Supplementary Data

Th2 and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in UBIOPRED

Chih-Hsi Scott Kuo[1.2], Stelios Pavlidis[3], Matthew Loza[3], Fred Baribaud[3], Anthony Rowe[3], Iaonnis Pandis[2], Ana Sousa[4], Julie Corfield[5], Ratko Djukanovic[6], Rene Lutter[7], Peter J. Sterk[7], Charles Auffray[8], Yike Guo[2], Ian M. Adcock[1†] & Kian Fan Chung[1†*] *on behalf of the U-BIOPRED consortium project team*[#]

[1]Airways Disease, National Heart & Lung Institute, Imperial College London, & Biomedical Research Unit, Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London, United Kingdom;

[2]Department of Computing & Data Science Institute, Imperial College London, United Kingdom;

[3]Janssen Research and Development, High Wycombe, Buckinghamshire, United Kingdom;

[4]Respiratory Therapeutic Unit, GSK, Stockley Park, United Kingdom;

[5]AstraZeneca R&D Molndal, Sweden and Areteva R&D, Nottingham, United Kingdom;

[6]Faculty of Medicine, Southampton University, Southampton, United Kingdom;

[7]Faculty of Medicine, University of Amsterdam, Amsterdam, Netherlands;

[8]European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL, Université de Lyon, France.

[†]Contributed equally

[#]Consortium project team members are listed under *Supplementary*

Materials

*To whom correspondence should be addressed: f.chung@imperial.ac.uk

List of the U-BIOPRED Consortium project team members

Uruj Hoda & Christos Rossios, Airways Disease, National Heart & Lung Institute, Imperial College London, UK & Biomedical Research Unit, Biomedical Research Unit, Royal Brompton & Harefield NHS Trust, London, UK; Elisabeth Bel, Faculty of Medicine, University of Amsterdam, Amsterdam, Netherlands; Navin Rao, Janssen Research and Development, High Wycombe, Buckinghamshire, United Kingdom; David Myles, Respiratory Therapy Area Unit, GlaxoSmithKline, Stockley Park, UK; Chris Compton, Discovery Medicine, GlaxoSmithKline, Stockley Park, UK; Marleen Van Geest, AstraZeneca R&D Molndal, Sweden; Peter Howarth & Graham Roberts, Faculty of Medicine, Southampton University, Southampton, UK and NIHR Southampton Respiratory Biomedical Research Unit, University Hospital Southampton, Southampton, UK; Diane Lefaudeux, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL, Université de Lyon, France; Bertrand De Meulder, European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL, Université de Lyon, France; Aruna T Bansal, Acclarogen Ltd, St John's Innovation Centre, Cambridge, CB4 0WS, UK; Richard Knowles, Knowles Consulting, Stevenage Bioscience Catalyst, Gunnels Wood Road, Stevenage SG1 2FX, UK; Damijn Erzen, Boehringer Ingelheim Pharma, Germany; Scott Wagers, BioSci Consulting, BioSci Consulting, Maasmechelen, Belgium; Norbert Krug, Immunology, Allergology and Clinical Inhalation, Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany; Tim Higenbottam, Corporate Clinical Development, Chiesi Pharmaceutics Ltd, Cheadle, UK. Current address: Allergy Therapeutics, West Sussex, UK; John Matthews, Genentech Inc, 1 DNA Drive, South San Francisco, CA 94080-4990, USA; Veit Erpenbeek, Translational Medicine - Respiratory Profiling, Novartis Institutes for BioMedical Research, Basel, Switzerland; Leon Carayannopoulos, Merck Inc. Kenilworth, New Jersey, USA;

Amanda Roberts, UBIOPRED Patient Input Platform, ELF, Sheffield, UK; David Supple, UBIOPRED Patient Input Platform, ELF, Sheffield, UK; Pim deBoer, UBIOPRED Patient Input Platform, ELF, Sheffield, UK; Massimo Caruso, Department of Clinical and Experimental Medicine Hospital University, University of Catania, Italy; Pascal Chanez, Département des Maladies Respiratoires, Laboratoire d'immunologie, Aix Marseille Université Marseille, France; Sven-Erik Dahlen, The Centre for Allergy Research, The Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden; Ildikó Horváth, Department of Pulmonology, Semmelweis University, Budapest, Hungary; Nobert Krug, Fraunhofer Institute for Toxicology and Experimental Medicine Hannover, Germany; Jacek Musial, Dept. of Medicine, Jagiellonian University Medical College, Krakow, Poland; Thomas Sandström, Dept of Medicine, Respiratory and Allergy unit, University Hospital, SE 901 85 Umeå, Sweden.

Methods:

Study design

This study analysed data from the recently-reported UBIOPRED cohort (1). 104 participants (**Supplementary Table S1**) with moderate-to-severe asthma and 16 healthy non-asthma volunteers (HV) from the U-BIOPRED cohort underwent sputum cell profile analysis (1). Pre- bronchodilator spirometry, exhaled nitric oxide (FeNO), skin prick tests, serum total IgE, serum periostin, and differential blood count were measured. The study was approved by the Ethics Committees of the recruiting centres. All participants gave written informed consent. The data and bioinformatic analyses are described below. Validation of the transcriptomic-associated clusters was performed using sputum transcriptomic data from the ADEPT asthma cohort (2).

