A Transcriptome-driven Analysis of Epithelial Brushings and Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED

DOI:
10.1164/rccm.201512-2452OC

Document Version
Accepted author manuscript

A transcriptome-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in U-BIOPRED

Chih-Hsi Scott Kuo[1,3], Stelios Pavlidis[2], Matthew Loza[2], Fred Baribaud[2], Anthony Rowe[2], Ioannis Pandis[1], Uruj Hoda[3], Christos Rossios[3], Ana Sousa[4], Susan J. Wilson[5], Peter Howarth[5], Barbro Dahlen[6], Sven-Erik Dahlen[6], Pascal Chanez[7], Dominick Shaw[8], Norbert Krug[9], Thomas Sandström[10], Bertrand De Meulder[11], Diane Lefaudeux[11], Stephen Fowler[12], Louise Fleming[3], Julie Corfield[13], Charles Auffray[11], Peter J. Sterk[14], Ratko Djukanovic[5], Yike Guo[1], Ian M. Adcock[3*] & Kian Fan Chung[3*] on behalf of the U-BIOPRED Project Team[#]

**Contributed equally,*

[#]Other Consortium Project Team members in Acknowledgements

[1]Department of Computing & Data Science Institute, Imperial College London, United Kingdom; [2]Janssen Research and Development, High Wycombe, Buckinghamshire, United Kingdom; [3]Airways Disease, National Heart & Lung Institute, Imperial College London, United Kingdom & Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London, UK; [4]Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park, UK; [5]Faculty of Medicine, Southampton University, Southampton, United Kingdom; [6] Centre for Allergy Research, Karolinska Institute, Stockholm, Sweden, [7]Université de la Méditerranee, Marseille, France, [8] Centre for Respiratory Research, University of Nottingham, UK, [9] Fraunhofer Institute of Toxicology & Experimental Medicine, Hannover, Germany, [10] Dept of Public Health and Clinical Medicine, Medicine, Umeå University, Umeå, Sweden, [11]European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL, Université de Lyon, France, [12]Centre for Respiratory Medicine and Allergy, The University of Manchester, UK; [13]AstraZeneca R&D Molndal, Sweden and Areteva R&D, Nottingham, UK; [14]Faculty of Medicine, University of Amsterdam, Amsterdam, Netherlands.

Word count: Abstract-245; Text- 3920

Figs: 7 **Tables:** 5

On line Supplementary Method, Table and Figure

Running title: Transcriptome-driven asthma phenotypes

Corresponding author

Professor K F Chung, National Heart & Lung Institute, Imperial College London,

Dovehouse St, London SW3 6LY, UK

f.chung@imperial.ac.uk

Author Contributions:

CHSK, DL, BDM and SP performed the analysis. CHSK, SP, ML, FB, AR, CA, IP and YG designed the analytical approaches taken and analysed the results. RD, PJS, PH, UH, SD, PC, DS, BD, NK, TS and KFC participated in the clinical characterisation of the patients. SJW and PH participated in the immunohistochemical studies. CR participated in experimental work and analysis. SF, LF, AS and JC were part of the data curation team. IMA, RD, PJS and KFC conceived & designed the study, and CHSK, IMA & KFC coordinated the data and drafted the manuscript. All have read the final version.

At a glance commentary

Scientific knowledge on the subject

Unbiased clustering of asthma using clinical features has derived phenotypes of limited pathobiological relevance. It is not known whether semi-biased clustering on transcriptomic data will provide additional information on underlying asthma biology.

What the study adds to the field

We show that semi-biased clustering of transcriptomic profiles from asthmatic airway samples produces biologically relevant clusters that have implications for directed or personalized therapy. We report subgroups of eosinophilic asthma driven by T-helper Type 2 cells and corticosteroid insensitivity signatures that featured either high submucosal eosinophils or sputum eosinophils. An inference scheme using currently-available inflammatory biomarkers can be used to predict the subtypes of gene expression.

Abstract (245 words)

Rationale and objectives: Asthma is a heterogeneous disease driven by diverse immunologic and inflammatory mechanisms. We used transcriptomic profiling of airway tissues to help define asthma phenotypes.

Methods: The transcriptome from bronchial biopsies and epithelial brushings of 107 moderate-to-severe asthmatics were annotated by gene-set variation analysis (GSVA) using 42 gene-signatures relevant to asthma, inflammation and immune function. Topological data analysis (TDA) of clinical and histological data was used to derive clusters and the nearest shrunken centroid algorithm used for signature refinement.

Results: 9 GSVA signatures expressed in bronchial biopsies and airway epithelial brushings distinguished two distinct asthma subtypes associated with high expression of T-helper type 2 (Th-2) cytokines and lack of corticosteroid response (Group 1 and Group 3). Group 1 had the highest submucosal eosinophils, high exhaled nitric oxide (FeNO) levels, exacerbation rates and oral corticosteroid (OCS) use whilst Group 3 patients showed the highest levels of sputum eosinophils and had a high BMI. In contrast, Group 2 and Group 4 patients had an 86% and 64% probability of having non-eosinophilic inflammation. Using machine-learning tools, we describe an inference scheme using the currently-available inflammatory biomarkers sputum eosinophilia and exhaled nitric oxide levels along with OCS use that could predict the subtypes of gene expression within bronchial biopsies and epithelial cells with good sensitivity and specificity.

Conclusion: This analysis demonstrates the usefulness of a transcriptomic-driven approach to phenotyping that segments patients who may benefit the most from specific agents that target Th2-mediated inflammation and/or corticosteroid insensitivity.

Key words: severe asthma, bronchial biopsies, bronchial brushings, corticosteroid insensitivity, T-helper Type 2(Th2), exhaled nitric oxide, gene-set variation analysis, asthma phenotype

Introduction

Asthma is a heterogeneous disease that presents with varying degrees of inflammation in the airways in association with airway wall structural changes. Both infiltrating and resident airway structural cells participate in the inflammatory and remodeling processes of asthma and various mechanisms have been proposed that could underlie the asthmatic inflammatory process (1). T-helper Type 2 (Th2) cell activation characterized by the secretion of IL-4, IL-5 and IL-13 is a key mechanism of allergic asthma with these cytokines being overexpressed in the bronchial submucosa of asthmatic patients (2, 3). Th2-high asthmatic patients are characterized by a greater degree of bronchial hyperresponsiveness, higher serum IgE levels, greater blood and airway eosinophilia, subepithelial fibrosis and airway mucin gene expression (4).

Th1 and Th17 cells, and their products are also important in asthma. Th1 cells are characterised by IFNγ production which plays a role in dealing with viral infections and in autoimmunity and both Th1 cells and IFN γ are overexpressed in asthma subjects (5-7). These have also been shown to be important in severe asthma (8, 9). Th-17 cells express IL-17A, IL-17E, IL-17F and IL-22, and may mediate corticosteroid-resistant airway inflammation and airway hyperresponsiveness in mice (10). IL-17A and IL-17F, have been localised in the airways of patients with severe asthma (11).

Oxidative stress is another important feature of asthma, particularly those with severe asthma, and is associated with increased levels of reactive oxygen species (ROS) and compromised antioxidant responses (12, 13). ROS may mediate various features of severe asthma such as the chronic inflammatory response, the hypercontractility of the airways and corticosteroid insensitivity (14-16). Finally, structural abnormalities can alter airway mechanics and contribute to bronchial hyperresponsiveness, while structural cells can also contribute to inflammatory processes through release of cytokines, chemokines, growth factors and extracellular matrix elements (17-19).

Therefore, asthma is a syndrome with many potential endotypes and defining these is an essential step towards providing personalized treatments (20, 21). Defining asthma phenotypes from clinical and physiological parameters, including currently-used inflammatory biomarkers, is limited in expressing the heterogeneity of underlying mechanisms (20, 22, 23). This may not be surprising considering the complexity of immunoinflammatory mechanisms and the diversity of cell types implicated in asthma (24-27). Deriving asthma clusters according to differentially-expressed genes may be a better approach to capture the diverse pathways of asthma pathobiology that could lead to the identification of important targets for therapy and offer insights towards achieving personalized medicine (20, 28, 29). This approach has been used to identify genes from sputum cells that linked pathway-based transcriptomic clusters to clinically-important features of asthma (30).

Conventional gene set enrichment (GSE) methods have limitations owing to the assumption of two classes (e.g. case versus control) (31). Gene Set Variation Analysis (GSVA) addresses this limitation by calculating GSE scores as a function of genes inside and outside a specifically-defined gene set (32) and can indicate functional activity across individual samples (33). Topological data analysis (TDA) that recognises the invariant shape of complex data sets (34, 35) has been used to overcome the challenges that stochastic gene expression presents on conventional hierarchical or k-means clustering (36).

In the current study, we analysed transcriptomic data from bronchial biopsies from asthma participants from the Unbiased BIOmarkers in PREDiction of respiratory disease outcomes (U-BIOPRED) project (37) and performed supervised clustering using 42 gene sets associated with asthma and immune/inflammatory pathways by GSVA. We then explored the relationship between these GSVA signatures and their unique clinico-histopathological features. This was then further investigated in the transcriptome from epithelial brushings of the same participants with the aim of examining whether transcriptomic signatures may help define subsets of Th2-high eosinophilic and non-Th2 asthma from distinct airway

compartments. Some of the results of these studies have been previously reported in the form of an abstract (38).

Materials and methods (500 words)

Details of analytical methods are provided in the *On-line Supplement*.

Clinical data

107 participants (**Table 1**) with moderate-to-severe asthma from the U-BIOPRED cohort underwent fiberoptic bronchoscopy for epithelial cell brushings and bronchial biopsies(39). Pre-bronchodilator spirometry, exhaled nitric oxide (FeNO), sputum differential cell count, skin prick tests, serum total IgE, and differential blood count were measured(39). Oral corticosteroid dependence was assessed by determining the lowest maintenance dose of daily prednisolone. The study was approved by the Ethics Committees for each of the 16 clinical recruiting centres. All participants gave written and signed informed consent.

Immunohistochemistry for submucosal cells in the bronchial biopsies

Samples from glycolmethacrylate embedded sections (2μm) were stained with monoclonal antibodies against CD3, CD4, CD8, neutrophil elastase and EG2. Cell counts were performed in a blinded fashion and expressed as positive cells/mm^2. These results have been presented previously in an abstract form [43]**.**

Microarray analysis of mRNA

Expression profiling was performed using AffymetrixU133 Plus 2.0 microarray (Affymetrix, Santa Clara, Calif) on total RNA extracted from bronchial biopsies and epithelial brushings. RNA purity (RIN >9.5) was measured by Agilent Bioanalyser (Agilent, Santa Clara, Calif). Raw data were quality assessed and pre-processed by robust multi-array average normalization using Almac Pipeline and Pre-processing Toolbox (Almac, Craigavon, United Kingdom). Probes of low expression were filtered according to robust multi-array signal values <5 and batch/technical effects were adjusted as covariates using a linear model.

Gene Set Variation Analysis (GSVA)

42 gene sets containing 2431 genes were manually curated: 6 for six key immune cell type-specific gene expression associated with their differentiation and activation, 6 for evolutionarily-conserved transcriptional signatures for Th1 and CD8 memory T cell differentiation, 5 of effector CD4+ T cell differentiation for Th2, Th17 and Treg cell subset, 6 for chronic effect of oxidative stress in response to ozone from a mouse model and genes associated with COPD from human lung, 7 from peripheral blood mononuclear cells (PBMCs) in four autoimmune diseases, 7 for asthma-specific mechanism driven by T cells and 5 from human lung biopsy, airway smooth muscle cells and PBMCs in response to corticosteroid treatment (Supplementary TableS1).

TDA and machine learning algorithm

TDA was used to explore clusters driven by gene signatures utilizing the Ayasdi Core™ (Ayasdi, Inc. California, USA) program and the nearest shrunken centroid algorithm was used to further define the optimal number of classifiers. Details of the methods are provided in Supplementary file.

Statistical analysis

All datasets needed for this analysis were uploaded from the tranSMART system, an open-source knowledge management platform for sharing research data(40) supported by the European Translational Information and Knowledge Management Services (eTRIKS) project. All categorical variables were analyzed using Fisher's exact test. Student's t test was used for continuous variables with normal distribution, otherwise Wilcoxon rank-sum test was used. A linear model for microarray data (Bioconductor R package limma) with false discovery rate (FDR) correction was used for differential expression gene analysis, and recursive partitioning (R package PARTY) for decision tree learning. An FDR<0.05 was considered

statistically significant.

Results

Transcriptomic profiles from 91 bronchial biopsies and 99 epithelial brushings were obtained, of which 83 samples were paired from the same subject.

