

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 Pepper, 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[‡]

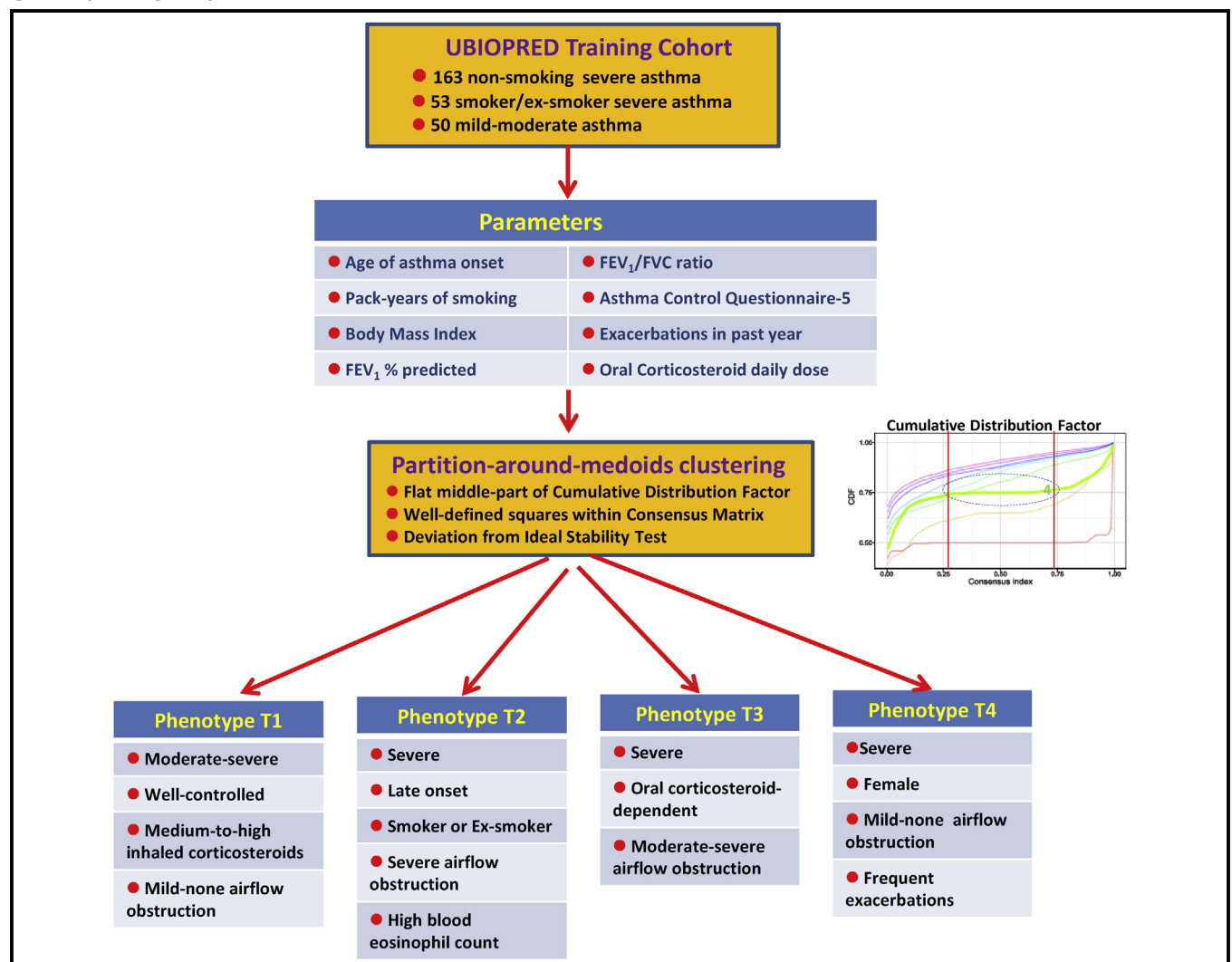
Lyon and Marseille, France; Spring House, Pa; High

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 clinicophysiological 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 clinicophysiological 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 clinicophysiological parameters yielded 4 stable and reproducible clusters that associate with different pathobiological pathways. (J Allergy Clin Immunol 2016;■■■:■■■-■■■.)

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 medications¹ 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 ^athe European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon; ^bJanssen Research and Development LLC, Spring House; ^cJanssen Research and Development Ltd, High Wycombe; ^dAcclarogen, St John's Innovation Centre, Cambridge; ^ethe Academic Medical Centre, University of Amsterdam; ^fthe Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park; ^gAstraZeneca R&D Molndal, and Aretiva R&D, Nottingham; ^hData Science Institute, Imperial College London; ⁱthe Department of Clinical Science, University of Bergen; ^jthe Department of Clinical and Experimental Medicine, University of Catania; ^kDépartement des Maladies Respiratoires, Aix Marseille Université Marseille; ^lthe Centre for Allergy Research, Karolinska Institutet, Stockholm; ^mNational Heart and Lung Institute, Imperial College & Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London; ⁿthe Centre for Respiratory Medicine and Allergy, University of Manchester; ^othe Department of Pulmonology, Semmelweis University, Budapest; ^pFraunhofer Institute for Toxicology and Experimental Medicine, Hannover; ^qthe Faculty of Medicine, Catholic University of the Sacred Heart, Rome; ^rthe Department of Medicine, Jagiellonian University Medical School, Krakow; ^sthe Department of Public Health and Clinical Medicine, Medicine, Umeå university; ^tthe Respiratory Research Unit, University of Nottingham; ^uUniversity Children's Hospital Bern; ^vNIHR Respiratory Biomedical Research Unit, Clinical and Experimental Sciences, Southampton; ^wthe 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. Lefaudeaux 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 from 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.

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.

0091-6749/\$36.00

© 2016 American Academy of Allergy, Asthma & Immunology

<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), FEV₁ as a percentage of predicted value (FEV₁ percent predicted), FEV₁/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 transformation¹⁰ 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 cytopsin preparations. Cytopsin assessments were performed centrally with the outcome of the cytopsin 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 package¹⁷ 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.somallogic.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 KEGG²² and Reactome²³ databases.

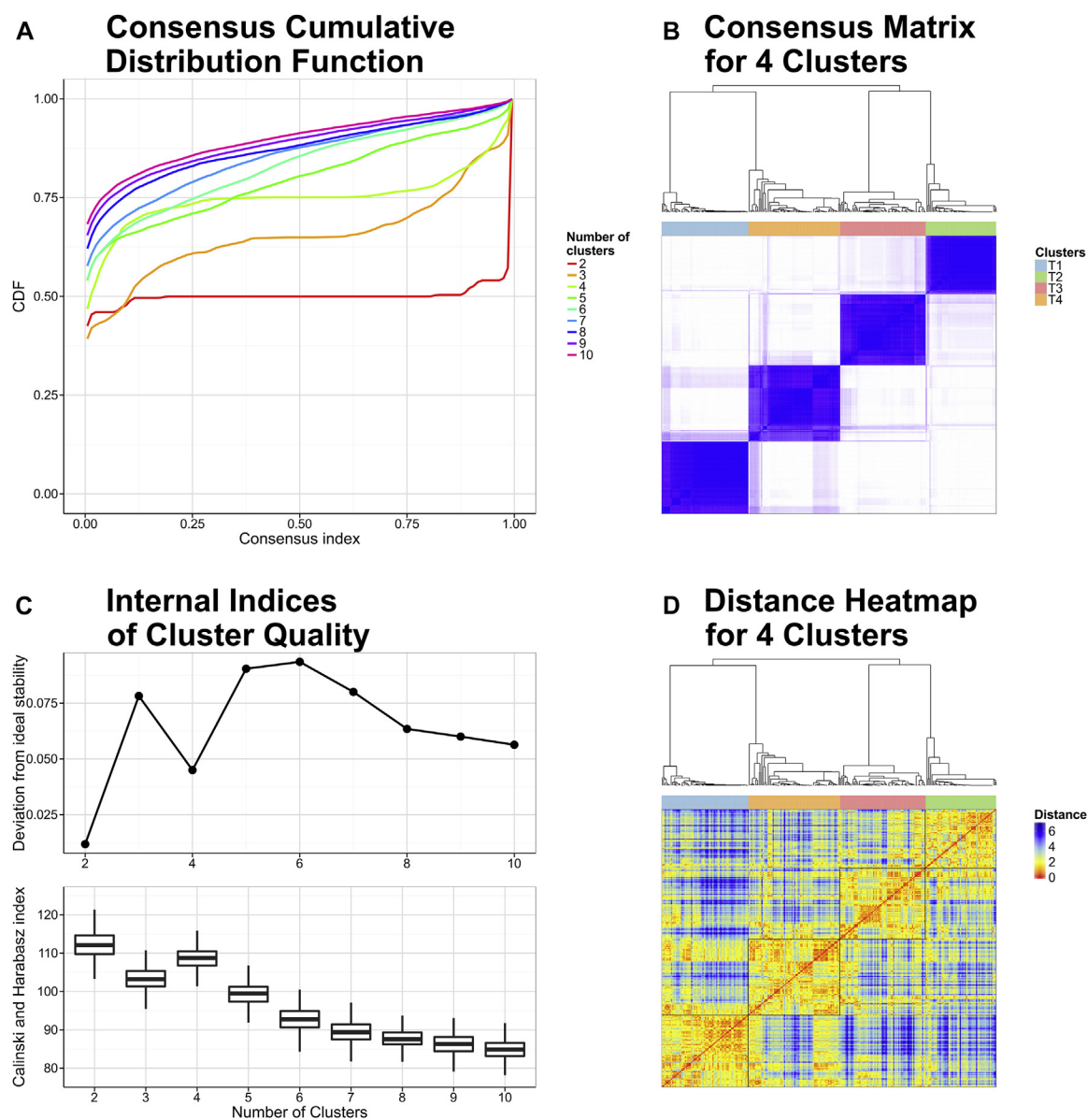


FIG 1. Clustering using the partition-around-medoids (PAM) algorithm on the training set. **A**, Consensus CDF of the consensus matrix shown in **B**. **C**, Two internal quality indexes: deviation from ideal stability and the Calinski and Harabasz index. **D**, Heat map of pairwise distances between 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 FEV₁ (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 FEV₁, 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/m² and 41% with BMI ≥ 30 kg/m²) patients with late-onset severe asthma who smoked and who had relatively poor control, severe airflow obstruction (mean FEV₁, 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/m², and 56% with BMI ≥ 30 kg/m²) 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 kernel²⁴ 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 xlsx 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 $\alpha 2$ (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 (× 10 ³ /μL)	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 (× 10 ³ /μL)	0/1/1/4	3.98 (3.10-4.87)	5.09 (3.97-7.15)	5.13 (3.74-7.61)	4.73 (3.62-6.70)	<.001§
			NA: 1	NA: 1	NA: 4	
Sputum eosinophils (%)	39/20/44/45	0.78 (0.252-5.88)	4.88 (1.42-20)	3.67 (1.01-29.6)	2.42 (0.288-7.06)	.0147§
			NA: 20	NA: 44	NA: 45	
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.

||χ² Test.