Microarray analysis of sputum transcriptome

Sputum was induced by inhalation of hypertonic saline solution and sputum plugs were collected from which sputum cells and sputum supernatants were obtained as previously described(3). Cell pellets were stored in RNA stabilization buffer (Norgen Biotek, Thorhold, Canada). RNA purity (RIN >6) was measured by Agilent Bioanalyser (Agilent, Santa Clara, Calif). Expression profiling was studied using Affymetrix U133 Plus 2.0 microarrays (Affymetrix, Santa Clara, Calif). Raw data were quality assessed and pre-processed by robust multi-array average normalization. Probes of low expression were filtered by robust multi-array signal analysis for values <5 and also for batch/technical effects. The intensity of the raw probe sets were log base 2 transformed and normalized by the robust multi-array average (RMA) method (4). A regression based method (R package limma) 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. False discovery rate (FDR) using the Benjamini and Hochberg method was applied for p-value adjustment in relation to multiple tests.

SomaLogic Proteomic Technique

The SOMAscan proteomic assay is an array-based method measuring 1,096 proteins each assay run which had its technique described comprehensively elsewhere(5, 6). All proteomic measurements for sputum supernatants were performed by SomaLogic Inc., (Boulder, CO) blinded to all subjects' clinical and transcriptomic data. Briefly, every protein measured in the assay has its own fluorophore-tagged SOMAmer (DNA) as a targeted reagent. SOMAmers that are in complexes with their cognate proteins are captured by automated partitioning steps. Using a custom Agilent hybridization chip designed as the antisense probe array specifically hybridizes to the SOMAmers, the measurement of proteins was transformed to the measurement of the fluorescent intensity of the hybridized SOMAmers. Protein concentrations were originally reported in relative fluorescence units (RFU) while this concentration were $\log_{10}$-transformed before statistical analysis to reduce heteroscedasticity.

Pathway analysis of transcriptomic features

We analysed 508 differentially-expressed genes (DEG) from a comparison of the three groups of the UBIOPRED cohort (**Fig 1A, B; Supplementary Table S1**). We defined a sputum eosinophil count ≥1.5% as being eosinophilic and a neutrophil count ≥74% as neutrophilic, while pauci-granulocytic and mixed-granulocytic counts were below and above these thresholds, respectively (1). Three sets of differentially expressed genes (DEGs) from pairwise contrasts of sputum EOS and non-EOS phenotypes, and healthy volunteers (HV) were analysed in order to obtain disease

driver genes. A filtering criteria with a false discovery rate (FDR) <0.05 and $\log_2$ fold change >0.5 was applied.

Computational and statistical analyses

Datasets were uploaded and curated in the tranSMART system(7). Statistical analysis was performed using R environment for statistical computing. False discovery rate was used to address multiple test correction. Hierarchical clustering based on Euclidean distance was used for cluster exploration and a resampling based technique was conducted as a measurement of cluster number optimization. Supervised learning algorithms using the shrunken centroid method (8) was applied to the cluster findings to determine predictive signatures for each cluster and feature reduction methods were implemented along with the learning algorithms to obtain a sparse model to facilitate interpretation. Kruskal-Wallis or ANOVA test was used for multiple group comparison of continuous variables. All categorical variables were analyzed using Fisher's exact test and p-value <0.05 was considered statistically significant.

Optimal cluster number determination

In order to perform clustering of asthma subjects using transcriptomic features, we first determined the optimal cluster number from these 508 DEGs. Consensus clustering, a resampling technique taking into account the cluster consensus across multiple runs of a clustering algorithm, was used to address the issue of optimal cluster number (9-11). This method analyzes the *N* subjects' cluster consensus distribution based on an (*N* x *N*) matrix built under the proportion of clustering runs in which two subjects are clustered together. The optimal cluster number is therefore determined by finding a cluster number *K* where consensus matrix histogram

approximates a bimodal distribution at *K* cluster and a relative small increase of area under curve (AUC) of cumulative distribution function (CDF) at *K*+1 cluster.

The consensus matrix for clusters between K=2 and K=5 are shown **(Figure S1, upper & middle panel)**. We noted that the CDF curve of the consensus index at cluster number K=2 approximated a bimodal distribution (**Figure S1, lower panel, left,** *red line***)** yet the increase of AUC at K=3 **(Figure S1, lower panel, right)** was very large. Cluster number K=3 (**lower panel, left,** *yellow line***)** was an optimal choice where the consensus index still approached bimodal distribution while the increase of AUC at K=4 **(Figure S1, lower panel, right)** was relatively small.

Shrunken centroid model to determine sputum gene and protein signatures

The nearest shrunken centroid method (8) was used as a supervised learning algorithm to refine the signatures for the identified TACs. The centroids (average expression of each gene) for each TAC as well as the overall samples were calculated. Standardization of the centroids of each TAC was performed through dividing the difference of the cluster centroids and overall centroids by the within-cluster standard deviation of each signature. This standardized value was treated as an absolute value which was later shrunken by an amount Δ (threshold value). If the value of a given standardized centroid was shrunken to zero for all TACs, then this gene did not contribute to the signature model. Otherwise, a non-zero value of a standardized centroid after shrinkage was retained as a classifier for the given TAC. The amount of shrinkage was chosen by iterative cross-validation on the performance in terms of the accuracy (or error rate) produced by a set of centroids from the prediction of TAC classification of each sample.

Signatures summarized by gene set variation analysis (GSVA)

We sought to evaluate gene expression related to a variety of disease mechanism of asthma using *a priori* knowledge. To this end, GSVA calculates sample-wise enrichment scores (ES) irrespective of any group labels thus enabled the implementation of null hypothesis based statistical analysis(12, 13). Therefore, by annotating each subject using summarization of the genes related to each disease mechanism GSVA addressed the need regarding to comparing the expression of a set of genes between groups. We compiled 9 gene sets each related to a specific aspect of asthma (T**able S4)** and the ES was calculated for each gene set for each subject. ANOVA was used to analyze the ES differences among group means and the Student's t-test was applied to compare the ES differences between two means.

