CHEST Online Supplement

Distinguishing Smoking-Related Lung Disease Phenotypes Via Imaging and Molecular Features

Ehab Billatos, MD; Samuel Y. Ash, MD; Fenghai Duan, PhD; Ke Xu, BS; Justin Romanoff, AM; Helga Marques, MS; Elizabeth Moses, PhD; MeiLan K. Han, MD; Elizabeth A. Regan, MD, PhD; Russell P. Bowler, MD, PhD; Stefanie E. Mason, MD; Tracy J. Doyle, MD; Rubén San José Estépar, MSc; Ivan O. Rosas, MD; James C. Ross, PhD; Xiaohui Xiao, BS; Hanqiao Liu, MD; Gang Liu, PhD; Gauthaman Sukumar, PhD; Matthew Wilkerson, PhD; Clifton Dalgard, PhD; Christopher Stevenson, PhD; Duncan Whitney, PhD; Denise Aberle, MD; Avrum Spira, MD; Raúl San José Estépar, PhD; Marc E. Lenburg, PhD; and George R. Washko, MD; on behalf of the DECAMP and COPDGene Investigators

CHEST 2021; 159(2):549-563

 DOI: 10.1016/j.chest.2020.08.2115

e-Appendix 1.

Supplemental Methods

Participants and Study Design

The COPDGene Study (NCT00608764) cohort has been described in great detail previously[1]. Briefly, it is a multicenter longitudinal observational investigation of smokers focused on the epidemiologic and genetic factors associated with chronic obstructive pulmonary disease (COPD). At baseline, all participants underwent an inspiratory and expiratory CT scan, six-minute walk test (6MWT), spirometry, assessments of dyspnea and health status via the St. George Respiratory Questionnaire (SGRQ), review of medication use, and self-report of physician diagnosed conditions. The baseline enrollment of the 10,306 baseline COPDGene participants occurred between October 2006 and January 2011. All participants were invited to return for five and ten year interval follow-up, and are currently being invited back for ten year follow-up visits. They are also followed longitudinally through the longitudinal follow-up (LFU) program. As LFU program, participants are contacted by telephone every 6 months and asked questions regarding diagnoses, medications, acute respiratory exacerbations and hospitalizations. Mortality data was obtained from the LFU and from the social security death index. For those individuals whose mortality was determined based the LFU, vital status was back censored six months prior to dataset generation. Those whose follow up time terminated in death were included if their contact in the prior six months indicated that they were being actively followed at the time of death. For those participants with vital status ascertained using the SSDI, deaths and vital status were back censored three months to account for the expected lag time between a death and its appearance in the SSDI dataset[2–4].

Biospecimen Collection in DECAMP

All individuals in the DECAMP study underwent bronchoscopy. Bronchial airway epithelial cells were obtained from brushings of the right mainstem bronchus collected during fiberoptic bronchoscopy with an endoscopic cytobrush (Cellebrity Endoscopic Cytology Brush, Boston Scientific, Boston). The brushes were immediately placed in 1 mL of RNAprotect Cell Reagent (Qiagen, Valencia, CA) and kept at -80$_0$C until RNA isolation was performed.

Imaging acquisition in COPDGene and DECAMP

For COPDGene participants, volumetric CT scans of the chest were performed at both maximal inflation and relaxed exhalation. Images were acquired with the following CT protocol: for General Electric (GE) LightSpeed-16, GE VCT-64, Siemens Sensation-16 and -64, and Philips 40- and 60-slice scanners with 120kVp, 200mAs, and 0.5s rotation time. Images were reconstructed using a standard algorithm at 0.625mm slice thickness and 0.625mm intervals for GE scanners; using a B31f algorithm at 0.625 (Sensation-16) or 0.75mm slice thickness and 0.5mm intervals for Siemens scanners; and using a B algorithm at 0.9mm slice thickness and 0.45mm intervals for Philips scanners.

DECAMP-1 utilized CT scans collected as part of routine clinical care while DECAMP-2 utilized a standardized protocol for image acquisition and reconstruction. DECAMP-2 scans were collected using low dose helical computed tomography on a minimum 16-slice scanner. The scans were acquired at 2.5 to 5 mm but reconstructed into 1 mm slice thickness using the soft tissue and lung algorithms. Images from all sites were then de-identified and submitted to the American College of Radiology Imaging Network (ACRIN) Core Laboratory for storage.

Quantitative CT Analysis

The objective imaging measurements used for cluster definition in both cohorts were obtained using previously defined methods. The breadth of possible quantitative imaging measures that could be used to define clusters of individuals with cigarette smoking related lung diseases is beyond the scope of this study. Briefly, cigarette smoking has been shown to have effects on both pulmonary and extrapulmonary tissues measurable by quantitative CT imaging and related to clinical outcomes and disease pathophysiology[5]. In the lungs alone, these include both scarring (fibrosis) and destruction (emphysema) of the lung parenchyma, thickening of the airway

wall and destruction of the small airways, and loss of the peripheral pulmonary vasculature[6–9]. Outside of the lung these changes include those that occur in the coronary vasculature, as well as changes to body composition including the loss of fat free mass, and the loss of bone density[10–13]. Based on prior experience and expertise in this area, we selected a parsimonious list of imaging features to attempt to represent the breadth of both pulmonary and extra-pulmonary quantitative CT metrics of lung disease[5,14–16]. These included 1) the objective characterization of interstitial features as well as emphysema-like tissue using a local histogram-based technique, 2) the measurement the of pectoralis muscle area (expressed in cm^2) performed on a single axial image above the level of the aortic arch and 3) airway wall thickness as defined by the mean thickness of 6 segmental airways from each subject [6,10,17–24]. Each of these metrics has been shown to be related to both pathobiological changes that occur in certain, but not all, individuals in response to cigarette smoking exposure and to smoking related lung disease outcomes, as such, we felt they were likely to define clusters of patients who not only had different clinical outcomes, but who also may have specific and differing patterns of gene expression.

