

# Type 2-low asthma phenotypes by integration of sputum transcriptomics and serum proteomics

## To the Editor

Asthma is a complex heterogeneous disease that presents with varying degrees of severity. Analysis of the transcriptome of sputum cells from asthma patients has defined one T2-high and two T2-low clusters.<sup>1</sup> To obtain a better granularity of the T2-low clusters, we have performed a combined analysis of sputum transcriptomics with serum proteomics using an integrative machine learning workflow (Figure S1). We studied 104 patients with asthma and 16 normal healthy controls from the U-BIOPRED cohort<sup>2</sup> and re-analysed the 508 differentially expressed genes (DEGs) of sputum cell transcriptomics between high eosinophil and low eosinophil counts<sup>1</sup> by applying consensus clustering and finite Gaussian mixture model (GMM) clustering (Table S1; Appendix S1). We identified four optimal clusters (TAC\*1, TAC\*2, TAC\*3a and TAC\*3b) (Figure S3), in agreement with our previous clustering<sup>1</sup> where TAC\*3a and TAC\*3b were combined. Similar clusterings were generated using complete linkage hierarchical agglomerative clustering, K-means and partitioning around medoids. TAC\*1 patients were more severe asthmatics with a greater use of oral corticosteroid therapy, more nasal polyps and a higher level of sputum eosinophils compared with TAC\*2 (Table S2; Figure S4). TAC\*3a included more severe asthmatics on oral corticosteroid therapy, with a higher sputum neutrophilia, serum C-reactive protein levels compared with TAC\*3b patients. Metacore pathway analysis (<http://metacore.com>) using overrepresentation analysis and shrunken centroid method<sup>3</sup> indicated key-regulated immune pathways that distinguished the 4 TAC\*s. Gene set variation analysis (GSVA),<sup>4</sup> an enrichment analysis of specific gene signatures shown in Table S3, indicated that TAC\*1 had the highest expression score (ES) for the T2-high gene signature while TAC\*2 showed upregulation of cell-killing pathway via CD8+ T-cells, NK cells and macrophages, inflammasome activation and toll-like receptor (TLR) pathogen-sensing pathways. While TAC\*3a showed upregulation of pathways similar to TAC\*2, TAC\*3b showed attenuation of TLR sensor pathway activation with downregulation of cellular responses in asthma, malignancy and autoimmunity compared with TAC\*3a.

GSVA indicated that the IL13 gene signature ES was highest in TAC\*1 while the ILC1 signature was highly enriched in TAC\*2 followed by TAC\*1 and also exhibited higher ES scores in TAC\*3a compared with TAC\*3b (Figure 1, Figures S5 and S6). The inflammasome pathway and activated neutrophil ES scores were highly enriched in TAC\*2 and TAC\*3a compared with TAC\*3b. In contrast, the KEGG

oxidative phosphorylation pathway showed higher ES score in TAC\*3b compared to TAC\*3a, with lower enrichment in both TAC\*1 and TAC\*2. The enrichment of activated polarised human peripheral blood Th17 cell ES was enhanced in TAC\*3b, but there was no difference for ILC3 cell ES across the TAC\*s while ILC2 cell ES were enriched in TAC\*1 compared with other TAC\*s (Figures S5 and S6). These clusters were validated in the independent ADEPT<sup>5</sup> and Australian<sup>6</sup> asthma cohorts by comparing the ESs of the signature genes of each of the 4 TAC\*s across the 3 cohorts. The distribution of the ESs was not significantly different (Figure S7).

Sputum transcriptomics and serum proteomics from 92 asthmatic subjects using SomaScan technology (SomaLogic)<sup>7</sup> and shotgun proteomics using liquid chromatography-mass spectrometry (LC-MS/MS)<sup>8</sup> were then integrated using similarity network fusion (SNF) (Figure S2). SNF is an efficient integrative method for different data sets that construct similarity networks based on each data set and then integrates these networks into a single network that represents all underlying data-types.<sup>9</sup> By pre-selecting the number of clusters to 4, we reproduced these 4 TAC\*s with a high degree of concordance (81%) when fusing with proteomic data, indicating the robustness of the SNF approach. Allowing the data to freely cluster, we identified 9 proteomic- and transcriptomic-associated clusters (PTACs) derived from the 4 TAC\* clusters (Figure 2; Figures S8 and S9). PTAC2a (n = 14) showed autoimmune pathways related to multiple sclerosis and systemic lupus erythematosus (SLE) while PTAC2b (n = 7) enrichment of TLR pathways, type I interferon and inflammasome activation, responses associated with infective pathogens. PTAC2a subjects had evidence for macrophage, NK and TC1 cell activation with enhancement of IL-2, IL-6 and apoptotic pathways while PTAC2b was associated with enhanced type I IFN and inflammasome activation. PTAC2a patients had the highest percentage of patients with severe asthma who were smokers and had a more frequent history of pneumonia and prevalence of atopy, compared with PTAC2b. PTAC\*3a1, PTAC\*3a2 and PTAC\*3a3 participants showed clear reduction in T2 pathways compared with TAC\*1. Furthermore, PTAC3a1 (n = 14) and PTAC3a2 (n = 6) participants were differentiated with respect to severe asthma, prevalence of nasal polyps, smoking history, prevalence of eczema and a history of pneumonia with PTAC3a2 with the highest percentage of severe asthma who were smokers or ex-smokers, with a more frequent history

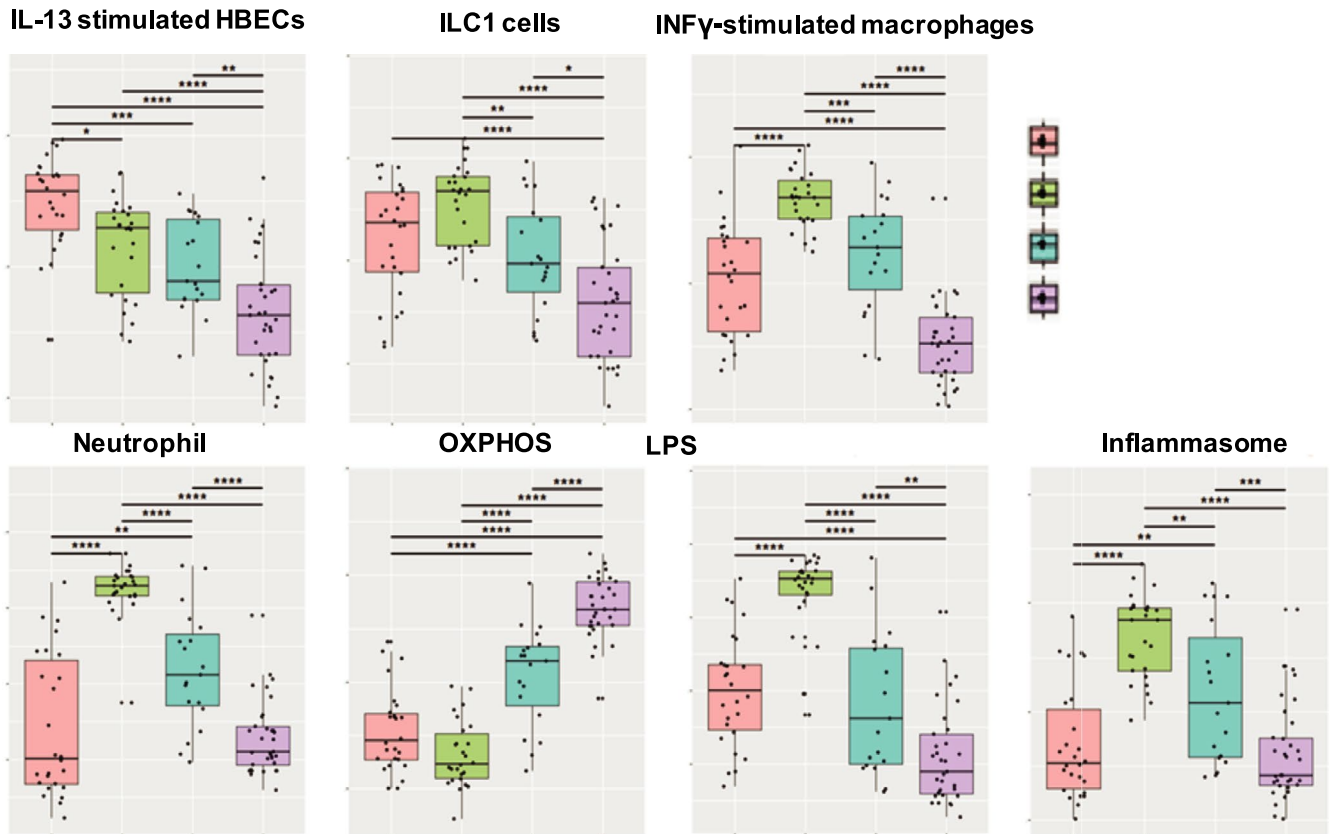


**FIGURE 1** Gene set variation analysis. Dot plot enrichment scores with box-and-whisker plots showing median and interquartile range for 7 different gene signatures in the transcriptome-associated clusters: TAC\*1 (n = 26), TAC\*2 (n = 26), TAC\*3a (n = 19) and TAC\*3b (n = 33). The signatures reflect activation pathways for IL-13 activation of human airway epithelial cells (IL-13-stimulated human bronchial epithelial cells), innate lymphoid cells (ILC1 cells), inflammasome activation, oxidative phosphorylation (OXPHOS), neutrophil activation, lipopolysaccharide (LPS)-stimulated macrophage and interferon- $\gamma$ -stimulated macrophage. \*\*\*\* $P < .0001$ , \*\*\* $P < .001$ , \*\* $P < .01$ , \* $P < .05$

of pneumonia, prevalence of atopy and nasal polyps, compared with PTAC3a1 subjects. PTAC3b1 (n = 10), PTAC3b2 (n = 6) and PTAC3b3 (n = 8) had different levels of blood leukocytes, blood neutrophils and C-reactive protein, and those on oral corticosteroids. TAC\*3 primarily comprised of subjects associated with OXPHOS and ageing pathways but their delineation remained less clear than PTAC2. Reduced OXPHOS compared with healthy controls in TAC1, TAC2 and TAC3a indicate a potential switch to glycolytic processes but TAC3b with significantly increased OXPHOS respond to distinct mitochondrial or metabolic treatment modalities (Figure S10). However, this analysis is limited by the small number of subjects within the PTAC groups and larger validation studies will be required to ascertain the identification of distinct pathways. Interestingly, TAC1 remains undivided, indicating that serum proteomics did not bring any factors that allowed for subclusters.