Differentially-expressed gene and gene set analysis

We performed differential expression analysis of individual genes and of gene sets comparing moderate asthma versus severe asthma in both epithelial brushings and bronchial biopsies. None of the individual genes or gene sets were significantly different between epithelial brushings from severe and non-severe asthma after multiple testing correction for FDR (**Fig 1A & B**). In addition, only 2 genes with marginally significant expression and three GSVA gene sets with a trend toward significant expression were found when comparing expression profiles in bronchial biopsies between severe and non-severe patients (**Fig 1C & D**). Therefore, the clinical classification of moderate versus severe asthma provides a limited framework for the identification of differential gene expression profiles.

There was a significant correlation between the ES of the Th2 signature and the relative steroid insensitivity signature in both biopsy and brushings (Fig S1), together with a significant relationship between the oxidative stress signature and T-cell, Th1 and Th2 signatures (Fig S2), and between the oxidative stress signatures and relative corticosteroid insensitivity signatures in both biopsy and brushings (Fig S3). These correlations indicate potential mechanistic interactions between Th2 and Th1 genes with oxidative stress and relative corticosteroid insensitivity which was defined by the signature of genes that are usually down-regulated by corticosteroids showing no change or being up-regulated.

In both biopsy and brushings, *CD44* as a constituent of corticosteroid insensitivity signature had the most extensive association/interaction with a subset of Th2 signature genes including *CCL26, IL1R2* and *CST2*. In addition, a highly correlated association between

NELFE (within the corticosteroid insensitivity signature) and *ATP5J* (within the Th2 signature) was also seen in both biopsy and brushing compartments.

Clinical clusters driven by bronchial biopsy GSVA analysis

We performed an analysis of the bronchial biopsy transcriptome using GSVA in order to investigate the immune cell-specific characteristics and distinctive pathophysiological processes in asthma. Unsupervised clustering of GSVA gene sets applied to biopsies using TDA identified a cluster characterized by high submucosal eosinophils (5.6 vs. 1.2/mm^2, p=6.6 x 10^{-5}), high FeNO (34.0 vs. 21.0 ppb, p=0.028) and high oral corticosteroid use (41.0 vs 15.4%, p=0.006) (**Fig 2 & Table 2**). Comparing Cluster A to Cluster non-A, 26 out of 42 gene sets were found to be differentially expressed (p<0.05) (**Fig 3**). A similar clustering was obtained when hierarchical clustering was performed using hierarchical clustering (**Fig 4**).

Finding signatures for optimal cluster classification

In order to determine the GSVA signatures that best distinguished Cluster A from Cluster non-A, the shrunken centroid algorithm was applied to these 26 GSVA gene-sets and produced a classification accuracy of 82.4% (Fig 5A). After feature reduction, 9 GSVA gene-sets of non-zero value which were expressed in the opposite direction in Cluster A and non-A were retained as centroid classifiers of discrimination (Fig 5B, Table 3). ROC curve analysis demonstrated a high performance of this 9-gene-set signature model (AUC: 0.866; 95%CI: 0.796-0.927, p=1.3x10-9) in differentiating membership of Cluster A, providing a sensitivity and specificity of 84.6% and 90.4%, respectively (Fig 5C).

Clusters from epithelial brushings and from combined epithelial brushings and bronchial biopsies

In order to test whether these 9 GSVA signatures were also able to discriminate clusters

in bronchial brushings, a similar analysis was performed on the 99 epithelial brushing samples. The 9 GSVA signatures from Cluster A determined in bronchial biopsies was found in 31 (31.3%) of subjects as calculated using the nearest shrunken centroids of the signatures. This subgroup reproduced the clinical characteristics noted in biopsies in terms of significantly higher submucosal eosinophils (5.8 vs. 1.5/mm^2, p=1.3x10^{-5}), higher exhaled NO (51.5 vs. 21.0 ppb, p=4.2x10^{-4}) and higher oral corticosteroid use (40.7 vs 17.9%, p=0.032; **Table 4**). In addition, higher eosinophil counts in blood and sputum and higher numbers of submucosal CD3, CD4 and CD8 T cells in bronchial submucosa were also noted.

The relative expression of these 9 GSVA signatures in bronchial biopsies and epithelial brushings is shown in **Table 5**: Group 1 (19 subjects, 23%) expressed these 9 signatures in both bronchial samples and epithelial brushing cells, Group 2 (17 patients, 20%) expressed these signatures in bronchial biopsies only, Group 3 (8 patients, 10%) in epithelial brushing cells alone and Group 4 (39 patients, 47%) expressed these 9 GSVA signatures in neither compartment.

Compared to Group 4 participants, two subtypes of patients with eosinophilic inflammation and relative steroid insensitivity were noted (Groups 1 and 3; **Table 6**). Group 1 had the highest submucosal eosinophils (6.1 vs. 1.0/mm^2, p=7.64x10^{-6}), CD3 (45.3 vs. 33.3/mm^2, p=0.048), CD8 (26.4 vs. 13.9/mm^2, p=0.020) T cells, FeNO (56.5 vs. 19.0 ppb, p=0.001), acute exacerbation rate (3.0 vs. 1.5 times/year, p=0.022), oral corticosteroid use (47.7 vs 12.8%, p=0.010) and intermediate-high sputum eosinophils (8.2 vs 0.6%, p=0.017). In contrast, Group 3 participants possessed the highest levels of sputum eosinophils (15.8 vs. 0.6%, p=0.014), high BMI (33.1 vs. 28.0, p=0.010) and a trend toward intermediate-high FeNO (35.5 vs. 19.0 ppb, p=0.058).

Levels of gene expression of IL-33 and TSLP but not IL-25 were relatively increased in Cluster A compared to Cluster non-A and in Group 1 compared to Groups 3 and 4 (for IL-33) and to Group 4 (for TSLP) in bronchial biopsies but not in epithelial brushings (Fig S4).

Inference of phenotype from non-invasive clinical measurements

The clinical features and non-invasive biomarkers from the patients classified into the 4 groups were used to build an inference tree framework. FeNO >55 ppb (Fig 6 Node 7, 58% for Group 1), sputum eosinophil >4.5% (Fig 6 Node 6, 42% for Group 1) or oral corticosteroid dependency (Fig 6 Node 4, 36% for Group 1) predicted subjects as the highest eosinophilic inflammation and steroid insensitivity subtype (Group 1), with an 84% sensitivity and 72% specificity. Patients with FeNO≤55 ppb, sputum eosinophil ≤4.5% and not dependent on oral corticosteroids had an 86% probability of having non-eosinophilic inflammation (Fig 6 Node 5, 22% for Group 2 and 64% for Group 4). Clustering analysis based on the 3 top gene sets in the 9-GSVA signature associated with these clinical traits confirmed the molecular association of Group 1 subjects with high Th2 activation and corticosteroid insensitivity in bronchial biopsy and epithelial brushing samples (Fig 7).

Discussion

We clustered gene expression data from bronchial biopsies and epithelial brushings obtained from patients with moderate-to-severe asthma in order to define clinical phenotypes of asthma. Direct comparison of the transcriptome obtained from severe asthma to moderately-severe asthma subjects led to the identification of few differentially-expressed genes in bronchial biopsies and epithelial brushings using false discovery rate. We then examined their gene expression profiles based on important disease drivers as described by specific gene-sets and derived phenotypes from the gene-set clusters. A collection of 9 gene-set signatures applied to genes expressed in bronchial biopsies and airway epithelial brushings led to two subtypes of patients with eosinophilic inflammation and relative corticosteroid insensitivity. Using machine-learning tools, we found that currently-available inflammatory biomarkers such as sputum eosinophilia and exhaled nitric oxide levels could be used to predict the subtypes of asthmatics described by gene expression profiling. This approach could ultimately define patients who may benefit the most from specific molecular agents that target Th2-mediated inflammation and/or relative corticosteroid insensitivity. It also defines the importance of the site of expression of these 9 gene-set signatures in either biopsies or brushings or in both in determining the phenotype.

It is now generally accepted that asthma is a heterogeneous disease with distinct phenotypes. Our analysis confirmed the presence of a previously-described Th2-high cluster derived from an analysis of Th2-gene expression in epithelial brushings which is indicative of Th2 activation (4, 41) and which was defined by two of the 42 GSVA gene-sets we used. This signature has been linked to other features of asthma, particularly corticosteroid responsiveness. The remaining signatures we used reflected the heterogeneous nature of asthma and represent many of the mechanisms that have been proposed to drive the asthmatic disease process in addition to the Th2 pathway.

Interestingly, GSVA gene-sets associated with steroid responsiveness accounted for the second and sixth most important signatures that were significantly up-regulated, when they

are expected to be down-regulated by corticosteroids in non-severe asthmatics. In addition, a moderate correlation was noted between Th2 and corticosteroid insensitivity signatures (*Supplementary Figure S1*) reflecting the clinical problem of clinical insensitivity to corticosteroid treatment in a subgroup of severe asthmatics. Further analysis of the genes that were associated in the corticosteroid insensitivity and Th2 signatures in the biopsies and brushings has shown an association between *CD44* and the Th2-associated genes *CCL26, IL1R2* and *CST2*, indicating potential underlying mechanisms. The patients in Group 1 where the corticosteroid insensitivity signatures were most highly expressed in the airway epithelium and bronchial biopsy were on the highest doses of oral corticosteroids. Therapeutic strategies other than corticosteroid dose escalation should be considered for these patients.

Relative corticosteroid insensitivity was defined at the molecular level by detecting the signature of genes usually down-regulated by corticosteroids as being up-regulated. However, such a definition assumes that the patient is adherent to corticosteroid therapy, particularly oral therapy. The patients in UBIOPRED have been followed-up for at least for 6 months in a specialist severe asthma clinic during which time their degree of compliance has been deemed to be satisfactory. Our observation that Group 1 where the expression of the relative corticosteroid insensitivity signature together with the Th2 high and oxidative stress signatures were found in both bronchial biopsies and brushings was associated with a greater level of oral steroid dependency would support the notion that relative corticosteroid insensitivity may be associated with corticosteroid dependency. A potential limitation of the analysis of relative corticosteroid insensitivity is the use of the gene signatures obtained from patients with COPD, where the gene response to corticosteroids may be different from that seen in asthma.

Two gene-sets used were related to ozone stimulation as a marker of elevated oxidative stress and other important GSVA signatures that were critical in differentiating asthma included those from monocytes relating to autoimmune diseases and to Th1 immunity as

reflected by the increase in the numbers of submucosal $CD3^+$ and $CD8^+$ T cells. Amongst them, we noted that the oxidative stress signature revealed a moderate to high correlation with Th1 signature and with the steroid insensitivity signature (*Supplementary Figure S2 & S3*). While ROS has been reported to play a role in promoting asthmatic inflammation (14, 15) and reducing steroid responsiveness (16), the mechanism remains largely controversial. Therefore, the implication of oxidative stress in our finding may shed light on the impact of ROS in asthma. Overall, we found that there were non-Th2 pathways that were associated with severe asthma in the analysis of both bronchial biopsies and epithelial brushings despite the presence of eosinophilic inflammation.

One potential concern regarding our approach in this study is the appropriateness of gene sets used for clustering. We endeavored to be as inclusive as possible by using gene sets reflecting the diversity of processes previously reported to be associated with various aspects of asthma pathobiology in addition to key mechanisms involved in chronic obstructive airway disease, baseline immune cell signatures and those reported in a number of systemic immune disorders. In keeping with this deployment of gene-sets, we found that Group 1 subjects were distinguished by diverse immune activation across airway compartments, whereas nearly half of Group 4 participants did not show pronounced activation of the immune cells selected. Interestingly, the enrichment of GSVA signatures representative of multiple immune pathways including Th2, Th1, Th17, neutrophil, dendritic cells and influenza response were largely lower in Group 4 compared to Group 1 (*Supplementary Figure S4*), suggesting the lack of a dominant disease driving immune pathway and may reflect a pauci-immune phenotype of asthma. Further unbiased bioinformatic analyses such as the use of weighted gene co-expression network analysis (WGCNA) will be necessary to determine other potential disease-driving mechanisms. In a preliminary WGCNA analysis of epithelial cell gene expression, we found several modules related to FeNO, sputum eosinophilia and oral corticosteroid use with distinct eigengenes associated with these clinical parameters.

Although the 9 GSVA gene-set signature was initially derived from gene expression in

bronchial biopsies, its expression in epithelial brushings was associated with similar clinical characteristics. The strength of this signature was greater when it was highly expressed in epithelial cells than when it was expressed solely in bronchial biopsies particularly when linked to clinical parameters such as FeNO, blood eosinophils, sputum eosinophils and subtypes of submucosal T cells. This finding echoes previous studies highlighting the key role of the bronchial epithelium in dictating Th2 asthma phenotype and emphasizes the importance of the bronchial epithelium in driving asthma pathobiology (42). However, cross-talk between these compartments must exist as the most severe asthma subtype we describe here featuring eosinophilic inflammation and corticosteroid insensitivity (Group 1) occurred in subjects where the 9 GSVA signatures were concomitantly highly expressed in both bronchial biopsies and epithelial brushings. Interestingly, the expression of the 9 GSVA gene-set signatures in epithelial cells was mainly associated with sputum eosinophilia (Group 3 participants).

