Clustering-based COPD Subtypes Have Distinct Longitudinal Outcomes and Multi-omics Biomarkers

3 Andrew Gregory¹, Zhonghui Xu¹, Katherine Pratte², Sool Lee¹, Congjian Liu³, Robert Chase¹,

4 Jeong H. Yun^{1,3}, Aabida Saferali¹, Craig P. Hersh^{1,3}, Russell P. Bowler⁴, Edwin K. Silverman^{1,3},

5 Peter J. Castaldi^{1,5}*, Adel Boueiz^{1,3}*.

6 *Contributed equally

¹Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical
 School, Boston, MA; ²Department of Biostatistics, National Jewish Health, Denver, CO;
 ³Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical
 School, Boston, MA; ⁴Division of Pulmonary, Critical Care and Sleep Medicine, National Jewish
 Health, Denver, Colorado; ⁵General Medicine and Primary Care, Brigham and Women's
 Hospital, Harvard Medical School, Boston, MA

13 Abstract

Introduction: Chronic obstructive pulmonary disease (COPD) can progress across several 14 15 domains, complicating the identification of the determinants of disease progression. In our 16 previous work, we applied k-means clustering to spirometric and chest radiologic measures to identify four COPD-related subtypes: "Relatively resistant smokers (RRS)", "mild upper lobe 17 predominant emphysema (ULE)", "airway-predominant disease (AD)", and "severe emphysema 18 19 (SE)". In the current study, we examined longitudinal spirometric and radiologic emphysema changes and prospective risks of COPD exacerbations, incident comorbidities, and mortality of 20 21 these clusters. We also compared their associations to protein and transcriptomic biomarkers.

<u>Methods</u>: We included 8,266 non-Hispanic white and African-American smokers from the COPDGene study. We used linear regression to investigate associations to five-year prospective changes in spirometric and radiologic measures and to plasma protein and blood gene expression levels. We used Cox-proportional hazard modeling to test for associations to prospective exacerbations, comorbidities, and mortality.

6 <u>**Results:**</u> The RRS, ULE, AD, and SE clusters represented 39%, 15%, 26%, and 20% of the 7 studied cohort at baseline, respectively. The SE cluster had the greatest 5-year FEV₁ and 8 emphysema progression, and the highest risks of exacerbations, cardiovascular disease (CVD), 9 and mortality. The AD cluster had the highest diabetes risk. After adjustments, only the ULE and 10 AD clusters had elevated CVD mortality risks, while only the ULE cluster had the highest 11 cancer-related mortality risk. These clusters also demonstrated differential protein and gene 12 expression biomarker associations.

<u>Conclusion</u>: COPD k-means subtypes demonstrate varying rates of disease progression,
 prospective comorbidities, mortality, and associations to proteomic and transcriptomic
 biomarkers.

16 **Funding Sources:**

This work was supported by NHLBI K08 HL141601, K08 HL146972, R01 HL116931, R01
HL124233, R01 HL126596, R01 HL116473, U01 HL089897, R01 HL147326, R01 HL130512,
and U01 HL089856. The COPDGene study (NCT00608764) is also supported by the COPD
Foundation through contributions made to an Industry Advisory Committee comprised of
AstraZeneca, Bayer, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer,
Siemens, and Sunovion.

2

1 Introduction

2 Chronic obstructive pulmonary disease (COPD) is a heterogeneous disorder, with a wide 3 variety of clinical manifestations ¹. Additionally, disease progression in COPD occurs across 4 multiple domains, such as lung function decline, worsening of emphysema, and development of 5 comorbidities ²⁻⁴. These challenges complicate COPD subtyping and the identification of the 6 determinants of COPD progression.

7 Several reports have shown that factors, such as COPD exacerbation history, reduced pulmonary function, and a low BMI, are associated with an elevated risk of respiratory 8 exacerbations, accelerated spirometric decline, and emphysema changes ^{2 5 6}. Additionally, 9 10 studies which have subtyped subjects with COPD based on spirometry, respiratory symptoms and other characteristics have revealed that these subgroups differ in their risks for 11 exacerbations, hospital admissions, and FEV_1 and emphysema changes ⁷⁻¹¹. However, many of 12 these studies have modest sample size and limited longitudinal follow-up. Furthermore, while 13 14 comorbid conditions such as cardiovascular disease (CVD) and type 2 diabetes mellitus are 15 common in subjects with COPD, the specific COPD-related characteristics associated with the risk of developing these comorbidities have not been fully described ^{12 13}. Moreover, different 16 17 protein biomarkers have been identified in subjects with COPD, but few studies have assessed subtype-specific proteomic signatures ¹⁴. Similarly, while transcriptomic signatures have been 18 identified in COPD, the differences in such patterns between COPD subtypes have not been 19 thoroughly investigated ¹⁵⁻¹⁷. 20

In our previous work, we applied k-means clustering to spirometric and chest radiologic measures and identified four COPD-related subtypes: "Relatively resistant smokers (i.e. no/mild airflow obstruction and minimal emphysema despite heavy smoking) (RRS)", "mild upper lobe

3

predominant emphysema (ULE)", "airway-predominant disease (AD)", and "severe emphysema 1 (SE)", which had differing cross-sectional profiles and genetic associations ¹⁸. In the current 2 study, we included up to 12.7 years of prospective data to investigate subtype-specific rates of 3 4 progression in spirometric measures and radiologic emphysema and to quantify the risks of prospective COPD exacerbations, CVD events, diabetes, and mortality. We also investigated 5 cross-sectional associations between subtypes and plasma protein and blood transcriptomic 6 7 biomarkers. We hypothesized that these subtypes would have different disease progression profiles and associations to biomarkers. Some of these results have been previously reported as 8 an abstract ¹⁹. 9

10 Methods

11 Study description

The COPDGene study is a prospective, multicenter, longitudinal study investigating the 12 13 genetic and epidemiological characteristics of COPD across 21 centers in the U.S. (NCT00608764, www.copdgene.org)²⁰. Institutional review board approval was obtained at 14 15 each study center. Patients or the public were not involved in the design, conduct, reporting, or 16 dissemination plans of our research. All data produced in the present study are available upon 17 reasonable request to the authors. This study enrolled non-Hispanic whites and African-18 Americans, who were 45-80 years old and had at least 10 pack-years of lifetime smoking history. 19 Subjects were recruited across the full spectrum of disease severity as defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) spirometric grading system²⁰. 20 COPDGene conducted two study visits (Visit 1 and Visit 2) approximately 5 years apart. 21 22 Subjects were also contacted every 3-6 months through the COPDGene Longitudinal Follow-up