Cluster derivation and statistical analysis

Cluster analysis was performed using a parsimonious set of variables selected to represent the breadth of airway, lung parenchyma and extrapulmonary processes evident in smokers. The imaging features were log-transformed and standardized as needed to address distribution skewness and range. K-means clustering was then applied to these variables to group the subjects into clusters. The optimum number of clusters was determined using the Silhouette (using Euclidean distance) and Elbow methods (e-Figure 1)[25].

ANOVA, pairwise t-tests and chi-squared tests were used to analyze differences in baseline clinical variables between the clusters as appropriate. Differences in longitudinal changes in lung emphysema, lung function and exercise capacity were analyzed using mixed effects models and limited to those individuals whose CT imaging was performed on the same scanner type at both visits, while differences in mortality were analyzed using Kaplan Meier analysis and the log-rank test. Finally, differences in the rate of acute respiratory disease (ARD) events were analyzed using multivariable zero-inflated negative binomial regression with adjustments made for age, sex, race, smoking status, percent predicted forced expiratory volume in one second (FEV1%), St. George's Respiratory Questionnaire (SGRQ) score, gastroesophageal reflux and prior exacerbation, and with the inclusion of a time scale factor to account for varying durations of follow-up. ARD events were assessed prospectively and occur in smokers with and without COPD. They were defined as intermittent episodes of increased shortness of breath, cough and/or change in sputum quality requiring a change in treatment, including antibiotics and/or steroids[26]. Because these analyses are exploratory no correction of multiple comparisons was implemented. All statistical tests were two sided and a p-value threshold of 0.05 was utilized to declare statistical significance. Statistical analyses were performed using R (version 3.5.0).

RNA isolation, sequencing and data pre-processing

Total RNA was isolated using the miRNeasy Mini Kit (Qiagen, Valencia, CA). RNA integrity was assessed by Agilent BioAnalyzer, and RNA purity confirmed using a NanoDrop spectrophotometer. Libraries were generated using the Illumina TruSeq Stranded Total RNA kit and sequenced on the Illumina NextSeq 500 and Illumina HiSeq3000 with 75 base-pair paired-end reads (Illumina, San Diego, CA).

For data preprocessing, we developed an automatic pipeline using the Nextflow framework[27]. Quality of FASTQ files was assessed with FastQC. Reads were aligned to the human genome with 2-pass STAR[28] and gene-level and isoform-level expression quantified with RSEM[29]. Splice junction saturation, transcript integrity, and biotype distributions were calculated for each sample with RSeQC. DESeq2[30] or edgeR[31] was used to identify associations between gene expression profiles and clinical variables while controlling for confounding covariates. Genetic variants were called using the Broad Institute's GATK RNA-seq best-practices workflow[32]. Briefly, duplicates were marked with Picard tools, splitting of intronic reads, realignment around indels, and base quality score recalibration were performed with GATK, and variants were called with Haplotypecaller.

Gene Expression Analysis

The LIMMA package in R (version 3.4.0) was used to assess the differential bronchial epithelial gene expression (DGE) by cluster. To do this, raw count matrix of gene expression was initially filtered by counts per million (CPM) such that a gene could only be included if its CPM was greater than 1 in 10% of the total number of patients. DGE analysis was then performed using a pairwise comparison between the de novo imaging clusters at a false discovery rate (FDR)[33] of 0.25. The differentially expressed genes (DEG) identified by LIMMA were further analyzed by Enrichr[34] for over-representation analysis. Heatmaps were used to visualize the data and identify unsupervised gene clusters using the "Ward.D2" algorithm[35]. Gene set enrichment analysis (GSEA) was performed on pre-ranked gene lists created by pairwise comparisons between imaging clusters. Hallmark gene sets from the Molecular Signature Database (MSigDB) curated by the Broad Institute[34,36], as well as a gene set correlated to COPD severity previously identified by the lab were used for querying. A FDR at 0.01 was applied to select significant hallmark gene sets.

In Silico Validation

To further understand the function of these differentially expressed genes, we utilized the CREED tool[37] to search it's library of manually curated signatures from the Gene Expression Ominbus (GEO) for experimental perturbations that lead to a pattern of gene expression alterations similar to the emphysema signature. Gene expression signatures from five published datasets involving the response to interferon-β were identified as concordant (GSE26104[38], GSE19392[39], GSE3920[40], GSE125066[41] and GSE48400[42]). We summarized the expression of genes increased in emphysema or the genes decreased with emphysema in these interferon-β datasets using gene set variation analysis (GSVA)[43], a gene set enrichment method that estimates variation of pathway activity over a sample population. We found the GSVA scores from the signature of genes increased in the emphysema cluster is significantly increased in PBMC following interferon-β treatment (GSE26104; Figure 5). We found a similar increase in the GSVA scores from the emphysema-increased signature in datasets examining the response of hepatocytes, fibroblasts, endothelial cells, and bronchial epithelial cells to interferon-β[39–42] (e-Figure 3).

