Supplemental Information

An interferon inducible signature of airway disease from blood gene expression profiling

Jeong H. Yun[1,2,3], Sool Lee[1], Pooja Srinivasa[1], Jarrett Morrow[1,3], Robert Chase[1], Aadbia Saferali[1,3], Zhonghui Xu[1,3], Russel Bowler[4], Michael Cho[1,2,3], Peter Castaldi[1,2,3], Craig P. Hersh[1,2,3]

1 Channing Division of Network Medicine, and 2 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA, 3 Harvard Medical School, Boston, MA, USA, 4 National Jewish Health, Denver, CO, USA

Supplementary Methods

COPDGene study

Institutional review board approval was obtained at 21 participating clinical centers, and written informed consent was obtained from all participants. Measurements were obtained during a period of clinical stability, with at least 30 days since their last exacerbation[1]. The current analysis included 2,396 subjects with quantitative chest CT scan analysis of airway phenotypes (AWT, Pi10, and WA%) and blood RNA sequencing and a subset of 23 subjects with bronchial epithelial brushings for gene expression which were performed following the Phase 2 visit[2]. Of these subjects, 2,181 (91%) participated in a longitudinal follow up program to prospectively assess exacerbations following the Phase 2 visit. Exacerbation was self-reported, and defined as an episode of increased dyspnea, cough and/or sputum production that required antibiotic and/or systemic steroid treatment. In the longitudinal follow up program, exacerbation information was collected by telephone call or web-based survey every 6 months.

Quantitative CT airway measurements

Volumetric CT was acquired at full inspiration and end expiration [1]. The detailed protocol and lung imaging findings from COPDGene have been reviewed previously [3]. Quantitative phenotyping of segmental airways using Thirona software was performed by measuring airway dimensions in 6 bronchial paths: right upper lobe apical bronchus (RB1), right middle lobe lateral bronchus (RB4), right lower lobe posterior basal bronchus (RB10), left upper lobe apicoposterior bronchus (LB1), superior lingular bronchus (LB4) and left lower lobe posterior basal bronchus (LB10) as per prior studies[4, 5]. The average airway wall thickness (AWT) and airway wall area

percent [WA% = total area of airway – area of airway lumen)/total area of airway x 100] were calculated. Pi10 was calculated by square root of the wall area at internal perimeter (Pi) of 10mm[6]. Percent emphysema was defined as total percentage of lung voxels with attenuation values less than -950 Hounsfield units on inspiratory scan.

RNA sequencing

At the COPDGene Phase 2 (5-year) visit, blood samples were collected into PAXgene blood RNA tubes and total RNA was extracted using PAXgene blood miRNA kit (Preanalytix)[7]. TruSeq Stranded Total RNA library prep kit (Illumina) was used to generate cDNA libraries depleted for globin genes using Ribo-Zero Globin kit. Sequencing was performed as 75 base pair paired-end runs on an Illumina HiSeq2000. Paired end reads were mapped to the human reference GRCh38 using STAR 2.4.0h, annotated to genes and exons using Ensembl version 81. Bronchial cell RNA sequencing has been published previously[8], which was obtained from flexible bronchoscopy with large airway brushing of right mainstem bronchus performed by a single physician.

Analysis of Gene Expression Omnibus (GEO) datasets

We chose datasets that were derived from peripheral blood that included healthy adult controls within each study to minimize the variability between different gene expression studies. We used the GEO2R tool (http://www.ncbi.nlm.nih.gov/geo/geo2r/) to identify the differentially expressed genes. Microarray data was normalized and log2 transformed. Differentially expressed genes were defined based on a false discovery rate (FDR) < 0.05 with absolute log2 fold change > 1.5. Overlap

of the same genes from differentially expressed gene lists were compared using a hypergeometric test. Full lists of GSE19491 and GSE60244 genes can be provided upon request.