1 program via phone or online surveys to collect data on incident COPD-related events, 2 comorbidities, and mortality. COPD-related events and comorbidities included self-reported COPD exacerbations (defined as the acute worsening of respiratory symptoms that required 3 systemic steroids and/or antibiotics²¹), CVD events (defined as a composite endpoint of stroke, 4 heart attack, coronary artery disease, coronary artery bypass graft surgery, peripheral artery 5 disease, and/or cardiac angina), and type 2 diabetes mellitus. All-cause mortality was determined 6 7 through a combination of longitudinal follow-up and a search of the social security death index. Cause-specific mortality was categorized as respiratory-related, CVD-related, cancer-related 8 9 (any type) or due to other causes and determined through systematic adjudication process based on the methods used in the Towards a Revolution in COPD Health (TORCH) trial²². 10

Demographics, spirometry, imaging, smoking burden, respiratory symptoms, and 11 12 comorbidities were collected at Visits 1 and 2. In addition to GOLD grades 0-4, we included 13 subjects with Preserved Ratio Impaired Spirometry (PRISm), defined as $FEV_1/FVC \ge 0.70$ but with $FEV_1 < 80\%$ predicted ²¹. Thirona software (www.thirona.eu) was used to quantify 14 15 emphysema as the percentage of lung voxels with an attenuation of < -950 HU at maximal inspiration (%LAA-950)²³. The Hounsfield units at the 15th percentile of the computed 16 tomography (CT) density histogram at end-inspiration corrected for the depth of inspiratory 17 variation (adjusted Perc15 density) were used for longitudinal changes in emphysema ^{24 25}. Per 18 convention, adjusted Perc15 density values are reported as HU + 1000. The levels of 1.305 19 20 protein biomarkers (SOMAscan Human Plasma 1.3K assay) were obtained from plasma samples collected at Visit 1²⁶. Total blood RNA was collected at Visit 2. 21

22 Cluster Generation

We used the k-means clusters that were generated in our previously published work ¹⁸. FEV₁ percent predicted, CT-quantified emphysema, percent airway wall thickness, and apicobasal emphysema distribution (log of the lung upper third to lower third ratio of emphysema) were the input features that were used for clustering at Visit 1. Using the same approach, we also performed k-means clustering at Visit 2 to assess cluster assignment stability between the two visits and to conduct differential gene expression analyses using RNA-Seq data available at Visit 2.

8 Statistical analyses

9 Data distributions were reported as medians with interquartile ranges or counts with percentages, where appropriate. We calculated FEV_1 and emphysema changes as either absolute 10 11 or relative annualized changes. We computed absolute annualized changes by subtracting Visit 1 from Visit 2 values and dividing the difference by the time in years between visits for each 12 subject. Relative annualized changes were obtained by dividing the absolute annualized changes 13 14 by Visit 1 values and multiplying by 100. Negative values indicate worsening of the disease 15 between visits. We used the Kruskal-Wallis and chi-square tests for continuous and categorical variables, respectively. We subsequently performed post-hoc pairwise comparisons between the 16 17 clusters using the Nemenyi and chi-square tests for continuous and categorical variables, respectively. Additionally, we constructed univariable and multivariable linear regression models 18 to relate changes in FEV_1 and emphysema as well as plasma proteins to cluster assignment. We 19 20 assessed risks of incident COPD exacerbations, CVD events, diabetes, and mortality using Cox proportional hazards models and obtained survival curves using the Kaplan-Meier method. For 21 22 the analyses of incident CVD events and diabetes, we excluded subjects who had a history of CVD or diabetes at Visit 1. We used the RRS cluster as the reference group. Linear regression 23

1 and Cox models were adjusted for relevant baseline physiological, clinical, and demographic 2 characteristics. We additionally adjusted for metabolic syndrome in the CVD multivariable Cox models and for body mass index, airflow obstruction and exercise capacity (BODE index) in the 3 4 mortality models. For the protein analyses, we adjusted for age, sex, race and current smoking status. To ensure that cluster associations to emphysema changes are not confounded by CT 5 scanner type, a sensitivity analysis was performed by adding scanner type as a covariate in a 6 7 subgroup analysis limited to subjects who underwent scans with the same scanner type between visits. 8

9 To test for differential gene expression between clusters, we used the linear modeling 10 approach implemented in the limma R package (v3.38.3) ²⁷ adjusting for age, sex, current 11 smoking, white blood cell count proportions, and library prep batch. Gene ontology (GO) 12 functional enrichment of the gene sets was calculated using the weighted Fisher test in the 13 topGO Bioconductor package that accounts for the dependency between terms in the GO 14 topology ²⁸. We reported only the GO pathways with at least 3 significant genes.

All tests for the clinical outcomes were two-tailed with a significance threshold of Pvalue < 0.05. For the protein and RNA-Seq analyses, we corrected for multiple comparisons using the Benjamini-Hochberg method and applied a threshold of significance of false discovery rate (FDR) of 10% ²⁹. Significantly enriched GO pathways were identified using the weighted Fisher P-value < 0.005.

20

Additional methods are available in the Supplement.

21 **<u>Results</u>**

The overview of the study is shown in Figure 1 and the study flow diagram is outlined in
Figure S1. Subjects who were not included in the analyses of longitudinal changes in FEV₁ and
emphysema, 935 of whom died between the first and second study visit, had a higher proportion
of GOLD spirometric grade 4 disease (Table S1).

5

Cluster characteristics at baseline and stability of cluster assignments between visits

6 The RRS cluster represented 39% of the studied population at the baseline visit and was characterized by a history of heavy smoking without significant airflow obstruction, emphysema, 7 or airway wall thickness compared to the other clusters (*P*-values < 0.05) (Table S2). 8 9 Additionally, subjects in the RRS cluster had predominantly GOLD 0-1 spirometry. At Visit 2, 10 RRS cluster membership was stable as 76.3% of subjects were still assigned to this cluster (Figure S2). The ULE cluster consisted of 15% of all subjects and had moderate airflow 11 obstruction and mild emphysema predominantly in the upper lung lobes. The ULE cluster was 12 relatively unstable, with only 38% of subjects remaining within this cluster at their follow-up 13 14 visits. The AD cluster, 26% of all participants, was characterized by high BMI and had the highest proportion of subjects with PRISm (*P-values* < 0.05). The AD cluster was more stable 15 than the ULE cluster with 53% of subjects staying within this cluster at their follow-up visits. 16 17 The SE cluster, 20% of the studied cohort, exhibited high emphysema and gas trapping and had the highest proportion of GOLD 3-4 smokers (*P*-values < 0.05). The SE cluster was very stable 18 as 92.2% of subjects remained within the SE cluster at Visit 2. For subsequent analyses of 19 20 longitudinal outcomes and protein data by k-means clusters, subjects were analyzed according to their cluster assignment at Visit 1. 21