Analysis of TACs in ADEPT cohort

The sputum signature findings predictive of each TAC from U-BIOPRED were applied to sputum transcriptomic data obtained from the Disease Profiling of Asthma and Chronic Obstructive Pulmonary Disease (ADEPT) cohort(2) using GSVA (**Table S5**). Sputum samples from 38 asthmatic subjects with a range of asthma severity and 9 healthy volunteers were analysed by Affymetrix U133 microarray (Affymetrix, Santa Clara, Calif). The baseline characteristics of study subjects is shown in **Table S5**. We annotated each subject in the ADEPT cohort using the TAC signatures derived in U-BIOPRED sputum samples. An ES was calculated for each TAC signature using GSVA. ANOVA was used to analyze the ES differences among group means and the Tukey HSD test was applied for subsequent pairwise comparison of ES differences between two means.

Table S1. Demographic and clinical characteristics of 104 asthmatics and 16 healthy volunteers

Variables†	**Asthmatics**	**Healthy volunteers**
Age (years)	51.3±13.4	38.1±13.3
Female	60 (57.7)	4 (25.0)
BMI	27.8±5.2	25.8±2.8
Nasal polyp	34 (32.7)	1 (6.3)
Allergic rhinitis	42 (40.4)	2 (12.5)
Eczema	33 (31.7)	NA
Severe asthma	84 (80.8)	NA
Oral corticosteroid use	38 (36.5)	NA
Atopy	73 (70.2)	4 (25.0)
Exacerbation numbers (per year)	1.0 (0-3.0)	NA
FEV_1 (% predicted)	69.8 (53.9-85.7)	104.0(98.9-113.2)
Total serum IgE (IU/ml)	102.0 (44.3-217.5)	39.5 (14.5-99.4)
Blood leukocyte ($10^3/\mu l$)	7.45 (6.06-9.88)	5.75 (4.73-7.75)
Blood eosinophil ($10^3/\mu l$)	0.25 (0.12-0.40)	0.10 (0.09-0.16)
Blood neutroophil ($10^3/\mu l$)	4.41 (3.50-6.47)	3.23 (2.78-5.11)
Sputum eosinophil (%)	2.4 (0.2-12.5)	0 (0-0.2)
Sputum neutrophil (%)	58.1 (34.8-78.7)	40.5 (19.6-68.9)
FeNO (ppb)	26.0 (16.0-46.5)	17.0 (13.5-26.1)
Serum periostin (ng/ml)	49.0 (39.7-59.4)	46.2 (43.9-51.5)
CRP (mg/l)	3.0 (1.0-6.0)	1.0 (1.0-2.0)

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

Table S2: Top 10 pathways from public ontology databases of the three DEG sets

DEG set	Database ID	Name	*p*-value[#]
EOS vs. HV			
	GO:0006955	immune response	1.09E-09
	GO:0001816	cytokine production	3.81E-06
	GO:0002684	positive regulation of immune system process	6.43E-06
	GO:0045321	leukocyte activation	6.62E-05
	GO:0031347	regulation of defense response	0.0002
	GO:0019221	cytokine-mediated signaling pathway	0.0007
	GO:1904018	positive regulation of vasculature development	0.0052
	CORUM:2790	ETS2-ETS1 complex	0.0167
	REAC:168249	Innate Immune System	0.0258
	CORUM:5465	IKB(epsilon)-RELA-cREL complex	0.0498
non-EOS vs. HV			
	GO:0045321	leukocyte activation	4.69E-06
	GO:0046649	lymphocyte activation	0.000153
	CORUM:2790	ETS2-ETS1 complex	0.00553
	REAC:198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	0.00556
	GO:0002252	immune effector process	0.00899
	GO:0044194	cytolytic granule	0.009
	KEGG:05202	Transcriptional misregulation in cancer	0.0125
	GO:0016337	single organismal cell-cell adhesion	0.0192
	GO:0007159	leukocyte cell-cell adhesion	0.0293
	GO:0070489	T cell aggregation	0.0474
EOS vs. non-EOS			
	GO:0045088	regulation of innate immune response	4.18E-09
	GO:2000116	regulation of cysteine-type endopeptidase activity	1.19E-06
	GO:0071723	lipopeptide binding	4.57E-06
	GO:0002221	pattern recognition receptor signaling pathway	1.55E-05
	GO:0034341	response to interferon-gamma	0.00109
	REAC:166054	Activated TLR4 signalling	0.00161
	GO:0072557	IPAF inflammasome complex	0.00241
	GO:0050702	interleukin-1 beta secretion	0.00269
	KEGG:04621	NOD-like receptor signaling pathway	0.00464
	GO:0043122	regulation of I-kappaB kinase/NF-kappaB signaling	0.00813

DEG: differentially expressed gene, EOS: eosinophilic, HV: healthy volenteer, #: p-value by Bonferroni correction, GO: Gene Ontology, CORUM: Comprehensive Resource of Mammalian protein complexes, REAC: Reactom, KEGG: Kyoto Encyclopedia of Genes and Genomes

