EUROPEAN RESPIRATORY journal
FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original research article

A Polygenic Risk Score and Age of Diagnosis of Chronic Obstructive Pulmonary Disease

Jingzhou Zhang, Hanfei Xu, Dandi Qiao, Dawn L. DeMeo, Edwin K. Silverman, George T. O'Connor, Brian D. Hobbs, Josée Dupuis, Michael H. Cho, Matthew Moll

Please cite this article as: Zhang J, Xu H, Qiao D, *et al*. A Polygenic Risk Score and Age of Diagnosis of Chronic Obstructive Pulmonary Disease. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.01954-2021).

A Polygenic Risk Score and Age of Diagnosis of Chronic Obstructive Pulmonary Disease

Jingzhou Zhang[1,2], Hanfei Xu[3], Dandi Qiao[1], Dawn L. DeMeo[1,4], Edwin K. Silverman[1,4], George T. O'Connor[2], Brian D. Hobbs[1,4], Josée Dupuis[3], Michael H. Cho[1,4]*, Matthew Moll[1,4]*

1 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115

2 The Pulmonary Center, Section of Pulmonary, Allergy, Sleep, and Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118

3 Department of Biostatistics, Boston University School of Public Health, MA 02118

4 Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115

*=co-corresponding authors

Correspondence:
Matthew Moll, M.D., M.P.H.
181 Longwood Avenue
Boston, MA 02115
remol@channing.harvard.edu

and/or

Michael H. Cho, M.D., M.P.H.
181 Longwood Avenue
Boston, MA 02115
remhc@channing.harvard.edu

Funding:
MM is supported by T32HL007427.
BDH is supported by NIH K08 HL136928, U01 HL089856, R01 HL147148, and R01 HL135142.
MHC is supported by NIH R01HL137927, R01HL135142, HL147148, and HL089856.
EKS is supported by NIH R01 HL137927, R01 HL147148, U01 HL089856, R01 HL133135, P01 HL132825, and P01 HL114501.
DLD is supported by P01 HL132825, and P01 HL114501 and a grant from the Alpha-1 Foundation.

The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board that has included AstraZeneca, Bayer Pharmaceuticals, Boehringer Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, and Sunovion.

The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195, HHSN268201500001I and 75N92019D00031).

Disclosures:
EKS received grant support from GlaxoSmithKline and Bayer. MHC has received grant support from GlaxoSmithKline and Bayer, consulting fees from Genentech and AstraZeneca, and speaking fees from Illumina. DLD has received support from Bayer and Honoraria from Novartis. JD received NIH funding for salary coverage paid to Boston University. JZ, HX, DQ, GTO, BDH, and MM have no conflict of interest to declare.

Author contributions:
Study Design: Jingzhou Zhang, Matthew Moll, Brian D. Hobbs, Michael H. Cho.

Acquisition, analysis, or interpretation of the data: All authors

Drafting of the manuscript: Jingzhou Zhang, Matthew Moll, Michael Cho

Critical revision of the manuscript for important intellectual content: All authors

Obtained funding: Edwin K. Silverman, Michael H. Cho.

Word counts:
Abstract: 248
Manuscript: 4017

Take home message:

A polygenic risk score is associated with earlier age of chronic obstructive pulmonary disease (COPD) diagnosis and is predictive for COPD occurring early in life, with clinical implications of individualized risk stratification and preventive measures.

Abstract

Genetic susceptibility may be associated with earlier onset of chronic obstructive pulmonary disease (COPD). We hypothesized that a polygenic risk score (PRS) for COPD would be associated with earlier age of diagnosis of COPD. In 6647 non-Hispanic white (NHW) and 2464 African American (AA) participants from COPDGene, and 6812 participants from the Framingham Heart Study (FHS), we tested the relationship of the PRS and age of COPD diagnosis. Age at diagnosis was determined by: 1) self-reported age at COPD diagnosis, or 2) age at visits when moderate-to-severe airflow limitation (GOLD 2-4) was observed on spirometry. We used Cox regression to examine the overall and time-dependent effects of the PRS on incident COPD. In the COPDGene study, we also examined the PRS's predictive value for COPD at age < 50 years (COPD50) using logistic regression and area-under-the-curve (AUC) analyses, with and without the addition of other risk factors present at early life (e.g.,

childhood asthma). In Cox models, the PRS demonstrated age-dependent associations with incident COPD, with larger effects at younger ages in both cohorts. The PRS was associated with COPD50 (OR [95% CI]: NHW 1.55 [1.41-1.71], AA 1.23 [1.05-1.43], FHS 2.47 [2.12-2.88]). In COPDGene, adding the PRS to known early-life risk factors improved prediction of COPD50 in NHW (AUC 0.69 *vs.* 0.74, p <0.0001) and AA participants (AUC 0.61 *vs.* 0.64, p =0.04). A COPD polygenic risk score is associated with earlier age of diagnosis of COPD and retains predictive value when added to known early-life risk factors.

Key words: age-dependent effect, age of diagnosis, incident COPD, polygenic risk score, risk prediction

Introduction

Characterized by persistent airflow limitation and respiratory symptoms, chronic obstructive pulmonary disease (COPD) is a leading cause of mortality and morbidity worldwide [1, 2]. Emerging evidence indicates that the pathogenesis of COPD can begin in early life and even the prenatal period [3, 4]. Many patients with COPD are undiagnosed or initially present with advanced disease, missing a potential opportunity for early intervention [5].

COPD in younger individuals is associated with poor clinical outcomes. In a recent large Danish population study, the prevalence of COPD early in life (defined as age <50 with a ratio of forced expiratory volume in 1s [FEV_1] to forced vital capacity [FVC] < lower limit of normal [LLN] and a cumulative cigarette smoking of 10 pack-years and greater) was estimated to be 15%, and these individuals had more frequent respiratory hospitalizations and increased risk for early death compared with age-adjusted controls [6]. In addition, COPD in young adults does not equate to mild disease, as severe airflow limitation and symptoms can be present [7]. The U.S. Preventive Services Task Force currently recommends against screening for COPD with spirometry in asymptomatic adults. Thus, a risk stratification tool that 1) can identify individuals who are at high risk for developing COPD at young ages and 2) does not depend on early-life

screening spirometry or imaging data, is of great importance for both individualized preventive measures and effective case-finding efforts.

COPD susceptibility is influenced by both environmental and genetic factors [8, 9]. Genetics explains a sizable proportion of phenotypic variance of adulthood lung function and risk of COPD (~30-60% for FEV_1, FEV_1/FVC, and moderate-to-severe COPD) in population-based, cohort, and family-based studies, as well as in families specifically identified through probands with severe, early-onset COPD [10–12]. In genome-wide association studies (GWASs) of lung function, each single nucleotide polymorphism (SNP) accounts for a small amount of phenotypic variance; however, variants can be combined into a polygenic risk score (PRS), which accounts for more phenotypic variability and is predictive of prevalent COPD and lung imaging phenotypes [13–16]. We recently derived a PRS based on lung function that was associated with prevalent COPD, emphysema, and lung function growth and decline trajectories [17]. However, it is unknown if the PRS can inform which patients are at high risk for COPD acquired early in life. Further, the magnitude of polygenic effects across age ranges are non-uniform in other diseases [18], and whether polygenic effects are uniform across age ranges in COPD remains unclear.

We hypothesized that a higher PRS would be associated with an earlier age of diagnosis of COPD and COPD before age of 50 years (COPD50) in an ascertained sample of smokers and a population-based cohort. We examined the effects of the PRS on the risk for incident COPD and examined age-dependent effects of the PRS. We additionally evaluated the predictive performance of the PRS for COPD50 compared to other early-life risk factors.

Materials and Methods

Study design and populations

All studies obtained approval from local institutional review boards and participants provided written informed consent.

COPDGene

We included non-Hispanic white (NHW) and African American (AA) participants from the Genetic Epidemiology of COPD (COPDGene) study, which has been previously described [19]. Briefly, COPDGene is an ongoing multicenter prospective cohort study which enrolled 10,198 NHW and AA participants aged 45-80 years, with a smoking history ≥10 pack-years at baseline, and without severe alpha-1 antitrypsin deficiency. Participants were followed up at 5- (phase 2 visit) and 10-year (phase 3 visit) intervals from baseline visits. Questionnaires for demographics and respiratory health conditions, and spirometry data were obtained at all visits. Further details can be found in the Supplementary Methods.

Framingham Heart Study

To additionally assess the effect of the PRS in a population-based cohort with longitudinal spirometry, we used participants of European ancestry from the Framingham Heart Study (FHS) Offspring cohort and Third Generation cohort. Briefly, FHS was a large longitudinal population-based cohort which has been previously described [20, 21]. Further details can be found in Supplementary Methods.

Age at diagnosis of COPD

At the COPDGene baseline visit, all participants were asked about physician diagnosis of COPD, emphysema, and chronic bronchitis, and for those who answered 'yes' they were further asked about age at diagnosis of those conditions. We considered correctly diagnosed COPD as participants with 1) a self-reported physician diagnosis of COPD and/or emphysema and/or chronic bronchitis and 2) moderate-to-severe airflow limitation (FEV_1/FVC <0.7 and FEV_1 percent predicted <80%) on baseline post-bronchodilator spirometry. For participants with correctly physician-diagnosed COPD and a self-reported age at diagnosis >30, we used self-reported age of diagnosis for COPD, emphysema, or chronic bronchitis, in that order of priority; otherwise, we used age at earliest visit when airflow limitation was identified as the age at diagnosis (supplementary Figures 1 and 2 for NHW and AA participants, respectively). In FHS, questionnaire data about the age of diagnosis was not available. Thus, age of diagnosis was defined using age at earliest exam when moderate-to-severe COPD (modified GOLD grades 2-4 using pre-bronchodilator spirometry) was observed. We defined COPD occurring early in life as an age at diagnosis before 50 years [5, 22].

Predictive variables

As described previously, we developed polygenic risk scores (PRSs) based on results from GWASs of FEV_1 and FEV_1/FVC in the UK Biobank and SpiroMeta Consortium participants of European ancestry [17]. Briefly, we calculated individual PRSs separately for FEV_1 and FEV_1/FVC based on weighted effects of 1.7 and 1.2 million SNPs, respectively. We then calculated a composite PRS as a weighted sum of the two individual PRSs and standardized the composite PRS to facilitate statistical analyses. Details regarding construction of the PRS can be found in the Supplementary Methods. The composite COPD PRS for this study was calculated in

COPDGene and FHS, which are external datasets that were not used in the derivation or tuning of the PRS.

A number of early-life risk factors have been identified for COPD [3, 4]. In COPDGene, we included previously described risk factors available by early adulthood that could potentially be used to guide risk stratification for acquiring COPD early in life. These available risk factors included maternal smoking during pregnancy, childhood asthma, active smoking during adolescence, childhood pneumonia, education, and family history of COPD. Details of definitions of the above risk factors can be found in the Supplementary Methods.

Statistical analysis

Continuous and categorical variables were shown as medians (interquartile ranges [IQRs]) and counts (percentages). Mann-Whitney U tests and Chi-Square tests (Fisher's exact tests when appropriate) were used to examine differences for continuous and categorical variables, respectively.

We first performed a time-to-event analysis to examine the effect of the PRS on incident COPD diagnosis in COPDGene. Age in years was used as the underlying timescale, and time was censored at age of COPD diagnosis or age at last follow-up. We plotted cumulative incidence curves of COPD diagnosis among tertiles (low *vs.* middle *vs.* high) of the PRS using the Kaplan-Meier estimator and tested differences between curves using a log-rank test. We fitted a Cox model and evaluated the proportional hazards (PH) assumption using diagnostics based on Schoenfeld residuals [23]. We additionally fitted Cox models separately for different age intervals (age <50, 50-59, 60-69, and ≥70) to allow time-dependent coefficients (β) of the PRS which was a fixed covariate (see the Supplementary Methods for more details). We tested for an association between the PRS and sex, age and pack-years of smoking at the baseline visit,

and age at smoking initiation to determine whether these variables could be potential confounding factors.