22 Cluster-specific rates of spirometric and emphysema progression

1 The 5-year change values for FEV_1 (measured as absolute change in FEV_1 and percent 2 change relative to baseline) and emphysema are shown in Figure 2, and the results from the 3 univariable and multivariable models are shown in Table 1. The RRS cluster, which has the least impaired spirometry and emphysema at baseline, had the greatest loss in absolute FEV_1 . The AD 4 cluster had the lowest absolute loss in FEV₁, significantly less than the RRS cluster in both 5 6 univariable and multivariable models. Both the ULE and SE clusters had similar absolute FEV_1 changes relative to the RRS cluster. However, when adjusted for relevant covariates, the SE 7 cluster had significantly less absolute FEV₁ decline than the RRS cluster. While absolute FEV₁ 8 9 changes were notably higher in the RRS cluster, percent changes in FEV_1 relative to baseline 10 were most pronounced in the SE cluster. Pairwise comparisons between all clusters showed that both emphysema-related clusters (ULE and SE) had significantly larger relative changes in 11 percent FEV₁ relative to baseline than both the RRS and AD clusters (*P-values* < 0.05, Table 12 S3). 13

Compared to FEV₁ changes, the pattern of CT-quantified emphysema progression was 14 15 less sensitive to the metric being used (absolute vs. relative). The SE cluster had the most rapid 16 while the RRS cluster had the least rapid relative progression in both univariable and 17 multivariable models. The ULE and AD clusters had significantly greater absolute and relative 18 emphysema changes than the RRS cluster. The ULE cluster had significantly greater relative 19 emphysema progression than the AD cluster (*Pairwise P-value: 0.03*, Table S3). We noted 20 similar cluster associations to both absolute and relative emphysema changes when we added 21 scanner type as a covariate in the subgroup analysis limited to subjects who underwent scans 22 with the same scanner type between visits (Table S4).

23 Cluster-specific risk of incident co-morbidities and COPD-related events

1 Starting from Visit 1, the median follow-up time was 9.2 years for prospective CVD and 2 diabetes, 5.3 years for respiratory exacerbations, 9.5 years for all-cause mortality, and and 7.8 years for cause-specific mortality. When we analyzed prospective risks of various health 3 4 outcomes by subtype, we observed that the SE cluster had the highest risk for prospective COPD exacerbations and incident CVD (Figure 3). After multivariable adjustment, the SE cluster had a 5 3 times higher likelihood of having a COPD exacerbation when compared to the RRS cluster 6 7 (*HR* 2.98 (SE: 0.05), P-value < 0.001) (Table S5). The AD cluster had the highest risk of incident diabetes (*HR 1.97 (SE: 0.09*), *P-value* < 0.001) and this association remained significant 8

risks were higher in the ULE cluster compared to the RRS cluster (*P*-values < 0.05).

after correcting for age, sex, race, BMI, and smoking pack-years. COPD exacerbation and CVD

9

Survival curves by subtype are shown for all-cause mortality (Figure 3) and for cause-11 12 specific mortality (Figure S3). Results of the survival models are presented in Table S6. In the 13 univariable models, subjects in the SE cluster had the highest risks of all-cause, respiratoryrelated, CVD-related, cancer-related and other causes-related mortality relative to the RRS 14 15 cluster (6, 50, 3, 3, and 2 times higher, respectively, *P-values* < 0.0001). The risks for all-cause, 16 respiratory, and CVD mortality were also elevated for the ULE and AD clusters relative to RRS 17 (P-values < 0.05). In multivariable models adjusting for age, sex, race, smoking pack-years, and BODE, statistical significance was maintained for the associations of the ULE, AD and SE 18 clusters with all-cause mortality, the association of the SE cluster with respiratory mortality, the 19 20 associations of the ULE and AD clusters with CVD mortality, and the association of the ULE cluster with cancer mortality (*P*-values < 0.05). 21

22 Clusters associations to protein and gene expression biomarkers

1 Using SOMAscan plasma protein measurements at baseline from 1.047 subjects, we 2 tested for differential protein associations between clusters. Using the RRS group as the 3 reference, we identified significant associations in adjusted models to 16, 65, and 219 proteins 4 for the ULE, AD, and SE clusters, respectively (FDR 10%) (Table S7). The most strongly associated proteins for the ULE cluster were related to mitochondrial function (ATP synthase 5 peripheral stalk subunit OSCP and glucokinase regulatory protein) and cytoskeleton 6 7 rearrangement (serine/threonine-protein kinases MRCK beta and PAK 6) (Table 2). Top proteins associations for the AD cluster were primarily involved in fatty acid metabolism, such as 8 9 elevated fatty acid-binding protein, leptin, and retinoic acid receptor responder protein 2 and decreased apolipoprotein M. For the SE cluster, top associated proteins were related to innate 10 immunity, such as bactericidal permeability-increasing protein, complement component C9, and 11 protein S100-A12. The overlap of the protein associations between subtypes is shown in Figure 12 S4. 13

Using blood RNA sequencing data from 2,072 subjects at Visit 2, we identified 14 15 significant associations to 3, 2,105, and 148 genes for the ULE, AD, and SE clusters, respectively with the RRS group as the reference. The Bland-Altman plots are shown in Figure 4 16 17 and the complete set of association results is reported in Table S8. Compared to the RRS cluster, the ULE cluster was associated with up-regulation of the GPR15, AHRR and GPR55 genes, and 18 GO pathway enrichment analysis results did not identify any significantly enriched pathways 19 20 (Table 3). The AD cluster had particularly strong differences in gene expression with 22 and 34 21 enriched pathways relative to the RRS and SE clusters, respectively. Compared to the RRS 22 cluster, the AD cluster showed significant associations to pathways involved in innate immunity, 23 cellular defense response and NF-kB signaling. Relative to the SE cluster, the AD cluster

demonstrated associations to processes involved in both innate and adaptive immunity (*adjusted P-values* < 0.005). The SE cluster also had many differentially expressed genes, with significant
pathway enrichment for positive regulation of synapse assembly and cell adhesion. The complete
set of pathway enrichment results is reported in Table S9.