Table S3. Signatures of genes and proteins characteristic of each TAC

TAC1		TAC2		TAC3	
Gene	**Protein**	**Gene**	**Protein**	**Gene**	**Protein**
IL1RL1	PAPPA	*CLEC4D*	TNFAIP6	*SCARB2*	CTSG
PRSS33	ENTPD1	*CXCR1*	PLCG1	*SUCLG2*	CTSB
CLC	CCL4L1	*IFITM1*	PSMA1	*ATP1B1*	
GPR42	APOA1	*MGAM*	CDH5	*ZYG11B*	
LGALS12	ITGAV	*FPR2*	ANP32B	*LINC01094*	
SOCS2	ARSB	*KRT23*	SRC	*TGOLN2*	
ALOX15	POSTN	*FAM65B*	CAST	*HLA-DMB*	
TARP	SERPINA1	*IL18RAP*	CAPG	*PLBD1*	
ATP2A3	HGFAC	*VNN3*	ARID3A	*SCOC*	
TRGV9	TPSB2	*VNN2*	NAMPT	*OAS1*	
FAM101B		*SMCHD1*	SERPING1	*CSTA*	
CD24		*CLEC4E*	MAPKAPK3	*TBC1D4*	
CRLF2		*DYSF*	ESD	*LSM6*	
TRGC2		*CREB5*	PDIA3	*PQLC3*	
TPSB2		*MSRB1*	PGLYRP1	*MRPL57*	
OLIG2		*CXCR2*	TNFSF14	*ZCRB1*	
HRH4		*LINC01093*		*PDCD2*	
CPA3		*CASP4*			
CCR3		*TSPAN2*			
VSTM1		*KCNJ15*			
		IDI2-AS1			
		SULT1B1			
		TREML2			
		IFIT2			
		TNFAIP3			
		SPATA13			
		TLR1			
		TNFSF10			
		NMI			
		LIMK2			
		UBE2D1			
		SAMSN1			
		WDFY3			
		REPS2			

		NAIP			
		DDIT4			
		IFITM3			
		MEFV			
		SLC7A5			

TAC: transcriptome-associated cluster

Table S4: Nine gene sets representing specific disease mechanisms of asthma

Name of signature	Details	Reference
IL13 Th2	*CST1, CCL26, PRB2, PRB1, PRB3, POSTN, PRB4, ITLN1, ALOX15, SH2D1B, CA2, NOS2, FCGBP, FOXA3, SPDEF, CAPN14, DUOXA2, 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, CCBL1, SDCBP2, TMPRSS2, HYAL1, CCDC109B, FAM83D, TRAK1, TPK1, SLC7A1, CYP2C18, CDC42EP5, KCNS3, ADRA2A, MRAP2, SLC2A10, PPARG, FAM26E, ADCY4, WNT3, SLCO4A1, ALDH1A2, C10orf99, WDFY2*	(14)
ILC1	*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*	(15)
ILC2	*HPGDS KRT1 IL17RB TNFRSF19 PTGDR2 HPGD IL1RL1 PKIB C10orf128 RP11-345M22.1 FSTL4 GAP43 MBOAT2 KLRG1 CSGALNACT1 FCRL3 CLIC4 IL10RA HLF LGALS12 ZP1 CHDH RP11-440I14.2 A2M BACE2 RP11-345M22.2 P2RY1 FASLG NRIP3 MSRB3 NTRK1 LINC00340 PZP PPARG TNFRSF9 UBXN10 A2MP1 IFIT3 UTS2 CALCRL RAP1GAP2 GRK5*	(15)
ILC3	*FCER1G SH2D1B NCR2 NRP1 LINC00299 KIAA1324 LST1 AMICA1 IL23R PCDH9 VWA5A XCL1 SIGLEC7 PLCG2 KLRC1 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 NSMCE1 SCN1B ID2 SLC43A2 ABCB5 ADAM28 CAT RP11-563D10.1 TNFSF11 STARD3NL TNS3 COL4A4 RNF152*	(15)

	CA2 TOX FES KIAA1217 B4GALNT1 VAV3 GPR82 DOCK5 CD33 S100A13 SOST TLE1 PECAM1 BCAS1 RORC LRRN3 TLE3 NCR1 SYK HDAC9 CRTAM DOK3 TNFSF4 CD300C RP11-330A16.1 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 TCF4 RP11-428G5.5 RGS9 KIAA0087 IRF4 DACH1 NCAM1 IKZF2 CDR2L 03-Mar FCRL4 PVRL2 LTBR NLRP7 JAG2 TRDJ2 RP11-91K8.1 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 SLCO2A1 GRAMD1B CARD6 ARSJ CHMP4C SNAP91 KCNQ5 TSPAN13 NLRP2 LYN RP11-256L6.2 CXCL16 CCDC102B GSN-AS1 C19orf35 MACC1 SORT1 GPR133 CCDC171 ORAI3 SERPINH1..1 NPTXR PTPRD DERL3 PPP1R9A B3GALT5 P4HA2 TRAJ44 MRC2 SLFN12 PLEKHN1 ACP6 TIE1 TMED8 UNC93B1 RP11-279F6.3 RP11-510J16.3 RP11-277P12.20 MS4A6A C9orf66 TRIO PRKAR2B CYP24A1 KLRK1..4 MMP25 TST CTD-2325P2.4 MAN1A1 RPPH1 PPFIBP1 KLRK1..6 UBE2E2 SORCS1 KIAA1456..1 COL24A1 LRP1 TRAIP GPR97 FGD6 LMLN EHHADH ZNF461 CSF2 TRPM8 MCAM CD38 FAM167A C1orf159 OSBPL6	
Th17	*KLRB1 RORC PLXND1 CTSH ALOX5 PTPN13 IL4I1 C11orf75 NEFL HLF JAKMIP2 DSE LIMS1 HLA-DRB1 LTK HLA-DRB4 USP10 NR1D1 LCAT SAMD3 HSPG2*	(16)
Neutrophil	*ABTB1,AMPD2,C5orf6,CCR3,CDA,CKLFSF2,CLC,CREB5, CTBS,DcR1,EST,FCGR2B,FCGR3B,FLJ10298,FPRL1,FRAT2, GPR27,GPR43,HSPA6,IL8RA,IL8RB,KIAA0779,KIAA1126, KRT23,LENG4,LENG5,MAD,MGC10500,MGC14126, MGC16353,MPPE1,MSCP,NCF4,NRBF-2,PHC2,PROK2,RALB, RNF141,SEC14L1,SEPX1,STX3A,TM4-B,VMP1,VNN2,XPO6*	(17)
Inflammasome	*IL1B, NLRP3, CASP1, CASP4, CASP5*	(18)
OXPHOS	*OXPHOS, ND1, ND2, ND3, ND4, ND4L, ND5, ND6, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS5, NDUFS6, NDUFS7,*	(19)

