Multi-Trait Polygenic Scores for COPD and COPD Exacerbations Implicate Druggable Proteins

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

Affiliations

[1] Channing Division of Network Medicine, Mass General Brigham, Boston, MA
[2] Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO
[3] Department of Epidemiology School of Public Health, UTHealth Houston, Houston, TX
[4] Pulmonary Center, Boston University School of Medicine, Boston, MA
[5] Harvard Medical School, Boston, MA
[6] Division of Pharmacoepidemiology and Pharmacoeconomics, Mass General Brigham, Boston, MA
[7] Department of Public Health Genomics, University of Virginia, Charlottesville, VA
[8] Department of Medicine, Dana-Farber Cancer Institute, Boston, MA
[9] Center for Computational Biomedicine, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[10] Genentech, South San Francisco, CA
[11] Department of Population Health Sciences, University of Leicester, Leicester, UK
[12] Division of Biostatistics and Bioinformatics, National Jewish Health, Denver, CO
[13] Department of Human Genetics, Department of Medicine, University of Chicago, Chicago, IL
[14] Section on Pulmonary, Critical Care, Allergy, and Sleep Medicine, VA Boston Healthcare System, West Roxbury, MA
[15] Public Health Genomics Program, University of Virginia, Charlottesville, VA
[16] Department of Genetic Medicine, Cleveland Clinic Lerner College of Medicine, Cleveland, OH
[17] Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Phoenix, AZ
[18] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; Program in Medical & Population Genetics, Broad Institute, Cambridge, MA

* These authors contributed equally and share first authorship
** These authors contributed equally and share last authorship

Correspondance:
Matthew Moll, MD, MPH
181 Longwood Avenue
Boston, MA 02115
remol@channing.harvard.edu

This article has an online data supplement.

ABSTRACT

Background. To construct multi-trait polygenic scores (PRS) predicting chronic obstructive pulmonary disease (COPD) and exacerbations, validate their performance in diverse cohorts, and identify PRS-related proteins for potential therapeutic targeting.

Methods. PRSmix+, a multi-trait PRS framework, is used to train a composite PRS (PRS_{multi}) in COPDGene non-Hispanic white participants (n=6,647). Associations of PRS_{multi} with COPD status (GOLD 2-4 vs. GOLD 0 or ICD) and exacerbation frequency were tested in COPDGene African American (n=2,466), ECLIPSE (n=1,858), MassGeneral 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_{multi} and were validated with proteomic data in COPDGene (n=5,173) and UK Biobank (n=5,012).

Results. PRSmix+ selected 7 traits for PRS_{multi}. In multivariable models, PRS_{multi} 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_{multi} outperformed traditional single-trait PRS in all tested cohorts. Using protein prediction models, we identified 73 proteins associated with the PRS that were also validated with measured protein levels in COPDGene 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. COPDGene: NCT00608764; ECLIPSE: 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 ~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 (PRS) for two measures of lung function (forced expiratory volume at one second: FEV_1 and the ratio of FEV_1 to forced vital capacity: FEV_1/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 COPD patients, 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). Further, 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 polygenic risk score and identification of proteins associated with exacerbations and genetic scores.

RESULTS

Characteristics of study participants

A total of 144,679 individuals were included in genetic analyses (**Table 1**), drawn from both COPD-enriched cohorts (COPDGene and ECLIPSE) and population-scale biobanks (MGBB and All of Us). As expected, smoking prevalence varied across cohorts.

Cohort characteristics for COPDGene and UKBB proteomic analyses are shown in **Table S1**. Participants had similar age, sex distribution, and spirometric measures. In COPDGene, individuals who were never smokers were excluded.

Development of multi-trait polygenic risk scores for COPD and related traits

We used 25 individual PRSs (**Table S2**), 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 (**Figure S1**).

We constructed multi-trait PRS models for COPD status (GOLD 2-4 vs. GOLD 0) and exacerbations, as well as CT measures of quantitative emphysema (adjusted lung density) and airway wall thickness (Pi10) using PRSmix+. Mixing weights for each model are displayed in **Table S3**. Final weighted models included between two and seven individual PRSs. Notably, PRSs for FEV_1/FVC (PRS_{ratio}) and a deep learning spirometry ($PRS_{DLspiro}$) trait contribute to all 4 traits. PRSs for FEV_1 (PRS_{FEV1}), BMI (PRS_{BMI}), and CRP (PRS_{CRP}) contribute to both Pi10 and COPD, and all seven PRSs, including smoking (cigarettes per day) (PRS_{smoke}) and idiopathic pulmonary fibrosis (PRS_{IPF}), contribute to COPD. Thus, while PRS_{ratio} is the largest contributor to COPD genetic risk, six additional PRSs also contribute to COPD risk.

All multi-trait PRSs showed strong associations with their respective training outcomes in multivariable models (**Table S4**). The PRS_{mix+} 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_{mix+} (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_{multi}. The PRS_{multi} explained 20.7% of the variance on the liability scale for COPD(12).

Testing the multi-trait polygenic risk score for COPD

We tested the predictive utility of PRS_{multi} for both COPD and COPD exacerbations compared with individual component PRSs. For COPD, PRS_{multi} consistently demonstrated larger absolute effect sizes in all cohorts except All of Us, where PRS_{BMI} showed the strongest effect (**Table S5**). For COPD exacerbations, PRS_{multi} outperformed other scores in COPDGene NHW and ECLIPSE, whereas PRS_{BMI} had higher effects in the remaining cohorts (**Table S6**). Notably, PRS_{multi} associations with COPD status remained significant after adjusting for baseline FEV_1 in COPDGene, but associations with exacerbations were attenuated in FEV_1-adjusted models.

In a random effects meta-analysis, PRS_{multi} was associated with increased odds of COPD (OR: 1.58 per SD (95% confidence interval (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_{IPF} and PRS_{BMI} were the only scores not significantly associated with COPD, and PRS_{multi} 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_{multi} was associated with an increase of 0.21 exacerbations per year per SD (95% CI: 0.11 - 0.31) (**Figure 3A**). However, this association

was attenuated in sensitivity analyses adjusting for baseline FEV_1. All component PRSs were significantly associated with exacerbations except for the PRS_{IPF} and PRS_{BMI} (**Figure 3B**). We note that PRS_{BMI} demonstrated the largest effects on exacerbations compared to other PRSs in some cohorts (**Table S6**), but it was not significantly associated with exacerbations in meta-analysis (**Figure S2**) due to a flipped direction of effect in COPDGene AA participants. Again, between-study heterogeneity was significant ($I^2 = 0.86$, $p < 0.0001$), though 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 ICD-based case definitions, respectively (**Figure S3, top**). In contrast, heterogeneity in exacerbation outcomes was lower and appeared to be influenced by outlier results from MGBB participants (**Figure S3, bottom**). In a leave-one-out sensitivity analysis excluding COPDGene NHW participants, we observed similar results for COPD and COPD exacerbations (**Figure S4**)

In AUC analyses among COPDGene NHW participants, PRS_{multi} outperformed PRS_{ratio} in predicting both COPD (AUC 0.67 vs 0.64, $p = 1.05 \times 10^{-8}$) and frequent exacerbations (≥ 2 per year) (AUC 0.569 vs 0.562, $p = 0.000001$). Both PRSs improve prediction when added to clinical covariates (**Table 2** & **Figure S5**). For COPD, the AUC increased from 0.756 (clinical only) to 0.8 with PRS_{multi} and 0.79 with the PRS_{ratio}. For exacerbation prediction, the AUC increased from 0.599 to 0.612 and 0.609, respectively. The PRS_{multi} also demonstrated higher predictive capacity for COPD and frequent exacerbations compared to any of the other tested single-trait PRSs or the previously published $PRS_{FEV1+FEV1/FVC}$ from Moll et al. (Table S9).

