Table of Contents Supplementary Appendix

List of Authors

Jose L. Gomez[1]**, Xiting Yan**[1], **Carole T. Holm**[1]**, Nicole Grant**[1]**, Qing Liu**[1]**, Lauren Cohn**[1]**, Vera Nezgovorova**[1]**, Deborah A. Meyers**[2]**, Eugene R. Bleecker**[2]**, Gina M. Crisafi**[3], **Nizar N. Jarjour**[3] **for the SARP Investigators, Linda Rogers**[4]**, Joan Reibman**[4]**, and Geoffrey L. Chupp**[1]

[1] Section of Pulmonary, Critical Care Medicine and Sleep and Department of Internal Medicine, Yale University School of Medicine, New Haven, CT

[2] Center for Human Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, NC

[3] Section of Pulmonary and Critical Care Medicine, University of Wisconsin School of Medicine, Madison, WI

[4] Department of Medicine and Environmental Medicine, New York University School of Medicine, New York, NY.

Methods

Subjects. Pulmonary function testing included baseline pre-bronchodilator spirometry and responsiveness to 2-8 puffs of short-acting beta-agonists. Atopy was assessed by skin prick testing and measurement of serum total immunoglobulin E (IgE). Fraction of exhaled nitric oxide (FeNO) was measured using a NIOX MINO device (Aerocrine, New Providence, NJ) at a constant flow rate. Blood was collected for complete blood count (CBC), IgE level quantitation (Healthpoint Diagnostix, Lincolnshire, IL) and additional analyses described below.

Medication use was recorded at the time of the study visit in all three cohorts and represented the current, stable dose of maintenance inhalers and oral therapies, including oral corticosteroids and leukotriene antagonists.

Measurement of Serum and Sputum YKL-40 Levels. Average values from duplicate measurements were used. Duplicate samples with coefficients of variation (CV) greater than 20% were re-assayed. The mean CV for all samples was 10%. The limit of detection was 5.4 ng/mL. Values under the limit of detection were defaulted to 0 ng/mL.

Sputum Induction and gene expression measurements. Using inhaled hypertonic saline as previously described. (1) Mucus plugs were removed using a dissecting microscope and washed to remove squamous cell contamination. The cellular and aqueous compartments were separated, and cell counts, cell differentials, and viability were determined (Diff-Quik and trypan blue exclusion). Aliquots of supernatants were stored until processing. Samples with >20% squamous cells were considered contaminated and were not processed further.

Total RNA was isolated from sputum cell pellets using the All-in-One purification kit (Norgen Biotek, Thorold, CAN), checked on an Agilent bioanalyzer and, if needed, treated again to remove DNA contamination (Qiagen, Gaithersburg, MD). Sputum RNA (10 ng) was amplified using the WT-Ovation Pico RNA Amplification System (NuGEN, San Carlos, CA) and processed per Affymetrix protocols. Samples were checked on an Agilent bioanalyzer. Samples with RNA integrity numbers less than 4.0 were rejected from the analysis. Purified total RNA from the sputum was processed for gene

expression using the Affymetrix HuGene 1.0 ST gene arrays following the manufacturer's protocols as previously described (Affymetrix, Santa Clara, CA). (2) The array data can be obtained from the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE56396.

Statistical Analysis. The following variables were used in the cluster analysis (Table E1): baseline FEV1% predicted, baseline FVC% predicted, baseline FEV1/FVC ratio, FEV1% predicted and FVC% predicted after maximal bronchodilation with albuterol, % of FEV1 bronchodilator response, age of asthma onset (years), asthma duration (years), gender (binary), frequency of β-agonist use (none; ≤1 time weekly; 2-3 times weekly; 1-2 times daily; or ≥3 times daily), inhaled corticosteroid (ICS) dose (mcg/day) and serum YKL-40 (ng/mL). ICS dose reporting was standardized to equivalent fluticasone doses in mcg/day. The subjects included in this analysis were required to have complete data available for all these variables.

In the YCAAD cohort, Ward's minimum-variance hierarchical clustering method was performed with the serum YKL-40 levels and eleven discriminant features identified by SARP. (3) The C-index was used to identify a cutoff in the number of clusters. (4). Following the identification of the four YKL-40 clusters in YCAAD, the rpart package (version 4.1-9) was used to generate a classifier model for these clusters. The tree generated using this approach was pruned to avoid overfitting the data. The classifier presented in Supplemental Figure E1 was then used to assign clusters in the SARP and NYUBAR cohorts.