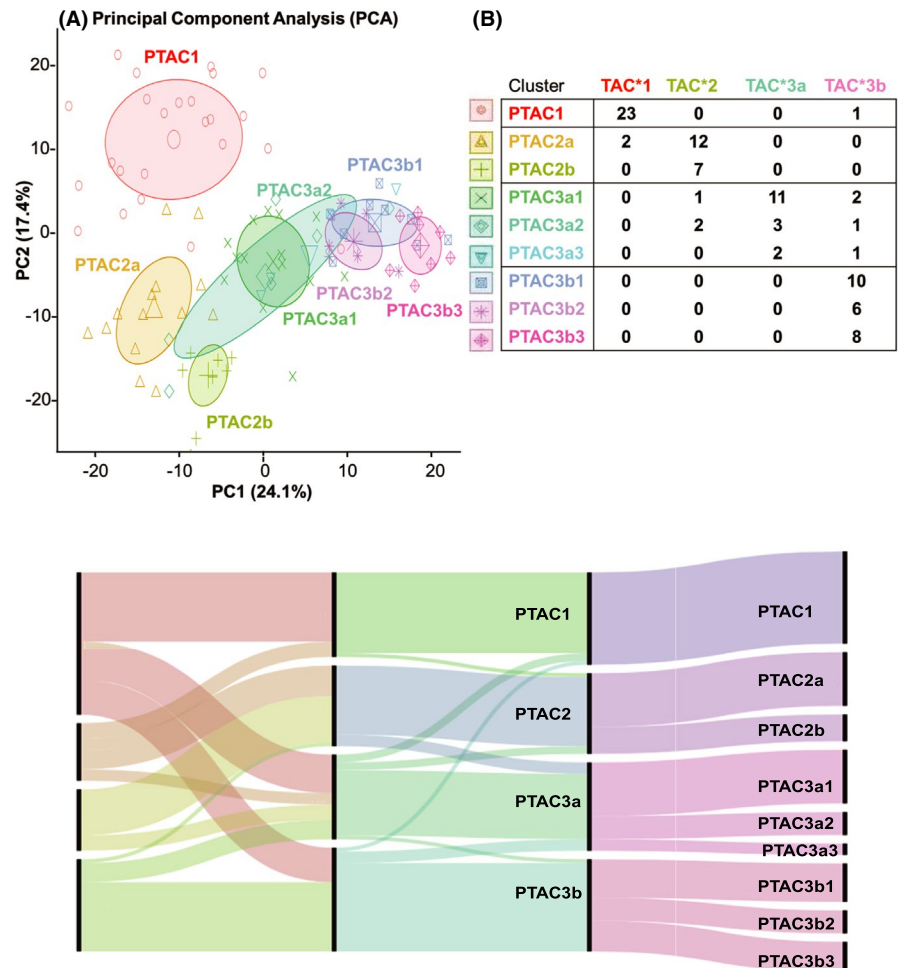
We have defined 9 molecular clusters, distinguished by different combinations of ES of distinct pathways that included T2-high and T2-low pathways, and associated with differences in clinical and inflammatory characteristics. T2-low asthma is a heterogeneous condition of 8 clusters compared with the homogeneity of

the single T2-high cluster. Thus, the addition of the 2 proteomic platforms to the analysis has allowed us to provide a greater granularity of the T2-low clusters. The existence of multiple T2-low asthma phenotypes is challenging with requirements for specific biomarkers and different targeted therapies compared to the single T2-high cluster.

Another important finding in this analysis is the repartition of various potential pathways involved in asthma in the subclusters generated in this study, as assessed by GSVA. Although the IL-13-T2-signature that we used indicated the highest expression in TAC1, it was also seen in individuals in TAC2. Other pathways that could potentially drive asthma pathobiology include ILC1 in TAC1 and TAC2, inflammasome, neutrophil and IL-17 activation in TAC2, OXPHOS in TAC3a and TAC3b and IFN $\gamma$  activation in TAC2 or TAC3a.

Finally, one limitation of our study is the unknown effects of corticosteroid therapy particularly oral prednisolone on different inflammatory pathways such as the T2-high, thus representing a confounding factor. There is also a degree of corticosteroid insensitivity since that T2-high TAC1 cluster had the highest level of eosinophilia despite 58% of the group being on daily oral

**FIGURE 2** Integration of sputum transcriptomics and serum proteomics. (A) Principal component analysis plot identified 9 transcriptomic-associated clusters. The dots represent patients and colour code based on cluster subgroups. TAC\*1 remains intact shown as PTAC\*1; TAC\*2 is divided into two subgroups PTAC\*2a and PTAC\*3b. TAC\*3a and TAC\*3b were divided into 3 subgroups each. (B) shows patients in each subgroup. (C) Sankey diagram of the directed flow of patients based on the granulocytic state (Level A), transcriptomic-associated clusters (TAC\*s) (Level B), proteomics- and transcriptomic-associated clusters (PTACs) (Level C) and PTACs with 9 clusters (Level D)



corticosteroids (OCS). Alternatively, this might represent the most severe patients with asthma. The effect of this therapy on molecular clustering can only be determined by studying the effect of OCS in this population.

## KEYWORDS

asthma, bioinformatics, endotypes, precision medicine, systems biology

## CONFLICT OF INTEREST

Ratko Djukanović has consulted and presented at symposia organised by TEVA, Novartis, GlaxoSmithKline and AstraZeneca and has shares in and consults for Synairgen; Dr Asa Wheelock report remuneration from AstraZeneca and Harvard Medical School for speaking engagements on SNF-clustering in COPD.; Charles Auffray reports grants from Innovative Medicine Initiative; Kian Fan Chung has received honoraria for participating in Advisory Board meetings of the pharmaceutical industry regarding treatments for asthma and chronic obstructive pulmonary disease and has also been remunerated for speaking engagements; Ian Adcock has received grants from Advisory Board meetings with pharmaceutical companies GSK, A-Z, Novartis, Boehringer Ingelheim and Vectura, and grants on asthma and COPD from Pfizer, GSK, MRC, EU, BI and IMI; Peter Sterk reports

grants from IMI Innovative Medicines Initiative, during the conduct of the study; Matthew Loza and Frederic Baribaud are Employees and Shareholders of Janssen Research and Development, a Johnson and Johnson company; John Riley and Ana R Sousa are employees of GSK; the rest of the authors have nothing to disclose.

## FUNDING INFORMATION

U-BIOPRED was supported by an Innovative Medicines Initiative Joint Undertaking (No.115010), resources from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution ([www.imi.europa.eu](http://www.imi.europa.eu)).

Nazanin Zounemat Kermani<sup>1</sup>

Mansoor Saqi<sup>1</sup>

Paul Agapow<sup>1</sup>

Stelios Pavlidis<sup>1</sup>

Chihhsi Kuo<sup>1,2</sup>

Kai Sen Tan<sup>3</sup>

Sharon Mumby<sup>2</sup>

Kai Sun<sup>1</sup>

Matthew Loza<sup>4</sup>

Frederic Baribaud<sup>4</sup>

Ana R. Sousa<sup>5</sup>

John Riley<sup>5</sup>  
Asa M. Wheelock<sup>6</sup>  
Craig E. Wheelock<sup>6</sup>  
Bertrand De Meulder<sup>7</sup>  
Jim Schofield<sup>8,9</sup>  
Stephany Sánchez-Ovando<sup>10</sup>  
Jodie Louise Simpson<sup>10</sup>  
Katherine Joanne Baines<sup>10</sup>  
Peter A. Wark<sup>10</sup>  
Charles Auffray<sup>7</sup>  
Sven-Erik Dahlen<sup>6</sup>  
Peter J. Sterk<sup>11</sup>  
Ratko Djukanovic<sup>8,9</sup>  
Ian M. Adcock<sup>1,2</sup>  
Yi-ke Guo<sup>1</sup>  
Kian Fan Chung<sup>1,2</sup>   
U-BIOPRED Project Team

<sup>1</sup>Department of Computing & Data Science Institute, Imperial College London, London, UK

<sup>2</sup>National Heart & Lung Institute, Imperial College London, London, UK

<sup>3</sup>Department of Otolaryngology, National University of Singapore, Singapore, Singapore

<sup>4</sup>Janssen Research and Development, High Wycombe, UK

<sup>5</sup>Respiratory Therapeutic Unit, GSK, Stockley Park, UK

<sup>6</sup>Department of Medicine Solna & Center for Molecular Medicine & Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden

<sup>7</sup>European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France

<sup>8</sup>Faculty of Medicine, Southampton University, Southampton, UK

<sup>9</sup>NIHR Southampton Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton, UK

<sup>10</sup>Priority Research Centre for Healthy Lungs, Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW, Australia

<sup>11</sup>Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

## Correspondence

Kian Fan Chung, National Heart & Lung Institute, Imperial College London, Dovehouse St, London SW3 6LY, UK.  
Email: f.chung@imperial.ac.uk

## ORCID

Kian Fan Chung  <https://orcid.org/0000-0001-7101-1426>

## REFERENCES

1. Kuo CS, Pavlidis S, Loza M, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. A Transcriptome-driven Analysis of Epithelial Brushings and Bronchial Biopsies to Define Asthma Phenotypes in U-BIOPRED. *Am J Respirat Crit Care Med*. 2017;195(4):443-455.
2. Shaw DE, Sousa AR, Fowler SJ, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J*. 2015;46(5):1308-1321.
3. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA*. 2002;99(10):6567-6572.
4. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
5. Loza MJ, Djukanovic R, Chung KF, et al. Validated and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. *Respir Res* 2016;17(1):165.
6. Baines KJ, Simpson JL, Wood LG, et al. Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. *J Allergy Clin Immunol*. 2014;133(4):997-1007.
7. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010;5(12):e15004.
8. Burg D, Schofield JPR, Brandsma J, et al. Large-Scale Label-Free Quantitative Mapping of the Sputum Proteome. *J Proteome Res*. 2018;17(6):2072-2091.
9. Wang BO, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods* 2014;11(3):333-337.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

## **Type 2-low asthma phenotypes by integration of sputum transcriptomics and serum proteomics**

### ***Supplementary Materials and Methods and Results, Supplementary Tables S1 to S3 & Supplementary Figures S1-S10***

#### **Supplementary Material and Methods**

##### **Datasets**

*Sputum transcriptomics.* Sputum plugs induced by hypertonic saline solution were collected. Expression profiling was obtained using Affymetrix U133 Plus 2.0 microarrays (Affymetrix, Santa Clara, USA). Analysis was performed as previously described (1).