We also performed logistic regression analysis with COPD diagnosed before age 50 (COPD50) as a binary outcome (see the Supplementary Methods for more details). We constructed three prediction models for COPD50 with predictive variables of 1) the PRS, 2) other early-life risk factors (see *Predictive Variables*) that were significantly associated with COPD50 in univariate analysis, and 3) a combination of 1) and 2). We evaluated the discriminatory accuracy of predictive models by comparing area-under-curve (AUC) of receiver-operating-characteristic (ROC) curves using DeLong tests [24].

To examine the effect of the PRS on incident COPD in FHS, we constructed similar models as in COPDGene. First, we built a Cox model in participants without COPD at baseline (offspring cohort exam 5 and generation 3 cohort exam 1), adjusting for baseline age, and then performed regressions across age intervals (age <50, 50-59, 60-69, and ≥70). We also performed linear regression between PRS and age of diagnosis of COPD in participants with COPD, and fitted a logistic regression model to examine the association between the PRS and COPD50.

All models including the PRS were adjusted for principal components of genetic ancestry. Models in FHS additionally accounted for study cohort and familial relatedness using mixed models and generalized estimating equations, as appropriate. We performed analyses using R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) and packages "survival", "survminer", "coxme", "gee" and "pROC". We considered a two-tailed p <0.05 as statistically significant for all tests.

Sensitivity analysis

To assess for potential bias from censoring and using two definitions of COPD diagnosis, we performed a range of sensitivity analyses including stratified analyses and analyses using varying COPD and control definitions in COPDGene NHW participants. See the Supplementary Methods for more details.

Results

Sample characteristics

In COPDGene, 6647 NHW (52.3% males) and 2464 AA (57.6% males) participants were included; 2115 and 444 had physician-diagnosed COPD prior to their baseline visit, and 696 and 309 were newly diagnosed at the time of study enrollment, respectively (Table 1). Compared to those with newly diagnosed COPD, participants with physician-diagnosed COPD were more likely to have GOLD spirometry grades 3 and 4 at baseline (supplemental Figure 3).

We included 6812 FHS participants of European ancestry who had at least two longitudinal spirometry measures from the offspring and generation 3 cohorts (Table 1). COPD diagnoses were ascertained in 811 participants, including 491 at baseline, and 154, 67, 65 and 34 at each follow-up visit, respectively.

Further details regarding COPDGene and FHS baseline characteristics can be found in the Supplementary Results.

PRS and incident COPD diagnosis

In COPDGene, cumulative incidence curves (Kaplan-Meier estimator) for COPD diagnosis stratified by tertiles of the PRS (Figure 1) demonstrate an increased risk for incident COPD diagnosis among individuals with higher PRSs across all ages (log-rank p <0.0001 for difference between curves for both NHW and AA participants). However, we found evidence against the

PH assumption for a Cox model of the PRS on incident COPD across all age ranges (p =0.048 for the scaled Schoenfeld residuals test) in COPDGene NHW participants. Therefore, we fitted a Cox model allowing different coefficients of the PRS within time intervals of ages <50, 50-59, 60-69, and ≥70 (p =0.88 for the PH assumption of the PRS on incident COPD); the hazard ratio (HR) estimates (95% confidence intervals [CIs]) per SD increase in the PRS were 1.51 (1.41-1.63), 1.39 (1.30-1.47), 1.36 (1.28-1.45), and 1.33 (1.21-1.47), respectively (Figure 2). In COPDGene AA participants, one SD increase of the PRS was associated with a 28% (HR 1.28, 95% CI 1.19-1.37) increased hazard for incident COPD across all ages (p =0.37 for testing PH assumption of the PRS on incident COPD).

Similarly, in FHS, we found evidence against the PH assumption for a Cox model examining the effects of the PRS on incident COPD across all age ranges (p =0.005 for the scaled Schoenfeld residuals test). For Cox models allowing different coefficients of the PRS within time intervals of ages <50, 50-59, 60-69, and ≥70 (p =0.45 for the PH assumption); the HR estimates (95% CIs) per SD increase in the PRS were 2.32 (1.81-2.97), 1.70 (1.41-2.04), 1.37 (1.11-1.67), and 1.28 (1.01-1.61), respectively (Figure 2).

In COPDGene, we did not detect a differential effect of the PRS by sex (p =0.07 and 0.91 for the interaction term for COPDGene NHW and AA participants, respectively). In a sex-stratified analysis in NHW participants, one SD increase of the PRS was associated with a 36% (HR 1.36, 95% CI 1.30-1.43) and a 44% (HR 1.44, 95% CI 1.37-1.52) increased hazard for incident COPD in male and female participants, respectively. We found no association between the PRS and sex, age, pack-years of smoking at baseline visit, or age at initiation of smoking. Results did not change significantly with further adjustment of these factors in the model.

Prediction for COPD diagnosed early in life

In COPDGene, COPD50 diagnoses were observed for 491 NHW and 194 AA participants. These individuals with COPD50 had higher PRSs and were more likely to have early-life risk factors compared to those without COPD50 (Table 2). The PRS was associated with an increased risk for COPD50 in univariable (odds ratio [OR] [95% CI] per SD: 1.60 [1.46-1.76] for NHW and 1.23 [1.06-1.43] for AA participants) and multivariable models including other early-life risk factors (OR [95% CI] per SD: 1.55 [1.41-1.71] for NHW and 1.23 [1.05-1.43] for AA participants) (Table 3 and Table 4). See the Supplementary Results for more details. In FHS, COPD50 diagnoses were observed in 186 participants, and one SD increase of the PRS was associated with an increased odds of COPD50 (OR 2.47, 95% CI 2.12-2.88, p <0.0001). In FHS participants with COPD, one SD increment of the PRS was associated with a 1.52-year (95% CI: 0.84-2.19, p <0.0001) earlier COPD diagnosis.

In COPDGene, ROC curves of the three prediction models for COPD50 (model 1: PRS; model 2: other early-life risk factors; model 3: PRS and other early-life risk factors) are shown in Figure 3, and the AUCs (95% CIs) were 0.659 (0.636-0.683), 0.692 (0.668-0.716), and 0.739 (0.718-0.761) for NHW participants and were 0.571 (0.530-0.612), 0.609 (0.572-0.646), and 0.635 (0.595-0.675) for AA participants, respectively. No significant difference was found between AUCs of model 1 and model 2 for NHW (p =0.055) or AA (p =0.17) participants. There were significant differences between AUCs of model 2 and model 3 for both NHW (p <0.0001) and AA (p = 0.043) participants.

Sensitivity analysis

All sensitivity analyses were performed in COPDGene NHW participants. In participants who had COPD at the baseline visit (n=2811), one SD increase in the PRS was associated with a 0.95-year (95% CI: 0.59-1.30, p <0.0001) earlier diagnosis of COPD. Compared with

participants with the lowest PRS tertile, those who were of the highest PRS tertile had a 1.89-year (95% CI: 1.01-2.78, p <0.0001) earlier in diagnosis. One SD increase of the PRS was associated with a 0.93-year (95% CI: 0.52-1.33, p <0.0001) and a 1.07-year (95% CI: 0.41-1.72, p =0.0016) earlier COPD diagnosis in participants with physician-diagnosed COPD and newly diagnosed COPD at the baseline visit, respectively. In participants who were ≥50 years old at the baseline visit (n=5929), COPD50 was reported in 353 participants. As shown in Table S1, the association between the PRS and COPD50 in univariable and multivariable models which included other early-life risk factors were similar compared to results of Table 3. No significant difference was found between AUCs of model 1 and model 2 (AUC [95% CI] 0.654 [0.626-0.682] *vs.* 0.692 [0.664-0.720], p =0.060). There was a significant difference between AUCs of model 2 and model 3 (AUC [95% CI] 0.692 [0.664-0.720] *vs.* 0.736 [0.711-0.762], p <0.0001). Additional sensitivity analysis results can be found in the Supplementary Results.

Discussion

In two cohorts – one of non-Hispanic white (NHW) and African-American (AA) smokers with and without COPD, predominately using self-reported age of diagnosis, and a second cohort, a population-based study relying on longitudinal lung function – we found that a polygenic risk score (PRS) for COPD is associated with an increased risk for earlier age of COPD diagnosis. In addition, we found an age-dependent effect of the PRS on risk for COPD, observing larger effect estimates at younger ages (negative age-dependency of the PRS).

Different lung function trajectories distinguished by combinations of failure to attain maximal lung function and/or accelerated decline of lung function may lead to airflow obstruction [25, 26]. Derived from more than a million SNPs from a large GWAS of lung

function in population-based studies, our PRS could theoretically represent the cumulative effect of common genetic variants associated with growth, plateau, and/or early decline of lung function. The genetic variants included in the PRS may affect gene regulation in fetal lung [13]. We previously demonstrated that the PRS was associated with reduced lung growth in children with asthma; these children with reduced lung growth developed spirometric obstruction early in adulthood [17]. Thus, it is possible that the observed age-dependent effects of the PRS are driven by variants important for lung growth and development. In addition, the PRS may be associated with genetic susceptibility to environmental exposure-induced injury, and consequently the rate of early lung function decline. For example, genetic factors may modify the effect of cigarette smoking on development of COPD [27–30].

Our findings may also be consistent with the increased relative effects of cigarette smoking in later age. While this question was not directly addressed in our study, it has been shown that heritability of lung function decreases with age in UK Biobank [31]. The underlying mechanisms of the time-varying effect of the PRS warrant further research.

COPD occurring at earlier ages is often underdiagnosed, and underdiagnosis is associated with unfavorable clinical outcomes [5, 32]. To address this issue, an international panel of experts defined early COPD as occurring in individuals aged <50 years with at least 10 pack-years cigarette smoking, with either abnormal lung function (FEV_1/FVC <LLN or accelerated FEV_1 decline) or abnormal chest imaging findings (visual emphysema, air trapping, or bronchial thickening) [22]. In the present study, we defined COPD occurring early in life using age, smoking exposure, and moderate-to-severe spirometric criteria for obstruction. Additional studies examining the association of the PRS with the development of destructive emphysema

and airway pathology in younger cohorts may help elucidate the specific phenotype identified by our PRS.

While a previous study suggested that the PRS could be used for early detection and prevention of COPD [17], the present study offers direct evidence that the PRS can predict COPD occurring in younger age groups. Under the current case-finding strategy for COPD, respiratory symptoms are essential for physicians to suspect a diagnosis of COPD. However, a substantial proportion of patients are underdiagnosed with already present or under-reported respiratory symptoms and have increased risk for respiratory hospitalizations and mortality compared to people without obstruction [32–34]. Thus, the current findings suggest that genetics offer a critical tool to identify young people at high risk for COPD occurring early in life. Knowledge of an individual's PRS may enable physicians to make targeted inquiries about patients' respiratory symptoms and prioritize pulmonary function testing for patients with symptoms. Also, a known unfavorable PRS might further motivate an individual to be compliant with medical interventions [35], although the behavioral and psychosocial reactions to the awareness of an individual's own genetic risk of COPD should be carefully studied prior to clinical application of the PRS.

Many early-life risk factors for COPD have been identified, including maternal smoking during pregnancy [36–38], childhood asthma [39–42], active smoking during adolescence [43], childhood pneumonia [44], socioeconomic status [45–47], and family history [48, 49]. The odds ratio of the PRS was comparable with other early-life risk factors (Table 3), but the PRS offers several distinct advantages compared to these clinical variables. First, the PRS can be determined at birth and earlier than most of the other risk factors, and interventions targeting modifiable early-life risk factors for COPD can be implemented. Second, the PRS is a continuous predictor;

as we have previously demonstrated, identifying those at the highest and lowest predicted risk as would likely be done in clinical implementation yields larger effect sizes. Third, the AUC estimates demonstrate that the PRS alone had a comparable performance for predicting COPD occurring early in life compared to a combination of early-life risk factors, and the addition of the PRS to early-life risk factors significantly improved the predictive performance. Notably, the PRS was derived from external cohorts, whereas early-life risk factors were modeled and tested in the same cohort, which might have resulted in an underestimated additive value of the PRS. However, early-life risk factors were self-reported and may suffer from recall bias, and the predictive performance of early-life risk factors may have been under- or over-estimated. The association of the PRS with COPD early in life was amplified when excluding participants with an age of COPD diagnosis ≥50 from controls, suggesting that the observed association of the PRS with COPD early in life could be a conservative estimate.

