CrossMark

T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED

Chih-Hsi Scott Kuo[1,2,3], Stelios Pavlidis[4], Matthew Loza[4], Fred Baribaud[4], Anthony Rowe[4], Iaonnis Pandis[3], Ana Sousa[5], Julie Corfield[6,7], Ratko Djukanovic[8], Rene Lutter[9], Peter J. Sterk[9], Charles Auffray[8,10], Yike Guo[3], Ian M. Adcock[1,2,11] and Kian Fan Chung[1,2,11] on behalf of the U-BIOPRED Study Group[12]

Affiliations: [1]Airways Disease, National Heart and Lung Institute, Imperial College London, London, UK. [2]Biomedical Research Unit, Royal Brompton and Harefield NHS Trust, London, UK. [3]Dept of Computing and Data Science Institute, Imperial College London, London, UK. [4]Janssen R&D, High Wycombe, UK. [5]Respiratory Therapeutic Unit, GSK, Stockley Park, UK. [6]AstraZeneca R&D, Mölndal, Sweden. [7]Areteva R&D, Nottingham, UK. [8]Faculty of Medicine, Southampton University, Southampton, UK. [9]Faculty of Medicine, University of Amsterdam, Amsterdam, The Netherlands. [10]European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL, Université de Lyon, Lyon, France. [11]These authors contributed equally to this work. [12]A full list of the U-BIOPRED Consortium project team member and their affiliations can be found in the Acknowledgements section.

Correspondence: K.F. Chung, National Heart and Lung Institute, Imperial College London, Dovehouse Street, London SW3 6LY, UK. E-mail: f.chung@imperial.ac.uk

ABSTRACT Asthma is characterised by heterogeneous clinical phenotypes. Our objective was to determine molecular phenotypes of asthma by analysing sputum cell transcriptomics from 104 moderate-to-severe asthmatic subjects and 16 nonasthmatic subjects.

After filtering on the differentially expressed genes between eosinophil- and noneosinophil-associated sputum inflammation, we used unbiased hierarchical clustering on 508 differentially expressed genes and gene set variation analysis of specific gene sets.

We defined three transcriptome-associated clusters (TACs): TAC1 (characterised by immune receptors *IL33R*, *CCR3* and *TSLPR*), TAC2 (characterised by interferon-, tumour necrosis factor-α- and inflammasome-associated genes) and TAC3 (characterised by genes of metabolic pathways, ubiquitination and mitochondrial function). TAC1 showed the highest enrichment of gene signatures for interleukin-13/T-helper cell type 2 (Th2) and innate lymphoid cell type 2. TAC1 had the highest sputum eosinophilia and exhaled nitric oxide fraction, and was restricted to severe asthma with oral corticosteroid dependency, frequent exacerbations and severe airflow obstruction. TAC2 showed the highest sputum neutrophilia, serum C-reactive protein levels and prevalence of eczema. TAC3 had normal to moderately high sputum eosinophils and better preserved forced expiratory volume in 1 s. Gene–protein coexpression networks from TAC1 and TAC2 extended this molecular classification.

We defined one Th2-high eosinophilic phenotype TAC1, and two non-Th2 phenotypes TAC2 and TAC3, characterised by inflammasome-associated and metabolic/mitochondrial pathways, respectively.

Introduction

Severe asthma is defined as asthma that remains partly or totally unresponsive to asthma treatments [1]. The inflammatory mechanisms underlying severe asthma involve multiple cellular compartments with a diversity of disease-driving mechanisms. The CD4 T-helper cell type 2 (Th2)-mediated pathway orchestrated by the airway epithelium has been recognised as a driving force in allergic asthma [2, 3]. The eosinophil count in induced sputum has been used as a surrogate biomarker for this pathway [4]. However, eosinophilic (EOS) asthma can also be underlain by a non-Th2 mechanism involving innate lymphoid cell type 2 (ILC2) [5, 6]. The driving mechanism for non-EOS asthma such as neutrophilic asthma has been associated with altered innate immune response and activation of Th17 cells [7, 8]. Gene expression analyses of sputum or blood cells from patients with neutrophilic asthma have reported high expression of genes related to pathogen recognition, neutrophil chemotaxis, protease activity and inflammasome assembly [9–11]. The disease driver(s) associated with paucigranulocytic asthma remain largely unclear [12, 13].

Clustering using clinical features alone has not yielded information on the underlying biology as similar inflammatory cell profiles have been seen between these clinical clusters [14]. We performed an unsupervised clustering of differentially expressed genes (DEGs) on EOS *versus* non-EOS asthma to categorise driving mechanisms that inform on the significance of the granulocytic inflammatory profile. We first defined a set of genes expressed in the sputum of EOS and non-EOS inflammatory phenotypes. Clustering on these genes led to the delineation of three new clusters distinguished by distinct sets of gene signatures. We define an interleukin (IL)-13/Th2-high predominantly EOS cluster and two non-Th2 phenotypes, which were characterised by interferon (IFN)/tumour necrosis factor (TNF)-α/inflammasome-associated, and metabolic and mitochondrial pathways, respectively.

Methods

A full description of methods is provided in the online supplementary material.

Study design

We obtained transcriptomic data from sputum cells obtained from 104 participants (online supplementary table S1) with moderate-to-severe asthma and 16 healthy volunteers (HV group) from the U-BIOPRED cohort [15]. The study was approved by the ethics committees of the recruiting centres. All participants gave written informed consent.

Analysis of sputum "omics"

Sputum was induced by inhalation of hypertonic saline solution and sputum plugs were collected from which sputum cells and sputum supernatants were obtained, as described previously [16]. Expression profiling was performed using Affymetrix U133 Plus 2.0 (Affymetrix, Santa Clara, CA, USA) microarrays with RNA extracted from sputum cells. Proteomic profiling of sputum supernatants was performed using the SOMAscan proteomic assay [17, 18].

Computational analysis of transcriptomic features

Data were uploaded and curated in the tranSMART system [19]. We analysed 508 DEGs from a pairwise comparison of gene expression in asthmatic subjects with high sputum eosinophil counts ($\geqslant 1.5\%$), in asthmatic subjects with low eosinophil counts ($<1.5\%$) and in the HV group (figure 1a and b, and online supplementary table S1). The intensity of the raw probe sets were $\log_2$ transformed and normalised by the robust multiarray average method [20]. A regression-based method (R package limma; www.r-project.org) was used to analyse DEGs with respect to the groups of interest, and batch/technical effects, age, sex and administration of oral corticosteroid were adjusted for as covariates in the linear model. The false discovery rate using the Benjamini–Hochberg method was applied for p-value adjustment in relation to multiple tests.

a)

```
Data pre-processing
1) Quality control
2) Technical/batch effect
```
↓
```
Sputum DEG analysis from EOS,
non-EOS and HV groups
```
↓
```
Identify TACs from sputum DEGs
```
↓
```
Determination of sputum
signature model for TACs
```
↓
```
Validation of sputum signature
model in ADEPT cohort n=47
```
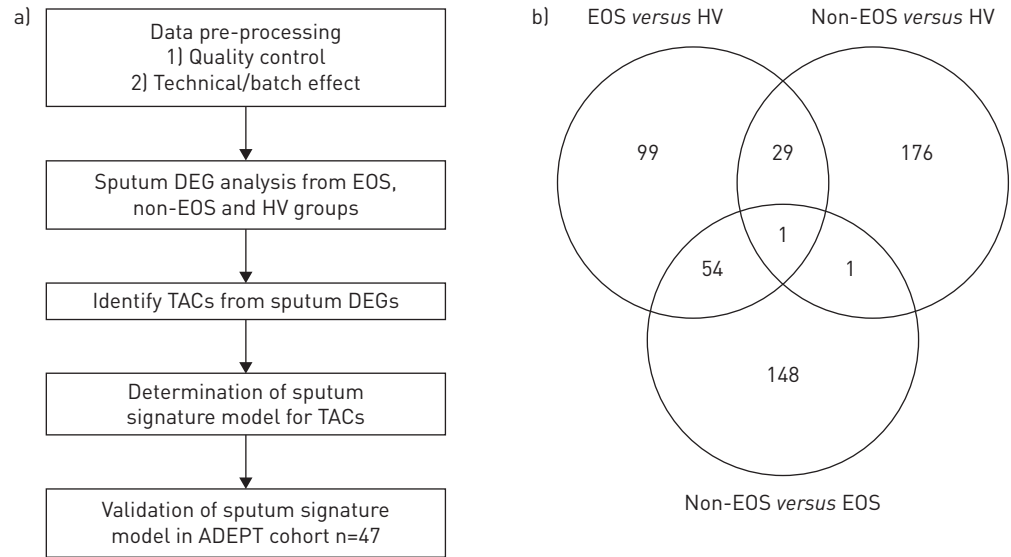
b)

FIGURE 1 Clustering approach. a) Workflow for identification of transcriptome-associated clusters (TACs), determination of signature classifiers, and validation of findings of signature and clinical feature of clusters. DEG: differentially expressed gene; EOS: eosinophilic; ADEPT: Airways Disease Endotyping for Personalized Therapeutics. b) Number of DEGs derived from comparison of EOS, non-EOS and healthy volunteer (HV) groups displayed using a Venn diagram.