*Serum proteomics.* The SOMAscan proteomics assay on serum is an array-based method measuring 1,129 proteins each assay run (2).

*Shotgun proteomics.* Sample preparation for shotgun proteomics, with preprocessing and annotation are described elsewhere (3).

##### **Disease subtyping workflow overview**

The disease subtyping workflow is shown on **Supplementary Figure S1**. The processed gene expression data matrix (104 patients x 508 genes) used in the study of Kuo et al (1) was the starting point for this study. Subtyping of the cohort was performed using a finite Gaussian mixture model (GMM) (from the R package *mclust*(4)) and consensus clustering (R package ConsensusClusterPlus (5)). Finite GMM clustering is a model-based clustering. The underlying rationale of this method is that the probability distribution of a sample set is a linear function of the number of component probability distributions (G). **Supplementary Table S1** describes the component probability distributions that were used.

##### **Integration of sputum transcriptomics and serum proteomics**

Data from asthmatic subjects with serum Somalogics (92 asthmatic subjects by 1129 features), serum LC-MS/MS (92 asthmatic subjects by 144 features) proteomics and sputum transcriptomics (92 asthmatic subjects by 508 features) were integrated using similarity network fusion (R package SNFTools(6)) as shown in **Supplementary Figure S2**.

##### **Clustering of baseline and follow-up transcriptomics data**

A finite GMM was applied to the transcriptomics data after DEGs analysis. It was tested for numbers of clusters ranging from 1 to 9, with nine being sufficiently large compared to the expected numbers of clusters within the data. The volume, shape and orientation are geometric characteristics of the within-group covariance matrix, e.g. EII; E stands for equal, I for Coordinate axes and V for variable (**Supplementary Table S1**). The majority of models identified 4 clusters as the most likely number of clusters in the dataset. The best model (VII) was spherical with varying volume comprising 4 components as indicated by the dotted red line that crosses the BIC score (**Supplementary Figure S3-A, B**). The DIS index applied on the outcome of the consensus clustering (**Supplementary Figure S2-C, D**) also identified the optimal number of clusters as four (**Supplementary Figure S3-E**). These four clusters were designated TAC\*1, TAC\*2, TAC\*3a and TAC\*3b. The validity of this clustering formation (i.e. TAC\*s) was then tested, as following: first, by comparing it to a previous study (1) and second, by comparing the TAC\*s with the clustering outcome of three other clustering techniques.

## Supplementary information

Hierarchical K-means and clustering (partitioning) around medoids are clustering techniques that have been successfully applied to many applications alongside Gaussian mixture clustering (7). Hierarchical clustering with complete linkage function and Euclidean distance were used. K-means and clustering around medoids with Euclidean distance were applied. Hierarchical clustering with four as the number of clusters inflated TAC\*3a and TAC\*3b while splitting TAC\*1. Therefore, to be able to make a more meaningful comparison, the number of clusters was set to four. For K-means and clustering around medoids, the number of clusters was fixed to four. A Chi-squared test for independence was used to determine whether there is a significant association between the two variables. It is applied to two categorical variables derived from a single population. The null hypothesis ( $H_0$ ) for this test is that two variables are independent, and the alternative two variables are not independent.

A random dataset was generated by randomly sampling in the range of [0-1] from a uniform distribution. Finite GMM, hierarchical, K-means and clustering around medoids methods were applied to the random data. The same parameters were used for clustering, and the number of clusters was fixed to 4. The labels generated for the patients were used for comparison with TAC\*s by the Chi-squared test and adjusted rand index.

The P values for the Chi-square test were  $<0.00001$ ,  $<0.00001$   $<0.00001$  between finite GMM clustering labels and hierarchical, K-means and clustering around medoids, respectively. The adjusted rand indices were 0.5812, 0.6193 and 0.6193 for comparison of finite GMM clustering with hierarchical, K-means and clustering around medoids, respectively. In contrast, the randomly generated dataset generated the P values of 0.5812, 0.6193 and 0.6193 for comparison of finite GMM clustering with hierarchical, K-means and clustering around medoids, respectively. Very small adjusted Rand indices equal to 0.0083, 0.0038 and 0.0078.

**Supplementary Table S1-** Parameterisations of the within-group covariance matrix  $\Sigma_k$  for multi-dimensional data and corresponding geometric characteristics, i.e. volume, shape and orientation.

| Model | $\Sigma_k$      | Distribution | Volume   | Shape    | Orientation     |
|-------|-----------------|--------------|----------|----------|-----------------|
| EII   | $\lambda I$     | Spherical    | Equal    | Equal    | -               |
| VII   | $\lambda_k I$   | Spherical    | Variable | Equal    | -               |
| EEI   | $\lambda A$     | Diagonal     | Equal    | Equal    | Coordinate axes |
| VEI   | $\lambda_k A$   | Diagonal     | Variable | Equal    | Coordinate axes |
| EVI   | $\lambda A_k$   | Diagonal     | Equal    | Variable | Coordinate axes |
| VVI   | $\lambda_k A_k$ | Diagonal     | Variable | Variable | Coordinate axes |

$\lambda_k$  is a scalar that determines the volume of the ellipsoid.  $A_k$  is a diagonal matrix controlling the shape of the density contours. Adapted from (8). EII; E stands for equal, I for Coordinate axes and V for variable

**Supplementary Table S2. Clinical features of four sputum transcriptome-associated clusters (TAC\*s)**

|                                                                    | TAC*1            | TAC*2               | TAC*3a           | TAC*3b             | Group comparisons Adjusted P-Value |        |        |        |        |         |
|--------------------------------------------------------------------|------------------|---------------------|------------------|--------------------|------------------------------------|--------|--------|--------|--------|---------|
|                                                                    |                  |                     |                  |                    | 1 v 2                              | 1 v 3a | 1 v 3b | 2 v 3a | 2 v 3b | 3a v 3b |
| Subjects                                                           | 26               | 26                  | 19               | 33                 |                                    |        |        |        |        |         |
| Age years                                                          | 53(46.2-62,26)   | 55(42.8-59.8,26)    | 52(43.5-65,19)   | 52(41-56,33)       | 0.6                                | 0.6    | 0.6    | 0.6    | 0.6    | 0.6     |
| Age of onset years                                                 | 37.5(17-49,26)   | 15(3.5-26,26)       | 27(13-46.5,19)   | 16(4.8-39.2,32)    | 0.02                               | 0.6    | 0.07   | 0.07   | 0.5    | 0.3     |
| Female                                                             | 16(62%,26)       | 16(62%,26)          | 10(53%,19)       | 18(55%,33)         | 1                                  | 0.6    | 0.6    | 0.6    | 0.6    | 1       |
| BMI                                                                | 25.5(22-31,26)   | 26.6(24-30.8,26)    | 29.6(26-32,19)   | 26(23.2-31.8,33)   | 0.5                                | 0.5    | 0.8    | 0.6    | 0.6    | 0.5     |
| FEV1 (% pred, L)                                                   | 52(37.3-66.9,26) | 66.2(47.1-87,26)    | 69.9(61.6-83,19) | 79.4(68.1-91.4,33) | 0.09                               | 0.01   | <.001  | 0.4    | 0.04   | 0.09    |
| Severe asthma                                                      | 26(100%,26)      | 22(85%,26)          | 16(84%,19)       | 20(61%,33)         | <.001                              | <.001  | <.001  | 1      | <.001  | 0.001   |
| Oral corticosteroid use (N, %)                                     | 14(54%,26)       | 9(35%,26)           | 6(33%,18)        | 6(18%,33)          | 0.02                               | 0.01   | <.001  | 0.9    | 0.02   | 0.02    |
| Exacerbations (previous year)                                      | 2(1-3.8,26)      | 2(1-3,26)           | 1(0-2,19)        | 1(0-2,33)          | 0.5                                | 0.09   | 0.09   | 0.2    | 0.1    | 1       |
| History of pneumonia                                               | 9(36%,25)        | 14(56%,25)          | 4(21%,19)        | 6(19%,32)          | 0.01                               | 0.04   | 0.02   | <.001  | <.001  | 0.9     |
| smoker                                                             | 15(58%,26)       | 19(73%,26)          | 10(53%,19)       | 20(61%,33)         | 0.1                                | 0.7    | 0.8    | 0.03   | 0.2    | 0.4     |
| Atopy                                                              | 19(73%,26)       | 17(65%,26)          | 12(71%,17)       | 27(82%,33)         | 0.4                                | 0.9    | 0.4    | 0.5    | 0.06   | 0.3     |
| Eczema                                                             | 6(25%,24)        | 14(58%,24)          | 3(16%,19)        | 10(31%,32)         | <.001                              | 0.2    | 0.4    | <.001  | <.001  | 0.03    |
| Nasal polyps                                                       | 12(50%,24)       | 7(29%,24)           | 8(42%,19)        | 7(21%,33)          | 0.008                              | 0.3    | <.001  | 0.1    | 0.3    | 0.006   |
| Allergic rhinitis                                                  | 10(48%,21)       | 12(55%,22)          | 6(35%,17)        | 14(44%,32)         | 0.5                                | 0.3    | 0.7    | 0.04   | 0.3    | 0.3     |
| Total IgE (IU·mL <sup>-1</sup> )                                   | 176(74-374,25)   | 87.8(22.5-160.5,26) | 105.3(37-306,18) | 86(45-167.5,31)    | 0.1                                | 0.4    | 0.1    | 0.9    | 0.9    | 0.9     |
| Sputum eosinophils (%)                                             | 34.3(22.5-53,26) | 0.6(0.2-2.2,26)     | 2.7(0.2-4.8,19)  | 0.8(0-4,33)        | <.001                              | <.001  | <.001  | 0.2    | 0.8    | 0.3     |
| Sputum neutrophils (%)                                             | 44.1(25-68.8,26) | 88.7(84.9-93.4,26)  | 60.8(50-71.5,19) | 41.1(29.5-50.9,33) | <.001                              | 0.08   | 0.6    | <.001  | <.001  | 0.001   |
| Sputum Macrophages (%)                                             | 15.1(9.7-22,26)  | 8.1(4.2-13.4,26)    | 33.4(24-41.6,19) | 51.1(40.6-64.3,33) | 0.009                              | <.001  | <.001  | <.001  | <.001  | <.001   |
| Blood leukocytes (x10 <sup>3</sup> μL <sup>-1</sup> ) <sup>1</sup> | 9(7.3-11,26)     | 8(6.1-9.8,26)       | 7.6(6.3-9.9,18)  | 6.4(5.7-7.7,32)    | 0.2                                | 0.2    | 0.02   | 0.8    | 0.2    | 0.2     |
| Blood eosinophils (x10 <sup>3</sup> μL <sup>-1</sup> )             | 0.4(0.2-0.6,26)  | 0.2(0.1-0.4,26)     | 0.3(0.2-0.4,18)  | 0.2(0.1-0.3,32)    | 0.02                               | 0.1    | 0.02   | 0.3    | 0.6    | 0.1     |
| Blood neutrophils (x10 <sup>3</sup> μL <sup>-1</sup> )             | 5.1(3.9-8.2,26)  | 4.6(3.9-6.5,26)     | 4.3(3.8-6.2,18)  | 3.7(3.1-4.9,32)    | 0.5                                | 0.2    | 0.05   | 0.5    | 0.07   | 0.2     |
| FeNO (ppb)                                                         | 29.5(20.5-61,26) | 22(14-26.5,25)      | 30.5(21-40,19)   | 27.5(17.8-44.5,31) | 0.2                                | 0.7    | 0.5    | 0.2    | 0.4    | 0.7     |
| Serum periostin (ng·mL <sup>-1</sup> )                             | 58.6(49-71.1,23) | 48.5(41.5-52.8,24)  | 47.8(39-63.7,12) | 42.5(36.3-53.2,27) | 0.1                                | 0.5    | 0.03   | 0.8    | 0.4    | 0.4     |
| CRP (mg·L <sup>-1</sup> )                                          | 1.9(0.9-3,26)    | 3.6(0.8-7.4,26)     | 2.6(1.4-5.5,18)  | 0.8(0.4-2,32)      | 0.3                                | 0.3    | 0.04   | 0.7    | 0.03   | 0.03    |