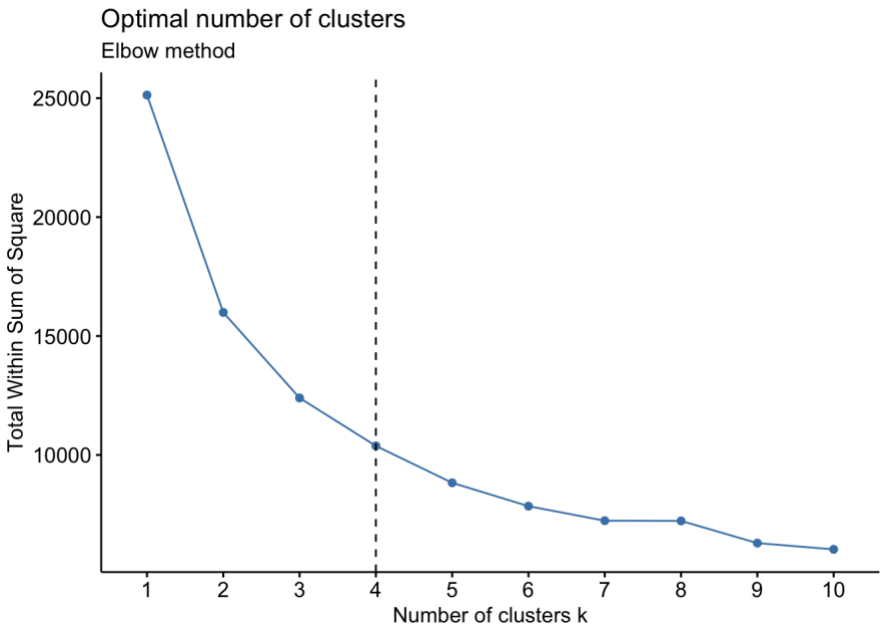
Supplemental References

1. Regan EA, Hokanson JE, Murphy JR, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD* 2010;7(1):32–43.
2. Standardization of Spirometry, 1994 Update. American Thoracic Society. *Am J Respir Crit Care Med* 1995;152(3):1107–1136.
3. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002;166(1):111–117.
4. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis* 1992;145(6):1321–1327.
5. Sanders KJC, Ash SY, Washko GR, Mottaghy FM, Schols AMWJ. Imaging approaches to understand disease complexity: chronic obstructive pulmonary disease as a clinical model. *J Appl Physiol* 2017;124(2):512–520.
6. Washko GR, Lynch DA, Matsuoka S, et al. Identification of early interstitial lung disease in smokers from the COPDGene Study. *Acad Radiol* 2010;17(1):48–53.
7. Estépar RSJ, Kinney GL, Black-Shinn JL, et al. Computed tomographic measures of pulmonary vascular morphology in smokers and their clinical implications. *Am J Respir Crit Care Med* 2013;188(2):231–239.

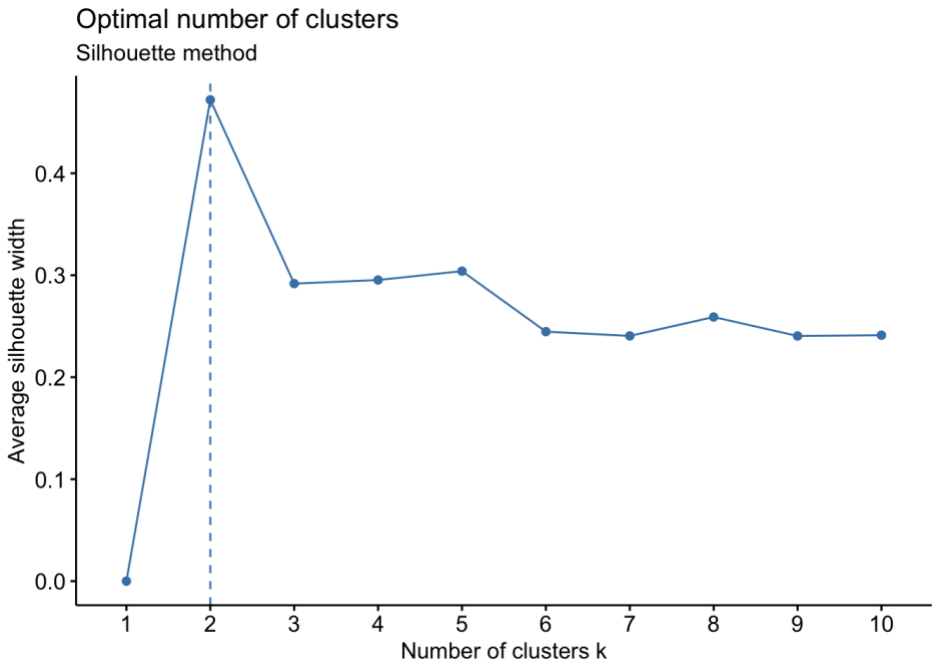
8. Bankier AA, De Maertelaer V, Keyzer C, Gevenois PA. Pulmonary emphysema: subjective visual grading versus objective quantification with macroscopic morphometry and thin-section CT densitometry. *Radiology* 1999;211(3):851–858.
9. Müller NL, Staples CA, Miller RR, Abboud RT. "Density mask". An objective method to quantitate emphysema using computed tomography. *Chest* 1988;94(4):782–787.
10. McDonald M-LN, Diaz AA, Ross JC, et al. Quantitative computed tomography measures of pectoralis muscle area and disease severity in chronic obstructive pulmonary disease. A cross-sectional study. *Ann Am Thorac Soc* 2014;11(3):326–334.
11. Budoff MJ, Hamirani YS, Gao YL, et al. Measurement of thoracic bone mineral density with quantitative CT. *Radiology* 2010;257(2):434–440.
12. Budoff MJ, Nasir K, Kinney GL, et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. *J Cardiovasc Comput Tomogr* 2011;5(2):113–118.
13. Oelsner EC, Lima JAC, Kawut SM, et al. Noninvasive tests for the diagnostic evaluation of dyspnea among outpatients: the Multi-Ethnic Study of Atherosclerosis lung study. *Am J Med* 2015;128(2):171-180.e5.
14. Albert RK, Connett J, Bailey WC, et al. Azithromycin for Prevention of Exacerbations of COPD. *N Engl J Med* 2011;365(8):689–698.
15. Anzueto A, Heijdra Y, Hurst JR. Controversies in COPD: ERS Monograph. European Respiratory Society; 2015.
16. Washko GR. Diagnostic Imaging in COPD. *Semin Respir Crit Care Med* 2010;31(3):276–285.
17. Diaz AA, Zhou L, Young TP, et al. Chest CT measures of muscle and adipose tissue in COPD: gender-based differences in content and in relationships with blood biomarkers. *Acad Radiol* 2014;21(10):1255–1261.
18. Kinsey CM, Estépar RSJ, Velden J van der, Cole BF, Christiani DC, Washko GR. Lower Pectoralis Muscle Area Is Associated with a Worse Overall Survival in Non–Small Cell Lung Cancer. *Cancer Epidemiol Prev Biomark* 2017;26(1):38–43.
19. Ash SY, Harmouche R, Putman RK, et al. Clinical and Genetic Associations of Objectively Identified Interstitial Changes in Smokers. *Chest* 2017;152(4):780–791.
20. Ash SY, Harmouche R, Ross JC, et al. The Objective Identification and Quantification of Interstitial Lung Abnormalities in Smokers. *Acad Radiol* 2017;24(8):941–946.
21. Ash SY, Harmouche R, Ross JC, et al. Interstitial Features at Chest CT Enhance the Deleterious Effects of Emphysema in the COPDGene Cohort. *Radiology* 2018;288(2):600–609.
22. Kim V, Desai P, Newell JD, et al. Airway wall thickness is increased in COPD patients with bronchodilator responsiveness. *Respir Res* 2014;15(1):84.
23. Putman RK, Hatabu H, Araki T, et al. Association Between Interstitial Lung Abnormalities and All-Cause Mortality. *JAMA* 2016;315(7):672–681.
24. Hunninghake GM, Hatabu H, Okajima Y, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013;368(23):2192–2200.
25. Kaufman L, Rousseeuw PJ. Finding groups in data. an introduction to cluster analysis [Internet]. 1990 [cited 2018 May 11]. Available from: http://adsabs.harvard.edu/abs/1990fgda.book.....K
26. Bowler RP, Kim V, Regan E, et al. Prediction of acute respiratory disease in current and former smokers with and without COPD. *Chest* 2014;146(4):941–950.
27. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;35(4):316–319.

28. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl* 2013;29(1):15–21.
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:323.
30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550.
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf Engl* 2010;26(1):139–140.
32. The GATK Best Practices for variant calling on RNAseq, in full detail [Internet]. GATK-Forum. [cited 2019 Feb 5];Available from: https://gatkforums.broadinstitute.org/gatk/discussion/3892/the-gatk-best-practices-for-variant-calling-on-rnaseq-in-full-detail
33. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* 1995;57(1):289–300.
34. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545–15550.
35. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* 2014;31(3):274–295.
36. Tan Y, Wu F, Tamayo P, Haining WN, Mesirov JP. Constellation Map: Downstream visualization and interpretation of gene set enrichment results. *F1000Research* 2015;4:167.
37. Wang Z, Monteiro CD, Jagodnik KM, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat Commun* [Internet] 2016 [cited 2019 Nov 18];7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052684/
38. Malhotra S, Bustamante MF, Pérez-Miralles F, et al. Search for specific biomarkers of IFNβ bioactivity in patients with multiple sclerosis. *PloS One* 2011;6(8):e23634.
39. Shapira SD, Gat-Viks I, Shum BOV, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 2009;139(7):1255–1267.
40. Indraccolo S, Pfeffer U, Minuzzo S, et al. Identification of genes selectively regulated by IFNs in endothelial cells. *J Immunol Baltim Md 1950* 2007;178(2):1122–1135.
41. Matta SK, Olias P, Huang Z, et al. Toxoplasma gondii effector TgIST blocks type I interferon signaling to promote infection. *Proc Natl Acad Sci U S A* 2019;116(35):17480–17491.
42. Bolen CR, Ding S, Robek MD, Kleinstein SH. Dynamic expression profiling of type I and type III interferon-stimulated hepatocytes reveals a stable hierarchy of gene expression. *Hepatol Baltim Md* 2014;59(4):1262–1272.
43. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
44. Steiling K, Berge M van den, Hijazi K, et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* 2013;187(9):933–942.