In the YCAAD cohort, sputum gene expression was analyzed with Genespring version 12.6 (Agilent Technologies, Santa Clara, CA) and MetaCore version 6.23 build 67496 (Thomson Reuters, New York, NY). A Type 2 gene expression signature in the sputum was generated with a previously described method that uses the mean of the combination of gene expression of IL-4, IL-5 and IL-13. (2, 5) Pairwise comparisons were performed between individuals in all YKL-40 clusters and a group of healthy controls (HC) that completed the YCAAD phenotyping protocol during the study period (September 2009 and June 2012). Transcripts with a fold change ≥2 were selected, and an unpaired

Mann-Whitney test with a false discovery rate (FDR) of 10% was used to identify expression signatures in the airway. The single experiment MetaCore workflow was used for biological pathway and network analyses of differentially expressed transcripts identified in the pairwise comparison analyses between each YKL-40 cluster and the HC group, as described above. Enrichment analyses with an FDR-adjusted p-value <0.05 were considered significant.

Figures

Figure E1. YKL-40 Asthma Cluster Classifier

Abbreviations:

FEV1: Forced Expiratory Volume 1 Second; BD: Bronchodilator; BDR: Bronchodilator Response.

Figure E2. Common Transcripts in YKL-40 Clusters C3 and C4

Abbreviations:

C3: Cluster 3; C4: Cluster 4; HC: Healthy Controls.

Figure E3. Innate and Adaptive Immunity Transcripts in YKL-40 Asthma Cluster 3 vs Healthy

Controls

Abbreviations:

C3: Cluster 3; HC: Healthy Controls.

Figure E4. Interleukin-18 (IL-18) Pathway Transcripts in YKL-40 Asthma Cluster 4 vs Healthy

Controls

Abbreviations:

C4: Cluster 4; HC: Healthy Controls.

Tables

Table E1. Twelve discriminating features in the YKL-40 asthma clusters

Feature	YCAAD p-value	SARP p-value	NYUBAR p-value
Gender (binary)	8.6×10^{-13}	2.2×10^{-16}	2.2×10^{-16}
Age of onset (years)	5.8×10^{-6}	5.5×10^{-5}	3.9×10^{-13}
Asthma duration (years)	5.9×10^{-5}	1.6×10^{-4}	4.0×10^{-8}
Baseline FEV1% predicted	2.2×10^{-16}	1.2×10^{-13}	2.2×10^{-16}
Baseline FVC% predicted	2.2×10^{-16}	3.4×10^{-8}	3.3×10^{-15}
Baseline FEV1/FVC	1.4×10^{-7}	2.1×10^{-13}	1.9×10^{-14}
Maximal FEV1% predicted	5.8×10^{-16}	1.9×10^{-10}	4.2×10^{-16}
Maximal FVC% predicted	4.1×10^{-13}	1.0×10^{-6}	8.5×10^{-13}
Change in % predicted FEV1 post-bronchodilator (BDR)	0.002	7.7×10^{-7}	0.002
Frequency of β-agonist use	9.8×10^{-5}	0.16	0.11
Inhaled corticosteroid dose (mcg/day)	0.0002	0.01*	4.3×10^{-7}
Serum YKL-40 (ng/mL)	2.4×10^{-10}	0.015	9.9×10^{-6}

Definition of abbreviations: YCAAD: Yale Center for Asthma and Airways Disease; SARP: Severe Asthma Research Program; NYUBAR: New York University/Bellevue Asthma Registry; FEV1: Forced Expiratory Volume 1 Second; FVC: Forced Vital Capacity; BDR: Bronchodilator Response

*By categorical group: No steroids; low-to-moderate inhaled corticosteroid (ICS); high ICS; systemic corticosteroids.

Table E2. Characteristics of the YKL-40 clusters in all three cohorts

	Total cohort	Cluster 1	Cluster 2	Cluster 3	Cluster 4	p-value
All Cohorts	664	365 (55%)	184 (28%)	49 (7%)	66 (10%)	
Female gender	472 (71%)	359 (98%)	37 (20%)	28 (57%)	48 (73%)	<0.01
Age of onset (years)	16 (6-30)	15 (7-27)	12 (5-27)	6 (3-20)	43 (36-52)	<0.01
Asthma duration (years)	18 (9-31)	16 (8-28)	19 (11-30)	36 (31-48)	11 (4-19)	<0.01
BMI	28 (24-33)	27 (23-33)	27 (24-32)	31 (27-37)	31 (26-35)	<0.01
Obesity (BMI ≥30)	260 (39%)	130 (36%)	65 (35%)	28 (57%)	37 (56%)	<0.01
Severe by EPR-3						
FEV1 Pre-BD*	82 (68-93)	88 (79-97)	75 (65-89)	39 (31-48)	68 (53-76)	<0.01
FEV1 Post-BD*	88 (76-98)	92 (83-102)	87 (77-96)	48 (41-54)	74 (64-82)	<0.01
Bronchodilator Response *	5 (2-10)	4 (2-7)	9 (4-17)	6 (3-11)	6 (1-11)	<0.01
Serum YKL-40 (ng/mL)	55 (38-85)	49 (35-70)	55 (37-79)	78 (49-110)	125 (64-239)	<0.01
Serum IgE (IU/mL)	120 (36-309)	96 (29-252)	172 (66-392)	154 (40-301)	118 (40-303)	0.002