While other early-life risk factors for COPD were not available for this study (e.g., low birth weight) [50], detailed knowledge of a person's early-life risk factors is often unknown or difficult to measure accurately (e.g., cumulative exposure to air pollution). By contrast, a PRS can be calculated using genome-wide genotyping which can be done once in an individual's lifetime, is of low cost, and potentially relevant for a large number of diseases. In addition, the performance of the PRS will likely improve with future genetic association studies.

Strengths of this study include a large sample of well-phenotyped smokers with post-bronchodilator spirometry data, which is essential to confirm irreversible airflow limitation and to differentiate COPD from asthma, especially in young participants. In addition, we were able to include a relatively extensive panel of known early-life risk factors to compare with the PRS for

the predictive performance of COPD occurring early in life. We were also able to replicate the findings in a population-based cohort with both smokers and non-smokers.

This study has several limitations. We acknowledge a potential measurement bias introduced from using two different COPD outcome definitions (physician-diagnosed vs. spirometry-defined); however, sensitivity analyses suggest that our findings are robust because 1) the linear association between the PRS and age of COPD diagnosis is consistent amongst participants with physician-diagnosed and spirometry-defined COPD, 2) removing participants with baseline COPD (i.e. those most likely to be physician-diagnosed) did not attenuate the association of the PRS with incident COPD, and 3) stratified analyses within individuals with physician-diagnosed COPD demonstrated similar results. In addition, our finding were further supported by FHS results which used only spirometry-defined COPD. We used age at diagnosis as a proxy of age at onset of COPD. We would expect a large variability in the interval between the two time points since the severity of COPD at the time of diagnosis is highly variable. This issue could result in misclassification bias of the outcome variable. However, our sensitivity analyses did not reveal an association between the PRS and either timing of physician diagnosis in the course of clinical COPD or participants' ages at baseline visit. Therefore, misclassifications would generally bias toward the null, and are unlikely to account the observed results.

Our case definition was for moderate-to-severe COPD (GOLD 2-4), and thus we included GOLD 1 and preserved ratio impaired spirometry (PRISm) subjects as controls in our calculations of incidence of moderate-to-severe disease. Having demonstrated age-dependent PRS effects, derivation and testing of PRSs in younger populations is needed. Findings of African American (AA) smokers were of the same direction of association compared with those

of NHW participants in COPDGene, although with appreciably smaller effect size estimates. This ~50% reduction in the magnitude of the regression coefficients (β) parallels the reduced predictive performance observed when applying the European-derived PRS for COPD to African Americans (OR [95% CI] per SD: 1.50 [1.37-1.65] for AA and 2.20 [2.03-2.37] for NHW participants). In addition to underrepresentation of non-European ancestry individuals in previous GWASs, the difference in the case-control ratio (milder disease in African Americans in COPDGene) and access to healthcare might also contribute the different performance of the PRS on age of diagnosis of COPD. These results highlight the need to improve multi-ancestry polygenic prediction.

In conclusion, in a large sample of smokers and a general population cohort, a higher COPD PRS is associated with an increased risk for incident COPD and the effect of the PRS is age-dependent and larger at younger ages. A higher PRS is associated with an earlier age of diagnosis of COPD. The PRS adds substantial value to other early-life risk factors in prediction for COPD occurring early in life.

Reference

1. Rabe KF, Watz H. Chronic obstructive pulmonary disease. *Lancet* 2017; 389: 1931–1940.
2. Lõpez-Campos JL, Tan W, Soriano JB. Global burden of COPD. *Respirology* 2016; 21: 14–23.
3. Postma DS, Bush A, Van Den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *Lancet* Elsevier Ltd; 2015; 385: 899–909.
4. Martinez FD. Early-Life Origins of Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* 2016; 375: 871–878.
5. Soriano JB, Polverino F, Cosio BG. What is early COPD and why is it important? *Eur. Respir. J.* 2018; 52.
6. Çolak Y, Afzal S, Nordestgaard BG, Vestbo J, Lange P. Prevalence, characteristics, and prognosis of early chronic obstructive pulmonary disease the Copenhagen general population study. *Am. J. Respir. Crit. Care Med.* 2020; 201: 671–680.
7. Hersh CP, DeMeo DL, Al-Ansari E, Carey VJ, Reilly JJ, Ginns LC, Silverman EK. Predictors of survival in severe, early onset COPD. *Chest* 2004; 126: 1443–1451.
8. Silverman EK. Genetics of COPD. *Annu. Rev. Physiol.* 2020; 82: 413–431.
9. Lange P, Ahmed E, Lahmar ZM, Martinez FJ, Bourdin A. Natural history and mechanisms of COPD. *Respirology* 2021; .
10. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am. J. Respir. Crit. Care Med.* 2013; 188: 941–947.
11. Klimentidis YC, Vazquez AI, de los Campos G, Allison DB, Dransfield MT, Thannickal VJ. Heritability of pulmonary function estimated from pedigree and whole-genome markers. *Front. Genet.* 2013; 4.
12. Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, O'Donnell WJ, Reilly JJ, Ginns L, Mentzer S, Wain J, Speizer FE. Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: Risk to relatives for airflow obstruction and chronic bronchitis. *Am. J. Respir. Crit. Care Med.* 1998; 157: 1770–1778.
13. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, Shrine N, Qiao D, Bartz TM, Kim DK, Lee MK, Latourelle JC, Li X, Morrow JD, Obeidat M, Wyss AB, Bakke P, Barr RG, Beaty TH, Belinsky SA, Brusselle GG, Crapo JD, de Jong K, DeMeo DL, Fingerlin TE, Gharib SA, Gulsvik A, Hall IP, Hokanson JE, Kim WJ, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat. Genet.* 2019; 51: 494–505.
14. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat M, Zhao JH, Wielscher M, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* 2019; 51: 481–493.

15. Wain L V., Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, Obeidat M, Henry AP, Portelli MA, Hall RJ, Billington CK, Rimington TL, Fenech AG, John C, Blake T, Jackson VE, Allen RJ, Prins BP, Campbell A, Porteous DJ, Jarvelin MR, Wielscher M, James AL, Hui J, Wareham NJ, Zhao JH, Wilson JF, Joshi PK, Stubbe B, Rawal R, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.* 2017; 49: 416–425.

16. Oelsner EC, Ortega VE, Smith BM, Nguyen JN, Manichaikul AW, Hoffman EA, Guo X, Taylor KD, Woodruff PG, Couper DJ, Hansel NN, Martinez FJ, Paine R, Han MK, Cooper C, Dransfield MT, Criner G, Krishnan JA, Bowler R, Bleecker ER, Peters S, Rich SS, Meyers DA, Rotter JI, Graham Barr R. A genetic risk score associated with chronic obstructive pulmonary disease susceptibility and lung structure on computed tomography. *Am. J. Respir. Crit. Care Med.* 2019; 200: 721–731.

17. Moll M, Sakornsakolpat P, Shrine N, Hobbs BD, DeMeo DL, John C, Guyatt AL, McGeachie MJ, Gharib SA, Obeidat M, Lahousse L, Wijnant SRA, Brusselle G, Meyers DA, Bleecker ER, Li X, Tal-Singer R, Manichaikul A, Rich SS, Won S, Kim WJ, Do AR, Washko GR, Barr RG, Psaty BM, Bartz TM, Hansel NN, Barnes K, Hokanson JE, Crapo JD, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir. Med.* 2020; 8: 696–708.

18. Archambault AN, Su YR, Jeon J, Thomas M, Lin Y, Conti D V., Win AK, Sakoda LC, Lansdorp-Vogelaar I, Peterse EFP, Zauber AG, Duggan D, Holowatyj AN, Huyghe JR, Brenner H, Cotterchio M, Bézieau S, Schmit SL, Edlund CK, Southey MC, MacInnis RJ, Campbell PT, Chang-Claude J, Slattery ML, Chan AT, Joshi AD, Song M, Cao Y, Woods MO, White E, et al. Cumulative Burden of Colorectal Cancer–Associated Genetic Variants Is More Strongly Associated With Early-Onset vs Late-Onset Cancer. *Gastroenterology* 2020; 158: 1274-1286.e12.

19. 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.* 2010; 7: 32–43.

20. Kannel WB, Feinleib M, Mcnamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: The framingham offspring study. *Am. J. Epidemiol.* 1979; 110: 281–290.

21. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, recruitment, and initial examination. *Am. J. Epidemiol.* 2007; 165: 1328–1335.

22. Martinez FJ, Han MK, Allinson JP, Graham Barr R, Boucher RC, Calverley PMA, Celli BR, Christenson SA, Crystal RG, Fageras M, Freeman CM, Groenke L, Hoffman EA, Kesimer M, Kostikas K, Paine R, Rafii S, Rennard SI, Segal LN, Shaykhiev R, Stevenson C, Tal-Singer R, Vestbo J, Woodruff GP, Curtis JL, Wedzicha JA. At the root: Defining and halting progression of early chronic obstructive pulmonary disease. *Am. J. Respir.*

Crit. Care Med. 2018; 197: 1540–1551.

23. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81: 515–526.

24. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* 1988; 44: 837.

25. Lange P, Celli B, Agustí A, Boje Jensen G, Divo M, Faner R, Guerra S, Marott JL, Martinez FD, Martinez-Camblor P, Meek P, Owen CA, Petersen H, Pinto-Plata V, Schnohr P, Sood A, Soriano JB, Tesfaigzi Y, Vestbo J. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* 2015; 373: 111–122.

26. Agusti A, Faner R. Lung function trajectories in health and disease. *Lancet Respir. Med.* 2019; 7: 358–364.

27. Rangasamy T, Cho CY, Thimmulappa RK, Zhen L, Srisuma SS, Kensler TW, Yamamoto M, Petrache I, Tuder RM, Biswal S. Genetic ablation of Nrf2 enhances susceptibility to cigarette smoke–induced emphysema in mice. *J. Clin. Invest.* 2004; 114: 1248–1259.

28. DeMeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, Rennard SI, Sandhaus RA, Stocks JM, Stoller JK, Strange C, Turino G, Silverman EK. IL10 polymorphisms are associated with airflow obstruction in severe α1-antitrypsin deficiency. *Am. J. Respir. Cell Mol. Biol.* 2008; 38: 114–120.

29. Molloy K, Hersh CP, Morris VB, Carroll TP, O'Connor CA, Lasky-Su JA, Greene CM, O'Neill SJ, Silverman EK, McElvaney NG. Clarification of the Risk of Chronic Obstructive Pulmonary Disease in α1-Antitrypsin Deficiency PiMZ Heterozygotes. *Am. J. Respir. Crit. Care Med.* 2014; 189: 419–427.

30. Kim W, Moll M, Qiao D, Hobbs BD, Shrine N, Sakornsakolpat P, Tobin MD, Dudbridge F, Wain L V, Ladd-Acosta C, Chatterjee N, Silverman EK, Cho MH, Beaty TH. Smoking Interaction with a Polygenic Risk Score for Reduced Lung Function. *medRxiv* 2021; : 2021.03.26.21254415.

31. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet.* 2017; 13: 070177.

32. Diab N, Gershon AS, Sin DD, Tan WC, Bourbeau J, Boulet LP, Aaron SD. Underdiagnosis and overdiagnosis of chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care Med. 2018.

33. Martinez CH, Mannino DM, Jaimes FA, Curtis JL, Han MLK, Hansel NN, Diaz AA. Undiagnosed obstructive lung disease in the United States associated factors and long-term mortality. *Ann. Am. Thorac. Soc.* 2015; 12: 1788–1795.

34. Gershon AS, Hwee J, Chapman KR, Aaron SD, O'Donnell DE, Stanbrook MB, Bourbeau J, Tan W, Su J, Victor JC, To T. Factors associated with undiagnosed and overdiagnosed COPD. *Eur. Respir. J.* 2016; 48: 561–564.

35. Carpenter MJ, Strange C, Jones Y, Dickson MM, Carter C, Moseley MA, Gilbert GE. Does genetic testing result in behavioral health change? Changes in smoking behavior following testing for alpha-1 antitrypsin deficiency. *Ann. Behav. Med.* 2007; 33: 22–28.