5 **Discussion**

In this study, we demonstrated that k-means subgroups of smokers enriched for COPD 6 have varying disease progression patterns, development of prospective comorbidities, and 7 8 distinct associations to plasma protein and blood transcriptomic biomarkers. The two clusters at 9 the extremes of the lung health spectrum (the RRS and SE clusters) showed high cluster assignment stability between the two visits. Spirometric progression was sensitive to the 10 11 progression metric being used (absolute vs. relative) with the SE subtype showing the most rapid rate of FEV_1 decline relative to baseline FEV_1 level. Emphysema progression however was less 12 sensitive to the use of absolute versus relative metrics of progression. In general, the SE cluster 13 had the highest risk for prospective adverse health events, though the AD cluster had the highest 14 risk of incident diabetes and the most distinct gene expression patterns. 15

16 A wide array of risk factors have been associated with spirometric decline, such as low BMI, higher baseline FEV₁ and FVC, smoking exposure, bronchodilator reversibility, African-17 American race, female sex, previous history of exacerbations, CT-quantified emphysema, upper 18 lobe predominant emphysema and small airway abnormalities ^{1 2 30-34}. Whereas previous studies 19 20 have shown that mild-to-moderate COPD (GOLD 1-2) is associated with an increased loss in absolute $FEV_1 \stackrel{2}{}^{30}$, our study also investigated changes in FEV_1 relative to baseline values, and 21 22 found that the SE cluster, which represents a subset of subjects with advanced disease, is associated with more rapid relative FEV₁ decline. To our knowledge, such differences in the 23

12

metrics of progression used to assess COPD progression have not been previously reported. This
finding emphasizes the fact that disease activity and disease severity are distinct concepts that
should be considered when assessing COPD patients ^{1 35}.

In contrast to FEV₁ changes, less is known about the factors that are associated with emphysema progression ³⁶⁻³⁸. In our study, we evaluated emphysema changes by COPD subtype, investigated both absolute and relative changes, and adjusted the analysis for age, sex, race, BMI, smoking pack-years, and CT scanner types. We showed that relative changes were the highest in the SE cluster, which has more advanced baseline emphysema, low BMI, and more COPD exacerbations.

In regard to prospective COPD events, prior reports have indicated that more severe 10 airflow obstruction and COPD exacerbation history are associated with higher risks of COPD 11 exacerbations and CVD ^{7 39-41}. We demonstrated that, when adjusted for covariates including 12 airflow limitation and COPD exacerbation history, the SE cluster had a 3 times higher risk of a 13 COPD exacerbation, while the AD and ULE clusters had a hazard ratio of ~1.35 relative to the 14 RRS group. The SE cluster was also associated with the highest risk of incident CVD, which 15 may be explained by atherosclerosis or arterial stiffness mediated by inflammatory markers ⁴². 16 17 Another novel finding from our study was that the AD cluster had the highest risk for the prospective development of diabetes, even after adjusting for BMI. This adds to the finding from 18 the study by Hersh et al., which revealed that COPD with limited emphysema and high airflow 19 obstruction is associated with diabetes in cross-sectional data ⁴³. Our study provides further 20 evidence linking airway-predominant COPD to diabetes and metabolic syndrome. 21

The age, dyspnea, and airflow obstruction (ADO) and BODE indices have been classically used to predict mortality in COPD, but less is known about the COPD characteristics

13

that contribute to cause-specific mortality ^{44 45}. CVD, airflow obstruction, low BMI, emphysema, 1 and poor exercise capacity were found to be associated with a high risk of all-cause mortality in 2 COPD ^{39 46-48}. We similarly observed that the SE cluster, which had these risk factors at baseline. 3 4 had the highest risk of all-cause mortality. With regards to the CVD mortality, high MMRC dyspnea score, exacerbation history, low FEV₁, and accelerated lung function decline have been 5 shown to be contributors ^{49 50}. In our paper, we showed that the SE, AD, and ULE clusters had 6 7 elevated risks of CVD-related mortality in unadjusted models, but when correcting for age, sex, race, BODE, and smoking pack-years, only the AD and ULE clusters had significant risks, which 8 9 indicates that airway disease and lobular emphysema distribution may be additional contributors 10 to this risk. While most previously published COPD studies have examined the association with prospective cancer rather than prospective mortality due to cancer, some reports have indicated 11 12 that COPD is associated with an elevated risk of both lung and extra-pulmonary cancer mortalities ^{51 52}. In our study, similar to CVD mortality, the ULE, AD, and SE clusters had 13 elevated cancer mortality risks, but after adjustments, only the ULE cluster had a significantly 14 increased cancer mortality risk. Differential genetic susceptibility, delayed clearance of inhaled 15 carcinogens, and chronic inflammation may be potential mechanisms underlying these 16 associations ⁵². 17

18 COPD subtype-specific associations to protein and transcriptomic biomarkers have only 19 been studied in a few studies of relatively small size ¹⁴. In our analysis, the AD cluster had 20 elevated plasma levels of leptin, supporting a previous report which showed that pro-21 inflammatory adipokines leptin and adiponectin are implicated in COPD ⁵³. While many 22 previous studies have shown that innate immunity is related to COPD ^{54 55}, we observed that it 23 was the SE cluster that demonstrated the strongest association to proteins involved in the innate

immune response. At the transcriptomic level, COPD has been associated with inflammation and
 sphingolipid metabolism ¹⁵⁻¹⁷, and our study demonstrates that it is the AD subtype that has the
 most distinct gene expression signature that is enriched for these pathways.

4 This study has a number of strengths. Compared to previous publications, our study included a larger sample size and longer follow-up, and investigated incident comorbidities in 5 addition to lung health outcomes. Additionally, we included a well-phenotyped cohort of 6 7 smokers across the full spectrum of disease severity, and we were able to test the association of subtypes to clinical, radiologic, and multi-omic molecular markers. When studying progression, 8 9 we considered both absolute and relative changes in lung function and CT-quantified emphysema. Furthermore, this is to our knowledge the first study demonstrating that airway-10 predominant COPD is independently associated with incident diabetes risk and has particularly 11 12 strong associations to inflammatory biomarkers.

One of the limitations of our study is that because of its observational design, the statistical associations observed may not reflect causal relationships. Our study sample was limited to subjects who survived the five-year observation period and as a result, our findings are not representative of subjects with very advanced COPD and limited life expectancy. Analyses that jointly model death and other aspects of disease progression would provide additional useful information. As further follow-up data from COPDGene is obtained, our hypotheses from this study can be validated over longer follow-up.

20

21 **Conclusions**

22 COPD-related subtypes defined by spirometric and radiologic measures at baseline have
 23 different rates of disease progression and are differentially associated to prospective health

15

outcomes. They also exhibit distinct biomarker profiles indicative of underlying biological
differences. In the future, these subtypes could be used as the basis for targeted drug
development, studies of differential treatment response, or the enrollment of specific subgroups
in clinical trials.