	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, UQCRFS1, CYTB, CYC1, UQCRC1, UQCRC2, UQCRH, UQCRHL, UQCRB, UQCRQ, UQCR10, UQCR11, COX10, COX3, COX1, COX2, COX4I2, COX4I1, COX5A, COX5B, COX6A1, COX6A2, COX6B1, COX6B2, COX6C, COX7A1, COX7A2, COX7A2L, 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	
Ageing	*MMACHC, PDE7B, CTSS, HLA-DRA, LUZP1, C3, C1QB, BRINP3, C1orf210, DENND6B, APOD, KHDRBS2, DHDDS, VWF, GPER1, CALHM2, MPEG1, FCGR2A, GPNMB, CLASP2, MSL3, C4A, MGST1, SHARPIN, APPBP2, AIP, IGJ, RNASET2, FCGR2B, ANTXR1, HIST1H1C, C1QA, RAB40B, CD74, LYZ, HMGN2, TLX3, SRPR, RORB, GFAP, ARRB1, MT2A, PTBP2, ABCB6, ARL11, KITLG, MMP10, UGT2B17, FXYD1, ANXA3, BIRC7, CDKN1A, AMH, NPC2, SH3GLB1, HBB, PCSK6, GSTA1*	(20)

ILC: innate lymphoid cell.

Table S5. Demographic and clinical characteristics of ADEPT cohort

Variables[†] N (%)	**Asthmatics (n=38)**	**HV (n=9)**
Age (years)	44.7±13.1	28.7±9.3
Female	23 (60.5)	3 (33.3)
Asthma severity		
Severe	17 (44.7)	NA
Moderate	11 (28.9)	NA
Mild	10 (26.4)	NA
Sputum cell profile		
EOS predominant	7 (18.4)	NA
NEU predominant	10 (26.3)	NA
Mixed granulocytic	7 (18.4)	NA
Pauci-granulocytic	14 (36.9)	NA

