Society of Medical Education & Research

2026, Volume 6, Issue 1, Page No: 1-17
ISSN: 3108-4826

Journal of Medical Sciences and Interdisciplinary Research

Multi-Omics Risk Stratification Identifies Clinically Distinct COPD Subtypes and Therapeutic Opportunities

Pieter Botha[1]*, Anel Van Wyk[1], Johan Marais[2]

[1] Department of Pulmonology and Multi-Omics, Faculty of Medicine, Stellenbosch University, Stellenbosch, South Africa.
[2] Department of COPD Subtyping and Personalized Therapy, Faculty of Medicine, University of Pretoria, Pretoria, South Africa.

***E-mail** ✉ pieter.botha@sun.ac.za

Abstract

Genetic factors and gene expression patterns are recognized predictors of susceptibility to chronic obstructive pulmonary disease (COPD). Nevertheless, the extent to which these elements shape the diverse clinical presentations of COPD is not well established. This investigation aimed to identify high-risk COPD subtypes by combining genetic risk assessment via polygenic risk score (PRS) with blood-based transcriptional risk score (TRS), and to examine variations in their associated clinical and molecular profiles. High-risk categories were constructed using quantiles derived from PRS and TRS distributions, selecting thresholds that produced the greatest separation in protein biomarker levels within the COPDGene training dataset. These definitions were then applied to independent COPDGene and ECLIPSE validation cohorts. Associations between subgroup membership and clinical endpoints were analyzed using multivariable models. Protein–protein interaction networks and potential drug repurposing strategies were also contrasted across the high-risk categories. Analyses focused on two omics-derived high-risk groups within separate test cohorts (n = 1133 non-Hispanic White participants from COPDGene, n = 299 African American participants from COPDGene, and n = 468 from ECLIPSE). The subgroups were labeled “high activity” (low PRS combined with high TRS) and “severe risk” (high PRS combined with high TRS). Relative to a low-risk reference group (low PRS and low TRS), participants in both high-risk subgroups had reduced body-mass index (BMI), impaired lung function, and dysregulated pathways involved in metabolism, growth, and immune responses. The “high activity” subgroup alone exhibited more rapid longitudinal decline in FEV1 (COPDGene: −51 mL/year; ECLIPSE: −40 mL/year). Proteomic signatures in this subgroup were notably enriched for gene sets influenced by 5-lipoxygenase inhibitors and angiotensin-converting enzyme (ACE) inhibitors. Considering polygenic and transcriptional risk scores together delineated clinically and molecularly distinct subsets within the high-risk COPD population. Proteomic characterization and drug-repurposing modeling highlighted therapy enrichments that differed by subtype, providing a plausible explanation for why earlier drug-repurposing initiatives may have failed due to insufficient consideration of patient heterogeneity.

Keywords: Polygenic risk scores, COPD, Transcriptomics, Endotyping, Drug repurposing

Introduction

Chronic obstructive pulmonary disease (COPD) ranks among the primary contributors to illness and death on a global scale [1]. Although defined primarily by fixed airflow limitation, the disorder varies widely among patients in terms of the degree of emphysema, airway abnormalities, exacerbation frequency, and the trajectory of lung function loss [2, 3]. Pinpointing individuals most likely to experience rapid disease progression or advance to severe stages is vital for tailoring therapeutic interventions to individual needs.

Extensive omics resources create opportunities to detect, using a routine blood sample, groups at elevated risk who share underlying disease processes amenable to intervention. Polygenic risk scores (PRSs), which summarise genetic liability, have successfully flagged

Access this article online https://smerpub.com/

Received: 26 September 2025; **Accepted:** 04 January 2026

How to cite this article: Botha P, Van Wyk A, Marais J. Multi-Omics Risk Stratification Identifies Clinically Distinct COPD Subtypes and Therapeutic Opportunities. J Med Sci Interdiscip Res. 2026;6(1):1-17. https://doi.org/10.51847/8c2Uv6ky5f