Gene set variation analysis

Gene set variation analysis (GSVA) estimates variation of gene set enrichments by deriving sample-wise scores. The gene by sample matrix is transformed to gene-set by sample matrix using a non-parametric and unsupervised method [9]. The airway signature gene set was defined as the 18 genes associated with CT airway wall thickness. Peripheral blood gene signatures for type 1 interferon signaling were obtained from 110 genes in previously published datasets (GSE38351 and GSE71634) that showed >1.5 log2 fold increase in expression after IFN-α2 or IFN-β treatment from healthy control subjects [10, 11]. Log counts per million (CPM) was used for calculation of enrichment score. The virus signature was calculated as the mean of the normalized, log2 transformed expression of the 161 overexpressed genes minus that of the 235 under expressed genes upon respiratory viral infections as defined previously[12].

Publicly available single cell RNA-seq data analysis

FACS sorted blood and lung single cell gene expression data from Travaglini et al. [13] was processed using the Seurat R package version (3.0). Expression of gene sets was calculated using addModuleScore function.

Supplementary Results

Biological pathways associated with AWT score

To further understand the relevant biological pathways associated with high vs. low AWT score, we stratified subjects with AWT score >0 (n= 1,039) vs. ≤0 (n = 1,357). As expected, those with high expression of AWT score had increased expression of 748 genes primarily enriched for interferon signaling. Interestingly, those with low expression of AWT score had increased expression of 457 genes involved in developmental pathways such as *GLI1, GLI3, PTCH1, WNT2B, WNT7A*, and *NOG* (Table E3).

Subgroup and sensitivity analysis

We performed stratified analysis for potential confounders that could influence gene expression profiles including sex, current smoking status, inhaled corticosteroid (ICS) use, and asthma history. Male sex and lack of ICS use were associated with stronger association of AWT score genes and airway wall thickness compared to females and ICS users. Smoking status did not have a significant effect on the differential expression, which may be influenced by smaller number of current smokers with COPD (Table E6). Sensitivity analysis by excluding participants with asthma (patient report of physician diagnosed asthma) or with asthma-COPD overlap (physician diagnosed asthma before age 40 in moderate to severe COPD) did not change the association of AWT score with reported airway wall thickness and lung function (Table E7).

These results are consistent with previous studies on sex difference in type 1 IFN response[14] and ICS [15]. In a stratified analysis of ICS users, none of the ISGs were significantly associated with airway wall thickness, and even showed negative associations (Table E8). Similarly, although the sample size was small, we found stronger association of the bronchial AWT score with AWT, WA%

and Pi10 after excluding five individuals using ICS from the bronchial gene expression data, suggesting that ICS diminishes the AWT score in bronchial epithelial cells (Table E9). The clinical characteristics of COPD subjects with high vs. low AWT score with and without ICS were largely comparable except those patients with high AWT score not on ICS had prospective increase in exacerbation frequency (Table E10).

Table E1. Summary of publicly available datasets

GSE	Analysis	Comparison groups (number of samples)	Total number of samples in dataset	Cells	Health status	Platform	PMID
GSE38351	IFN1 signature	Healthy control (n=7) vs. healthy control stimulated with IFNα2 for 1.5 hr (n= 10)	74	Peripheral blood monocytes	Healthy	Affymetrix	22610275
GSE71634	IFN1 signature	Healthy control (n=20) without vs. with IFNβ simulation for 18hr (n=20)	40	PBMC	Healthy	Illumina	26644207
GSE19491	SLE associated genes	Healthy control (n= 24) vs. SLE patients (n=28)	498	Whole blood	Healthy vs. SLE vs. bacteremia	Illumina	20725040
	Infection associated genes	Healthy control (n=24) vs. Strep and/or staph bacteremia (n= 75)					
GSE60244	Respiratory bacterial infection	Healthy control (n= 40) vs. bacterial infection (n= 22)	158	Whole blood	Healthy vs. Hospitalized for lower respiratory infection	Illumina	26598315
	Respiratory viral infection	Healthy control (n= 40) vs. viral infection (n= 71)					