EOS: eosinophil, NEU: neutrophil

Figure S1. Shrunken medoid analysis of TACs

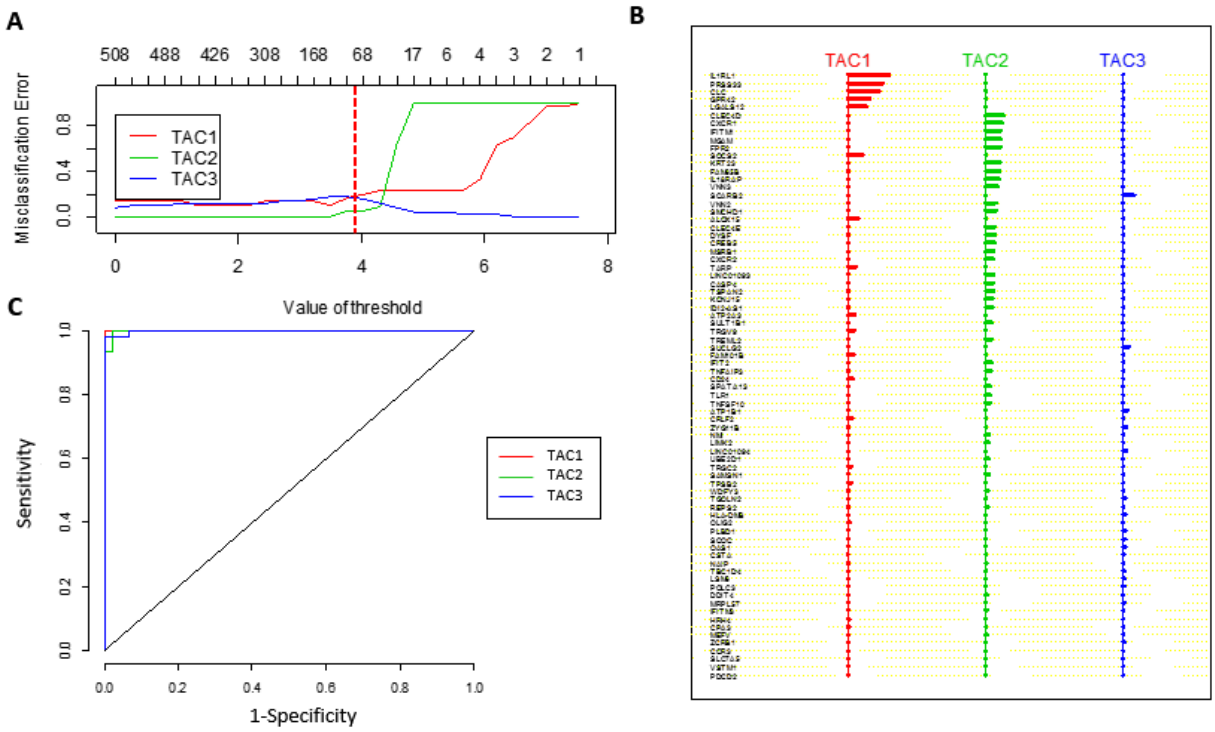


Figure S1 <u>Panel A</u>: Training of classifiers for the 3 clusters was evaluated by classification error using 10-fold cross-validation. A threshold of 3.95 (red broken line) was selected which enabled the reduction of classifiers to 76 genes at a cross-validated error <5%. **<u>Panel B</u>:** Centroid profile of the 76 signatures. Length of the centroid in each cluster denotes the relative amount the expression was deviating from the overall mean expression for each signature. Hence, the longer the bar in a given cluster, the higher the gene expressed with respect to the others. From top down, the centroids of each cluster were ranked in descending ordeer of magnitude. **<u>Panel C</u>:** Receiver Operating characteristic (ROC) curve showing the discriminative performance of the 76 signatures (mean AUC: 0.999) based on the probability model of cluster classification and one-vs.-rest approach. AUC under red, green and blue line

indicated classification performance of genes belonging to TAC1 (AUC: 1.000, p=8.1x10^{-17}), TAC2 (AUC: 0.998, p=1.0x10^{-16}) and TAC3 (AUC: 0.998, p=5.2x10^{-21}). Protein signatures characteritic of each TAC were analyzed using the same method for 71 subjects who also had samples available for supernatant protein analysis and 28 proteins were identified. The details of the 76 genes and 28 proteins characteritics of each TAC were shown in **Table S3.**

Figure S2. Enrichment of three TAC signatures according to asthma severity in ADEPT cohort.

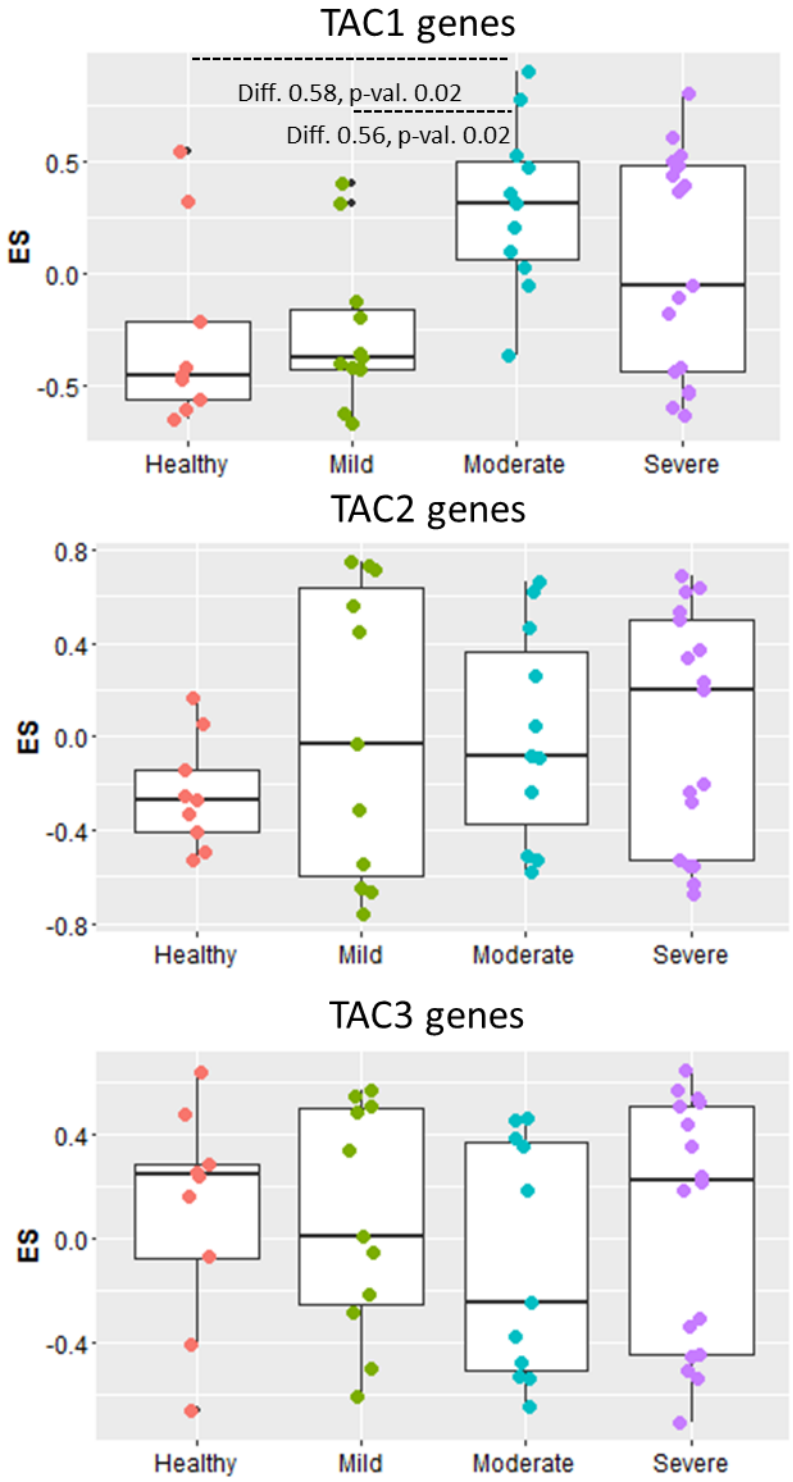


Figure S3. Correlation matrices built from all TAC signatures-related genes and proteins in sputum samples.