Hierarchical clustering based on Euclidean distance was used for cluster exploration. A supervised learning algorithm using the shrunken centroid method [21] was applied to the cluster findings to determine the predictive signatures for each cluster and feature reduction methods were implemented along with learning algorithms to obtain a sparse model to facilitate interpretation. Consensus clustering, a resampling technique taking into account the cluster consensus across multiple runs of a clustering algorithm, was used to determine the optimal cluster number by finding a cluster number k where the consensus matrix histogram approximates a bimodal distribution at the k cluster and a relative small increase of the area under curve of the cumulative distribution function at the k+1 cluster [22–24]. The nearest shrunken centroid method [22] was used as a supervised learning algorithm to refine the signatures for the identified transcriptome-associated clusters (TACs).

Signatures summarised by gene set variation analysis
Gene set variation analysis (GSVA) calculates sample-wise enrichment scores (ESs) [25, 26]. We compiled nine gene sets each related to a specific aspect of asthma (online supplementary table S4) and the ES was calculated for each gene set for each subject. ANOVA was used to analyse the ES differences among group means and the t-test was applied to compare the ES differences between the two means.

Validation study
The sputum signature findings predictive of each TAC from U-BIOPRED were applied to sputum transcriptomic data obtained from a study for disease profiling of asthma and chronic obstructive pulmonary disease (ADEPT (Airways Disease Endotyping for Personalized Therapeutics)) cohort) [27] using GSVA (online supplementary table S5). Sputum samples from 38 asthmatic subjects and nine HV group subjects were analysed by Affymetrix U133 microarray.

Data deposition
The transcriptomic data have been deposited in the GEO (Gene Expression Omnibus) database (www.ncbi.nlm.nih.gov/geo) with accession number GSE76262.

Results
Pathway analysis of transcriptomic features
We defined subgroups of asthmatic patients by analysing 508 DEGs from a comparison of the three groups of the U-BIOPRED cohort defined by EOS (⩾1.5%) versus non-EOS, EOS versus HV and non-EOS versus HV (figure 1a and b, and online supplementary table S1). Online supplementary table S2 shows the top 10 significant pathways for the three sets of DEGs from available public ontology databases [28]. Most of the DEGs in each comparison set were enriched in biological processes or pathways related to immune activation and cytokine production, while DEGs from EOS versus non-EOS presented enrichment in some

specific categories, such as regulation of cysteine-type endopeptidase activity ($p=1.19\times10^{-6}$), pattern recognition receptor signalling pathway ($p=1.55\times10^{-5}$), response to IFN-γ ($p=0.001$), IPAF inflammasome complex ($p=0.002$) and NOD-like receptor signalling pathway ($p=0.004$).

Definition of three transcriptomic-associated clusters

The consensus matrices for clusters between k=2 and k=5 (figure 2a–d) showed that the cumulative distribution function curve of the consensus index at cluster k=2 approximated a bimodal distribution (figure 2e), yet the increase of the area under the curve at k=3 (figure 2f) was very large.
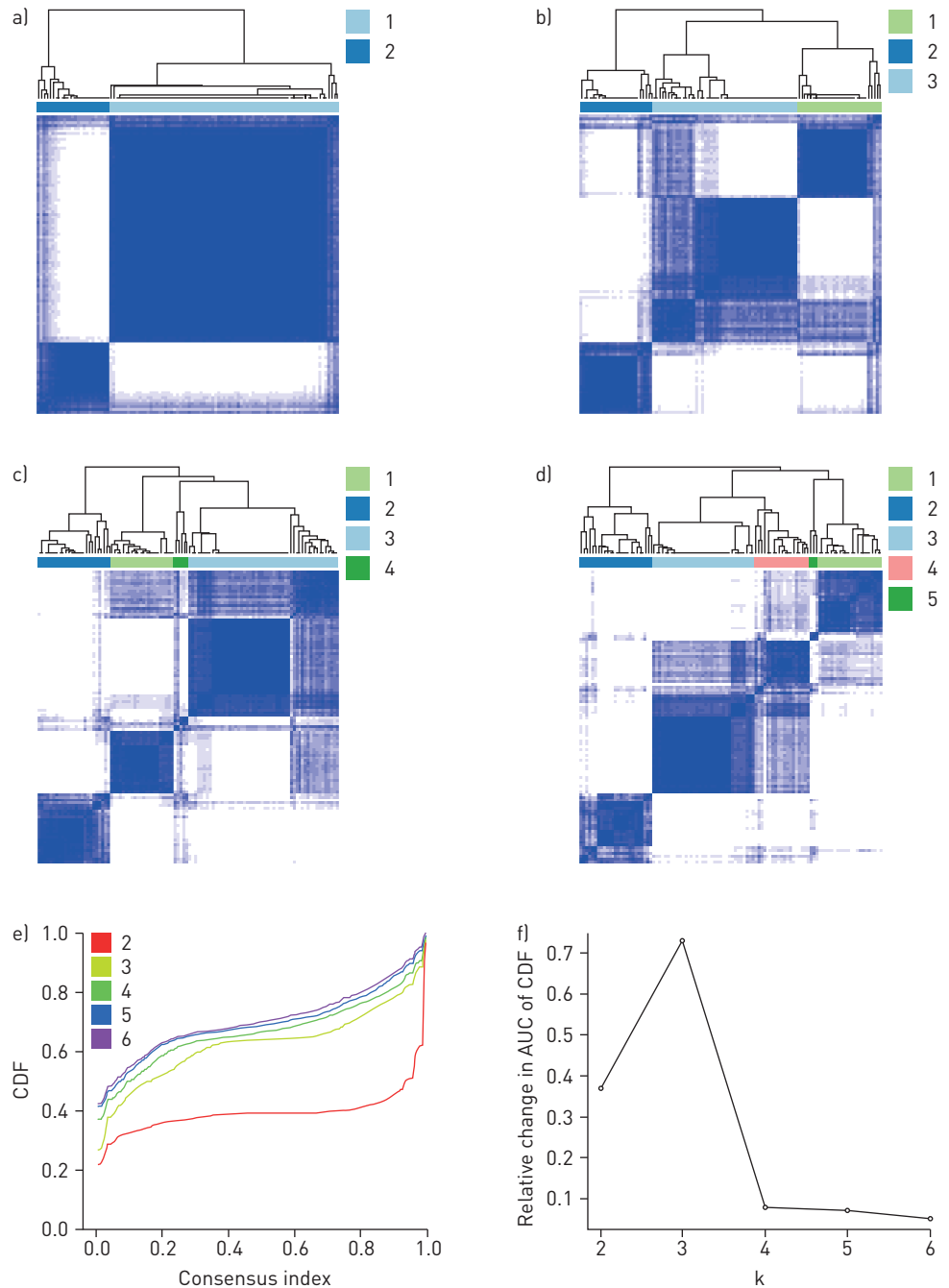

FIGURE 2 Consensus clustering to determine optimal number of matrices. AUC: area under the curve; CDF: cumulative distribution function. The optimal cluster number was determined by finding a cluster number k where the consensus matrix histogram approximates a bimodal distribution at the k cluster and a relative small increase of the AUC of the CDF at the k+1 cluster. a–d) Consensus matrices for clusters number a) k=2, b) k=3, c) k=4 and d) k=5. e) CDF curves of the consensus index for k=2–6, where k=2 approximated a bimodal distribution, while f) the increase of AUC at k=3 was maximal. Cluster number k=3 was the optimal choice.