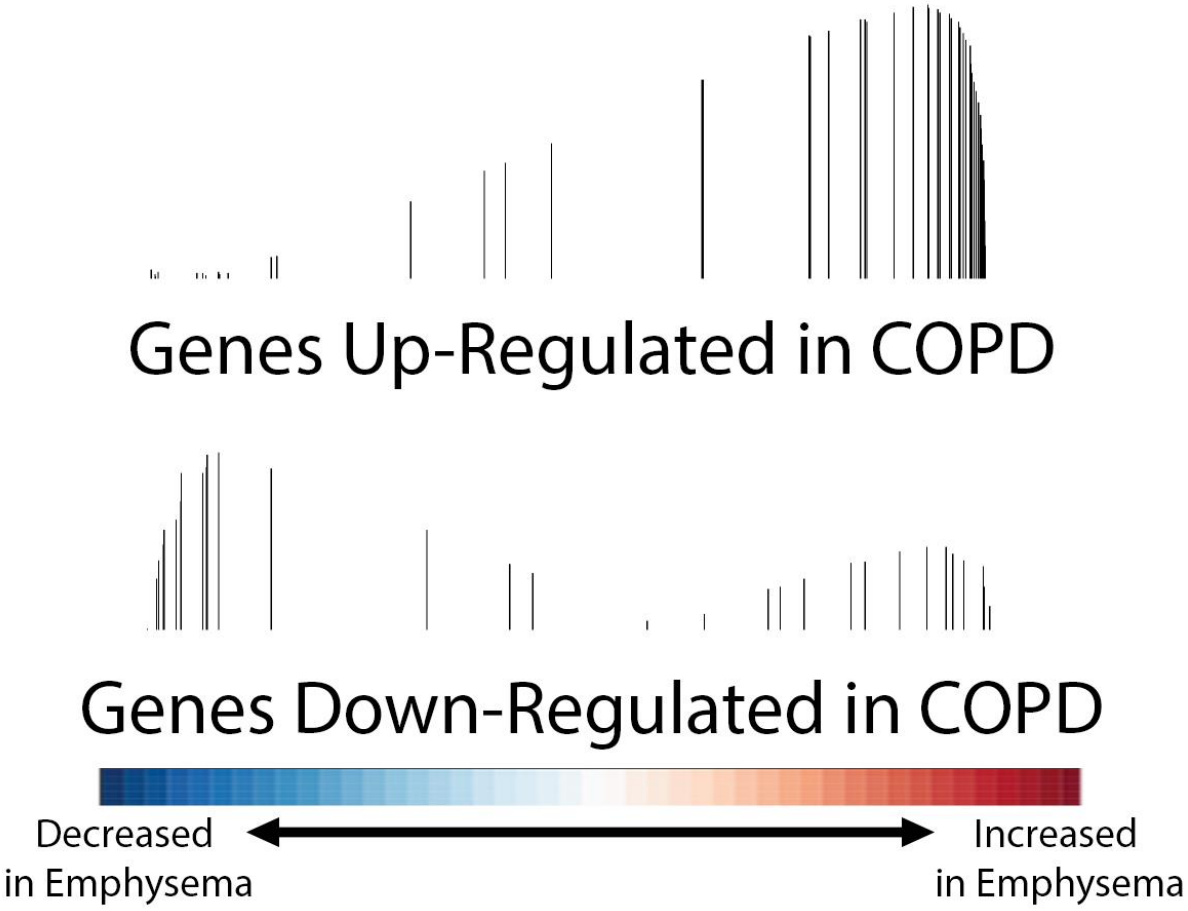
e-Figure 1A: *Elbow Method*



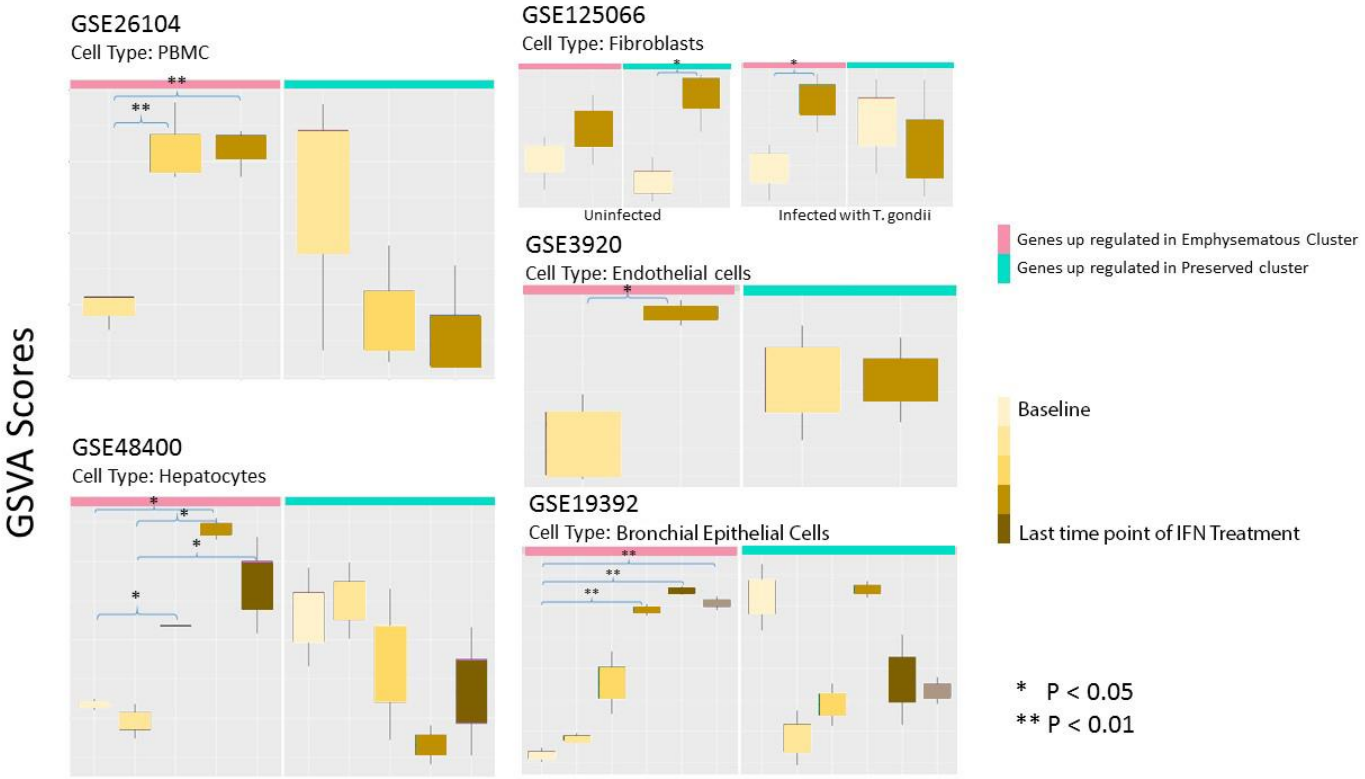
e-Figure 1B*: Silhouette Method*



e-Figure 1: Cluster derivation. Cluster analysis was performed using a parsimonious set of variables selected to represent the breadth of airway, lung parenchyma and extrapulmonary processes evident in smokers. The imaging features were log-transformed and standardized as needed to address distribution skewness and range. K-means clustering was then applied to these variables to group the subjects into clusters. The optimum number of clusters was determined using the Silhouette (using Euclidean distance) and Elbow methods.



e-Figure 2. Genes that are changed in subjects with COPD are enriched among the genes changed in the emphysema cluster relative to the preserved cluster. The distribution of genes identified by Steiling et al. [44] as being either up-regulated (top) or down-regulated (bottom) in subjects with COPD were examined in a list of all genes ranked by their expression difference between the emphysema cluster and the preserved cluster by Gene Set Enrichment Analysis (GSEA). Genes that are up-regulated in COPD are enriched among the genes most increased in individuals from the emphysema cluster ($q < 0.001$). Genes that are down-regulated in COPD are enriched among the genes most decreased in individuals from the emphysema cluster ($q < 0.001$).



e-Figure 3: Gene set variation analysis (GSVA) of emphysema signature gene clusters in various cell types following interferon-β treatment. GSVA was used to summarize the expression of each emphysema signature gene cluster in a number of previously published datasets involving interferon-β treatment: PBMCs (GSE26104), hepatocytes (GSE48400), fibroblasts (GSE125066), endothelial cells (GSE3920), and bronchial epithelial cells (GSE19392). Post-hoc Tukey's HSD was applied to examine the pairwise differences in GSVA scores between groups. Symbols for pairwise comparisons: * = P ≤ 0.05; ** = P ≤ 0.01.

e-Table 1

	Genes	Associated Functions
Genes up-regulated in Preserved Group	TRAC, THEMIS	T-cell receptor signaling pathway
	SIT1, TRAC	Regulation of T-cell activation via T-cell receptor contact with antigen bound to MHC molecule on antigen presenting cells
Genes up-regulated in Emphysema Group	CEBPB, TNFA1P6, CCL20, IL1B, CXCL3	Inflammatory response
	CEBPB, CCL20, IL1B, CXCL3, PTGS2	TNF-signaling pathway