Examining other COPD-related outcomes, we found that a higher PRS_{BMI} was the PRS most associated with all-cause mortality, while the $PRS_{DLspiro}$ was most highly associated with

183 respiratory mortality (Table S10). However, the PRS_{multi} had the largest effects on severe
184 exacerbations and antibiotic or steroid use compared to other tested PRSs (Table S10).
185
186 *Genetically-predicted proteins based on shared genetic risks amongst traits*
187 Following the identification of seven traits contributing to COPD genetic risk, we investigated
188 shared genetic mechanisms by identifying proteins whose levels are predicted to be altered based
189 on the genetic architecture of these traits. Using GWAS summary statistics for each trait, we
190 applied S-PrediXcan(10) with ARIC PredictDB protein expression models(13) to infer genetically
191 regulated protein levels (full results available upon request). We then tested whether measured
192 plasma protein levels were associated with COPD exacerbations using multivariable negative
193 binomial regression models in COPDGene and UKBB. Models were adjusted for COPD case-
194 control status and other confounders (see **Methods**). We examined only proteins meeting FDR-
195 adjusted significance in S-PrediXcan, COPDGene SomaScan (measured protein levels), and
196 UKBB Olink (measured protein levels) data (**Table S7**). We further restricted the protein list to
197 those with concordant directions of effects in COPDGene and UKBB, resulting in a final set of 73
198 genetically predicted proteins associated with COPD exacerbations (**Table 3**). The number of
199 proteins at each filtering step for all 7 traits are detailed in Table S11.
200 We ranked these proteins by the number of traits for which all criteria were met. The receptor
201 for advanced glycation end-products and its soluble form (RAGE/sRAGE), interleukin 1 receptor-
202 like 1 (IL1RL1), and scavenger receptor class F member 2 (SCARF2) were represented in 5 out
203 of 7 tested traits. SERPINA1 was represented in two out of 7 tested traits. We performed gene
204 name lookups in the OpenTargets Platform(14) to identify potential drug repurposing candidates.
205 We identified 25 proteins with existing clinical trials for 46 drugs (**Table 3 and Table S8**).

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 eight proteins (Agrin (AGRN), CD300C, CFB, GM2 Ganglioside Activator (GM2A), IL1RL1, INHBB, Leukocyte Immunoglobulin Like Receptor A5 (LILRA5), TIMP4) that demonstrated consistent associations with COPD status, frequent exacerbations, and the highest vs. lowest quintile of PRS_{multi} in COPDGene NHW participants (**Figure S6**). Compared to IL1RL1, RAGE/sRAGE, and SCARF2 protein levels, PRS_{multi} demonstrated higher predictive capacity for frequent exacerbations (Figure S8; 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 multi-trait polygenic risk score (PRS_{multi}) for COPD and exacerbations. PRS_{multi} outperformed single-trait PRSs, the previously published $PRS_{FEV1+FEV1/FVC}$ from Moll et al., 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 seven component traits that share a genetic basis with COPD, including spirometry measures such as FEV_1/FVC and FEV_1, 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 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 seven traits contributing to the genetic risk of COPD, emphasizing its complex genetic architecture. Among these, three are spirometry-based, including a deep learning-derived spirometry phenotype, indicating that traditional measures like FEV_1 and FEV_1/FVC alone may not fully capture genetic risk or phenotypic variability. Our decision to use the COPD PRS_{mix+} 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_{mix+} 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_{mix+} (Pi10) 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_{mix+} models. The

reason for this result is unclear, but when tuned to quantitative emphysema, PRS_{ratio} and $PRS_{DLspiro}$ were selected, suggesting that the effects of emphysema might be largely captured by the genetics of low lung function.

The PRS_{BMI} was most associated with all-cause mortality which likely reflects higher cardiovascular risks in higher BMI individuals; as a corollary, it was the deep learning spirometry PRS that was actually most highly associated with respiratory mortality, suggesting that variants associated with low lung function beyond FEV_1 and FEV_1/FVC are important predictors of COPD mortality. The PRS_{multi} was also associated with severe exacerbations, suggesting that it could be used to enrich a population for severe exacerbators, though further validation is needed.

Although the PRS_{multi} outperformed single-trait PRSs and the previously published $PRS_{FEV1+FEV1/FVC}$ 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 seven traits, we used genetic protein prediction models derived from protein quantitative trait loci from the Atherosclerosis Risk in Communities (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 to a standard PRS(27).

Instead of focusing on a single trait, we applied a genetic protein prediction approach across seven traits and validated our findings using measured protein levels in two 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 (≥2 exacerbations per year), and individuals in the highest quintile of PRS_{multi}; we identified eight key proteins: AGRN, CD300C, CFB, GM2A, IL1RL1, INHBB, LILRA5, and TIMP4. These findings suggest that PRS_{multi} 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 to 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.

Based on 5 out of 7 traits, three 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 3 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

297 inflammation. The proteomic assays used quantify sRAGE rather than membrane-bound RAGE.
298 RAGE in plasma arises from two sources: 1) the secretory isoform generated by alternative
299 splicing, and 2) ectodomain shedding of RAGE by protease activity, which are not distinguished
300 by the analyzed assays. Further, as full-length RAGE can be cleaved by multiple proteases (e.g.,
301 MMP9, ADAM17), alterations in protease pathways are not captured in the current study.
302 RAGE/sRAGE has been implicated in several diseases, including glioblastoma multiforme
303 (NCT05986851), triple-negative breast cancer(32), and Alzheimer's disease(33). Azeliragon is an
304 antagonist of RAGE/sRAGE and was well tolerated in a phase III trial of Alzheimer's disease, but
305 the trial (NCT02080364) was terminated due to lack of efficacy. The complicated issue in using
306 azeliragon in COPD is that rs2070600 C→T variant, associated with higher COPD risk, is
307 associated with lower soluble RAGE (sRAGE) levels, and lower sRAGE levels are associated with
308 more emphysema(34). The effects of this variant on exacerbations are less clear, and further
309 investigation needs to be done on which patients might benefit from repurposing azeliragon for
310 COPD exacerbations.
311 SCARF2 polymorphisms have been implicated in COPD risk based on Mendelian
312 randomization(35), and we now report that this genetically predicted protein is associated with
313 exacerbations at the measured protein level in two cohorts. We did not identify any drug
314 repurposing candidates in OpenTargets based on SCARF2, making this protein a potential novel
315 therapeutic target for exacerbations.
316 Strengths of this study are that it combines many participants from both research and biobank
317 cohorts with carefully defined COPD and COPD exacerbation criteria. The use of PRSmix+
318 facilitated the identification of key traits contributing to genetic risk and the discovery of
319 genetically predicted protein biomarkers. While individual S-PrediXcan studies have been

published as stand-alone analyses(10), our study extends this framework by performing seven proteome-wide association analyses and validating key signals with measured protein levels in two independent cohorts, which identified several promising biomarkers and therapeutic targets.