Data are shown as n (%), mean± Standard deviation, or median (interquartile range 25%-75%). BMI: body mass index; FEV<sub>1</sub>: forced expiratory volume in 1 second; FeNO: exhaled nitric oxide fraction; CRP: C-reactive protein; p value:  $\chi^2$  test was used for categorical variables, for numerical variable ANOVA test (Kruskal-Wallis test) for normally distributed variable (and non-normally distributed variable) was used. P values were corrected for multiple comparisons using Benjamini-Hochberg correction. A value of 0 indicates  $p \leq 10^{-4}$ .

**Supplementary Table S3-** Gene sets representing specific disease mechanisms or pathways.

| Signature names             | Genes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Ref  |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| IL-13, stimulate d, HBECS   | CST1, CCL26, PRB2, PRB1, PRB3, POSTN, PRB4, ITLN1, ALOX15, SH2D1B, CA2, NOS2, FCGBP, FOXA3, SPDEF, CAPN14, DUOX2, CLDN5, PADI3, TSPAN8, ALPL, KCNJ16, FETUB, B3GNT6, CDH26, LRRC31, MUC13, VSIG2, CSTA, FAM3B, SLC9B2, NTRK1, KLF4, HPDL, SOCS1, TRNP1, HS3ST1, VWF, DUOX2, CISH, ATP13A5, ZNF808, RNASE4, CCB1, SDCBP2, TMPRSS2, HYAL1, CCDC109B, FAM83D, TRAK1, TPK1, SLC7A1, CYP2C18, CDC42EP5, KCNS3, ADRA2A, MRAP2, SLC2A10, PPARG, FAM26E, ADCY4, WNT3, SLC04A1, ALDH1A2, C10orf99, WDFY2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | (9)  |
| ILC1, cells                 | SIT1, CD3D, CD3G, CD4, CD6, TRAV13-1, CD5, CD27, C14orf64, COTL1, CD8A, IKZF3, LITAF, CCR7, TRAV8-2, TRAV4, SYNE2..1, MAL, GZMM, GZMK, TC2N, GZMA, SH2D1A, IFNG-AS1, PASK, TRBV5-1, ADTRP, TRAV9-2, CACNA1I, CCL5, ACTN1, CXCR3, BCL11B, PYHIN1, CH25H, LBH, FBLN7, LINC00402, TRAV2, TRBV2, IGFBP3, ANK3, IL6R, LDLRAP1, ACSL6, TRAV41, MIAT, TRBV20-1, LAG3, IFNG, TRAV26-2, GABBR1, TSHZ2, SLC25A4, AP000569.8..1, RASGRF2, TRAV8-4, RP11-664D1.1, TNFRSF10D, PLEKHB1, TRAV12-2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | (10) |
| Inflammasome                | IL1B, NLRP3, CASP1, CASP4, CASP5                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | (11) |
| OxPHOS                      | ND1, ND2, ND3, ND4, ND4L, ND5, ND6, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS5, NDUFS6, NDUFS7, NDUFS8, NDUFV1, NDUFV2, NDUFV3, NDUFA1, NDUFA2, NDUFA3, NDUFA4, NDUFA4L2, NDUFA5, NDUFA6, NDUFA7, NDUFA8, NDUFA9, NDUFA10, NDUFAB1, NDUFA11, NDUFA12, NDUFA13, NDUFB1, NDUFB2, NDUFB3, NDUFB4, NDUFB5, NDUFB6, NDUFB7, NDUFB8, NDUFB9, NDUFB10, NDUFB11, NDUFC1, NDUFC2, NDUFC2-KCTD14, SDHA, SDHB, SDHC, SDHD, UQCRCF51, CYTB, CYC1, UQCRC1, UQCRC2, UQCRH, UQCRHL, UQCRB, UQCRC, UQCRC10, UQCRC11, COX10, COX3, COX1, COX2, COX4I2, COX4I1, COX5A, COX5B, COX6A1, COX6A2, COX6B1, COX6B2, COX6C, COX6D, COX7A1, COX7A2, COX7B, COX7B2, COX7C, COX8C, COX8A, COX11, COX15, COX17, ATP5A1, ATP5B, ATP5C1, ATP5D, ATP5E, ATP5O, ATP6, ATP5F1, ATP5G1, ATP5G2, ATP5G3, ATP5H, ATP5I, ATP5J2, ATP5L, ATP5J, ATP8, ATP6V1A, ATP6V1B1, ATP6V1B2, ATP6V1C2, ATP6V1C1, ATP6V1D, ATP6V1E2, ATP6V1E1, ATP6V1F, ATP6V1G1, ATP6V1G3, ATP6V1G2, ATP6V1H, TCIRG1, ATP6V0A2, ATP6V0A4, ATP6V0A1, ATP6V0C, ATP6V0B, ATP6V0D1, ATP6V0D2, ATP6V0E1, ATP6V0E2, ATP6AP1, ATP4A, ATP4B, ATP12A, PPA2, PPA1, LHPP                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | (12) |
| Neutrophil                  | ABT81, AMPD2, C5orf6, CCR3, CDA, CKLFSF2, CLC, CREB5, CTBS, DcR1, EST, FCGR2B, FCGR3B, FLJ10298, FRP11, FRAT2, GPR27, GPR43, HSPA6, IL8RA, IL8RB, KIAA0779, KIAA1126, KRT23, LENG4, LENG5, MAD, MGC10500, MGC14126, MGC16353, MPPE1, MSCP, NCF4, NRBF2, PHC2, PROK2, RALB, RNF141, SEPX1, STX3A, TM4-B, VMP1, VNN2, XPO6                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | (13) |
| LPS-stimulated, macrophage  | MLLT4, ETL2, NMT2, SLC43A3, IL4I1, MCOLN2, MAOA, ZFP36L1, APOBEC3A, TXN, FTH1, CCL7, ACSL1, CCL23, VMH8, RBM47, RPLP1, HSP90B3P, RHBDF2, RPN2, HN RNPC, HMGN2P46, SDF4, FABP5, ANXA2P1, TNIP1, TNIP3, LILRB1, TIMP1, PSTPIP2, MARCKSL1, ADAMDEC1, SLC3A2, SRC, NMI, CYP27B1, KYNU, WSB2, SLC41A2, IL3RA, S100A12, VRK2, YWHAG, CXCL5, ATP2C1, NFS1, BID, MTF1, PPBP, PPP2CB, PDE4DIP, CCL5, SLC2A6, CXCL6, CD40, FABP5P3, IFIH1, LAD1, SDC2, GPR84, C1orf122, LILRA3, CFLAR, CAPZA1, LGALS3, GCH1, EIF1AX, P2RX4, NFE2L3, SLBP, CLIC4, MMP12, ACP2, MYC, ARL6IP1, ATP11B, ZSWIM4, VPS37A, SRGN, FTH1P3, TDP2, OAS3, C6orf211, SMAD7, SMARCA5, HSPA13, GBBP1, PARP10, BOLA2, NUP62, IL7R, G3BP1, LARP4, LILRB4, ILK, SLC16A3, GRINA, C1QTNF1, SOD2, FAM129B, RPS29, TPI1, LOC440704, HSPB1, MT1H, EXOG, CD68, PTMA, MT1X, KIAA1147, PDIA3P1, MT1P, MT1E, GYPC, TPT1, SLC29A1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | (14) |