people at heightened susceptibility to coronary artery disease, facilitating earlier statin use than suggested by conventional criteria [4]. In oncology, merging genetic and transcriptomic data has improved treatment guidance and prognosis [5]. Our earlier research showed that both a PRS and a transcriptional risk score (TRS) independently anticipate COPD development [6, 7]. The TRS yielded an odds ratio of 3.3 and an area under the curve of 0.79 when predicting COPD [7]. Although both instruments rely on spirometric parameters, they show little correlation [7] and appear to reflect separate dimensions of lung pathology. COPD-related genetic variants are notably enriched in pathways governing lung development and show stronger influence in early disease phases [6, 8]. At the same time, the TRS aligns with inflammatory indicators and progressive functional impairment, possibly signaling active disease processes and vulnerability to further deterioration [7]. This complementarity suggests that combining PRS and TRS information could help distinguish clinically and biologically discrete COPD subtypes.

Although omics-based prediction models have advanced, several practical issues for real-world implementation persist. These scores are generally adjusted for statistical modeling, which complicates decisions about how best to apply them to classify individual risk [9]. As continuous traits that often follow a normal distribution, they pose challenges for subtype definition along a gradient—a difficulty noted in earlier work [10]. Even so, categorization is required to associate omics-highlighted high-risk populations—who stand to gain from targeted treatments—with particular mechanistic pathways and clinical management strategies. New and repurposed medications for COPD frequently show disappointing results in trials [11], yet whether these shortcomings arise from limitations in selecting appropriate participants, particularly without omics or biomarker guidance, remains unresolved.

While the PRS performed well in forecasting COPD severity and new cases, the TRS excelled at predicting FEV1 decline. Their joint use may therefore reveal subgroups susceptible to several key adverse COPD outcomes. Accordingly, we proposed that our established PRS and TRS—both grounded in spirometry—could delineate subtypes reflecting heterogeneity within high-risk ever-smoker populations across two cohorts, manifesting as differences in clinical presentation and biological features. The study objectives included devising a method to deploy omics risk scores in population contexts and using them to advance precision medicine. Proteomics provided an independent layer of biological characterization for the resulting subgroups, and computational drug repurposing analyses identified therapy candidates with potential subgroup selectivity.

Materials and Methods

Study Populations

COPDGene

The analysis included participants from the Genetic Epidemiology of COPD (COPDGene) study (ClinicalTrials.gov Identifier: NCT00608764) with available single nucleotide polymorphism (SNP) genotyping, RNA-sequencing, and SomaScan proteomic measurements to enable PRS calculation, TRS derivation, and assessment of differential protein expression. In summary, COPDGene enrolled 10,198 non-Hispanic White (NHW) and African American (AA) adults between 45 and 80 years of age with a minimum smoking exposure of 10 pack-years [12]. Originally structured as a cross-sectional case–control investigation, the study evolved into a longitudinal effort with evaluations at 5- and 10-year intervals. Each visit involved comprehensive anthropometric assessments, spirometric testing, and computed tomography (CT) scans. We performed baseline genotyping, with RNA sequencing and proteomic profiling at the 5-year follow-up.

We used the Illumina (San Diego, CA) HumanOmniExpress platform for SNP genotyping. Separate testing for Z and S alpha-1 antitrypsin alleles excluded participants with severe deficiency states. Imputation relied on the Michigan Imputation Server, drawing from the Haplotype Reference Consortium [13] and 1000 Genomes Phase I v3 Cosmopolitan panels for non-Hispanic White and African American subjects, respectively. Variants exhibiting $r^2 \leq 0.3$ were discarded.

ECLIPSE

Consistent with prior reports,7 the present analysis included participants from the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) study (ClinicalTrials.gov Identifier: NCT00292552) with SNP genotyping, whole-blood microarray expression data, and at least two FEV1 recordings. The ECLIPSE cohort comprised 2,140 individuals diagnosed with COPD, aged 40–75 years, each with at least 10 pack-years of smoking history [14].

Initial data collection encompassed anthropometric, spirometric, and CT imaging variables. Blood specimens were drawn at entry, with concurrent genotype and microarray expression data available for selected participants. Those fulfilling Global Initiative for Chronic Obstructive Lung Disease (GOLD) [15] stage 2–4 criteria at baseline attended semiannual spirometry visits over three years.