36. Hanrahan JP, Tager IB, Segal MR, Tosteson TD, Castile RG, Van Vunakis H, Weiss ST, Speizer FE. The effect of maternal smoking during pregnancy on early infant lung function. *Am. Rev. Respir. Dis.* 1992; 145: 1129–1135.

37. Gilliland FD, Berhane K, McConnell R, Gauderman WJ, Vora H, Rappaport EB, Avol E, Peters JM. Maternal smoking during pregnancy, environmental tobacco smoke exposure and childhood lung function. *Thorax* 2000; 55: 271–276.

38. Hayatbakhsh MR, Sadasivam S, Mamun AA, Najman JM, Williams GM, O'Callaghan MJ. Maternal smoking during and after pregnancy and lung function in early adulthood: A prospective study. *Thorax* 2009; 64: 810–814.

39. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, Cho MH, Croteau-Chonka DC, Castaldi PJ, Jain G, Sanyal A, Zhan Y, Lajoie BR, Dekker J, Stamatoyannopoulos J, Covar RA, Zeiger RS, Adkinson NF, Williams P V., Kelly HW, Grasemann H, Vonk JM, Koppelman GH, Postma DS, Raby BA, Houston I, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N. Engl. J. Med.* 2016; 374: 1842–1852.

40. Sears MR, Greene JM, Willan AR, Wiecek EM, Taylor DR, Flannery EM, Cowan JO, Herbison GP, Silva PA, Poulton R. A Longitudinal, Population-Based, Cohort Study of Childhood Asthma Followed to Adulthood. *N. Engl. J. Med.* 2003; 349: 1414–1422.

41. Morgan WJ, Stern DA, Sherrill DL, Guerra S, Holberg CJ, Guilbert TW, Taussig LM, Wright AL, Martinez FD. Outcome of asthma and wheezing in the first 6 years of life follow-up through adolescence. *Am. J. Respir. Crit. Care Med.* 2005; 172: 1253–1258.

42. James AL, Palmer LJ, Kick E, Maxwell PS, Lagan SE, Ryan GF, Musk AW. Decline in lung function in the Busselton health study: The effects of asthma and cigarette smoking. *Am. J. Respir. Crit. Care Med.* 2005; 171: 109–114.

43. Gold DR, Wang X, Wypij D, Speizer FE, Ware JH, Dockery DW. Effects of Cigarette Smoking on Lung Function in Adolescent Boys and Girls. *N. Engl. J. Med.* 1996; 335: 931–937.

44. Hayden LP, Hobbs BD, Cohen RT, Wise RA, Checkley W, Crapo JD, Hersh CP. Childhood pneumonia increases risk for chronic obstructive pulmonary disease: The COPDGene study. *Respir. Res.* 2015; 16.

45. Prescott E, Lange P, Vestbo J. Socioeconomic status, lung function and admission to hospital for COPD: Results from the Copenhagen City Heart Study. *Eur. Respir. J.* 1999; 13: 1109–1114.

46. Yin P, Zhang M, Li Y, Jiang Y, Zhao W. Prevalence of COPD and its association with socioeconomic status in China: Findings from China Chronic Disease Risk Factor Surveillance 2007. *BMC Public Health* 2011; 11.

47. Kanervisto M, Vasankari T, Laitinen T, Heliövaara M, Jousilahti P, Saarelainen S. Low socioeconomic status is associated with chronic obstructive airway diseases. *Respir. Med.* 2011; 105: 1140–1146.

48. Hersh CP, Hokanson JE, Lynch DA, Washko GR, Make BJ, Crapo JD, Silverman EK. Family history is a risk factor for COPD. *Chest* 2011; 140: 343–350.

49. Moll M, Lutz SM, Ghosh AJ, Sakornsakolpat P, Hersh CP, Beaty TH, Dudbridge F, Tobin MD, Mittleman MA, Silverman EK, Hobbs BD, Cho MH. Relative contributions of family history and a polygenic risk score on COPD and related outcomes: COPDGene and ECLIPSE studies. *BMJ Open Respir. Res.* 2020; 7.

50. Baumann S, Godtfredsen NS, Lange P, Pisinger C. The impact of birth weight on the level of lung function and lung function decline in the general adult population. the Inter99 study. *Respir. Med.* 2015; 109: 1293–1299.

Table 1. Sample characteristics by status of COPD diagnosis at baseline visit. COPDGene participants with previously diagnosed COPD reported a physician diagnosis of COPD prior to the baseline visit.

	COPDGene NHW participants			COPDGene AA participants			FHS participants	
	Previously diagnosed (n=2115)	Newly diagnosed (n=696)	No COPD* (n=3836)	Previously diagnosed (n=444)	Newly diagnosed (n=309)	No COPD* (n=1711)	COPD* (n=491)	No COPD* (n=6321)
Polygenic risk score	0.22 (1.34)	0.22 (1.32)	-0.17 (1.30)	0.18 (1.30)	0.27 (1.37)	-0.10 (1.33)	0.66 (1.34)	-0.08 (1.31)
Age at baseline visit	65.8 (11.6)	62.8 (12.8)	60.0 (13.7)	60.6 (12.1)	55.4 (10.9)	51.9 (7.8)	57.0 (17.0)	46.0 (16.0)
Sex (male)	1176 (55.6%)	388 (55.7%)	1913 (49.9%)	228 (51.4%)	184 (59.5%)	1008 (58.9%)	257 (52.3%)	2940 (46.5%)
Smoking status (ever-smokers)**	100%	100%	100%	100%	100%	100%	384 (78.2%)	3240 (51.3%)
Pack-years of smoking	51.3 (34.5)	45.0 (28.1)	37.0 (25.0)	38.3 (26.2)	37.3 (26.4)	33.3 (21.3)	35.5 (32.1)	12 (21.5)
Age at smoking initiation	16.0 (4.0)	16.0 (4.0)	16.0 (4.0)	16.0 (4.0)	16.0 (5.0)	16.0 (5.0)	18.0 (3.0)	17.0 (3.0)
Baseline FEV_1 % predicted	44.3 (27.1)	66.4 (18.9)	90.5 (18.3)	44.5 (27.2)	63.4 (21.5)	96.0 (17.2)	69.9 (15.1)	98.5 (16.1)

Continuous variables were shown as medians (interquartile ranges). Categorical variables were shown as numbers (percentages). NHW: non-Hispanic white; AA: African American, FHS: Framingham Heart Study; FEV_1: forced expiratory volume in 1s.

* COPD: GOLD spirometry grades 2-4

** All COPDGene participants had a smoking history of 10 pack-years or greater

Table 2. Distribution of risk factors for COPD occurring early in life (age < 50; COPD50) in COPDGene non-Hispanic white and African American participants.

Predictive variable	COPD50					
	Non-Hispanic white participants			African American participants		
	Yes (n=491)	No (n=5870)	P value	Yes (n=194)	No (n=1974)	P value
Polygenic risk score	0.45 (1.27)	-0.058 (1.32)	<0.0001	0.25 (1.27)	-0.018 (1.37)	0.0058
Sex (male)	252 (51.3%)	3062 (52.2%)	0.76	103 (53.1%)	1129 (57.2%)	0.31
Maternal smoking during pregnancy	187 (38.1%)	1457 (24.8%)	<0.0001	37 (19.1%)	294 (14.9%)	0.15
Active adolescent smoking	425 (86.6%)	4365 (74.4%)	<0.0001	161 (83.0%)	1409 (71.4%)	0.0008
Childhood asthma	58 (11.8%)	291 (5.0%)	<0.0001	33 (17.0%)	157 (8.0%)	<0.0001
Childhood pneumonia	60 (12.2%)	589 (10.0%)	0.14	15 (7.7%)	88 (4.5%)	0.062
Family history of COPD	248 (50.5%)	1799 (30.6%)	<0.0001	44 (22.7%)	304 (15.4%)	0.011
Education			<0.0001			0.11
High school or below	224 (45.6%)	1652 (28.1%)		119 (61.3%)	1054 (53.4%)	
College	243 (49.5%)	3439 (58.6%)		72 (37.1%)	880 (44.6%)	
Graduate	24 (4.9%)	779 (13.3%)		3 (1.5%)	40 (2.0%)	

Continuous variables were shown as median (interquartile range). Categorical variables were shown as count (percentage).

Table 3. Associations between the PRS and other early-life risk factors and COPD occurring early in life (age < 50) in COPDGene non-Hispanic white participants.

Predictive variable	OR (95% CI)		
	model 1	model 2	model 3
Polygenic risk score *	1.60 (1.46-1.76)		1.55 (1.41-1.71)
Maternal smoking during pregnancy		1.48 (1.21-1.81)	1.51 (1.23-1.86)
Active smoking during adolescence		1.86 (1.42-2.43)	1.92 (1.46-2.52)
Childhood asthma		2.34 (1.72-3.19)	2.07 (1.51-2.83)
Family history of COPD		1.99 (1.64-2.41)	1.86 (1.52-2.26)
Education			
High school or below (reference)		1.00	1.00
College		0.53 (0.44-0.64)	0.55 (0.45-0.67)
Graduate		0.25 (0.16-0.39)	0.28 (0.18-0.44)

PRS: polygenic risk score; OR: odds ratio; CI: confidence interval.
model 1: PRS + principal components of genetic ancestry; model 2: other early-life risk factors; model 3: PRS + other early-life risk factors + principal components of genetic ancestry
* ORs were calculated for per standard deviation increase of the PRS
All associations were statistically significant with p <0.001

Table 4. Associations between the PRS and other early-life risk factors and COPD occurring early in life (age < 50) diagnosis in COPDGene African American participants.

Predictive variable	OR (95% CI)		
	model 1	model 2	model 3
Polygenic risk score *	1.23 (1.06-1.43)		1.23 (1.05-1.43)
Active smoking during adolescence		1.91 (1.29-2.81)	1.90 (1.29-2.80)
Childhood asthma		2.32 (1.54-3.50)	2.28 (1.51-3.45)
Family history of COPD		1.58 (1.10-2.26)	1.57 (1.09-2.25)

PRS: polygenic risk score; OR: odds ratio; CI: confidence interval.
model 1: PRS + principal components of genetic ancestry; model 2: other early-life risk factors; model 3: PRS + other early-life risk factors + principal components of genetic ancestry
* ORs were calculated for per standard deviation increase of the PRS
All associations were statistically significant with p <0.05

Table S1. Associations between the PRS and other early-life risk factors and COPD occurring early in life (age < 50) in COPDGene non-Hispanic white participants aged ≥ 50 at the baseline visit.