*Percent of predicted

Table E3. YKL-40 asthma clusters vs SARP clusters demonstrate a different distribution of the cluster assignment (p<0.001)

	YKL-40 Cluster 1	YKL-40 Cluster 2	YKL-40 Cluster 3	YKL-40 Cluster 4
Denominator	354*	184	49	65*
SARP Cluster 1	18% (62)	11% (20)	0% (0)	2% (1)
SARP Cluster 2	63% (223)	43% (80)	0% (0)	23% (15)
SARP Cluster 3	8% (28)	13% (24)	4% (2)	18% (12)
SARP Cluster 4	5% (19)	20% (37)	8% (4)	18% (12)
SARP Cluster 5	6% (22)	13% (23)	88% (43)	38% (25)

*Information to determine SARP cluster assignment was not available in nine C1 subjects and one C4 subject, all in NYUBAR. Combined number of individuals analyzed in the three cohorts: C1=365; C2=184; C3=49; C4=66.

Table E4. Demographic, physiologic and sputum characteristics of healthy controls

	Total cohort
Number of subjects	10
Age (years)	32 (25-42)
Female gender	7 (70%)
Race White	10 (100%)
Hispanic	0
BMI (kg/m^2)	23.9 (22.0-25.4)
Baseline lung function	
FEV1% predicted	97 (92-102)
FVC% predicted	103 (97-108)
FEV1/FVC	78 (77-83)
Maximal lung function	
FEV1% predicted	103 (91-108)
FVC% predicted	104 (95-108)
FEV1/FVC	81 (77-83)
Change in % predicted FEV1	3 (1-5)
Atopy status	6 (60%)
Sputum characterization (n)	
Eosinophil-predominant sputum ≥2%	1
Neutrophil-predominant sputum ≥40%	3
Paucigranulocytic sputum Eosinophils <2% and neutrophils <40%	4
Mixed granulocytic sputum Eosinophils ≥2% and Neutrophils ≥40%	2

Data are n (%) or median (interquartile range) unless otherwise specified.

Definition of Abbreviations: FEV1: Forced Expiratory Volume 1 second; FVC: Forced Vital Capacity

Table E5. Sputum transcriptome analysis: C1 vs Healthy Controls

Probeset ID	FDR-adjusted p-value	p-value	Fold Change C1 vs HC	Gene Symbol
7934898	0.01	0.00	3.36	ANKRD22
8040080	0.01	0.00	3.29	RSAD2
7917576	0.02	0.01	3.08	GBP5
8049528	0.02	0.01	2.88	LRRFIP1
8058765	0.01	0.00	2.85	FN1
8091511	0.01	0.00	2.81	P2RY14
8035779	0.03	0.01	2.65	ZNF626///ZNF93///ZNF253
8146092	0.04	0.02	2.55	IDO1
7917561	0.01	0.00	2.51	GBP4
8160559	0.02	0.01	2.50	DDX58
8112327	0.08	0.08	2.50	CKS1B
7979698	0.04	0.02	2.40	ATP6V1D
8149126	0.04	0.03	2.39	DEFA1B///DEFA3///DEFA1
8006594	0.03	0.02	2.38	CCL18///CCL18
8154233	0.02	0.01	2.36	CD274
8068713	0.01	0.00	2.29	MX1
8121515	0.01	0.00	2.29	SLC16A10
8036755	0.02	0.01	2.26	CLC
8166500	0.02	0.01	2.24	ZFX
7995783	0.00	0.00	2.23	MT2A
8091141	0.02	0.01	2.06	XRN1
8117219	0.08	0.08	2.03	ACOT13
8149097	0.02	0.01	2.02	DEFB1
8017843	0.08	0.08	2.02	SLC16A6
8000636	0.06	0.05	2.02	
8026490	0.06	0.06	-2.04	UCA1
8015349	0.04	0.03	-2.06	KRT19
7934997	0.01	0.00	-2.07	PPP1R3C
7954065	0.03	0.01	-2.11	GPRC5A
8120833	0.02	0.01	-2.17	SH3BGRL2
8100827	0.03	0.02	-2.27	IGJ
7953775	0.01	0.00	-2.32	A2ML1
7977621	0.02	0.01	-2.32	NDRG2
8177931	0.06	0.06	-2.35	MUC21
8118069	0.07	0.06	-2.36	MUC21
8152703	0.01	0.00	-2.36	FBXO32
8100675	0.01	0.00	-2.36	TMPRSS11A
7902738	0.01	0.00	-2.38	CLCA4
8015323	0.05	0.05	-2.39	KRT13
7922414	0.05	0.04	-2.40	SNORD76///GAS5
7903414	0.05	0.04	-2.40	AMY2B///AMY2A///AMY1C///AMY1B///AMY1A

7903440	0.04	0.02	-2.43	AMY2B///AMY2A///AMY1C///AMY1B///AMY1A
7918134	0.04	0.02	-2.43	AMY2B///AMY2A///AMY1C///AMY1B///AMY1A
8126905	0.01	0.00	-2.43	CRISP3
8037053	0.05	0.04	-2.51	CEACAM7
8029086	0.07	0.06	-2.57	CEACAM5
8043504	0.03	0.02	-2.58	MAL
8053797	0.02	0.01	-3.03	
8158167	0.01	0.00	-3.04	LCN2
8015060	0.02	0.01	-3.06	KRT24
8126820	0.01	0.01	-3.13	GPR110
8176624	0.01	0.00	-4.22	DDX3Y
8095380	0.01	0.00	-4.35	TMPRSS11E
7933327	0.00	0.00	-6.87	FAM25G///FAM25C///FAM25A

Abbreviations:
FDR: False Discovery Rate; HC: Healthy Controls.