SNP genotyping in ECLIPSE utilized the Illumina HumanHap 550 V3 array (Illumina, San Diego, CA). We removed individuals and variants with call rates below 95%. Imputation was executed on the Michigan Imputation Server utilizing the Haplotype Reference Consortium [13] panel.

Cohort expression data

COPDGene: RNA sequencing data

Whole blood specimens were collected and preserved in PAXgene Blood RNA tubes during the 5-year follow-up assessment of the COPDGene study. Earlier reports outline the procedures for acquiring and processing RNA-sequencing data [7, 16]. In short, total RNA was isolated using the Qiagen PreAnalytiX PAXgene Blood miRNA Kit (Qiagen, Valencia, CA). After quality control, samples underwent globin depletion before cDNA library construction. Sequencing generated 75 bp reads, achieving an average depth of 20 million reads per sample on an Illumina HiSeq 2500 instrument. We filtered raw count data to retain transcripts with counts exceeding 1 per million (CPM) in at least 99% of samples. We normalized data using log-CPM transformation with the edgeR package in R [17]. We adjusted library size, then corrected batch effects using the removeBatchEffects function from the limma package [18].

ECLIPSE: Microarray data

We collected blood samples from ECLIPSE participants at study entry. We isolated total RNA with PAXgene Blood miRNA kits and hybridized it to the Affymetrix Human Gene 1.1 ST array. When multiple probes corresponded to the same transcript, we selected the probe with the largest interquartile range. Batch effects were mitigated using the removeBatchEffects function in limma [18]. Additional information on RNA data handling and preprocessing is available in our previous publication [7].

Before downstream analyses, we restricted transcripts to those common to both datasets, based on HGNC gene symbols. For the ECLIPSE microarray results, we retained the probe with the highest interquartile range when multiple probes were present per gene. We standardized both RNA-sequencing count data and microarray expression values to a mean of 0 and a standard deviation of 1.

Proteomic data: COPDGene

Plasma proteomic profiling was conducted with the SomaScan v4.0 assay, which employs aptamers (SOMAmers) to measure 4,776 distinct human proteins. On the SomaScan 5K platform, plate hybridization was followed by median signal normalization, plate scaling, and SOMAmer calibration to minimize technical variation arising from array signals, run-to-run differences, inter-assay variability, and plate-to-plate batch effects. We have previously reported comprehensive descriptions of SomaScan data generation and preprocessing [19].

Polygenic and transcriptional risk scores

Both the COPD PRS and TRS were constructed from spirometric traits and have been detailed in prior publications [6, 7].

Polygenic risk score

A PRS for COPD was previously developed6 based on genome-wide association studies (GWAS) of FEV1 and FEV1/FVC involving roughly 500,000 participants from the UK Biobank and SpiroMeta consortium [20]. We computed separate PRSs for FEV1 and FEV1/FVC using lassosum [21], a shrinkage regression method that reduces multicollinearity, performs variable selection, and adjusts for linkage disequilibrium. We then combined these two component scores into an overall risk score, consistent with earlier methodology [6]. Neither COPDGene nor ECLIPSE contributed to PRS construction; both served as independent validation cohorts.

To optimize separation by genetic risk and reduce confounding from population structure, we examined non-Hispanic White (NHW) and African American (AA) individuals separately. We regressed out principal components reflecting genetic ancestry before applying the PRS.

Transcriptional risk score

The TRS was originally developed [7] using least absolute shrinkage and selection operator (LASSO)-

penalized regression [22] on a training set of 1,374 COPDGene participants, with performance evaluated in an independent held-out group of 674 individuals. Subsequent RNA-sequencing data from TOPMed Freeze 4 became available for an additional 459 NHW and 143 AA participants, who were incorporated into the COPDGene testing set. We ensured that no training samples were included in the current COPDGene test cohort.

Statistics

Overview of study design

To delineate high-risk subgroups from continuous risk metrics, we used the previously established COPDGene training cohort.7 Different combinations of PRS and TRS quantiles were evaluated to identify partitions that maximized the detection of differentially expressed proteins across the resulting subtype categories (**Figure 1**). Only participants of European ancestry were included in the training phase, consistent with the ancestry used in PRS development. We then applied raw (non-standardized) score thresholds corresponding to the selected percentile cut-offs from the training set to assign individuals in the COPDGene test set and the independent ECLIPSE validation cohort to specific omics-defined subtypes. We then characterized these newly classified subtypes through proteomic network analysis and drug repurposing modeling. We used multivariable linear regression models to evaluate associations between subtype membership and COPD-related clinical outcomes.