Table E2. Association of airway wall thickness score and quantitative CT measurements

Variables	WA%*		Pi10*		Emphysema %†		PRM^{fSAD}‡	
	beta (95% CI)	adj. p	beta (95% CI)	adj. p	beta (95% CI)	adj.p	beta (95% CI)	adj.p
All subjects (n= 2396)	0.24(0.14-0.34)	<0.001	0.22(0.12-0.32)	<0.001	0.24(0.15-0.33)	<0.001	0.10 (0.03-0.16)	0.1
COPD (n= 1028)	0.25(0.10-0.40)	0.05	0.25(0.08-0.41)	0.1	0.19(0.07-0.31)	0.07	0.17 (0.06-0.27)	0.07
Control (n= 1368)	0.11(0-0.22)	1.0	0.05(-0.05-0.15)	1.0	0.09(0.00-0.19)	1.0	0.008(-0.07-0.08)	1.0

CT measurements were standardized using z score

*Effect estimates (beta) from linear regression adjusted for age, gender, race, body mass index, smoking status, CT scanner model, emphysema%

†Log transformed. Effect estimates from linear regression adjusted for age, gender, race, BMI, smoking status, CT scanner model, and airway wall thickness

‡ Cube root transformed. Effect estimates from linear regression adjusted for age, gender, race, BMI, smoking status, CT scanner model and cube root transformed parametric response mapping emphysema %.

Abbreviations: WA% = wall area percent, PRM^{fSAD} = parametric response mapping functional small airway disease. adj.p = adjusted p value using Holm-Bonferroni method

Table E3. Differential expression analysis of high AWT vs. low AWT score

		High AWT score (n=1,039)	Low AWT score (n=1,357)
Number of upregulated genes		748	457
Enrichment Analysis	GO BP	Defense response to virus Response to Biotic Stimulus Response to Virus	Animal organ morphogenesis Biological adhesion Cell-cell signaling Neurogenesis
	MSigDB	Genes transcriptionally modulated in the blood of multiple sclerosis patinets in response to recombinant IFNB1 Genes upregulated in SARS-CoV-2 infection Genes upregulated in parainfluenza virus 3 infection in A549 cells Hallmark Interferon gamma response	Hallmark myogenesis

Genes associated with AWT score binary variable (AWT score > 0 vs. ≤0) at false discovery rate <0.05. Top 100 genes by false discovery rate were used for enrichment analysis. GO BP = GO biologic process, MSigDB = molecular signature database

Table E4. COPDGene bronchoscopy study subject characteristics

	Control (n=14)	COPD (n=9)	P value
Age, yr (mean(SD))	64.00 (5.72)	63.49 (3.41)	0.81
Female (%)	9 (64.3)	4 (44.4)	0.61
Non-Hispanic white (%)	8 (57.1)	8 (88.9)	0.26
FEV1% predicted (mean(SD))	101.25 (13.03)	73.98 (17.55)	<0.01
Bronchodilator response (%)	3 (21.4)	3 (33.3)	0.88
Inhaled corticosteroid use (%)	2 (14.3)	3 (33.3)	0.57
Airway wall thickness (mean(SD))	0.93 (0.18)	1.13 (0.16)	0.01
Pi10 (mean(SD))	1.84 (0.37)	2.34 (0.40)	0.01
Wall area %(mean(SD))	44.51 (7.4)	52.53 (5.1)	0.01
Current smokers (%)	2 (14.3)	4 (44.4)	0.26
Smoking history (PY, mean (SD))	25.48 (13.22)	38.72 (10.46)	0.02
WBC ($x10^3$, mean(SD))	6.41 (1.29)	7.03 (1.06)	0.25
Neutrophil % (mean(SD))	54.86 (9.84)	60.22 (5.97)	0.16
Lymphocyte % (mean(SD))	33.71 (8.96)	27.36 (6.21)	0.08
Monocyte % (mean(SD))	8.00 (1.4)	8.58 (1.95)	0.42
Eosinophil % (mean(SD))	2.71 (1.86)	2.70 (1.18)	0.98
Basophil % (mean(SD))	0.79 (0.43)	0.92 (0.23)	0.39