This observation may reflect priming of the asthmatic airway upon which Group 1 subtype was dependent when the signature was also found in the bronchial biopsies. This priming is further supported by the observation that subjects lacking this 9 GSVA signature expression in epithelial cells do not show eosinophilic inflammation irrespective of their signature expression in bronchial biopsies (Groups 2 and 4 participants). This highlights the need to further clarify the role played by airway epithelial cells in driving the development of airway inflammation in asthma. Indeed, recent work indicates that the airway epithelium is an important upstream cell that controls the regulation of Th2 cytokines through the production of cytokines such as thymic stromal lymphopoietin (TSLP), IL-25 and IL-33, which can be stimulated by exposure of epithelial cells to external stimuli such as pollutants, viruses and allergens (42, 43). The epithelium in severe asthma is also reported to be thicker than in mild-to-moderate asthma with increased proliferation, apoptosis and release of pro-inflammatory factors (44). However, our results indicate that IL-33 and TSLP gene expression is increased in the biopsies not in the brushings in asthmatic patients in Cluster A

and in Group 1 who also show the highest Th2 gene signature with bronchial eosinophilia.

The subtypes of asthma reported here, based on transcriptomic rather than clinical data, share some overlapping clinical and biomarker features with phenotypes driven by clinical parameters from the Leicester (23) and Severe Asthma Research Program (SARP) (22) clustering studies. The small Group 3 phenotype (10% of patients) reported here is very similar to Cluster 3 in SARP (8%) in terms of higher BMI and higher sputum eosinophils but lower biopsy eosinophils, relative to the remaining asthmatic groups. In addition, our Group 1 phenotype (23%) has some resemblance to Cluster 5 in SARP (16%) regarding the lower $FEV_{1,}$ whereas they were discordant with respect to its neutrophilic profile. Indeed, we found no difference in the level of either blood or sputum neutrophilia in the 4 groups. These findings provide a bridge between phenotypes derived from clustering of clinical parameters and our approach of using gene expression profiles within key airway structural cells along with tissue histopathology. We have provided a path to exploring the mechanisms underlying different subsets of patients where discordant inflammation is found.

A key question is whether there is any clinical utility or application of the phenotypes described here particularly as they were determined using transcriptomic analysis of samples obtained using bronchoscopy. The clinical utility is addressed by the tree-based recursive partitioning algorithm selecting significant non-invasive parameters or clinical traits based on their conditional distribution statistics against phenotypes using an unbiased permutation method (45). This approach successfully translated our phenotypes originally defined by gene signatures and by histopathological results to a clinically inferable scheme. Following a stepwise binary value split of FeNO (>55ppb), sputum eosinophils (>4.5%) or oral corticosteroid dependency, this scheme identified Group 1 subjects with an 84 % sensitivity and 72% specificity. This is clinically important as this group bears the Th2 and steroid insensitivity signatures across airway structural cells that justifies the need for molecular agents that target Th2-specific pathobiology (20). However, other groups such as Group 2 and 3 may also benefit from these agents with the possibility that different targets within the

Th2 phenotype may be more appropriate for each of the 3 groups. Thus, epithelial-dependent release of different eosinophilic cytokines/chemokines might be appropriate for Groups 1 and 3, while in Group 2, a non-eosinophilic Th2 or T2 targets might be more useful.

In conclusion, transcriptomic profiling in bronchial biopsies and epithelial brushing cells showed heterogeneity of underlying mechanisms with respect to Th2 and relative steroid insensitivity signatures which allowed the definition of different asthma phenotypes. Routinely-used clinical biomarkers successfully translated these findings, and used in this way may help in taking therapeutic decisions for patients with severe eosinophilic inflammation and relative corticosteroid insensitivity.

Acknowledgements

U-BIOPRED is supported through an Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115010, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution (www.imi.europa.eu). We thank all the members of each recruiting Centres for the recruitment and assessment of the participants. Chih-Hsi Scott Kuo is supported by the Department of Thoracic Medicine, Linkou Chang-Gung Memorial Hospital through a grant from Chang-Gung Medical Foundation, Taoyuan, Taiwan.

.

[#]U-BIOPRED consortium study group members

Nora Adriaens[1], Hassan Ahmed[2], Antonios Aliprantis[3], Kjell Alving[4], Philipp Badorek[5], David Balgoma6, Clair Barber[7], An Bautmans[8], Annelie F. Behndig[9], Elisabeth Bel[1], Jorge Beleta[10], Ann Berglind[6,11], Alix Berton[12], Jeanette Bigler[13], Hans Bisgaard[14], Grazyna Bochenek[15], Michael J. Boedigheimer[13], Klaus Bøonnelykke[14], Joost Brandsma[16], Armin Braun[5], Paul Brinkman[1], Dominic Burg[17], Davide Campagna[18], Leon Carayannopoulos[19], João P. Carvalho da Purfição Rocha[20], Amphun Chaiboonchoe[2], Romanas Chaleckis[6], Courtney Coleman[21], Chris Compton[22], Arnaldo D'Amico[23], Barbro Dahlén[6, 24], Jorge De Alba[10], Pim de Boer[25], Inge De Lepeleire[8], Tamara Dekker[1], Ingrid Delin[6], Patrick Dennison[6], 26, Annemiek Dijkhuis[1], Aleksandra Draper[27], Jessica Edwards[21], Rosalia Emma[18], Magnus Ericsson[24], Veit Erpenbeck[28], Damijan Erzen[29], Cornelia Faulenbach[5], Klaus Fichtner[29], Neil Fitch[27], Breda Flood[21], Urs Frey[30], Martina Gahlemann[31], Gabriella Galffy[32], Hector Gallart[6], Trevor Garret[27], Thomas Geiser[33], Jilaiha Gent[20], Maria Gerhardsson de Verdier[12], David Gibeon[34], Cristina Gomez[6], Kerry Gove[7], Neil Gozzard[35], Yi-Ke Guo[36], Simone Hashimoto[1], John Haughney[37], Gunilla Hedlin[6, 11], Pieter-Paul Hekking[1], Elisabeth Henriksson[24], Lorraine Hewitt[7], Tim Higgenbottam[38], Uruj Hoda[20], Jans Hohlfeld[5], Cecile Holweg[39], Peter Howarth[7], Richard Hu[13], Sile Hu[34], Xugang Hu[13], Val Hudson[21], Anna J. James[6], Juliette Kamphuis[25], Erika J. Kennington[21], Dyson Kerry[40], Matthias Klüglich[29], Hugo Knobel[41], Richard Knowles[42], Alan Knox[43], Johan Kolmert[6], Jon Konradsen[6,11], Maxim Kots[44], Linn Krueger[30], Scott Kuo[34], Maciej Kupczyk[6], Bart Lambrecht[45], Ann-Sofie Lantz[6, 11], Lars Larsson[12], Nikos Lazarinis[24], Saeeda Lone-Satif[1], Lisa Marouzet[7], Jane Martin[7], Sarah Masefield[46], Caroline Mathon[6], John G.Matthews[39], Alexander Mazein[2], Sally Meah[34], Andrea Maiser[34], Andrew Menzies-Gow[20], Leanne Metcalf[21], Roelinde Middelveld[6], Maria Mikus[47], Montse Miralpeix[10], Philips Monk[48], Nadia Mores[49], Clare S. Murray[50, 51], Jacek Musial[15], David Myles[22], Shama Naz[6], Katja Nething[29], Ben Nicholas[52], Ulf Nihlen[12], Peter Nilsson[47], Björn Nordlund[4, 6], Jörgen Östling[12], Antonio Pacino[53], Laurie Pahus[54], Susanna Palkonnen[55], Stelios Pavlidis[34], Giorgio Pennazza[23], Anne Petrén[6], Sandy Pink[7], Anthony Postle[52], Pippa Powel[46],

Malayka Rahman-Amin[21], Navin Rao[56], Lara Ravanetti[1], Emma Ray[7], Stacey Reinke[6], Leanne Reynolds[21], Kathrin Riemann[29], John Riley[22], Martine Robberechts[8], Amanda Roberts[21], Christos Rossios[34], Kirsty Russell[34], Michael Rutgers[25], Giuseppe Santini[49], Marco Sentoninco[26], Corinna Schoelch[29], James P.R. Schofield1[7], Wolfgang Seibold[29], Ralf Sigmund[29], Marcus Sjödin[6], Paul J.Skipp[17], Barbara Smids[1], Caroline Smith[7], Jessica Smith[21], Katherine M. Smith[43], Päivi Söderman[11], Adesimbo Sogbesan[20], Doroteya Staykova[57], Karin Strandberg[24], Kai Sun[34], David Supple[21], Marton Szentkereszty[32], Lilla Tamasi[32], Kamran Tariq[7, 26], John-Olof Thörngren[24], Bob Thornton[19], Jonathan Thorsen[14], Salvatore Valente[23], Wim van Aalderen[1], Marianne van de Pol[1], Kees van Drunen[1], Marleen van Geest[12], Jenny Versnel[21], Jorgen Vestbo[50, 51], Anton Vink[41], Nadja Vissing[14], Christophe von Garnier[33], Arianne Wagerner[1], Scott Wagers[27], Frans Wald[29], Samantha Walker[21], Jonathan Ward[58], Zsoka Weiszhart[32], Kristiane Wetzel[29], Craig E. Wheelock[6], Coen Wiegman[34], Siân Williams[37], Susan J. Wilson[58], Ashley Woodcock[50, 51], Xian Yang[34], Elizabeth Yeyashingham[59], Wen Yu[13], Wilhelm Zetterquist[4, 6], Koos Zwinderman[11]: Academic Medical Centre, University of Amsterdam, The Netherlands; [2]: European Institute for Systems Biology and Medicine, CIRI UMR5308, CNRS-ENS-UCBL-INSERM, Lyon, France; [3]: Merck Research Laboratories, Boston, USA; [4]: Department of Women's & Children's Health, Uppsala University, Sweden; [5]: Fraunhover Institute for Toxicology and Experimental Medicine, Hannover, Germany; [6]: Centre for Allergy Research, Karolinska Institutet, Stockhlom, Sweden; [7]: NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, UK; [8]: MSD, Brussels, Belgium; [9]: Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; [10]: Almirall S.A., Barcelona, Spain; [11]: Department of Women's & Childern's Health, Karolinska Institutet, Stockholm, Sweden; [12]: AstraZeneca, Mölndal, Sweden; [13]: Amgen Inc. Seattle, USA; [14]: Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Genofte Hospital, University of Copenhagen, Copenhagen, Denmark; [15]: Department of Internal Medicine, Jagiellonian University Medical College, Krakow, Poland; [16]: Faculty of Medicine, Southampton University, Southampton, UK; [17]: Centre for Proteomics Research, Institute of Life Sciences, University of Southampton, Southampton, UK; [18]: Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy; [19]: MSD, Kenilworth, USA; [20]: Royal Brompton and Harefield NHS Fondation Trust, London, UK; [21]: Asthma UK, London, UK; [22]: Respiratory Therapeutic Unit, GSK, London, UK; [23]: University of Rome 'Tor Vergata', Rome, Italy; [24]: University Hospital, Karolinska Institutet, Stockholm, Sweden; [25]: Longfonds, Amersfoort, The Netherlands; [26]: NIHR-Wellcome Trust Clinical Research Facility, Faculty of Medicine, University of Southampton, Southampton, UK; [27]: BioSci Consulting, Maasmechelen, Belgium; [28]: Translational Medicine, Respiratory Profiling, Novartis Institute for Biomedical Research, Basel, Switzerland; [29]: Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany; [30]: University Children's Hospital, Basel, Switzerland; [31]: Boehringer Ingelheim (Schweiz) GmbH, Basel, Switzerland; [32]: Semmelweis Universty, Budapest, Hungary; [33]: Department of Respiratory Medicine, University Hospital