There are several limitations. PRSmix+ 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 non-Hispanic white and African American 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_{multi} 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 $PRS_{FEV1+FEV1/FVC}$ 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_{multi} associations with exacerbations were attenuated when adjusting for baseline FEV_1 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

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 Genetic Epidemiology of COPD (COPDGene) study (ClinicalTrials.gov Identifier: NCT00608764) is a smoking-enriched cohort of self-identified non-Hispanic White (NHW) and African American (AA) participants aged 45-80 years with ≥ 10 pack-years of smoking history(37). COPDGene had enrollment, 5- and 10-year visits. Proteomic data were collected at visit 2 (5-year follow-up).

ECLIPSE. The Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study (ClinicalTrials.gov Identifier: 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 Mass General Brigham Biobank (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 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 one 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.

Additional cohort details, genotyping information, and proteomic details are in the **Supplementary Appendix**.

Statistical analysis

Overview of study design

We integrated genetic data from COPDGene, ECLIPSE, MGBB, UKBB, and All of Us to develop and validate a multi-trait PRS (PRS_{multi}) for both COPD and exacerbations. Following genotype imputation and PRS construction, we applied PRSmix+ 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 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. An overview of the study design is shown in **Figure 1**.

Outcomes

COPD

COPD in Research Cohorts. COPD was defined using the GOLD (Global Initiative for Chronic Obstructive Lung Disease) 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 ($FEV_1/FVC\geq0.7$ and $FEV_1 \geq$ 80% predicted).

COPD in Biobanks. For genetic analyses, we included individuals aged ≥40 years 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 inpatient code or ≥3 outpatient diagnostic codes within the past three years(41) to ensure identification of individuals with active disease. We identified COPD patients using ICD-9-CM codes 491.xx, 492.xx, 496 and ICD-10-CM codes J41.x, J42, J43.x, 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 Figure S7.

411

412 **COPD Exacerbations**

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

423 In COPDGene, we additionally assessed quantitative computed tomography (CT) measures of
424 emphysema (volume-noise-bias-adjusted lung density(45)) and airway pathology (Pi10: square
425 root of the wall area of a hypothetical airway with an internal perimeter of 10 millimeters).

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

431

432 *Predictors*

433 **Single-trait polygenic risk scores**

PRSs were computed using GWAS summary statistics or obtained from the PGS 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 Table S2 and the **Supplementary Appendix**. In cases where a PGS 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 polygenic risk score

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_{FEV1} 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+(40) is in the **Supplementary Appendix**.

We trained PRSmix+ using COPDGene non-Hispanic white 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, Pi10, and emphysema, adjusting for age, sex, smoking pack-years, BMI, five genetic principal components, 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_{multi}). All PRSs were centered and scaled prior to analysis, ensuring a mean of 0 and a standard deviation (SD) of 1.

457

458 *Testing of the multi-trait polygenic risk score*

459 We tested the association of PRS_{multi} with COPD and exacerbations using multivariable logistic
460 regression models and negative binomial models, respectively. All models were adjusted for age,
461 sex, smoking pack-years, BMI, and five genetic principal components. Exacerbation models
462 included a log-offset for follow-up time. For comparison, we performed these same regression
463 analyses using each individual component PRS included in PRS_{multi}. To account for multiple
464 comparisons, statistical significance was defined as a Benjamini-Hochberg (BH) adjusted p-value
465 less than 0.05. As a sensitivity analysis, we further adjusted models for baseline FEV_1 in research
466 cohorts. We performed area under the receiver operating characteristic curve (AUC) analyses,
467 which are detailed in the **Supplementary Appendix**.

468 We examined the multivariable associations of single-trait PRSs, the previously published
469 $PRS_{FEV1+FEV1/FVC}$, and the PRS_{multi} with all-cause and respiratory mortality (Cox models), severe
470 exacerbations requiring ER visit or hospitalization and antibiotic or steroid use (negative binomial
471 models)

472

473 *Meta-analysis*

474 Following multivariable regression analyses, we performed random effects meta-analyses using
475 the meta R package for PRS_{multi} and component PRSs(48). Heterogeneity was assessed using the
476 I^2 statistic, and funnel plots were generated to visualize inter-cohort variability and assess selection
477 bias.

478

479 *Protein prediction*

480 We applied the S-PrediXcan framework, which leverages GWAS summary statistics to infer
481 associations between predicted protein expression and complex traits or diseases, to evaluate
482 genetically predicted protein levels. Summary statistics from the PRSs contributing to PRS_{multi}(10)
483 served as inputs. S-PrediXcan combines protein-wide association models with GWAS summary
484 statistics, enabling the identification of protein-trait associations without requiring individual-level
485 data. Prior to analysis, we harmonized GWAS variants to the GTEx v8 reference panel(49) and
486 imputed summary statistics to address ambiguous or missing variants. We then applied multi-
487 ancestry protein prediction models from the Atherosclerosis Risk in Communities (ARIC)
488 study(13) to estimate genetically regulated proteins with significantly altered levels.
489 For significant proteins (BH FDR-adjusted p-value < 0.1), we assessed whether directly
490 measured protein levels were associated with COPD status and exacerbations in both COPDGene
491 and UK Biobank. In COPDGene, models were adjusted for age, sex, race, pack-years of smoking,
492 COPD case-control status, and log offset of time. In UK Biobank, analyses used propensity score
493 matching on these variables, except the log offset of time since exacerbations were followed for 1
494 year, to account for case-control imbalance.
495
496 *Drug repurposing analysis*
497 Based on the above analyses, we identified a list of targetable proteins that satisfied the following
498 criteria: 1) significantly predicted to have altered levels based on GWAS/S-PrediXcan results, 2)
499 significant association with COPD exacerbations, 3) concordant direction of effect between UK
500 Biobank and COPDGene. From the resulting list of proteins, we further prioritized proteins by the
501 number of PRS_{multi}-associated traits for which all three criteria were met. See the Supplementary
502 Appendix 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 data" and "Supporting data values". Weights for newly developed PRSs and full S-PrediXcan results are available upon request.

Author Contributions:

Study Design: Chengyue Zhang, Iain Konigsberg, Michael H. Cho, Matthew Moll

Acquisition, analysis, or interpretation of the data: Chengyue Zhang, Iain R. Konigsberg, Yixuan He, Jingzhou Zhang, Yi Ding, Bogdan Pasaniuc, Qingwen Chen, Jessica A. Lasky-Su, Sean Kalra, Katherine A. Pratte, Russell P. Bowler, Michael H. Cho, Matthew R. Moll

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

Statistical analysis: Chengyue Zhang, Iain Konigsberg, Michael H. Cho, Matthew Moll

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

Order of co-first authorship: Chengyue Zhang led and carried out the analyses, and Chengyue Zhang and Iain Konigsberg co-wrote the manuscript.