| IFNG-stimulated, macrophage | SELM, EHD1, IL10RA, NAMPT, MCTP1, C15orf48, SLC7A11, SIGLEC10, CDK5RAP2, EREG, TERF2IP, IL1A, ISG20, MYH11, MCOLN2, CLCF1, DDIT4, ANKRD12, AARS, SQSTM1, FLNB, NUB1, DCUN1D3, CXCL8, RILPL2, DRAM1, ATOX1, PTGS2, MAP3K4, UPB1, ADARB1, CREG1, PTPN1, RNF19B, IER3, NKG7, FAM60A, IL2RA, SLC7A1, MT1H, SAMSNA1, ASCC1, FYN, MT2A, ETS1, PPP3CC, MT1A, LAT, RHBDF2, G0S2, CEP135, PANX2, C5orf15, HMGN2P46, ORM1, TRAF1, C17orf58, MB21D2, LPAR1, TRAF3IP2, TNFAIP3, SEMA4A, ARHGEF2, LMO4, TNIP1, TBC1D9, ACSL5, CYP2S1, BTG3, TNIP3, EXOG, NLRP3, SEMA6B, PSTPIP2, PTGES, MARCKSL1, CD86, NFKB1, LSS, AQP9, TANK, CCR7, MT1G, LAMB3, SLC25A37, RUNX2, ADAMDEC1, NFKBIE, SYT11, PVRL2, MLLT6, MT1F, ETS2, SLC7A5, CA12, MRP56, GK, SLC3A2, BHLHE41, WARS, TNFAIP2, CXCL16, PILRA, CCDC115, TMEM205, HIST2H2BE, SIK3, ITGB8, PPP1R18, PTPRE, NBN, HIVEP1, GRAMD1A, KYNU, SLC1A3, NFE2L1, PCNX, CD14, CYP27B1, APBB3, FPR2, CEP350, CTTN, IRAK2, DDX59, CCM2L, PIM2, TSPAN33, ACO1, TAP1, RNF144B, CXCL5, IKKB, EMR3, APOL3, H1FO, NADK, ZMIZ2, NFS1, HLA-F, BATF3, HIF1A, MTF1, YARS, STOM, RASGRP1, HIST2H2AC, GJB2, SLAMF1, CLEC4E, CCL5, KIAA0226L, GNA15, CFB, AMPD3, PNKD, MT1X, SMOX, IL1B, BIRC3, MAP1LC3A, ZMYND15, SLC2A6, SPECC1L, OSM, PRKCH, C1S, LAD1, PDE4B, STAT4, GPR84, TNFAIP6, C1orf122, PLAGL2, SLC43A2, HELZ2, TSC22D1, SLC5A3, CXCL1, HEY1, ABHD17C, CRISPLD2, DNAJB6, PGD, NDP, MMP9, ASNS, IRF7, ACKR3, RCN1, SLC9A1, RFTN1, EB13, TGFA, PPA1, ADORA2A, CLEC4D, MUCL1, GPR132, RELB, FAM20A, GCH1, CTSL, NINJ1, ST3GAL1, SMAP2, MYO1B, TYMP, TNFRSF4, NCF1C, GBP5, HBEGF, CYB561A3, MT1P, ZSWIM4, IL36G, LAMP3, MT1E, ZHX2, PFKFB3, N4BP1, MXD1, GNA12, CD38, PARP10, DTX2, LILRB2, TMEM132A, EMR2, SPHK1, SIRPA, PLAUR, VEGFA, PTPRF, C1QTNF1, STEAP3, MPZL1, CD82, CLEC4A, DUSP10, GLIS3, SOD2, ISCU | (14) |
| ILC2_up_PM ID26878113       | HPGD5, KRT1, IL17RB, TNFRSF19, PTGDR2, HPGD, IL1RL1, PKIB, C10orf128, RP11345M22.1, FSTL4, GAP43, MBOAT2, KLRG1, CSGALNACT1, FCRL3, CLIC4, IL10RA, HLF, LGALS12, ZP1, CHDH, RP11440I14.2, A2M, BACE2, RP11345M22.2, P2RY1, FASLG, NRIP3, MSRB3, NTRK1, LINC00340, PZP, PPARG, TNFRSF9, UBXN10, A2MP1, IFIT3, UTS2, CALCRL, RAP1GAP2, GRK5                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | (10) |
| Ageing                      | MMACHC, PDE7B, CTSS, HLADRA, LUZP1, C3, C1QB, BRINP3, C1orf210, DENND6B, APOD, KHDRBS2, DHDDS, VWF, GPER1, CALHM2, MPEG1, FCGR2A, GPNMB, CLASP2, MSL3, C4A, MGST1, SHARPIN, APPBP2, AIP, IGF, RNASET2, FCGR2B, ANTXR1, HIST1H1C, C1QA, RAB40B, CD74, LYZ, HMGN2, TLX3, SRPR, RORB, GFAP, ARRB1, MT2A, VTB2, ABCB6, ARL11, KITLG, MMP10, UGT2B17, FXYP1, ANXA3, BIRC7, CDKN1A, AMH, NPC2, SH3GLB1, HBB, PCSK6, GSTA1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | (15) |
| Th17.Zhang                  | KLRB1, RORC, PLXND1, CTSH, ALOX5, PTPN13, IL4I1, C11orf75, NEFL, HLF, JAKMIP2, DSE, LIMS1, HLA-DRB1, LTK, HLA-DRB4, USP10, NR1D1, LCAT, SAMD3, HSPG2, PTGIR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | (16) |
| ILC3_up_PM ID26878113       | FCER1G, SH2D1B, NCR2, NRP1, LINC00299, KIAA1324, LST1, AMICA1, IL23R, PCDH9, VWASA, XCL1, SIGLEC7, PLCG2, KLRC1, IL4I1, SLC4A10, KIT, GSN, IL1R1, TOX2, CD300LF, TYROBP, PTGDR, KRT81, XCL2, LTA4H, SPINK2, STAC, MYO7A, OTUD5, KLRF2, LIF, AFF3, ATP8B4, ENPP1, IL2RB, TMIGD2, PRR5, ELOVL6, AC092580.4, TNFSF13B, AFAP1L1, TGM2, B3GNT7, COL23A1, ENG, LDLRAD3, FGR, TNFRSF18, HIP1, APOL4, RP11-845M18.6, ITM2C, PLCH1, LPAR1, RHOC, NSMCE1, SCN1B, ID2, SLC43A2, ABCB5, ADAM28, CAT, RP11-563D10.1, TNFSF11, STARD3NL, TNS3, COL4A4, RNF152, CA2, TOX, FES, KIAA1217, B4GALNT1, VAV3, GPR82, DOCK5, CD33, S100A13, SOST, TLE1, PECAM1, BCAS1, RORC, LRRN3, TLE3, NCR1, SYK, HDAC9, CRTAM, DOK3, TNFSF4, CD300C, RP11330A16.1, FSD1, GNLY, CLNK, CXXC5, HPN, GOLIM4, AP000476.1, SPRY1, FAM179A, MPG, ABHD15, SPRED2, TRAJ45, GPR68, PRAM1, PDE6G, MATN2, AE000661.37, DOK7, ARMC9, HOXA10, SERPINA11, PDZK1, ENTPD1, LINGO4, TTN, CCL20, CD81, HOXA5, AC017104.6, SERP2, TRGJP1, SNCA, NEK10, TCF4, RP11-428G5.5, RGS9, KIAA0087, IRF4, DACH1, NCAM1, IKZF2, CDR2L, O3-Mar, FCRL4, PVRL2, LTBR, NLRP7, JAG2, TRDJ2, RP1191K8.1, S100B, LDLRAD4..4, HOXA7..1, CHKA, EFCAB4A, SUOX, KCTD11, RP11-31E23.1..1, RP11-98D18.9..2, LIMK1, JUP, RP11-15B24.5, KLRK1..1, KIFC3, RP11-98G13.1, IGFBP4, BANK1, OPCML..1, DTWD2, REEP1, MEF2C, ZFYVE9, MECOM, CARD9, GPRC5C, MUC2, INPP1, MPV17L, MYB, SLC2A10, SLC02A1, GRAMD1B, CARD6, ARSJ, CHMP4C, SNAPR1, KCNQ5, TSPAN13, NLRP2, LYN, R P11256L6.2, CXCL16, CCDC102B, GSAS1, C19orf35, MACC1, SORT1, GPR133, CCDC171.1, ORA3, SERPINH1..1, NPTXR, PTPRD, DERL3, PPP1R9A, B3GALT5, P4HA2, TRAJ44, MRC2, SLFN12, PLEKHN1, ACP6, TIE1, TMED8, UNC93B1                                                                                                                               | (10) |