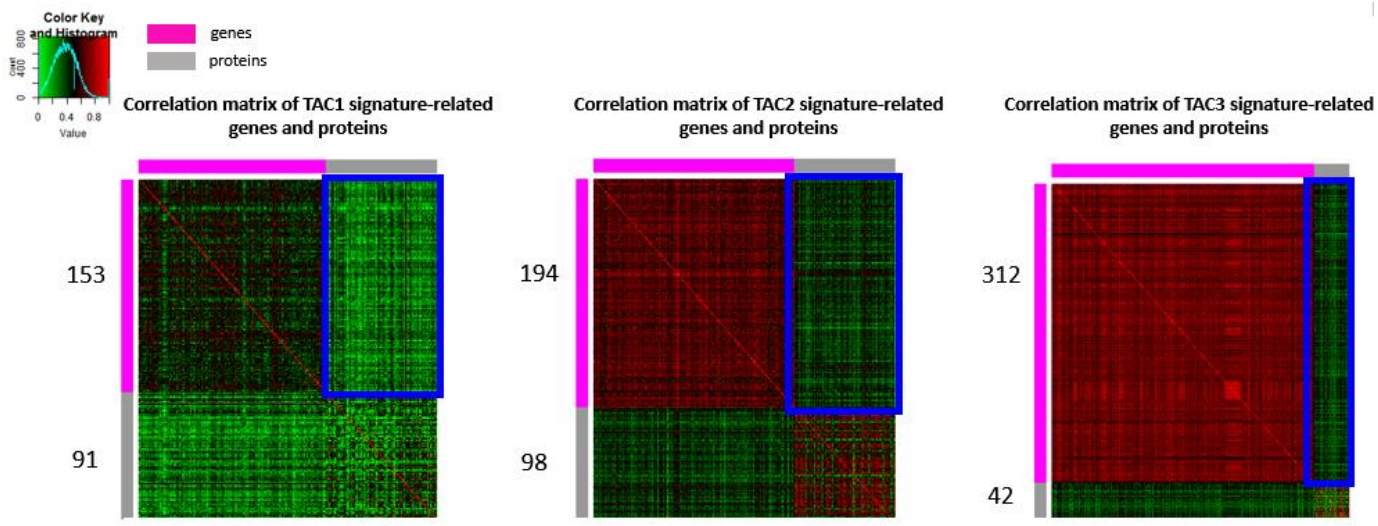
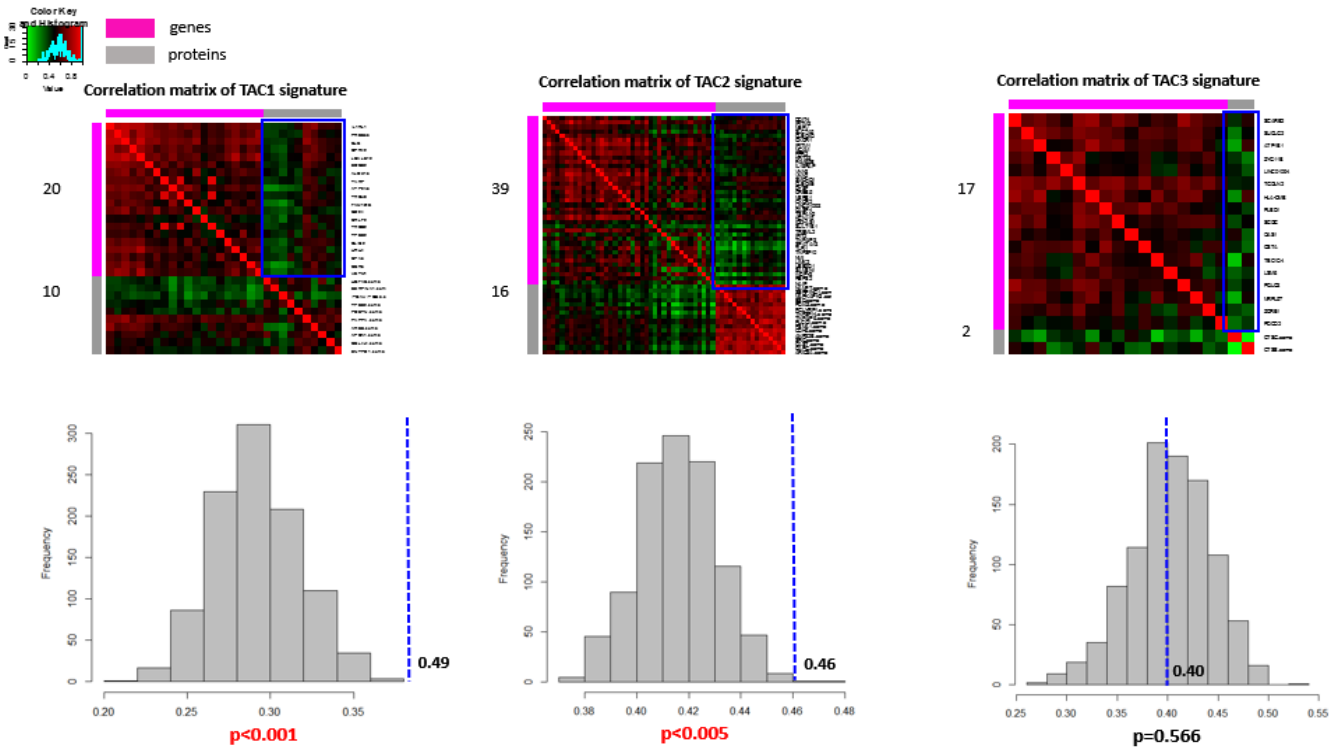


Figure S4. Distribution of mean gene-protein correlation following 1000 iterations of random samplings without replacement.