Funding:

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.

MM is supported by NIH grant K08HL159318. EKS is supported by NIH grants R01 HL152728, P01 HL114501, U01 HL089856, R01 HL147148, and R01 HL133135. MHC is supported by NIH

grants R01 HL153248, R01 HL135142, R01 HL149861, R01 HL137927, and R01 HL137148. 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.

Disclosures:

MM discloses consulting fees from 2ndMD, TheaHealth, Axon Advisors, Dialectica, Sanofi, Verona Pharma; grant support from Genentech; service as a medical expert; and honoraria from ATS 2024 and NYSTS 2024.

REFERENCES

1. Mannino DM, Roberts MH, Mapel DW, Zhang Q, Lunacsek O, Grabich S, 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):1093-106.
2. Fabbri LM, Celli BR, Agusti A, Criner GJ, Dransfield MT, Divo M, et al. COPD and multimorbidity: recognising and addressing a syndemic occurrence. *Lancet Respir Med.* 2023;11(11):1020-34.
3. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, and Laird NM. Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *American Journal of Respiratory and Critical Care Medicine.* 2013;188(8):941-7.
4. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet.* 2019;51(3):481-93.
5. Moll M, Sakornsakolpat P, Shrine N, Hobbs BD, DeMeo DL, John C, 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, Moll M, Hobbs BD, Bakke P, Regan EA, Xu H, et al. Genetically Predicted Body Mass Index and Mortality in Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine.* 2024;210(7):890-9.
7. He Y, Lu W, Jee YH, Wang Y, Tsuo K, Qian DC, et al. 2024.
8. Wedzicha JA, and Seemungal TAR. COPD exacerbations: defining their cause and prevention. *The Lancet.* 2007;370(9589):786-96.

9. Lomenick B, Olsen RW, and Huang J. Identification of Direct Protein Targets of Small Molecules. *ACS Chemical Biology.* 2010;6(1):34-46.

10. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications.* 2018;9(1).

11. Donoghue LJ, Benner C, Chang D, Irudayanathan FJ, Pendergrass RK, Yaspan BL, et al. Integration of biobank-scale genetics and plasma proteomics reveals evidence for causal processes in asthma risk and heterogeneity. *Cell Genomics.* 2025;5(5).

12. Lee SH, Goddard ME, Wray NR, and Visscher PM. A Better Coefficient of Determination for Genetic Profile Analysis. *Genetic Epidemiology.* 2012;36(3):214-24.

13. Zhang J, Dutta D, Köttgen A, Tin A, Schlosser P, Grams ME, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nature Genetics.* 2022;54(5):593-602.

14. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Research.* 2023;51(D1):D1353-D9.

15. Castaldi PJ, Xu Z, Young KA, Hokanson JE, Lynch DA, Humphries SM, et al. Heterogeneity and Progression of Chronic Obstructive Pulmonary Disease: Emphysema-Predominant and Non–Emphysema-Predominant Disease. *American Journal of Epidemiology.* 2023;192(10):1647-58.

16. Young KA, Strand MJ, Ragland MF, Kinney GL, Austin EE, Regan EA, et al. Pulmonary Subtypes Exhibit Differential Global Initiative for Chronic Obstructive Lung Disease

Spirometry Stage Progression: The COPDGene® Study. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation.* 2019;6(5):414-29.

17. Moll M, Sordillo JE, Ghosh AJ, Hayden LP, McDermott G, McGeachie MJ, et al. Polygenic risk scores identify heterogeneity in asthma and chronic obstructive pulmonary disease. *Journal of Allergy and Clinical Immunology.* 2023;152(6):1423-32.

18. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, 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, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med.* 2014;2(3):214-25.

20. Marston NA, Pirruccello JP, Melloni GEM, Koyama S, Kamanu FK, Weng L-C, et al. Predictive Utility of a Coronary Artery Disease Polygenic Risk Score in Primary Prevention. *JAMA Cardiology.* 2023;8(2).

21. Zhang J, Xu H, Qiao D, DeMeo DL, Silverman EK, O'Connor GT, et al. A Polygenic Risk Score and Age of Diagnosis of Chronic Obstructive Pulmonary Disease. *Eur Respir J.* 2022.

22. Zhang J, Hobbs BD, Silverman EK, Sparrow D, Ortega VE, Xu H, et al. Polygenic Risk Score Added to Conventional Case Finding to Identify Undiagnosed Chronic Obstructive Pulmonary Disease. *Jama.* 2025;333(9).

23. Jia Y, and Jannink J-L. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. *Genetics.* 2012;192(4):1513-22.

24. Do R, Julienne H, Laville V, McCaw ZR, He Z, Guillemot V, et al. Multitrait GWAS to connect disease variants and biological mechanisms. *PLOS Genetics.* 2021;17(8).

25. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature.* 2023;622(7982):329-38.

26. Xu Y, Ritchie SC, Liang Y, Timmers PRHJ, Pietzner M, Lannelongue L, et al. An atlas of genetic scores to predict multi-omic traits. *Nature.* 2023;616(7955):123-31.

27. Hu X, Qiao D, Kim W, Moll M, Balte PP, Lange LA, et al. Polygenic transcriptome risk scores for COPD and lung function improve cross-ethnic portability of prediction in the NHLBI TOPMed program. *The American Journal of Human Genetics.* 2022;109(5):857-70.

28. King EA, Davis JW, and Degner JF. 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, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015;47(8):856-60.

30. Krishnamurthy N, Grimshaw AA, Axson SA, Choe SH, and Miller JE. Drug repurposing: a systematic review on root causes, barriers and facilitators. *BMC Health Serv Res.* 2022;22(1):970.

31. Pratte KA, Curtis JL, Kechris K, Couper D, Cho MH, Silverman EK, et al. Soluble receptor for advanced glycation end products (sRAGE) as a biomarker of COPD. *Respir Res.* 2021;22(1):127.