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

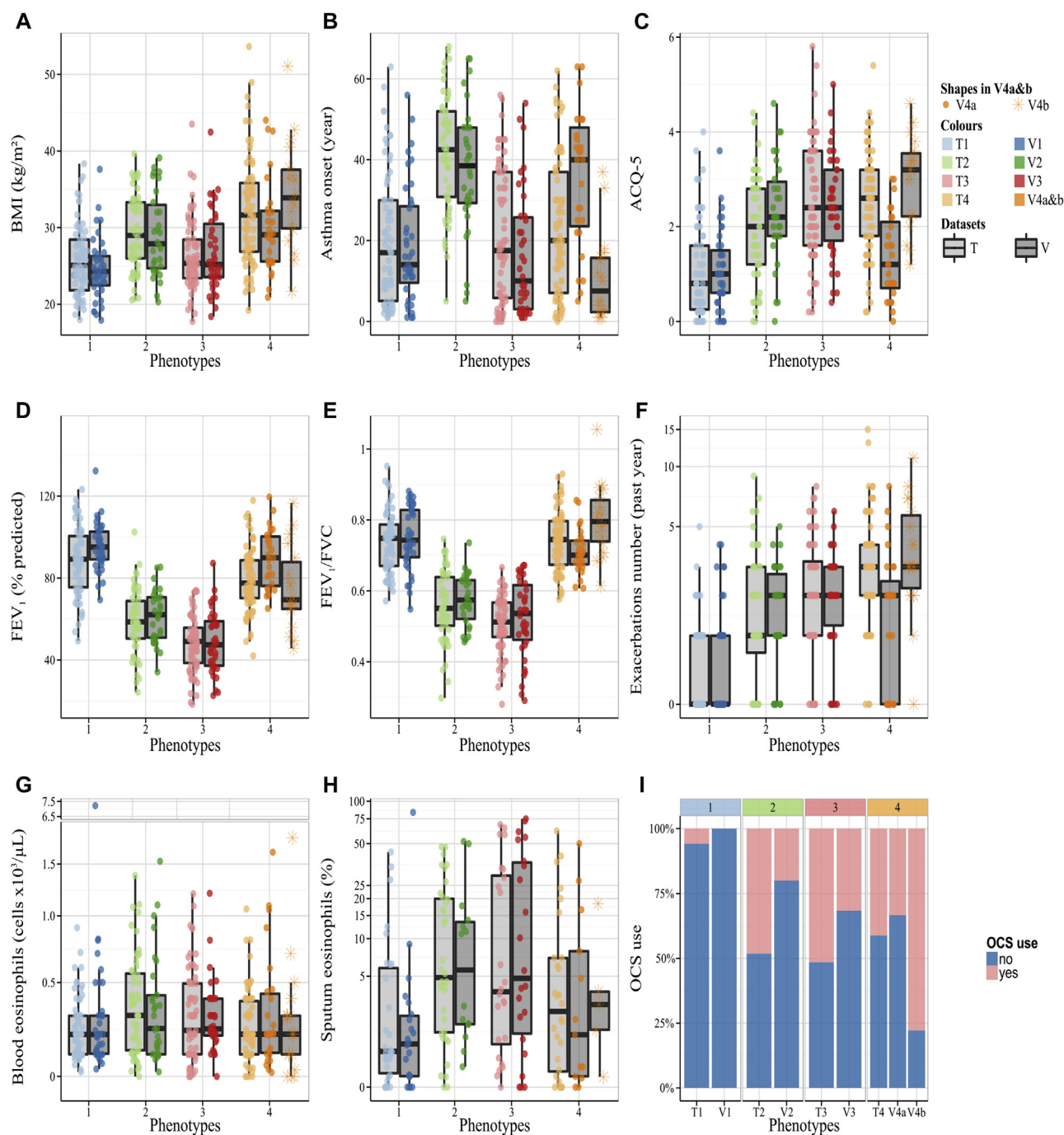
Differential transcript expression in sputum cells

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



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 clinicophysiological 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

Target name	Gene symbol	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)	
		log ₂ (FC) [‡]	P value ^{*†}	log ₂ (FC) [‡]	P value ^{*†}	log ₂ (FC) [‡]	P value ^{*†}	log ₂ (FC) [‡]	P value ^{*†}
IL-16	<i>IL16</i>	0.968	9.87e-6 (4.18e-3)	1.034	2.04e-5 (.023)				
CTAP-III	<i>PPBP</i>	1.096	4.61e-4 (9.13e-3)						
Trypsin 2	<i>PRSS2</i>	0.930	1.85e-4 (6.72e-3)						
GM-CSF	<i>CSF2</i>	0.898	6.69e-4 (0.0111)						
HPLN1	<i>HAPLN1</i>	−0.571	6.68e-5 (5.39e-3)						
Cathepsin G	<i>CTSG</i>			−2.088	3.91e-5 (.0221)			1.556	2.29e-3 (.590)
ARSB	<i>ARSB</i>			0.966	9.53e-4 (.131)				
PSA2	<i>PSMA2</i>			0.955	7.82e-4 (.126)				
LYN	<i>LYN</i>					−2.376	4.32e-5 (.0488)		
FUT5	<i>FUT5</i>					−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, FEV₁ (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 study³ 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 clinicophysiological 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 al⁸), 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,^c David Balgoma,^f Clair Barber,^g An Bautmans,^h Annelie F. Behndig,ⁱ Elisabeth Bel,^a Jorge Beleta,^j Ann Berglund,^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,^c Paul Brinkman,^a Dominic Burg,^q Davide Campagna,^r Leon Carayannopoulos,^s João P. Carvalho da Purificação 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,^c Klaus Fichtner,^{cc} Neil Fitch,^{aa} Breda Flood,^u Urs Frey,^{dd} Martina Gahlemann,^{cc} Gabriella Galfy,^{ff} Hector Gallart,^f Trevor Garret,^{aa} Thomas Geiser,^{gg} Jilaila Gent,^l 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,^{fk} Pieter-Paul Hekking,^a Elisabeth Henriksen,^x Lorraine Hewitt,^g Tim Higgenbottom,^{ll} Uruj Hoda,^t Jans Hohlfeld,^c Cecile Holweg,^{mmm} 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 Klüglic,^{oo} Hugo Knobel,^{oo} Richard Knowles,^{pp} Alan Knox,^{qq} Johan Kolmert,^f Jon Konradsen,^{fk} Maxim Kots,^{rr} Linn Krueger,^{dd} Scott Kuo,^{hh} Maciej Kupczyk,^f Bart Lambrecht,^{ss} Ann-Sofie Lantz,^{fk} 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,^{mmm} Alexander Mazein,^b Sally Meah,^{hh} Andrea Maiser,^{hh} Andrew Menzies-Gow,^t Leanne Metcalf,^u Roelinde Middelvel,^f Maria Mikus,^{uu} Montse Miralpeix,^j Philips Monk,^{vv} Nadia Mores,^{www} 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,^{df} Jörgen Östling,^l Antonio Pacino,^{aaa} Laurie Pahu,^{bbb} Susanna Palkonnen,^{ccc} Stelios Pavlidis,^{hh} Giorgio Pennazza,^w Anne Petré,^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 Robbrechts,^h Amanda Roberts,^h Christos Rossios,^{hh} Kirsty Russell,^{hh} Michael Rutgers,^y Giuseppe Santini,^{ww} Marco Sentoninco,^z Corinna Schoelch,^{cc} James P. R. Schofield,^{ag} 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,^{ccc} 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 Aalderen,^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 Wagener,^a Scott Wagers,^{aa} Frans Wald,^{cc} Samantha Walker,^u Jonathan Ward,^{fff} Zsoka Weiszhar,^{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,^{df} and Koos Zwinderman^a

From ^aAcademic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; ^bthe European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon, France; ^cMerck Research Laboratories, Boston, Mass; ^dthe Department of Women's & Children's Health, Uppsala University, Uppsala, Sweden; ^eFraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany; ^fthe Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; ^gNIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, United Kingdom; ^hMSD, Brussels, Belgium; ⁱthe Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; ^jAlmirall S.A., Barcelona, Spain; ^kthe Department of Women's & Children's Health, Karolinska Institutet, Stockholm, Sweden; ^lAstraZeneca, Mölndal, Sweden; ^mAmgen, Seattle, Wash; ⁿCopenhagen Prospective Studies on Asthma in Childhood, Herlev and Genofte Hospital, University of Copenhagen, Copenhagen, Denmark; ^othe Department of Internal Medicine, Jagiellonian University Medical College, Krakow, Poland; ^pthe Faculty of Medicine, Southampton University, Southampton, United Kingdom; ^qthe Centre for Proteomics Research, Institute of Life Sciences, University of Southampton, Southampton, United Kingdom; ^rthe Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy; ^sMSD, Kenilworth, NJ; ^tRoyal Brompton and Harefield NHS Foundation Trust, London, United Kingdom; ^uAsthma UK, London, United Kingdom; ^vthe Respiratory Therapeutic Unit, GlaxoSmithKline, London, United Kingdom; ^wUniversity of Rome Tor Vergata, Rome, Italy; ^xUniversity Hospital, Karolinska Institutet, Stockholm, Sweden; ^yLongfonds, Amersfoort, The Netherlands; ^zNIHR-Wellcome Trust Clinical Research Facility, Faculty of

Medicine, University of Southampton, Southampton, United Kingdom; ^{aa}Bio-Sci 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 University, 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; ⁱⁱUCB, 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; ⁿⁿCromSource, 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 Ethique Méditerranéen, 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, Agustí 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.

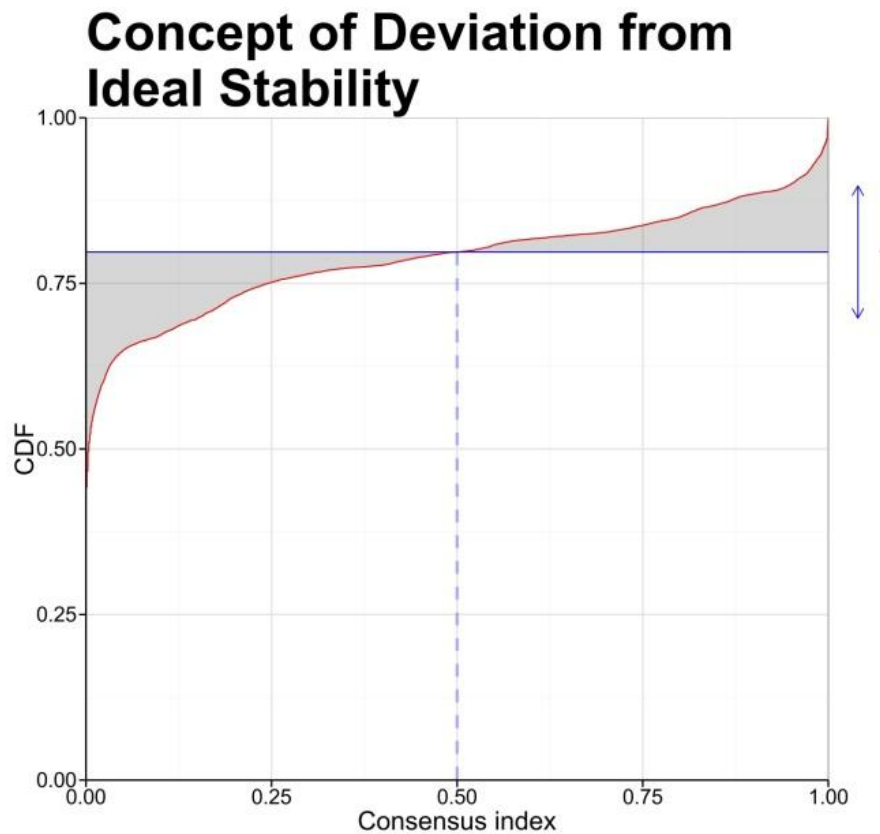
9. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015;46:1308-21.
10. Box GEP, Cox DR. An analysis of transformations (with discussion). *J R Stat Soc B* 1964;26:211-52.
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, Casagrande 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.