Bern, Bern, Switzerland; [34]: National Hearth and Lund Institute, Imperial College, London, UK; [35]: UCB, Slough, UK; [36]: Data Science Institute, Imperial College, London, UK; [37]: International Primary Care Respiratory Group, Aberdeen, Scotland; [38]: Allergy Therapeutics, West Sussex, UK; [39]: Respiratory and Allergy Diseases, Genentech, San Francisco, USA; [40]: CromSource, Stirling, UK; [41]: Philips Research Laboratories, Eindhoven, The Netherlands; [42]: Arachos Pharma, Stevenage, UK; [43]: Respiratory Research Unit, University of Nottingham, UK; [44]: Chiesi Pharmaceuticals, SPA, Parma, Italy; [45]: University of Gent, Gent, Belgium; [46]: European Lung Foundation, Sheffield, UK; [47]: Science for Life Laboratory & The Royal Institute of Technology, Stockholm, Sweden; [48]: Synairgen Research Ltd, Southampton, UK; [49]: Università Cattolica del Sacro Cuore, Roma, Italy; [50]: Centre for Respiratory Medicine and Allergy, Institute of Inflammation and Repair, University of Manchester, Manchester, UK; [51]: University Hospital of South Manchester, Manchester Academic Health Sciences Centre, Manchester, UK; [52]: Faculty of Health Science, Southampton University, Southampton, UK; [53]: Lega Italiana Anti Fumo, Catania, Italy; [54]: Assistance Publique des Hôpitaux de Marseille, Clinique des bronches, allergies et sommeil, Espace EthiqueMéditerranéen, Aix-Marseille Université, Marseille, France; [55]: European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium; [56]: Janssen Research & Development, USA; [57]: Centre for Biological Sciences, University of Southampton, Southampton, UK; [58]: Histochemistry Research Unit, Faculty of Medicine, University of Southampton, Southampton, UK; [59]: UK Clinical Operations, GSK, Stockley Park, UK

Reference

1. Trejo Bittar HE, Yousem SA, Wenzel SE. Pathobiology of severe asthma. *Annual review of pathology* 2015; 10: 511-545.
2. Bentley AM, Meng Q, Robinson DS, Hamid Q, Kay AB, Durham SR. Increases in activated T lymphocytes, eosinophils, and cytokine mRNA expression for interleukin-5 and granulocyte/macrophage colony-stimulating factor in bronchial biopsies after allergen inhalation challenge in atopic asthmatics. *American journal of respiratory cell and molecular biology* 1993; 8: 35-42.
3. Robinson DS, Ying S, Bentley AM, Meng Q, North J, Durham SR, Kay AB, Hamid Q. Relationships among numbers of bronchoalveolar lavage cells expressing messenger ribonucleic acid for cytokines, asthma symptoms, and airway methacholine responsiveness in atopic asthma. *The Journal of allergy and clinical immunology* 1993; 92: 397-403.
4. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *American journal of respiratory and critical care medicine* 2009; 180: 388-395.
5. Boniface S, Koscher V, Mamessier E, El Biaze M, Dupuy P, Lorec AM, Guillot C, Badier M, Bongrand P, Vervloet D, Magnan A. Assessment of T lymphocyte cytokine production in induced sputum from asthmatics: a flow cytometry study. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2003; 33: 1238-1243.
6. Cho SH, Stanciu LA, Holgate ST, Johnston SL. Increased interleukin-4, interleukin-5, and interferon-gamma in airway CD4+ and CD8+ T cells in atopic asthma. *American journal of respiratory and critical care medicine* 2005; 171: 224-230.
7. Krug N, Madden J, Redington AE, Lackie P, Djukanovic R, Schauer U, Holgate ST, Frew AJ, Howarth PH. T-cell cytokine profile evaluated at the single cell level in BAL and blood in allergic asthma. *American journal of respiratory cell and molecular biology* 1996; 14: 319-326.
8. Raundhal M, Morse C, Khare A, Oriss TB, Milosevic J, Trudeau J, Huff R, Pilewski J, Holguin F, Kolls J, Wenzel S, Ray P, Ray A. High IFN-gamma and low SLPI mark severe asthma in mice and humans. *J Clin Invest* 2015; 125: 3037-3050.
9. Chambers ES, Nanzer AM, Pfeffer PE, Richards DF, Timms PM, Martineau AR, Griffiths CJ, Corrigan CJ, Hawrylowicz CM. Distinct endotypes of steroid-resistant asthma characterized by IL-17A(high) and IFN-gamma(high) immunophenotypes: Potential benefits of calcitriol. *J Allergy Clin Immunol* 2015; 136: 628-637.e624.
10. McKinley L, Alcorn JF, Peterson A, Dupont RB, Kapadia S, Logar A, Henry A, Irvin CG, Piganelli JD, Ray A, Kolls JK. TH17 cells mediate steroid-resistant airway inflammation and airway hyperresponsiveness in mice. *J Immunol* 2008; 181: 4089-4097.
11. Doe C, Bafadhel M, Siddiqui S, Desai D, Mistry V, Rugman P, McCormick M, Woods J, May R, Sleeman MA, Anderson IK, Brightling CE. Expression of the T helper 17-associated

cytokines IL-17A and IL-17F in asthma and COPD. *Chest* 2010; 138: 1140-1147.

12. Comhair SA, Ricci KS, Arroliga M, Lara AR, Dweik RA, Song W, Hazen SL, Bleecker ER, Busse WW, Chung KF, Gaston B, Hastie A, Hew M, Jarjour N, Moore W, Peters S, Teague WG, Wenzel SE, Erzurum SC. Correlation of systemic superoxide dismutase deficiency to airflow obstruction in asthma. *American journal of respiratory and critical care medicine* 2005; 172: 306-313.
13. Kirkham P, Rahman I. Oxidative stress in asthma and COPD: antioxidants as a therapeutic strategy. *Pharmacology & therapeutics* 2006; 111: 476-494.
14. Comhair SA, Bhathena PR, Farver C, Thunnissen FB, Erzurum SC. Extracellular glutathione peroxidase induction in asthmatic lungs: evidence for redox regulation of expression in human airway epithelial cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2001; 15: 70-78.
15. Dweik RA, Comhair SA, Gaston B, Thunnissen FB, Farver C, Thomassen MJ, Kavuru M, Hammel J, Abu-Soud HM, Erzurum SC. NO chemical events in the human airway during the immediate and late antigen-induced asthmatic response. *Proceedings of the National Academy of Sciences of the United States of America* 2001; 98: 2622-2627.
16. Ito K, Hanazawa T, Tomita K, Barnes PJ, Adcock IM. Oxidative stress reduces histone deacetylase 2 activity and enhances IL-8 gene expression: role of tyrosine nitration. *Biochemical and biophysical research communications* 2004; 315: 240-245.
17. James AL, Bai TR, Mauad T, Abramson MJ, Dolhnikoff M, McKay KO, Maxwell PS, Elliot JG, Green FH. Airway smooth muscle thickness in asthma is related to severity but not duration of asthma. *The European respiratory journal* 2009; 34: 1040-1045.
18. Kaminska M, Foley S, Maghni K, Storness-Bliss C, Coxson H, Ghezzo H, Lemiere C, Olivenstein R, Ernst P, Hamid Q, Martin J. Airway remodeling in subjects with severe asthma with or without chronic persistent airflow obstruction. *The Journal of allergy and clinical immunology* 2009; 124: 45-51 e41-44.
19. Macedo P, Hew M, Torrego A, Jouneau S, Oates T, Durham A, Chung KF. Inflammatory biomarkers in airways of patients with severe asthma compared with non-severe asthma. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2009; 39: 1668-1676.
20. Chung KF. New treatments for severe treatment-resistant asthma: targeting the right patient. *The Lancet Respiratory medicine* 2013; 1: 639-652.
21. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nature medicine* 2012; 18: 716-725.
22. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D'Agostino R, Jr., Castro M, Curran-Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair SA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleecker ER. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. *American journal of respiratory and critical care medicine* 2010;

181: 315-323.

23. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, Wardlaw AJ, Green RH. Cluster analysis and clinical asthma phenotypes. *American journal of respiratory and critical care medicine* 2008; 178: 218-224.
24. Bel EH, Sousa A, Fleming L, Bush A, Chung KF, Versnel J, Wagener AH, Wagers SS, Sterk PJ, Compton CH. Diagnosis and definition of severe refractory asthma: an international consensus statement from the Innovative Medicine Initiative (IMI). *Thorax* 2011; 66: 910-917.
25. Miller MK, Johnson C, Miller DP, Deniz Y, Bleecker ER, Wenzel SE. Severity assessment in asthma: An evolving concept. *The Journal of allergy and clinical immunology* 2005; 116: 990-995.
26. Wenzel SE. Asthma: defining of the persistent adult phenotypes. *Lancet* 2006; 368: 804-813.
27. Chung KF, Adcock IM. Clinical phenotypes of asthma should link up with disease mechanisms. *Current opinion in allergy and clinical immunology* 2015; 15: 56-62.
28. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, Ortega H, Chanez P. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. *Lancet* 2012; 380: 651-659.
29. Chung KF. Defining phenotypes in asthma: a step towards personalized medicine. *Drugs* 2014; 74: 719-728.
30. Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, Perez MF, Zhao H, Mane S, Martinez FD, Ober C, Nicolae DL, Barnes KC, London SJ, Gilliland F, Weiss ST, Raby BA, Cohn L, Chupp GL. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. *American journal of respiratory and critical care medicine* 2015; 191: 1116-1125.
31. Edelman E, Porrello A, Guinney J, Balakumaran B, Bild A, Febbo PG, Mukherjee S. Analysis of sample set enrichment scores: assaying the enrichment of sets of genes for individual samples in genome-wide expression profiles. *Bioinformatics* 2006; 22: e108-116.
32. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* 2013; 14: 7.
33. Bao Z, Zhang C, Yan W, Liu Y, Li M, Zhang W, Jiang T. BMP4, a strong better prognosis predictor, has a subtype preference and cell development association in gliomas. *Journal of translational medicine* 2013; 11: 100.
34. Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences of the United States of America* 2011; 108: 7265-7270.
35. Sarikonda G, Pettus J, Phatak S, Sachithanantham S, Miller JF, Wesley JD, Cadag E, Chae J, Ganesan L, Mallios R, Edelman S, Peters B, von Herrath M. CD8 T-cell reactivity to

islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes. *Journal of autoimmunity* 2014; 50: 77-82.

36. Thalamuthu A, Mukhopadhyay I, Zheng X, Tseng GC. Evaluation and comparison of gene clustering methods in microarray analysis. *Bioinformatics* 2006; 22: 2405-2412.
37. Auffray C, Adcock IM, Chung KF, Djukanovic R, Pison C, Sterk PJ. An integrative systems biology approach to understanding pulmonary diseases. *Chest* 2010; 137: 1410-1416.
38. Kuo C. S. PS, Loza M., Baribaud F., Rowe A., Pandis I., Rossios C., Wilson S.J., Djukanovic R., Sterk P.J., Chung K.F., Adcock I.M. & Guo Y-K. . Asthma phenotypes from a semi-supervised machine-learning approach of bronchial biopsy and brush transcriptomic in U-BIOPRED. *American Journal of Respiratory Critical Care Medicine* 2015; 191: A2205.
39. Dominick E Shaw ARS, Stephen Fowler, Louise Fleming, Graham Roberts, Julie Corfield, Ioannis Pandis, Aruna T Bansal, Elisabeth H Bel, Charles Auffray, Chris H Compton, Hans Bisgaard, Enrica Bucchioni, Massimo Caruso, Pascal Chanez, Barbro Dahlén, Sven-Erik Dahlen, Urs Frey, Thomas Geiser, Maria Gerhardsson de Verdier, David Gideon, Yi-ke Guo, Simone Hashimoto, Gunilla Hedlin, Elizabeth Jeyasingham, Tim Higenbottam, Ildikó Horváth, Alan Knox, Nobert Krug, Lars X Larsson, Nikos Lazarinis, Roelinde Middelveld, Paolo Montuschi, Jacek Musial, David Myles, Laurie Pahus, Thomas Sandström, Wolfgang Seibold, Florian Singer, Karin Strandberg, Jorgen Vestbo, Nadja Vissing, Christophe von Garnier, Ian M Adcock, Scott Wagers, Anthony Rowe, Peter Howarth, Ratko Djukanovic, Peter J Sterk, Kian Fan Chung on behalf of the U-BIOPRED Consortium. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *The European respiratory journal* 2015; in press.
40. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science* 2013; 2013: 6-8.
41. Abbas AR, Jackman JK, Bullens SL, Davis SM, Choy DF, Fedorowicz G, Tan M, Truong BT, Gloria Meng Y, Diehl L, Miller LA, Schelegle ES, Hyde DM, Clark HF, Modrusan Z, Arron JR, Wu LC. Lung gene expression in a rhesus allergic asthma model correlates with physiologic parameters of disease and exhibits common and distinct pathways with human asthma and a mouse asthma model. *The American journal of pathology* 2011; 179: 1667-1680.
42. Holgate ST. The sentinel role of the airway epithelium in asthma pathogenesis. *Immunological reviews* 2011; 242: 205-219.
43. Licona-Limon P, Kim LK, Palm NW, Flavell RA. TH2, allergy and group 2 innate lymphoid cells. *Nature immunology* 2013; 14: 536-542.
44. Cohen L, E X, Tarsi J, Ramkumar T, Horiuchi TK, Cochran R, DeMartino S, Schechtman KB, Hussain I, Holtzman MJ, Castro M. Epithelial cell proliferation contributes to airway

remodeling in severe asthma. *American journal of respiratory and critical care medicine* 2007; 176: 138-145.