624 32. Magna M, Hwang GH, McIntosh A, Drews-Elger K, Takabatake M, Ikeda A, et al.
625 RAGE inhibitor TTP488 (Azeliragon) suppresses metastasis in triple-negative breast
626 cancer. *npj Breast Cancer.* 2023;9(1).
627 33. Burstein AH, Sabbagh M, Andrews R, Valcarce C, Dunn I, and Altstiel L. Development
628 of Azeliragon, an Oral Small Molecule Antagonist of the Receptor for Advanced
629 Glycation Endproducts, for the Potential Slowing of Loss of Cognition in Mild
630 Alzheimer's Disease. *The Journal of Prevention of Alzheimer's Disease.* 2018;5(2):149-
631 54.
632 34. Serveaux-Dancer M, Jabaudon M, Creveaux I, Belville C, Blondonnet R, Gross C, et al.
633 Pathological Implications of Receptor for Advanced Glycation End-Product (AGER)
634 Gene Polymorphism. *Disease Markers.* 2019;2019:1-17.
635 35. Wang S, Yue Y, Wang X, Tan Y, and Zhang Q. SCARF2 is a target for chronic
636 obstructive pulmonary disease: Evidence from multi-omics research and cohort
637 validation. *Aging Cell.* 2024;23(10).
638 36. Celli BR, Fabbri LM, Aaron SD, Agusti A, Brook R, Criner GJ, et al. An Updated
639 Definition and Severity Classification of Chronic Obstructive Pulmonary Disease
640 Exacerbations: The Rome Proposal. *American Journal of Respiratory and Critical Care*
641 *Medicine.* 2021;204(11):1251-8.
642 37. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic
643 epidemiology of COPD (COPDGene) study design. *COPD.* 2010;7(1):32-43.
644 38. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of
645 COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *European*
646 *Respiratory Journal.* 2008;31(4):869-73.

39. Bick AG, Metcalf GA, Mayo KR, Lichtenstein L, Rura S, Carroll RJ, et al. Genomic data in the All of Us Research Program. *Nature.* 2024;627(8003):340-6.

40. Truong B, Hull LE, Ruan Y, Huang QQ, Hornsby W, Martin H, et al. Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics.* 2024;4(4).

41. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, and To T. Identifying Individuals with Physcian Diagnosed COPD in Health Administrative Databases. *COPD: Journal of Chronic Obstructive Pulmonary Disease.* 2009;6(5):388-94.

42. Feldman WB, Kesselheim AS, Avorn J, Russo M, and Wang SV. Comparative Effectiveness and Safety of Generic Versus Brand-Name Fluticasone–Salmeterol to Treat Chronic Obstructive Pulmonary Disease. *Annals of Internal Medicine.* 2023;176(8):1047-56.

43. Stewart JI, Moyle S, Criner GJ, Wilson C, Tanner R, Bowler RP, et al. Automated Telecommunication to Obtain Longitudinal Follow-up in a Multicenter Cross-sectional COPD Study. *COPD: Journal of Chronic Obstructive Pulmonary Disease.* 2012;9(5):466-72.

44. Wan ES, Balte P, Schwartz JE, Bhatt SP, Cassano PA, Couper D, et al. Association Between Preserved Ratio Impaired Spirometry and Clinical Outcomes in US Adults. *Jama.* 2021;326(22).

45. Vegas Sánchez-Ferrero G, Díaz AA, Ash SY, Baraghoshi D, Strand M, Crapo JD, et al. Quantification of Emphysema Progression at CT Using Simultaneous Volume, Noise, and Bias Lung Density Correction. *Radiology.* 2024;310(1).

669 46. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score
670 Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet.*
671 2021;53(4):420-5.
672 47. Lambert SA, Wingfield B, Gibson JT, Gil L, Ramachandran S, Yvon F, et al. Enhancing
673 the Polygenic Score Catalog with tools for score calculation and ancestry normalization.
674 *Nature Genetics.* 2024;56(10):1989-94.
675 48. Balduzzi S, Rücker G, and Schwarzer G. How to perform a meta-analysis with R: a
676 practical tutorial. *Evidence Based Mental Health.* 2019;22(4):153-60.
677 49. Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, et al. The GTEx
678 Consortium atlas of genetic regulatory effects across human tissues. *Science.*
679 2020;369(6509):1318-30.
680

FIGURE LEGENDS

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

Figure 2. PRS_{multi} is associated with COPD. A. Forest plot from a meta-analysis of the association between PRS_{multi} and COPD status across multiple cohorts. The figure presents odds ratios (ORs) with 95% confidence intervals (CIs) for each cohort, demonstrating the overall effect size and heterogeneity (I^2). **B.** Forest plot comparing the effects of individual PRSs and PRS_{multi} on COPD risk. PRS_{multi} 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.

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

Supplementary Figure 1. Heatmap showing hierarchical clustering of Pearson correlation coefficients between PRSs input into PRSmix+.

Supplementary Figure 2. Forest plot showing the random effects meta-analysis of multivariable negative binomial associations of PRS_{BMI} on exacerbations.

Supplementary Figure 3. Funnel plots of meta-analysis results for COPD and exacerbations. Funnel plots displaying meta-analysis results for (A) COPD and (B) COPD exacerbations.

Supplementary Figure 4. Leave-one-out meta-analysis for COPD (A) and exacerbations (B). Meta-analyses excluding COPDGene NHW participants were performed, as this cohort was used to tune PRS_{multi} weights and could be prone to overfitting.

Supplementary Figure 5. ROC curves for PRS-based models of COPD and exacerbations. Receiver operating characteristic (ROC) curves for models predicting (A) COPD and (B) exacerbations. Models include combinations of clinical covariates, PRS_{multi}, PRS_{ratio}, and genetic principal components.

Supplementary Figure 6. Protein Associations with COPD Outcomes. Of the 73 genetically-predicted proteins identified to be associated with COPD exacerbations, eight proteins were associated with COPD affection status (left), frequent exacerbations ($\geq$ 2 exacerbations/year vs. no exacerbations) (middle), and significantly differentially expressed between the top and bottom quintile of the PRS_{multi} (right) in COPDGene NHW participants. Boxplots for each association with interquartile ranges are displayed. P-values for Student t-tests are also displayed.

TABLE LEGENDS

727

728 **Table 1. Characteristics of study participants.** Baseline demographic and clinical characteristics
729 of participants across cohorts (COPDGene, ECLIPSE, MGBB, UKBB, and All of Us). COPDGene
730 = Genetic Epidemiology of COPD. ECLIPSE = Evaluation of COPD Longitudinally to Identify
731 Predictive Surrogate Endpoints. MGBB = MassGeneral Brigham Biobank. NHW = non-Hispanic
732 white. AA = African American.

733

734 **Table 2. Area-under-the-curve (AUC) analyses comparing PRS_{multi} to single PRSs for COPD**
735 **and exacerbations.** AUC comparisons for PRS_{multi} versus the best-performing single PRSs for
736 predicting COPD and frequent exacerbations (≥2 per year) in COPDGene NHW participants. The
737 table also reports AUC changes when adding PRS to a clinical model (age, sex, BMI, smoking
738 pack-years), with DeLong test p-values for model comparison.

739

740 **Table 3. Druggable proteins associated with COPD exacerbations.** List of targetable proteins
741 satisfying three criteria: (1) significant genetic association via S-PrediXcan, (2) measured protein
742 level significantly associated with exacerbations in UKBB and COPDGene, and (3) concordant
743 effect directions in both cohorts. Prioritization is based on the number of PRS_{multi}-associated traits
744 involved. The table also includes drug targets from OpenTargets. Only proteins for which 2 or
745 more traits were implicated are shown.

746

747 **Supplementary Table 1. Characteristics of study participants with proteomics data.**
748 Summary statistics for participants with available plasma proteomics data from COPDGene and
749 UKBB.