## Supplementary information

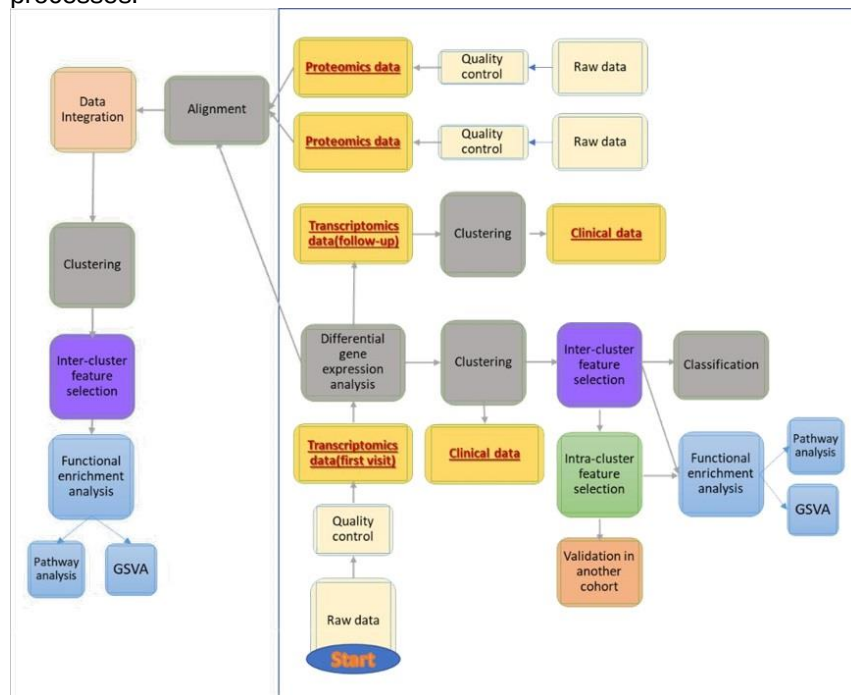
|                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |      |
|------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
|                                          | ,RP11279F6.3,RP11510J16.3,RP11277P12.20,MS4A6A,C9orf66,TRIO,PRKAR2B,CYP24A1,KLRK1..4,MMP25,TST,CTD2325P2.4,MAN1A1,RPPH1,PPFIBP1,KLRK1..6,UBE2E2,SORCS1,KIAA1456..1,COL24A1,LRP1,TRAIP,GPR97,FGD6,LMLN,EHHADH,ZNF461,CSF2,TRPM8,MCAM,CD38,FAM167A,C1orf159,OSBPL6                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |      |
| IL17a                                    | NFKBIZ,ZC3H12A,RELB,BIRC3,CXCL1,ZC3H12C,TNFAIP3,MAP3K8,CXCL2,CASZ1,IER3,LENG1,SOD2,RRAD,NFKB2,ID2,TATDN1,CBR3,CCL3,ICAM1,PIK3R6,GADD45B,ADAM21,ZBTB25                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | (17) |
| Th2<br>activated                         | ABCA2,ACTN1,ASPH,BAI2,BCL2L11,BEGAIN,BEST4,BTBD11,C16orf45,C6orf105,CADM1,CAPG,CCR7,CD27,DUSP23,E2F3,FAM55C,FANK1,FLJ32810,FLJ45983,FLJ46446,GAB2,GALT,GATA3,GBP5,GPA33,HOMER3,IGF1R,ITPR1P1,KBTBD11,KLRB1,LMCD1,LOC100128031,LOC100131176,LOC729041,MBD5,METRN1,MRC2,NECAB3,NEFL,NOD2,NOS3,NRIP1,NRIP3,PELI2,PPAPDC1B,PPP1R9B,PRAGMIN,PRMT10,RORC,SEMA5A,SLC12A6,SLC25A33,SNED1,SNN,SNTN,SUSD4,TMEM2,TMIGD2,TNFRSF18,TPPP,TULP4,TKX,ZBTB16,ZDHHC11,ZDHHC11B,ZP1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | (18) |
| Treg<br>activated                        | ,ACSL6,ANK3,AP3M2,APP,ASPH,AXIN2,BASP1,BFSP2,C1orf228,CACNA1I,CD40LG,CD70,CKS2,CPNE2,CTLA4,CTNNAL1,CXCR5,DENND5A,DPP4,DUSP10,DUSP4,F5,FANK1,FLJ46446,FOXP3,GBP5,GPR35,HLA-DMA,HLA-DMB,HLA-DOA,HLA-DRB1,HLA-DRB4,HS3ST3B1,ICA1,ID2,IKZF2,IKZF4,IL2RA,IL7R,IRF4,KLRB1,LMCD1,LOC100131646,LOC100132345,LOC284475,LOC541471,LOC643988,LOC650392,LOC730631,MGST2,MSX2P1,NELL2,NLRP3,NOG,NOSIP,NR3C2,NUSAP1,PCTP,PDIA5,PMAIP1,PTPLA,RBMS1,RECK,RHOXF1,RORC,RP3-527F8.2,RPL39L,SCML4,SETD7,SHMT2,SLC2A8,SLC40A1,SMC6,SMPD3,SOS2,SOX8,SPATS2L,STAT4,THEMIS,TIGIT,TMEM71,TNFRSF1B,TOX,TRIB1,TRIM16L,TSPAN18,VAV3,VIPR1,WNT10B,WNT7A,WNT7B,ZFP36L2                                                                                                                                                                                                                                                                                                                                                                                                                | (18) |
| CD4                                      | ALOX5,ANXA4,BIVM,BPI,BRD7,CAMP,CASK,CD63,CEACAM6,CEACAM8,CFLAR,DEFA4,GCA,LCN2,LOC23117,LOC728358,LTf,MBOAT1,MMP8,MNDA,PDE4D,POLR2J2,PPP1R16B,PRKCB1,RETN,RNASE3,RUNX2,S100A12,S100A8,S100A9,S100P,SLC38A2,SPATA13,SRGN,SRGN,STOM,TNIP1,TTYH2,UGCG,ZNF749                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | (19) |
| Eosinophils                              | CLC,SIGLEC8,EMR1,EMR4P,LGAL312,HRH4,CEBPE,DACH1,VSTM1                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | (20) |
| Mast cell                                | MS4A2,TPSAB1,TPSAB2,CPA3,RGS13,C1orf186                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | (20) |
| Macrophage<br>.GM_CSF.IL4                | CBR3,CLEC10A,FZD2,CD1C,SIGLEC10,PARM1,CCL26,FCER2,MAOA,MAP4K1,FOXQ1,HOMER2,PDXK,GALNT18,NAGPA,CD209,INSIG2,RTKN,SUCNR1,NFXL1,FAM60A,FCER1A,DUOX1,ASAP1,CXCR2P1,FGL2,GPD1L,PA1LLD,QPRT,C17orf58,C1orf54,CMTM8,CARD9,ADAM19,LIPA,SLC27A3,CTNNAL1,ASCL2,CD1A,RAB33A,DHRS2,RASGRP3,GATM,DTNA,RASSF7,ZNF366,BLOC1S6,PTPRE,ACOT7,STARD7,CCDC6,RASL10A,INTS3,ST6GAL1,DHRS11,ADAM15,PTRF,SGK223,PPP1R14A,BATF3,RAMP1,CERK,MMP12,SLAMF1,IL1R2,KIAA1671,SOC1,RAP1GAP,ALOX15,CD1E,CDKN1A,SOX8,DENND1B,VCL,SPINT2,PFKP,RHBDP1,FCSCN1,RRP1B,SNX8,LOC283070,PPM1F,GOLGA8B,ESPNL,CDR2L,NMNAT3,CCL22,TTC9C,CLEC4G,AUH,PTPRO,CLEC4A,CHN2                                                                                                                                                                                                                                                                                                                                                                                                                                 | (14) |
| Macrophage<br>.GM_CSF.TN<br>Fa.HS.IVS.UP | ZFYVE16,HNRNPDL,MCTP1,SLC7A11,Mar-03,IL1A,MCOLN2,NEK6,DDIT4,MAOA,NUB1,PHLDB1,TNS3,DRAM1,COX10,SLC20A1,VNN1,INSIG2,OCSTAMP,PTPN1,FAM60A,IL2RA,SLC7A1,ETS1,RHBDP2,G0S2,TSC22D1,EEF1B2,IDH3A,C17orf58,TRAF3IP2,PACSIN2,LMO4,TNIP1,MPO,ACSL5,CS,KLHL5,BTG3,OTUD4,PLGRKT,F5,SLAH1,NFKB1,AQP9,LAMB3,RPL32,ADAMDEC1,PFND1,SYT11,MT1F,TNC,SLC7A5,CA12,METTL21B,MRPS6,GK,BHLHE41,WARS,EFTUD1,SRG,CCDC115,FILIP1L,ARFGAP3,ATP1A1,PTPRE,NBN,HIVEP1,GRAMD1A,KYNU,SLC1A3,NFE2L1,PCNX,CYP27B1,ANXA5,QPCT,SH3PXD2A,SERPINE1,CTTN,CLINT1,TTC13,IRAK2,SLC41A2,TSPAN33,ACO1,RNF144B,HNRNPA0,FAM213A,NFS1,MTF1,BCAT1,UCK2,SLAMF1,STRN3,GPR124,CCL5,IL1B,BIRC3,SLC50A1,SLC2A6,CLEC5A,LINC00938,PRKCH,LAD1,SLC25A3,GPR84,DENND5A,TNFAIP6,SLC35F2,CRISPLD2,PRKCDBP,DNAJB6,NDP,PLEK,MMP9,SPARC,F3,ANKRD22,SLC9A1,RFTN1,TSKU,EBI3,PPA1,ADORA2A,RGS16,MYOF,ATP5L,METTL1,TNPO1,AK4,ESYT2,SCG5,COA1,FNBP4,TNFRSF4,HBEGF,MT1H,TMEM194A,ZSWIM4,IL36G,RPS4X,ACVR1B,BCL2,FYN,DTX2,HOPX,TMEM132A,PTGES2,EMR2,SPHK1,SFXN4,BID,VEGFA,PTPRF,C1QTNF1,HSD11B1,C6orf48,CD82,FMNL3,MTHFD2,SOD2 | (14) |

## References

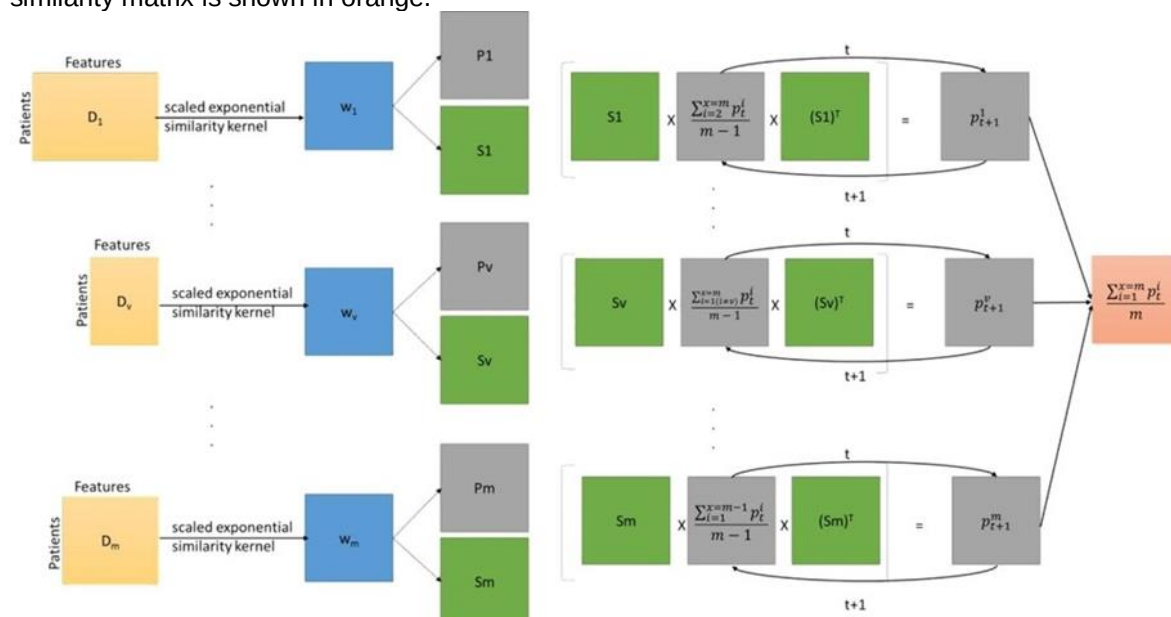
- Kuo CS, Pavlidis S, Loza M, Baribaud F, Rowe A, Pandis I, et al. T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur Respir J*. 2017;49(2):443-55.
- Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS one*. 2010;5(12):e15004.
- Burg D, Schofield JPR, Brandsma J, Staykova D, Folisi C, Bansal A, et al. Large-scale label-free quantitative mapping of the sputum proteome. *Journal of proteome research*. 2018;17(6):2072-91.
- Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J*. 2016;8(1):289-317.
- Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics (Oxford, England)*. 2010;26(12):1572-3.
- Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*. 2014;11(3):333-7.
- Webb AR. Statistical pattern recognition: John Wiley & Sons; 2003.
- Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*. 2016;8(1):289.
- Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT, et al. IL-13-induced airway mucus production is attenuated by MAPK13 inhibition. *J Clin Invest*. 2012;122(12):4555-68.
- Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nature immunology*. 2016;17(4):451-60.
- Simpson JL, Phipps S, Baines KJ, Oreo KM, Gunawardhana L, Gibson PG. Elevated expression of the NLRP3 inflammasome in neutrophilic asthma. *Eur Respir J*. 2014;43(4):1067-76.
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)*. 2011;27(12):1739-40.
- Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun*. 2005;6(4):319-31.
- Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. 2014;40(2):274-88.
- de Magalhaes JP, Curado J, Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics (Oxford, England)*. 2009;25(7):875-81.
- Zhang H, Nestor CE, Zhao S, Lentini A, Bohle B, Benson M, et al. Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1. *J Allergy Clin Immunol*. 2013;132(4):1005-8.
- Dragon S, Hirst SJ, Lee TH, Gounni AS. IL-17A mediates a selective gene expression profile in asthmatic human airway smooth muscle cells. *American journal of respiratory cell and molecular biology*. 2014;50(6):1053-63.
- Zhang H, Nestor CE, Zhao S, Lentini A, Bohle B, Benson M, et al. Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1. *Journal of Allergy and Clinical Immunology*. 2013;132(4):1005-8.
- Tsitsiou E, Williams AE, Moschos SA, Patel K, Rossios C, et al. Transcriptome analysis shows activation of circulating CD8+ T cells in patients with severe asthma. *Journal of Allergy and Clinical Immunology*. 2012;129(1):95-103.
- Leaker BR, Malkov VA, Mogg R, et al. The nasal mucosal late allergic reaction to grass pollen involves type 2 inflammation (IL-5 and IL-13), the inflammasome (IL-1 $\beta$ ), and complement. *Mucosal immunology*. 2017;10(2):408-20.