45. Hothorn T HK, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 2006; 15: 651-674.

Figure legends:

Figure 1: Differential expression analysis comparing moderate versus severe asthma from bronchial biopsies and epithelial brushing transcriptomic profiles on the basis of individual genes or Gene set variation analysis (GSVA). Given the FDR correction (broken line), no differentially expressed genes (A) or gene sets (B) in epithelial brushing was noted. Two genes were marginally significant (C) and three gene sets (D) showed a trend towards significance in bronchial biopsies. ES: enrichment score.

Figure 2: Topological network analysis of gene-set variation analysis using 42 gene sets of relevance to the pathogenesis of asthma in bronchial biopsies revealing one large signature-driven cluster (Cluster A) and 4 smaller distinct clusters. Cluster A was distinguished by higher submucosal eosinophil counts, higher levels of nitric oxide in exhaled breath and higher oral corticosteroid use. The network is coloured on the basis of the submucosal eosinophil count with low cell count in blue and high in red. The colour key histogram (bottom left) showed the number of subjects with respect to the given cell count (x-axis) from low to high. Subjects of high submucosal eosinophil count were noted to be highly enriched in Cluster A as compared to the rest.

Figure 3: Heat map of differential gene set expression from bronchial biopsies between Cluster A (blue bar) and Cluster non-A (grey bar). High expression is denoted by red and low expression by green. Twenty-six of 42 gene sets used were significantly differentially-expressed in Cluster A with 23 over-expressed and 3 under-expressed.

Figure 4: Hierarchical clustering showing 91 bronchial biopsies on 42 GSVA signatures and the color bar revealed the clustering result from TDA on the same dataset. The Cluster A (red bar) identified by TDA method overlapped substantially with the subgroup of patients identified by hierarchical clustering that shaded in grey. The agreement on two clustering

methods measured by Rand and adjusted Rand Index were 0.7 and 0.4, respectively.

Figure 5: Machine-learning algorithm using shrunken centroid method on 26 differentially-expressed gene sets in bronchial biopsies. Panel A: Training of classifiers for Cluster A was evaluated for classification error using 10-fold cross-validation. A threshold of 1.82 (red broken line) was selected which reduced classifiers to 9 gene-sets at an 82% cross-validated accuracy. Panel B: Centroid profile of the 9-gene-set signatures. Length of the centroid denoted the relative amount the expression was away from the overall mean expression for each given gene-set signature. From top down, the centroids of Cluster A and Cluster non-A were ranked in decreasing amount. Panel C: Discrimination performance of the 9 gene-set signatures (AUC: 0.866; 95% CI:0.796-0.927, p=1.3e^{-09}) based on the estimation of the Cluster A probability from Gaussian density distribution. The probability cut-off at 0.5 gave a sensitivity and specificity of 84.6% and 90.4% respectively, with the confidence intervals shown.

Figure 6: Tree-based inference scheme for the probability of Group 1 to Group 4 using two non-invasive measurements (FeNO: p=0.014, sputum eosinophil counts: p=0.012) and one clinical parameter (oral corticosteroid dependency: p=0.030). Group 1 (red bar) represented a phenotype of high eosinophilic inflammation and steroid insensitivity. Group 3 (pink bar) represented a phenotype of medium eosinophilic inflammation and Group 2 & 4 (black bar) low eosinophilic inflammation.

Figure 7: Three signatures related to Th2 and steroid insensitivity in epithelial (above the broken line) and biopsy cells (below the broken line) across 83 asthmatics were used for clustering. Subject's membership as Group 1 to 4 were shown as color bar and mapped underneath the dendrogram. Group 1 (magenta) patients with the highest and most extensive expression of these signatures were mainly clustered to the right. Group 3 (purple) subjects

with the signatures solely in the epithelium also clustered closely. Group 2 (light blue) and 4 (grey) subjects interlaced with each other and clustered to the left. These 3 out of the 9 signatures were the major representative features of Group 1 and 3 asthma subtypes.

Table 1. Demographic and clinical characteristics of 107 asthmatic subjects

Variables†	
Age (years)	45.9 (13.9)
Female	59 (55.1%)
BMI	28.7 (5.8)
Severe asthma category	67 (62.6%)
Oral corticosteroid use (number)	29 (27.1%)
Exacerbation numbers (per year)	2 (1-4)
Atopy	76 (71.0%)
Allergic rhinitis	53 (49.5%)
Eczema	42 (39.3%)
Nasal polyp	27 (25.2%)
FEV_1 (% predicted)	74.9 (62.3-95.1)
Total serum IgE (IU/ml)	115.6 (43.0-316.0)
Blood eosinophil (%)	3.1 (1.7-4.4)
Sputum eosinophil (%)	1.5 (0.4-7.5)
Sputum neutrophil (%)	52.5 (36.7-70.1)
FeNO (ppb)	31.5 (19.0-53.8)
Bronchial biopsy available	91 (85.0%)
Epithelial brushing available	99 (92.5%)

†: Data presented as N (%) and mean (SD) or median (IQR).BMI: Body mass index, FEV_1: Forced expiratory volume in 1 second, FeNO: Fractional exhaled nitric oxide

Table 2. Clinical characteristics of Cluster A versus non-A

Variables† **No (%)**	**Cluster A N= 39**	**Cluster non-A N=52**	**p-value**
Age (year)	48.4±13.6	45.8±13.1	0.355
Female	18 (46.2%)	31 (59.6%)	0.288
BMI	28.1±4.9	28.6±5.8	0.661
Severe asthma	26 (66.7%)	30 (57.7%)	0.258
Oral corticosteroid use	16 (41.0%)	8 (15.4%)	0.006
Exacerbation number (time/year)	2.5 (2.0-3.75)	2.0 (1.0-3.0)	0.139
Atopy	25 (64.1%)	40 (76.9%)	0.356
Allergic rhinitis	17 (43.6%)	26 (50.0%)	0.793
Eczema	15 (38.5%)	21 (40.4%)	1.000
Nasal polyp	14 (35.9%)	11 (21.2%)	0.092
FEV_1 (%)	74.2 (59.8-91.6)	74.7 (62.3-94.6)	0.630
Total serum IgE (IU/ml)	125.0 (41.0-231.5)	104.0 (44.0-364.5)	0.813
Blood eosinophil (%)	3.2 (1.6-4.8)	2.7 (1.7-3.8)	0.523
Sputum eosinophil (%)	3.9 (0.6-12.1)	1.1 (0.2-4.7)	0.156
Sputum neutrophil (%)	53.2 (42.7-76.9)	52.4 (34.6-63.4)	0.421
FeNO (ppb)	34.0 (20.0-61.0)	21.0 (14.5-40.0)	0.028
Submucosal cells (mm^{-2})			
Eosinophils	5.6 (2.7-11.4)	1.2 (0-3.3)	6.6×10^{-5}
Neutrophils	11.0 (6.6-18.1)	12.1 (6.8-18.4)	0.833
CD3 T cells	36.3 (22.6-50.0)	35.4 (16.6-49.5)	0.471
CD4 T cells	11.2 (5.9-18.9)	9.5 (4.8-15.8)	0.545
CD8 T cells	17.8 (9.5-30.4)	14.2 (4.9-27.7)	0.531

†: Data presented as N (%) and mean (SD) or median (IQR).

BMI: Body mass index, FEV_1: Forced expiratory volume in 1 second, FeNO: Fractional exhaled nitric oxide

Table 3. Nine GSVA signatures differentiating Cluster A from the rest

GSVA signature	ES difference	p-value	FDR
Lung.brushings.Th2high.asthma.HS.IVV.UP	0.578	$6.0x10^{-10}$	$2.5x10^{-8}$
Lung.biopsy.COPD.FLU.SAL.HS.IVV.DOWN	0.308	$5.3x10^{-6}$	$7.4x10^{-5}$
Lung.biopsy.HDM.Rhesus.IVV.UP	0.275	$7.8x10^{-7}$	$1.6x10^{-5}$
PBMC.Systemic immune disorders.HS.IVV.UP	0.203	0.001	0.005
Th1.activated.HS.IVS.UP.2	0.199	$7.2x10^{-6}$	$7.6x10^{-5}$
PBMC.asthma.GC.HS.IVS	0.234	$8.4x10^{-5}$	$7.1x10^{-4}$
Tcell.activated.HS.IVS	0.192	$1.5x10^{-4}$	0.001
Ozone.Air.MM.IVV.Up	0.157	$8.9x10^{-4}$	0.004
Ozone.MM.IVV.Down	-0.159	$2.1x10^{-4}$	0.001

ES: enrichment score, FDR: false discovery rate, HS: Homo sapiens, MM: Mus musculus, IVS: in vitro soil-less, IVV: in vivo, COPD: chronic obstructive pulmonary disease, FLU: fluticasone, SAL: salmeterol, HDM: house dust mite, PBMC: peripheral blood mononuclear cell

Table 4. Clinical characteristics of the 9-gene-set signature expressed in epithelial brushings

Variables† **No (%)**	**Expressed N= 31**	**Not expressed N=68**	**p-value**
Age (year)	48.0±14.8	44.4±14.0	0.253
Female	12 (38.7%)	40 (58.8%)	0.101
BMI	28.8±5.4	29.0±6.3	0.895
Severe asthma	23 (74.2%)	40 (58.8%)	0.211
Oral corticosteroid use	11 (40.7%)	10 (17.9%)	0.032
Exacerbation number (time/year)	3.0 (2.0-4.0)	2.0 (1.0-3.0)	0.025
Atopy	17 (63.0%)	40 (75.0%)	0.419
Allergic rhinitis	19 (61.3%)	30 (44.1%)	0.171
Eczema	10 (32.3%)	32 (47.1%)	0.245
Nasal polyp	10 (32.3%)	15 (22.1%)	0.404
FEV_1 (%)	74.2 (57.3-95.8)	87.9 (68.5-95.5)	0.264
Total serum IgE (IU/ml)	184.0 (44.0-373.0)	87.0 (39.6-268.5)	0.067
Blood eosinophil (%)	3.5 (2.7-7.4)	2.3 (1.5-3.8)	0.008
Sputum eosinophil (%)	12.0 (2.6-26.2)	0.6 (0-1.7)	$5.2x10^{-4}$
Sputum neutrophil (%)	53.2 (42.7-76.9)	52.4 (34.6-63.4)	0.786
FeNO (ppb)	51.5 (20.8-75.8)	21.0 (13.3-37.8)	$4.2x10^{-4}$
Submucosal cells (mm^{-2})			
Eosinophils	5.8 (3.0-12.5)	1.5 (0-4.8)	$1.3x10^{-5}$
Neutrophils	11.5 (4.6-17.1)	12.1 (6.8-19.2)	0.780
CD3 T cells	44.7 (32.2-61.7)	24.4 (14.8-46.8)	0.003
CD4 T cells	13.8 (7.9-23.7)	8.3 (3.2-14.6)	0.016
CD8 T cells	24.9 (15.7-37.6)	10.8 (4.2-20.2)	$2.0x10^{-4}$

†: Data presented as N (%) and mean (SD) or median (IQR).