Hierarchical clustering (figure 3) and resampling (figure 2) yielded three TACs. TAC1, with the highest eosinophilia, exhaled nitric oxide fraction (F_{NO}) and serum periostin, defined exclusively severe asthma patients with high oral corticosteroid dependency, acute exacerbation, nasal polyps and severe airflow obstruction (table 1). The shrunken centroid algorithm [21] (figure 4 and online supplementary figure S1) defined 20 genes for TAC1 related to multiple cytokine receptors and signalling (*IL1RL1*, *SOCS2*, *CCR3*, *CRLF2*), enzymes found in macrophages, mast cells and eosinophils (*PRSS33*, *CLC*, *ALOX15*, *TPSB2*, *CPA3*), and a cell adhesion molecule on granulocytes and B-cells involved in the damaged-induced adaptive immune response (*CD24*) [29]. TAC2 was characterised by the highest sputum neutrophil counts, serum C-reactive protein (CRP) and greater prevalence of eczema, and was defined by 39 genes associated with the IFN and TNF superfamilies (*IFIT2*, *TNFSF10*, *IFIH1*, *TNFAIP3*, *IFITM1*, *IL18RAP*), leukocyte surface receptors mediating innate immunity (*FPR2*, *TREML2*, *TLR1*), neutrophil chemotaxis and migration (*CXCR1*, *CXCR2*, *VNN2*, *VNN3*), inflammasomes (*CASP4*, *MEFV*, *NAIP*), and pattern recognition (*CLEC4D*, *CLEC4E*). TAC3 had normal to moderately high sputum eosinophils, better preserved forced expiratory volume in 1 s, the lowest prevalence of severe asthma and the least oral corticosteroid dependency. The TAC3 signature comprised 17 genes linked to glucose and succinate metabolism (*SUCLG2*, *TBC1D4*), ubiquitination (*ZYG11B*), mitochondrial function (*MRPL57*, *PDCD2*), energy consumption (*ATP1B1*), and endo/lysosomal function and transport (*SCARB2*, *TGOLN2*, *SCOC*) (figure 4).

Using GSVA [25], there was a significant difference in the enrichment of the nine signatures associated with asthma (figure 5), *i.e.* the activation of IL-13/Th2, ILC1, ILC2, ILC3, Th17, neutrophil activation, inflammasome, oxidative phosphorylation (OXPHOS) and ageing signatures. TAC1 showed the highest ESs for IL-13/Th2 and ILC2, and low ESs for Th17, neutrophil activation, inflammasome, OXPHOS and ageing signatures, while TAC2 had the highest ESs for ILC1, neutrophil activation and inflammasome
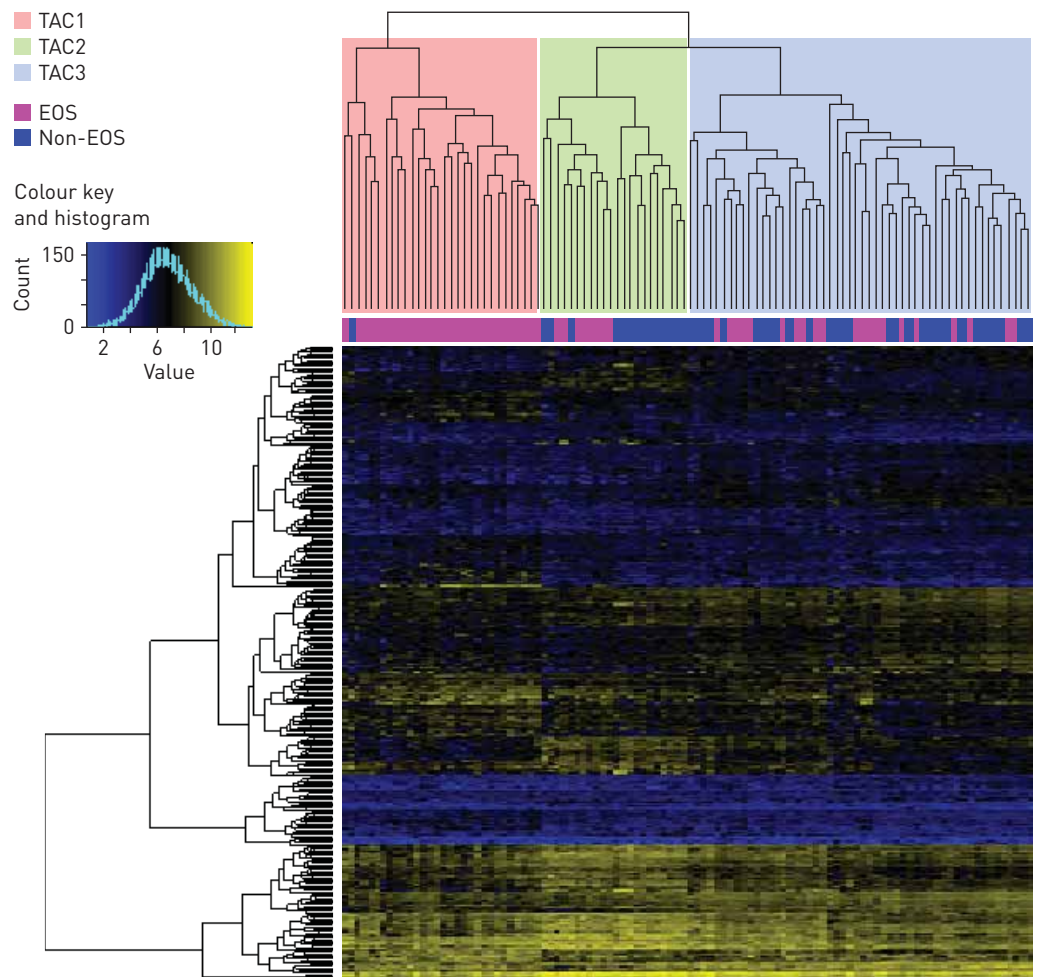

FIGURE 3 Heatmap of hierarchical clustering on 104 asthmatic subjects (columns) with 508 transcriptomic features (rows). Clustering results in three transcriptome-associated clusters: TAC1, TAC2 and TAC3. The sputum granulocyte status for each participant is mapped underneath the column dendrogram. EOS: eosinophilic.

TABLE 1 Comparison of the clinical features of the three sputum transcriptome-associated clusters (TACs)

	TAC1	TAC2	TAC3	p-value[#]		
				TAC1 *versus* TAC2	TAC1 *versus* TAC3	TAC2 *versus* TAC3
Subjects	30	22	52			
Age years	56.5±11.3	49.6±13.6	49.1±13.8	0.059	0.011	0.907
Age of onset years	35.0±19.2	17.3±16.2	23.5±19.5	7.3×10^{-4}	0.012	0.163
Female	17 (56.7)	15 (68.2)	28 (53.8)	0.579	0.987	0.376
BMI kg·m^{-2}	27.3±5.0	28.5±5.4	27.8±5.4	0.396	0.646	0.606
Smoker	7 (23.3)	3 (13.6)	13 (25.0)	0.603	1.000	0.437
Nasal polyps	14 (46.7)	6 (27.3)	14 (26.9)	0.258	0.115	1.000
Allergic rhinitis	12 (40.0)	10 (45.5)	20 (38.5)	0.913	1.000	0.763
Eczema	7 (23.3)	13 (59.1)	13 (25.0)	0.020	1.000	0.110
Severe asthma	29 (96.7)	18 (81.8)	37 (71.2)	0.187	0.012	0.503
Oral corticosteroid use	17 (56.7)	8 (36.4)	13 (25.0)	0.243	0.009	0.478
Atopy	21 (70.0)	14 (63.6)	38 (73.1)	0.854	0.965	0.593
Exacerbations year^{-1}	2.0 (1.0–3.0)	1.0 (1.0–2.8)	1.0 (0.0–2.0)	0.144	0.012	0.328
FEV$_1$ % pred	55.1 (41.9–74.1)	62.9 (47.1–83.5)	74.2 (63.7–89.8)	0.303	3.1×10^{-4}	0.025
Total IgE IU·mL^{-1}	152.0 (70.5–272.2)	90.5 (31.3–187.5)	89.4 (43.0–170.0)	0.241	0.091	0.852
Blood leukocytes ×10^3 μL^{-1}	9.04 (7.07–10.82)	7.60 (6.15–9.87)	6.81 (6.00–8.45)	0.266	0.006	0.248
Blood eosinophils ×10^3 μL^{-1}	0.43 (0.25–0.58)	0.25 (0.10–0.40)	0.20 (0.12–0.30)	0.013	3.9×10^{-4}	0.532
Blood neutrophils ×10^3 μL^{-1}	5.12 (3.89–7.41)	4.46 (3.90–6.48)	4.01 (3.32–4.90)	0.613	0.016	0.111
Sputum eosinophils %	30.9 (15.0–51.4)	0.6 (0.2–2.4)	1.0 (0.2–4.5)	3.2×10^{-8}	9.2×10^{-11}	0.593
Sputum neutrophils %	49.0 (25.9–70.5)	90.9 (84.9–94.4)	48.8 (33.3–63.5)	7.4×10^{-8}	0.784	3.4×10^{-10}
Sputum eosinophils >10%	25 (83.3)	0	6 (11.5)	1.5×10^{-8}	4.9×10^{-10}	0.232
F_{eNO} ppb	29.5 (20.0–57.5)	22.0 (14.0–26.0)	27.5 (18.1–42.4)	0.090	0.421	0.110
Serum periostin ng·mL^{-1}	60.2 (50.2–72.4)	49.9 (43.2–52.9)	42.6 (37.0–53.2)	0.051	8.8×10^{-4}	0.076
CRP mg·L^{-1}	2.5 (1.0–3.6)	5.4 (3.0–7.2)	1.9 (1.0–5.0)	0.022	0.756	0.010

