

# Multi-trait polygenic scores for COPD and COPD exacerbations implicate druggable proteins

Chengyue Zhang,<sup>1</sup> Iain R. Konigsberg,<sup>2</sup> Yixuan He,<sup>3</sup> Jingzhou Zhang,<sup>4</sup> Tinashe Chikowore,<sup>1,5</sup> William B. Feldman,<sup>5,6</sup> Xiaowei Hu,<sup>7</sup> Yi Ding,<sup>8</sup> Bogdan Pasaniuc,<sup>9</sup> Diana Chang,<sup>10</sup> Qingwen Chen,<sup>1,5</sup> Jessica A. Lasky-Su,<sup>1,5</sup> Julian Hecker,<sup>1,5</sup> Martin D. Tobin,<sup>11</sup> Jing Chen,<sup>11</sup> Sean Kalra,<sup>1,5</sup> Katherine A. Pratte,<sup>12</sup> Hae Kyung Im,<sup>13</sup> Emily S. Wan,<sup>5,14</sup> Ani Manichaikul,<sup>15</sup> Edwin K. Silverman,<sup>1,5</sup> Russell P. Bowler,<sup>16</sup> Leslie A. Lange,<sup>2</sup> Victor E. Ortega,<sup>17</sup> Alicia R. Martin,<sup>5,18,19</sup> Michael H. Cho,<sup>1,5</sup> and Matthew R. Moll<sup>1,5,14</sup>

<sup>1</sup>Channing Division of Network Medicine, Mass General Brigham, Boston, Massachusetts, USA. <sup>2</sup>Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA. <sup>3</sup>Department of Epidemiology School of Public Health, UTHealth Houston, Houston, Texas, USA. <sup>4</sup>Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts, USA. <sup>5</sup>Harvard Medical School, Boston, Massachusetts, USA. <sup>6</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Mass General Brigham, Boston, Massachusetts, USA. <sup>7</sup>Department of Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA. <sup>8</sup>Department of Medicine, DanaFarber Cancer Institute, Boston, Massachusetts, USA. <sup>9</sup>Center for Computational Biomedicine, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>10</sup>Genentech, South San Francisco, California, USA. <sup>11</sup>Department of Population Health Sciences, University of Leicester, Leicester, United Kingdom. <sup>12</sup>Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, Colorado, USA. <sup>13</sup>Department of Human Genetics, Department of Medicine, University of Chicago, Chicago, Illinois, USA. <sup>14</sup>Section on Pulmonary, Critical Care, Allergy, and Sleep Medicine, VA Boston Healthcare System, West Roxbury, Massachusetts, USA. <sup>15</sup>Public Health Genomics Program, University of Virginia, Charlottesville, Virginia, USA. <sup>16</sup>Department of Genetic Medicine, Cleveland Clinic Lerner College of Medicine, Cleveland, Ohio, USA. <sup>17</sup>Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Phoenix, Arizona, USA. <sup>18</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. <sup>19</sup>Program in Medical & Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.

**Authorship note:** CZ and IRK contributed equally and have been designated as co-first authors. MHC and MRM contributed equally and have been designated as co-last authors.

**Conflict of interest:** MM discloses consulting fees from 2ndMD, TheaHealth, Axon Advisors, Dialectica, Sanofi, and Verona Pharma; grant support from Genentech; service as a medical legal expert for multiple law firms; and honoraria from ATS 2024 and NYSTS 2024.

**Copyright:** © 2026, Zhang et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

**Submitted:** September 10, 2025

**Accepted:** February 10, 2026

**Published:** February 19, 2026

**Reference information:** *JCI Insight*. 2026;11(7):e199951.  
<https://doi.org/10.1172/jci.insight.199951>.

**BACKGROUND.** We constructed multi-trait polygenic risk scores (PRSs) predicting chronic obstructive pulmonary disease (COPD) and exacerbations, validated their performance in diverse cohorts, and identified PRS-related proteins for potential therapeutic targeting.

**METHODS.** PRSmix\*, a multi-trait PRS framework, is used to train a composite PRS (PRS<sub>multi</sub>) in COPD Gene non-Hispanic White participants ( $n = 6,647$ ). Associations of PRS<sub>multi</sub> with COPD status (GOLD 2–4 vs. GOLD 0 or ICD) and exacerbation frequency were tested in COPD Gene African American ( $n = 2,466$ ), ECLIPSE ( $n = 1,858$ ), Mass General Brigham Biobank ( $n = 15,152$ ), and All of Us ( $n = 118,566$ ). Protein prediction models were applied to GWAS summary statistics from traits contributing to PRS<sub>multi</sub> and were validated with proteomic data in COPD Gene ( $n = 5,173$ ) and UK Biobank ( $n = 5,012$ ).

**RESULTS.** PRSmix\* selected 7 traits for PRS<sub>multi</sub>. In multivariable models, PRS<sub>multi</sub> was associated with COPD status (meta-analysis random effects [RE] OR 1.58 [95% CI: 1.28–1.94]) and exacerbation frequency (meta-analysis RE  $\beta$  0.21 [95% CI: 0.11–0.31]), with higher effect sizes observed in smoking-enriched cohorts. PRS<sub>multi</sub> outperformed traditional single-trait PRS in all tested cohorts. Using protein prediction models, we identified 73 proteins associated with the PRSs that were also validated with measured protein levels in COPD Gene and UK Biobank. Of these proteins, 25 were linked to approved or investigational drugs. Notable targets include RAGE/sRAGE, IL1RL1, and SCARF2, all implicated in COPD pathogenesis and exacerbations.

**CONCLUSIONS.** Multi-trait PRS improves prediction of COPD and exacerbation risk. Integration with proteomic data identifies druggable protein targets, offering a promising avenue for precision medicine in COPD management.

**TRIAL REGISTRATION.** COPD Gene: ClinicalTrials.gov NCT00608764; ECLIPSE: ClinicalTrials.gov NCT00292552.

## Introduction

Chronic obstructive pulmonary disease (COPD), a leading global cause of morbidity and mortality, is marked by persistent airflow limitation and an enhanced inflammatory lung response to harmful inhaled particles or gases. COPD is highly heterogeneous in presentation, progression, therapy response, and exacerbations, the latter of which are major drivers of morbidity (1). Comorbid conditions, such as cardiovascular and metabolic diseases, are important predictors of COPD-related mortality. However, susceptibility to COPD and its comorbidities is determined by complex gene-by-environment interactions (2). Notably, only a minority of smokers develop COPD, with genetics accounting for approximately 30%–40% of the variability in susceptibility (3).

Genome-wide association studies (GWAS) have identified numerous genetic variants linked to lung function and COPD (4). Summing GWAS variants together, polygenic risk scores (PRSs) for 2 measures of lung function (forced expiratory volume at 1 second [FEV<sub>1</sub>] and the ratio of FEV<sub>1</sub> to forced vital capacity [FEV<sub>1</sub>/FVC]) were combined into a COPD PRS that was highly predictive of COPD in multiple cohorts, outperforming single PRSs in predicting COPD and COPD-related phenotypes (5). Recent findings highlight that BMI genetics can predict mortality in patients with COPD, underscoring the importance of multiple-trait (multi-trait) genetic analyses in understanding COPD risk and outcomes (6).

Multi-trait genetic analyses, which identify shared mechanisms among traits, have gained traction with the rise of a multitude of PRSs for various traits. He et al. developed a multi-trait PRS for COPD that performed exceptionally well in biobanks (7). Furthermore, this multi-trait PRS and the prior lung function–based COPD PRS were associated with exacerbations. However, in the latter, the COPD PRS effects were attenuated when accounting for baseline lung function, consistent with the known clinical association between COPD severity and exacerbations. It remains unclear whether combining PRSs for spirometry and broader phenotypes can better predict COPD and exacerbations across diverse cohorts.