Table E6. Enrichment analyses of the sputum transcriptome in C1 vs Healthy Controls

Enrichment by pathway maps

#	Maps	Total	In Data	FDR
1	Cytoskeleton remodeling_Keratin filaments	36	4	<0.01

Enrichment by diseases (by biomarkers)

#	Diseases	Total	In Data	FDR
1	Connective Tissue Diseases	2213	35	<0.01
2	Respiratory Tract Infections	515	16	<0.01
3	Sjogren's Syndrome	507	15	<0.01
4	Xerostomia	507	15	<0.01
5	Arthritis	1650	26	<0.01
6	Colonic Diseases	4412	45	<0.01
7	Joint Diseases	1680	26	<0.01
8	Arthritis, Rheumatoid	1446	24	<0.01
9	Rheumatic Diseases	1563	25	<0.01
10	Dry Eye Syndromes	563	15	<0.01

Enrichment by GO processes

#	Processes	Total	In Data	FDR
1	immune response	1761	28	<0.01
2	innate immune response	1211	21	<0.01
3	defense response	2018	27	<0.01
4	keratinocyte differentiation	148	8	<0.01
5	antibacterial humoral response	65	6	<0.01
6	antimicrobial humoral response	67	6	<0.01
7	epidermal cell differentiation	224	9	<0.01
8	defense response to other organism	559	13	<0.01
9	keratinization	75	6	<0.01
10	epidermis development	408	11	<0.01

Enrichment by Process Networks

#	Networks	Total	In Data	FDR
1	Cytoskeleton_Intermediate filaments	81	4	0.01
2	Immune response_Innate immune response to RNA viral infection	85	4	0.01

Abbreviations:
FDR: False Discovery Rate.

Table E7. Sputum transcriptome analysis: C2 vs Healthy Controls

Probeset ID	FDR-adjusted p-value	p-value	Fold Change C2 vs HC	Gene Symbol
8098439	0.05	0.00	5.65	EPCAM
7975453	0.08	0.04	3.77	SNORD56B
7914342	0.05	0.01	3.56	FABP3
8149126	0.08	0.04	3.52	DEFA1B///DEFA3///DEFA1
8066493	0.05	0.00	3.24	SLPI
8102800	0.09	0.05	3.15	SLC7A11
7964927	0.05	0.01	3.12	TSPAN8
7951034	0.08	0.04	2.86	SNORA8///TAF1D
8091511	0.05	0.01	2.59	P2RY14
8085716	0.05	0.01	2.58	SATB1
8137709	0.06	0.03	2.54	ZFAND2A
7943413	0.05	0.00	2.54	BIRC3
7908347	0.06	0.02	2.47	OCLM
8043995	0.06	0.02	2.45	IL1R1
8023696	0.06	0.02	2.31	SERPINB3
8124262	0.08	0.04	2.28	TDP2
7963946	0.05	0.00	2.25	MMP19
7906764	0.05	0.01	2.23	HSPA6
7940191	0.05	0.01	2.13	STX3
8016980	0.06	0.02	2.09	MIR142
8124492	0.09	0.05	2.08	HIST1H2BK
8133770	0.05	0.01	2.07	CCDC146
8017212	0.10	0.06	2.06	USP32
8014650	0.06	0.03	2.04	LOC440434///NPEPPS
8018902	0.05	0.01	2.02	DNAH17
8121429	0.09	0.05	-2.01	FIG4
8144758	0.05	0.01	-2.03	ZDHHC2
8072946	0.10	0.06	-2.06	EIF3L
8101429	0.07	0.03	-2.06	PLAC8
7941505	0.06	0.02	-2.07	CST6
8023191	0.09	0.05	-2.07	SMAD2
8106193	0.05	0.01	-2.08	UTP15
8133309	0.06	0.02	-2.08	PMS2P9///PMS2P6///DTX2P1-UPK3BP1-PMS2P11///LOC441259///PMS2P5///PMS2L2///PMS2P1
7900576	0.09	0.05	-2.10	PPIH
8143028	0.09	0.05	-2.10	CHCHD3
7987369	0.05	0.01	-2.10	DPH6
7964834	0.07	0.03	-2.28	CPM
7904303	0.05	0.00	-2.28	CD101
7939087	0.06	0.02	-2.32	ARL14EP
8150266	0.05	0.01	-2.36	LSM1

8011826	0.08	0.04	-2.38	SCIMP
7991357	0.09	0.05	-2.41	AP3S2
8125775	0.06	0.03	-2.42	UQCC2
7940473	0.05	0.00	-2.65	TMEM138
7905220	0.07	0.03	-2.91	ECM1///ECM1
7927363	0.06	0.02	-4.78	FAM25G///FAM25C///FAM25A
7933327	0.06	0.02	-4.78	FAM25G///FAM25C///FAM25A

Abbreviations:
FDR: False Discovery Rate; HC: Healthy Controls.