Predictive variable	OR (95% CI)		
	model 1	model 2	model 3
Polygenic risk score *	1.55 (1.39-1.72)		1.49 (1.34-1.67)
Maternal smoking during pregnancy		1.46 (1.15-1.84) $	1.50 (1.18-1.90)
Active smoking during adolescence		1.87 (1.37-2.56)	1.93 (1.41-2.64)
Childhood asthma		2.76 (1.98-3.86)	2.46 (1.75-3.46)
Family history of COPD		2.10 (1.67-2.63)	1.96 (1.56-2.46)
Education			
High school or below (reference)		1.00	1.00
College		0.60 (0.48-0.76)	0.63 (0.50-0.79)
Graduate		0.33 (0.20-0.52)	0.37 (0.23-0.59)

PRS: polygenic risk score; OR: odds ratio; CI: confidence interval.
model 1: PRS + principal components of genetic ancestry; model 2: other early-life risk factors; model 3: PRS + other early-life risk factors + principal components of genetic ancestry
* ORs were calculated for per standard deviation increase of PRS
$ p =0.002
All other associations were statistically significant with p <0.001

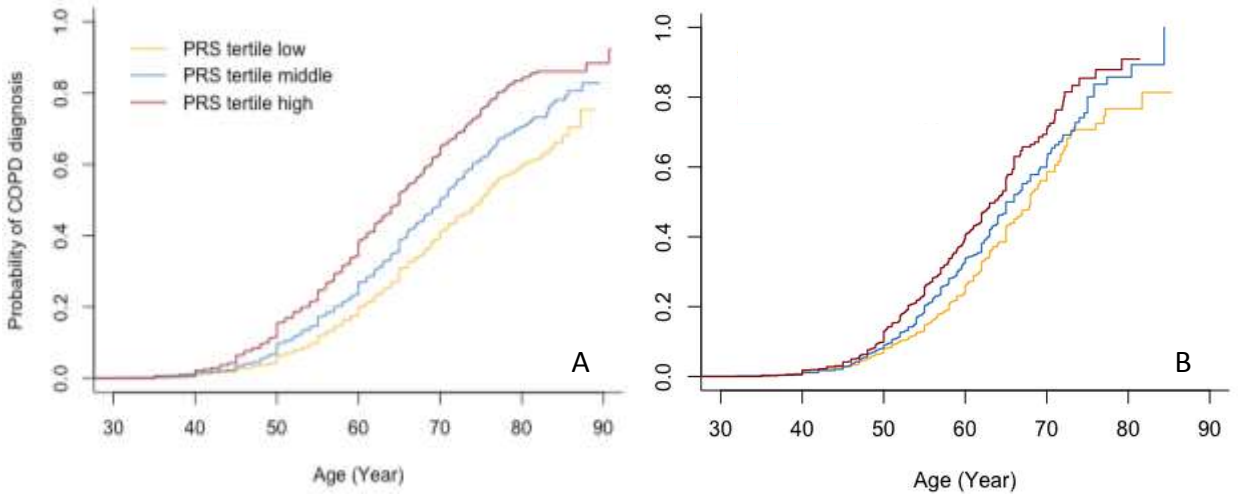

Figure 1. Cumulative incidence curves of COPD diagnosis by tertiles of polygenic risk scores in COPDGene non-Hispanic white (A) and African American (B) participants.

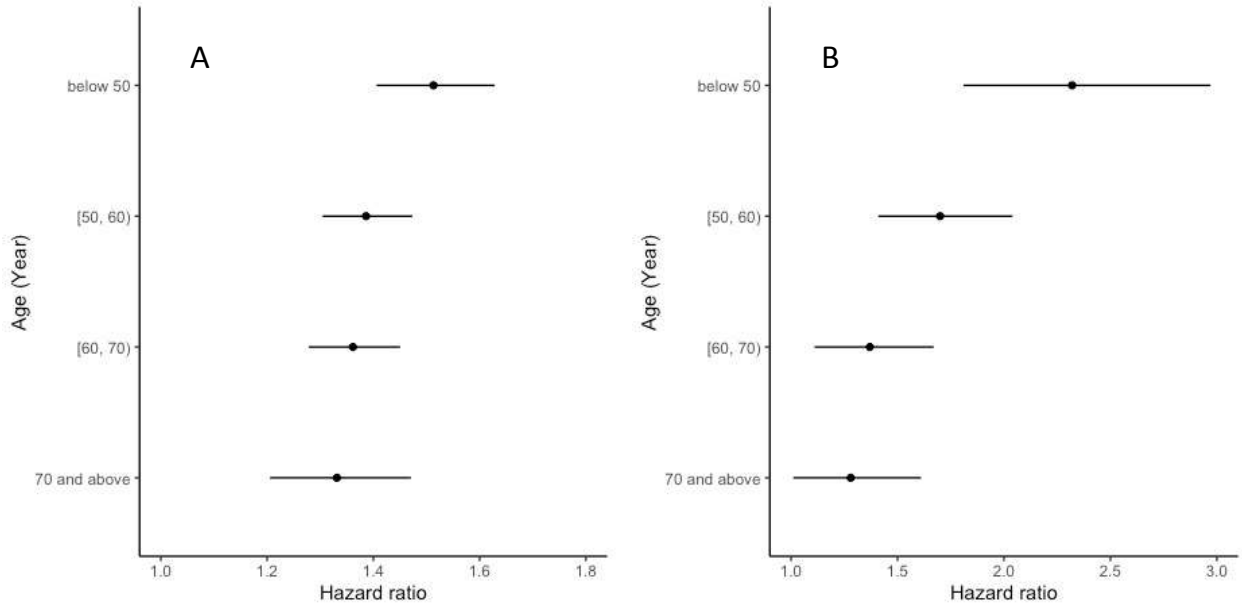

Figure 2. Risk estimates (points: hazard ratios, bars: 95% confidence intervals) associated with the polygenic risk score (per standard deviation) for incident COPD among age intervals of below 50, 50-60, 60-70, and 70 years and beyond in COPDGene non-Hispanic white (A) and Framingham Heart Study (B) participants.

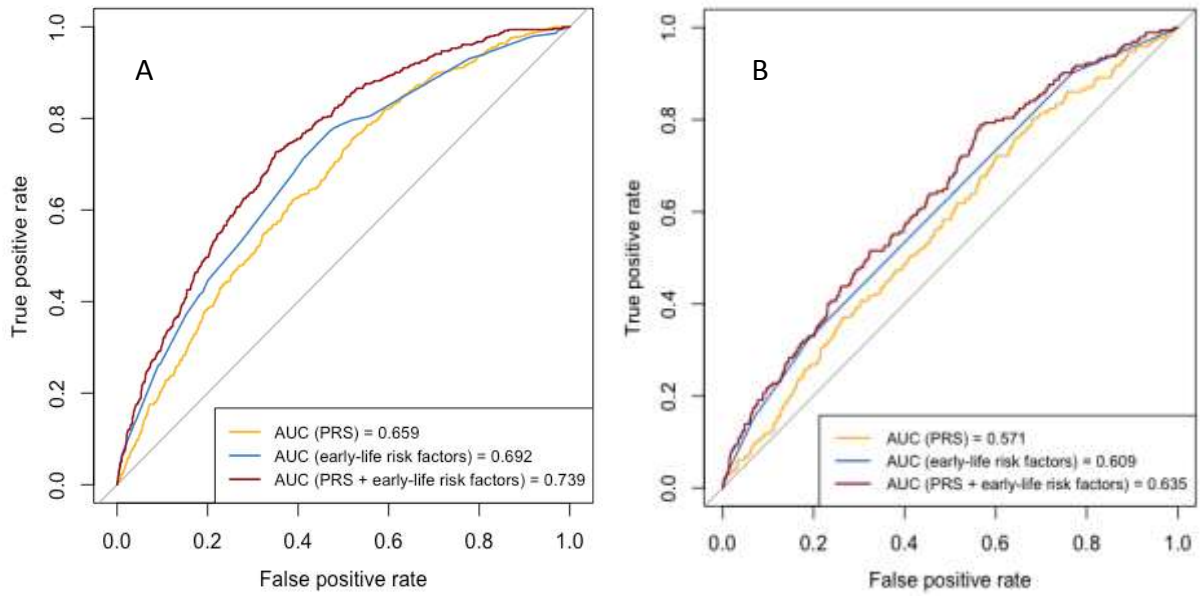

Figure 3. Receiver operating characteristic curves of prediction models for COPD occurring early in life (age < 50) with predictive variables of 1) PRS, 2) early-life risk factors, and 3) PRS and early-life risk factors in COPDGene non-Hispanic white (A) and African American (B) participants. Models with PRS were adjusted for principal components of genetic ancestry. AUC: area under curve, PRS: polygenic risk score.

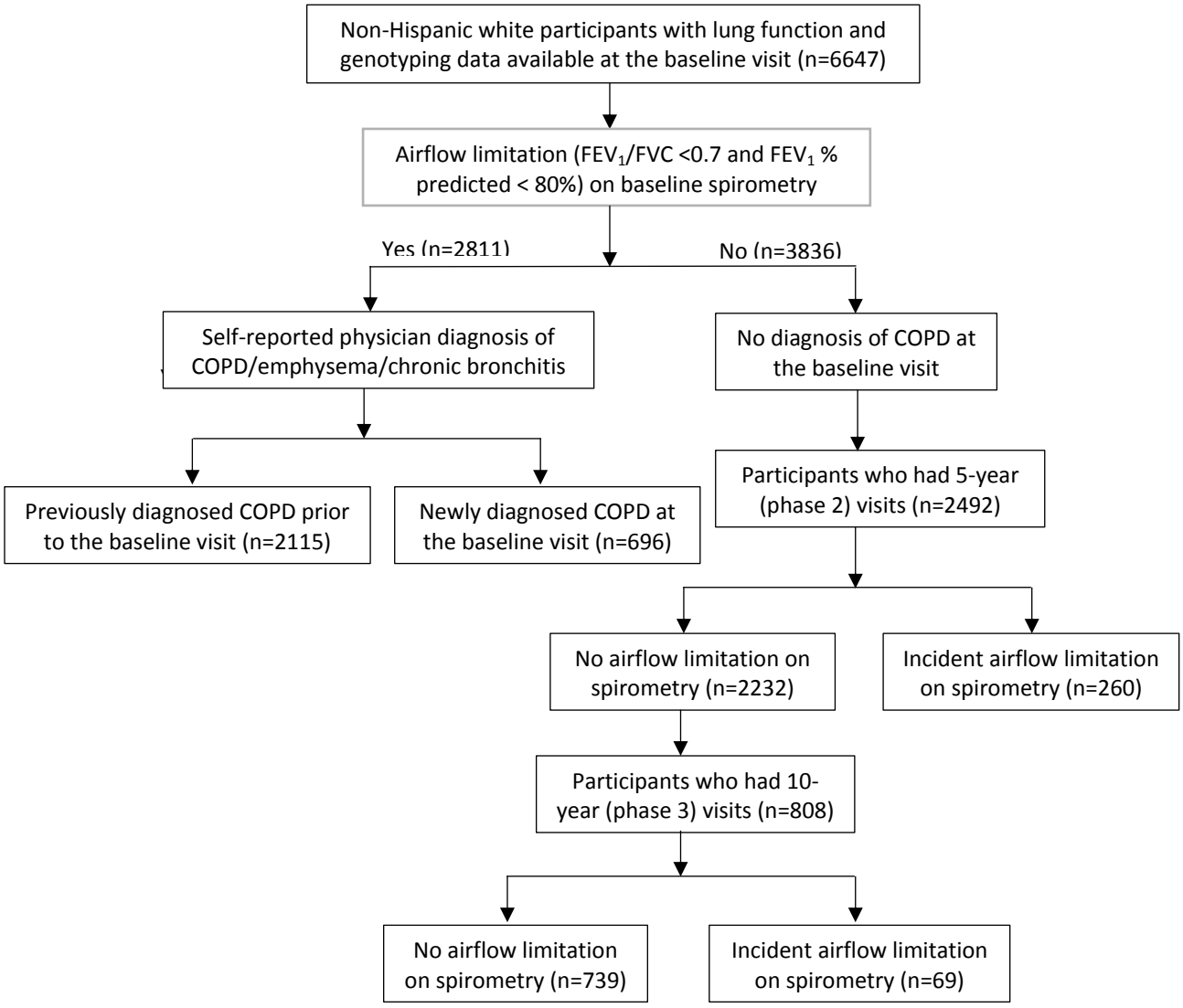

Figure S1. Flowchart of diagnosis of moderate-to-severe COPD (GOLD 2-4) at baseline and follow-up visits for COPDGene non-Hispanic white participants.