	CCL20, IL1R2, IL1B, OSM, CXCL3	Cytokine-cytokine receptor interaction

e-Table 1: Functional gene expression analysis using Enrichr. Using linear modeling, 41 genes were found to be differentially expressed between the preserved and emphysema cluster (FDR < 0.25). We used Enrichr to identify over-represented functional categories. MHC = major histocompatibility complex; TNF = tumor necrosis factor.

e-Table 2

	Preserved vs Interstitial Predominant	Preserved vs Emphysema Predominant	Interstitial Predominant vs Emphysema Predominant
Up Regulated Pathways	**INTERFERON GAMMA RESPONSE**	**PANCREAS BETA CELLS**	PANCREAS BETA CELLS
	INTERFERON ALPHA RESPONSE	**OXIDATIVE PHOSPHORYLATION**	HEME METABOLISM
	ALLOGRAFT REJECTION	**INTERFERON ALPHA RESPONSE**	MYOGENESIS
	OXIDATIVE PHOSPHORYLATION	**MYC TARGETS V1**	KRAS SIGNALING DN
	PROTEIN SECRETION	HEDGEHOG SIGNALING	UV RESPONSE DN
	FATTY ACID METABOLISM	MYOGENESIS	ESTROGEN RESPONSE EARLY
	PANCREAS BETA CELLS	FATTY ACID METABOLISM	HEDGEHOG SIGNALING
	COMPLEMENT	PEROXISOME	WNT BETA CATENIN SIGNALING
	MYC TARGETS V1	DNA REPAIR	NOTCH SIGNALING
	ADIPOGENESIS	PROTEIN SECRETION	MYC TARGETS V1
	REACTIVE OXYGEN SPECIES PATHWAY		
Down Regulated Pathways	**TNFA SIGNALING VIA NFKB**	**TNFA SIGNALING VIA NFKB**	**INFLAMMATORY RESPONSE**
	EPITHELIAL MESENCHYMAL TRANSITION	**INFLAMMATORY RESPONSE**	**TNFA SIGNALING VIA NFKB**
	HYPOXIA	**KRAS SIGNALING UP**	**INTERFERON GAMMA RESPONSE**
	WNT BETA CATENIN SIGNALING	**IL6 JAK STAT3 SIGNALING**	**KRAS SIGNALING UP**
	TGF BETA SIGNALING	**COMPLEMENT**	**ALLOGRAFT REJECTION**
	ESTROGEN RESPONSE LATE	**COAGULATION**	**IL6 JAK STAT3 SIGNALING**
	ESTROGEN RESPONSE EARLY	**TGF BETA SIGNALING**	**COMPLEMENT**
	HEME METABOLISM	**HYPOXIA**	**INTERFERON ALPHA RESPONSE**
	KRAS SIGNALING DN	**IL2 STAT5 SIGNALING**	**MTORC1 SIGNALING**
	P53 PATHWAY	**UV RESPONSE UP**	**IL2 STAT5 SIGNALING**

e-Table 2: Gene Set Enrichment Analysis used to identify pathway-related genesets. To better characterize the biology of the differentially expressed genes, GSEA was performed on pre-ranked gene lists created by the comparison of the emphysema cluster and the preserved cluster to identify enrichment of pathway-related genesets from the KEGG, Reactome, and Gene Ontology databases. Gene sets with significant enrichment (GSEA $q < 0.05$) are bold.

e-Table 3: IRB committee names and project approval numbers for each center.