Abbreviations: NHW = non Hispanic white, the remainder was African American. PY = pack years, WBC = white blood cell

P values are based on t test, or chi-square test for proportion

Table E5. Association of meta-virus signature with prospective exacerbations

	Incidence Rate Ratio	95% CI	Adj.p
Exacerbation frequency	2.8	1.6-4.8	0.01

Negative binomial regression adjusted for age, gender, race, current smoking status, baseline FEV_1 % predicted, history of prior exacerbations

Abbreviations: adj.p = adjusted p value using Holm-Bonferroni method

Table E6. Effect of inhaled corticosteroids on the association between blood AWT score and CT airway measurements

	AWT		WA%		Pi10	
Groups	ICS users	ICS non-users	ICS users	ICS non-users	ICS users	ICS non-users
All subjects	0.03 (NS)	0.06***	1.44 (NS)	1.87**	0.08 (NS)	0.11**
COPD	0.02 (NS)	0.07 (NS)	0.56 (NS)	2.75(NS)	0.02 (NS)	0.1*
Control	0.04 (NS)	0.04 (NS)	2.2 (NS)	0.8 (NS)	0.08 (NS)	0.02 (NS)

Effect estimates (beta) from linear regression adjusted for age, gender, race, BMI, smoking status, CT scanner model, and emphysema %. N=2357. Excluded 39 subjects without inhaled corticosteroid usage information. ICS users N=526 (402 COPD, 124 controls); ICS non-users N= 1831 (599 COPD, 1232 controls)

Abbreviations: ICS = inhaled corticosteroid, NS = not significant, *** adjusted p <0.001, ** adjusted p <0.01, *adjusted p <0.05

Table E7. Effect of asthma and asthma-COPD overlap on the association between blood AWT score and phenotypes

Outcome variables	Excluding asthma		Excluding ACO		Original Analysis	
	Beta	P value	Beta	P value	Beta	P value
airway wall thickness*	0.28	<0.001	0.27	<0.001	0.25	<0.001
FEV_1% predicted†	-3.2	<0.01	-3.6	<0.001	-3.4	<0.01
FEV_1/FVC†	-0.03	<0.001	-0.03	<0.001	-0.03	<0.001
ΔFEV_1(ml) ‡	-55.7	<0.001	-57.9	<0.001	-35.5	<0.01

Effect estimates (beta) from linear regression adjusted for *age, gender, race, BMI, smoking status, CT scanner model, and emphysema %, †age, gender, race, BMI, smoking status, smoking history (pack years), emphysema %, ‡age, gender, race, BMI, smoking status, smoking history (pack years), baseline FEV1 % predicted, emphysema %.
Abbreviations: ACO = asthma-COPD overlap, defined by physician diagnosed asthma before age 40 in moderate to severe COPD, ΔFEV_1 = change in FEV_1(ml) over 5 years.
Subjects with asthma n= 385, subjects with ACO n= 77

Table E8. Subgroup analysis of blood AWT gene expression by sex, smoking status and inhaled corticosteroid use.