Online repository

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

D. Lefaudeux, MSc, B. De Meulder, PhD, M. J. Loza, PhD, N. Peffer, BS, A. Rowe, PhD, F. Baribaud, PhD, A. T. Bansal, PhD, R. Lutter, PhD, A. R. Sousa, PhD, J. Corfield, MSc, I. Pandis, PhD, P.S. Bakke MD, M.Caruso MD, P Chanez MD, S.E. Dahlen MD , L.J. Fleming MD, S.J. Fowler MD, I. Horvath MD, N. Krug MD, P. Montuschi MD, M. Sanak MD, T. Sandstrom MD, D.E. Shaw MD, F. Singer MD, P.J. Sterk, MD, PhD, G. Roberts, MD, I. M. Adcock, PhD, R. Djukanovic, MD, C. Auffray, PhD, K. F. Chung, MD and the U-BIOPRED Study Group

Figure E1: Deviation from Ideal Stability Index



The figure E1 shows the concept of deviation from ideal stability. It corresponds to the grey areas; the smaller it is, the closer to ideal stability it is.

This index defines an objective measure for stability based on a consensus cumulative distribution function from consensus clustering. Ideal stability is reached when each pair of objects is always clustered the same way, either always separated or always together. This corresponds to a constant cumulative distribution function as shown by the blue line. The idea behind the deviation from ideal stability (DIS) is to measure the minimal area between the red curve (the cumulative distribution function from the consensus matrix) and the blue line, which scans all values from 0 to 1. Since the cumulative distribution is increasing it can be shown that the minimal value is always reached for $\alpha = \text{CDF}(0.5)$. This index varies from 0 to 0.5, and 0 is reached for ideally stable clusters.

This concept is formalized as follow:

$$DIS = \min_{\alpha \in [0,1]} \int_{x=0}^1 |CDF(x) - \alpha| dx \stackrel{\text{min}}{=} \int_{x=0}^1 |CDF(x) - CDF(0.5)| dx$$

It is possible to increase the importance of the central values, which corresponds to random assignment, by adding a power on the distance between the CDF and α that would be higher for values of x close to 0 or 1.

Table E1: Characteristics of the training and validation sets[□]

Variables	Training set (n=266)	Validation set (n=152)	P-Value
Cohort [◇]	SAn: 163 ; SAsm: 53; MMA: 50	SAn: 90; SAsm: 33; MMA: 29	0.895 ³
Age	49.6 ± 14.7	50.0 ± 14.2	0.819 ¹
Gender (female)	58.5%	61.2%	0.657 ³
Total daily OCS dose (normalised to prednisolone) [†]	0 (0 – 7.5)	0 (0 – 5)	0.058 ²
Asthma onset (years) [†]	24 (8.25 – 40)	21.5 (7.75 – 39.2)	0.358 ²
FEV ₁ % predicted [†]	69.4 ± 22.2	73.6 ± 23.5	0.0717 ¹
FEV ₁ /FVC [†]	0.645 ± 0.136	0.657 ± 0.136	0.246 ¹
ACQ_5 [†]	2 (1 – 3)	2 (1 – 2.8)	0.553 ²
Exacerbations number past year [†]	2 (0 – 3)	2 (0 – 3)	0.587 ²
BMI [†]	27.2 (23.9 – 32.3)	26.5 (23.7 – 32.0)	0.554 ²
Pack Years [†]	0 (0 – 1.95)	0 (0 – 4)	0.268 ²

[□] Values are presented as mean ± sd, median (1st – 3rd quartiles) or percentages.

¹ ANOVA

² Kruskal-Wallis test

³ χ^2 test

[†] Variables included in the clustering

[◇] SAn: Severe Asthmatics non smoker; SAsm: Severe Asthmatics smokers; MMA: Mild to Moderate Asthmatics

Table E2: Detailed characteristics of the 4 clusters from the training set [□]

Variables		Missing or uncertain T1 / T2 / T3 / T4	T1 (n = 69)	T2 (n = 56)	T3 (n = 68)	T4 (n = 73)	P-Value
Demographics	Cohort [◊] (SAn / SAsm / MMA)		22 / 3 / 44	15 / 40 / 1	62 / 4 / 2	64 / 6 / 3	< 0.001 ³
	Pseudo Sites (a / b / c / d / e / f / g / h / i / j / k / l / m / n / o)		3 / 5 / 13 / 6 / 5 / 6 / 1 / 13 / 1 / 3 / 5 / 2 / 3 / 3 / 0	2 / 5 / 3 / 15 / 7 / 0 / 8 / 5 / 0 / 5 / 4 / 1 / 0 / 1 / 0	2 / 3 / 10 / 2 / 7 / 5 / 6 / 16 / 1 / 9 / 3 / 0 / 0 / 4 / 0	6 / 3 / 12 / 1 / 2 / 4 / 8 / 20 / 0 / 4 / 5 / 2 / 4 / 1 / 1	< 0.001 ³
	Age		42.9 ± 15.6	57.4 ± 10.1	52.5 ± 15	47.5 ± 13.6	< 0.001 ¹
	Gender (female)		55.1%	57.1%	47.1%	83.6%	< 0.001 ³
	BMI		25.1 (21.8 – 28.4)	29 (26 – 33.3)	25.4 (23.4 – 28.5)	31.6 (26.8 – 35.8)	< 0.001 ²
Blood cell count	WBC (x10 ³ /μl)	0 / 1 / 1 / 4	6.9 (5.59 – 7.9)	8.83 (7.29 – 10.6)	8.3 (6.45 – 10.8)	7.7 (6.4 – 9.7)	< 0.001 ²
	Platelets (x10 ³ /μl)	0 / 1 / 1 / 4	245 (215 – 282)	292 (244 – 324)	267 (236 – 310)	276 (221 – 324)	0.00644 ²
	Neutrophils (x10 ³ /μl)	0 / 1 / 1 / 4	3.99 (3.1 – 4.87)	5.1 (3.98 – 7.16)	5.13 (3.75 – 7.62)	4.74 (3.62 – 6.7)	< 0.001 ²
	Neutrophils (%)	0 / 1 / 1 / 4	59.1 ± 8.23	61.3 ± 12.4	65.5 ± 11.6	63.3 ± 12.4	0.0077 ¹
	Eosinophils (x10 ³ /μl)	0 / 1 / 1 / 4	0.199 (0.1 – 0.3)	0.301 (0.119 – 0.560)	0.22 (0.1 – 0.494)	0.2 (0.0997 – 0.385)	0.0273 ²
	Eosinophils (%)	0 / 1 / 1 / 4	2.61 (1.78 – 4.35)	3.75 (1.38 – 6.87)	3.57 (1.02 – 6.71)	2.77 (1.30 – 4.35)	0.342 ²
	Basophils (x10 ³ /μl)	0 / 1 / 1 / 4	0.0199 (0 – 0.0921)	0.0444 (0.0288 – 0.0855)	0.0398 (0.0101 – 0.0997)	0.02 (0 – 0.0995)	0.0413 ²
	Monocytes (x10 ³ /μl)	0 / 1 / 1 / 4	0.43 (0.398 – 0.58)	0.58 (0.441 – 0.726)	0.565 (0.348 – 0.7)	0.499 (0.385 – 0.694)	0.00222 ²
	Lymphocytes (x10 ³ /μl)	0 / 1 / 1 / 4	1.9 (1.6 – 2.37)	2.16 (1.59 – 2.85)	1.7 (1.45 – 2.2)	1.98 (1.48 – 2.31)	0.105 ²
	Lymphocytes (%)	0 / 1 / 1 / 4	29.8 ± 6.8	26.1 ± 10.2	22.4 ± 8.42	26.3 ± 9.76	< 0.001 ¹
Sputum cell counts	Neutrophils (%)	39 / 20 / 44 / 45	54.4 (27.4 – 64.2)	59.7 (44.1 – 70.6)	63.4 (40.2 – 86.3)	49.4 (31.6 – 70.0)	0.268 ²

	Eosinophils (%)	39 / 20 / 44 / 45	0.78 (0.252 – 5.88)	4.88 (1.42 – 20)	3.67 (1.01 – 29.6)	2.42 (0.288 – 7.06)	0.0147 ²
	Lymphocytes (%)	39 / 20 / 44 / 45	0.925 (0.582 – 2)	1 (0.662 – 1.61)	0.915 (0.59 – 1.54)	0.945 (0.355 – 1.4)	0.821 ²
	Macrophages (%)	39 / 20 / 44 / 45	39.9 (21.6 – 66.5)	22.2 (13.1 – 42.2)	15.2 (7.75 – 32.6)	40.9 (17.9 – 54.3)	0.00127 ²
Lung function	FEV1% predicted		88.5 ± 16.9	58.9 ± 15.6	48.5 ± 13.8	79.2 ± 15.4	< 0.001 ¹
	FEV1 relative change		8.3 (5.09 – 14.9)	17.6 (11.8 – 23.9)	16.5 (7.75 – 27.1)	11.6 (6.72 – 19.7)	< 0.001 ²
	FEV1 absolute change		6.99 (4.74 – 12.2)	9.99 (6.98 – 14.6)	8.86 (3.72 – 13.5)	8.55 (5.32 – 14.8)	0.167 ²
	FVC% predicted		101 ± 19.4	86.9 ± 18.7	78.6 ± 18.1	91.2 ± 18.8	< 0.001 ¹
	FVC relative change		2.55 (-0.11 – 7.46)	10.5 (6.62 – 19.1)	9.95 (4.03 – 16.9)	7.6 (1.83 – 13.8)	< 0.001 ²
	FVC absolute change		2.73 (-0.15 – 6.65)	10.7 (5.97 – 16.3)	9.04 (2.88 – 13)	7 (2.19 – 11.7)	< 0.001 ²
	FEF 25-75 % predicted		57.7 (42.6 – 73.9)	21.8 (16.7 – 32.4)	16.4 (12.2 – 22.5)	51.7 (33 – 62.7)	< 0.001 ²
	FEF 25-75 relative change	0 / 0 / 0 / 1	29.5 (10.8 – 43.2)	23.8 (8.86 – 37.6)	22.8 (7.22 – 53.7)	27.8 (5.66 – 50.1)	0.941 ²
	FEF 25-75 absolute change	0 / 0 / 0 / 1	12.4 (6.69 – 23.3)	5.03 (1.89 – 9.45)	3.51 (0.93 – 11.3)	10.2 (3.04 – 21.4)	< 0.001 ²
	FEV1/FVC		0.737 ± 0.0836	0.557 ± 0.0978	0.505 ± 0.0787	0.741 ± 0.0832	< 0.001 ¹
	FEV1/FVC (post salbutamol)		76.2 ± 9.71	58.5 ± 9.98	54.7 ± 10.6	75.3 ± 9.1	< 0.001 ¹
	PEF% predicted		93.1 ± 22.6	63.2 ± 19	53.5 ± 18.4	85.2 ± 19.5	< 0.001 ¹
	TLC% predicted		108 ± 22	109 ± 21.5	114 ± 25	93.3 ± 17.6	< 0.001 ¹
	Residual volume (L)	12 / 9 / 19 / 27	2.09 (1.62 – 2.66)	3.05 (2.47 – 3.4)	2.85 (2.43 – 3.49)	2.2 (1.44 – 2.58)	< 0.001 ²
Airflow limitation	SGaw	12 / 10 / 19 / 27	1.21 (0.67 – 2.06)	0.71 (0.505 – 1.11)	0.43 (0.28 – 0.85)	0.825 (0.492 – 1.67)	< 0.001 ²
	PEF relative change	1 / 2 / 0 / 1	8.88 (1.49 – 16.9)	16 (6.96 – 33.6)	12.2 (4.87 – 31.7)	11.8 (5.66 – 21)	0.0122 ²
	Ten Brinke et al (positive)		8.7%	50.0%	51.5%	13.7%	< 0.001 ³
	Quanjer (positive)		1.4%	23.2%	38.2%	0.0%	< 0.001 ³
Co	Nasal polyps diagnosed (positive)	8 / 2 / 5 / 7	13.1%	38.9%	41.2%	36.6%	0.00272 ³