Table E8. Enrichment analyses of the sputum transcriptome in C2 vs Healthy Controls

Enrichment by pathway maps

#	Maps	Total	In Data	FDR
1	Immune response_IL-5 signaling via JAK/STAT	57	4	<0.01

Enrichment by diseases (by biomarkers)

#	Diseases	Total	In Data	FDR
1	Hypersensitivity	1679	32	<0.01
2	Bronchial Diseases	1365	28	<0.01
3	Asthma	1312	27	<0.01
4	Respiratory Hypersensitivity	1319	27	<0.01
5	Hypersensitivity, Delayed	252	13	<0.01
6	Crohn's Disease	588	18	<0.01
7	Hypersensitivity, Immediate	1503	27	<0.01
8	Colitis, Ulcerative	666	18	<0.01
9	Dermatitis, Allergic Contact	251	12	<0.01
10	Colitis	689	18	<0.01

Enrichment by GO processes

#	Processes	Total	In Data	FDR
1	immunoglobulin production	91	7	<0.01
2	innate immune response in mucosa	35	5	<0.01
3	production of molecular mediator of immune response	110	7	<0.01
4	disruption of cells of other organism	40	5	<0.01
5	killing of cells of other organism	40	5	<0.01
6	mismatch repair	42	5	<0.01
7	hematopoietic stem cell migration	6	3	<0.01
8	regulation of T-helper 1 type immune response	45	5	<0.01
9	immune system process	3154	33	<0.01
10	mucosal immune response	49	5	<0.01

Enrichment by Process Networks

#	Networks	Total	In Data	FDR
1	Development_Regulation of angiogenesis	222	8	<0.01
2	Development_Blood vessel morphogenesis	228	6	0.02

Abbreviations:
FDR: False Discovery Rate.

Table E9. Common transcripts in the sputum transcriptome analysis: C3 & C4 vs Healthy Controls

FDR-adjusted p-value	Fold Change C3 vs HC	FDR-adjusted p-value	Fold Change C4 vs HC	Gene Symbol	Previous studies in asthma and airway biology
0.02	2.9	0.05	2.2	TNFAIP3	(6-8)
0.02	2.8	0.02	2.6	MIR21	(9-11)
0.02	2.8	0.01	2.9	HLA-DQA1	(12, 13)
0.02	2.6	0.03	2.4	IL1RAP	(14)
0.09	2.6	0.06	3.5	SNORD56B	
0.03	2.3	0.06	2	XPO6	
0.04	2.3	0.06	2.2	RRN3P3	
0.04	2.2	0.04	2.3	GPR97	
0.10	2	0.06	2.6	CCL18	(15-17)
0.05	-2.8	0.04	-2.6	IGJ	
0.02	-6.4	0.02	-7.5	RPS4Y1	

Abbreviations:
FDR: False Discovery Rate; HC: Healthy Controls.

Table E10. Sputum transcriptome analysis: C3 vs Healthy Controls

Probeset ID	FDR-adjusted p-value	p-value	Fold Change C3 vs HC	Gene Symbol
8149126	0.02	0.00	5.10	DEFA1B///DEFA3///DEFA1
7943158	0.02	0.01	3.87	SCARNA9
7934898	0.02	0.00	3.48	ANKRD22
7905079	0.02	0.00	3.27	HIST2H2AA4///HIST2H2AC///HIST2H2AA3
7919619	0.02	0.00	3.27	HIST2H2AA4///HIST2H2AC///HIST2H2AA3
7937335	0.02	0.00	2.93	IFITM1
8122265	0.02	0.01	2.92	TNFAIP3
7917576	0.04	0.02	2.84	GBP5
8145291	0.03	0.01	2.79	SLC25A37
8008885	0.02	0.00	2.79	MIR21
8118556	0.02	0.00	2.76	HLA-DQA1
8084794	0.02	0.00	2.62	IL1RAP
7975453	0.09	0.07	2.58	SNORD56B
8006123	0.04	0.02	2.39	CPD
8043995	0.04	0.03	2.31	IL1R1
8097903	0.03	0.01	2.30	TLR2
7922889	0.02	0.01	2.29	IVNS1ABP
8000482	0.03	0.02	2.28	XPO6
8000192	0.04	0.02	2.27	RRN3P3///RRN3P1///RRN3P2///RRN3
8080973	0.02	0.00	2.26	PPP4R2
8160531	0.03	0.01	2.23	C9orf72
7953749	0.04	0.02	2.21	CLEC4D
7996100	0.04	0.03	2.21	GPR97
8091778	0.06	0.05	2.18	SCARNA7
8011077	0.02	0.00	2.18	PITPNA
8125512	0.02	0.01	2.16	TAP1///PSMB8-AS1
8180061	0.02	0.01	2.16	TAP1///PSMB8-AS1
8152845	0.02	0.00	2.14	FAM49B
7995783	0.02	0.00	2.14	MT2A
8151711	0.03	0.01	2.14	NBN
8009653	0.02	0.00	2.13	CD300A
7987180	0.04	0.02	2.07	KATNBL1P6///KATNBL1
8062927	0.05	0.03	2.06	PI3
8006594	0.10	0.08	2.04	CCL18///CCL18
8088958	0.03	0.01	2.04	GBE1
8146934	0.04	0.03	2.04	LY96
8117106	0.03	0.01	2.03	RNF144B
7914603	0.04	0.03	2.02	RNF19B
7951034	0.06	0.04	2.02	SNORA8///TAF1D
8043993	0.02	0.00	2.01	LOC100131131
8116910	0.03	0.01	2.00	HIVEP1