While shared genetics can inform phenotype prediction and reveal insights into disease mechanisms, proteins offer several advantages. First, proteins are influenced by both genetics and environmental factors, such as infections — the leading cause of exacerbations (8). Second, proteins can serve as direct therapeutic targets for small molecules (9). Third, recent advances in statistical tools can use genetics to predict protein expression levels (10), which are more likely causally related to a trait than non–genetically regulated proteins (11). We hypothesized that integrating PRSs for spirometry and comorbid traits could enhance the prediction of COPD and exacerbations in both research and biobank cohorts. Furthermore, we sought to identify and validate genetically predicted protein levels linked to shared genetic architectures, using measured protein levels to confirm key findings. Our main goal was to leverage polygenic risk to identify protein drug targets associated with exacerbations through generation of an improved multi-trait PRS and identification of proteins associated with exacerbations and genetic scores.

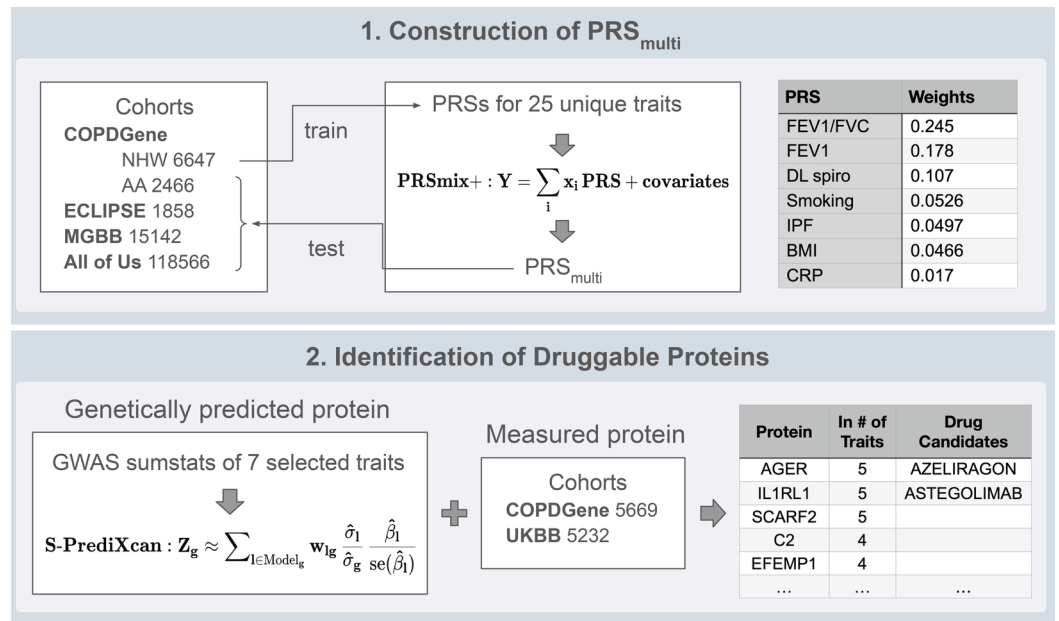
## Results

An overview of the study design is shown in Figure 1 and described in Methods.

*Characteristics of study participants.* A total of 144,679 individuals were included in genetic analyses (Table 1), drawn from both COPD-enriched cohorts (Genetic Epidemiology of COPD [COPDGene] and Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints [ECLIPSE]) and population-scale biobanks (Mass General Brigham Biobank [MGBB] and All of Us). As expected, smoking prevalence varied across cohorts.

Cohort characteristics for COPDGene and UK Biobank (UKBB) proteomic analyses are shown in Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.199951DS1>. Participants had similar age, sex distribution, and spirometric measures. In COPDGene, individuals who were never smokers were excluded.

*Development of multi-trait PRSs for COPD and related traits.* We used 25 individual PRSs (Supplemental Table 2), chosen based on clinician input and literature review, to develop multi-trait PRSs for COPD and related traits. Four new PRSs were developed for this analysis (emphysema, peak expiratory flow, eosinophils, and smoking) which are available upon request. Clustering analysis revealed that PRSs for spirometry were highly correlated and that the emphysema and type 2 diabetes scores were correlated, as were asthma/allergic rhinitis and venous thromboembolism/cor pulmonale (Supplemental Figure 1).



**Figure 1. Overview of study design.** Schematic representation of the study workflow. Plots included in this figure are intended for illustrative purposes only.

We constructed multi-trait PRS models for COPD status (Global Initiative for Chronic Obstructive Lung Disease [GOLD] stages 2–4 vs. GOLD 0) and exacerbations, as well as computed tomography (CT) measures of quantitative emphysema (adjusted lung density) and airway wall thickness (square root of the wall area of a hypothetical airway with an internal perimeter of 10 millimeters, Pi10) using PRSmix+. Mixing weights for each model are displayed in Supplemental Table 3. Final weighted models included between 2 and 7 individual PRSs. Notably, PRSs for FEV<sub>1</sub>/FVC (PRS<sub>ratio</sub>) and a deep learning spirometry (PRS<sub>DLspiro</sub>) trait contribute to all 4 traits. PRSs for FEV<sub>1</sub> (PRS<sub>FEV1</sub>), BMI (PRS<sub>BMI</sub>), and CRP (PRS<sub>CRP</sub>) contribute to both Pi10 and COPD, and all 7 PRSs, including smoking (cigarettes per day) (PRS<sub>smoke</sub>) and idiopathic pulmonary fibrosis (PRS<sub>IPF</sub>), contribute to COPD. Thus, while PRS<sub>ratio</sub> is the largest contributor to COPD genetic risk, 6 additional PRSs also contribute to COPD risk.

All multi-trait PRSs showed strong associations with their respective training outcomes in multivariable models (Supplemental Table 4). The PRS<sub>mix+</sub> models for emphysema (adjusted lung density) and airway thickness (Pi10) were associated with COPD and exacerbations, and inversely associated with their respective outcomes (e.g., higher PRS<sub>mix+</sub> [Pi10] is associated with lower quantitative emphysema). As our a priori decision was to select the multi-trait PRS that represents the highest number of comorbid traits, the PRS for COPD derived from 7 separate PRSs was carried forward for further testing and will hereafter be referred to as the PRS<sub>multi</sub>. The PRS<sub>multi</sub> explained 20.7% of the variance on the liability scale for COPD (12).

*Testing the multi-trait PRS for COPD.* We tested the predictive utility of PRS<sub>multi</sub> for both COPD and COPD exacerbations compared with individual component PRSs. For COPD, PRS<sub>multi</sub> consistently demonstrated larger absolute effect sizes in all cohorts except All of Us, where PRS<sub>BMI</sub> showed the strongest effect (Supplemental Table 5). For COPD exacerbations, PRS<sub>multi</sub> outperformed other scores in COPDGene non-Hispanic White (NHW) and ECLIPSE, whereas PRS<sub>BMI</sub> had higher effects in the remaining cohorts (Supplemental Table 6). Notably, PRS<sub>multi</sub> associations with COPD status remained significant after adjusting for baseline FEV<sub>1</sub> in COPDGene, but associations with exacerbations were attenuated in FEV<sub>1</sub>-adjusted models.

In a random effects meta-analysis, PRS<sub>multi</sub> was associated with increased odds of COPD (OR: 1.58 per SD [95% CI: 1.28–1.94]) (Figure 2A). We observed significant between-study heterogeneity ( $I^2 = 0.98$ ,  $P < 0.01$ ), with stronger effects in smoking-enriched research cohorts. Among the component PRSs, PRS<sub>IPF</sub> and PRS<sub>BMI</sub> were the only scores not significantly associated with COPD, and PRS<sub>multi</sub> had a larger effect on COPD status than any individual score (Figure 2B).