	Nasal polyps surgery diagnosed (positive)	8 / 2 / 1 / 5	11.5%	33.3%	32.8%	27.9%	0.0200 ³
	Sinusitis diagnosed (positive)	10 / 2 / 2 / 8	22.0%	37.0%	28.8%	40.0%	0.136 ³
	Sinus surgery diagnosed (positive)	8 / 2 / 1 / 5	14.8%	29.6%	20.9%	27.9%	0.191 ³
	Allergic rhinitis diagnosed (positive)	11 / 4 / 8 / 7	58.6%	32.7%	58.3%	66.7%	0.00213 ³
	Non allergic rhinitis diagnosed (positive)	9 / 4 / 5 / 8	13.3%	17.3%	14.3%	15.4%	0.944 ³
	Hay fever diagnosed (positive)	11 / 5 / 2 / 7	50.0%	35.3%	56.1%	54.5%	0.115 ³
	Hypertension diagnosed (positive)	8 / 4 / 1 / 6	18.0%	42.3%	31.3%	26.9%	0.0396 ³
	Eczema diagnosed (positive)	9 / 3 / 2 / 6	31.7%	22.6%	33.3%	34.3%	0.520 ³
	GERD diagnosed (positive)	10 / 5 / 3 / 6	22.0%	47.1%	38.5%	58.2%	< 0.001 ³
	Vocal chord diagnosed (positive)	8 / 2 / 3 / 5	1.6%	3.7%	4.6%	10.3%	0.145 ³
	Coronary diagnosed (positive)	8 / 2 / 2 / 5	0.0%	7.4%	3.0%	5.9%	0.176 ³
	Congestive diagnosed (positive)	8 / 3 / 3 / 5	0.0%	0.0%	1.5%	1.5%	0.630 ³
	Osteoporosis diagnosed (positive)	9 / 5 / 2 / 4	6.7%	33.3%	30.3%	20.6%	0.00227 ³
	Diabetes diagnosed (positive)	8 / 2 / 1 / 6	1.6%	16.7%	7.5%	11.9%	0.0353 ³
	Bronchitis (positive)	3 / 1 / 4 / 3	25.8%	36.3%	53.1%	57.1%	< 0.001 ³
	Chronic dronchitis (positive)	2 / 0 / 4 / 2	7.5%	17.9%	12.5%	8.5%	0.254 ³
	Emphysema or COPD (positive)	0 / 1 / 0 / 2	4.3%	7.3%	4.4%	2.8%	0.697 ³
	Pneumonia (positive)	4 / 2 / 4 / 3	1.5%	29.6%	43.8%	35.7%	0.00466 ³
History	Smoking status		Current smoker: 2 Ex smoker: 6 Non smoker: 61	Current smoker: 13 Ex smoker: 38 Non smoker: 5	Current smoker: 3 Ex smoker: 6 Non smoker: 69	Current smoker: 4 Ex smoker: 7 Non smoker: 62	< 0.001 ²

	Smoking years		0 (0 – 0)	20 (10 – 32.3)	0 (0 – 0)	0 (0 – 0)	< 0.001 ²
	Pack Years		0 (0 – 0)	16 (4.94 – 25.9)	0 (0 – 0)	0 (0 – 0)	< 0.001 ²
	Second hand smoke (positive)	0 / 0 / 0 / 1	18.8%	19.6%	16.2%	19.4%	0.952 ³
	Asthma onset (years)		17 (5 – 30)	42.5 (30.8 – 52)	17.5 (5.75 – 37)	20 (7 – 37)	< 0.001 ²
	Excacerbations number past year		0 (0 – 1)	1 (0.75 – 3)	2 (1 – 3.25)	3 (2 – 4)	< 0.001 ²
	Icu admission ever (positve)	0 / 0 / 0 / 1	8.7%	5.4%	27.9%	26.4%	< 0.001 ³
	Icu admission past year (positive)	0 / 0 / 0 / 1	0.0%	3.6%	1.5%	6.9%	0.0876 ³
	Icu admission times ever	0 / 0 / 0 / 1	Never: 63 Once: 2 2 to 5: 4 More than 5: 0	Never: 53 Once: 2 2 to 5: 0 More than 5: 0	Never: 49 Once: 8 2 to 5: 8 More than 5: 3	Never: 53 Once: 6 2 to 5: 12 More than 5: 1	< 0.001 ²
	Emergency room (positive)	2 / 0 / 1 / 0	31.3%	53.6%	70.1%	76.8%	< 0.001 ³
	Asthma mother (positive)	1 / 3 / 8 / 0	16.2%	20.8%	23.3%	26.0%	0.5417 ³
	Asthma father (positive)	4 / 3 / 8 / 3	7.7%	9.4%	18.3%	17.1%	0.197 ³
	Parental asthma (positive)	4 / 4 / 9 / 2	24.6%	30.7%	39.0%	40.8%	0.177 ³
Atopy	FeNO	1 / 1 / 0 / 10	24 (14 – 42.9)	29 (15.5 – 53)	33.8 (21.2 – 56.5)	26.5 (15.5 – 48)	0.332 ²
	Total IgE (iu/ml)	0 / 1 / 2 / 3	131 (54 – 316)	140 (47 – 310)	113 (41.5 – 352)	124 (41.2 – 319)	0.995 ²
	Atopy combined SPT and Ige results (positive)	0 / 2 / 0 / 1	84.1%	55.6%	70.6%	73.6%	0.0058 ³
Atopy triggers	Stress (positive)	2 / 3 / 5 / 2	35.8%	50.9%	63.5%	70.4%	< 0.001 ³
	Menstrual cycle (positive, female only)	33 / 29 / 37 / 17	2.8%	0.0%	16.1%	14.3%	0.0476 ³
	Routine physical activities (positive)	0 / 1 / 2 / 2	33.3%	72.7%	65.0%	78.9%	< 0.001 ³
	Dust (positive)	4 / 5 / 8 / 2	64.6%	52.9%	81.7%	73.2%	0.00814 ³

	Pollen (positive)	5 / 3 / 7 / 3	54.7%	35.8%	68.9%	72.9%	< 0.001 ³
	Wood smoke (positive)	6 / 3 / 8 / 13	25.4%	47.2%	43.3%	53.3%	0.0122 ³
	Pets (positive)	3 / 4 / 6 / 4	47.0%	17.3%	50.0%	58.0%	< 0.001 ³
	Fungus (positive)	7 / 15 / 9 / 11	27.4%	29.2%	52.5%	56.4%	0.00109 ³
	Perfumes (positive)	4 / 1 / 2 / 4	35.4%	49.1%	45.4%	71.0%	< 0.001 ³
	Respiratory infections (positive)	3 / 0 / 3 / 0	78.8%	78.6%	95.4%	91.8%	0.00574 ³
	Pollutants (positive)	5 / 3 / 4 / 7	42.2%	66.0%	68.8%	72.7%	0.00141 ³
	Aspirin (positive)	8 / 13 / 23 / 15	6.6%	9.3%	24.4%	34.5%	< 0.001 ³
	Barns (positive)	15 / 6 / 21 / 16	35.2%	20.0%	42.6%	43.9%	0.0460 ³
	Physical exercise (positive)	4 / 2 / 7 / 4	70.8%	83.3%	90.2%	86.7%	0.0209 ³
	Cold air (positive)	1 / 4 / 3 / 0	50.0%	53.8%	80.0%	80.8%	< 0.001 ³
Medication	Rescue inhaler use		Never: 25 Once a month: 10 Weekly: 15 ≥2 a week: 13 Once a day: 6	Never: 14 Once a month: 5 Weekly: 11 ≥2 a week: 8 Once a day: 18	Never: 8 Once a month: 6 Weekly: 15 ≥2 a week: 16 Once a day: 23	Never: 5 Once a month: 13 Weekly: 4 ≥2 a week: 16 Once a day: 35	< 0.001 ²
	OCS (positive)		5.8%	48.2%	51.5%	41.1%	< 0.001 ³
	Total daily OCS dose (normalised to prednisolone)		0 (0 – 0)	0 (0 – 10)	4 (0 – 10)	0 (0 – 10)	< 0.001 ²
	Injectable CS (positive)	0 / 1 / 1 / 0	0.0%	1.8%	11.9%	6.8%	0.00966 ³
	Leukotriene modifiers (positive)		17.4%	33.9%	36;8%	47.9%	0.00183 ³
	ICS (positive)		100.0%	92.9%	98.5%	94.5%	0.0866 ³
	ICS dose range (high)	2 / 5 / 2 / 5	37.3%	98.0%	97.0%	95.6%	< 0.001 ³
	Combination ICS-LABA (positive)		33.3%	87.5%	83.8%	87.7%	< 0.001 ³
	Oral Beta agonist (positive)	0 / 1 / 1 / 0	0.0%	0.0%	6.0%	2.7%	0.0694 ³