Data are presented as n, mean±SD, n (%) or median (interquartile range), unless otherwise stated. BMI: body mass index; FEV$_1$: forced expiratory volume in 1 s; F_{eNO}: exhaled nitric oxide fraction; CRP: C-reactive protein. [#]: Kruskal–Wallis or ANOVA test.

signatures. TAC3 showed the highest ESs for ILC3, Th17, OXPHOS and ageing signatures, but low ESs for IL-13/Th2, ILC1, neutrophil activation and inflammasome signatures (figure 5).

TACs and sputum granulocytic inflammation

EOS inflammation was present in all three TACs, with 96.7% (29 out of 30) of TAC1 with high levels of sputum eosinophilia, and with 36.4% (eight out of 22) of TAC2 and 40.4% (21 out of 52) of TAC3. The TAC1 subtype was enriched for both IL-13/Th2 and ILC2 signatures (p=10^{-7} and p=0.01, respectively; figure 6a and b). Neutrophilic inflammation was found mainly in TAC2 and also in TAC3 (figure 6a and b), and the paucigranulocytic group was mainly TAC3 (figure 6a). In neutrophilic inflammation, the neutrophil and inflammasome signatures were highest for TAC2 compared with TAC3 (p=0.01 and p<0.02, respectively). Mixed inflammation was composed of TAC1 and TAC2, with TAC1 patients in this mixed group showing enrichment for IL-13/Th2, but TAC2 subjects were relatively enriched for neutrophil and inflammasome signatures (p<0.01 and p=0.002, respectively; figure 6a and c). These findings support the view that the mixed granulocytic phenotype is less biologically distinct [12, 13], but dependent on the pathobiological drive determined by each TAC.

Analysis of TAC signatures in the ADEPT cohort

The ES of TAC1 signatures (online supplementary figure S2) was significantly higher in moderate asthma compared with mild asthma (mean difference −0.52, p=0.010) or healthy controls (mean difference −0.48, p=0.030) in the ADEPT cohort. The TAC2 and TAC3 signatures (online supplementary figure S2) among healthy controls, mild-to-moderate and severe asthma were similar. When subjects were classified according to granulocytic inflammatory phenotypes, the ADEPT samples recapitulated the findings from the U-BIOPRED cohort. We confirmed for the eosinophil-predominant phenotype that the mean ES was positive for both TAC1 and TAC3 signatures, but not for TAC2 (figure 7). The mixed granulocytic phenotype was enriched for TAC1 and TAC2 signatures, but not for TAC3 (figure 7). In addition, the neutrophil-predominant phenotype was enriched for the TAC2 signature, but not for TAC1 or TAC3, and the paucigranulocytic phenotype was enriched for the TAC3 signature, but not for TAC1 or TAC2 (figure 7).
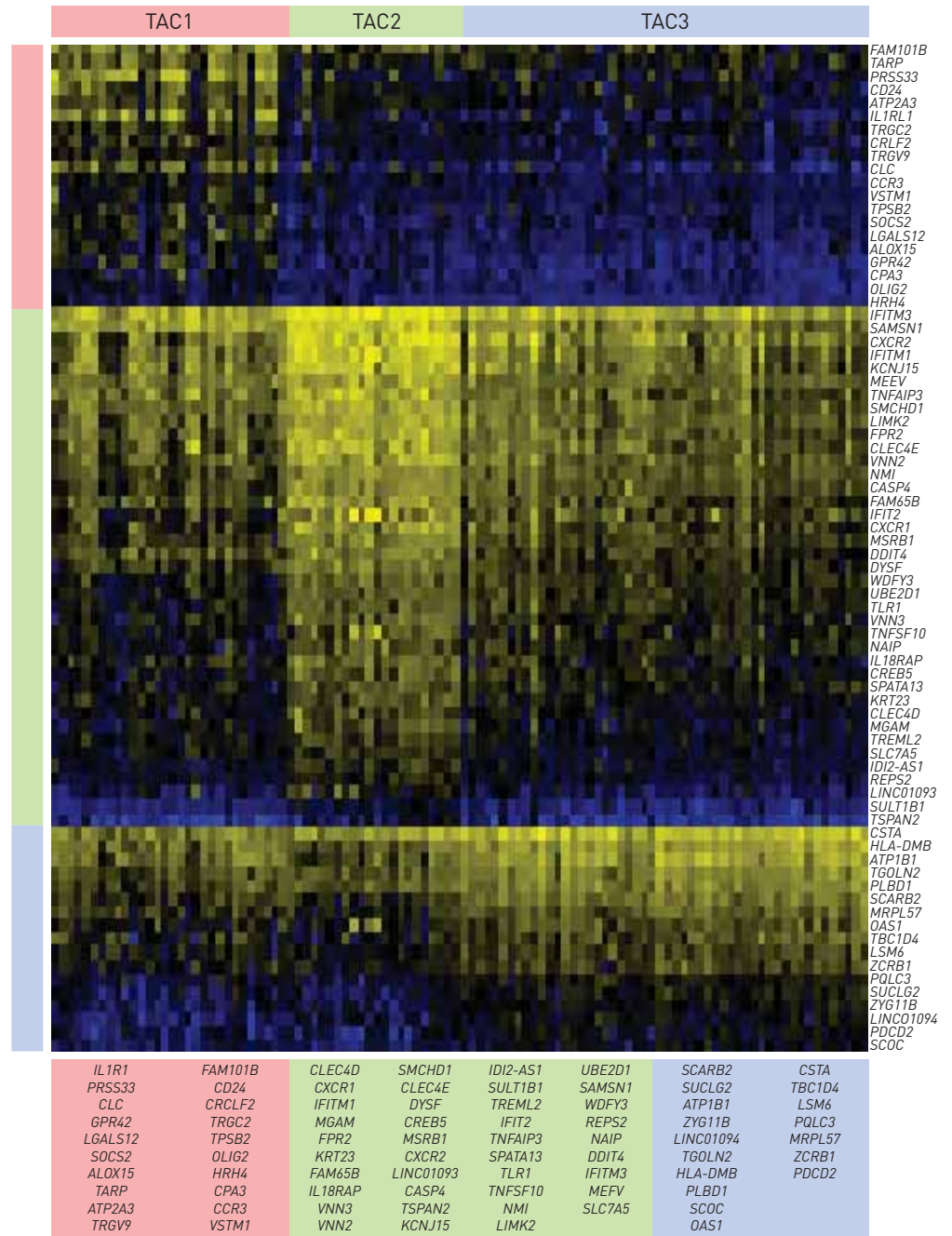
FIGURE 4 Heat map showing the signature of genes that best discriminate each transcriptome-associated cluster (TAC) derived using the shrunken centroid method. Columns represents 104 asthmatic subjects and rows represent 76 genes. The signatures of genes in the corresponding colours of each TAC (TAC1: 20 genes; TAC2: 39 genes; TAC3: 17 genes) are shown.