Supplementary Table 2. PRSs included in PRSmix+ analysis. List of PRSs used in PRSmix+ modeling, along with their source (GWAS, PGS Catalog, or newly developed) and references.

Supplementary Table 3. PRS-mix weights for COPDGene, weighted by different phenotypes. Summary of PRS-mix weights assigned to different traits in the construction of PRS_{multi}. This table presents the contribution of each PRS to the final model predicting COPD and exacerbations.

Supplementary Table 4. Association of PRSmix+ values with different weights in COPDGene. Results of multivariable models testing PRSmix+ scores with different weights for their association with COPD and exacerbations.

Supplementary Table 5. Multivariable associations of PRSs in PRSmix+ with COPD in each cohort. Logistic regression results for PRSs in PRSmix+ predicting COPD across individual cohorts. Blue shading indicates the highest performing PRS in a cohort.

Supplementary Table 6. Multivariable associations of PRSs in PRSmix+ with exacerbations in each cohort. Negative binomial regression results for PRSs in PRSmix+ predicting exacerbation frequency across individual cohorts. Blue shading indicates the highest performing PRS in a cohort.

Supplementary Table 7. S-PrediXcan proteins significantly associated with COPD and exacerbations. List of proteins with significant associations based on S-PrediXcan using GWAS

773 summary statistics for PRS_{multi} traits that were also significant in measuring proteomic association
774 analysis using COPDGene SomaScan and UKBB Olink data.
775
776 **Supplementary Table 8. Clinical trial information from OpenTargets for drugs targeting S-**
777 **PrediXcan proteins.** Summary of drugs targeting proteins identified via S-PrediXcan analysis.
778 Includes drug names, indications, and clinical trial phase information.
779

780 **Table 1. Characteristics of study participants.** Baseline demographic and clinical characteristics of participants across cohorts
781 (COPDGene, ECLIPSE, MGBB, UKBB, and All of Us). COPDGene = Genetic Epidemiology of COPD. ECLIPSE = Evaluation of
782 COPD Longitudinally to Identify Predictive Surrogate Endpoints. MGBB = MassGeneral Brigham Biobank. NHW = non-Hispanic
783 white. AA = African American.

Characteristic	*COPDGene NHW*	*COPDGene AA*	*ECLIPSE*	*All of Us*	*MGBB*
N	6647	2466	1858	118566	15142
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)
Sex (No. % Female)	3170 (47.7)	1045 (42.4)	625 (33.6)	71119 (60.0)	8918 (58.9)
smoking status (No. %)					
Never	0	0	0	75915 (64.9)	8489 (56.1)
Former	4045 (60.9)	529 (21.5)	1217 (65.5)	22650 (19.4)	5835 (38.5)
Current	2602 (39.1)	1937 (78.5)	641 (34.5)	18486 (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)	
FEV1 % predicted (mean (sd))	77.85 (27.49)	74.15 (23.19)	52.20 (22.53)		
FEV1/FVC (mean (sd))	0.64 (0.17)	0.71 (0.15)	0.47 (0.15)		
COPD (No. % Cases)	2634 (48.6)	803 (40.6)	1711 (92.1)	3980 (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)

784

785 **Table 2. Area-under-the-curve (AUC) analyses comparing PRS_{multi} to single PRSs for COPD and exacerbations.** AUC
786 comparisons for PRS_{multi} versus the best-performing single PRSs for predicting COPD and frequent exacerbations (≥2 per year) in
787 COPDGene NHW participants. The table also reports AUC changes when adding PRS to a clinical model (age, sex, BMI, smoking
788 pack-years), with DeLong test p-values for model comparison.

Model	*COPD AUC*	*Frequent Exacerbations AUC*
Clinical model	0.756	0.599
PRS_{multi} + PCs	0.666	0.569
Clinical + PRS_{multi} + PCs	0.8	0.612
PRS_{ratio} + PCs	0.643	0.562
Clinical + PRS_{ratio} + PCs	0.79	0.609

COPD AUC p-value [PRS_{multi} + PCs vs. PRS_{ratio} + PCs] = 1.05e-08
COPD AUC p-value [Clinical + PRS_{multi} + PCs vs. Clinical + PRS_{ratio} + PCs] = 1.35e-06
COPD AUC p-value [Clinical + PRS_{multi} + PCs vs. Clinical] = 5.35e-26
COPD AUC p-value [Clinical + PRS_{ratio} + PCs vs. Clinical] = 1.29e-19

Frequent exacerbations AUC pvalue [PRS_{multi} + PCs vs. PRS_{ratio} + PCs] = 1.09e-06
Frequent exacerbations AUC pvalue [Clinical + PRS_{multi} + PCs vs. Clinical + PRS_{ratio} + PCs] = 0.00138
Frequent exacerbations AUC pvalue [Clinical + PRS_{multi} + PCs vs. Clinical] = 6.13e-09
Frequent exacerbations AUC pvalue [Clinical + PRS_{ratio} + PCs vs. Clinical] = 2.75e-07
789

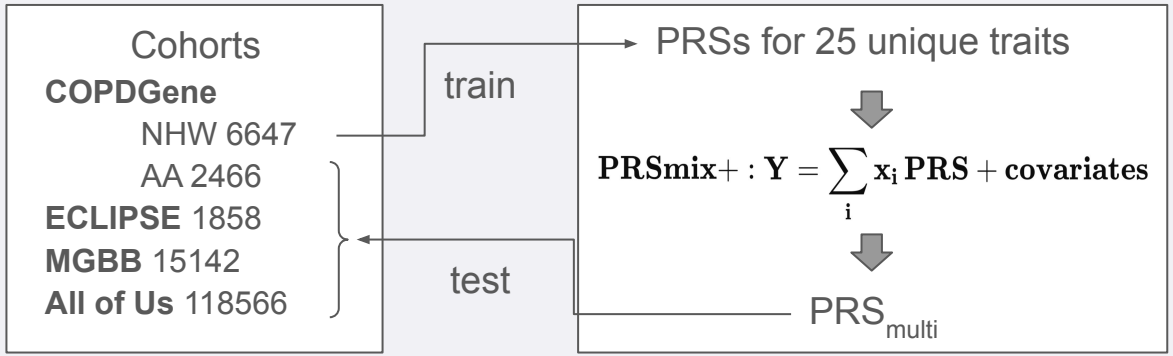
790 **Table 3. Druggable proteins associated with COPD exacerbations.** List of targetable proteins
791 satisfying three criteria: (1) significant genetic association via S-PrediXcan, (2) measured protein
792 level significantly associated with exacerbations in UKBB and COPDGene, and (3) concordant
793 effect directions in both cohorts. Prioritization is based on the number of PRS_{multi}-associated
794 traits involved. The table also includes drug targets from OpenTargets. Only proteins for which 2
795 or more traits were implicated are shown.