REFERENCES:

1. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, Pandis I, Bansal AT, Bel EH, Auffray C, Compton CH, Bisgaard H, Bucchioni E, Caruso M, Chanez P, Dahlen B, Dahlen SE, Dyson K, Frey U, Geiser T, Gerhardsson de Verdier M, Gibeon D, Guo YK, Hashimoto S, Hedlin G, Jeyasingham E, Hekking PP, Higenbottam T, Horvath I, Knox AJ, Krug N, Erpenbeck VJ, Larsson LX, Lazarinis N, Matthews JG, Middelveld R, Montuschi P, Musial J, Myles D, Pahus L, Sandstrom T, Seibold W, Singer F, Strandberg K, Vestbo J, Vissing N, von Garnier C, Adcock IM, Wagers S, Rowe A, Howarth P, Wagener AH, Djukanovic R, Sterk PJ, Chung KF. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. *Eur Respir J* 2015; 46: 1308-1321.
2. Silkoff PE, Strambu I, Laviolette M, Singh D, FitzGerald JM, Lam S, Kelsen S, Eich A, Ludwig-Sengpiel A, Hupp GC, Backer V, Porsbjerg C, Girodet PO, Berger P, Leigh R, Kline JN, Dransfield M, Calhoun W, Hussaini A, Khatri S, Chanez P, Susulic VS, Barnathan ES, Curran M, Das AM, Brodmerkel C, Baribaud F, Loza MJ. Asthma characteristics and biomarkers from the Airways Disease Endotyping for Personalized Therapeutics (ADEPT) longitudinal profiling study. *Respir Res* 2015; 16: 142.
3. Green RH, Brightling CE, Woltmann G, Parker D, Wardlaw AJ, Pavord ID. 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.
4. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 2003; 31: e15.
5. Hathout Y, Brody E, Clemens PR, Cripe L, DeLisle RK, Furlong P, Gordish-Dressman H, Hache L, Henricson E, Hoffman EP, Kobayashi YM, Lorts A, Mah JK, McDonald C, Mehler B, Nelson S, Nikrad M, Singer B, Steele F, Sterling D, Sweeney HL, Williams S, Gold L. Large-scale serum protein biomarker discovery in Duchenne muscular dystrophy. *Proceedings of the National Academy of Sciences of the United States of America* 2015; 112: 7153-7158.
6. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE, Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, Heil J, Heilig J, Hicke B, Husar G, Janjic N, Jarvis T, Jennings S, Katilius E, Keeney TR, Kim N, Koch TH, Kraemer S, Kroiss L, Le N, Levine D, Lindsey W, Lollo B, Mayfield W, Mehan M, Mehler R, Nelson SK, Nelson M, Nieuwlandt D, Nikrad M, Ochsner U, Ostroff RM, Otis M, Parker T, Pietrasiewicz S, Resnicow DI, Rohloff J, Sanders G, Sattin S, Schneider D,

Singer B, Stanton M, Sterkel A, Stewart A, Stratford S, Vaught JD, Vrkljan M, Walker JJ, Watrobka M, Waugh S, Weiss A, Wilcox SK, Wolfson A, Wolk SK, Zhang C, Zichi D. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS one* 2010; 5: e15004.

7. Athey BD, Braxenthaler M, Haas M, Guo Y. tranSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research. *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science* 2013; 2013: 6-8.
8. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 2002; 99: 6567-6572.
9. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572-1573.
10. Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, Socinski MA, Perou C, Meyerson M. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2006; 24: 5079-5090.
11. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell* 2010; 17: 98-110.
12. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* 2013; 14: 7.
13. Bao Z, Zhang C, Yan W, Liu Y, Li M, Zhang W, Jiang T. BMP4, a strong better prognosis predictor, has a subtype preference and cell development association in gliomas. *Journal of translational medicine* 2013; 11: 100.
14. Alevy YG, Patel AC, Romero AG, Patel DA, Tucker J, Roswit WT, Miller CA, Heier RF, Byers DE, Brett TJ, Holtzman MJ. IL-13-induced airway mucus production is attenuated by MAPK13 inhibition. *J Clin Invest* 2012; 122: 4555-4568.
15. Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, Sandberg R, Mjosberg J. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nature immunology* 2016; 17: 451-460.

16. Zhang H, Nestor CE, Zhao S, Lentini A, Bohle B, Benson M, Wang H. Profiling of human CD4+ T-cell subsets identifies the TH2-specific noncoding RNA GATA3-AS1. *The Journal of allergy and clinical immunology* 2013; 132: 1005-1008.
17. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM, Chan AC, Clark HF. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes and immunity* 2005; 6: 319-331.
18. Simpson JL, Phipps S, Baines KJ, Oreo KM, Gunawardhana L, Gibson PG. Elevated expression of the NLRP3 inflammasome in neutrophilic asthma. *The European respiratory journal* 2014; 43: 1067-1076.
19. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics (Oxford, England)* 2011; 27: 1739-1740.
20. 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: 875-881.