7952046	0.03	0.02	-2.03	MPZL2
8152703	0.07	0.05	-2.05	FBXO32
8148553	0.02	0.01	-2.05	LY6K
7981724	0.06	0.04	-2.08	IGHG1///IGHD
8173941	0.08	0.06	-2.34	TSPAN6
8163908	0.05	0.03	-2.45	GGTA1P
7941505	0.02	0.00	-2.49	CST6
8100827	0.05	0.03	-2.77	IGJ
7899455	0.04	0.02	-2.86	PHACTR4
7997188	0.07	0.05	-3.12	HP
8116734	0.02	0.00	-3.29	LY86
8095364	0.04	0.02	-3.48	TMPRSS11E
7927363	0.02	0.00	-5.91	FAM25G///FAM25C///FAM25A
7933327	0.02	0.00	-5.91	FAM25G///FAM25C///FAM25A
8176375	0.02	0.00	-6.44	RPS4Y1
8100664	0.03	0.01	-6.59	TMPRSS11D
8176719	0.02	0.00	-8.62	EIF1AY

Abbreviations:
FDR: False Discovery Rate; HC: Healthy Controls.

Table E11. Enrichment analyses of the sputum transcriptome in C3 vs Healthy Controls

Enrichment by pathway maps

#	Maps	Total	In Data	FDR
1	NETosis in SLE	31	4	<0.01
2	Immune response_Bacterial infections in normal airways	49	4	<0.01
3	Immune response_HSP60 and HSP70/TLR signaling pathway	54	4	<0.01
4	Inhibition of neutrophil migration by pro-resolving lipid mediators in COPD	70	4	0.01
5	IL-1 beta-dependent CFTR expression	31	3	0.01
6	Immune response_TLR ligands	34	3	0.01
7	Cytoskeleton remodeling_Keratin filaments	36	3	0.01
8	Th17 cells in CF (mouse model)	49	3	0.02
9	Signal transduction_NF-kB activation pathways	51	3	0.02
10	TLRs-mediated IFN-alpha production by plasmacytoid dendritic cells in SLE	53	3	0.02

Enrichment by diseases (By Biomarkers)

#	Diseases	Total	In Data	FDR
1	Hypersensitivity	1535	41	<0.01
2	Hypersensitivity, Immediate	1362	35	<0.01
3	Bronchial Diseases	1223	33	<0.01
4	Lung Diseases, Obstructive	1640	38	<0.01
5	Asthma	1172	32	<0.01
6	Respiratory Hypersensitivity	1187	32	<0.01
7	Warts	27	8	<0.01
8	Arthritis, Rheumatoid	1446	34	<0.01
9	Arthritis	1650	36	<0.01
10	Rheumatic Diseases	1563	35	<0.01

Enrichment by GO processes

#	Processes	Total	In Data	FDR
1	immune response	1761	46	<0.01
2	immune system process	2863	54	<0.01
3	defense response	2018	45	<0.01
4	innate immune response	1211	35	<0.01
5	immune effector process	655	23	<0.01
6	humoral immune response	258	15	<0.01
7	response to biotic stimulus	1216	28	<0.01
8	defense response to other organism	559	19	<0.01
9	response to other organism	1170	27	<0.01
10	response to external biotic stimulus	1172	27	<0.01

Enrichment by Process Networks

#	Networks	Total	In Data	FDR
1	Inflammation_Innate inflammatory response	180	9	<0.01
2	Chemotaxis	137	7	<0.01
3	Development_Regulation of angiogenesis	223	7	0.02
4	Immune response_Antigen presentation	197	6	0.03
5	Immune response_T helper cell differentiation	140	5	0.03
6	Immune response_Innate immune response to RNA viral infection	85	4	0.03
7	Cell adhesion_Leucocyte chemotaxis	205	6	0.03
8	Autophagy_Autophagy	55	3	0.05
9	Inflammation_Interferon signaling	110	4	0.05
10	Inflammation_IL-4 signaling	115	4	0.06

Abbreviations:
FDR: False Discovery Rate; GO: Gene Ontology.