HGNC symbol	*UniProt ID*	*Protein name*	*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 CLOSTRIDIUM HISTOLYTICUM, OCRIPLASMIN
HGNC:3386	P29317	EPHA2	2	DASATINIB, REGORAFENIB, VANDETANIB
HGNC:3395	P54760	EPHB4	2	JI-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	

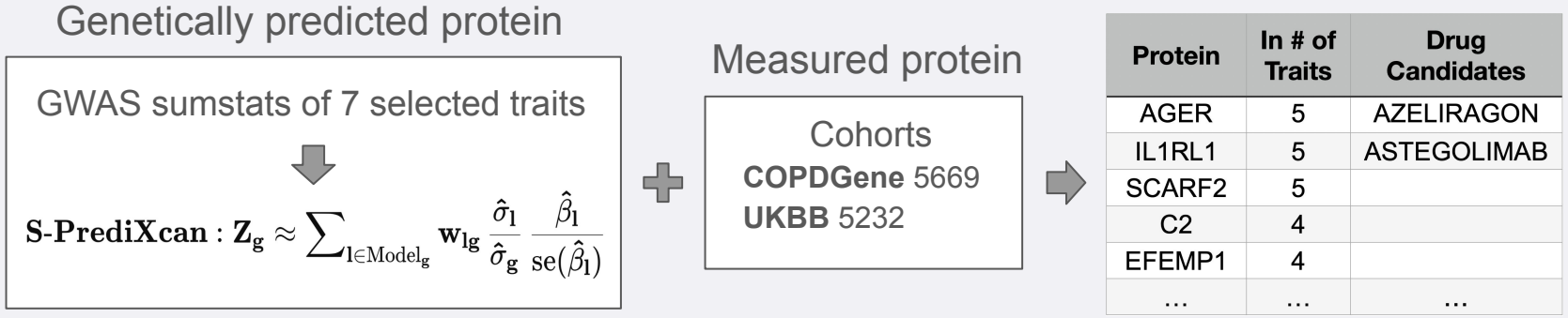
796

1. Construction of PRS_{multi}



PRS	Weights
FEV1/FVC	0.245
FEV1	0.178
DL spiro	0.107
Smoking	0.0526
IPF	0.0497
BMI	0.0466
CRP	0.017

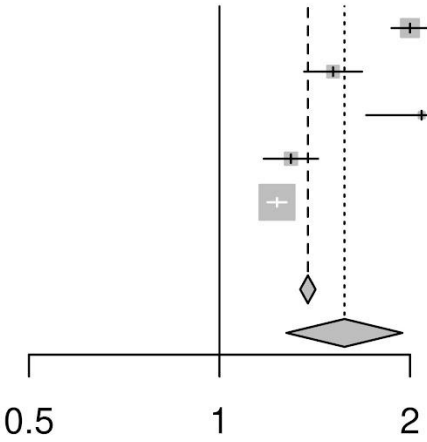
2. Identification of Druggable Proteins



Protein	In # of Traits	Drug Candidates
AGER	5	AZELIRAGON
IL1RL1	5	ASTEGOLIMAB
SCARF2	5	
C2	4	
EFEMP1	4	
…	…	…

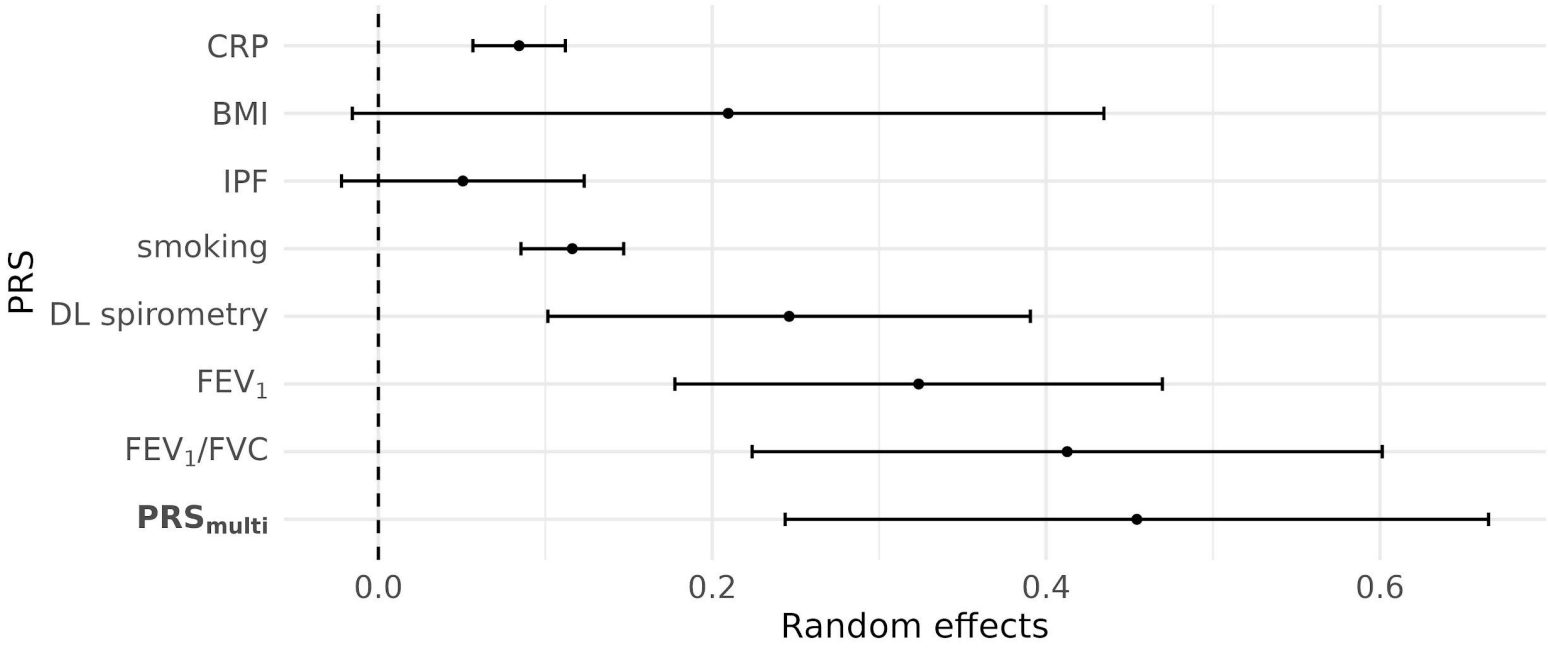
A)

Study	logOR	SE(logOR)	OR	95%-CI
COPDGene NHW	0.6929	0.0343	2.00	[1.87; 2.14]
COPDGene AA	0.4131	0.0536	1.51	[1.36; 1.68]
ECLIPSE	0.7356	0.1029	2.09	[1.71; 2.55]
MGBB	0.2599	0.0504	1.30	[1.17; 1.43]
All of Us	0.2092	0.0176	1.23	[1.19; 1.28]
Common effect model			**1.38**	**[1.34; 1.42]**
Random effects model			**1.58**	**[1.28; 1.94]**