We next performed a similar analysis for COPD exacerbations but with multivariable negative binomial modeling of count data. In meta-analysis, PRS<sub>multi</sub> was associated with an increase of 0.21 exacerbations per year per SD (95% CI: 0.11–0.31) (Figure 3A). However, this association

**Table 1. Characteristics of study participants**

Characteristic	COPDGene NHW	COPDGene AA	ECLIPSE	All of Us	MGBB
N	6647	2466	1858	118,566	15,142
Age in years, mean (SD)	62.03 (8.83)	54.90 (7.44)	63.14 (7.52)	60.71 (11.37)	64.22 (12.08)
Population, no. (% female) <sup>1</sup>	3,170 (47.7)	1,045 (42.4)	625 (33.6)	71,119 (60.0)	8,918 (58.9)
Smoking status, no. (%)					
Never	0	0	0	75,915 (64.9)	8,489 (56.1)
Former	4,045 (60.9)	529 (21.5)	1,217 (65.5)	22,650 (19.4)	5,835 (38.5)
Current	2602 (39.1)	1,937 (78.5)	641 (34.5)	18,486 (15.8)	818 (5.4)
BMI, mean (SD)	28.69 (6.04)	28.76 (6.48)	26.77 (5.53)	30.12 (7.30)	27.84 (6.02)
Pack years of smoking, mean (SD)	47.26 (26.02)	38.47 (21.05)	48.92 (27.75)	6.79 (15.37)	
FEV <sub>1</sub> % predicted, mean (SD)	77.85 (27.49)	74.15 (23.19)	52.20 (22.53)		
FEV <sub>1</sub> /FVC, mean (SD)	0.64 (0.17)	0.71 (0.15)	0.47 (0.15)		
COPD, no. (% cases)	2,634 (48.6)	803 (40.6)	1,711 (92.1)	3,980 (3.4)	459 (3.0)
Exacerbation rate (no./year)	0.12 (0.51)	0.14 (0.66)	0.75 (0.98)	0.04 (0.29)	0.01 (0.17)

Baseline demographic and clinical characteristics of participants across cohorts (COPDGene, ECLIPSE, MGBB, UKBB, and All of Us). COPDGene, Genetic Epidemiology of COPD; ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints; MGBB, Mass General Brigham Biobank; NHW, non-Hispanic White; AA, African American. <sup>1</sup>The rest of the cohort population was male.

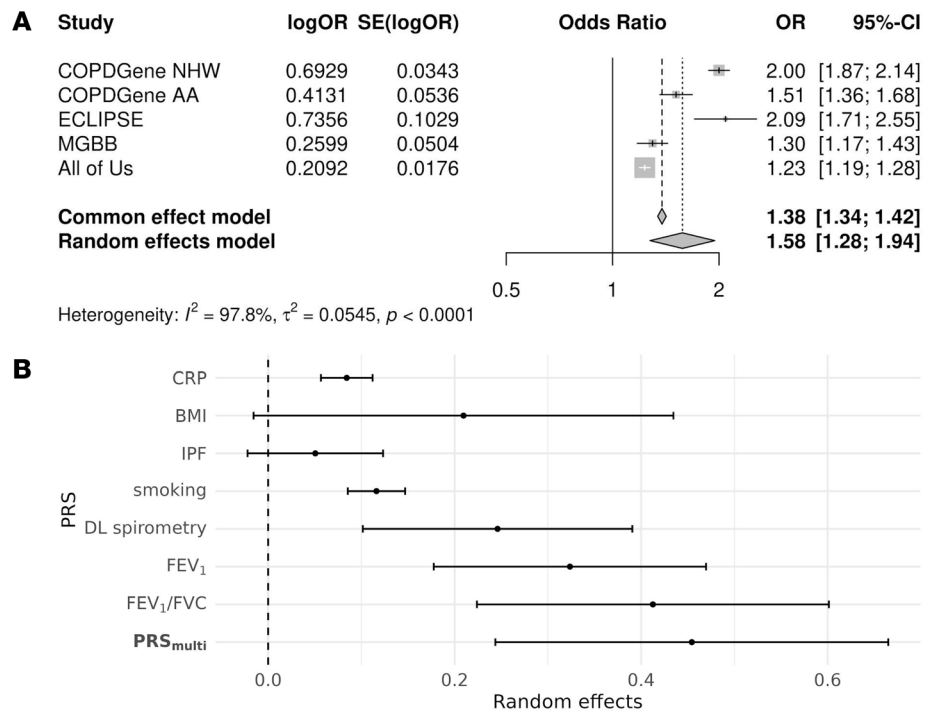
was attenuated in sensitivity analyses adjusting for baseline FEV<sub>1</sub>. All component PRSs were significantly associated with exacerbations except for the PRS<sub>IPF</sub> and PRS<sub>BMI</sub> (Figure 3B). We note that PRS<sub>BMI</sub> demonstrated the largest effects on exacerbations compared with other PRSs in some cohorts (Supplemental Table 6), but it was not significantly associated with exacerbations in meta-analysis (Supplemental Figure 2) due to a flipped direction of effect in COPDGene African American (AA) participants. Again, between-study heterogeneity was significant ( $I^2 = 0.86$ ,  $P < 0.0001$ ), although less pronounced than for COPD status.

We also observed that heterogeneity in COPD status associations was driven largely by differences between research and biobank cohorts, reflecting the use of spirometry- versus International Classification of Diseases–based (ICD-based) case definitions, respectively (Supplemental Figure 3A). In contrast, heterogeneity in exacerbation outcomes was lower and appeared to be influenced by outlier results from MGBB participants (Supplemental Figure 3B). In a leave-one-out sensitivity analysis excluding COPDGene NHW participants, we observed similar results for COPD and COPD exacerbations (Supplemental Figure 4).

In AUC analyses among COPDGene NHW participants, PRS<sub>multi</sub> outperformed PRS<sub>ratio</sub> in predicting both COPD (AUC 0.67 vs. 0.64,  $P = 1.05 \times 10^{-8}$ ) and frequent exacerbations ( $\geq 2$  per year) (AUC 0.569 vs. 0.562,  $P = 0.000001$ ). Both PRSs improve prediction when added to clinical covariates (Table 2 and Supplemental Figure 5). For COPD, the AUC increased from 0.756 (clinical only) to 0.8 with PRS<sub>multi</sub> and 0.79 with the PRS<sub>ratio</sub>. For exacerbation prediction, the AUC increased from 0.599 to 0.612 and 0.609, respectively. The PRS<sub>multi</sub> also demonstrated higher predictive capacity for COPD and frequent exacerbations compared with any of the other tested single-trait PRSs or the previously published PRS<sub>FEV1+FEV1/FVC</sub> from Moll et al. (5) (Supplemental Table 9).

Examining other COPD-related outcomes, we found that a higher PRS<sub>BMI</sub> was the PRS most associated with all-cause mortality, while the PRS<sub>DLSpiro</sub> was most highly associated with respiratory mortality (Supplemental Table 10). However, the PRS<sub>multi</sub> had the largest effects on severe exacerbations and antibiotic or steroid use compared with other tested PRSs (Supplemental Table 10).