## Supplementary figures

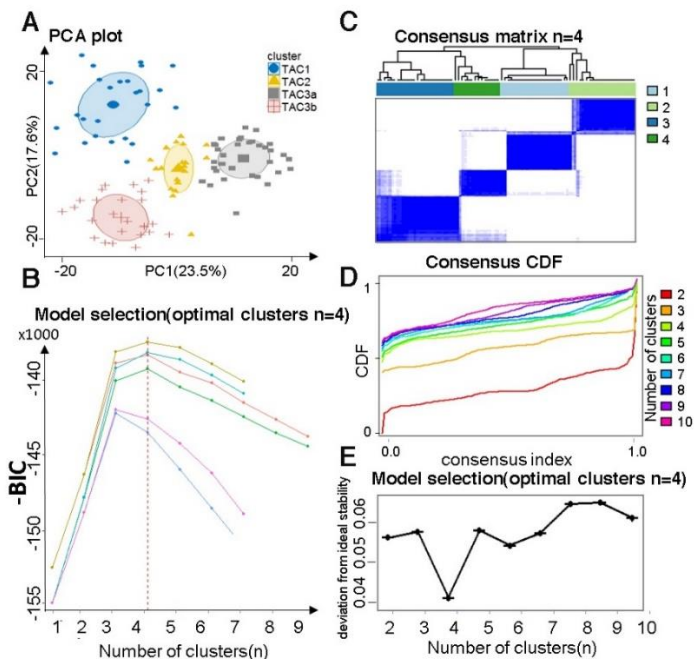
**Supplementary Figure S1-** Overview of disease subtyping workflow. The panel on the right shows analyses processes which are based on a single omics data. The grey panels indicate the integration processes.



**Supplementary Figure S2** -Schematic of similarity network fusion (SNF) - SNF steps are shown. The rectangles in yellow represent datasets. The blue boxes show the adjacency matrices. For each data set two matrices were generated based on the similarity kernels; these are shown by grey and green boxes. The boxes in the brackets show how individual similarity matrices are updated iteratively and the formula in the centre of the box shows the function for updating the underlying matrix. The fused similarity matrix is shown in orange.



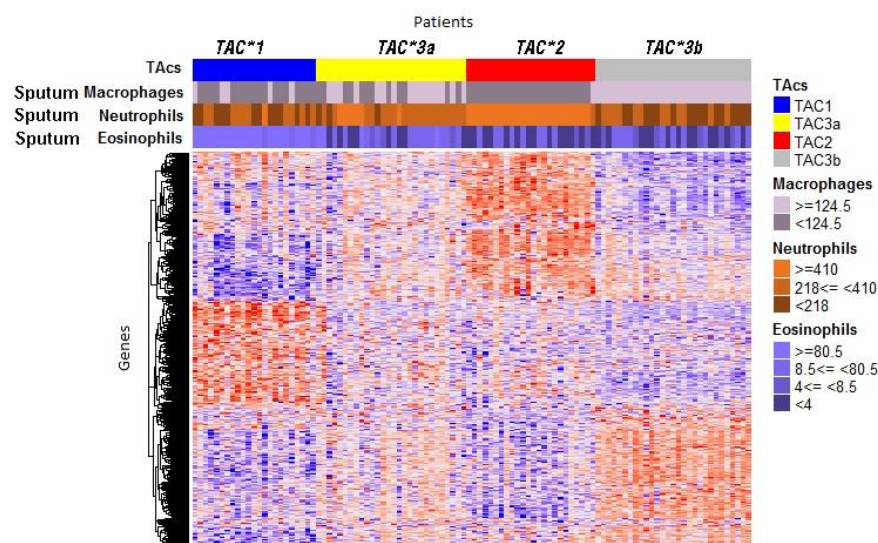
**Supplementary Figure S3-** Assessment of clustering tendency and determining the optimal number of disease subgroups. A is the principal component analysis (PCA) plot of the data colored according to the estimated clusters. The first principal component (PC1) is plotted versus the second principal component (PC2). B illustrates the clustering quality index based on the traces of Bayesian Information Criterion (BIC) as the function of number of clusters (1:9). C illustrates the consensus matrix based on the four clusters. Finite Gaussian mixture model clustering method was used as the base clustering method in the consensus clustering framework. D the Cumulative distribution function (CDF) of consensus matrices are shown as for a predefined number of clusters (1:9). E shows the deviation



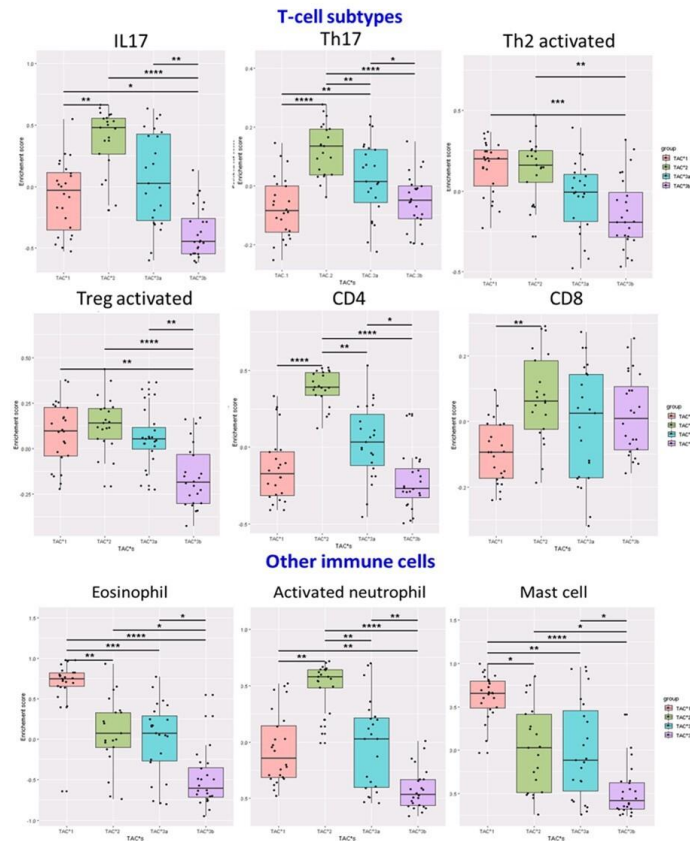
from ideal stability measurements

(DIS) as a function of number of clusters.

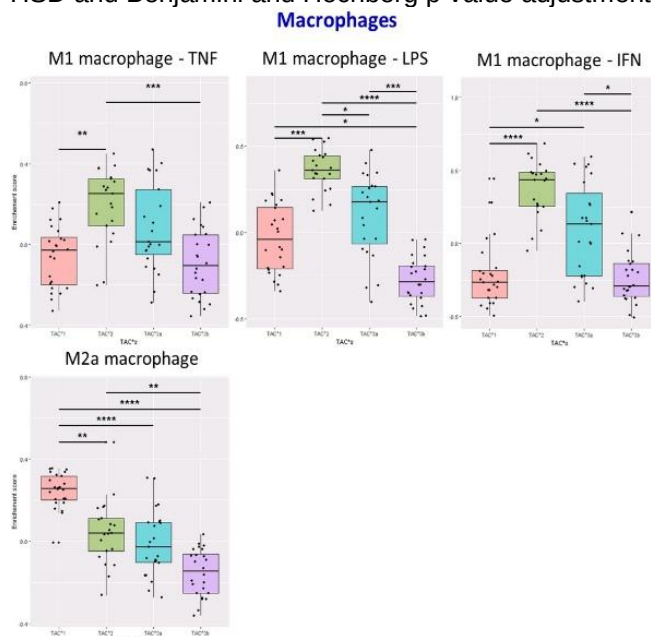
**Supplementary Figure S4-** Sputum inflammatory cells and TAC\*s- Heatmap of sputum transcriptomics data from 104 asthmatic subjects (columns) with 508 transcriptomic features (rows). Clustering results in four transcriptome-associated clusters: TAC\*1, TAC\*2, TAC\*3a and TAC\*3b. The sputum inflammatory cell counts for each participant is mapped on top of the column.