Site Name	ACRIN Site #	CTEP Inst #	FWA	FWA Exp Date	DECAMP 1 IRB Approval	DECAMP 1 HRPO Log #	DECAMP 2 IRB Approval	DECAMP 2 HRPO Log #
University of Pennsylvania	4202	PA141	FWA00004028	2/2/21	Protocol #: 816341	A-17242.1o	Protocol #: 818746 Review Board: IRB #2	A-17242.2o
Brooke Army Medical Center	4238	TX055	FWA00004092	12/5/21	Project #: 376127 Reference #: C.2012.135	A-17242.1m	Project #: 385999 Reference #: C.2013.107	A-17242.2m
Roswell Park Memorial Institute	4278	NY158	FWA00006731	5/4/23	IRB ID: I 217812	A-17242.1g	IRB ID: I 251914	A-17242.2g
VA Greater Los Angeles Health Care System	4438	CA221	FWA00000734	11/1/22	VA Project #: 0051	A-17242.1h	VA Project #: 0052	A-17242.2h
UCLA	4494	CA006	FWA00004642	6/22/23	IRB#12-000926	A-17242.1p	N/A	N/A
Philadelphia VA	4714	PA082	FWA00001311	4/13/22	ID: 01405 | Prom #: 0002	A-17242.1e	ID: 01428 | Prom #: 0003	A-17242.2e
VA Boston Healthcare System	4790	MA139	FWA00001270	7/6/23	IRB# 2661	A-17242.1c	IRB# 2802	A-17242.2c
VA North Texas Health Care System	4791	TX002	FWA00001338	1/2/23	IRB# 12-035	A-17242.1d	IRB# 13-050	A-17242.2d
VA Eastern Colorado Health Care System	4792	CO015	FWA00005070	10/26/22	COMIRB Protocol 12-0707	A-17242.1f	COMIRB Protocol 12-1662	A-17242.2f
Nashville VA Medical Center	4793	TN004	FWA00003772	12/1/20	Study ID#: 331929	A-17242.1j	Study ID#: 470111	A-17242.2j
VA Pittsburgh Healthcare System	4794	PA022	FWA00001282	2/9/21	Pro00000495	A-17242.1i	Pro00000576	A-17242.2i
Walter Reed National Military Medical Center	4795	MD001	FWA00017749	9/9/21	IRBnet #: 376221	A-17242.1n	IRBnet #: 387954	A-17242.2n
Naval Medical Center San Diego	4796	CA074	FWA00002342	9/22/21	Protocol CIP # NMCSD.2012.0025	A-17242.1l	Protocol CIP # NMCSD.2013.0078	A-17242.2l
Naval Medical Center Portsmouth	4797	VA024	FWA00006001	11/23/21	Protocol CIP # NMCP.2013.0006	A-17242.1k	Protocol CIP # NMCP.2013.0046	A-17242.2k
Boston Medical Center	4798	MA136	FWA00000301	11/12/20	IRB Number: H-31755	A-17242.1a	IRB Number: H-32479	A-17242.1a

DECAMP Investigator Directory

Executive Center – Boston University
Avi Spira, MD *†
Ehab Billatos, MD *†
Elizabeth Moses, PhD *
Marc Lenburg, PhD †
Jennifer Beane, PhD
Josh Campbell, PhD
Jack Cunningham
Gang Liu, PhD
Hanqiao Liu
Sherry Zhang
Jiarui Zhang
Xu Ke
Xingyi Shi
Carter Merenstein, MS
Yue Zhao

Coordinating Center – American College of Radiology
Denise Aberle, MD *
Mitchell Schnall, MD, PhD *
Charles Apgar, MBA *
Irene Mahon, RN, MPH *
Lindsey Dymond
Joe Bauza
Sarah Gevo

Data Coordinating Center – Brown University
Constantine Gastonis, PhD *
Fenghai Duan, PhD *†
Helga Marquez
David Elashoff, PhD †

Pathology Core – MD Anderson Cancer Center
Ignacio Wistuba, MD
Humam Kadara, PhD
Junya Fujimoto, MD, PhD

Sequencing Core - Uniformed Services University of the Health Science
Clifton Dalgard, PhD
Matthew Wilkerson, PhD

Imaging Core
Denise Aberle, MD
George Washko, MD *
Charles M Kinsey, MD, MPH
Fenghai Duan, PhD

