypes of cigarette smokers identified by cluster analysis of patients with severe asthma. Ann Am Thor Soc 2015;12:1771-80.
28. Weatherall M, Travers J, Shirtcliffe PM, Marsh SE, Williams MV, Nowitz MR, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. Eur Respir J 2009;34:812-8.
29. Kim TB, Jang AS, Kwon HS, Park JS, Chang YS, Cho SH, et al. Identification of asthma clusters in two independent Korean adult asthma cohorts. Eur Respir J 2013;41:1308-14.
30. Bellini A, Yoshimura H, Vittori E, Marini M, Mattoli S. Bronchial epithelial cells of patients with asthma release chemoattractant factors for T lymphocytes. J Allergy Clin Immunol 1993;92:412-24.
31. Cruikshank WW, Center DM, Nisar N, Wu M, Natke B, Theodore AC, et al. Molecular and functional analysis of a lymphocyte chemoattractant factor: association of biologic function with CD4 expression. Proc Natl Acad Sci U S A 1994;91:5109-13.
32. Laberge S, Ernst P, Ghaffar O, Cruikshank WW, Kornfeld H, Center DM, et al. Increased expression of interleukin-16 in bronchial mucosa of subjects with atopic asthma. Am J Respir Cell Mol Biol 1997;17:193-202.
33. Vlahos R, Bozinovski S, Chan SP, Ivanov S, Linden A, Hamilton JA, et al. Neutralizing granulocyte/macrophage colony-stimulating factor inhibits cigarette smoke-induced lung inflammation. Am J Respir Crit Care Med 2010;182:34-40.
34. Yee J, Sadar MD, Sin DD, Kuzyk M, Xing L, Kondra J, et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. J Clin Oncol 2009;27:2787-92.
35. Di Stefano A, Caramori G, Gnemmi I, Contoli M, Bristot L, Capelli A, et al. Association of increased CCL5 and CXCL7 chemokine expression with neutrophil activation in severe stable COPD. Thorax 2009;64:968-75.
36. Rovina N, Koutsoukou A, Koulouris NG. Inflammation and immune response in COPD: where do we stand? Mediat Inflamm 2013;413735; http://dx.doi.org/10.1155/2013/413735.
37. Charles N, Watford WT, Ramos HL, Hellman L, Oettgen HC, Gomez G, et al. Lyn kinase controls basophil GATA-3 transcription factor expression and induction of Th2 cell differentiation. Immunity 2009;30:533-43.
38. Wang W, Ye Y, Li J, Li X, Zhou X, Tan D, et al. Lyn regulates cytotoxicity in respiratory epithelial cells challenged by cigarette smoke extracts. Curr Mol Med 2014;14:663-72.
39. Beavitt SJ, Harder KW, Kemp JM, Jones J, Quilici C, Casagranda F, et al. Lyn-deficient mice develop severe, persistent asthma: Lyn is a critical negative regulator of Th2 immunity. J Immunol 2005;175:1867-75.