DOI: 10.1111/all.14573

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.[1] 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[2] and re-analysed the 508 differentially expressed genes (DEGs) of sputum cell transcriptomics between high eosinophil and low eosinophil counts [1] 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[1] 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[3] indicated key-regulated immune pathways that distinguished the 4 TAC*s. Gene set variation analysis (GSVA),[4] 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[5] and Australian[6] 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)[7] and shot-gun proteomics using liquid chromatography-mass spectrometry (LC-MS/MS)[8] 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.[9] 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 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

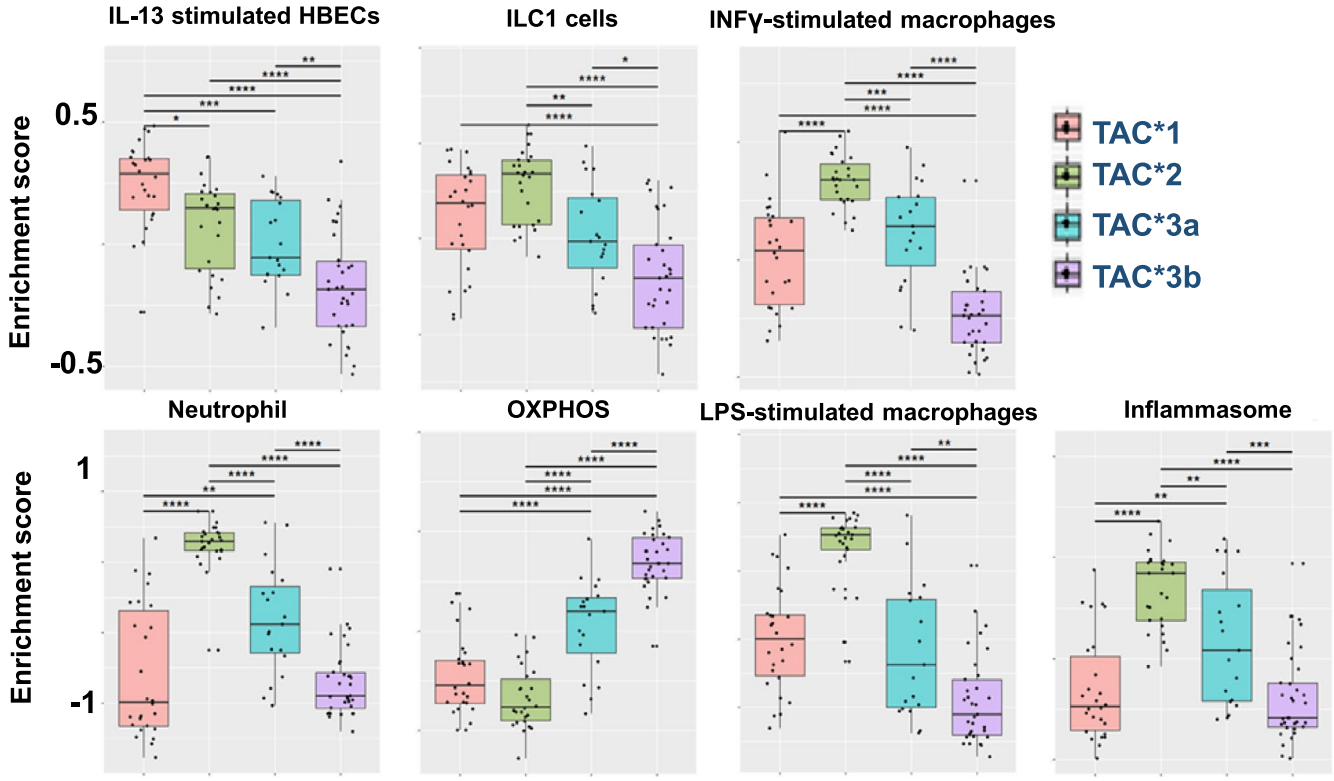


FIGURE 1 Gene set variation analysis. Dot plot enrichment scores with box-and-whisker plots showing median and interquartile range for 7 different gene signatures in the transcriptome-associated clusters: TAC*1 (n = 26), TAC*2 (n = 26), TAC*3a (n = 19) and TAC*3b (n = 33). The signatures reflect activation 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-γ-stimulated macrophage. ****$P < .0001$, ***$P < .001$, **$P < .01$, *$P < .05$

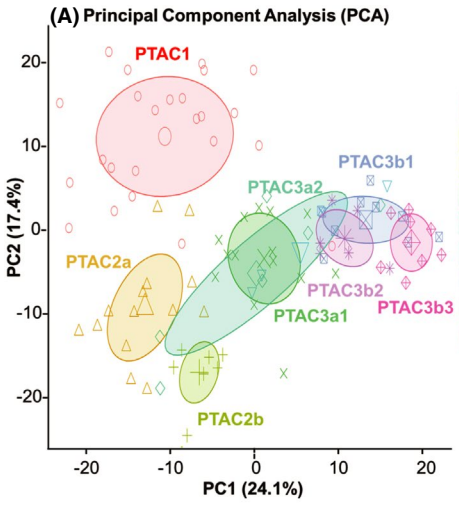
associated with OXPHOS and ageing pathways but their delineation remained less clear than PTAC2. Reduced OXPHOS compared with healthy controls in TAC1, TAC2 and TAC3a indicate a potential switch to glycolytic processes but TAC3b with significantly increased OXPHOS respond to distinct mitochondrial or 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 IFNγ 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 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