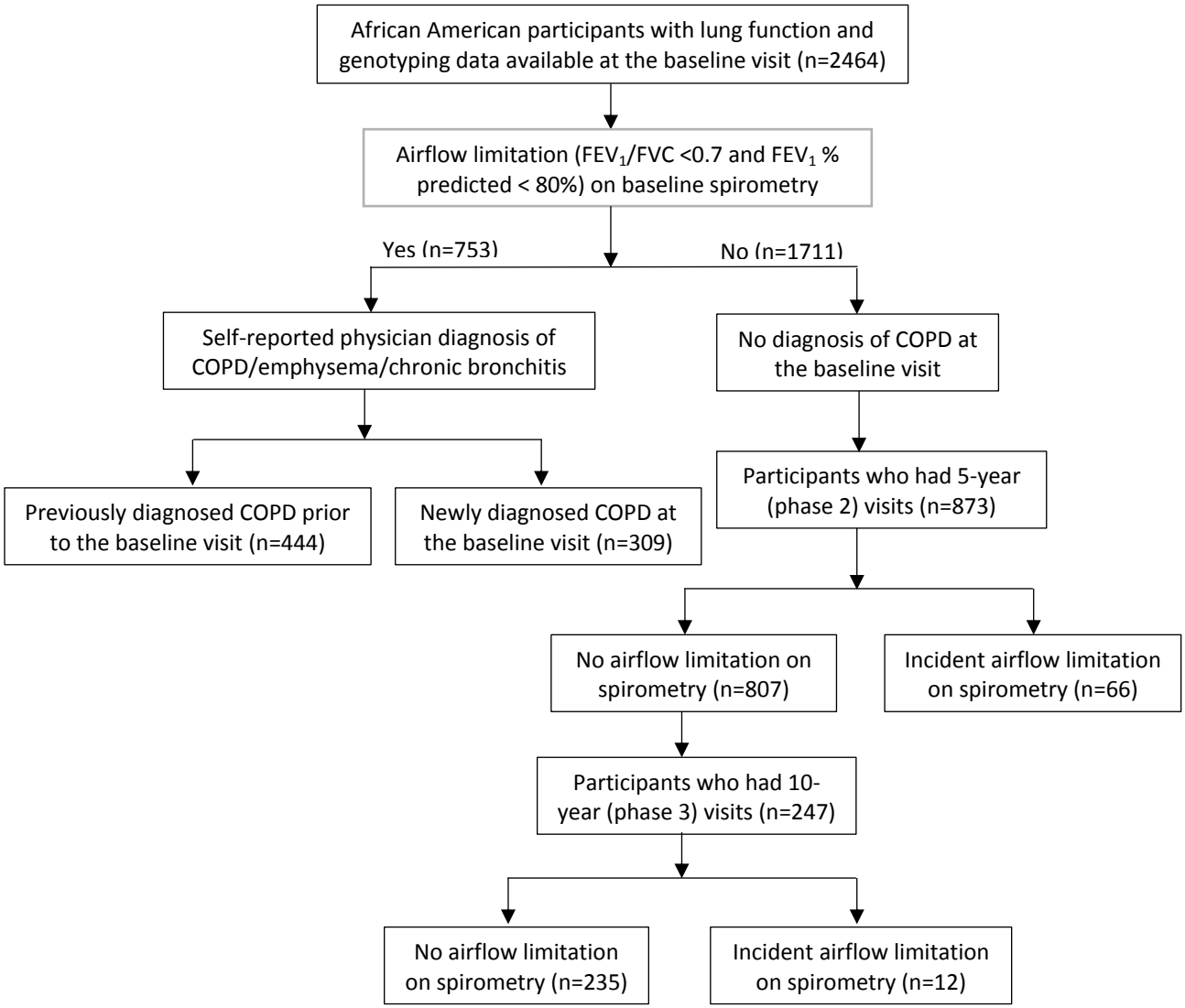

Figure S2. Flowchart of diagnosis of moderate-to-severe COPD (GOLD 2-4) at baseline and follow-up visits for COPDGene African American participants.

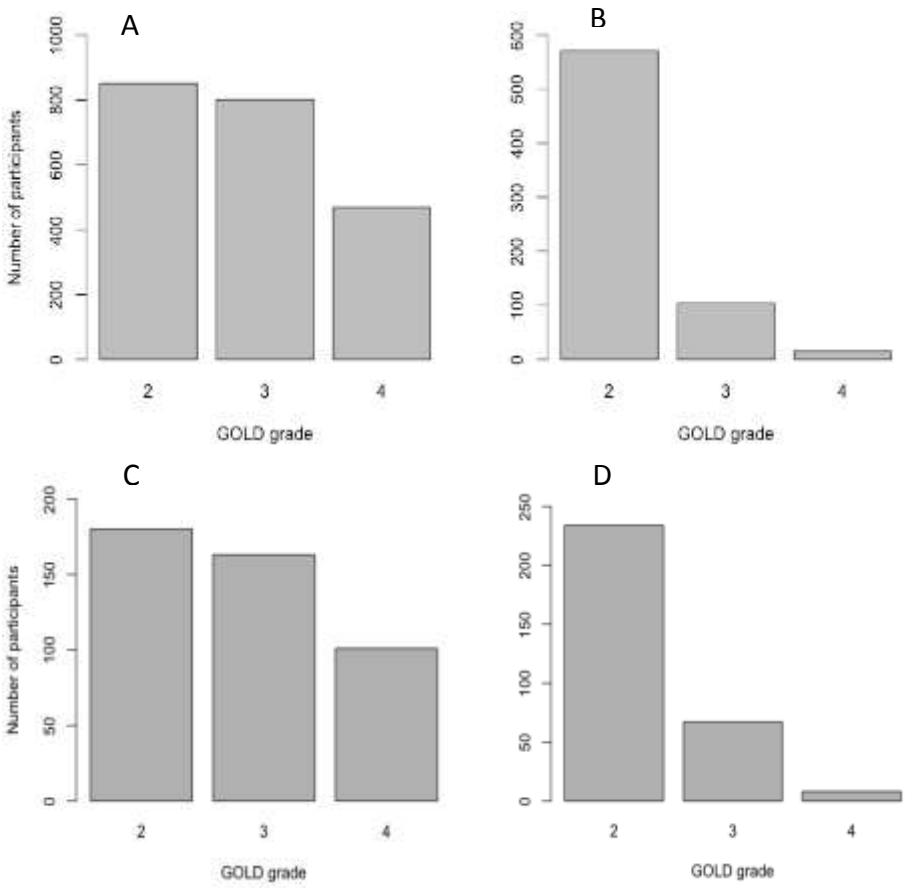

Figure S3. Severity of airflow limitation on baseline spirometry in subjects with previously diagnosed COPD and newly diagnosed COPD at the baseline visit in COPDGene non-Hispanic white (A and B) and African American (C and D) participants. Subjects with previously diagnosed COPD reported a physician diagnosis of COPD prior to the baseline visit. GOLD: Global Initiative for Chronic Obstructive Lung Disease.

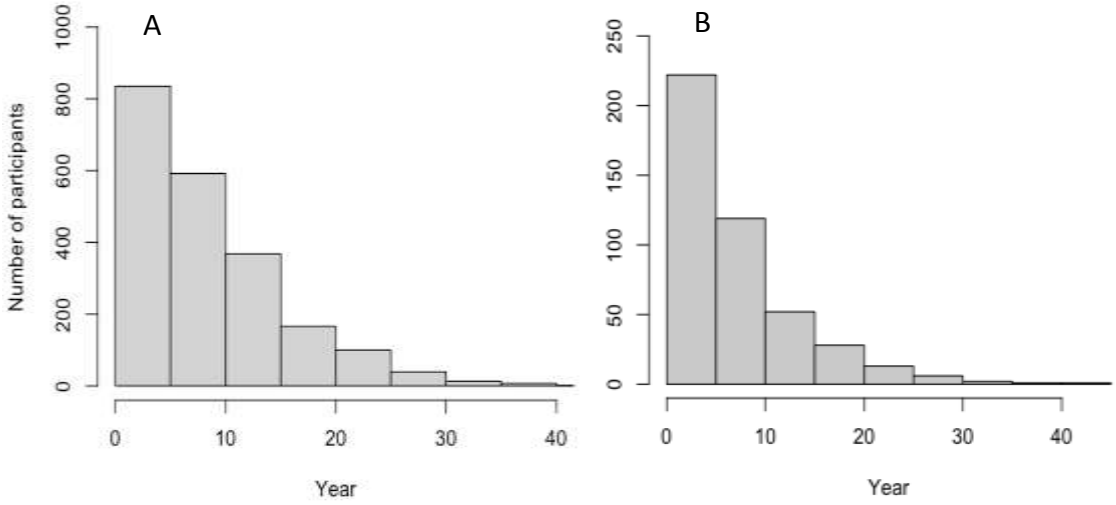

Figure S4. Age difference between baseline visit and diagnosis of COPD in COPDGene participants with physician diagnosed COPD. A: non-Hispanic white, B: African American.

Supplementary Appendix to "A Polygenic Risk Score and Age of Diagnosis of Chronic Obstructive Pulmonary Disease"

Jingzhou Zhang, Hanfei Xu, Dandi Qiao, Dawn L. DeMeo, Edwin K. Silverman, George T. O'Connor, Brian D. Hobbs, Josée Dupuis, Michael H. Cho*, Matthew Moll*

Table of Contents

Funding and Acknowledgements

COPDGene Phase 3

Grant Support and Disclaimer

The project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

COPD Foundation Funding

COPDGene is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer-Ingelheim, Genentech, GlaxoSmithKline, Novartis, Pfizer, Siemens, and Sunovion.

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 Beaty, PhD; Ferdouse Begum, PhD; Peter J. Castaldi, MD, MSc; Michael Cho, MD; Dawn L. DeMeo, MD, MPH; Adel R. Boueiz, MD; Marilyn G. Foreman, MD, MS; Eitan Halper-Stromberg; Lystra P. Hayden, MD, MMSc; Craig P. Hersh, MD, MPH; Jacqueline Hetmanski, MS, MPH; Brian D. Hobbs, MD; John E. Hokanson, MPH, PhD; Nan Laird, PhD; Christoph Lange, PhD; Sharon M. Lutz, PhD; Merry-Lynn McDonald, PhD; Margaret M. Parker, PhD; Dmitry Prokopenko, Ph.D; Dandi Qiao, PhD; Elizabeth A. Regan, MD, PhD; Phuwanat Sakornsakolpat, MD; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Sungho Won, PhD

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 Sanchez-Ferrero, 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; Gregory Kinney, MPH, PhD; Sharon M. Lutz, PhD; Kendra A. Young, PhD

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 Banaei-Kashani, Ph.D

Supplementary Methods

Study design and populations

COPDGene

The COPDGene study aimed to enroll 10,000 former and current smoking participants from 21 clinical centers across the U.S., with a goal of 4,000 control participants ($FEV_1/FVC \geq 0.7$, $FEV_1 \geq 80\%$ predicted), 2,000 GOLD 1 and preserved ratio impaired spirometry (PRISm) participants, and 4,000 GOLD 2-4 participants. NHW and AA participants were planned to be 2/3 and 1/3 of each group, respectively (www.copdgene.org/study-design). Recruitment for COPDGene was directed by the individual centers and monitored by the Data Coordinating Center. The full study protocol and manual of procedures, including details of recruitment, are available at http://www.copdgene.org.

Framingham Heart Study

The Framingham Heart Study (FHS) is a U.S. community-based cohort in Framingham, Massachusetts, established in 1948 (the Original cohort). The Offspring cohort began in 1971 and is comprised of children of the Original cohort and spouses of these children [1]. The Third Generation (Gen3) cohort started in 2002 and is comprised of children from large Offspring cohort families [2]. Spirometry was performed at exams 5-9 of the Offspring cohort and exams 1-2 of the Gen3 cohort. The time intervals between exams were approximately 4-6 years. This study included participants from the Offspring and Gen3 cohorts with both spirometry and genetic data available.

Predictive variables

Polygenic risk score

The construction of the polygenic risk score is as previously described [3]. Briefly, we constructed a polygenic risk to predict moderate-to-severe COPD, which is characterized both by a low FEV_1 and FEV_1/FVC ratio, based on results of GWASs of FEV_1 and FEV_1/FVC. Regression coefficients were used as weights to calculate a weighted sum of the two PRSs into a composite COPD PRS, which was a continuous variable. The composite PRS captured more disease-related heterogeneity and generally performed better than either individual PRS, which was another reason we chose to use the composite score in this study.

Early-life risk factors

Childhood asthma was determined by self-reported physician-diagnosed asthma with age at diagnosis equal to or less than twelve or “as a child” [4]. Adolescent smoking was defined by self-reported smoking initiation before the age of nineteen. In concordance with prior analyses in COPDGene [5–7], we defined childhood pneumonia using self-reported age of physician-diagnosed pneumonia below sixteen or “as a child”, and defined family history of COPD as self-reported maternal or paternal history of COPD, emphysema, and/or chronic bronchitis. Participants reported highest degree or level of school completed and we categorized them into three levels of education: high school or below, college, and graduate school.