BMI: Body mass index, FEV_1: Forced expiratory volume in 1 second, FeNO: Fractional exhaled nitric oxide

Table 5. Expression status of the 9 gene-set signature of Cluster A in 83vbronchial biopsy and epithelial brushing

Group	Bronchial biopsy	Epithelial brushing	N (%)
Group 1	Expressed	Expressed	19 (23%)
Group 2	Expressed	Not expressed	17 (20%)
Group 3	Not expressed	Expressed	8 (10%)
Group 4	Not expressed	Not expressed	39 (47%)

Table 6. Clinical features of the expression of 9 gene-set signatures in bronchial biopsy and/or epithelial brush

Variables[†]	**Group 1**	**Group 2**	**Group 3**	**Group 4**	**1 vs 4**	**3 vs 4**	**1 vs 3**
n (%)	19 (23%)	17 (20%)	8 (10%)	39 (47%)		*p* value	
Age (year)	48.0 ±13.6	47.4±14.1	50.6±13.4	44.0±13.5	0.284	0.187	0.644
Age of asthma onset	24.7±13.4	22.2±14.4	32.0±11.6	22.1±13.7	0.485	0.039	0.176
Female	7 (36.8%)	8 (47.1%)	3 (37.5%)	24 (61.5%)	0.136	0.390	1.000
BMI	27.8±4.8	28.6±4.9	33.1±3.8	28.0±5.9	0.986	0.010	0.007
Nasal polyps	8 (42.1%)	6 (35.3%)	2 (25.0%)	9 (23.1%)	0.102	0.898	0.357
Rhinitis	8 (42.1%)	8 (47.1%)	6 (75.0%)	16 (41.0%)	0.996	0.100	0.150
Eczema	6 (31.6%)	9 (52.9%)	2 (25.0%)	17 (43.6%)	0.604	0.525	0.900
Severe asthma	14 (73.7%)	10 (58.8%)	6 (75.0%)	20 (51.3%)	0.179	0.402	1.000
Oral corticosteroid use	9 (47.4%)	5 (29.4%)	2 (25.0%)	5 (12.8%)	0.010	0.737	0.515
Atopy	11 (57.9%)	12 (70.6)	6 (75.0%)	30 (76.9%)	0.312	0.476	0.466
Exacerbations/year	3.0 (2.0-5.0)	2.0 (1.3-2.8)	3.0 (2.0-3.0)	1.5 (1.0-3.0)	0.022	0.238	0.563
FEV_1 (%)	68.1(54.9-78.3)	90.4 (70.9-98.0)	80.3 (65.9-101.6)	79.7(65.5-94.8)	0.133	0.708	0.260
Total IgE (IU/ml)	163 (42–231.5)	86.6 (24.0-196.0)	93.9 (41.3-879.8)	96 (44.3-278.5)	0.659	0.839	0.815
Blood eosinophil (%)	3.5 (2.4-8.0)	1.7 (1.4-3.6)	3.4 (3.2-4.2)	2.4 (1.4-3.9)	0.070	0.100	0.915
Blood neutrophil (%)	58.8 (50.4-64.2)	61.9 (56.6-68.3)	65.5 (61.9-69.5)	58.6 (53.2-68.9)	0.596	0.380	0.217
Sputum eosinophil (%)	8.2 (1.8-27.0)	0.6 (0.6-1.0)	15.8(12.4-16.2)	0.6 (0-1.6)	0.017	0.014	0.863
Sputum neutrophil (%)	52.3 (33.4-77.4)	57.0 (54.2-76.7)	45.3 (41.4-50.8)	52.7 (34.3-65.4)	0.835	0.763	0.727
FeNO (ppb)	56.5 (34.1-74.6)	22.0 (14.5-32.5)	35.5 (20.0-50.4)	19 (12.5-36.1)	0.001	0.058	0.312
Submucosal cells (mm^{-2})							
Eosinophil	6.1 (4.3-13.6)	2.2 (0-5.9)	2.7 (1.5-4.6)	1.0 (0-3.5)	7.6E-06	0.109	0.019
Neutrophil	11.0 (4.6-18.2)	9.8 (7.1-18.7)	12.1 (6.5-13.5)	12.1 (6.8-19.2)	0.853	0.500	0.727
CD4 T cell	15.4 (9.2-24.7)	5.8 (1.5-11.1)	9.2 (6.2-16.3)	10.2 (4.6-15.5)	0.070	0.839	0.187
CD3 T cell	45.3 (32.7-77.2)	21.8 (16.0-37.0)	41.9 (36.9-70.9)	33.3(16.4-49.9)	0.048	0.243	0.874
CD8 T cell	26.4 (15.7-37.0)	9.4 (4.6-14.2)	27.1 (18.1-39.8)	13.9 (4.7-22.1)	0.020	0.150	0.852

†: Data presented as N (%) and mean (SD) or median (IQR). BMI: Body mass index, FEV_1: Forced expiratory volume in 1 second, FeNO: fractional exhaled nitric oxide

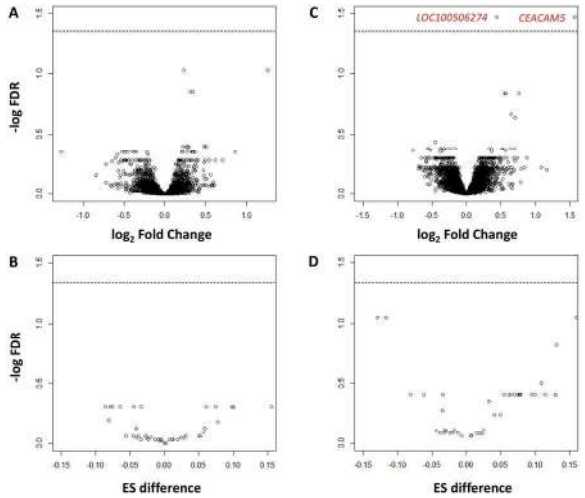


Figure 1

190x107mm (300 x 300 DPI)

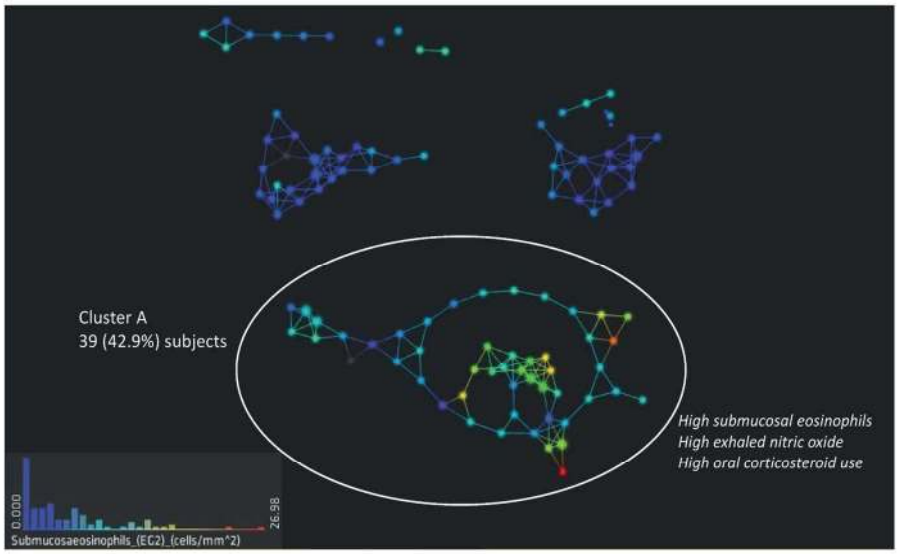


Figure 2

128x76mm (300 x 300 DPI)

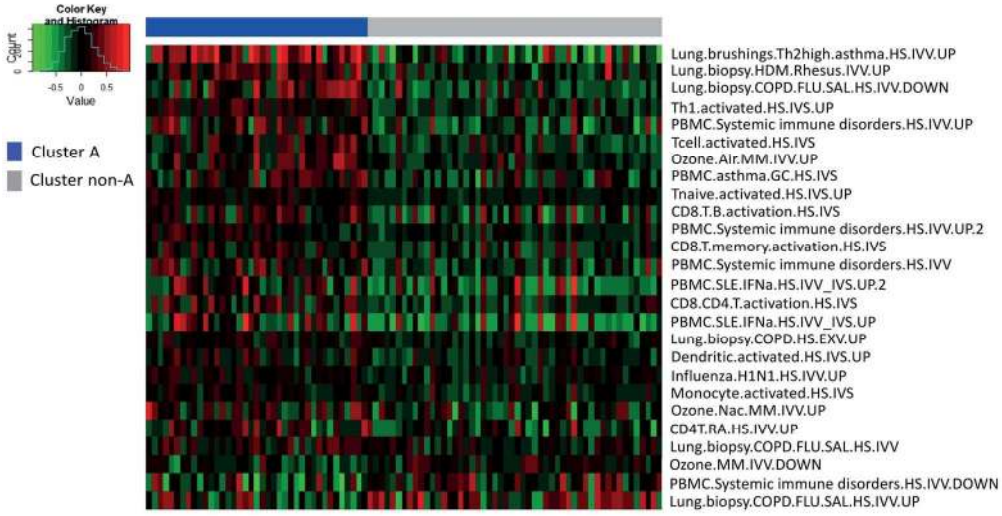


Figure 3

157x85mm (300 x 300 DPI)

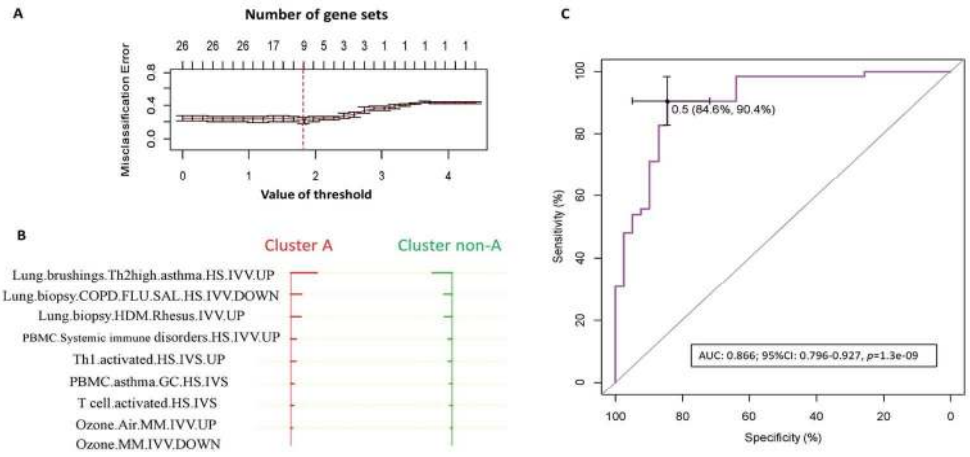


Figure 4

134x70mm (300 x 300 DPI)

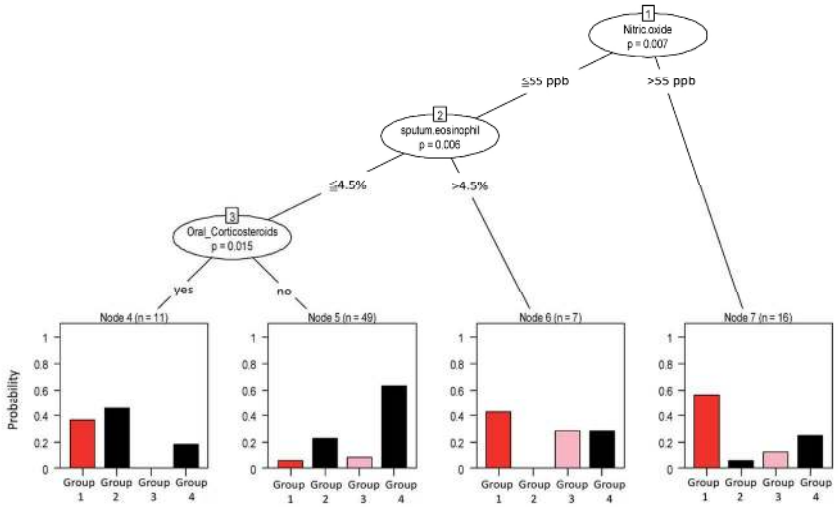


Figure 5

127x67mm (300 x 300 DPI)

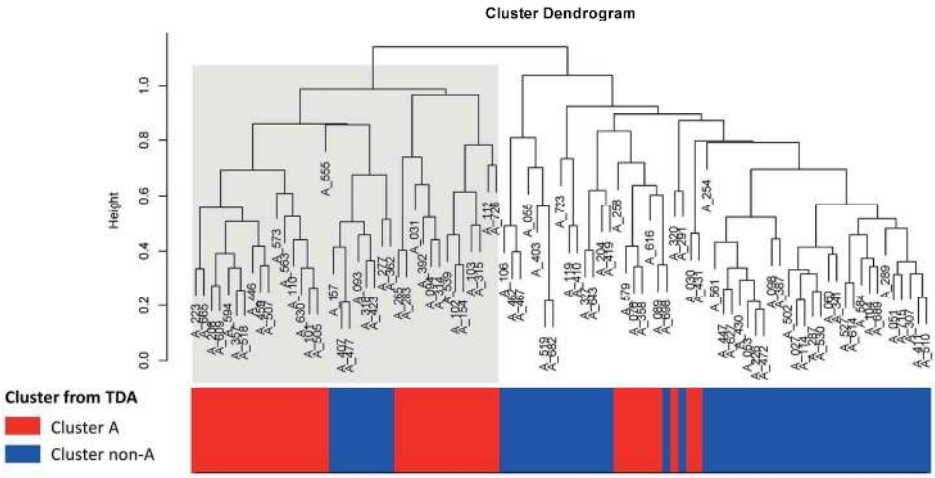


Figure 6

138x75mm (300 x 300 DPI)