8

9 **Figure legends:**

Figure 1. Study design. The goal of the study was to analyze COPD progression, differential plasma protein associations and blood gene expression, and gene ontology (GO) enrichment characteristics of the four clusters that we identified in our previous k-means clustering analysis in the COPDGene study (*Castaldi et al, Thorax 2014*).

Figure 2. Disease progression by k-means cluster. (*A*) Absolute change in FEV₁ (mL/year). (*B*) Relative change in FEV₁ (change as % of baseline value/year). (*C*) Absolute change in emphysema measured as adjusted Perc15 density change/year. (*D*) Relative change in emphysema measured as adjusted Perc15 density change (% of baseline value/year). P-values <0.05 are indicated by an asterisk. Abbreviations: RRS = Relatively resistant smokers; ULE = Upper lobe predominant emphysema; AD = Airway-predominant disease; SE = Severe emphysema.

Figure 3. Kaplan-Meier plots of COPD-related events by k-means cluster. (A) COPD
 exacerbation, defined as the acute worsening of respiratory symptoms that required antibiotics

1	and/or systemic steroids. (B) Cardiovascular disease (CVD) event, defined as a composite
2	endpoint of stroke, heart attack, coronary artery disease, coronary artery bypass graft surgery,
3	peripheral artery disease, and/or cardiac angina. (C) Diabetes. (D) All-cause mortality. For CVD
4	events and diabetes, subjects who had a history of CVD or diabetes at Visit 1 were excluded
5	from the analysis. Abbreviations: RRS = Relatively resistant smokers; ULE = Upper lobe
6	predominant emphysema; AD = Airway-predominant disease; SE = Severe emphysema.
7	Figure 4. Bland-Altman (MA) plots of the log ratio versus mean gene expression for the
8	differential expression analysis results between k-means clusters. The cluster following the "vs."
9	is the reference group. (A) ULE vs. RRS. (B) AD vs. RRS. (C) SE vs. RRS. (D) AD vs. ULE. (E)
10	SE vs. ULE. (F) SE vs. AD. Abbreviations: $RRS = Relatively resistant smokers; ULE = Upper$
11	lobe predominant emphysema; $AD = Airway$ -predominant disease; $SE = Severe emphysema$.
12	
13	
14	
15	
16	
17	
18	
19	
20	

REFERENCES

- 1. Casanova C, de Torres JP, Aguirre-Jaíme A, et al. The progression of chronic obstructive pulmonary disease is heterogeneous: the experience of the BODE cohort. *Am J Respir Crit Care Med* 2011;184(9):1015-21.
- 2. Tantucci C, Modina D. Lung function decline in COPD. Int J Chron Obstruct Pulmon Dis 2012;7:95-9.
- 3. Tanabe N, Muro S, Hirai T, et al. Impact of exacerbations on emphysema progression in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2011;183(12):1653-9.
- 4. Cavaillès A, Brinchault-Rabin G, Dixmier A, et al. Comorbidities of COPD. *Eur Respir Rev* 2013;22(130):454-75.
- 5. Sun Y, Milne S, Jaw JE, et al. BMI is associated with FEV(1) decline in chronic obstructive pulmonary disease: a meta-analysis of clinical trials. *Respir Res* 2019;20(1):236.
- 6. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. *N Engl J Med* 2010;363(12):1128-38.
- 7. Kim S, Lim MN, Hong Y, et al. A cluster analysis of chronic obstructive pulmonary disease in dusty areas cohort identified three subgroups. *BMC Pulm Med* 2017;17(1):209.
- 8. Arostegui I, Esteban C, García-Gutierrez S, et al. Subtypes of patients experiencing exacerbations of COPD and associations with outcomes. *PLoS One* 2014;9(6):e98580.
- 9. Esteban C, Arostegui I, Aburto M, et al. Chronic Obstructive Pulmonary Disease Subtypes. Transitions over Time. *PLoS One* 2016;11(9):e0161710.
- 10. Garcia-Aymerich J, Gómez FP, Benet M, et al. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. *Thorax* 2011;66(5):430-7.
- 11. Rennard SI, Locantore N, Delafont B, et al. Identification of five chronic obstructive pulmonary disease subgroups with different prognoses in the ECLIPSE cohort using cluster analysis. *Ann Am Thorac Soc* 2015;12(3):303-12.
- 12. Sode BF, Dahl M, Nordestgaard BG. Myocardial infarction and other co-morbidities in patients with chronic obstructive pulmonary disease: a Danish nationwide study of 7.4 million individuals. *Eur Heart J* 2011;32(19):2365-75.
- 13. Lee CT, Mao IC, Lin CH, et al. Chronic obstructive pulmonary disease: a risk factor for type 2 diabetes: a nationwide population-based study. *Eur J Clin Invest* 2013;43(11):1113-9.
- 14. Regan EA, Hersh CP, Castaldi PJ, et al. Omics and the Search for Blood Biomarkers in Chronic Obstructive Pulmonary Disease. Insights from COPDGene. *Am J Respir Cell Mol Biol* 2019;61(2):143-49.
- 15. Bahr TM, Hughes GJ, Armstrong M, et al. Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol* 2013;49(2):316-23.
- 16. Spira A, Beane J, Pinto-Plata V, et al. Gene expression profiling of human lung tissue from smokers with severe emphysema. *Am J Respir Cell Mol Biol* 2004;31(6):601-10.
- 17. Boudewijn IM, Faiz A, Steiling K, et al. Nasal gene expression differentiates COPD from controls and overlaps bronchial gene expression. *Respir Res* 2017;18(1):213.
- 18. Castaldi PJ, Dy J, Ross J, et al. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* 2014;69(5):415-22.
- 19. A. Gregory SL, J. H. Yun, A. Saferali, C. P. Hersh, R. P. Bowler, E. K. Silverman, A. R. El Boueiz, P. Castaldi. K-Means Subtypes of Smokers in the COPDGene Study Have Distinct Patterns of COPD Progression and Distinct Serum Protein Biomarker and Transcriptomic Profiles *Am J Respir Crit Care Med* 2020
- 20. Regan EA, Hokanson JE, Murphy JR, et al. Genetic epidemiology of COPD (COPDGene) study design. *Copd* 2010;7(1):32-43.