Table E12. Sputum transcriptome analysis: C4 vs Healthy Controls

Probeset ID	FDR-adjusted p-value	p-value	Fold Change C4 vs HC	Gene Symbol
7914342	0.02	0.00	4.40	FABP3
8036755	0.04	0.01	4.33	CLC
8091511	0.02	0.00	4.10	P2RY14
7961075	0.06	0.03	3.74	CD69
7908347	0.03	0.01	3.70	OCLM
7975453	0.06	0.04	3.50	SNORD56B
8058765	0.02	0.00	3.25	FN1
8044049	0.05	0.02	3.16	IL18RAP
8044035	0.04	0.02	2.96	IL18R1
8043981	0.03	0.01	2.95	IL1R2
8118556	0.01	0.00	2.93	HLA-DQA1
8040080	0.08	0.06	2.83	RSAD2
8125447	0.04	0.02	2.82	HLA-DRB1///HLA-DQB1///HLA-DRB5///HLA-DRB4///HLA-DRB3///HLA-DRB1
8000636	0.04	0.01	2.79	
7928999	0.04	0.01	2.76	LIPN
8006594	0.06	0.03	2.64	CCL18///CCL18
8008885	0.02	0.00	2.57	MIR21
8113591	0.05	0.02	2.56	PGGT1B
8084794	0.03	0.01	2.44	IL1RAP
8127987	0.05	0.02	2.38	SNORD50A
8055980	0.04	0.01	2.38	CYTIP
7953737	0.04	0.02	2.37	CLEC6A
7967322	0.08	0.06	2.35	HCAR2///HCAR3
7995552	0.03	0.01	2.32	CYLD
7956878	0.03	0.01	2.31	IRAK3
8027862	0.10	0.08	2.28	FFAR2
8113305	0.04	0.02	2.27	CHD1
7996100	0.04	0.01	2.27	GPR97
8000192	0.06	0.03	2.23	RRN3P3///RRN3P1///RRN3P2///RRN3
8122265	0.05	0.02	2.18	TNFAIP3
7963946	0.02	0.00	2.17	MMP19
8014361	0.04	0.01	2.17	CCL23
8084173	0.02	0.00	2.16	ATP11B
7922404	0.08	0.06	2.15	GAS5///SNORD80
8064868	0.05	0.02	2.13	GPCPD1
7957186	0.08	0.06	2.12	TBC1D15
8113113	0.01	0.00	2.08	KIAA0825
8091141	0.04	0.01	2.08	XRN1
8019885	0.09	0.06	2.07	SMCHD1
8086498	0.02	0.00	2.06	KIAA1143

8127234	0.04	0.01	2.03	DST
8084704	0.10	0.08	2.03	EIF4A2
7901299	0.06	0.04	2.03	CMPK1
8083260	0.06	0.04	2.02	CPA3
8149097	0.02	0.00	2.02	DEFB1
8066275	0.06	0.04	2.02	LOC100127886
8068713	0.06	0.03	2.02	MX1
8000482	0.06	0.04	2.02	XPO6
8148293	0.09	0.06	2.01	NSMCE2
8126905	0.06	0.03	-2.02	CRISP3
8176698	0.02	0.00	-2.06	TXLNGY
8036777	0.01	0.00	-2.06	FBL
7968234	0.07	0.05	-2.08	SNORA27
7913237	0.04	0.01	-2.08	CAMK2N1
7963555	0.09	0.06	-2.11	KRT78
8044391	0.02	0.00	-2.18	MERTK
8126820	0.10	0.08	-2.21	GPR110
8037053	0.09	0.06	-2.30	CEACAM7
8100827	0.04	0.01	-2.61	IGJ
7905220	0.04	0.01	-3.33	ECM1///ECM1
7920252	0.06	0.03	-3.37	S100A7
7933423	0.01	0.00	-6.97	FAM25G///FAM25C///FAM25A
8176375	0.02	0.00	-7.48	RPS4Y1

Abbreviations:
FDR: False Discovery Rate; HC: Healthy Controls.