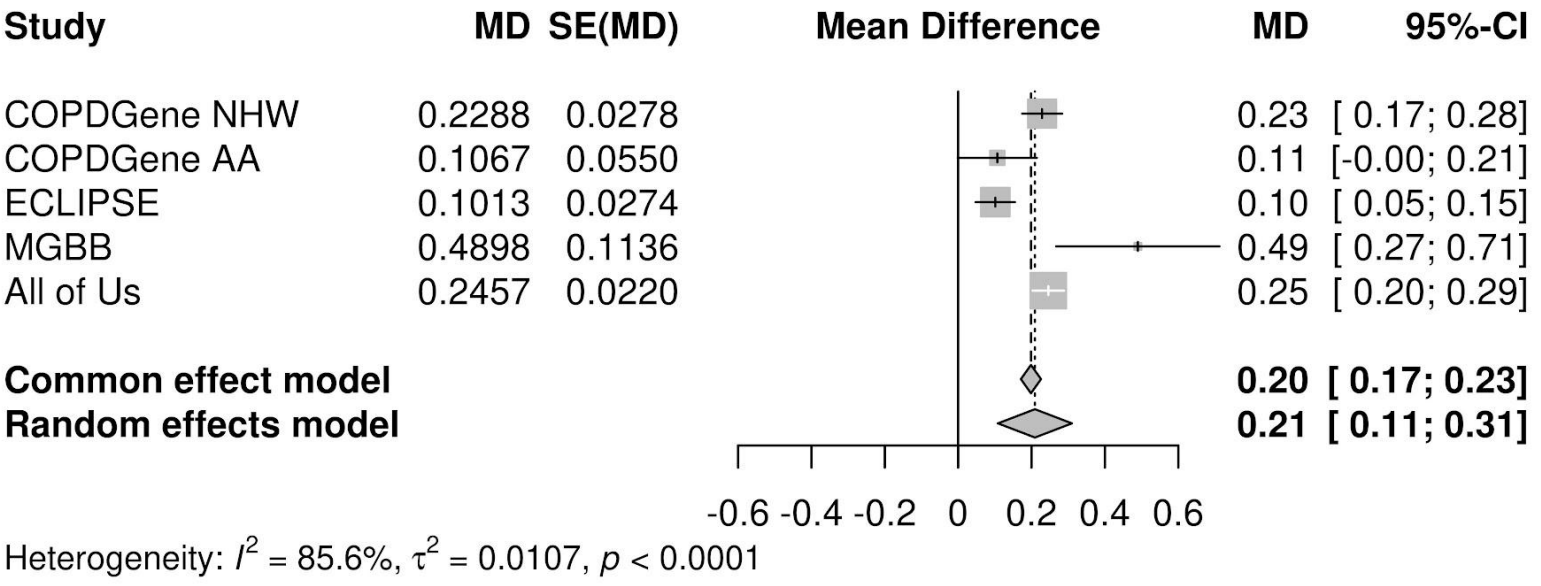


Heterogeneity: I^2 = 97.8%, τ^2 = 0.0545, p < 0.0001

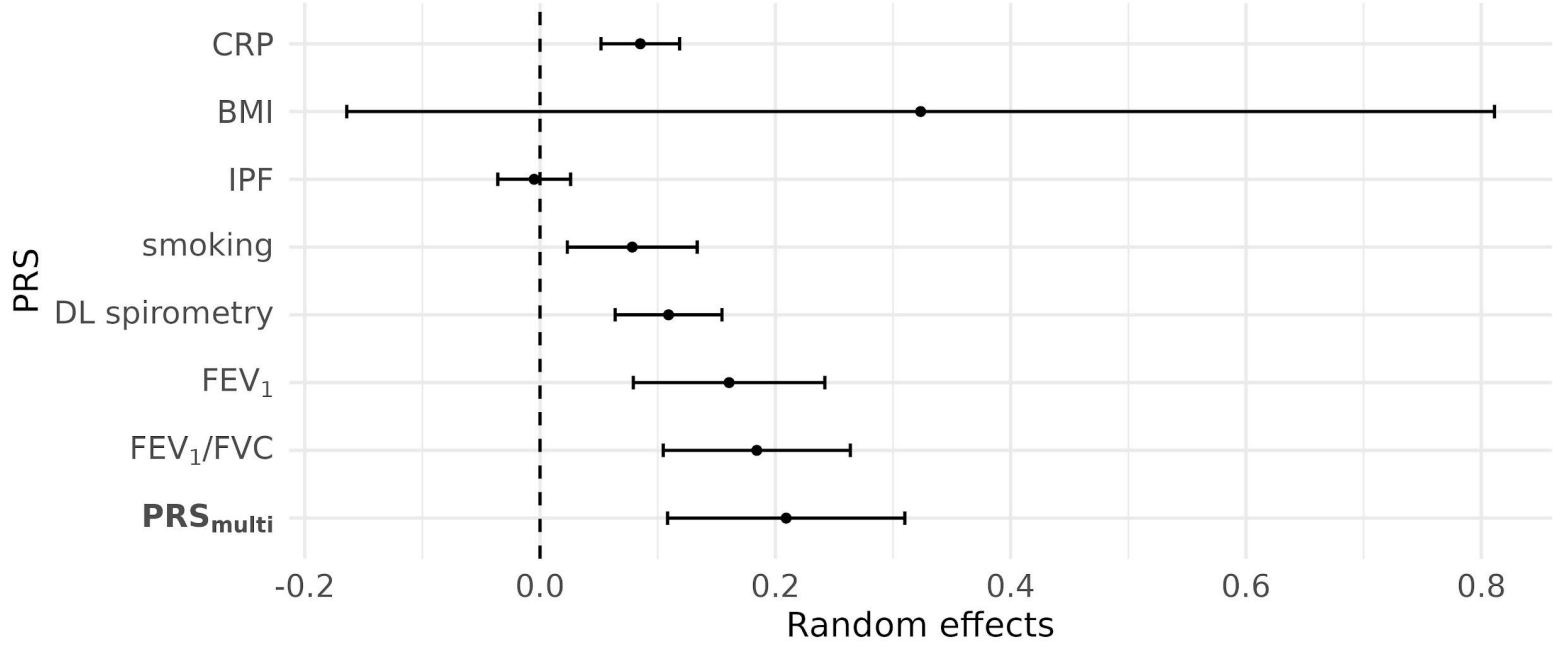
B)



A)


Study
MD
SE(MD)
Mean Difference
MD
95%-CI
COPDGene NHW 0.2288 0.0278 0.23 [0.17; 0.28]
COPDGene AA 0.1067 0.0550 0.11 [-0.00; 0.21]
ECLIPSE 0.1013 0.0274 0.10 [0.05; 0.15]
MGBB 0.4898 0.1136 0.49 [0.27; 0.71]
All of Us 0.2457 0.0220 0.25 [0.20; 0.29]
Common effect model 0.20 [0.17; 0.23]
Random effects model 0.21 [0.11; 0.31]
-0.6 -0.4 -0.2 0 0.2 0.4 0.6
Heterogeneity: I^2 = 85.6%, τ^2 = 0.0107, p < 0.0001

B)


CRP
BMI
IPF
smoking
DL spirometry
FEV1
FEV1/FVC
PRSmulti
PRS
-0.2
0.0
0.2
0.4
0.6
0.8
Random effects