(B)

Cluster	TAC*1	TAC*2	TAC*3a	TAC*3b
PTAC1	23	0	0	1
PTAC2a	2	12	0	0
PTAC2b	0	7	0	0
PTAC3a1	0	1	11	2
PTAC3a2	0	2	3	1
PTAC3a3	0	0	2	1
PTAC3b1	0	0	0	10
PTAC3b2	0	0	0	6
PTAC3b3	0	0	0	8

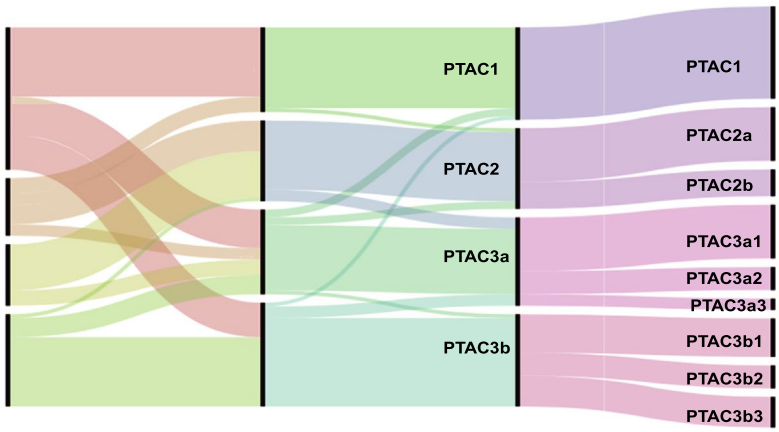


FIGURE 2 Integration of sputum transcriptomics and serum proteomics. (A) Principal component analysis plot identified 9 transcriptomic-associated clusters. The dots represent patients and colour code based on cluster subgroups. TAC*1 remains intact shown as PTAC*1; TAC*2 is divided into two subgroups 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)

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 AstraZenica 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).

Nazanin Zounemat Kermani[1]
Mansoor Saqi[1]
Paul Agapow[1]
Stelios Pavlidis[1]
Chihhsi Kuo[1,2]
Kai Sen Tan[3]
Sharon Mumby[2]
Kai Sun[1]
Matthew Loza[4]
Frederic Baribaud[4]
Ana R. Sousa[5]
John Riley[5]
Asa M. Wheelock[6]
Craig E. Wheelock[6]
Bertrand De Meulder[7]
Jim Schofield[8,9]
Stephany Sánchez-Ovando[10]
Jodie Louise Simpson[10]
Katherine Joanne Baines[10]
Peter A. Wark[10]

Charles Auffray[7]
Sven-Erik Dahlen[6]
Peter J. Sterk[11]
Ratko Djukanovic[8,9]
Ian M. Adcock[1,2]
Yi-ke Guo[1]
Kian Fan Chung[1,2]
U-BIOPRED Project Team

[1]*Department of Computing & Data Science Institute, Imperial College London, London, UK*
[2]*National Heart & Lung Institute, Imperial College London, London, UK*
[3]*Department of Otolaryngology, National University of Singapore, Singapore, Singapore*
[4]*Janssen Research and Development, High Wycombe, UK*
[5]*Respiratory Therapeutic Unit, GSK, Stockley Park, UK*
[6]*Department of Medicine Solna & Center for Molecular Medicine & Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden*
[7]*European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Lyon, France*
[8]*Faculty of Medicine, Southampton University, Southampton, UK*
[9]*NIHR Southampton Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton, UK*
[10]*Priority Research Centre for Healthy Lungs, Faculty of Health and Medicine, University of Newcastle, Newcastle, NSW, Australia*
[11]*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. GSVA: 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.

DOI: 10.1111/all.14518

2S albumins and 11S globulins, two storage proteins involved in pumpkin seeds allergy

To the Editor,

Members from *Cucurbitaceae* family have been reported to induce food allergy.[1] Pumpkin (*Cucurbita maxima*) pulp has been mostly the allergenic source but few studies are focused on the allergenic potential of its seeds.[2] Pumpkin seed may be consumed as snacks or as components in other food products, becoming hidden allergens, eliciting infrequent but severe cases of allergy with life-threatening reactions.[3] The nature of those allergens has not been investigated in detail so far. This study aimed to identify two allergens, from pumpkin seeds a 2S albumin and an 11S globulin, involved in severe allergic reactions of four patients allergic to these seeds and evaluate the cross-reactivity with other seeds and nuts containing homologous proteins. General characteristics of selected individuals, extracted from their clinical histories, are shown in Table S1. All patients described immediate allergic reactions with severe and systemic symptoms as anaphylaxis, showing a positive specific IgE (ImmunoCAP, Thermo Fisher) and Skin prick testing (SPT) to pumpkin seeds extract.

Immunological profiles of allergic patients were obtained by testing their sera with pumpkin seed extract under reducing and