- 21. Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2013;187(4):347-65.
- 22. McGarvey LP, John M, Anderson JA, et al. Ascertainment of cause-specific mortality in COPD: operations of the TORCH Clinical Endpoint Committee. *Thorax* 2007;62(5):411-5.
- 23. Coxson HO, Rogers RM, Whittall KP, et al. A quantification of the lung surface area in emphysema using computed tomography. *Am J Respir Crit Care Med* 1999;159(3):851-6.
- 24. Parr DG, Sevenoaks M, Deng C, et al. Detection of emphysema progression in alpha 1-antitrypsin deficiency using CT densitometry; methodological advances. *Respir Res* 2008;9(1):21.
- 25. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012;40(6):1324-43.
- 26. Carolan BJ, Hughes G, Morrow J, et al. The association of plasma biomarkers with computed tomography-assessed emphysema phenotypes. *Respir Res* 2014;15(1):127.
- 27. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.
- 28. Alexa A RJ. topGO: Enrichment Analysis for Gene Ontology. R package version 2.40.0. 2020.
- 29. Strimmer K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* 2008;24(12):1461-2.
- Bhatt SP, Soler X, Wang X, et al. Association between Functional Small Airway Disease and FEV1 Decline in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2016;194(2):178-84.
- 31. Dransfield MT, Kunisaki KM, Strand MJ, et al. Acute Exacerbations and Lung Function Loss in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2017;195(3):324-30.
- 32. Vestbo J, Edwards LD, Scanlon PD, et al. Changes in forced expiratory volume in 1 second over time in COPD. *N Engl J Med* 2011;365(13):1184-92.
- 33. Nishimura M, Makita H, Nagai K, et al. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185(1):44-52.
- 34. Mohamed Hoesein FA, van Rikxoort E, van Ginneken B, et al. Computed tomography-quantified emphysema distribution is associated with lung function decline. *Eur Respir J* 2012;40(4):844-50.
- 35. Carter RI, Stockley RA. Disease 'activity', 'severity' and 'impact': interrelationships in COPD; is a measure of disease 'activity' the Holy Grail for COPD, or a variable impossible to quantify? *Copd* 2014;11(4):363-7.
- 36. Boueiz A, Chang Y, Cho MH, et al. Lobar Emphysema Distribution Is Associated With 5-Year Radiological Disease Progression. *Chest* 2018;153(1):65-76.
- 37. Coxson HO, Dirksen A, Edwards LD, et al. The presence and progression of emphysema in COPD as determined by CT scanning and biomarker expression: a prospective analysis from the ECLIPSE study. *Lancet Respir Med* 2013;1(2):129-36.
- 38. Bhavani S, Tsai CL, Perusich S, et al. Clinical and Immunological Factors in Emphysema Progression. Five-Year Prospective Longitudinal Exacerbation Study of Chronic Obstructive Pulmonary Disease (LES-COPD). Am J Respir Crit Care Med 2015;192(10):1171-8.
- 39. Müllerova H, Maselli DJ, Locantore N, et al. Hospitalized exacerbations of COPD: risk factors and outcomes in the ECLIPSE cohort. *Chest* 2015;147(4):999-1007.
- 40. Ford ES, Wheaton AG, Mannino DM, et al. Elevated cardiovascular risk among adults with obstructive and restrictive airway functioning in the United States: a cross-sectional study of the National Health and Nutrition Examination Survey from 2007-2010. *Respir Res* 2012;13(1):115.

- 41. Kunisaki KM, Dransfield MT, Anderson JA, et al. Exacerbations of Chronic Obstructive Pulmonary Disease and Cardiac Events. A Post Hoc Cohort Analysis from the SUMMIT Randomized Clinical Trial. *Am J Respir Crit Care Med* 2018;198(1):51-57.
- 42. Morgan AD, Zakeri R, Quint JK. Defining the relationship between COPD and CVD: what are the implications for clinical practice? *Ther Adv Respir Dis* 2018;12:1753465817750524.
- 43. Hersh CP, Make BJ, Lynch DA, et al. Non-emphysematous chronic obstructive pulmonary disease is associated with diabetes mellitus. *BMC Pulm Med* 2014;14:164.
- 44. Marin JM, Alfageme I, Almagro P, et al. Multicomponent indices to predict survival in COPD: the COCOMICS study. *Eur Respir J* 2013;42(2):323-32.
- 45. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009;374(9691):704-11.
- 46. Moll M, Qiao D, Regan EA, et al. Machine Learning and Prediction of All-Cause Mortality in COPD. *Chest* 2020;158(3):952-64.
- 47. Esteban C, Quintana JM, Aburto M, et al. Predictors of mortality in patients with stable COPD. *J Gen Intern Med* 2008;23(11):1829-34.
- 48. Huiart L, Ernst P, Suissa S. Cardiovascular morbidity and mortality in COPD. *Chest* 2005;128(4):2640-6.
- 49. Whittaker HR, Bloom C, Morgan A, et al. Accelerated FEV(1) decline and risk of cardiovascular disease and mortality in a primary care population of COPD patients. *Eur Respir J* 2021;57(3)
- 50. Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest* 2005;127(6):1952-9.
- 51. Wasswa-Kintu S, Gan WQ, Man SF, et al. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. *Thorax* 2005;60(7):570-5.
- 52. van Gestel YR, Hoeks SE, Sin DD, et al. COPD and cancer mortality: the influence of statins. *Thorax* 2009;64(11):963-7.
- 53. Ali Assad N, Sood A. Leptin, adiponectin and pulmonary diseases. *Biochimie* 2012;94(10):2180-9.
- 54. Borger JG, Lau M, Hibbs ML. The Influence of Innate Lymphoid Cells and Unconventional T Cells in Chronic Inflammatory Lung Disease. *Front Immunol* 2019;10:1597.
- 55. Shaykhiev R, Crystal RG. Innate immunity and chronic obstructive pulmonary disease: a mini-review. *Gerontology* 2013;59(6):481-9.

	K-means cluster	Univariable models		Multivariable models	
		Beta (Std Err)	P-value	Beta (SE)	Beta (Std Err)
	ULE	3.07 (2.21)	0.2	3.63 (2.23)	0.1
Absolute annualized FEV_1 (mL/year) changes	AD	11.64 (1.86)	< 0.0001	8.84 (1.90)	< 0.0001
	SE	2.84 (2.17)	0.2	4.98 (2.25)	0.03
Deleting annualized FEV shanges (noncert shanges from	ULE	-0.27 (0.11)	0.01	-0.18 (0.11)	0.1
baseline) Relative annualized $F \ge v_1$ changes (percent changes from	AD	0.09 (0.09)	0.3	0.004 (0.09)	1.0
	SE	-1.35 (0.10)	< 0.0001	-1.24 (0.11)	< 0.0001
	ULE	-0.99 (0.11)	< 0.0001	-0.89 (0.11)	< 0.0001
changes	AD	-0.69 (0.09)	< 0.0001	-0.82 (0.09)	< 0.0001
	SE	-0.89 (0.11)	< 0.0001	-0.77 (0.11)	< 0.0001
	ULE	-1.15 (0.16)	< 0.0001	-1.01 (0.16)	< 0.0001
Relative annualized emphysema (adjusted Perc15 density) changes (percent changes from baseline)	AD	-0.82 (0.13)	< 0.0001	-0.99 (0.14)	< 0.0001
	SE	-1.19 (0.16)	< 0.0001	-1.07 (0.16)	< 0.0001

<u>**Table 1.**</u> Associations of k-means clusters with absolute and relative annualized FEV_1 and emphysema changes

Absolute annualized changes were computed by subtracting Visit 1 values from Visit 2 values and dividing by the time in years between both visits for each subject. Relative annualized changes were calculated by dividing absolute annualized changes by Visit 1 values and multiplying by 100. Negative values indicate worsening of the disease between visits.