AWT genes	logFC					
	Current Smokers	Former smokers	Male	Female	ICS users	ICS nonusers
GPR15	**0.96**	**1.02**	**0.95**	**0.73**	0.28	**1.16**
IGLV3-21	**1.09**	0.93	**0.81**	0.24	0.87	**0.88**
IGHG3	0.96	**0.85**	**0.75**	0.39	1.1	**0.80**
CTD-2083E4.5	0.59	0.55	0.63	0.36	0.22	0.51
RIN2	0.23	0.08	**0.16**	0.13	0.09	0.17
IFI44	0.80	0.69	**0.66**	0.35	0.00	0.75
DAB2IP	-0.75	-0.68	**-0.67**	-0.67	-0.67	-0.51
IFI44L	0.96	1.04	**0.73**	0.35	-0.12	**0.89**
OAS1	0.49	0.29	0.33	0.16	-0.01	0.42
RSAD2	0.95	1.05	**0.79**	0.33	-0.20	**0.88**
DDX60	0.40	0.28	0.30	0.13	0.02	0.34
OAS3	0.65	0.57	0.49	0.22	-0.31	**0.61**
CMPK2	0.65	0.61	**0.53**	0.23	-0.20	0.64
OAS2	0.40	0.21	0.49	0.10	-0.10	**0.34**
KIF13A	-0.16	-0.08	-0.10	-0.04	-0.18	-0.16
CLEC9A	-0.42	-0.43	**-0.47**	-0.42	-0.42	-0.47
STEAP4	-0.18	-0.09	-0.12	-0.12	-0.10	-0.21
HERC5	0.47	0.52	0.38	0.17	-0.24	**0.48**

Abbreviations: logFC = log fold change, ICS = inhaled corticosteroid use
Log fold changes are shown. Bold indicates adjusted p value <0.05

Table E9. Association between bronchial epithelial AWT score and CT airway measurements after excluding inhaled corticosteroid users

Outcome variables*	Excluding ICS		Original analysis	
	Beta	p value	Beta	p value
Airway wall thickness	3.96	0.003	3.22	0.004
Wall Area %	3.33	0.02	2.82	0.01
Pi10	2.29	0.045	1.39	0.2

* CT airway measurements are standardized using z-score

Abbreviations: AWT = segmental airway wall thickness, WA% = wall area percent, ICS = inhaled corticosteroid use

Linear regression adjusted for age, sex, race, body mass index, smoking status, and emphysema (%). CT scanner model was identical for all subjects. Excluded 3 COPD and 2 control subjects who were using inhaled corticosteroids.

Table E10. Clinical characteristics of COPD subjects stratified by ICS use and AWT score

	ICS non-users			ICS users		
	Low AWT score (n=313)	High AWT score (n=286)	P value	Low AWT score (n=196)	High AWT score (n=206)	P value
Age (mean(SD))	69.1 (8.4)	66.5 (8.5)	<0.01	68.5 (8.3)	68.9 (7.5)	0.59
Female sex (%)	119 (38)	125 (44)	0.18	87 (44)	100 (49)	0.46
NHW race (%)	262 (84)	225 (78)	0.14	152 (78)	171 (83)	0.21
AWT score, mean(SD)	-0.27 (0.2)	0.36 (0.2)	<0.01	-0.27 (0.2)	0.33 (0.2)	<0.01
Exacerbations, mean(SD)*	0.08 (0.6)	0.26 (0.8)	0.03	0.7 (0.9)	0.7 (1.1)	0.70

*In subset of subject (n= 628) who had prospective exacerbation data.
Abbreviations: ICS = inhaled corticosteroid use, NHW = Non Hispanic White