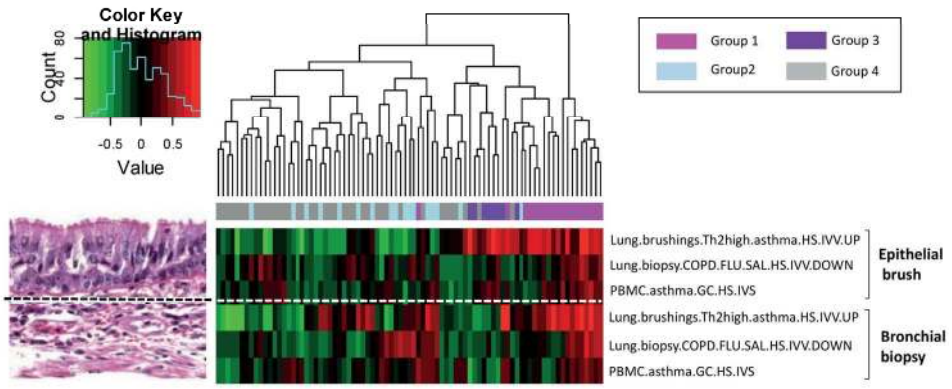


Figure 7

153x85mm (300 x 300 DPI)

Supplementary data

A transcriptomic-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in UBIOPRED

Chih-Hsi Scott Kuo, Stelios Pavlidis, Matthew Loza, Fred Baribaud, Anthony Rowe, Iaonnis Pandis, Uruj Hoda, Christos Rossios, Ana Sousa, Susan J. Wilson, Peter Howarth, Barbro Dahlen, Sven-Eric Dahlen, Pascal Chanez, Dominic Shaw, Norbert Krug, Thomas Sandström, Bertrand De Muelder, Charles Auffray, Peter J. Sterk, Ratko Djukanovic, Yike Guo, Ian M. Adcock & Kian Fan Chung on behalf of the U-BIOPRED Project Team

Table of Contents

Methods for Cluster Analysis and Machine Learning Algorithm …….....1

Topology Data analysis (TDA)…….....2

Nearest Shrunken Centroid Learning Algorithm……..…….....3

Supplementary Table and Figures……..4

Table S1…… ……....5

Figure S1……6

Figure S2……7

Figure S3……8

Figure S4……9

REFERENCES……..……..…….……..*11*

Methods for Cluster Analysis and Machine Learning Algorithm

Topology Data Analysis

Software used for TDA was provided by Ayasdi Core™ (Ayasdi, Inc. California, USA). TDA method takes three clustering parameters as input, namely *dissimilarity (distance) metric*, *filter function* and *resolution parameters* (resolution and overlap). A user-defined distance function computes the distance between any pairs of the data points and thus determines a metric space of the data set that serves as the starting point of TDA. The distance metric using correlation distance was a consensus setting in the UBIOPRED project across each individual analysis when applying the TDA approach for clustering. The filter function in TDA serves to generate a real valued quantity associated with each data point that reflects a unique view of the data set. There are various sort of filter functions in TDA that include functions that depend only on the distance metric itself, such as a density estimator or a measure of centrality; functions that produce linear projections on a data matrix, such as principle component analysis or multidimensional scaling; or functions that retrieved from input parameters by researchers, such as clinical metadata. In UBIOPRED, the filter function was an open option across each analysis. In this analysis, filter function of linear projection was chosen as the 1st and 2nd principle component of the 42 GSVA signatures in order to unbiasedly focus on the major variation within the data matrix.

The value of each data point computed by the filter function which spanned from low to high was segmented into overlapping intervals (bins). User-defined resolution parameters in TDA are used to control the size of the bins, a high resolution means fine-grained partitions and a low resolution suggests coarse-grained ones. A network in high resolution will preserve the strongest connections between data points and data points that are more weakly associated tend to break apart as the bins are split. An overlap parameter is used to define the extent to which the data will be

oversampled. This parameter allows users to set the number of bins that are shared by most data points thus a higher overlap parameter is often useful to preserve the relationship of continuity within a data set. In this analysis, we set a moderate-high resolution (Resolution=30) and overlap parameter (Gain=3x) in order to reveal the detailed structure of the network while not compromising the inherent continuity of the data. Using these input set up, data points having their filter function value sat in the same bin were combined when clustering was performed based on the correlation dissimilarity. By clustering similar data points into nodes and joining nodes that had shared data points with edges, TDA produced a coordinate-free network with 3D graphic visualization to enable cluster exploration. Subsequent analysis of these graphs is used to identify specific network features such as flares or separated clusters that represent subgroups within the data set. Ayasdi Core™ also provides a histogram-based coloring scheme for the constructed network from all available filter functions or metadata which was input by users. This flexible coloring scheme enables a quick overview of the characteristic of interest for all identified clusters or subgroups within the network [1, 2].

Nearest Shrunken Centroid learning algorithm

The nearest shrunken centroid method was used as a machine learning algorithm to further define the optimal number of GSVA signatures that best differentiated between identified clusters. The centroids (average expression of each GSVA signature) for each cluster as well as the overall samples were calculated. Standardization of the centroids of each cluster was performed through dividing the difference of the cluster centroids and overall centroids by the within-cluster standard deviation of each signature. This standardized value was treated as an absolute value which was later shrunken by an amount (Δ, threshold value). If the value of a given standardized

centroid was shrunken to zero by Δ for all clusters, then this signature did not contribute to the nearest-centroid classification. Otherwise, a non-zero value of a standardized centroid after shrinkage was retained as a centroid classifier for the given cluster. The inherent property of this method is that many signatures will be eliminated from cluster prediction as Δ increases. Given that each amount of Δ shrunken will result in a set of surviving centroids for each cluster, the amount of shrinkage is chosen by iterative cross-validation on the performance that each set of surviving centroids correctly predicts the cluster classification of each sample. This algorithm is available as a free R software as prediction analysis for microarrays in The Comprehensive R Archive Networks (CRAN-pamr package) [3].

Table S1. 42 signatures used for gene set variation analysis

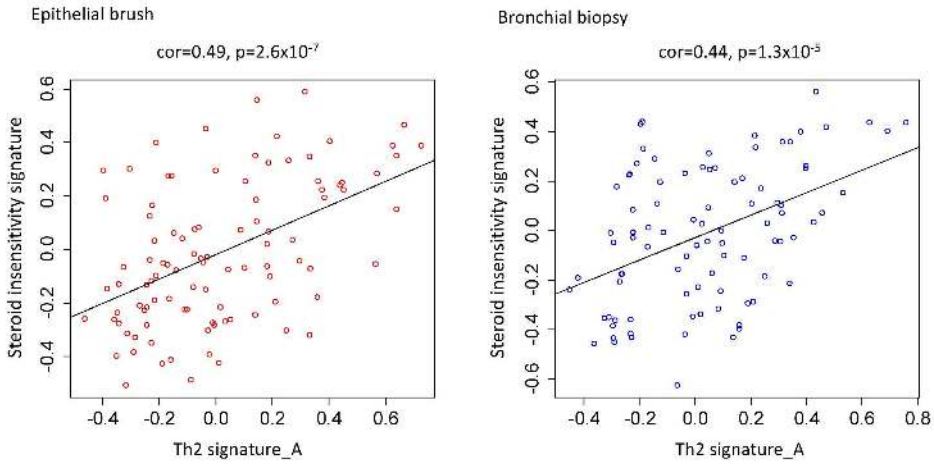
Name of gene-sets	Disease/Treatment	Species	Cell Type	Study Type	Reference
Omnibus immune cells activation					
Tcell.activated.HS.IVS	activated	HS	Tcell	IVS	[4]
Bcell.activated.HS.IVS	activated	HS	Bcell	IVS	[4]
Monocyte.activated.HS.IVS	activated	HS	Monocyte	IVS	[4]
NKcell.activated.HS.IVS	activated	HS	NKcell	IVS	[4]
Dentritic.activated.HS.IVS	activated	HS	Dentritic	IVS	[4]
Neutrophil.activated.HS.IVS	activated	HS	Neutrophil	IVS	[4]
Th1 and CD8 T cell immunity					
CD8.T.memory.activation.HS.IVS	activated	HS	peripheral blood	IVS	[5]
CD8.T.B.activation.HS.IVS	activated	HS	peripheral blood	IVS	[5]
CD8.CD4.T.activation.HS.IVS	activated	HS	peripheral blood	IVS	[5]
Th1.activated.HS.IVS.UP	activated	HS	peripheral blood	IVS	[6]
Th1.activated.HS.IVS.UP.2	activated	HS	peripheral blood	IVS	[7]
Influenza.H1N1.HS.IVV.UP	H1N1trivalent influenza vaccine	HS	peripheral blood	IVV	[8]
Th2, Th17 and T-reg cell immunity					
Th17.activated.HS.IVS.UP	activated	HS	peripheral blood	IVS	[6]
Th17.activated.HS.IVS.UP.2	activated	HS	peripheral blood	IVS	[7]
Th2.activated.HS.IVS.UP	activated	HS	peripheral blood	IVS	[7]
Treg.activated.HS.IVS.UP	activated	HS	peripheral blood	IVS	[7]
Tnaive.activated.HS.IVS.UP	activated	HS	peripheral blood	IVS	[7]
Oxidative stress					
Ozone.MM.IVV.Down	Ozone	MM	lung biopsy	IVV	#
Ozone.Air.MM.IVV.UP	Ozone followed by air	MM	lung biopsy	IVV	#
Ozone.Air.MM.IVV.DOWN	Ozone followed by air	MM	lung biopsy	IVV	#
Ozone.Nac.MM.IVV.UP	Ozone followed by N-acetylcysteine	MM	lung biopsy	IVV	#
Smoking.signature.Mouse	smoke	MM	lung biopsy	IVV	#
Lung.biopsy.COPD.HS.EXV.UP	COPD	HS	lung biopsy	EXVIVO	[9]
Systemic immune disorders					
PBMC.Systemic immune disorders.HS.IVV	4 autoimmune diseases	HS	peripheral blood	IVV	[10]
PBMC.Systemic immune	4 autoimmune	HS	peripheral blood	IVV	[10]

disorders.HS.IVV.UP	diseases				
PBMC.Systemic immune disorders.HS.IVV.DOWN	4 autoimmune diseases	HS	peripheral blood	IVV	[10]
PBMC.Systemic immune disorders.HS.IVV.2	3 autoimmune diseases	HS	peripheral blood	IVV	[10]
PBMC.SLE.IFNa.HS.IVV_IVS.UP	SLE (Bonferroni)	HS	peripheral blood	IVV_IVS	[11]
PBMC.SLE.IFNa.HS.IVV_IVS.UP.2	SLE (FDR)	HS	peripheral blood	IVV_IVS	[11]
CD4T.RA.HS.IVV.UP	RA	HS	peripheral blood	IVV	[12]
Asthma specific T cell immunity					
CD4.severe.asthma.HS.IVV.UP	Severe Asthma	HS	peripheral blood	IVV	[13]
CD4.severe.asthma.HS.IVV.UP	Severe Asthma	HS	peripheral blood	IVV	[13]
CD8.severe.asthma.HS.IVV.UP	Severe Asthma	HS	peripheral blood	IVV	[13]
CD8.severe.asthma.HS.IVV.DOWN	Severe Asthma	HS	peripheral blood	IVV	[13]
Lung.biopsy.HDM.Rhesus.IVV.UP	HDM	Rhesus	lung biopsy	IVV	[14]
Lung.biopsy.Th2high.asthma.HS.IVV.UP	Th2 high/low asthma	HS	lung biopsy	IVV	[15]
Lung.brushings.Th2high.asthma.HS.IVV.UP	Th2 high/low asthma	HS	lung brushings	IVV	[16]
Glucocorticoid response					
Lung.biopsy.COPD.FLU.SAL.HS.IVV	COPD fluticasone & salmeterol	HS	lung biopsy	IVV	[17]
Lung.biopsy.COPD.FLU.SAL.HS.IVV.UP	COPD fluticasone & salmeterol	HS	lung biopsy	IVV	[17]
Lung.biopsy.COPD.FLU.SAL.HS.IVV.DOWN	COPD.fluticasone & salmeterol	HS	lung biopsy	IVV	[17]
Lung.ASM.asthma.Prednisolone.HS.IVV.UP	Asthma, Prednisolone	HS	lung biopsy ASM	IVV	[18]
PBMC.asthma.GC.HS.IVS	Asthma,IL-1β, TNF-α, with or without GC	HS	peripheral blood	IVS	[19]

HS: Homo sapiens, MM: Mus musculus, IVS: in vitro soil-less, IVV: in vivo, INF: interferon, COPD: chronic obstructive pulmonary disease, FLU: fluticasone, SAL: salmeterol, HDM: house dust mite, PBMC: peripheral blood mononuclear cell, SLE: systemic lupus erythematosus, RA: rheumatoid arthritis, GC: glucocorticoid, #: unpublished data from Philip Hansbro et al.