Table E13. Enrichment analyses of the sputum transcriptome in C4 vs Healthy Controls

Enrichment by pathway maps

#	Maps	Total	In Data	FDR
1	SLE genetic marker-specific pathways in T cells	101	5	0.02

Enrichment by diseases (By Biomarkers)

#	Diseases	Total	In Data	FDR
1	Polyendocrinopathies, Autoimmune	12	9	<0.01
2	Uveomeningoencephalitic Syndrome	15	9	<0.01
3	Hemoglobinuria, Paroxysmal	16	9	<0.01
4	Hemoglobinuria	17	9	<0.01
5	Aortitis	7	7	<0.01
6	Peanut Hypersensitivity	7	7	<0.01
7	Vulvar Lichen Sclerosus	7	7	<0.01
8	Berylliosis	18	9	<0.01
9	Jaundice	22	9	<0.01
10	Anti-Glomerular Basement Membrane Disease	9	7	<0.01

Enrichment by GO processes

#	Processes	Total	In Data	FDR
1	immune response	1761	47	<0.01
2	innate immune response	1211	37	<0.01
3	defense response	2018	44	<0.01
4	immune system process	2863	52	<0.01
5	protein-carbohydrate complex subunit organization	8	7	<0.01
6	polysaccharide assembly with MHC class II protein complex	8	7	<0.01
7	protein-carbohydrate complex assembly	8	7	<0.01
8	antigen processing and presentation of polysaccharide antigen via MHC class II	8	7	<0.01
9	peptide antigen assembly with MHC class II protein complex	11	7	<0.01
10	MHC class II protein complex assembly	11	7	<0.01

Enrichment by Process Networks

#	Networks	Total	In Data	FDR
1	Inflammation_Innate inflammatory response	180	7	0.01
2	Immune response_Antigen presentation	197	7	0.01
3	Inflammation_Interferon signaling	110	5	0.02
4	Translation_Translation initiation	171	6	0.02
5	Inflammation_IL-4 signaling	115	5	0.02

Abbreviations:
FDR: False Discovery Rate; GO: Gene Ontology.

References

1.   Pizzichini E, Pizzichini MM, Efthimiadis A, Hargreave FE, Dolovich J. Measurement of inflammatory indices in induced sputum: effects of selection of sputum to minimize salivary contamination. Eur Respir J. 1996;9(6):1174-80.

2.   Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, et al. Noninvasive Analysis of the Sputum Transcriptome Discriminates Clinical Phenotypes of Asthma. Ann Am Thorac Soc. 2016;13 Suppl 1:S104-5.

3.   Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med. 2010;181(4):315-23.

4.   LJ H, JR L. A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin1976. p. 1072-80.

5.   Peters MC, Mekonnen ZK, Yuan S, Bhakta NR, Woodruff PG, Fahy JV. Measures of gene expression in sputum cells can identify TH2-high and TH2-low subtypes of asthma. J Allergy Clin Immunol. 2014;133(2):388-94.

6.   Kang NI, Yoon HY, Lee YR, Won M, Chung MJ, Park JW, et al. A20 attenuates allergic airway inflammation in mice. J Immunol. 2009;183(2):1488-95.

7.   Simpson LJ, Patel S, Bhakta NR, Choy DF, Brightbill HD, Ren X, et al. A microRNA upregulated in asthma airway T cells promotes TH2 cytokine production. Nat Immunol. 2014;15(12):1162-70.

8.   Schuijs MJ, Willart MA, Vergote K, Gras D, Deswarte K, Ege MJ, et al. Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. Science. 2015;349(6252):1106-10.

9.   Lu TX, Munitz A, Rothenberg ME. MicroRNA-21 is up-regulated in allergic airway inflammation and regulates IL-12p35 expression. J Immunol. 2009;182(8):4994-5002.

10.   Lu TX, Hartner J, Lim EJ, Fabry V, Mingler MK, Cole ET, et al. MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity. J Immunol. 2011;187(6):3362-73.

11.   Liu Y, Yang K, Shi H, Xu J, Zhang D, Wu Y, et al. MiR-21 modulates human airway smooth muscle cell proliferation and migration in asthma through regulation of PTEN expression. Exp Lung Res. 2015;41(10):535-45.

12.   Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, et al. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol. 2015;135(6):1502-10.

13.   Lasky-Su J, Himes BE, Raby BA, Klanderman BJ, Sylvia JS, Lange C, et al. HLA-DQ strikes again: genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. Clin Exp Allergy. 2012;42(12):1724-33.

14.   Savenije OE, Mahachie John JM, Granell R, Kerkhof M, Dijk FN, de Jongste JC, et al. Association of IL33-IL-1 receptor-like 1 (IL1RL1) pathway polymorphisms with wheezing phenotypes and asthma in childhood. J Allergy Clin Immunol. 2014;134(1):170-7.

15.   de Nadaï P, Charbonnier AS, Chenivesse C, Sénéchal S, Fournier C, Gilet J, et al. Involvement of CCL18 in allergic asthma. J Immunol. 2006;176(10):6286-93.

16.   Kim HB, Kim CK, Iijima K, Kobayashi T, Kita H. Protein microarray analysis in patients with asthma: elevation of the chemokine PARC/CCL18 in sputum. Chest. 2009;135(2):295-302.

17.   Gavala ML, Kelly EA, Esnault S, Kukreja S, Evans MD, Bertics PJ, et al. Segmental allergen challenge enhances chitinase activity and levels of CCL18 in mild atopic asthma. Clin Exp Allergy. 2013;43(2):187-97.