Network of each TAC-related gene and protein

As biological processes are commonly regulated by coexpressed genes, each TAC signature only represents the most characteristic genes as a subset of all the genes involved in each biological process. We therefore attempted to understand the coexpression relationship of all TAC-related genes and proteins potentially functioning in the biological processes associated with asthma. We first filtered the individual genes based on a moderate-to-high positive correlation with the GSVA ES of each TAC gene signature. This produced three sets of TAC-related genes, containing 153 TAC1-related genes, 194 TAC2-related genes and 312 TAC3-related genes. A similar filtering scheme for individual proteins led to the identification of three sets of TAC-related proteins, containing 91 TAC1-related proteins, 98 TAC2-related proteins and 42 TAC3-related proteins. Following this, each TAC-related gene–protein network was displayed as a correlation
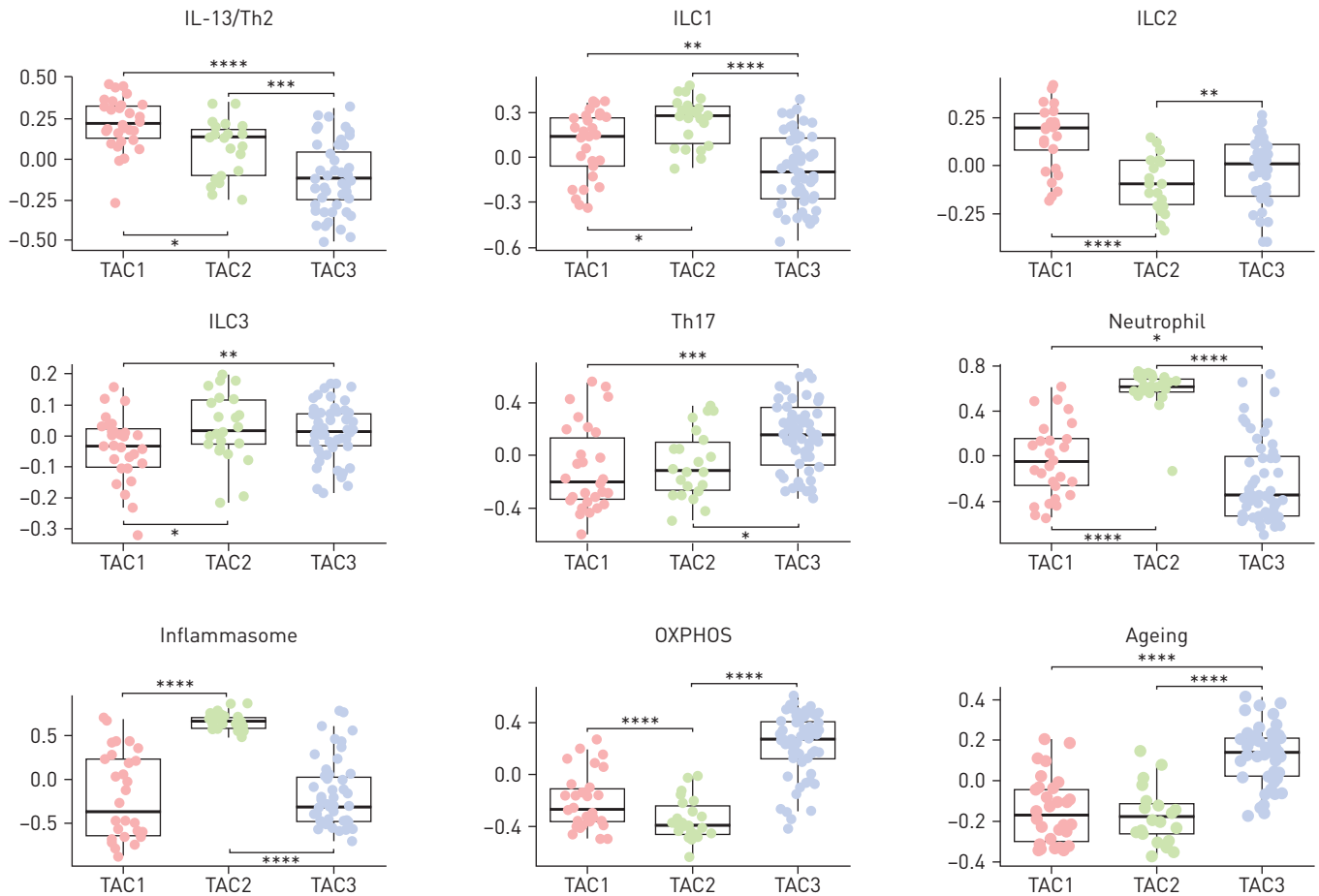
FIGURE 5 Gene set variation analysis (GSVA). Dot blot expression scores for nine different gene signatures as shown in online supplementary table S4 in the three transcriptome-associated clusters (TAC1, TAC2 and TAC3). IL: interleukin; Th: T-helper cell; ILC: innate lymphoid cell; OXPHOS: oxidative phosphorylation. Data are presented as individual data points, and box-and-whisker plots showing median and interquartile range. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 10^{-3}$; ****: $p < 10^{-5}$.

matrix (online supplementary figure S3). We found that the three TAC-related gene–protein networks presented mild-to-moderate mean gene–protein correlations (online supplementary figure S3, blue frame; mean Pearson's correlation: TAC1 r=0.292, TAC2 r=0.416 and TAC3 r=0.403), suggesting a functional coherence of the three sets of TAC-related genes and proteins.

Gene–protein relationships in TACs
We dissected the mechanistic implication of the gene signatures and the related proteins in sputum using the shrunken centroid algorithm. We defined 10 proteins for TAC1 (including a metalloproteinase (PAPPA), a chemokine (CCL4L1) and a sulfatase (ARSB)), 16 proteins for TAC2 (including those related to the proteasome (PSMA1), phospholipase (PLCG1) and TNF-α (TNFAIP6)) and two proteins (cathepsins CTSG and CTSB) for TAC3 (online supplementary table S3). We established three coexpression networks using each TAC signature (figure 8) and showed that the mean gene–protein correlations of the TAC1 (r=0.49, p<0.001) and TAC2 (r=0.46, p<0.005) networks were significantly higher than random gene–protein relationships (online supplementary figure S4), indicating that the genes and proteins of TAC1 and TAC2 are regulated by similar mechanisms. Highly coexpressed *IL33R*/ARSB (r=0.78, p=4.2×10^{-16}), *IL33R*/PAPPA (r=0.73, p=2.3×10^{-13}) and *CLC*/PAPPA (r=0.79, p=9.7×10^{-17}) featured in TAC1, and *IFITM3*/PGLYRP1 (r=0.67, p=9.0×10^{-11}), *IFITM1*/PGLYRP1 (r=0.63, p=2.3×10^{-9}) and *MEFV*/PLCG1 (r=0.67, p=9.0×10^{-11}) featured in TAC2. The mean gene–protein correlation of the TAC3 network was similar to random gene–protein relationships (r=0.40, p=0.566; online supplementary figure S4), suggesting that TAC3 signatures were underlain by more diverse mechanisms, particularly post-transcriptional mechanisms.