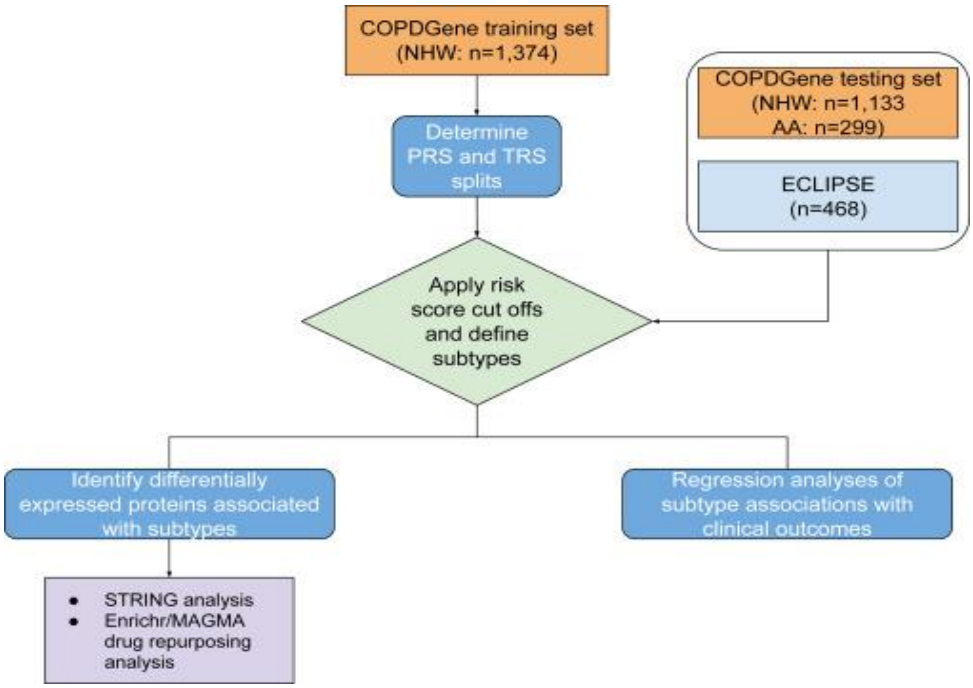

Figure 1. Schematic overview of the study design. COPD, chronic obstructive pulmonary disease. COPDGene, Genetic Epidemiology of COPD study. ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints study. PRS, polygenic risk score. TRS, transcriptional risk score. STRING, Search Tool for the Retrieval of Interacting Genes/Proteins. MAGMA, Multi-marker Analysis of GenoMic Annotation.

Determining risk score divisions and identifying omics-defined subtypes

Omics-derived risk scores are routinely standardized for statistical modeling, resulting in a normal distribution that inherently limits their suitability for conventional clustering methods. Nevertheless, clinical translation requires assignment of patients to discrete categories. To assess whether one or multiple clusters would be most appropriate, we computed the gap statistic for the PRS and TRS using the clusGap function from the cluster package in R, specifying up to 8 k-means clusters and 500 bootstrap replications. This statistic evaluates within-cluster dispersion relative to what would be expected under a null reference distribution [23]; the "gap"

represents the difference between observed and expected dispersion, and the largest gap across varying cluster numbers identifies the optimal cluster solution.

As a practical alternative to clustering, participants are frequently assigned to quantiles of omics risk scores, with each quantile compared against the lowest-risk reference group [4, 6, 24]. Extending this strategy to two independent omics scores introduces added complexity: fewer divisions yield larger subgroups and greater statistical power, whereas finer divisions enable more extreme contrasts. Because genetic and transcriptomic measures defined the subtypes, proteomic data served as an orthogonal biological perspective. Therefore, PRS and TRS were each divided into 2 to 3 quantiles (yielding between 4 and 9 total groups), and the combinations that maximized the number of significantly differentially expressed proteins—determined via limma [18] by comparing each category against the lowest quantile—were selected. We further evaluated the sensitivity of the partitions by inspecting clinical features across the various group combinations. Significance was defined as Benjamini-Hochberg [25] false discovery rate (FDR)-adjusted p-values below 0.05. We summed the total count of significantly differentially expressed proteins for each omics risk score category.