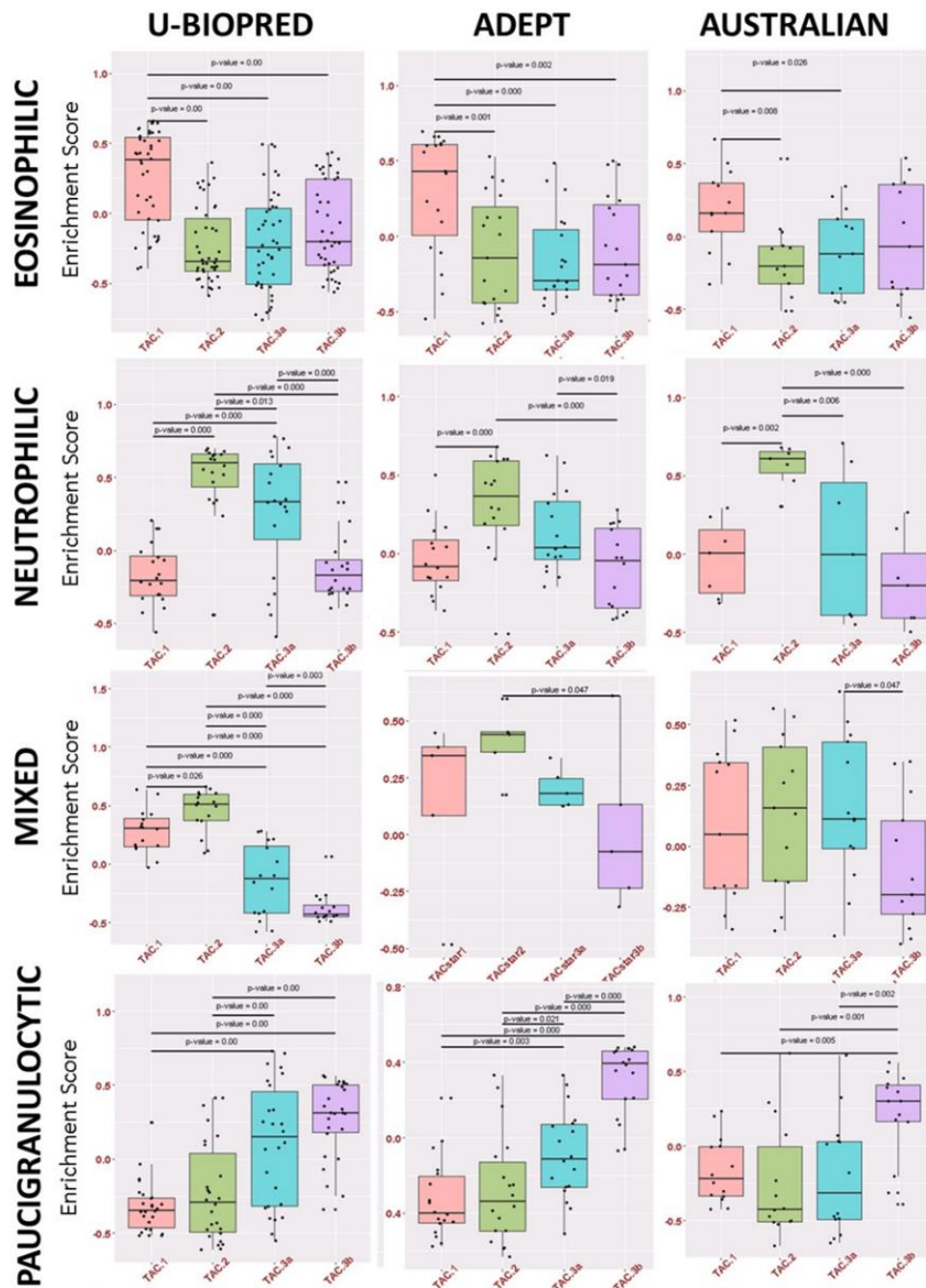
**Supplementary Figure S5-** Gene set variation analysis. Dot plot expression scores for T-cells and other immune cells gene signatures as shown in four transcriptome-associated clusters (TAC\*1, TAC\*2, TAC\*3a and TAC\*3b). Data are presented as individual data points, and box-and-whisker plots showing median and interquartile range respectively. Significance codes \*\*\*\* $P < 10^{-4}$ , \*\*\* $P < 10^{-3}$ , \*\* $P < 10^{-2}$ , \* $P < 0.05$ . P values were generated by Tukey's HSD and corrected for by Benjamini and Hochberg P-value adjustment.



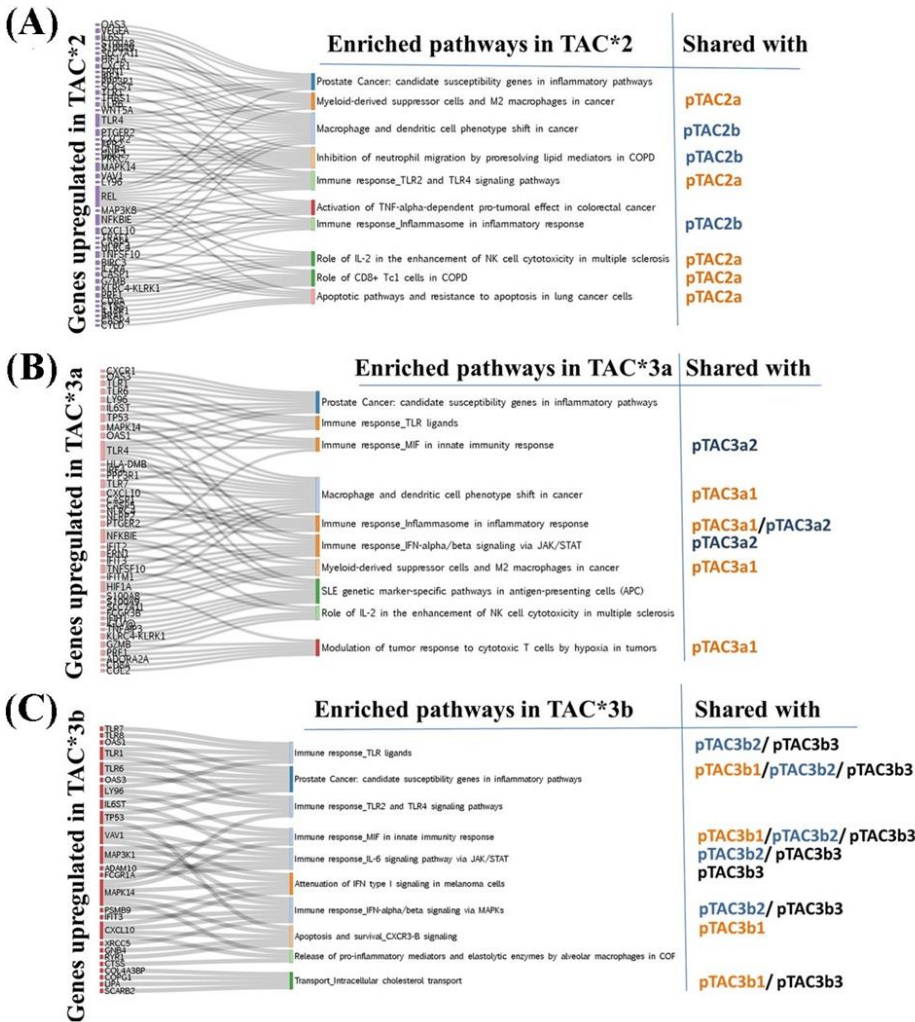
**Supplementary Figure S6-** Gene set variation analysis. Dot plot expression scores for macrophage gene signatures as shown in four transcriptome-associated clusters (TAC\*1, TAC\*2, TAC\*3a and TAC\*3b). Data are presented as individual data points, and box-and-whisker plots showing median and interquartile range respectively. Significance codes \*\*\*\* $P < 10^{-4}$ , \*\*\* $P < 10^{-3}$ , \*\* $P < 10^{-2}$ , \* $P < 0.05$ . Tukey's HSD and Benjamini and Hochberg p-value adjustment.



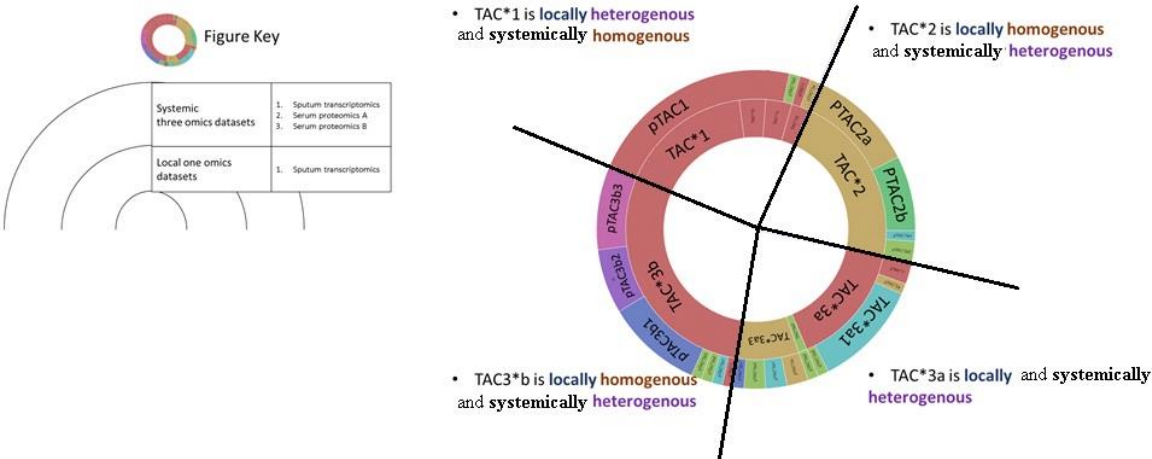
**Supplementary Figure S7. Validation of UBIOPRED signatures in ADEPT and Australian cohorts.** Dot-blot enrichment scores with box-and-whisker plots showing median and interquartile range of each of the TAC gene signatures according to the type of sputum granulocytic inflammation (eosinophilic, neutrophilic, or paucigranulocytic) in U-BIOPRED (Left panels), ADEPT (Middle panels) and Australian (Right panels) cohorts. Definitions: Eosinophilic: eosinophils  $\geq 1.49\%$ ; Neutrophilic: neutrophils  $\geq 73\%$ ; Mixed: eosinophils  $\geq 1.49\%$ ; and neutrophils  $\geq 73\%$ ; Paucigranulocytic: eosinophils  $< 1.49\%$  and neutrophils  $< 73\%$ . \*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .



**Supplementary Figure S8. TAC\*2 and TAC\*3 transcriptomic pathways split into PTACs.** Sankey diagram shows the associations between genes and the overrepresented pathways for TAC\*2 (Panel A), for TAC\*3a (Panel B) and for TAC\*3b (Panel C) after integration of transcriptomic and proteomic pathways into the PTACs.



**Supplementary Figure S9- Sunburst chart of sputum microarray and integrative clustering results.** The inner ring represents nine clusters derived from sputum microarray data. The outer ring represents clustering derived from the data integration workflow.



**Supplementary Figure S10.** Gene set variation analysis. Dot plot expression scores for OXPHOS gene signatures in sputum as shown in four transcriptome-associated clusters (TAC\*1, TAC\*2, TAC\*3a and TAC\*3b) compared to healthy volunteers (HV). Data are presented as individual data points, and box-and-whisker plots showing median and interquartile range respectively. Significance codes \*\*\*\* $P < 10^{-4}$ . Tukey's HSD and Benjamini and Hochberg p-value adjustment.