Discussion
We have described three new TACs of severe asthma derived from a semibiased analysis of sputum transcriptomics using hierarchical clustering. TAC1 contains a greater enrichment for IL-13/Th2 and ILC2
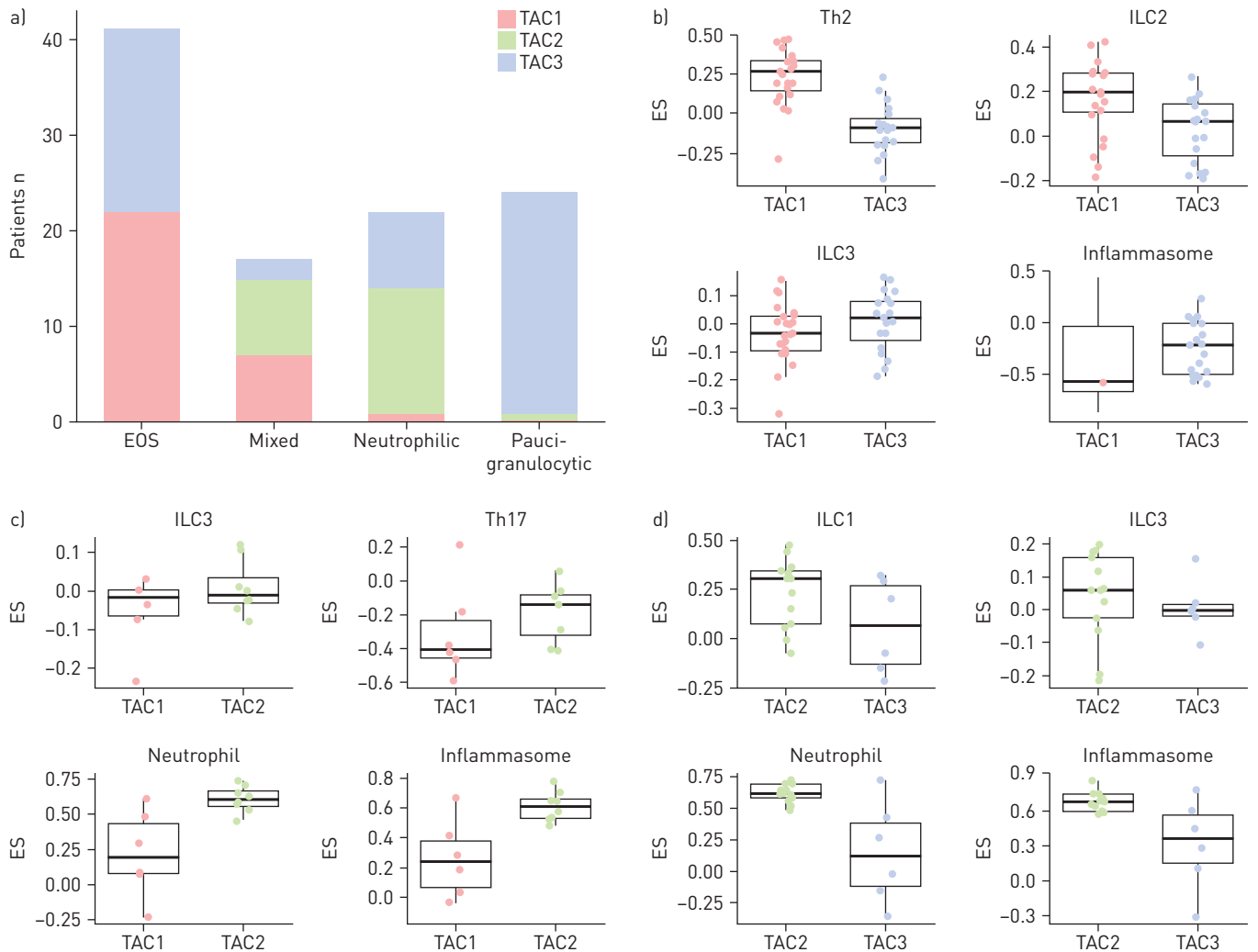
FIGURE 6 Distribution of transcriptome-associated clusters (TACs) according to granulocytic inflammation. a) Eosinophilic (EOS) inflammation links equally to TAC1 and TAC3, and mixed granulocytic (Mixed) links mainly to TAC1 and TAC2. Neutrophilic (Neu) and paucigranulocytic (Pauci) phenotypes match, respectively, to TAC2 and TAC3 (Chi-squared test $p=3.2\times10^{-13}$). b) Within the EOS-predominant phenotype, the TAC1 subtype (n=23) had a higher T-helper cell type 2 (Th2) (enrichment score (ES) difference 0.33, $p=10^{-7}$) and innate lymphoid cell type 2 (ILC2) (ES difference 0.14, $p=0.01$) signature, while the TAC3 subtype (n=19) showed a higher but nonsignificant inflammasome signature. ILC3 was similar between the two TAC subtypes. c) Within the mixed granulocytic phenotype, the TAC2 subtype (n=8) was slightly enriched in the neutrophil (ES difference 0.39, $p=0.02$) and inflammasome (ES difference 0.35, $p=0.004$) signatures, but the ILC3 signature was not enriched for the TAC1 subtype. d) Within the neutrophil-predominant phenotype, the TAC2 subtype (n=13) was enriched in the neutrophil (ES difference 0.47, $p=0.001$) and inflammasome (ES difference 0.37, $p=0.011$) signatures, with no significant differences for the ILC1 signature. Data in b–d) are presented as individual data points, and box-and-whisker plots showing median and interquartile range.

signatures, and is associated with blood and sputum eosinophilia, reflecting severe asthma characterised by mast cell and eosinophil activation and upregulation of receptors for TSLP, IL-33, IL-3 and CCL11 (CCR3). TAC2 is inflammasome-dominant with IFN and TNF superfamily upregulation and high expression of DAMPs (damage-associated molecular patterns), and is associated predominantly with neutrophilic inflammation and highest CRP levels and with chronic airflow obstruction of a lesser severity than that found in TAC1. The molecular signature for TAC3 highlighted metabolic, ubiquitination enzymes and mitochondrial energy metabolic genes, with the highest expression scores for mitochondrial oxidative stress (OXPHOS) and ageing gene signatures associated with paucigranulocytic and mild EOS inflammation. TAC3 is characterised by the lowest oral corticosteroid use, mild airflow obstruction and less frequent exacerbations than TAC1. Thus, the molecular phenotyping based on sputum cells yielded three distinct clinical clusters.

The gene signatures for TAC1 and TAC2 were highly coexpressed with the corresponding protein signature, indicating that similar levels of linked transcription–translation occur in each of the two classifications, and provided both gene and/or protein biomarkers and targets for potential new therapies for severe asthma. The highly coexpressed IL33R/ARSB, IL33R/PAPPA and CLC/PAPPA found in TAC1 emphasise the link
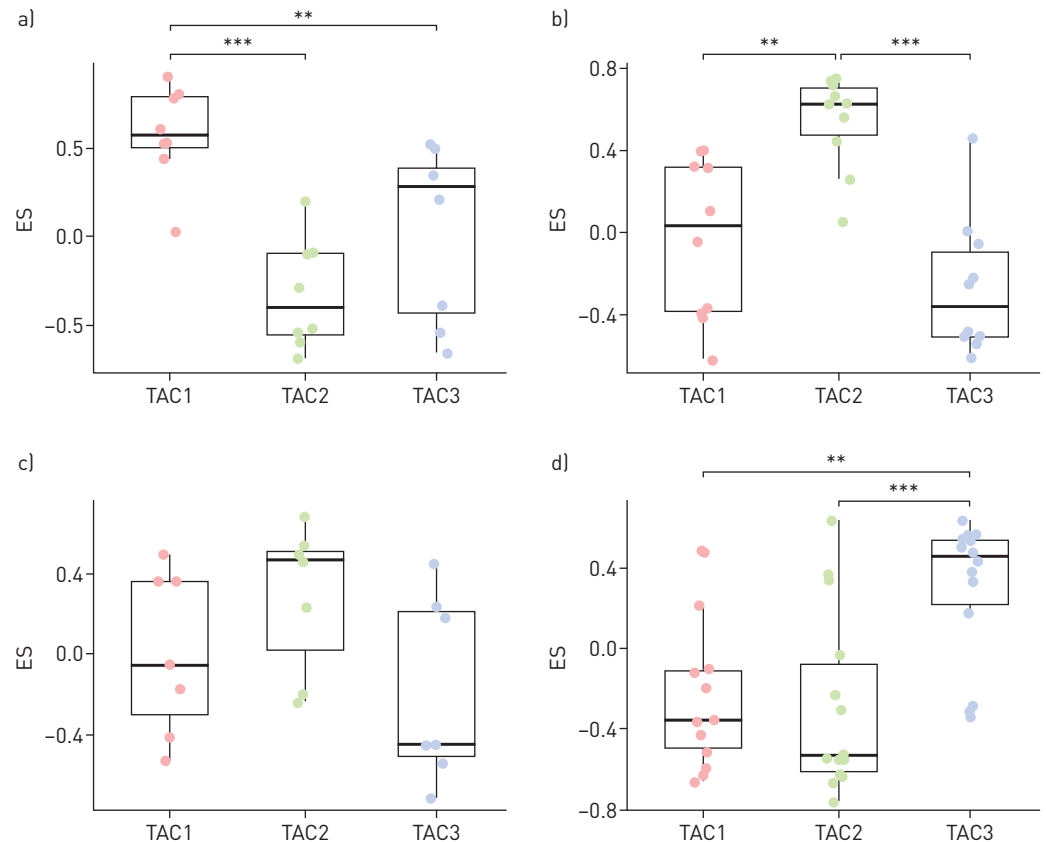
FIGURE 7 Enrichment of transcriptome-associated cluster (TAC) signatures in the ADEPT (Airways Disease Endotyping for Personalized Therapeutics) cohort according to sputum granulocytic inflammation. a) Eosinophil-predominant: ⩾1.5%. b) Neutrophil-predominant: ⩾74%. c) Mixed granulocytic: eosinophils ⩾1.5% and neutrophils ⩾74%. d) Paucigranulocytic: eosinophils <1.5% and neutrophils <74%. Data are presented as individual data points, and box-and-whisker plots showing median and interquartile range. *: p<0.05; **: p<0.01; ***: p<10^{-3}; ****: p<10^{-5}.

between IL-33R and eosinophil activation markers, which is in accord with the alarmin IL-33 being able to exacerbate eosinophil-mediated airway inflammation through the release of the Th2-based cytokines IL-5 and IL-3 [30]. In the TAC2 phenotype, the highly coexpressed *IFITM3*/PGLYRP1, *IFITM1*/PGLYRP1 and *MEFV*/PLCG1 reflect innate host defence responses to viruses and bacteria [31, 32].