Reference

1. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, Curran-Everett D, Silverman EK, Crapo JD. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J. Chronic Obstr. Pulm. Dis.* 2011; 7: 32–43.
2. Morrow J, Qiu W, DeMeo DL, Houston I, Pinto Plata VM, Celli BR, Marchetti N, Criner GJ, Bueno R, Washko GR, Glass K, Quackenbush J, Choi AMK, Silverman EK HC. Network Analysis of Gene Expression in Severe COPD Lung Tissue Samples. *Am J Respir Crit Care Med* 2015; 191: A1253.
3. Lynch DA, Newell JD, Washko GR, Bhatt SP, Galban CJ, Hoffman EA, Silverman EK, Bodduluri S, Diaz AA, San José Estépar R. Imaging Advances in Chronic Obstructive Pulmonary Disease. Insights from the Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene) Study. *Am. J. Respir. Crit. Care Med.* 2018; 199: 286–301.
4. Nakano Y, Wong JC, De Jong PA, Buzatu L, Nagao T, Coxson HO, Elliott WM, Hogg JC, Paré PD. The prediction of small airway dimensions using computed tomography. *Am. J. Respir. Crit. Care Med.* 2005; 171: 142–146.
5. Diaz AA, Han MK, Come CE, Estépar RSJ, Ross JC, Kim V, Dransfield MT, Curran-Everett D, Schroeder JD, Lynch DA, Tschirren J, Silverman EK, Washko GR. Effect of emphysema on CT scan measures of airway dimensions in smokers. *Chest* 2013; 143: 687–693.
6. Patel BD, Coxson HO, Pillai SG, Agustí AGN, Calverley PMA, Donner CF, Make BJ, Müller NL, Rennard SI, Vestbo J, Wouters EFM, Hiorns MP, Nakano Y, Camp PG, Nasute Fauerbach P V., Screaton NJ, Campbell EJ, Anderson WH, Paré PD, Levy RD, Lake SL, Silverman EK, Lomas DA. Airway Wall Thickening and Emphysema Show Independent Familial Aggregation in Chronic Obstructive Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 2008; 178: 500–505.
7. Parker MM, Chase RP, Lamb A, Reyes A, Saferali A, Yun JH, Himes BE, Silverman EK, Hersh CP, Castaldi PJ. RNA sequencing identifies novel non-coding RNA and exon-specific effects associated with cigarette smoking. *BMC Med. Genomics* 2017; 10: 58.
8. Morrow JD, Chase RP, Parker MM, Glass K, Seo M, Divo M, Owen CA, Castaldi P, Demeo DL, Silverman EK, Hersh CP. RNA-sequencing across three matched tissues reveals shared and tissue-specific gene expression and pathway signatures of COPD. *Respir. Res.* 2019; 20: 65.
9. Hänzelmann S, Castelo R, Guinney J, Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013; 14: 7.
10. Smiljanovic B, Grün JR, Biesen R, Schulte-Wrede U, Baumgrass R, Stuhlmüller B, Maslinski W, Hiepe F, Burmester GR, Radbruch A, Häupl T, Grützkau A. The multifaceted balance of TNF-α and type I/II interferon responses in SLE and RA: How monocytes manage the impact of cytokines. *J. Mol. Med.* 2012; 90: 1295–1309.
11. Clarelli F, Liberatore G, Sorosina M, Osiceanu AM, Esposito F, Mascia E, Santoro S, Pavan G, Colombo B, Moiola L, Martinelli V, Comi G, Martinelli-Boneschi F. Pharmacogenetic study of long-term response to interferon-β treatment in multiple sclerosis. *Pharmacogenomics J.* 2017; 17: 84–91.
12. Andres-Terre M, McGuire HM, Pouliot Y, Bongen E, Sweeney TE, Tato CM, Khatri P.

Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. *Immunity* 2015; 43: 1199–1211.

13. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit R V., Chang S, Conley SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020; 587: 619–625.
14. Bunders MJ, Altfeld M. Implications of Sex Differences in Immunity for SARS-CoV-2 Pathogenesis and Design of Therapeutic Interventions. *Immunity* 2020. p. 487–495.
15. Singanayagam A, Glanville N, Girkin JL, Ching YM, Marcellini A, Porter JD, Toussaint M, Walton RP, Finney LJ, Aniscenko J, Zhu J, Trujillo-Torralbo MB, Calderazzo MA, Grainge C, Loo SL, Veerati PC, Pathinayake PS, Nichol KS, Reid AT, James PL, Solari R, Wark PAB, Knight DA, Moffatt MF, Cookson WO, Edwards MR, Mallia P, Bartlett NW, Johnston SL. Corticosteroid suppression of antiviral immunity increases bacterial loads and mucus production in COPD exacerbations. *Nat. Commun.* 2018; 9: 2229.