*Genetically predicted proteins based on shared genetic risks among traits.* Following the identification of 7 traits contributing to COPD genetic risk, we investigated shared genetic mechanisms by identifying proteins whose levels are predicted to be altered based on the genetic architecture of these traits. Using GWAS summary statistics for each trait, we applied S-PrediXcan (10) with Atherosclerosis Risk in Communities (ARIC) PredictDB protein expression models (13) to infer genetically regulated protein levels (full results available upon request). We then tested whether measured plasma protein levels were associated with COPD exacerbations using multivariable negative binomial regression models in COPDGene and UKBB. Models were adjusted for COPD case-control status and other confounders (see Methods). We examined only proteins meeting FDR-adjusted significance in S-PrediXcan, COPDGene SomaScan



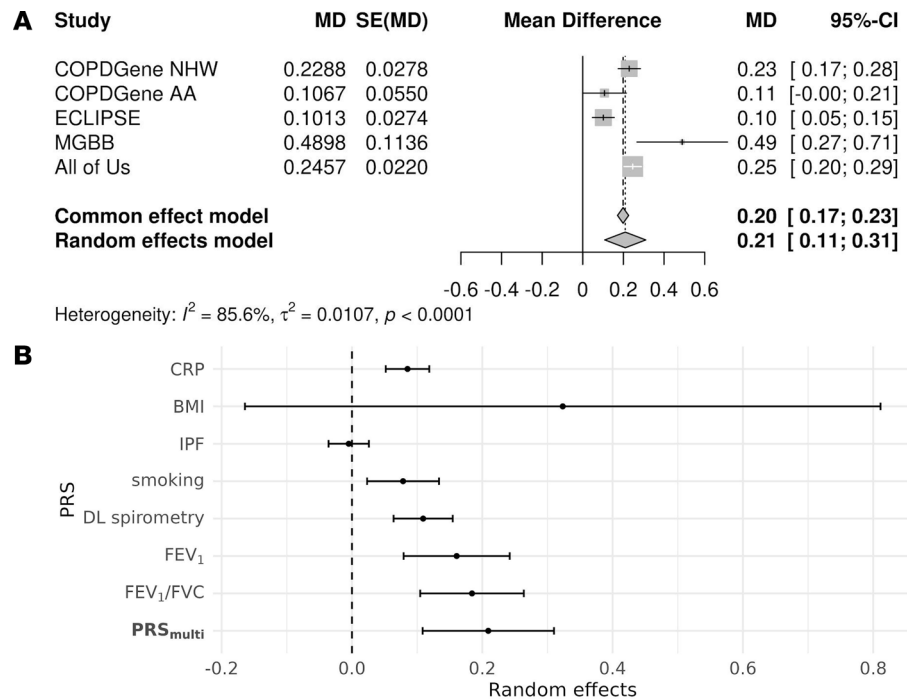
**Figure 2. PRS<sub>multi</sub> is associated with COPD.** (A) Forest plot from a meta-analysis of the association between PRS<sub>multi</sub> and COPD status across multiple cohorts. The figure presents ORs with 95% CIs for each cohort, demonstrating the overall effect size and heterogeneity ( $I^2$ ). (B) Forest plot comparing the effects of individual PRSs and PRS<sub>multi</sub> on COPD risk. PRS<sub>multi</sub> outperforms single-trait PRSs, particularly in smoking-enriched cohorts. CRP, C-reactive protein; BMI, body mass index; IPF, idiopathic pulmonary fibrosis; DL, deep learning; PRS, polygenic risk score. Cohort abbreviations are in the legend for Table 1.

(measured protein levels), and UKBB Olink (measured protein levels) data (Supplemental Table 7). We further restricted the protein list to those with concordant directions of effects in COPDGene and UKBB, resulting in a final set of 73 genetically predicted proteins associated with COPD exacerbations (Table 3). The number of proteins at each filtering step for all 7 traits are detailed in Supplemental Table 11.

We ranked these proteins by the number of traits for which all criteria were met. The receptor for advanced glycation end products (RAGE) and its soluble form (sRAGE), IL-1 receptor-like 1 (IL1RL1), and scavenger receptor class F member 2 (SCARF2) were represented in 5 out of 7 tested traits. SERPINA1 was represented in 2 out of 7 tested traits. We performed gene name lookups in the OpenTargets Platform (14) to identify potential drug repurposing candidates. We identified 25 proteins with existing clinical trials for 46 drugs (Table 3 and Supplemental Table 8). Notably, we found drugs targeting RAGE/sRAGE and IL1RL1 that are being tested in clinical trials, with the latter being tested in COPD. Finally, among the 73 proteins, we identified 8 (agrin [AGRN], CD300C, CFB, GM2 ganglioside activator [GM2A], IL1RL1, INHBB, leukocyte immunoglobulin-like receptor A5 [LILRA5], and TIMP4) that demonstrated consistent associations with COPD status, frequent exacerbations, and the highest versus lowest quintile of PRS<sub>multi</sub> in COPDGene NHW participants (Supplemental Figure 6). Compared with IL1RL1, RAGE/sRAGE, and SCARF2 protein levels, PRS<sub>multi</sub> demonstrated higher predictive capacity for frequent exacerbations (Supplemental Figure 8; AUC 0.77, all  $P < 0.05$ ).

### Discussion

In this study, we leveraged genetic data from over 140,000 individuals across smoking-enriched research cohorts and large-scale biobanks to develop a composite multi-trait PRS (PRS<sub>multi</sub>) for COPD and exacerbations. PRS<sub>multi</sub> outperformed single-trait PRSs, the previously published PRS<sub>FEV1+FEV1/FVC</sub> from Moll et al. (5), and protein levels of RAGE/sRAGE, IL1RL1, and SCARF2 for predicting both COPD risk and exacerbation frequency, demonstrating the advantage of leveraging multiple trait genetics to understand drivers of disease activity. Using elastic net modeling, we developed a multi-trait PRS and identified 7 component traits that share a genetic basis with COPD, including spirometry measures such as FEV<sub>1</sub>/FVC and FEV<sub>1</sub>, a deep learning-derived spirometry phenotype, cigarettes per day, pulmonary fibrosis, C-reactive protein (CRP), and BMI. Using the genetic associations for these traits, we applied protein prediction models and identified 73 genetically predicted



**Figure 3. PRS is associated with COPD exacerbations.** (A) Meta-analysis forest plot showing the association between PRS<sub>multi</sub> and COPD exacerbations, with estimates from individual cohorts and the overall effect. The PRS<sub>multi</sub> is significantly associated with increased exacerbation risk. (B) Comparison of individual PRSs and PRS<sub>multi</sub> for exacerbation prediction. Effect sizes and confidence intervals are presented, highlighting PRS<sub>multi</sub> as the most predictive score. Abbreviations are in the legends of Figure 2 and Table 1.

proteins whose measured levels were associated with exacerbations in both a smoking-enriched and a biobank cohort, independent of COPD status. These findings provide critical insights into the shared genetic architecture of COPD and related traits and nominate a set of proteins as potential biomarkers and therapeutic targets.

Although COPD comorbidities are known predictors of exacerbations and mortality, their genetic contributions to COPD risk have been underexplored. We identified 7 traits contributing to the genetic risk for COPD, emphasizing its complex genetic architecture. Among these, 3 are spirometry-based, including a deep learning–derived spirometry phenotype, indicating that traditional measures like FEV<sub>1</sub> and FEV<sub>1</sub>/FVC alone may not fully capture genetic risk or phenotypic variability. Our decision to use the COPD PRS<sub>mix+</sub> model was based on an a priori principle of taking the model with the greatest number of traits in order to capture the greatest subphenotypic genetic variation. Nonetheless, the other PRS<sub>mix+</sub> models demonstrated interesting findings in research cohorts. For example, the inverse association of airway- and emphysema-based PRSs with the other outcome (e.g., higher PRS<sub>mix+</sub> [P110] is associated with lower quantitative emphysema) is consistent with mounting evidence that COPD exists along emphysema- and airway-predominant axes (15, 16). Future studies can examine how these PRSs for imaging traits can be used to dissect disease heterogeneity.