Univariable linear regression models included only visit 1 k-means cluster assignment. Multivariable models also included adjustments for age, sex, race, BMI, and smoking pack-years. The reference group was the relatively resistant smokers cluster (RRS) cluster. P-values < 0.05 are italicized.

Abbreviations: ULE = Upper lobe predominant emphysema; AD = Airway-predominant disease; SE = Severe emphysema.

	Protein	Protein ID	Beta coefficient (standard error)	FDR
	ATP synthase peripheral stalk subunit OSCP	P48047	0.33 (0.07)	6.3*10 ⁻³
	Glucokinase regulatory protein	Q14397	0.14 (0.04)	4.2*10 ⁻²
ULE vs. RRS	Serine/threonine-protein kinase MRCK beta	Q9Y5S2	0.09 (0.02)	4.2*10 ⁻²
	Serine/threonine-protein kinase PAK 6	Q9NQU5	0.37 (0.10)	4.6*10 ⁻²
	Membrane frizzled-related protein	Q9BY79	0.14 (0.04)	5.5*10 ⁻²
	Fatty acid-binding protein, heart	P05413	0.20 (0.03)	1.3*10 ⁻⁶
	Leptin	P41159	0.34 (0.06)	4.8*10 ⁻⁶
AD vs. RRS	Renin	P00797	0.25 (0.05)	1.6*10 ⁻⁴
	Retinoic acid receptor responder protein 2	Q99969	0.08 (0.02)	2.6*10 ⁻⁴
	Apolipoprotein M	O95445	-0.15 (0.03)	3.9*10 ⁻⁴
	Bactericidal permeability-increasing protein	P17213	0.40 (0.06)	1.8*10 ⁻¹¹
	Complement component C9	P02748	0.15 (0.02)	1.9*10 ⁻¹¹
SE vs. RRS	Troponin T, cardiac muscle	P45379	0.20 (0.03)	2.1*10-10
	Protein S100-A12	P80511	0.19 (0.03)	3.4*10 ⁻¹⁰
	Oxidized low-density lipoprotein receptor 1	P78380	0.23 (0.04)	1.1*10 ⁻⁹

SOMAscan plasma proteins significantly associated to visit 1 k-means cluster membership from multivariable linear regression modeling (*FDR 10%*). We used linear regression and adjusted for age, sex, race, and current smoking status. The top 5 significantly differentially associated proteins unique to each of the 3 comparisons (ULE *vs.* RRS, AD *vs.* RRS, and SE *vs.* RRS) are reported in this table. The cluster following the "*vs.*" is the reference group. The units of all proteins are relative fluorescence units. Abbreviations: RRS = Relatively resistant smokers; ULE = Upper lobe predominant emphysema; AD = Airway-predominant disease; SE = Severe emphysema.

	Top 10 significant genes	Top 10 significant GO terms
ULE vs. RRS	↑ <i>GPR15</i> ↑ <i>AHRR</i> ↑ <i>GPR55</i>	NS
AD vs. RRS	↑ GPR15 ↓ SNRK ↑ PTPN13 ↓ ALG1L13P ↑ CD36 ↓ MYLIP ↓ EXTL3 ↓ IL18R1 ↓ ADPRHL2 ↑ PTPRA	Innate immune response Mitotic chromosome movement towards spindle pole Negative regulation of NIK/NF-kappaB signaling Cellular defense response Cell proliferation Regulation of mitochondrial depolarization Negative regulation of defense response to virus Mitotic cytokinesis Fc-epsilon receptor signaling pathway Vascular endothelial growth factor receptor signaling pathway

Table 3. Top 10 significantly differentially expressed genes and enriched gene ontology (GO) terms between k-means clusters

SE vs. RRS	↑ <i>GPR15</i> ↑ <i>MS4A2</i> ↑ <i>TTLL7</i> ↑ <i>GCSAML</i> ↑ <i>AKAP12</i> ↑ <i>LINC02458</i> ↓ <i>COG2</i> ↑ <i>HDC</i> ↓ <i>DLEU1</i> ↑ <i>ENSG00000261055</i> (LncRNA)	Positive regulation of synapse assembly Cell adhesion
------------	--	--

For the gene expression analysis, covariates used were age, sex, race, current smoking status, white blood cell count proportions, and library batch effects. A false discovery rate (FDR) of 10% was used for multiple testing corrections.

For the GO analysis, we only reported the pathways with at least 3 significant genes. Enriched GO terms between clusters were identified using the weighted Fisher's test P-values < 0.005. The cluster following the "vs." is the reference group.

Abbreviations: RRS = Relatively resistant smokers; ULE = Upper lobe predominant emphysema; AD = Airway-predominant disease; SE = Severe emphysema. NS = Non-significant. \uparrow = positive log fold change; \downarrow = negative log fold change.

Figure 1



Four k-means clusters (Castaldi et al, Thorax 2014):

- Relatively resistant smokers
- Upper lobe predominant emphysema
- Airway-predominant disease
- Severe emphysema

Data collection: - COPDGene Visits 1 and 2 (5-year follow-up) - Longitudinal Follow-up Program: Surveys every 3-6 months from Visit 1



COPD progression measurements:

FEV₁ and emphysema changes between visits
COPD exacerbations, cardiovascular events, diabetes, and survival over 8 years

Differential plasma protein associations between clusters - Visit 1



Differential blood gene expression and gene ontology enrichment analyses between clusters - Visit 2



Figure 2







Figure 4





ACKNOWLEDGEMENTS

COPDGene Investigators -Core Units:

Administrative Center: James D. Crapo, MD (PI); Edwin K. Silverman, MD, PhD (PI); Barry J. Make, MD; Elizabeth A. Regan, MD, PhD