Clinical Centers
Denise Fine, *Boston University*
Ron Goldstein, MD, *Boston VA*
Kathleen LaCerda, *Boston VA*
John Battaile, MD, *Dallas VA*
Mitchell Kroll, *Dallas VA*
Bob Keith, MD, *Denver VA*
Mary Jackson, *Denver VA*
Steve Dubinett, MD, *UCLA*
Gina Lee, MD, *UCLA*
Babak Aryanfar, *UCLA*
Rafael Corona, *UCLA*
Anil Vachani, MD, *University of Pennsylvania* ŦT
Sam Soloman, *University of Pennsylvania*
Charles Atwood, MD, *Pittsburgh VA*
Gregory Owens, *Pittsburgh VA*
Hanna Edvardsson *Pittsburgh VA*
Pierre Massion, MD, *Vanderbilt University*
Trey Helton, *Nasvhille VA*
Mary Reid, *Roswell Park Cancer Institute*
Chris Kuzniewski, MD, *Naval Medical Center Portsmouth*
Jacob Carmichael, MD, *Naval Medical Center Portsmouth*
Holly LaPerriere, *Naval Medical Center Portsmouth*
J Scott Parrish, MD, *Naval Medical Center San Diego*
Lindsey White, *Naval Medical Center San Diego*
Anna Kaur, *Naval Medical Center San Diego*
Robert Browning Jr, MD, *Walter Reed Army Medical Center* *Ŧ
Maggie Nelissery, *Walter Reed Army Medical Center*
Folashade Akanni, *Walter Reed Army Medical Center*

**Denotes DECAMP Executive Committee Member*
Ŧ *Denotes DECAMP Data Access Committee Member*

CHEST™ Online Supplement

Investigator Directory

Administrative Center
James D. Crapo, MD
Edwin K. Silverman, MD, PhD
Gisselle Gonzalez
Kelley Madden
Barry J. Make, MD
Emily Port
Elizabeth A. Regan, MD, PhD
Joshua Ryon
Lori Stepp
Shandi Watts, MS
Michael Weaver

Executive Committee
Terri Beaty, PhD
Russell P. Bowler, MD, PhD
James D. Crapo, MD
Jeffrey L. Curtis, MD
Douglas Everett, PhD
MeiLan K. Han, MD, MS
John E. Hokanson, MPH, PhD
David A. Lynch, MB
Barry J. Make, MD
Elizabeth A. Regan, MD, PhD
Edwin K. Silverman, MD, PhD

External Advisory Committee
Gary Anderson, PhD
Eugene R. Bleecker, MD
Harvey O. Coxson, PhD
Ronald G. Crystal, MD
James C. Hogg, MD
Michael A. Province, PhD
Stephen I. Rennard, MD

NHLBI
Thomas Croxton, MD, PhD
Weiniu Gan, PhD
Lisa A. Postow, PhD
Lisa M. Viviano, BSN

COPD Foundation
Corinne Costa Davis
Elisha Malanga
Delia Prieto

Biorepository Visit 1 (Baltimore)
Homayoon Farzadegan, PhD
Akila Hadji

Biorepository Visit 2 (Boston)
Leanna Farnam

Data Coordinating Center
Director: Douglas Everett, PhD
Grace Chen
James Crooks, PhD
Ruthie Knowles, MSW, CCRP
Katherine Pratte, PhD
Matthew J. Strand, PhD
Carla Wilson, MS
Pearlanne T. Zelarney, MS

National Jewish Analysis Group
James D. Crapo, MD
Russell P. Bowler, MD, PhD
Katerina J. Kechris, PhD
Sonia Leach, PhD
Elizabeth A. Regan, MD, PhD
Ryan J. Webster, MD

Epidemiology Center
Director: John E. Hokanson, MPH, PhD
Erin E. Austin, PhD
Gregory Kinney, MPH, PhD
Sharon M. Lutz, MPH, PhD
Margaret F. Ragland, MD, MS
Kendra A. Young, MSPH, PhD

Sequencing and Bioinformatics Center
Director: Michael Cho, MD
Peter J. Castaldi, MD, MSc
Kimberly Glass, PhD
Craig Hersh, MD, MPH
Wonji Kim, PhD
Yang-Yu Liu, PhD
Edwin K. Silverman, MD, PhD

Genotyping Centers

Genome-Wide Center
Director: Terri Beaty, PhD

Candidate Genotyping Center
Craig P. Hersh, MD, MPH
Edwin K. Silverman, MD, PhD

Genetic Analysis Center
Director: Edwin K. Silverman, MD, PhD
Terri Beaty, PhD
Christopher J. Benway, PhD
Jacqueline Bidinger, MS
Peter J. Castaldi, MD, MSc
Michael H. Cho, MD
Douglas Conrad, MD
Dawn L. DeMeo, MD, MPH
Adel R. El-Boueiz, MD
Marilyn G. Foreman, MD, MS
Nadia N. Hansel, MD, MPH
Lystra P. Hayden, MD
Craig Hersh, MD, MPH
Brian Hobbs, MD
Wonji Kim, PhD
Woori Kim, MPH
Christoph Lange, PhD
Sharon M. Lutz, MPH, PhD
Merry-Lynn McDonald, PhD
Matthew Moll, MD
Melody Morris, PhD
Nikolaos A. Patsopoulos, MD, PhD
Dandi Qiao, PhD
Elizabeth A. Regan, MD, PhD
Emily S. Wan, MD

Imaging Center
Director: David A. Lynch, MB
Harvey O. Coxson, PhD
Jennifer G. Dy, PhD
Sean B. Fain, PhD
Shoshana Ginsburg, MS
Eric A. Hoffman, PhD
Stephen Humphries, PhD
Philip F. Judy, PhD
Alex Kluiber
Stefanie Mason, MD
Joseph M. Reinhardt, PhD
James Ross, PhD
Raul San Jose Estepar, PhD
Joyce D. Schroeder, MD
Arkadiusz Sitek, Ph.D.
Robert M. Steiner, MD
Edwin van Beek, MD, PhD, MEd
Bram van Ginneken, PhD
Eva van Rikxoort, PhD
George R. Washko, MD