Notably, the PRSs of several traits known to be associated with exacerbations were not significant in elastic net modeling, including eosinophils and asthma, which have previously demonstrated genetic overlap with COPD (17, 18). This observation could reflect inclusion of relevant overlapping loci into other risk scores (e.g., asthma risk loci included in COPD or lung function), phenotypic specificity (e.g., data on eosinophilic COPD was not uniformly available), or other factors. These findings emphasize the need to understand the shared genetic architecture of COPD and inflammation-related traits. The emphysema PRS was developed from the largest GWAS of quantitative CT emphysema (19) but was not chosen in any of our PRS<sub>mix+</sub> models. The reason for this result is unclear, but when tuned to quantitative emphysema, PRS<sub>ratio</sub> and PRS<sub>DLspiro</sub> were selected, suggesting that the effects of emphysema might be largely captured by the genetics of low lung function.

The PRS<sub>BMI</sub> was most associated with all-cause mortality, which likely reflects higher cardiovascular risks in individuals with a higher BMI; as a corollary, it was the PRS<sub>DLspiro</sub> that was actually most highly associated with respiratory mortality, suggesting that variants associated with low lung function beyond

**Table 2. AUC analyses comparing PRS<sub>multi</sub> to single PRSs for COPD and exacerbations**

Model	COPD AUC	Frequent exacerbations AUC
Clinical model	0.756	0.599
PRS <sub>multi</sub> + PCs	0.666	0.569
Clinical + PRS <sub>multi</sub> + PCs	0.8	0.612
PRS <sub>ratio</sub> + PCs	0.643	0.562
Clinical + PRS <sub>ratio</sub> + PCs	0.79	0.609
COPD AUC <i>P</i> value (PRS <sub>multi</sub> + PCs vs. PRS <sub>ratio</sub> + PCs) = $1.05 \times 10^{-8}$		
COPD AUC <i>P</i> value (clinical + PRS <sub>multi</sub> + PCs vs. clinical + PRS <sub>ratio</sub> + PCs) = $1.35 \times 10^{-6}$		
COPD AUC <i>P</i> value (clinical + PRS <sub>multi</sub> + PCs vs. clinical) = $5.35 \times 10^{-26}$		
COPD AUC <i>P</i> value (clinical + PRS <sub>ratio</sub> + PCs vs. clinical) = $1.29 \times 10^{-19}$		
Frequent exacerbations AUC <i>P</i> value (PRS <sub>multi</sub> + PCs vs. PRS <sub>ratio</sub> + PCs) = $1.09 \times 10^{-6}$		
Frequent exacerbations AUC <i>P</i> value (clinical + PRS <sub>multi</sub> + PCs vs. clinical + PRS <sub>ratio</sub> + PCs) = 0.00138		
Frequent exacerbations AUC <i>P</i> value (clinical + PRS <sub>multi</sub> + PCs vs. clinical) = $6.13 \times 10^{-9}$		
Frequent exacerbations AUC <i>P</i> value (clinical + PRS <sub>ratio</sub> + PCs vs. clinical) = $2.75 \times 10^{-7}$		

AUC comparisons for PRS<sub>multi</sub> versus the best-performing single PRSs for predicting COPD and frequent exacerbations ( $\geq 2$  per year) in COPDGene NHW participants. The table also reports AUC changes when adding PRS to a clinical model (age, sex, BMI, smoking pack-years), with DeLong test *P* values for model comparison.

FEV<sub>1</sub> and FEV<sub>1</sub>/FVC are important predictors of COPD mortality. The PRS<sub>multi</sub> was also associated with severe exacerbations, suggesting that it could be used to enrich a population for severe exacerbators, although further validation is needed.

Although the PRS<sub>multi</sub> outperformed single-trait PRSs and the previously published PRS<sub>FEV1+FEV1/FVC</sub> from a statistical perspective, we acknowledge that this improved performance likely translates to incremental clinical gains; similar observations have been made in PRSs for cardiovascular diseases, yet the utility for both coronary disease and COPD seem to be for earlier age ranges (20, 21) and for case finding (22). In addition to this clinical context, the purpose of the current study was not to develop a better genetic prediction tool, but rather to leverage the shared genetic architecture across multiple traits to gain biological insights into risk of COPD and exacerbations. Indeed, there is broad interest in both multi-trait genetics (23, 24) and genetic-protein prediction (25, 26), and here we apply these principles and domain knowledge in the case of COPD.

To investigate shared genetic mechanisms between these 7 traits, we used genetic protein prediction models derived from protein quantitative trait loci from the ARIC study (13). Using genetics to predict other omics has been shown to enhance prediction of complex traits; for example, polygenic transcriptome risk scores for spirometry using PrediXcan demonstrated greater portability across self-identified race and ethnicity groups compared with a standard PRS (27).

Instead of focusing on a single trait, we applied a genetic protein prediction approach across 7 traits and validated our findings using measured protein levels in 2 cohorts. As a result, we identified 73 genetically predicted proteins associated with exacerbations in both cohorts after adjusting for COPD case-control status. To find the most relevant potential biomarkers for exacerbations, we analyzed proteins that were elevated in COPD cases, frequent exacerbators ( $\geq 2$  exacerbations per year), and individuals in the highest quintile of PRS<sub>multi</sub>; we identified 8 key proteins: AGRN, CD300C, CFB, GM2A, IL1RL1, INHBB, LILRA5, and TIMP4. These findings suggest that PRS<sub>multi</sub> and these proteins could be measured concurrently to predict COPD exacerbation risk. Developing an integrated prediction model and subsequent prospective validation are needed.

As drugs with genetically backed targets lead to approval rates twice that of non-genetically backed compounds, and drug repurposing agents achieve markedly higher approval rates (30% compared with 10% for de novo drugs), we used the genetically predicted proteins associated with exacerbations to identify drug repurposing candidates (28–30). Our analysis revealed 25 proteins with trials involving 46 drugs.

**Table 3. Druggable proteins associated with COPD exacerbations**

HGNC symbol	UniProt ID	Protein symbol	Number of traits in which criteria were met	OpenTargets drug
HGNC:320	Q15109	AGER (RAGE/sRAGE)	5	Azeliragon
HGNC:5998	Q01638	IL1RL1	5	Astegolimab
HGNC:19869	Q96GP6	SCARF2	5	
HGNC:1248	P06681	C2	4	
HGNC:3218	Q12805	EFEMP1	4	
HGNC:1037	P00751	CFB	3	Iptacopan
HGNC:4114	P22466	GAL	3	
HGNC:4367	P17900	GM2A	3	
HGNC:4453	P78333	GPC5	3	
HGNC:5344	P05362	ICAM1	3	Alicaforsen, BI-505
HGNC:5988	Q13478	IL18R1	3	Iboctadekin
HGNC:6168	Q06033	ITIH3	3	
HGNC:16936	Q86VZ4	LRP11	3	
HGNC:9618	Q13308	PTK7	3	
HGNC:11927	O43508	TNFSF12	3	BIIB-023, RO-5458640
HGNC:329	O00468	AGRN	2	
HGNC:1628	P08571	CD14	2	IC14, VB-201
HGNC:18219	Q9HCU0	CD248	2	Ontuxizumab
HGNC:11891	P05452	CLEC3B	2	
HGNC:2195	P39060	COL18A1	2	Collagenase, <i>Clostridium histolyticum</i> , ocriplasmin
HGNC:3386	P29317	EPHA2	2	Dasatinib, regorafenib, vandetanib
HGNC:3395	P54760	EPHB4	2	J1-101, tesevatinib, TG100-801, vandetanib
HGNC:5391	P35475	IDUA	2	
HGNC:6000	P18510	IL1RN	2	
HGNC:6563	P17931	LGALS3	2	Belapectin, davanat, olitigaltin
HGNC:9868	Q99969	RARRES2	2	
HGNC:8941	P01009	SERPINA1	2	
HGNC:11823	Q99727	TIMP4	2	

List of targetable proteins satisfying 3 criteria: (a) significant genetic association via S-PrediXcan, (b) measured protein level significantly associated with exacerbations in UKBB and COPDgene, and (c) concordant effect directions in both cohorts. Prioritization is based on the number of PRS<sub>multi</sub>-associated traits involved. The table also includes drug targets from OpenTargets. Only proteins for which 2 or more traits were implicated are shown. HGNC, HUGO Gene Nomenclature Committee.