Clinical comparisons of omics-defined subtypes

We assessed differences in clinical parameters between the omics-defined subtypes using the tableone R package. Because transcriptomic and proteomic profiling in COPDGene occurred at the 5-year follow-up examination (Phase 2), analyses focused on anthropometric indices—including yearly change in body mass index (kg/m²/year) from study entry to the 5-year assessment—along with spirometry results (including the rate of FEV1 decline from the 5- to 10-year visits) and CT-derived measures of emphysema (% LAA < −950 HU on inspiratory scans [26] and the 15th percentile lung density [Perc15] [27]) as well as airway wall thickening (WA% [26] and Pi10 [28]) obtained at the 5-year time point. Parallel outcomes were evaluated in ECLIPSE, with follow-up data spanning from enrolment through the 3-year visit. Both studies used post-bronchodilator spirometry. All regression models accounted for age, sex, current smoking, cumulative pack-year exposure, genetic ancestry principal components, and CT scanner model (restricted to imaging endpoints). We cross-checked self-reported sex and race against genetic determinations derived from X/Y chromosomes and ancestry components, in line with established protocols from prior genetic research [20, 29]. We examined potential interaction effects between PRS and TRS by testing the statistical significance of their product term in fully adjusted models.

In COPDGene, we calculated FEV1 change as the difference between 10-year and 5-year values divided by the intervening years. In ECLIPSE, we used the slope of the linear regression of FEV1 over time, consistent with earlier descriptions [7, 30]. Additional COPDGene-specific evaluations included early-onset disease (GOLD 2–4 before age 55 years [31]) and circulating white blood cell differential counts.

Sensitivity analyses restricted the sample to individuals meeting COPD criteria (FEV1/FVC < 0.7) at baseline or at the 5-year visit before rerunning the regression models.

Regression model specifications

We carried out multivariable linear regression analyses within the COPDGene validation set, contrasting each identified subtype against the reference category. Key endpoints included FEV1 percent predicted, FEV1/FVC ratio, % LAA < −950 HU, Perc15, Pi10, and WA%. Covariate adjustment encompassed age, sex, current smoking status, pack-years, genetic ancestry principal components, and CT scanner (for radiographic variables). Outcome selection reflected clinical importance and adequate data completeness (< 20% missing). We determined adjustment covariates through clinical consultation and based on universal availability. Interaction testing involved including PRS, TRS, and their multiplicative term (PRS ∗ TRS) simultaneously in regression models.

Biological characterization of omics-defined subtypes

Differential expression of genes and proteins (FDR-adjusted P < 0.05) was determined by contrasting each high-risk subtype with the reference category, defined by the lowest quantiles of both PRS and TRS. Identified proteins were projected onto the human protein–protein interactome [32], and subsequent enrichment analyses were conducted using Reactome [33], Enrichr [34-36], and STRING [37] tools.

Differential expression analysis

We used limma to detect proteins whose expression varied by PRS status, declaring significance at FDR-adjusted P-values < 0.05. We also performed direct contrasts of protein profiles between high-risk

subgroups. To clarify the influence of genetic risk on the transcriptome, we assessed differential gene expression linked to COPD affection status (GOLD 2–4 versus preserved spirometry), adjusting for age, sex, smoking status, and pack-years. We then refitted the model to adjust for PRS and ancestry principal components. We selected covariates based on clinical judgment and complete data availability across the sample.

STRING, pathway enrichment, and drug repurposing analyses

Networks were generated in STRING (www.string-db.org) by querying the human interactome and retaining up to ten first-shell and five second-shell interactors around each seed protein [37], limited to high-confidence edges (score ≥ 0.7).