	LABA (positive)		36.2%	91.1%	82.6%	87.7%	< 0.001 ³
	Anti IgE	0 / 1 / 1 / 0	4.3%	18.2%	11.9%	19.2%	0.0411 ³
Questionnaires	AQLQ Mean	3 / 3 / 1 / 1	6.13 (5.33 – 6.47)	4.72 (3.56 – 5.47)	4.22 (3.22 – 5.14)	4.39 (3.52 – 5.17)	< 0.001 ²
	AQLQ Symptom	3 / 3 / 1 / 1	6.08 (5.29 – 6.5)	4.67 (3.5 – 5.5)	4.17 (3.08 – 5.12)	4.29 (3.4 – 5.19)	< 0.001 ²
	AQLQ Activity Limitation	3 / 3 / 1 / 1	6 (4.93 – 6.55)	4.45 (3.73 – 5.55)	3.91 (3.32 – 5.09)	4.22 (3.34 – 5)	< 0.001 ²
	AQLQ Environment	3 / 3 / 1 / 1	6 (5.25 – 6.69)	5 (3.75 – 6)	4.75 (3.5 – 5.62)	4.5 (3.5 – 5.75)	< 0.001 ²
	AQLQ Emotional	3 / 3 / 1 / 1	6.4 (5.05 – 6.8)	4.8 (3.6 – 5.6)	4.6 (3.1 – 5.6)	4.6 (3.2 – 5.6)	< 0.001 ²
	ESS Mean	4 / 1 / 3 / 3	0.63 (0.38 – 1.13)	0.88 (0.5 – 1.25)	0.88 (0.5 – 1.25)	0.88 (0.5 – 1.5)	0.0445 ²
	ACQ-5		0.8 (0.25 – 1.6)	2 (1.2 – 2.8)	2.4 (1.6 – 3.6)	2.6 (1.8 – 3.2)	< 0.001 ²
	ACQ-7		0.86 (0.43 – 1.86)	2.5 (1.68 – 3.14)	3.14 (2.11 – 4.04)	2.71 (2.0 – 3.86)	< 0.001 ²
	MARS Total	3 / 2 / 0 / 5	23 (21 – 24)	24 (22 – 24.8)	23 (20.8 – 24)	23 (22 – 25)	0.117 ²
	HADS Anxiety	12 / 11 / 17 / 23	4.0 (1.0 – 8.0)	6.0 (3.0 – 9.0)	6.0 (3.0 – 9.5)	7.0 (3.25 – 10.75)	0.0258 ²
	HADS Depression	12 / 11 / 17 / 23	2 (0 – 5)	5 (1 – 7)	5 (2 – 7.5)	5 (2 – 8.75)	0.00322 ²
	HADS Total	12 / 11 / 17 / 23	6 (3 – 13)	11 (5 – 18)	10 (5.5 – 17.5)	11.5 (6.25 – 18)	0.00778 ²
Biomarkers	Periostin assay	10 / 12 / 10 / 12	47.8 (38.9 – 53.5)	49.5 (40.7 – 62.6)	50.9 (44 – 56.8)	49.1 (40.4 – 57.2)	0.33 ²

□ Values are presented as mean $\pm$ SD, median (1st – 3rd quartiles) or percentages. NA corresponds to missing values.

¹ ANOVA

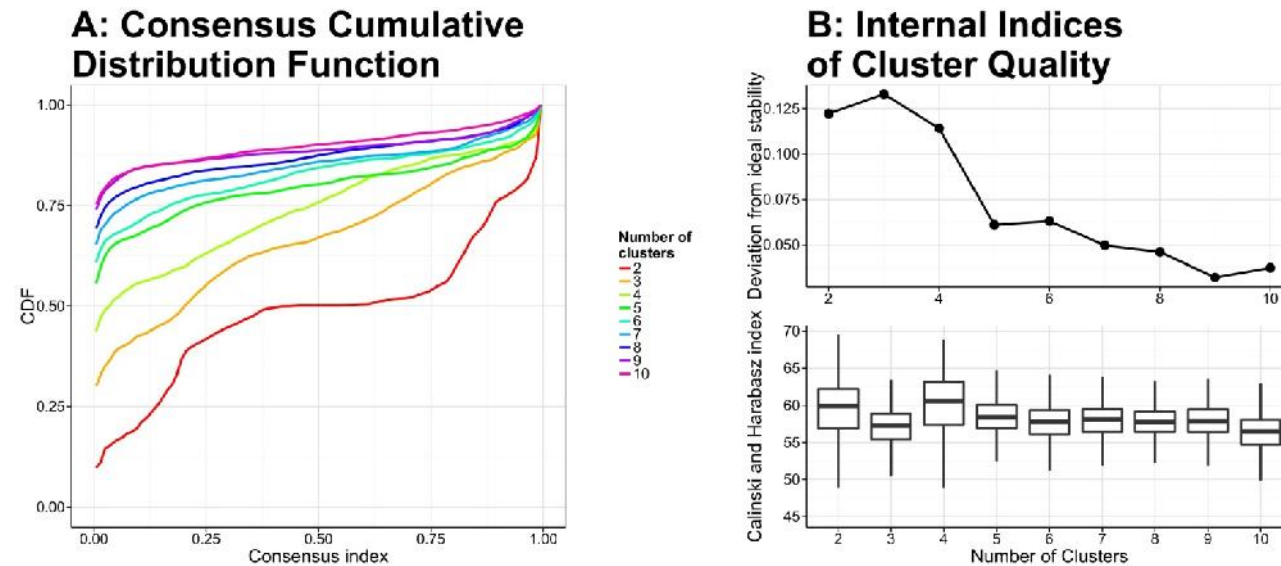
² Kruskal-Wallis test

³ χ^2 test

[†] Variables included in the clustering

ACQ: Asthma control questionnaire; AQLQ: Asthma quality of life questionnaire; BMI: Body Mass index; ESS: Epworth Sleep Score; FEF: Forced expiratory flow; FEV1: Forced expiratory volume in one second; FVC: Forced vital capacity; HADS: Health Anxiety and Depression score; ICS: Inhaled corticosteroid; LABA: long-acting β -agonist; MMA: Mild-moderate asthma; OCS: Oral corticosteroids; PEF: Peak expiratory flow; San: Severe non-smoking asthma; SAsm: severe smoking asthma; SGaw: specific airway conductance; TLC: Total lung capacity; WBC: White blood cell .

Figure E2: PAM clustering results of the validation set.



As for the training set, the algorithm was run 1,000 times removing randomly 10% of the data at each iteration in order to assess stability of the clusters. As it can be seen in panel A and B (top), 5 clusters are more stable (flat CDF and low Deviation from Ideal Stability) than 4 for the validation set even though the Calinski and Harabasz index in panel B (bottom) is slightly better for 4 clusters. Those clusters are however not as stable and well defined than the ones in the training set, potentially due to smaller number of participants in this set.

Table E3: Characteristics of the 5 clusters from the validation set. □

Variables	Missing or uncertain V1 / V2 / V3 / V4a / V4b	V1 (n = 39) similar to T1	V2 (n = 30) similar to T2	V3 (n = 38) similar to T3	V4a (n = 27) similar to T4	V4b (n = 18)	P-Value
Age		40.2 ± 15.6	57.7 ± 10.1	54.1 ± 10.8	55.7 ± 11.7	41.2.7 ± 13.1	< 0.001 ¹
Gender (female)		56.4%	56.7%	57.9%	44.4%	88.9%	0.057 ³
Total daily OCS dose (normalised to prednisolone) [†]		0 (0 – 0)	0 (0 – 0)	0 (0 – 5)	0 (0 – 7.5)	10 (5.25 – 18.8)	< 0.001 ²
Asthma onset (years) [†]		14 (9.5 – 28.5)	38.8 (29.2 – 48)	10 (3 – 25.8)	40 (23.5 – 48)	7.5 (2.25 – 18.8)	< 0.001 ²
FEV1% predicted [†]		95.4 ± 11.5	61.4 ± 12.9	48.9 ± 15	89.5 ± 14.4	75.2 ± 20.2	< 0.001 ¹
FEV1/FVC [†]		0.75 ± 0.0885	0.574 ± 0.0729	0.523 ± 0.109	0.71 ± 0.0626	0.8 ± 0.101	< 0.001 ¹
ACQ_5 [†]		1 (0.6 – 1.5)	2.2 (1.8 – 2.95)	2.4 (1.7 – 3.2)	1.2 (0.7 – 2.1)	3.2 (2.21 – 3.55)	< 0.001 ²
Exacerbations number past year [†]		0 (0 – 1)	2 (1 – 2.75)	2 (1.25 – 3)	2 (0 – 2.5)	3 (2.25 – 5.75)	< 0.001 ²
BMI [†]		24.3 (22.5 – 26.3)	27.9 (24.7 – 33)	25.2 (23.5 – 30.5)	29.1 (25.6 – 32.2)	33.9 (29.9 – 37.6)	< 0.001 ²
Pack Years [†]		0 (0 – 0)	16.2 (10 – 22.5)	0 (0 – 0)	1.4 (0 – 4.75)	0 (0 – 0)	< 0.001 ²
ICS dose range (high)	2 / 2 / 3 / 1 / 0	35.1%	96.4%	94.3%	92.3%	100.0%	< 0.001 ³
Blood eosinophils (x10 ³ /μl)	1 / 1 / 0 / 1 / 0	0.199 (0.103 – 0.3)	0.23 (0.1 – 0.42)	0.228 (0.193 – 0.4)	0.2 (0.107 – 0.43)	0.2 (0.0998 – 0.3)	0.620 ²
Blood neutrophils (x10 ³ /μl)	1 / 1 / 0 / 0 / 0	3.53 (2.9 – 4.18)	4.47 (3.73 – 5.94)	4.05 (3.18 – 6.67)	4.97 (3.62 – 5.87)	6.70 (4.75 – 8.38)	< 0.001 ²
Sputum eosinophils (%)	17 / 16 / 16 / 14 / 13	1.01 (0.192 – 2.17)	6.8 (1.76 – 13.4)	4.84 (1.4 – 36.7)	1.33 (0.19 – 8.01)	2.8 (1.52 – 3.68)	0.030 ²
Sputum neutrophils (%)	17 / 16 / 16 / 14 / 13	39.8 (20.2 – 55.9)	55.0 (46.2 – 59.3)	50.6 (26.4 – 82.71)	50.7 (34.2 – 73.8)	56.4 (39.9 – 64.4)	0.352 ²
FeNO	1 / 2 / 4 / 1 / 1	21 (15 – 59.7)	21.2 (11.8 – 38.7)	28 (17.1 – 51.5)	22.1 (16.2 – 40.2)	16 (12 – 27)	0.2247 ²
Total IgE (iu/ml)	2 / 1 / 0 / 0 / 0	104 (52.8 – 379)	106 (72.6 – 316)	95.5 (53.5 – 245)	136 (70.4 – 288)	133 (86.8 – 267)	0.922 ²
Atopy combined SPT and specific IgE results (positive)		89.7%	66.7%	81.6%	66.7%	72.2%	0.104 ³

□ Values are presented as mean $\pm$ sd or median (1st – 3rd quartiles)

¹ ANOVA

² Kruskal-Wallis test

³ χ^2 test

[†] Variables included in the clustering

Table E4: Comparison of the 4 clusters of the training set and the 5 clusters of the validation set [□].

Variables	P-Val T1-V1	P-Val T2-V2	P-Val T3-V3	P-Val T4-V4a&b [◇]
Cohort ^a	0.362	0.725	0.83	0.29
Age ^c	0.392	0.9	0.556	0.355
Gender ^a	1	1	0.386	0.0166
Total daily OCS (normalised to prednisolone) ^b	0.127	0.0171	0.0357	0.4
Asthma onset ^b	0.848	0.283	0.216	0.502
FEV1% predicted ^c	0.0238	0.456	0.876	0.145
FEV1/FVC ^c	0.469	0.406	0.344	0.761
ACQ-5 ^b	0.271	0.24	0.756	0.021
Exacerbation number (past year) ^b	0.218	0.827	0.568	0.0791
BMI ^b	0.242	0.741	0.913	0.801
Cigarette smoking (Pack Years) ^b	0.491	0.935	0.88	0.00178

[□] Significant p-values show statistical difference between the training and validation clusters.

[◇] The validation seem to have define a further separation in the 4th cluster from the training.