Based on 5 out of 7 traits, 3 proteins were genetically predicted to affect COPD and exacerbation risk: IL1RL1, RAGE/sRAGE, and SCARF2. IL1RL1 is targeted by astegolimab, which is already in phase III trials for COPD. This finding offers a proof-of-concept that our method can highlight potential drug repurposing candidates for COPD exacerbations. *AGER* (RAGE/sRAGE), a highly replicable GWAS locus and well-established COPD biomarker (31), is expressed in alveolar epithelial cells and appears to have a broad role in regulating immunity and inflammation. The proteomic assays used quantify sRAGE rather than membrane-bound RAGE. RAGE in plasma arises from 2 sources: (a) the secretory isoform generated by alternative splicing, and (b) ectodomain shedding of RAGE by protease activity, which are not distinguished by the analyzed assays. Furthermore, as full-length RAGE can be cleaved by multiple proteases (e.g., MMP9, ADAM17), changes in protease pathways are not captured in the current study. RAGE/sRAGE has been implicated in several diseases, including glioblastoma multiforme (ClinicalTrials.gov NCT05986851), triple-negative breast cancer (32), and Alzheimer disease (33). Azeliragon is an antagonist of RAGE/sRAGE and was well tolerated in a phase III trial of Alzheimer disease, but the trial (ClinicalTrials.gov NCT02080364) was terminated due to lack of efficacy. The complicated issue in using azeliragon in COPD is that the rs2070600 C→T variant, associated with higher COPD risk, is associated with lower sRAGE levels, and lower sRAGE levels are associated with more emphysema (34). The effects of this variant on exacerbations are less clear, and further investigation needs to be done on which patients might benefit from repurposing azeliragon for COPD exacerbations.

SCARF2 polymorphisms have been implicated in COPD risk based on Mendelian randomization (35), and we now report that this genetically predicted protein is associated with exacerbations at the measured protein level in 2 cohorts. We did not identify any drug repurposing candidates in OpenTargets based on SCARF2, making this protein a potential novel therapeutic target for exacerbations.

Strengths of this study are that it combines many participants from both research and biobank cohorts with carefully defined COPD and COPD exacerbation criteria. The use of PRSmix<sup>+</sup> facilitated the identification of key traits contributing to genetic risk and the discovery of genetically predicted protein biomarkers. While individual S-PrediXcan studies have been published as stand-alone analyses (10), our study extends this framework by performing 7 proteome-wide association analyses and validating key signals with measured protein levels in 2 independent cohorts, which identified several promising biomarkers and therapeutic targets.

There are several limitations. PRSmix<sup>+</sup> studies typically aim to improve prediction, and while including more PRSs as inputs may enhance predictive power, we are approaching the upper bound of heritability for COPD, limiting further gains. The list of traits analyzed is not exhaustive, and translating findings from European to non-European cohorts remains a challenge. Cohort design limitations, such as recruiting only NHW and AA individuals in COPDGene, may also influence results. However, we found similar results in All of Us when including all groups, emphasizing the value of diverse representative cohorts. Definitions of COPD and exacerbations differ between research cohorts and biobanks, and parameters that would help reduce heterogeneity and improve uniformity of these definitions, such as exacerbation type (36) and lung function (in biobanks), was not readily available. However, the fact that the PRS<sub>multi</sub> can predict COPD and exacerbations across both research and biobank cohorts suggests that our genetic and proteomic signals generalize across case definitions rather than being confined to a single operationalization of COPD; that is, our findings support generalizability and transferability. We previously tested the PRS<sub>FEV1+FEV1/FVC</sub> in biobanks using validated machine learning phenotypes for COPD, and observed smaller effect sizes than we found in the current study (17). Future studies with additional phenotyping may allow increased effect size or specificity of our results.

Importantly, PRS<sub>multi</sub> associations with exacerbations were attenuated when adjusting for baseline FEV<sub>1</sub> in research cohorts, suggesting the effects are primarily driven by disease severity; by contrast, measured protein associations were adjusted for COPD status, suggesting that these genetically predicted protein targets are also important for disease activity and exacerbation risk. Finally, the integration of genetic and protein biomarkers, coupled with prospective validation and clinical implementation studies, is crucial to establish the practical utility and clinical relevance of these findings.

In conclusion, multi-trait PRSs for COPD and exacerbations identify genetic contributions to disease heterogeneity and druggable protein targets. These findings suggest that genetic risk prediction can be linked to specific therapeutic strategies for COPD precision medicine.

## Methods

Additional cohort details, genotyping information, and proteomic details are in the Supplemental Methods.

### Sex as a biological variable

Our study included both male and female participants and analyses were conducted jointly with sex included as an adjustment variable.

### Study cohorts

*COPDGene*. The COPDGene study (ClinicalTrials.gov NCT00608764) is a smoking-enriched cohort of self-identified NHW and AA participants aged 45–80 years with 10 or more pack-years of smoking history (37). COPDGene had enrollment and 5- and 10-year visits. Proteomic data were collected at visit 2 (5-year follow-up).

*ECLIPSE*. The ECLIPSE study (ClinicalTrials.gov NCT00292552) was a multicenter cohort study designed to explore biomarkers and clinical phenotypes of COPD (38). Participants include individuals with moderate to severe COPD (GOLD stages 2–4), aged 40–75 years, recruited across 12 countries.

*MGBB*. The MGBB is a large-scale biorepository linked to electronic health records (EHRs) of over 117,000 participants from the Mass General Brigham healthcare system since 2009. Participants aged 18 years or older were genotyped using the Illumina Global Screening Array.

*All of Us.* The All of Us Research Program is a population-scale biobank designed to advance precision medicine by collecting genetic, environmental, and health data from over 1 million participants in the United States. Whole-genome sequencing was conducted using the Illumina NovaSeq platform, as previously described (39). We utilized the allele call/allele frequency (ACAF) dataset for analyses.

## Statistics

*Overview of study design.* We integrated genetic data from COPDGene, ECLIPSE, MGBB, UKBB, and All of Us to develop and validate a PRS<sub>multi</sub> for both COPD and exacerbations. Following genotype imputation and PRS construction, we applied PRSmix<sup>+</sup> to combine trait-specific PRSs relevant to COPD susceptibility and exacerbation risk (40). We assessed associations with COPD case-control status and exacerbations. In COPDGene, NHW and AA participants were analyzed separately due to differences in demographic and disease characteristics, partly shaped by recruitment strategies. In other cohorts, racial/ethnic groups were analyzed together, adjusting for self-identified race and principal components (PCs) of genetic ancestry.