This highly coexpressed relationship was not seen in the gene–protein signature of TAC3, indicating that it was likely driven by more complex regulatory factors such as post-translational modifications and altered metabolic enzyme activity rather than by classical cellular activation mechanisms. Sputum cells comprising macrophages and granulocytes are at the interface between the environment and the lung epithelial cell layer, and these TACs may reflect this interaction that is prominent in TAC2 with inflammasome activation, as reported previously [33]. Finally, these signatures were only partly validated in a separate smaller asthma cohort (ADEPT), thus extending the applicability of these signatures.

To try and understand the biological basis of these inflammatory phenotypes, we examined the distribution of the three TACs in relation to their sputum inflammatory phenotypes. The eosinophil-predominant (TAC1 or TAC3), neutrophil-dependent (TAC2 or TAC3) and mixed granulocytic-dominant (TAC1 or TAC2) phenotypes were split into two main biological classifications, whereas the paucigranulocytic-dominant (TAC3) phenotype was representative of only one TAC, *i.e.* TAC3. A major finding of this study was that at least two distinct transcriptional signatures were associated with sputum eosinophilia. The eosinophil-predominant TAC1 had a higher expression of IL-13/Th2 and ILC2 signatures, while the eosinophil-associated TAC3 was higher in the inflammasome signature, indicating a similarity of TAC1 to Th2-mediated allergic asthma and ILC2-mediated EOS asthma. This was also reflected by some of the genes (*e.g. IL33R, TSLPR*) involved in the ILC2 mechanism [34, 35] being in the top ranks of the TAC1 signature.

TAC1 was differentiated by higher blood eosinophils and serum periostin, but not by *F*eNO. We did not see a correlation with the traditional Th2 cytokines IL-4, IL-5 and IL-13, but TAC1 did associate with the signature obtained from IL-13-stimulated epithelial cells. However, this signal enrichment was only one
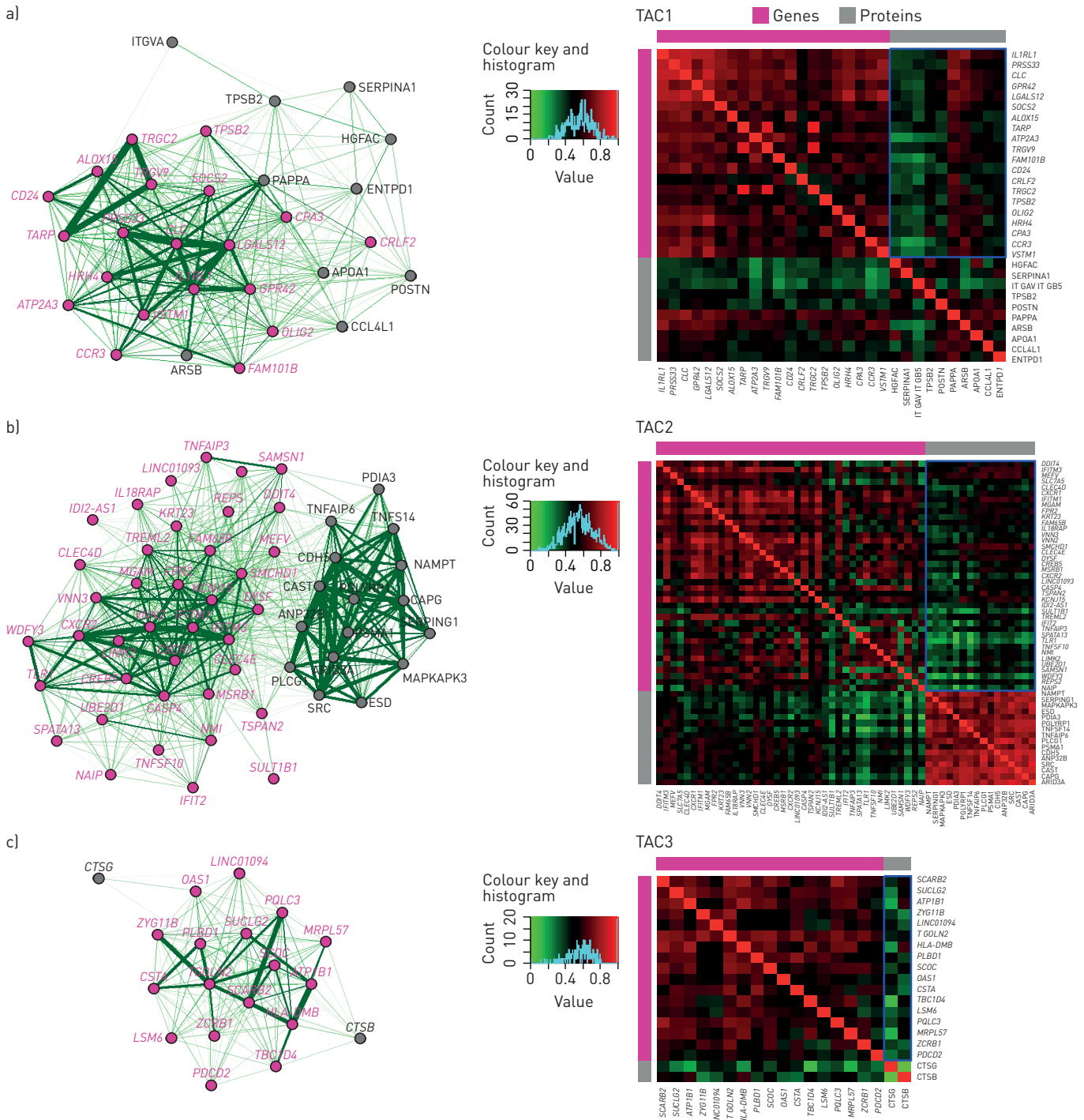
FIGURE 8 Coexpression network of each transcriptome-associated cluster (TAC) signature of genes (magenta nodes) and proteins (grey nodes). The linkage between each pair of nodes by an edge (green line) corresponds to the mutual correlation between two nodes as gene–gene, gene–protein or protein–protein relationships. Only mutual correlations of node pairs >0.5 are displayed in the network, where the thicker the edge, the closer the mutual correlation approaches 1.0. a) TAC1 has the highest gene–protein coexpression relationship with a high correlation of *IL33R*/ARSB, *IL33R*/PAPPA and *CLC*/PAPPA. b) TAC2 shows the second highest gene–protein coexpression relationship featuring modest correlation of *IFITM3*/PGLYRP1, *IFITM1*/PGLYRP1 and *MEFV*/PLCG1. TAC2 displays the highest protein–protein coexpression. c) TAC3 does not show high gene–protein coexpression although it is characterised focally by a modest correlation of *ATP1B1*/cathepsin B and *TBC1D4*/cathepsin B. A low correlation of protein–protein relationships is noted in TAC3 as the two proteins are distributed away from each other.

component of the TAC1 phenotype. Although TAC1 was uniquely EOS, which was not surprising considering that the clustering analysis was performed on DEGs from EOS compared with non-EOS and healthy subjects, EOS inflammation was found to different extents in all three TACs. This is likely to be

due to the fact that the signatures were driven by macrophage-derived genes in the sputum cells rather than eosinophils alone.

A higher enrichment of the inflammasome signature was observed in the TAC3 patients with eosinophil-predominant asthma compared with the TAC1 eosinophil-high patients. This was in line with our finding of ubiquitination enzyme genes that are essential for inflammasome assembly [36]. Moreover, activation of the inflammasome pathway can lead to the suppression of IL-33-associated EOS inflammation [37]. The gene set encoding multiple metabolic enzymes in the TAC3 signature was also echoed by a recent study linking NLRP3 inflammasome activation to the dysregulated metabolism of fatty acid and cholesterol in a mouse obesity-associated asthma model [38].