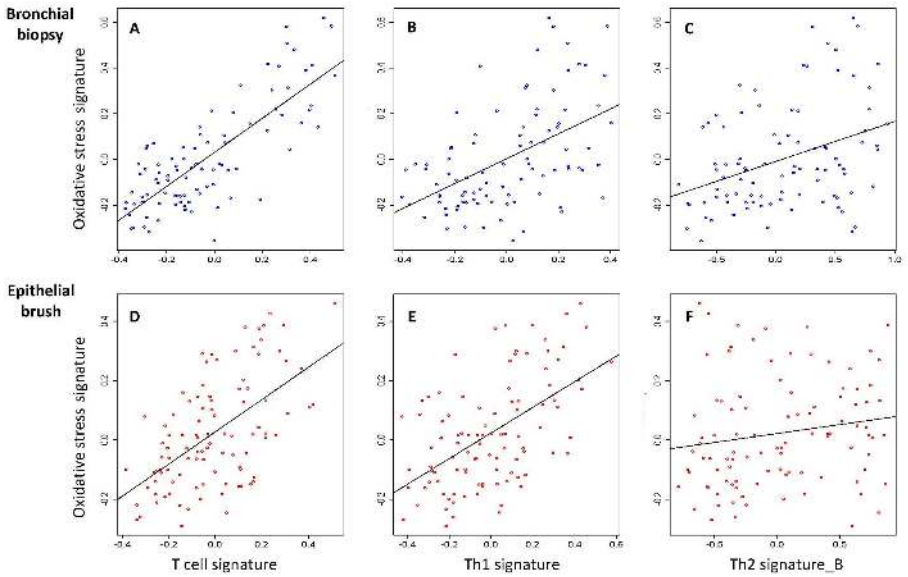
Supplementary Figures

Figure S1



Relationship between steroid insensitivity (PBMC.asthma.GC.HS.IVS) and Th2 signature-A (Lung.biopsy.HDM.Rhesus.IVV.UP) showing a moderate correlation in epithelial brushings (Pearson's *r*: 0.49, p=2.6x10^{-7}) and bronchial biopsies (Pearson's *r*: 0.44, p=1.3x10^{-5}). There was also a significant correlation between the steroid insensitivity signature and the Th2 signature-B (Lung.brushings.Th2high.asthma.HS.IVV.UP) in epithelial brushings and bronchial biopsies (data not shown).

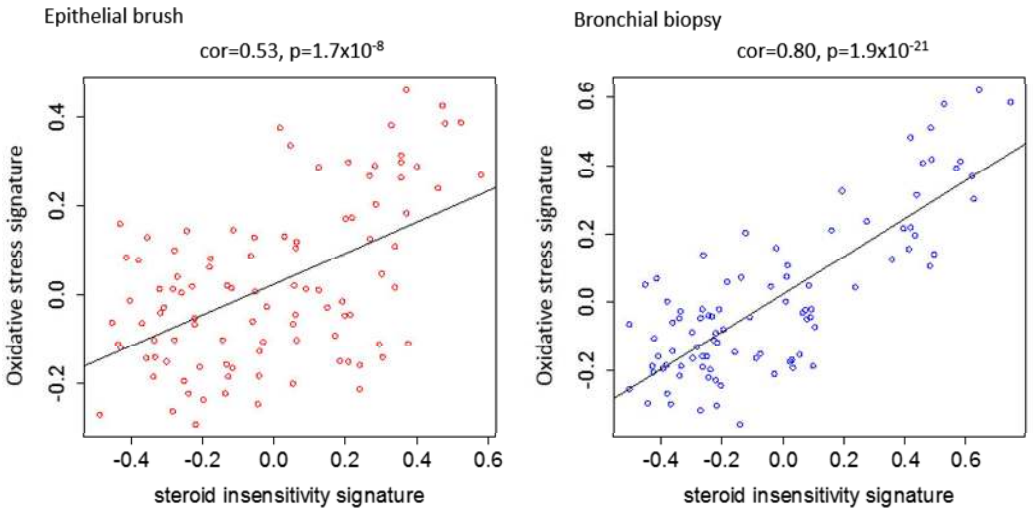
Figure S2



Relationship between oxidative stress (Ozone.Air.MM.IVV.UP) and adaptive immunity signatures. In bronchial biopsy, oxidative stress showed high correlation with (A) T cell signature (Tcell.activated.HS.IVS; Pearson's *r*: 0.79, $p=1.3\text{x}10^{-20}$), moderate correlation with (B) Th1 signature (Th1.activated.HS.IVS.UP; Pearson's *r*: 0.50, $p=4.5\text{x}10^{-7}$) and mild correlation with (C) Th2 signature_B (Lung.brushings.Th2high.asthma.HS.IVV.UP; Pearson's *r*: 0.36, $p=4.6\text{x}10^{-7}$). In epithelial brushing, oxidative stress showed moderate correlation with (D) T cell signature (Tcell.activated.HS.IVS; Pearson's *r*: 0.58, $p=3.1\text{x}10^{-10}$), moderate correlation with (E) Th1 signature (Th1.activated.HS.IVS.UP; Pearson's *r*: 0.55, $p=3.7\text{x}10^{-9}$) and weak to negligible correlation with (F) Th2 signature_B (Lung.brushings.Th2high.asthma.HS.IVV.UP; Pearson's *r*: 0.16, $p=0.110$). There was also a significant correlation between the oxidative stress signature and the Th2

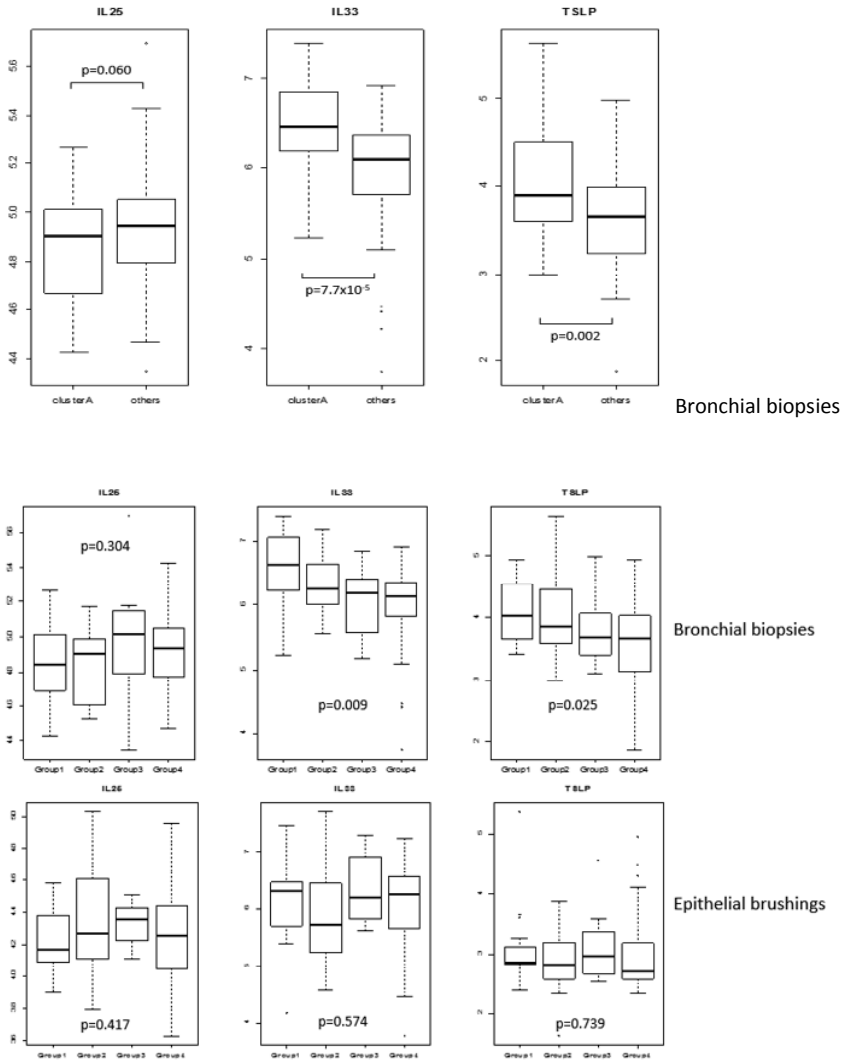
signature-A (Lung.biopsy.HDM.Rhesus.IVV.UP) in bronchial biopsies and epithelial brushings (data not shown).

Figure S3



Relationship between steroid insensitivity (Lung.biopsy.COPD.FLU.SAL.HS.IVV.DOWN) and oxidative stress (Ozone.Air.MM.IVV.UP) signature showing a moderate correlation in epithelial brushings (Pearson's *r*: 0.53, p=1.7x10^{-8}) and high correlationin bronchial biopsies (Pearson's *r*: 0.80, p=1.9 x10^{-21}).

Figure S4



Gene expression of IL-25, IL-33 and TSLP in bronchial biopsies according to cluster A and non-A (upper panel), and in bronchial biopsies and epithelial brushings according to the 4 Groups (lower panels). For bronchial biopsies, there was a significant difference in the level of gene expression for IL-33 and TSLP between Group1 and Group4 (p=0.003 for both) and between Group2 and Group4 (p=0.014 and p= 0.051, respectively). There was a difference in the level of gene expression of IL-33 between Group1 and Group3 (p=0.053).

Reference:

1. Lum, P.Y., G. Singh, A. Lehman, et al., *Extracting insights from the shape of complex data using topology.* Sci Rep, 2013. **3**: 1236.
2. Nicolau, M., A.J. Levine, and G. Carlsson, *Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival.* Proc Natl Acad Sci U S A, 2011. **108**: 7265-70.
3. Tibshirani, R., T. Hastie, B. Narasimhan, and G. Chu, *Diagnosis of multiple cancer types by shrunken centroids of gene expression.* Proc Natl Acad Sci U S A, 2002. **99**: 6567-72.
4. Abbas, A.R., D. Baldwin, Y. Ma, et al., *Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data.* Genes Immun, 2005. **6**: 319-31.
5. Haining, W.N., B.L. Ebert, A. Subrmanian, et al., *Identification of an evolutionarily conserved transcriptional signature of CD8 memory differentiation that is shared by T and B cells.* J Immunol, 2008. **181**: 1859-68.
6. Zhang, W., J. Ferguson, S.M. Ng, et al., *Effector CD4+ T cell expression signatures and immune-mediated disease associated genes.* PLoS One, 2012. **7**: e38510.
7. Zhang, H., C.E. Nestor, S. Zhao, et al., *Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1.* J Allergy Clin Immunol, 2013. **132**: 1005-8.
8. Bucasas, K.L., L.M. Franco, C.A. Shaw, et al., *Early patterns of gene expression correlate with the humoral immune response to influenza vaccination in humans.* J Infect Dis, 2011. **203**: 921-9.
9. Campbell, J.D., J.E. McDonough, J.E. Zeskind, et al., *A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK.* Genome Med, 2012. **4**: 67.
10. Tuller, T., S. Atar, E. Ruppin, M. Gurevich, and A. Achiron, *Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases.* Genes Immun, 2013. **14**: 67-82.
11. Bennett, L., A.K. Palucka, E. Arce, et al., *Interferon and granulopoiesis signatures in systemic lupus erythematosus blood.* J Exp Med, 2003. **197**: 711-23.
12. Pratt, A.G., D.C. Swan, S. Richardson, et al., *A CD4 T cell gene signature for early rheumatoid arthritis implicates interleukin 6-mediated STAT3 signalling,*

particularly in anti-citrullinated peptide antibody-negative disease. Ann Rheum Dis, 2012. **71**: 1374-81.

13. Tsitsiou, E., A.E. Williams, S.A. Moschos, et al., *Transcriptome analysis shows activation of circulating CD8+ T cells in patients with severe asthma.* J Allergy Clin Immunol, 2012. **129**: 95-103.
14. Abbas, A.R., J.K. Jackman, S.L. Bullens, et al., *Lung gene expression in a rhesus allergic asthma model correlates with physiologic parameters of disease and exhibits common and distinct pathways with human asthma and a mouse asthma model.* Am J Pathol, 2011. **179**: 1667-80.
15. Choy, D.F., B. Modrek, A.R. Abbas, et al., *Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways.* J Immunol, 2011. **186**: 1861-9.
16. Woodruff, P.G., H.A. Boushey, G.M. Dolganov, et al., *Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids.* Proc Natl Acad Sci U S A, 2007. **104**: 15858-63.
17. van den Berge, M., K. Steiling, W. Timens, et al., *Airway gene expression in COPD is dynamic with inhaled corticosteroid treatment and reflects biological pathways associated with disease activity.* Thorax, 2014. **69**: 14-23.
18. Yick, C.Y., A.H. Zwinderman, P.W. Kunst, et al., *Glucocorticoid-induced changes in gene expression of airway smooth muscle in patients with asthma.* Am J Respir Crit Care Med, 2013. **187**: 1076-84.
19. Hakonarson, H., U.S. Bjornsdottir, E. Halapi, et al., *Profiling of genes expressed in peripheral blood mononuclear cells predicts glucocorticoid sensitivity in asthma patients.* Proc Natl Acad Sci U S A, 2005. **102**: 14789-94.