^a χ^2 -test

^b Kruskal-Wallis test

^c ANOVA

[†] Variables included in the clustering

Table E5: Comparison of clustered participants with and without sputum proteomics or transcriptomics [□]

Variables	Proteomics				Transcriptomics			
	Missing or uncertain	Without sputum (n=332)	With sputum (n=86)	P-Value	Missing or uncertain	Without sputum (n=324)	With sputum (n=94)	P-Value
Age		49.6 ± 14.8	50.4 ± 13.6	0.646 ¹		49.4 ± 14.8	51.3 ± 13.6	0.271 ¹
Gender (female)		60,8%	58,1%	0.608 ³		60,5%	59,6%	0.968 ³
Total daily OCS dose (normalised to prednisolone) [†]		0 (0 – 7.1)	0 (0 – 6.9)	0.982 ²		0 (0 – 5.25)	0 (0 – 7.5)	0.610 ²
Asthma onset (years) [†]		22.0 (7.75 – 39.3)	26.5 (11.0 – 43.0)	0.100 ²		22.5 (8 – 40)	23.0 (7.25 – 40)	0.581 ²
FEV1% predicted [†]		71.0 ± 22.8	70.8 ± 22.7	0.932 ¹		71.4 ± 22.9	69.4 ± 22.5	0.456 ¹
FEV1/FVC [†]		0.655 ± 0.139	0.618 ± 0.122	0.0246 ¹		0.659 ± 0.138	0.606 ± 0.125	9.25E-4 ¹
ACQ_5 [†]		2 (0 – 3)	2 (0 – 2.95)	0.942 ²		2 (1 – 3)	1.8 (1 – 2.8)	0.515 ²
Exacerbations number past year [†]		2 (0 – 3)	1 (0 – 2.75)	0.217 ²		2 (0 – 3)	2 (0 – 3)	0.823 ²
BMI [†]		27.2 (24.0 – 32.1)	26.6 (23.5 – 32.1)	0.485 ²		27.0 (23.9 – 32.4)	26.4 (23.9 – 30.8)	0.403 ²
Pack Years [†]		0 (0 – 1.4)	0 (0 – 12.7)	0.0057 ²		0 (0 – 2.03)	0 (0 – 3.88)	0.454 ²
ICS dose range (high)	22 / 0	81.6%	76.7%	0.392 ³	21 / 1	81.2%	78.5%	0.671 ³
Blood eosinophils (x10 ³ /μl)	8 / 1	0.200 (0.100 – 0.400)	0.243 (0.120 – 0.415)	0.254 ²	8 / 1	0.200 (0.100 – 0.400)	0.286 (0.148 – 0.400)	0.0505 ²
Blood neutrophils (x10 ³ /μl)	8 / 1	4.57 (3.47 – 6.30)	4.20 (3.28 – 6.51)	0.361 ²	19 / 1	4.55 (3.4 – 6.21)	4.42 (3.50 – 6.51)	0.992 ²
FeNO	18 / 3	25.0 (15.0 – 48.0)	27.5 (16.8 – 53.5)	0.576 ²	19 / 2	25.5 (15.0 – 49.0)	26.8 (16.9 – 50)	0.466 ²
Total IgE (iu/ml)	7 / 2	122 (52.0 – 318)	116 (50.5 – 320)	0.916 ²	7 / 2	126 (52.8 – 324)	105 (49.8 – 230)	0.349 ²
Atopy combined SPT and specific IgE results (positive)	1 / 2	73.7%	73.8%	1 ³	1 / 2	73.7%	73.9%	1 ³

[□] Values are presented as mean ± sd or median (1st – 3rd quartiles).

¹ χ^2 -test

² Kruskal-Wallis test

³ ANOVA

[†] Variables included in the clustering

Table E6: Consistently differentially expressed probesets for sputum transcriptomic in a subset.

PROBESETID	Gene Symbol*	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 2 (n = 19 vs 24)		Phenotype 4 vs 3 (n = 19 vs 20)	
		log2(FC) ³	P-value ^{1,2}	log2(FC) ³	P-value ^{1,2}	log2(FC) ³	P-value ^{1,2}	log2(FC) ³	P-value ^{1,2}	log2(FC) ³	P-value ^{1,2}
211839_PM_s_at	CSF1	+ 1.51	0.000115 (1)	+ 1.44	0.000304 (0.0696)						
224596_PM_at	SLC44A1	+ 0.77	0.0000657 (1)								
227620_PM_at	SLC44A1	+ 1.05	0.000502 (1)								
209921_PM_at	SLC7A11	+ 1.33	0.000504 (1)								
206749_PM_at	CD1B	+ 1.00	0.00144 (1)								
238193_PM_at	---	+ 1.07	0.00036 (1)								
244375_PM_at	---	+ 0.75	0.00039 (1)			- 0.88	0.000196 (0.0765)				
1562512_PM_at	SMPDL3A	- 0.62	0.00146 (1)	- 0.84	0.0000188 (0.0239)						
214433_PM_s_at	SELENBP1	- 0.24	0.000753 (1)								
207618_PM_s_at	BCS1L	- 0.55	0.000142 (1)								
1564474_PM_at	LOC728723	- 0.81	0.000252 (1)								
232691_PM_at	SFXN5			+ 0.60	0.00000102 (0.00790)	+ 0.62	0.00000493 (0.0378)				
214079_PM_at	DHRS2			+ 0.51	0.0000158 (0.0216)	+ 0.52	0.00000277 (0.0303)			- 0.50	0.0000532 (0.506)
230453_PM_s_at	ATP2A3			+ 1.34	0.000000142 (0.00245)					- 0.91	0.00198 (1)
213036_PM_x_at	ATP2A3			+ 1.66	0.000000179 (0.00245)					- 1.24	0.000266 (1)
226905_PM_at	FAM101B			+ 1.94	0.00000622 (0.0124)					- 1.62	0.000222 (1)
226876_PM_at	FAM101B			+ 1.49	0.0000252 (0.0264)						

207522_PM_s_at	ATP2A3		+ 1.59	0.000000828 (0.00245)			
207521_PM_s_at	ATP2A3		+ 1.48	0.0000165 (0.0220)			
229971_PM_at	GPR114		+ 0.76	0.00000121 (0.00790)			
208304_PM_at	CCR3		+ 1.33	0.00000145 (0.00793)			
242809_PM_at	IL1RL1		+ 3.22	0.00000376 (0.0108)			
213825_PM_at	OLIG2		+ 1.25	0.0000046 (0.0108)			
1554717_PM_a_at	PDE4D		+ 1.49	0.00000534 (0.0112)			
226794_PM_at	STXBP5		+ 0.87	0.00000657 (0.0124)			
235789_PM_at	KDM4B		+ 1.33	0.0000137 (0.0216)			
216379_PM_x_at	CD24		+ 1.66	0.0000158 (0.0216)			
238583_PM_at	MSRB3		+ 0.49	0.0000239 (0.0264)			
216370_PM_s_at	TKTL1		+ 0.84	0.0000285 (0.0274)			
236796_PM_at	BACH2		+ 0.82	0.0000395 (0.0322)			
1555486_PM_a_at	PRR5L		+ 1.10	0.0000465 (0.0352)			
211516_PM_at	IL5RA		+ 0.60	0.0000575 (0.0393)			
207902_PM_at	IL5RA		+ 0.60	0.000777 (0.0951)			
221341_PM_s_at	OR1D4		+ 0.46	0.0000766 (0.0411)			
220404_PM_at	GPR97		+ 1.47	0.0000777 (0.0443)			
1552349_PM_a_at	PRSS33		+ 1.23	0.0000707 (0.0431)			
1552782_PM_at	SLC44A5		+ 0.33	0.0000911 (0.0475)			
237921_PM_at	---		+ 0.33	0.000102 (0.0490)			
208145_PM_at	---		+ 0.39	0.0001030 (0.0490)			
1567101_PM_at	---		+ 0.71	0.0001130 (0.0499)			
1554615_PM_at	---		+ 1.11	0.00000475 (0.0108)			

237904_PM_at	---		+ 1.02	0.0000250 (0.0264)			
1569933_PM_at	---		+ 0.40	0.0000725 (0.00436)			
204947_PM_at	E2F1		+ 0.49	0.000444 (0.0834)			- 0.44 0.000639 (1)
203787_PM_at	SSBP2		+ 0.79	0.000178 (0.0583)			
227747_PM_at	MPZL3		+ 0.87	0.000202 (0.0627)			
228170_PM_at	OLIG1		+ 1.17	0.000303 (0.0696)			
1553176_PM_at	SH2D1B		+ 0.52	0.000228 (0.0647)			
216408_PM_at	OR2B2		+ 0.60	0.000229 (0.0647)			
222755_PM_s_at	CHD7		+ 0.48	0.000333 (0.0717)			
237342_PM_at	TOLLIP		+ 0.58	0.000233 (0.0647)			
1568627_PM_at	SMEK2		+ 0.48	0.000239 (0.0647)			
201858_PM_s_at	SRGN		+ 0.50	0.000356 (0.0738)			
205529_PM_s_at	RUNX1T1		+ 0.38	0.000360 (0.0738)			
204642_PM_at	S1PR1		+ 1.22	0.000414 (0.0805)			
206535_PM_at	SLC2A2		+ 0.41	0.000441 (0.0834)			
229332_PM_at	HPDL		+ 0.33	0.000584 (0.0910)			
214783_PM_s_at	ANXA11		+ 0.70	0.000690 (0.0935)			
229789_PM_at	TIGD3		+ 0.47	0.000610 (0.0925)			
1569775_PM_at	RNF157		+ 0.47	0.000719 (0.0946)			
244721_PM_at	TP53INP1		+ 0.67	0.000858 (0.0983)			
1553395_PM_a_at	CD200R1		+ 0.66	0.000861 (0.0983)			
210140_PM_at	CST7		+ 1.23	0.000882 (0.0997)			
235809_PM_at	LIN54		+ 0.47	0.000970 (0.104)			
225364_PM_at	STK4		+ 0.81	0.00138 (0.117)			

209606_PM_at	CYTIP		+ 0.79	0.00149 (0.121)			
41577_PM_at	PPP1R16B		+ 0.55	0.00779 (0.226)			
212736_PM_at	C16orf45		+ 0.29	0.000465 (0.0834)			
211351_PM_at	C6orf54		+ 0.52	0.000451 (0.0834)			
1558046_PM_x_at	LOC100132864		+ 0.58	0.0000673 (0.0431)			
244408_PM_at	---		+ 0.90	0.00393 (0.180)			
241584_PM_at	---		+ 0.57	0.000222 (0.0647)			
243189_PM_at	---		+ 0.47	0.000143 (0.0532)			
1556238_PM_at	---		+ 0.64	0.000119 (0.0513)			
241540_PM_at	---		+ 0.75	0.000166 (0.0578)			
241279_PM_at	---		+ 1.03	0.000167 (0.0578)			
1562342_PM_at	---		+ 0.81	0.000322 (0.0716)			
1566500_PM_at	---		+ 0.51	0.000244 (0.0647)			
240612_PM_at	---		+ 0.35	0.000698 (0.0935)			
237082_PM_at	---		+ 0.84	0.000701 (0.0935)			
229074_PM_at	EHD4		- 1.91	0.000000154 (0.00245)	- 1.47	0.0000302 (0.0580)	
204110_PM_at	F2R		- 1.51	0.00000361 (0.0108)	- 1.13	0.000215 (0.0784)	
238546_PM_at	SLC8A1		- 1.39	0.000000768 (0.00700)	- 0.95	0.000299 (0.0871)	
203139_PM_at	DAPK1		- 1.19	0.000027 (0.0274)	- 1.06	0.0000713 (0.0600)	
229905_PM_at	RAP1GDS1		- 1.25	0.0000367 (0.0314)	- 1.16	0.0000406 (0.0580)	
1555543_PM_a_at	CLCC1		- 0.55	0.0000617 (0.0412)	- 0.58	0.0000516 (0.0580)	
219944_PM_at	CLIP4		- 1.08	0.0000873 (0.0473)	- 1.03	0.0000542 (0.0580)	
228259_PM_s_at	EPB41L4A		- 1.98	0.000000237 (0.00259)	- 1.63	0.000000369 (0.0202)	
203148_PM_s_at	TRIM14		- 1.67	0.00000181 (0.00825)			