Genetic Analysis Center: Terri H. Beaty, PhD; Peter J. Castaldi, MD, MSc; Michael H. Cho, MD, MPH; Dawn L. DeMeo, MD, MPH; Adel Boueiz, MD, MMSc; Marilyn G. Foreman, MD, MS; Auyon Ghosh, MD; Lystra P. Hayden, MD, MMSc; Craig P. Hersh, MD, MPH; Jacqueline Hetmanski, MS; Brian D. Hobbs, MD, MMSc; John E. Hokanson, MPH, PhD; Wonji Kim, PhD; Nan Laird, PhD; Christoph Lange, PhD; Sharon M. Lutz, PhD; Merry-Lynn McDonald, PhD; Dmitry Prokopenko, PhD; Matthew Moll, MD, MPH; Jarrett Morrow, PhD; Dandi Qiao, PhD; Elizabeth A. Regan, MD, PhD; Aabida Saferali, PhD; Phuwanat Sakornsakolpat, MD; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Jeong Yun, MD, MPH

Imaging Center: Juan Pablo Centeno; Jean-Paul Charbonnier, PhD; Harvey O. Coxson, PhD; Craig J. Galban, PhD; MeiLan K. Han, MD, MS; Eric A. Hoffman, Stephen Humphries, PhD; Francine L. Jacobson, MD, MPH; Philip F. Judy, PhD; Ella A. Kazerooni, MD; Alex Kluiber; David A. Lynch, MB; Pietro Nardelli, PhD; John D. Newell, Jr., MD; Aleena Notary; Andrea Oh, MD; Elizabeth A. Regan, MD, PhD; James C. Ross, PhD; Raul San Jose Estepar, PhD; Joyce Schroeder, MD; Jered Sieren; Berend C. Stoel, PhD; Juerg Tschirren, PhD; Edwin Van Beek, MD, PhD; Bram van Ginneken, PhD; Eva van Rikxoort, PhD; Gonzalo Vegas SanchezFerrero, PhD; Lucas Veitel; George R. Washko, MD; Carla G. Wilson, MS

PFT QA Center, Salt Lake City, UT: Robert Jensen, PhD

Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO: Douglas Everett, PhD; Jim Crooks, PhD; Katherine Pratte, PhD; Matt Strand, PhD; Carla G. Wilson, MS

Epidemiology Core, University of Colorado Anschutz Medical Campus, Aurora, CO: John E. Hokanson, MPH, PhD; Erin Austin, PhD; Gregory Kinney, MPH, PhD; Sharon M. Lutz, PhD; Kendra A. Young, PhDVersion Date: March 26, 2021

Mortality Adjudication Core: Surya P. Bhatt, MD; Jessica Bon, MD; Alejandro A. Diaz, MD, MPH; MeiLan K. Han, MD, MS; Barry Make, MD; Susan Murray, ScD; Elizabeth Regan, MD; Xavier Soler, MD; Carla G. Wilson, MS

Biomarker Core: Russell P. Bowler, MD, PhD; Katerina Kechris, PhD; Farnoush BanaeiKashani, PhD

COPDGene Investigators -Clinical Centers:

Ann Arbor VA: Jeffrey L. Curtis, MD; Perry G. Pernicano, MD

Baylor College of Medicine, Houston, TX: Nicola Hanania, MD, MS; Mustafa Atik, MD; Aladin Boriek, PhD; Kalpatha Guntupalli, MD; Elizabeth Guy, MD; Amit Parulekar, MD

Brigham and Women's Hospital, Boston, MA: Dawn L. DeMeo, MD, MPH; Craig Hersh, MD, MPH; Francine L. Jacobson, MD, MPH; George Washko, MD

Columbia University, New York, NY: R. Graham Barr, MD, DrPH; John Austin, MD; Belinda D'Souza, MD; Byron Thomashow, MD

Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD; H. Page McAdams, MD; Lacey Washington, MD

HealthPartners Research Institute, Minneapolis, MN: Charlene McEvoy, MD, MPH; Joseph Tashjian, MD

Johns Hopkins University, Baltimore, MD: Robert Wise, MD; Robert Brown, MD; Nadia N. Hansel, MD, MPH; Karen Horton, MD; Allison Lambert, MD, MHS; Nirupama Putcha, MD, MHS

Lundquist Institute for Biomedical Innovation at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi, PhD, MD; Alessandra Adami, PhD; Matthew Budoff, MD; Hans Fischer, MD; Janos Porszasz, MD, PhD; Harry Rossiter, PhD; William Stringer, MD

Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh, MD, PhD; Charlie Lan, DO

Minneapolis VA: Christine Wendt, MD; Brian Bell, MD; Ken M. Kunisaki, MD, MS

Morehouse School of Medicine, Atlanta, GA: Eric L. Flenaugh, MD; Hirut Gebrekristos, PhD; Mario Ponce, MD; Silanath Terpenning, MD; Gloria Westney, MD, MS

National Jewish Health, Denver, CO: Russell Bowler, MD, PhD; David A. Lynch, MB

Reliant Medical Group, Worcester, MA: Richard Rosiello, MD; David Pace, MD

Temple University, Philadelphia, PA: Gerard Criner, MD; David Ciccolella, MD; Francis Cordova, MD; Chandra Dass, MD; Gilbert D'Alonzo, DO; Parag Desai, MD; Michael Jacobs, PharmD; Steven Kelsen, MD, PhD; Victor Kim, MD; A. James Mamary, MD; Nathaniel Marchetti, DO; Aditi Satti, MD; Kartik Shenoy, MD; Robert M. Steiner, MD; Alex Swift, MD; Irene Swift, MD; Maria Elena Vega-Sanchez, MD

University of Alabama, Birmingham, AL: Mark Dransfield, MD; William Bailey, MD; Surya P. Bhatt, MD; Anand Iyer, MD; Hrudaya Nath, MD; J. Michael Wells, MD

University of California, San Diego, CA: Douglas Conrad, MD; Xavier Soler, MD, PhD; Andrew Yen, MD

University of Iowa, Iowa City, IA: Alejandro P. Comellas, MD; Karin F. Hoth, PhD; John Newell, Jr., MD; Brad Thompson, MD

University of Michigan, Ann Arbor, MI: MeiLan K. Han, MD MS; Ella Kazerooni, MD MS; Wassim Labaki, MD MS; Craig Galban, PhD; Dharshan Vummidi, MD University of Minnesota, Minneapolis, MN: Joanne Billings, MD; Abbie Begnaud, MD; Tadashi Allen, MD

University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD; Jessica Bon, MD; Divay Chandra,

MD, MSc; Joel Weissfeld, MD, MPH

University of Texas Health, San Antonio, San Antonio, TX: Antonio Anzueto, MD; Sandra Adams, MD; Diego Maselli-Caceres, MD; Mario E. Ruiz, MD; Harjinder Singh