Statistical analysis

We tested whether sex modified the effect of the PRS on incident COPD by including a cross-product interaction term of PRS-by-sex in multivariable regression models.

We additionally fitted Cox models separately for different age intervals (age <50, 50-59, 60-69, and ≥70) to allow time-dependent coefficients (β) of the PRS which was a fixed covariate. The covariates of the Cox model included the PRS, sex, age and pack-years of smoking at the baseline visit, age at smoking initiation and first 5 principal components of ancestry, which remained unchanged amongst age intervals. For example, a participant who was diagnosed COPD at age of 75 would contribute to the Cox models for the first three age intervals (age <50, 50-60 and 60-70) as a "censor" at the end of each age interval and contribute to the last age interval (age ≥70) as an "event" occurring at 75. We used the "survSplit" function in the "survival" R package to break up the data into time-dependent parts [8].

The logistic regression model used COPD50 as cases and used participants with age of COPD diagnosis ≥50 (COPD ≥50) and those who were followed at least until 50 and had no COPD (without COPD ≥50) as controls. We performed an additional analysis excluding those with age of COPD diagnosis ≥50 from the controls. We used variance inflation factors (VIFs) to examine potential multicollinearity among other early-life risk factors, accepting VIFs less than 3.3 [9]. We also evaluated the correlation of active smoking in adolescence with education level and examined the effect of including education level as a covariate in a model with active smoking during adolescence.

Sensitivity analysis

To assess for potential bias from censoring, we performed linear regression analyses to examine the association between the PRS and age at diagnosis of COPD (as a continuous outcome variable) in participants with a diagnosis of COPD at the baseline visit; we also repeated the logistic regression analyses for COPD50 in participants aged ≥50 at the baseline visit. To assess for potential bias from using two definitions of COPD diagnosis, we additionally performed linear regression analyses examining the association between the PRS and age of diagnosis separately in participants with physician-diagnosed COPD and newly diagnosed COPD at the baseline visit. We performed a time-to-event analysis using only those with physician-diagnosed COPD and fitted a Cox model using age as the underlying timescale and time was censored at age of physician diagnosis or age of baseline visit. We additionally fitted a Cox model for incident COPD in participants who had no COPD at the baseline visit, using baseline visit as the start time and adjusting for age at baseline visit. We evaluated an age-dependent effect of PRS by testing the interaction between PRS and age of the baseline visit. To assess for potential bias from misdiagnosed COPD by physicians, we repeated the above-mentioned linear regression for age at diagnosis by excluding participants with physician-diagnosed asthma from physician-diagnosed COPD. For time-to-event analyses estimating the effect of the PRS on incident COPD, we fitted a Weibull accelerated failure time model in addition to the Cox PH models described above.

Supplementary Results

Sample characteristics

In COPDGene, participants with COPD were older, had lower FEV_1 % predicted, more pack-years of smoking, and higher polygenic risk scores, but similar age of smoking onset, compared to those without COPD at baseline. Among those with physician-diagnosed COPD, lead years of diagnosis (age difference between baseline visit and diagnosis) showed a right skewed distribution (supplemental Figure 4) and had a median (IQR) of 6.5 (8.3) years for NHW and 5.1 (7.0) years for AA participants. Individuals with higher grades of GOLD airflow limitation on baseline spirometry had longer lead years of diagnosis (5.5 years (GOLD 2) *vs.* 6.9 years (GOLD 3) *vs.* 8.0 years (GOLD 4), p <0.0001 for NHW; 4.2 years (GOLD 2) *vs.* 4.7 years (GOLD 3) *vs.* 6.6 years (GOLD 4), p =0.017 for AA participants).

Of 3836 NHW and 1711 AA participants without COPD at baseline, 2492 and 873 were followed up at 5 years (i.e., Phase 2), and 260 and 66 were newly diagnosed with COPD, respectively. At the 10-year follow up visit (i.e., Phase 3), 808 NHW and 247 AA participants without COPD were followed, and 69 and 12 were found to have new-onset COPD, respectively. Combining physician-diagnosed COPD prior to the baseline visit and newly diagnosed COPD at each visit, the median (range) of age of diagnosis were 60.0 (31.0-90.7) and 55.0 (32.0-84.4) years in NHW and AA participants, respectively. For those without COPD, the median (range) of age of last follow-up were 64.6 (44.0-90.9) in NHW and 55.8 (45.0- 85.3) years in AA participants.

In FHS, compared to those without COPD, participants who had COPD at baseline were older and more likely to be ever-smokers (*vs.* life-time non-smokers), had more pack-years of smoking in ever-smokers, and higher polygenic risk scores, but similar age of smoking initiation. A total of 1761 participants were lost to follow-up during four follow-up visits.

Prediction for COPD diagnosed early in life

In the multivariable logistic regression analysis after excluding participants with an age of COPD diagnosis ≥50 from controls, we observed that the effect of the PRS on COPD50 was larger (OR [95% CI] per SD 1.94 [1.74-2.16] for NHW) compared to that observed in the main analysis.

We did not find evidence of collinearity among early-life risk factors based on variance inflation factors. We also found that active smoking during adolescence was associated with lower educational attainment (Table S2). Comparing multivariable logistic regression models without and with education as a covariate, there was an 11.8% change on an additive scale in the regression coefficient (β =0.738 and 0.651, respectively) of active smoking during adolescence.

Sensitivity analysis

All sensitivity analyses were performed in COPDGene NHW participants. In time-to-event analyses using only those with physician-diagnosed COPD, one SD increase of the PRS was associated with a 37% (HR 1.37, 95% CI 1.31-1.43, p <0.0001) increased hazard for COPD. As there was a statistical trend suggesting potential violation of the proportional hazards (PH) assumption (p=0.054), we additionally fitted Cox models separately among age intervals (age <50, 50-60, and ≥60) to allow different coefficients of the PRS. We combined age intervals (60-

69 and ≥70) due to fewer participants contributing to age ≥70 in this subsample. The HR estimates (95% CI) per SD increase of the PRS were 1.49 (1.38-1.61), 1.36 (1.27-1.46), 1.28 (1.19-1.38) for the above three age intervals, respectively.

Among 2492 NHW participants who had no COPD at the time of enrollment and had follow-up visits, 329 incident COPD diagnoses occurred during a median follow-up time of 5.9 years (range: 3.8-12.0 years), and every increase of the PRS by 1 SD was associated with a 37% increased risk (HR 1.37, 95% CI 1.23-1.54, p <0.0001) for incident COPD. We did not find a significant interaction between PRS and age at baseline visit (β = -0.010, p = 0.13 for the interaction term), possibly due to a lack of power from both a decreased sample size with fewer incident COPD cases and a narrowing range of age of diagnosis in this subsample. However, a negative estimate of the regression coefficient of the interaction term is consistent with a negative age-dependency of PRS (β = -0.010 indicating an increase in the hazard of the PRS by 10.5% on a multiplicative scale for every 10-year younger in age).

In a subsample of participants with physician-diagnosed COPD and without self-reported asthma at baseline (n =1621), one SD increase of the PRS was associated with a 1.00-year (95% CI 0.54-1.46, p <0.0001) earlier diagnosis of COPD.

For an accelerated failure time model, one SD increase in the PRS was associated with a younger age at diagnosis of COPD by an acceleration factor of 0.94 (95% CI: 0.94-0.95) or was associated with a 6% earlier age of diagnosis.

Supplementary Tables

Table S1. Associations between the PRS and other early-life risk factors and COPD occurring early in life (age < 50) in COPDGene non-Hispanic white participants aged ≥ 50 at the baseline visit.

Predictive variable	OR (95% CI)		
	model 1	model 2	model 3
Polygenic risk score *	1.55 (1.39-1.72)		1.49 (1.34-1.67)
Maternal smoking during pregnancy		1.46 (1.15-1.84) $	1.50 (1.18-1.90)
Active smoking during adolescence		1.87 (1.37-2.56)	1.93 (1.41-2.64)
Childhood asthma		2.76 (1.98-3.86)	2.46 (1.75-3.46)
Family history of COPD		2.10 (1.67-2.63)	1.96 (1.56-2.46)
Education			
High school or below (reference)		1.00	1.00
College		0.60 (0.48-0.76)	0.63 (0.50-0.79)
Graduate		0.33 (0.20-0.52)	0.37 (0.23-0.59)

PRS: polygenic risk score; OR: odds ratio; CI: confidence interval.
model 1: PRS + principal components of genetic ancestry; model 2: other early-life risk factors; model 3: PRS + other early-life risk factors + principal components of genetic ancestry
* ORs were calculated for per standard deviation increase of PRS
$ p =0.002
All other associations were statistically significant with p <0.001

Table S2. Association between active smoking in adolescence and educational attainment.

		Educational attainment		
		High school or below	College	Graduate
Smoking in Adolescence	Yes	1641 (81.8%)	2831 (73.9%)	547 (67.6%)
	No	366 (18.2%)	1000 (26.1%)	262 (32.4%)

Supplementary Figures

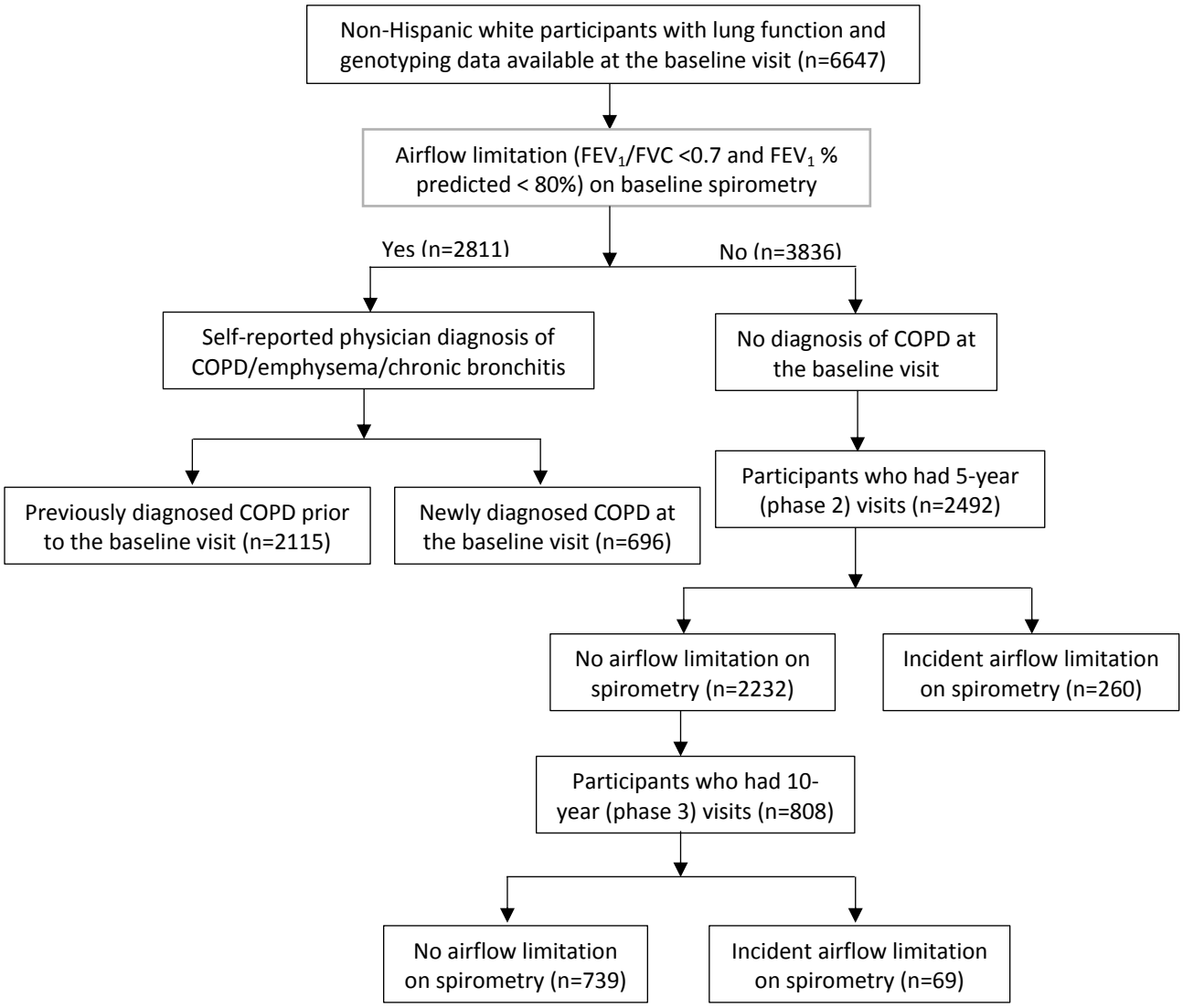

Figure S1. Flowchart of diagnosis of moderate-to-severe COPD (GOLD 2-4) at baseline and follow-up visits for COPDGene non-Hispanic white participants.