203147_PM_s_at	TRIM14		- 1.62	0.00000439 (0.0108)			+ 1.26	0.000212 (1)
204502_PM_at	SAMHD1		- 1.18	0.00000130 (0.00790)				
218632_PM_at	HECTD3		- 1.13	0.00000162 (0.00806)				
217760_PM_at	TRIM44		- 1.19	0.00000217 (0.0913)				
44673_PM_at	SIGLEC1		- 1.59	0.00000329 (0.0108)				
219519_PM_s_at	SIGLEC1		- 1.60	0.00000523 (0.0112)				
212796_PM_s_at	TBC1D2B		- 0.88	0.00000380 (0.0108)				
202944_PM_at	NAGA		- 1.00	0.00000428 (0.0108)				
217751_PM_at	GSTK1		- 1.00	0.00000451 (0.0108)				
1552302_PM_at	TMEM106A		- 1.26	0.00000452 (0.0108)				
228059_PM_x_at	MRPS22		- 0.77	0.00000648 (0.0124)				
223448_PM_x_at	MRPS22		- 0.61	0.0000981 (0.0488)				
229937_PM_x_at	LILRB1		- 1.03	0.0000144 (0.0216)				
201570_PM_at	SAMM50		- 1.63	0.0000183 (0.0238)				
200598_PM_s_at	HSP90B1		- 0.78	0.0000238 (0.0264)				
225440_PM_at	AGPAT3		- 0.93	0.0000278 (0.0274)				
212681_PM_at	EPB41L3		- 0.83	0.0000280 (0.0274)				
217752_PM_s_at	CNDP2		- 0.91	0.0000445 (0.0343)				
210644_PM_s_at	LAIR1		- 0.87	0.0000470 (0.0352)				
209555_PM_s_at	CD36		- 1.66	0.0000501 (0.0362)				
218802_PM_at	CCDC109B		- 0.92	0.0000565 (0.0391)				
226200_PM_at	VAR52		- 0.54	0.0000608 (0.0411)				
200617_PM_at	MLEC		- 0.77	0.0000632 (0.0417)				
41644_PM_at	SASH1		- 1.09	0.0000706 (0.0431)				

201647_PM_s_at	SCARB2		- 1.47	0.0000750 (0.0441)			
201646_PM_at	SCARB2		- 1.22	0.000676 (0.0935)			
208639_PM_x_at	PDIA6		- 0.75	0.0000765 (0.0441)			
221802_PM_s_at	KIAA1598		- 1.16	0.0000844 (0.0465)			
218856_PM_at	TNFRSF21		- 0.89	0.0000850 (0.0465)			
204972_PM_at	OAS2		- 1.35	0.0000893 (0.0474)			
206553_PM_at	OAS2		- 1.35	0.000497 (0.0858)			
1567627_PM_at	CD74		- 1.33	0.0000953 (0.0483)			
1567628_PM_at	CD74		- 1.42	0.000140 (0.0528)			
203547_PM_at	CD4		- 1.02	0.0000972 (0.0488)			
225710_PM_at	GNB4		- 0.94	0.0000996 (0.0490)			
208874_PM_x_at	PPP2R4		- 0.43	0.000101 (0.0490)			
210996_PM_s_at	YWHAЕ		- 0.59	0.000108 (0.0497)			
1552316_PM_a_at	GIMAP1		- 1.10	0.000111 (0.0498)			
1559883_PM_s_at	SAMHD1		- 1.09	0.000521 (0.0883)			
241742_PM_at	PRAM1		- 0.90	0.000444 (0.0834)	- 0.98	0.0000337 (0.0580)	
218400_PM_at	OAS3		- 1.33	0.000148 (0.0543)			+ 1.01 0.00348 (1)
222832_PM_s_at	MFF		- 0.70	0.000118 (0.0512)			
206335_PM_at	GALNS		- 0.51	0.000127 (0.0515)			
226436_PM_at	RASSF4		- 1.31	0.000127 (0.0515)			
49306_PM_at	RASSF4		- 0.91	0.000243 (0.0647)			
212192_PM_at	KCTD12		- 0.96	0.000132 (0.0520)			
209263_PM_x_at	TSPAN4		- 0.57	0.000132 (0.0520)			
208977_PM_x_at	TUBB4B		- 0.75	0.000135 (0.0524)			

224335_PM_s_at	BACE1		- 1.15	0.000135 (0.0524)			
222463_PM_s_at	BACE1		- 1.04	0.000484 (0.0850)			
218315_PM_s_at	CDK5RAP1		- 0.49	0.000136 (0.0524)			
200929_PM_at	TMED10		- 0.73	0.000144 (0.0532)			
222644_PM_s_at	GLT25D1		- 0.96	0.000144 (0.0532)			
219505_PM_at	CECR1		- 1.05	0.000152 (0.0551)			
208969_PM_at	NDUFA9		- 0.93	0.000163 (0.0578)			
202154_PM_x_at	TUBB3		- 0.78	0.000169 (0.0578)			
209684_PM_at	RIN2		- 1.18	0.000187 (0.0602)			
200785_PM_s_at	LRP1		- 1.04	0.000191 (0.0607)			
227798_PM_at	SMAD1		- 1.48	0.000205 (0.0627)			
241946_PM_at	ZDHHC21		- 1.12	0.000207 (0.0627)			
201740_PM_at	NDUFS3		- 0.80	0.000223 (0.0647)			
241937_PM_s_at	WDR4		- 1.00	0.000227 (0.0647)			
202838_PM_at	FUCA1		- 1.41	0.000231 (0.0647)			
216250_PM_s_at	LPXN		- 1.07	0.000235 (0.0647)			
203472_PM_s_at	SLCO2B1		- 1.07	0.000236 (0.0647)			
1555501_PM_s_at	RSRC1		- 1.19	0.000237 (0.0647)			
212953_PM_x_at	CALR		- 0.62	0.000245 (0.0647)			
209485_PM_s_at	OSBPL1A		- 1.01	0.000252 (0.0657)			
201900_PM_s_at	AKR1A1		- 1.02	0.000254 (0.0659)			
229175_PM_at	SMYD4		- 0.94	0.000267 (0.0667)			
201770_PM_at	SNRPA		- 0.79	0.000298 (0.0696)			
201326_PM_at	CCT6A		- 1.05	0.000302 (0.0696)			

221894_PM_at	ADCK2		- 0.47	0.000325 (0.0717)			
238422_PM_at	LOC151534		- 0.71	0.000332 (0.0717)			
1553906_PM_s_at	FGD2		- 1.01	0.000336 (0.0717)			
210427_PM_x_at	ANXA2		- 0.90	0.000351 (0.0730)			
238516_PM_at	BMPR2		- 1.12	0.000452 (0.0834)			
239314_PM_at	NHLRC3		- 1.20	0.000489 (0.0852)			
200986_PM_at	SERPING1		- 1.24	0.000496 (0.0858)			
208612_PM_at	PDIA3		- 0.69	0.000548 (0.0897)			
225286_PM_at	ARSD		- 0.61	0.000581 (0.0910)			
201143_PM_s_at	EIF2S1		- 0.97	0.000588 (0.0911)			
226039_PM_at	MGAT4A		- 0.93	0.000602 (0.0923)			
218204_PM_s_at	FYCO1		- 0.71	0.000607 (0.0925)			
236347_PM_at	MMAA		- 0.63	0.000686 (0.0935)			
213274_PM_s_at	CTSB		- 0.93	0.000714 (0.0946)			
224357_PM_s_at	MS4A4A		- 1.11	0.000725 (0.0946)			
242665_PM_at	FMNL2		- 1.19	0.000771 (0.0951)			
215049_PM_x_at	CD163		- 1.06	0.000933 (0.102)			
224632_PM_at	GPATCH4		- 0.82	0.000999 (0.105)			
203909_PM_at	SLC9A6		- 1.08	0.00100 (0.105)			
225002_PM_s_at	SUMF2		- 1.01	0.00115 (0.111)			
218543_PM_s_at	PARP12		- 0.75	0.00128 (0.113)			
201537_PM_s_at	DUSP3		- 0.80	0.00153 (0.122)			
226948_PM_at	RHBDD1		- 1.12	0.00268 (0.155)			
222613_PM_at	C12orf4		- 0.69	0.000626 (0.0929)			

203216_PM_s_at	MYO6		- 1.40 0.000204 (0.0627)				
203215_PM_s_at	MYO6			- 0.88	0.00492 (0.187)		
210650_PM_s_at	PCLO			+ 0.66	0.0000147 (0.0443)		
1558888_PM_x_at	ZNF321P			+ 0.80	0.0000154 (0.0443)		
1558887_PM_at	ZNF321P			+ 0.97	0.0000122 (0.0443)		
214294_PM_at	KIAA0485			+ 0.43	0.00000518 (0.0378)		
227174_PM_at	WDR72			+ 0.59	0.00000124 (0.0226)		
240589_PM_at	C21orf84			+ 0.49	0.0000151 (0.0443)		
229224_PM_x_at	LOC643085			+ 0.91	0.00000116 (0.0226)		
1560238_PM_at	---			+ 0.64	0.0000085 (0.0378)		
1562970_PM_at	---			+ 0.49	0.00000723 (0.0378)		
229024_PM_at	---			+ 0.51	0.000015 (0.0443)		
216130_PM_at	---			+ 0.58	0.0000174 (0.0445)		
1556518_PM_at	---			+ 0.52	0.00000785 (0.0378)		
205609_PM_at	ANGPT1			+ 0.53	0.000118 (0.0643)		
215558_PM_at	UBR2			+ 0.59	0.000779 (0.112)		
1552808_PM_at	ADAMTSL1			+ 0.30	0.0000881 (0.0612)		
220535_PM_at	FAM90A1			+ 0.44	0.0000994 (0.0614)		
242774_PM_at	SYNE2			+ 0.87	0.00121 (0.129)		
220402_PM_at	TP53AIP1			+ 0.70	0.0000486 (0.0580)		
214688_PM_at	TLE4			+ 0.60	0.000535 (0.102)		
1560773_PM_at	DLGAP1			+ 0.44	0.0000315 (0.0580)		
205654_PM_at	C4BPA			+ 0.49	0.0000494 (0.0580)		
211480_PM_s_at	SLCO1A2			+ 0.49	0.000224 (0.0808)		