To identify underlying biological mechanisms, we integrated genetically predicted protein levels (via S-PrediXcan) with measured plasma proteomic data from COPDGene and UKBB to identify proteins linked to COPD and exacerbations (10). These candidate proteins were then cross-referenced with drug databases to highlight potential targets for therapeutic repurposing.

## Outcomes

### COPD

*COPD in research cohorts.* COPD was defined using the GOLD classification system. Two groups were identified: GOLD 2–4, representing those with moderate to very severe airflow limitation ( $FEV_1/FVC < 0.7$  and  $FEV_1 < 80\%$  predicted), and GOLD 0, referring to those with normal spirometry ( $FEV_1/FVC \geq 0.7$  and  $FEV_1 \geq 80\%$  predicted).

*COPD in biobanks.* For genetic analyses, we included individuals aged 40 years or older with available smoking data. In All of Us and MGBB, we excluded people with ICD codes for interstitial lung disease (J84) and heart failure (I50). The index date was defined as the date of the first COPD exacerbation or the first recorded observation for controls. COPD was defined by 1 or more inpatient code or 3 or more outpatient diagnostic codes within the past 3 years (41) to ensure identification of individuals with active disease. We identified COPD patients using ICD-9-CM codes 491.xx, 492.xx, and 496 and ICD-10-CM codes J41.x, J42, J43.x, and J44.x (42).

In UKBB, COPD case-control status was defined using spirometry to align with the definition in research cohorts (see below). A CONSORT diagram of participant inclusion and exclusion is shown in Supplemental Figure 7.

### COPD exacerbations

*Exacerbations in research cohorts.* In COPDGene, annualized exacerbation rates were calculated from the first study visit using data from the Longitudinal Follow-up Program, which was designed to continuously collect data on clinical outcomes, including exacerbations, comorbidities, and mortality, through surveys and electronic health records (43). As COPDGene was enriched for individuals who smoke, and individuals without COPD in this cohort experience respiratory events and increased mortality, we included these individuals in analyses (44). In ECLIPSE, exacerbation data were collected at each study visit using standardized questionnaires that recorded the number of exacerbations and use of steroids or antibiotics since the prior visit. As a result of study inclusion and exclusion criteria, nearly all participants with genetic data had had moderate-to-severe COPD.

In COPDGene, we additionally assessed quantitative CT measures of emphysema (volume-noise-bias-adjusted lung density) (45) and airway pathology (Pi10).

*Exacerbations in biobanks.* COPD exacerbation numbers were tracked for 1 year following the index date. Events were identified using ICD-9-CM codes 491.21 and 493.22 and ICD-10-CM codes J44.0 and J44.1. Multiple events within a 15-day period were considered a single exacerbation. For proteomic analysis, only exacerbations occurring after proteomic sampling were included in the analysis.

## Predictors

*Single-trait PRSs.* PRSs were computed using GWAS summary statistics or obtained from the Polygenic Score Catalog (46, 47). Selection of PRSs was based on biological relevance according to clinician input and literature review, available summary statistics, and the exclusion of testing cohorts from the original GWAS to avoid overfitting.

Details of the GWAS used for PRS construction are in Supplemental Table 2 and the Supplemental Methods. In cases where a Polygenic Score Catalog accession number is not listed, we developed our own PRS either in a prior study or for the current study. For the current study, we developed PRSs for emphysema, deep learning spirometry, peak expiratory flow, smoking, and eosinophils. Variant weights are available upon request.

*Development of a multi-trait PRS.* Prior to constructing a multi-trait PRS, we oriented directions of effects of individual PRSs such that a higher PRS is associated with higher COPD risk (e.g., higher PRS<sub>FEV<sub>1</sub></sub> is associated with higher COPD risk) since it is not clear what the expected effects are for all traits. Several methods exist for calculating multi-trait PRSs. A detailed description of PRSmix and PRSmix<sup>+</sup> (40) is in the Supplemental Methods.

We trained PRSmix<sup>+</sup> using COPDGene NHW participants, as most component PRSs were originally developed in populations with high genetic similarity to European reference panels. Models were trained to predict COPD, COPD exacerbations, P<sub>110</sub>, and emphysema, adjusting for age, sex, smoking pack-years, BMI, 5 genetic PCs, and CT scanner model (as appropriate). We selected the model that incorporated the largest number of PRSs and used the resulting weights to construct the final multi-trait PRS (PRS<sub>multi</sub>). All PRSs were centered and scaled prior to analysis, ensuring a mean of 0 and a standard deviation (SD) of 1.

## Testing of the multi-trait PRS

We tested the association of PRS<sub>multi</sub> with COPD and exacerbations using multivariable logistic regression models and negative binomial models, respectively. All models were adjusted for age, sex, smoking pack-years, BMI, and 5 genetic PCs. Exacerbation models included a log-offset for follow-up time. For comparison, we performed these same regression analyses using each individual component PRS included in PRS<sub>multi</sub>. To account for multiple comparisons, statistical significance was defined as a Benjamini-Hochberg-adjusted (BH-adjusted) *P* value of less than 0.05. As a sensitivity analysis, we further adjusted models for baseline FEV<sub>1</sub> in research cohorts. We performed area under the receiver operating characteristic curve (AUC) analyses, which are detailed in the Supplemental Methods.

We examined the multivariable associations of single-trait PRSs, the previously published PRS<sub>FEV<sub>1</sub>+FEV<sub>1</sub>/FVC</sub>, and the PRS<sub>multi</sub> with all-cause and respiratory mortality (Cox models), severe exacerbations requiring ER visit or hospitalization and antibiotic or steroid use (negative binomial models).

## Meta-analysis

Following multivariable regression analyses, we performed random effects meta-analyses using the meta R package for PRS<sub>multi</sub> and component PRSs (48). Heterogeneity was assessed using the *I*<sup>2</sup> statistic, and funnel plots were generated to visualize intercohort variability and assess selection bias.

## Protein prediction

We applied the S-PrediXcan framework, which leverages GWAS summary statistics to infer associations between predicted protein expression and complex traits or diseases, to evaluate genetically predicted protein levels. Summary statistics from the PRSs contributing to PRS<sub>multi</sub> (10) served as inputs. S-PrediXcan combines protein-wide association models with GWAS summary statistics, enabling the identification of protein-trait associations without requiring individual-level data. Prior to analysis, we harmonized GWAS variants to the GTEx v8 reference panel (49) and imputed summary statistics to address ambiguous or missing variants. We then applied multi-ancestry protein prediction models from the ARIC study (13) to estimate genetically regulated proteins with significantly altered levels.

For significant proteins (BH FDR-adjusted *P* < 0.1), we assessed whether directly measured protein levels were associated with COPD status and exacerbations in both COPDGene and UKBB. In COPDGene, models were adjusted for age, sex, race, pack-years of smoking, COPD case-control status, and log offset of time. In UKBB, analyses used propensity score matching on these variables, except the log offset of time since exacerbations were followed for 1 year, to account for case-control imbalance.

## Drug repurposing analysis

Based on the above analyses, we identified a list of targetable proteins that satisfied the following criteria: (a) significantly predicted to have altered levels based on GWAS/S-PrediXcan results, (b) significant association with COPD exacerbations, and (c) concordant direction of effect between UKBB and COPDGene. From the resulting list of proteins, we further prioritized proteins by the number of PRS<sub>multi</sub>-associated traits for which all 3 criteria were met. See the Supplemental Methods for further details.

## Data availability

All code used in this study is available at <https://github.com/CynthiaCZ/PRSmix.git>. Additional data are available in the supplemental material and Supporting Data Values file. Weights for newly developed PRSs and full S-PrediXcan results are available upon request.