The TAC2 phenotype within the mixed granulocytic phenotype had a greater enrichment of neutrophil and inflammasome signatures, while the TAC1 subjects with mixed granulocytic asthma had a trend towards higher expression of the IL-13/Th2 signature. This suggests that TAC2 is dominated by neutrophil-driving mechanisms [39]. In contrast, TAC1 revealed a dominant IL-13/Th2 signature with a high EOS component. However, the mixed granulocytic TAC1 revealed a higher Th2 signature expression, suggesting that its EOS component was relatively dominant over its neutrophilic component, acting biologically in a similar way to the predominantly EOS phenotype. The biological distinction of a mixed granulocytic phenotype has been considered as a transitional phenotype with less stability [12, 13]. Our study supports the view that a mixed granulocytic phenotype is less biologically distinct, but rather dependent on the biology determined by either neutrophils or eosinophils.

The definition of TAC3 as being associated with mitochondrial oxidative stress and with ageing genes is new, and is of considerable interest because of the potential contribution of mitochondria to aspects of the ageing process, including cellular senescence and chronic inflammation [40]. Mitochondrial dysfunction has been reported in airway smooth muscle cells from patients with severe asthma, with evidence of oxidative phosphorylation [41], but how mitochondrial oxidative stress and ageing signatures drive asthma with little evidence of inflammation (paucigranulocytic inflammation and low serum CRP) is unclear. Further research examining the role of other cell types (*e.g.* macrophages and epithelial cells) may help determine these mechanisms.

In summary, our approach provides a fresh framework on which to phenotype asthma and a more precise targeting of specific treatments [42]. Future work is required to validate the biology of the non-Th2 pathways demonstrated here. As the stratification of these TACs is not entirely predicted by measuring sputum granulocytic inflammation and not all subjects were able to produce sputum, a point-of-care, preferably blood-based, biomarker of these TACs will be needed.

References

1   Chung KF, Wenzel SE, Brozek JL, *et al.* International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* 2014; 43: 343–373.

2   Choy DF, Modrek B, Abbas AR, *et al.* Gene expression patterns of Th2 inflammation and intercellular communication in asthmatic airways. *J Immunol* 2011; 186: 1861–1869.

3   Woodruff PG, Modrek B, Choy DF, *et al.* T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180: 388–395.

4   Peters MC, Mekonnen ZK, Yuan S, *et al.* Measures of gene expression in sputum cells can identify TH2-high and TH2-low subtypes of asthma. *J Allergy Clin Immunol* 2014; 133: 388–394.

5   Brusselle GG, Maes T, Bracke KR. Eosinophils in the spotlight: eosinophilic airway inflammation in nonallergic asthma. *Nat Med* 2013; 19: 977–979.

6   Lambrecht BN, Hammad H. The immunology of asthma. *Nat Immunol* 2015; 16: 45–56.

7   Cosmi L, Liotta F, Maggi E, *et al.* Th17 cells: new players in asthma pathogenesis. *Allergy* 2011; 66: 989–998.

8   Doe C, Bafadhel M, Siddiqui S, *et al.* Expression of the T helper 17-associated cytokines IL-17A and IL-17F in asthma and COPD. *Chest* 2010; 138: 1140–1147.

9   Baines KJ, Simpson JL, Wood LG, *et al.* Systemic upregulation of neutrophil α-defensins and serine proteases in neutrophilic asthma. *Thorax* 2011; 66: 942–947.

10   Wood LG, Baines KJ, Fu J, *et al.* The neutrophilic inflammatory phenotype is associated with systemic inflammation in asthma. *Chest* 2012; 142: 86–93.

11   Brusselle GG, Provoost S, Bracke KR, *et al.* Inflammasomes in respiratory disease: from bench to bedside. *Chest* 2014; 145: 1121–1133.

12   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: 997–1007.

13   Baines KJ, Simpson JL, Wood LG, *et al.* Transcriptional phenotypes of asthma defined by gene expression profiling of induced sputum samples. *J Allergy Clin Immunol* 2011; 127: 153–160.

14   Moore WC, Fitzpatrick AM, Li X, *et al.* Clinical heterogeneity in the severe asthma research program. *Ann Am Thorac Soc* 2013; 10: Suppl., S118–S124.

15   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: 1308–1321.

16   Green RH, Brightling CE, Woltmann G, *et al.* Analysis of induced sputum in adults with asthma: identification of subgroup with isolated sputum neutrophilia and poor response to inhaled corticosteroids. *Thorax* 2002; 57: 875–879.

17   Hathout Y, Brody E, Clemens PR, *et al.* Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proc Natl Acad Sci USA* 2015; 112: 7153–7158.

18   Gold L, Ayers D, Bertino J, *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010; 5: e15004.

19   Athey BD, Braxenthaler M, Haas M, *et al.* tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Jt Summits Transl Sci Proc.* 2013; 2013: 6–8.

20   Irizarry RA, Bolstad BM, Collin F, *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003; 31: e15.

21   Tibshirani R, Hastie T, Narasimhan B, *et al.* Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* 2002; 99: 6567–6572.

22   Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572–1573.

23   Hayes DN, Monti S, Parmigiani G, *et al.* Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol* 2006; 24: 5079–5090.

24   Verhaak RG, Hoadley KA, Purdom E, *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* 2010; 17: 98–110.

25   Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.

26   Bao Z, Zhang C, Yan W, *et al.* BMP4, a strong better prognosis predictor, has a subtype preference and cell development association in gliomas. *J Transl Med* 2013; 11: 100.

27   Silkoff PE, Strambu I, Laviolette M, *et al.* Asthma characteristics and biomarkers from the Airways Disease Endotyping for Personalized Therapeutics (ADEPT) longitudinal profiling study. *Respir Res* 2015; 16: 142.

28   Reimand J, Arak T, Vilo J. g:Profiler – a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 2011; 39: W307–W315.

29   Chen GY, Tang J, Zheng P, *et al.* CD24 and Siglec-10 selectively repress tissue damage-induced immune responses. *Science* 2009; 323: 1722–1725.

30   Stolarski B, Kurowska-Stolarska M, Kewin P, *et al.* IL-33 exacerbates eosinophil-mediated airway inflammation. *J Immunol* 2010; 185: 3472–3480.

31   Herbert C, Zeng QX, Shanmugasundaram R, *et al.* Response of airway epithelial cells to double-stranded RNA in an allergic environment. *Transl Respir Med* 2014; 2: 11.

32   Yao X, Gao M, Dai C, *et al.* Peptidoglycan recognition protein 1 promotes house dust mite-induced airway inflammation in mice. *Am J Respir Cell Mol Biol* 2013; 49: 902–911.

33   Simpson JL, Phipps S, Baines KJ, *et al.* Elevated expression of the NLRP3 inflammasome in neutrophilic asthma. *Eur Respir J* 2014; 43: 1067–1076.

34   McKenzie AN, Spits H, Eberl G. Innate lymphoid cells in inflammation and immunity. *Immunity* 2014; 41: 366–374.

35   Walker JA, Barlow JL, McKenzie AN. Innate lymphoid cells – how did we miss them? *Nat Rev Immunol* 2013; 13: 75–87.

36   Minton K. Inflammasomes: ubiquitin lines up for inflammasome activity. *Nat Rev Immunol* 2014; 14: 580–581.

37   Madouri F, Guillou N, Fauconnier L, *et al.* Caspase-1 activation by NLRP3 inflammasome dampens IL-33-dependent house dust mite-induced allergic lung inflammation. *J Mol Cell Biol* 2015; 7: 351–365.

38   Kim HY, Lee HJ, Chang YJ, *et al.* Interleukin-17-producing innate lymphoid cells and the NLRP3 inflammasome facilitate obesity-associated airway hyperreactivity. *Nat Med* 2014; 20: 54–61.

39   Moore WC, Hastie AT, Li X, *et al.* Sputum neutrophil counts are associated with more severe asthma phenotypes using cluster analysis. *J Allergy Clin Immunol* 2014; 133: 1557–1563.

40   Sun N, Youle RJ, Finkel T. The mitochondrial basis of aging. *Mol Cell* 2016; 61: 654–666.

41   Trian T, Benard G, Begueret H, *et al.* Bronchial smooth muscle remodeling involves calcium-dependent enhanced mitochondrial biogenesis in asthma. *J Exp Med* 2007; 204: 3173–3181.

42   Chung KF. Defining phenotypes in asthma: a step towards personalized medicine. *Drugs* 2014; 74: 719–728.