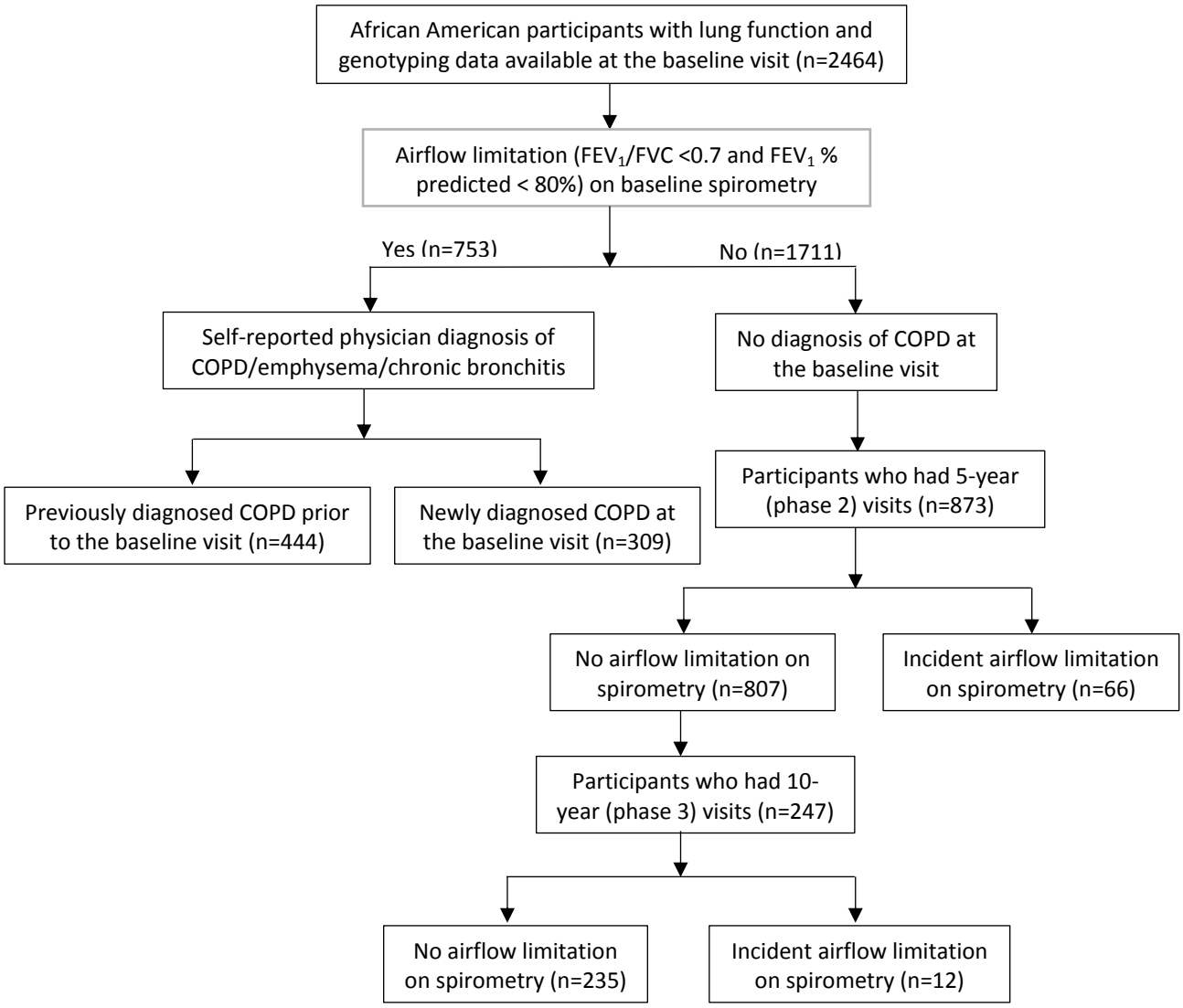

Figure S2. Flowchart of diagnosis of moderate-to-severe COPD (GOLD 2-4) at baseline and follow-up visits for COPDGene African American participants.

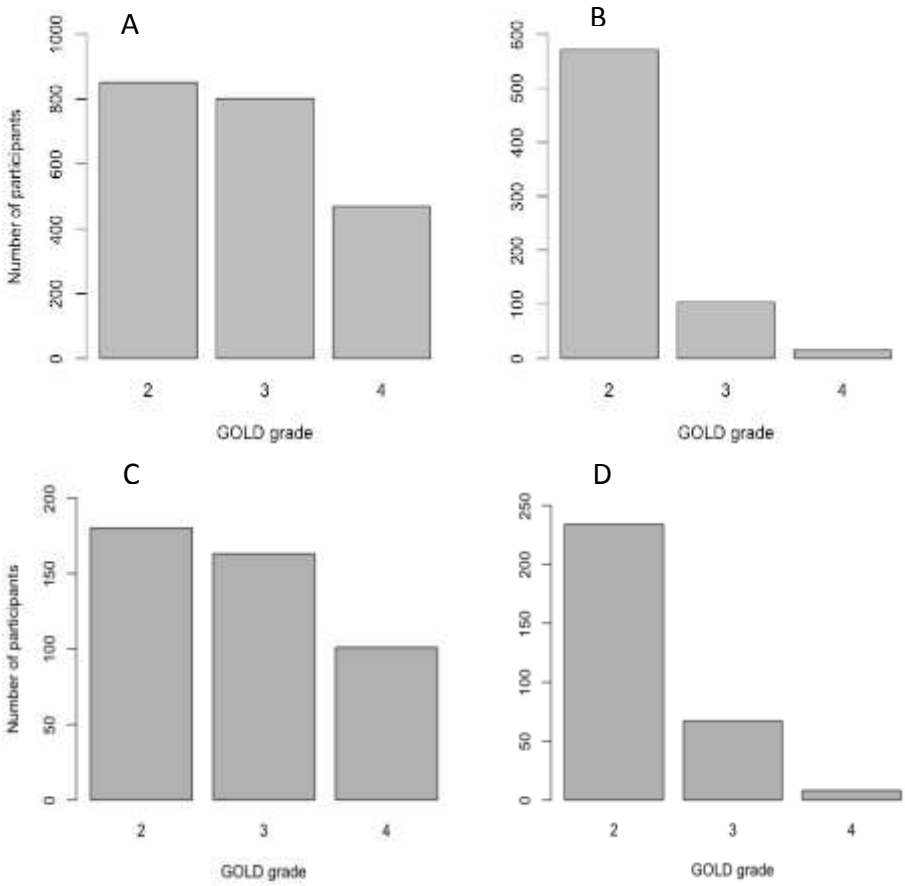

Figure S3. Severity of airflow limitation on baseline spirometry in subjects with previously diagnosed COPD and newly diagnosed COPD at the baseline visit in COPDGene non-Hispanic white (A and B) and African American (C and D) participants. Subjects with previously diagnosed COPD reported a physician diagnosis of COPD prior to the baseline visit. GOLD: Global Initiative for Chronic Obstructive Lung Disease.

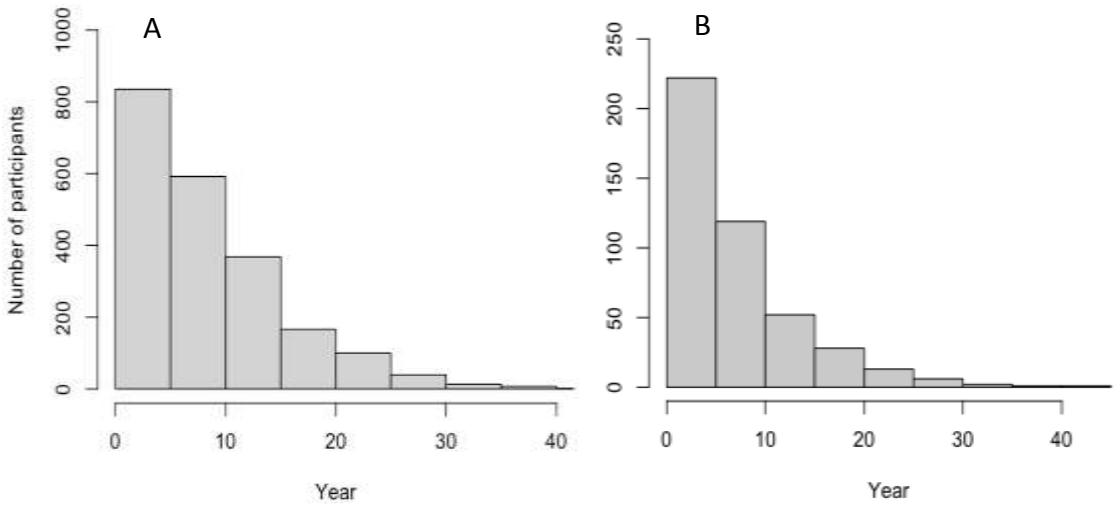

Figure S4. Age difference between baseline visit and diagnosis of COPD in COPDGene participants with physician diagnosed COPD. A: non-Hispanic white, B: African American.

References

1. Kannel WB, Feinleib M, Mcnamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: The framingham offspring study. *Am. J. Epidemiol.* 1979; 110: 281–290.
2. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, recruitment, and initial examination. *Am. J. Epidemiol.* 2007; 165: 1328–1335.
3. Moll M, Sakornsakolpat P, Shrine N, Hobbs BD, DeMeo DL, John C, Guyatt AL, McGeachie MJ, Gharib SA, Obeidat M, Lahousse L, Wijnant SRA, Brusselle G, Meyers DA, Bleecker ER, Li X, Tal-Singer R, Manichaikul A, Rich SS, Won S, Kim WJ, Do AR, Washko GR, Barr RG, Psaty BM, Bartz TM, Hansel NN, Barnes K, Hokanson JE, Crapo JD, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir. Med.* 2020; 8: 696–708.
4. McGeachie MJ, Yates KP, Zhou X, Guo F, Sternberg AL, Van Natta ML, Wise RA, Szefler SJ, Sharma S, Kho AT, Cho MH, Croteau-Chonka DC, Castaldi PJ, Jain G, Sanyal A, Zhan Y, Lajoie BR, Dekker J, Stamatoyannopoulos J, Covar RA, Zeiger RS, Adkinson NF, Williams P V., Kelly HW, Grasemann H, Vonk JM, Koppelman GH, Postma DS, Raby BA, Houston I, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N. Engl. J. Med.* 2016; 374: 1842–1852.
5. Hayden LP, Hobbs BD, Cohen RT, Wise RA, Checkley W, Crapo JD, Hersh CP. Childhood pneumonia increases risk for chronic obstructive pulmonary disease: The COPDGene study. *Respir. Res.* 2015; 16.
6. Hersh CP, Hokanson JE, Lynch DA, Washko GR, Make BJ, Crapo JD, Silverman EK. Family history is a risk factor for COPD. *Chest* 2011; 140: 343–350.
7. Moll M, Lutz SM, Ghosh AJ, Sakornsakolpat P, Hersh CP, Beaty TH, Dudbridge F, Tobin MD, Mittleman MA, Silverman EK, Hobbs BD, Cho MH. Relative contributions of family history and a polygenic risk score on COPD and related outcomes: COPDGene and ECLIPSE studies. *BMJ Open Respir. Res.* 2020; 7.
8. Therneau T, Crowson C, Atkinson E. Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. 2021; Available from: https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf.
9. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: A Global Perspective. 7th ed. Pearson Prentice Hall Up. Saddle River, NJ 2010.