229953_PM_x_at	LCA5			+ 0.41	0.000247 (0.0829)		
230339_PM_at	CCDC138			+ 0.48	0.0000762 (0.0605)		
206211_PM_at	SELE			+ 0.33	0.0000996 (0.0614)		
219843_PM_at	IPP			+ 0.36	0.00101 (0.123)		
237425_PM_at	LOC474358			+ 0.53	0.0000382 (0.0580)		
206953_PM_s_at	LPHN2			+ 0.43	0.00133 (0.131)		
217483_PM_at	FOLH1			+ 0.56	0.0000516 (0.0580)		
230271_PM_at	ONECUT2			+ 0.45	0.000367 (0.0920)		
1552715_PM_a_at	RXFP1			+ 0.54	0.0000761 (0.0605)		
204941_PM_s_at	ALDH3B2			+ 0.53	0.0000564 (0.0580)		
1554914_PM_at	PLA2G4D			+ 0.61	0.000043 (0.0580)		
1553620_PM_at	TRIM42			+ 0.37	0.000146 (0.0683)		
204624_PM_at	ATP7B			+ 0.41	0.0000928 (0.0612)		
208019_PM_at	ZNF157			+ 0.42	0.00025 (0.0834)		
1553277_PM_at	RTTN			+ 0.60	0.000569 (0.103)		
220503_PM_at	SLC13A1			+ 0.40	0.000301 (0.0871)		
1553674_PM_at	LRRIQ3			+ 0.48	0.0000778 (0.0608)		
1554679_PM_a_at	LAPTM4B			+ 0.53	0.0014 (0.132)		
1563014_PM_at	RPS15			+ 0.49	0.000048 (0.0580)		
223030_PM_at	TRAF7			+ 0.29	0.00048 (0.0989)		
244297_PM_at	ANKRD18A			+ 0.54	0.0000881 (0.0612)		
203862_PM_s_at	ACTN2			+ 0.46	0.000283 (0.0866)		
1570039_PM_at	TMSB15B			+ 0.58	0.00028 (0.0866)		
215074_PM_at	MYO1B			+ 0.60	0.000393 (0.0931)		

234003_PM_at	ENOX2			+ 0.33	0.000381 (0.0925)		
239949_PM_at	THNSL2			+ 0.55	0.00121 (0.129)		
220356_PM_at	CORIN			+ 0.42	0.000329 (0.0900)		
1553350_PM_at	AMER3			+ 0.40	0.0000311 (0.0580)		
1553817_PM_at	POM121L10P			+ 0.47	0.000653 (0.107)		
235488_PM_at	RASL10B			+ 0.45	0.000378 (0.0925)		
1568924_PM_a_at	IQUB			+ 0.53	0.0000973 (0.0614)		
230964_PM_at	FREM2			+ 0.46	0.000171 (0.0724)		
226469_PM_s_at	GGT7			+ 0.36	0.000614 (0.105)		
1552564_PM_at	NUDT9P1			+ 0.41	0.000506 (0.101)		
1552774_PM_a_at	SLC25A27			+ 0.48	0.00139 (0.131)		
218353_PM_at	RGS5			+ 0.34	0.000272 (0.0865)		
235355_PM_at	CSRNP3			+ 0.44	0.00121 (0.129)		
220982_PM_s_at	SPACA1			+ 0.39	0.00120 (0.129)		
235219_PM_at	C5orf55			+ 0.44	0.000327 (0.0900)		
237363_PM_at	C9orf68			+ 0.52	0.0000789 (0.0608)		
238047_PM_at	ARHGAP36			+ 0.39	0.000154 (0.0701)		
1570451_PM_at	C1orf61			+ 0.45	0.000238 (0.0811)		
1560537_PM_at	LOC100129662			+ 0.42	0.00135 (0.131)		
1558844_PM_at	LOC387790			+ 0.43	0.000186 (0.0743)		
230743_PM_at	LOC404266			+ 0.43	0.000239 (0.0811)		
240472_PM_at	---			+ 0.36	0.000404 (0.0932)		
244724_PM_at	---			+ 0.68	0.0000237 (0.0540)		
243655_PM_x_at	---			+ 0.50	0.000761 (0.111)		

1561226_PM_at	---			+ 0.50	0.00014 (0.0677)		
1569193_PM_at	---			+ 0.37	0.000103 (0.0619)		
1557276_PM_at	---			+ 0.51	0.000273 (0.0865)		
1570054_PM_at	---			+ 0.43	0.000388 (0.0925)		
244073_PM_at	---			+ 0.42	0.000804 (0.115)		
1562737_PM_at	---			+ 0.48	0.000348 (0.0909)		
1566622_PM_at	---			+ 0.49	0.00105 (0.124)		
237633_PM_at	---			+ 0.39	0.000401 (0.932)		
231233_PM_at	---			+ 0.53	0.000284 (0.866)		
234641_PM_at	---			+ 0.66	0.00131 (0.131)		
239767_PM_at	---			+ 0.47	0.000408 (0.932)		
1570098_PM_at	---			+ 0.46	0.00148 (0.134)		
1570124_PM_at	---			+ 0.35	0.00174 (0.143)		
237355_PM_at	---			+ 0.39	0.000621 (0.106)		
243729_PM_at	---			+ 0.52	0.0000369 (0.0580)		
244794_PM_at	---			+ 0.46	0.000874 (0.117)		
243942_PM_at	---			+ 0.37	0.000197 (0.0765)		
236943_PM_at	---			+ 0.58	0.0000362 (0.0580)		
1563557_PM_at	---			+ 0.50	0.000039 (0.0580)		
231168_PM_at	---			+ 0.46	0.0000763 (0.0605)		
1555963_PM_x_at	B3GNT7			- 1.05	0.0000154 (0.0443)		
201809_PM_s_at	ENG			- 0.79	0.00000228 (0.0303)		
202556_PM_s_at	MCRS1			- 0.68	0.00000677 (0.0378)		
204765_PM_at	ARHGEF5			- 0.67	0.0000205 (0.0488)		

220570_PM_at	RETN			- 1.02	0.0000172 (0.0445)		
235598_PM_at	SMAD2			- 0.42	0.00256 (0.158)		
221853_PM_s_at	NOMO1			- 0.84	0.0000688 (0.0600)		
217225_PM_x_at	NOMO1			- 0.91	0.00017 (0.0724)		
238959_PM_at	LARP4			- 1.24	0.000054 (0.0580)		
224700_PM_at	STT3B			- 0.85	0.0000589 (0.0580)		
204146_PM_at	RAD51AP1			- 1.25	0.000548 (0.102)		
225602_PM_at	GLIPR2			- 0.48	0.000496 (0.100)		
214190_PM_x_at	GGA2			- 0.93	0.000091 (0.0612)		
1559361_PM_at	MACC1			- 1.34	0.0000623 (0.0580)		
225163_PM_at	FRMD4A			- 1.26	0.0000709 (0.0600)		
213221_PM_s_at	SIK2			- 1.09	0.00161 (0.139)		
1556175_PM_at	MTSS1L			- 0.94	0.0000631 (0.0580)		
209796_PM_s_at	CNPY2			- 1.23	0.000185 (0.0743)		
213646_PM_x_at	TUBA1B			- 0.60	0.00155 (0.136)		
211945_PM_s_at	ITGB1			- 0.80	0.000288 (0.0866)		
203704_PM_s_at	RREB1			- 0.96	0.000180 (0.0743)		
204767_PM_s_at	FEN1			- 0.63	0.00144 (0.133)		
229632_PM_s_at	INTS10			- 0.54	0.00119 (0.129)		
243_PM_g_at	MAP4			- 1.26	0.000119 (0.643)		
217892_PM_s_at	LIMA1			- 1.45	0.000388 (0.0925)		
242871_PM_at	PAQR5			- 1.60	0.000848 (0.116)		
213375_PM_s_at	N4BP2L1			- 0.91	0.000047 (0.0580)		
221619_PM_s_at	MTCH1			- 0.53	0.000274 (0.0865)		

203179_PM_at	GALT			- 0.57	0.00284 (0.163)		
221900_PM_at	COL8A2			- 0.82	0.0000843 (0.0612)		
205816_PM_at	ITGB8			- 1.51	0.000208 (0.0769)		
242982_PM_x_at	ITGB8			- 0.90	0.00123 (0.129)		
214550_PM_s_at	TNPO3			- 0.44	0.000677 (0.107)		
1558199_PM_at	FN1			- 1.05	0.00105 (0.124)		
205726_PM_at	DIAPH2			- 0.86	0.000172 (0.0724)		
206621_PM_s_at	EIF4H			- 0.54	0.0000999 (0.0614)		
209609_PM_s_at	MRPL9			- 0.95	0.0019 (0.146)		
209764_PM_at	MGAT3			- 0.59	0.00216 (0.151)		
230806_PM_s_at	FAM65A			- 1.00	0.00176 (0.143)		
205012_PM_s_at	HAGH			- 0.67	0.0000952 (0.0614)		
218482_PM_at	ENY2			- 0.50	0.00287 (0.163)		
201778_PM_s_at	EFCAB14			- 0.63	0.00297 (0.164)		
1554503_PM_a_at	OSCAR			- 0.84	0.000413 (0.0934)		
228559_PM_at	CENPN			- 1.07	0.000924 (0.119)		
34868_PM_at	SMG5			- 0.43	0.000847 (0.116)		
209233_PM_at	EMG1			- 0.70	0.000205 (0.0768)		
223277_PM_at	C3orf75			- 0.43	0.000132 (0.0673)		
224727_PM_at	C19orf63			- 0.77	0.000420 (0.0940)		
225887_PM_at	C13orf23			- 0.96	0.000134 (0.0673)		
203197_PM_s_at	C1orf123			- 0.62	0.000959 (0.121)		
227576_PM_at	---			- 1.18	0.0000322 (0.0580)		
228586_PM_at	---			- 0.88	0.0000322 (0.0580)		

242951_PM_at	---			- 0.55	0.000488 (0.0995)		
1564077_PM_at	---			- 1.09	0.000120 (0.0643)		
242299_PM_at	---			- 0.99	0.000498 (0.100)		
1565913_PM_at	---					+ 1.10	0.00686 (1)
236939_PM_at	---					+ 0.65	0.000416 (1)
225832_PM_s_at	DAGLB					- 0.62	0.00216 (1)
201539_PM_s_at	FHL1						+ 0.74 0.000901 (1)
210298_PM_x_at	FHL1						+ 0.88 0.000307 (1)
211429_PM_s_at	SERPINA1						+ 0.40 0.0000301 (0.506)
205660_PM_at	OASL						+ 0.99 0.000578 (1)
212271_PM_at	MAPK1						+ 0.58 0.00117 (1)
226584_PM_s_at	FAM110A						- 0.58 0.00187 (1)
1559127_PM_x_at	RRP12						- 0.49 0.000132 (1)
218122_PM_s_at	SENP2						- 0.56 0.000847 (1)
204011_PM_at	SPRY2						- 0.94 0.0000201 (0.506)

* Annotations are given with the Affymetrix[®] annotation file HT-HGU133+PM v31.

¹ Tukey p-value adjusted for age and gender in the training and validation set pooled together of the transcripts that were consistently found both in the training and the validation for that specific contrast.

² P-values (FDR p-values). Bold value indicates that the difference is still significant even when correcting for false discovery rate using Benjamini-Hochberg method.

³ Fold-change is given in base 2 logarithm, if it is positive in X vs Y it means that the transcript is more expressed in X than Y, on the contrary negative values mean less expressed.