## Author contributions

Study design: CZ, IRK, MHC, and MRM. Acquisition, analysis, or interpretation of the data: CZ, IRK, YH, JZ, YD, BP, QC, JALS, SK, KAP, RPB, MHC, and MRM. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: CZ, IRK, MHC, and MRM. Obtained funding: MHC, EKS, and MRM. Order of co-first authorship: CZ led and carried out the analyses, and CZ and IRK co-wrote the manuscript WBF, XH, DC, JH, MDT, JC, HKI, ESW, AM, LAL, VEO, TC, and ARM helped with interpretation of the data and critical manuscript revisions.

## Funding support

This work is the result of NIH funding, in whole or in part, and is subject to the NIH Public Access Policy. Through acceptance of this federal funding, the NIH has been given a right to make the work publicly available in PubMed Central.

- NIH grant K08 HL159318 (to MM).
- NIH grants R01 HL152728, P01 HL114501, U01 HL089856, R01 HL147148, and R01 HL133135 (to EKS).
- NIH grants R01 HL153248, R01 HL135142, R01 HL149861, R01 HL137927, and R01 HL137148 (to MHC).
- 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) to the COPDGene study.

## Acknowledgments

Address correspondence to: Matthew Moll, Channing Laboratory, 181 Longwood Avenue, Boston, Massachusetts, 02115, USA. Email: [remol@channing.harvard.edu](mailto:remol@channing.harvard.edu).

1. Mannino DM, et al. National and local direct medical cost burden of COPD in the United States from 2016 to 2019 and projections through 2029. *Chest*. 2024;165(5):1093–1106.
2. Fabbri LM, et al. COPD and multimorbidity: recognising and addressing a syndemic occurrence. *Lancet Respir Med*. 2023;11(11):1020–1034.
3. Zhou JJ, et al. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. *Am J Respir Crit Care Med*. 2013;188(8):941–947.
4. Shrine N, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019;51(3):481–493.
5. Moll M, et al. Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet Respir Med*. 2020;8(7):696–708.
6. Zhang J, et al. Genetically predicted body mass index and mortality in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2024;210(7):890–899.
7. He Y, et al. 2024. Multi-trait and multi-ancestry genetic analysis of comorbid lung diseases and traits improves genetic discovery and polygenic risk prediction. *Nat Genet*. 2026 Feb;58(2):289–298.
8. Wedzicha JA, Seemungal TAR. COPD exacerbations: defining their cause and prevention. *Lancet*. 2007;370(9589):786–796.
9. Lomenick B, et al. Identification of direct protein targets of small molecules. *ACS Chem Biol*. 2010;6(1):34–46.
10. Barbeira AN, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*. 2018;9(1):1825.
11. Donoghue LJ, et al. Integration of biobank-scale genetics and plasma proteomics reveals evidence for causal processes in asthma risk and heterogeneity. *Cell Genom*. 2025;5(5):100840.

12. Lee SH, et al. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol.* 2012;36(3):214–224.
13. Zhang J, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet.* 2022;54(5):593–602.
14. Ochoa D, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Res.* 2023;51(d1):D1353–D1359.
15. Castaldi PJ, et al. Heterogeneity and progression of chronic obstructive pulmonary disease: emphysema-predominant and non-emphysema-predominant disease. *Am J Epidemiol.* 2023;192(10):1647–1658.
16. Young KA, et al. Pulmonary subtypes exhibit differential global initiative for chronic obstructive lung disease spirometry stage progression: The COPDGene study. *Chronic Obstr Pulm Dis.* 2019;6(5):414–429.
17. Moll M, et al. Polygenic risk scores identify heterogeneity in asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol.* 2023;152(6):1423–1432.
18. Sakornsakolpat P, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet.* 2019;51(3):494–505.
19. Cho MH, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med.* 2014;2(3):214–225.
20. Marston NA, et al. Predictive utility of a coronary artery disease polygenic risk score in primary prevention. *JAMA Cardiol.* 2023;8(2):130–137.
21. Zhang J, et al. A polygenic risk score and age of diagnosis of chronic obstructive pulmonary disease. *Eur Respir J.* 2022;60(3):2101954.
22. Zhang J, et al. Polygenic risk score added to conventional case finding to identify undiagnosed chronic obstructive pulmonary disease. *JAMA.* 2025;333(9):784–792.
23. Jia Y, Jannink JL. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics.* 2012;192(4):1513–1522.
24. Do R, et al. Multitrait GWAS to connect disease variants and biological mechanisms. *PLoS Genet.* 2021;17(8):e1009713.
25. Sun BB, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature.* 2023;622(7982):329–338.
26. Xu Y, et al. An atlas of genetic scores to predict multi-omic traits. *Nature.* 2023;616(7955):123–131.
27. Hu X, et al. Polygenic transcriptome risk scores for COPD and lung function improve cross-ethnic portability of prediction in the NHLBI TOPMed program. *Am J Hum Genet.* 2022;109(5):857–870.
28. King EA, et al. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 2019;15(12):e1008489.
29. Nelson MR, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015;47(8):856–860.
30. Krishnamurthy N, et al. Drug repurposing: a systematic review on root causes, barriers and facilitators. *BMC Health Serv Res.* 2022;22(1):970.
31. Pratte KA, et al. Soluble receptor for advanced glycation end products (sRAGE) as a biomarker of COPD. *Respir Res.* 2021;22(1):127.
32. Magna M, et al. RAGE inhibitor TTP488 (azeliragon) suppresses metastasis in triple-negative breast cancer. *NPJ Breast Cancer.* 2023;9(1):59.
33. Burstein AH, et al. Development of azeliragon, an oral small molecule antagonist of the receptor for advanced glycation end-products, for the potential slowing of loss of cognition in mild alzheimer's disease. *J Prev Alzheimers Dis.* 2018;5(2):149–154.
34. Serveaux-Dancer M, et al. Pathological implications of receptor for advanced glycation end-product (AGER) gene polymorphism. *Dis Markers.* 2019;2019:2067353.
35. Wang S, et al. SCARF2 is a target for chronic obstructive pulmonary disease: evidence from multi-omics research and cohort validation. *Aging Cell.* 2024;23(10):e14266.
36. Celli BR, et al. An updated definition and severity classification of chronic obstructive pulmonary disease exacerbations: the rome proposal. *Am J Respir Crit Care Med.* 2021;204(11):1251–1258.
37. Regan EA, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD.* 2010;7(1):32–43.
38. Vestbo J, et al. Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur Respir J.* 2008;31(4):869–873.
39. Bick AG, et al. Genomic data in the All of Us research program. *Nature.* 2024;627(8003):340–346.
40. Truong B, et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genom.* 2024;4(4):100523.
41. Gershon AS, et al. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD.* 2009;6(5):388–394.
42. Feldman WB, et al. Comparative effectiveness and safety of generic versus brand-name fluticasone-salmeterol to treat chronic obstructive pulmonary disease. *Ann Intern Med.* 2023;176(8):1047–1056.
43. Stewart JI, et al. Automated telecommunication to obtain longitudinal follow-up in a multicenter cross-sectional COPD study. *COPD.* 2012;9(5):466–472.
44. Wan ES, et al. Association between preserved ratio impaired spirometry and clinical outcomes in US adults. *JAMA.* 2021;326(22):2287–2298.
45. Vegas Sánchez-Ferrero G, et al. Quantification of emphysema progression at CT using simultaneous volume, noise, and bias lung density correction. *Radiology.* 2024;310(1):e231632.
46. Lambert SA, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet.* 2021;53(4):420–425.
47. Lambert SA, et al. Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet.* 2024;56(10):1989–1994.
48. Balduzzi S, et al. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health.* 2019;22(4):153–160.
49. Aguet F, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318–1330.