Pathway over-representation testing and Markov cluster (MCL) analysis (inflation parameter set at 3) [38] were performed on the resulting protein–protein interaction (PPI) networks. Differentially expressed proteins were submitted to Enrichr (maayanlab.cloud/Enrichr) to interrogate the MAGMA drugs and diseases resource [39]. We also used the Enrichr Appyter platform to propose candidate drugs for repurposing in specific omics-defined subtypes. We updated subtype nomenclature post-analysis to better capture the associated clinical traits and mechanistic signatures.

Ethics

We obtained written informed consent from all participants, and the original studies received approval from relevant institutional review boards. Approval for the present investigation was granted by the Mass General Brigham institutional review board (IRB #2007P000554).

Role of funders

GlaxoSmithKline participated in the design of the ECLIPSE study and in acquiring genotype and phenotype information. Beyond this contribution, sponsors had no role in the design, conduct, analysis, interpretation, or reporting of the current work. The corresponding author, with unrestricted data access, assumed final responsibility for manuscript submission.

Results and Discussion

Characteristics of study populations

We included 3,274 ever-smokers from two distinct cohorts in the analyses. The COPDGene training and testing subsets showed close alignment in baseline demographics and lung function parameters (**Table 1**). In comparison with COPDGene, the ECLIPSE cohort consisted of individuals who were, on average, younger, more often male, had accumulated greater lifetime tobacco exposure, presented with more impaired FEV1 percent predicted and FEV1/FVC ratios, and included a smaller proportion of current smokers.

Table 1. Features of the participating cohorts.

Variable	COPDGene training cohort	COPDGene validation cohort (NHW)	COPDGene validation cohort (AA)	ECLIPSE cohort	P-value
Participants, n	1374	1133	299	468	—
Age, years (mean ± SD)	67.38 ± 8.25	68.13 ± 8.30	61.01 ± 6.95	64.43 ± 6.09	<0.001
Female participants, n (%)	677 (49.3)	560 (49.4)	153 (51.2)	156 (33.3)	<0.001
Smoking exposure, pack-years (mean ± SD)	45.47 ± 24.61	45.47 ± 23.65	39.55 ± 20.74	49.33 ± 26.87	<0.001
Current smokers, n (%)	343 (25.0)	289 (25.5)	182 (60.9)	70 (15.0)	<0.001
Predicted FEV1 (%) (mean ± SD)	78.37 ± 24.21	77.85 ± 24.81	81.79 ± 23.54	44.22 ± 14.64	<0.001
Predicted FVC (%) (mean ± SD)	87.25 ± 17.34	86.79 ± 17.55	88.68 ± 16.63	80.63 ± 19.37	<0.001
FEV1/FVC ratio (mean ± SD)	0.67 ± 0.15	0.66 ± 0.15	0.71 ± 0.14	0.45 ± 0.11	<0.001

Abbreviations: COPD, chronic obstructive pulmonary disease. COPDGene, Genetic Epidemiology of COPD. ECLIPSE, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints study. FEV1, forced expiratory volume in 1 s. FEV1/FVC, FEV1/forced vital capacity. NHW, non-Hispanic white. AA, African American.

Defining polygenic and transcriptional risk score divisions

Mapping study participants by PRS and TRS values revealed a continuous spectrum without clear separation,

similar to the distribution commonly observed for standard measures of COPD severity such as FEV1 and FEV1/FVC [10, 40]. Evaluation of the gap statistic across multiple k-means cluster solutions in the COPDGene training data reinforced this observation: the single-cluster solution produced the maximum gap value, indicating no natural clustering structure.

Given the limitations of clustering for these continuous risk metrics, we pursued a quantile-based stratification strategy. Because the most appropriate cut-points balancing meaningful risk elevation with sufficient statistical power were not obvious, we systematically compared four partitioning schemes that combined the two omics scores with an independent layer of information—differential protein expression. The configuration that identified the most differentially expressed proteins involved splitting the PRS into two groups and the TRS into three. These thresholds were then applied to classify participants in the COPDGene testing cohort and the ECLIPSE validation sample (**Figure 2**). Stability of the derived subgroups was verified by exploring alternative divisions ranging from 4 to 9 groups; clinical traits remained consistent for both the highest-risk categories ("low PRS, high TRS" and "high PRS, high TRS") and the lowest-risk reference ("low PRS, low TRS") across these variations.

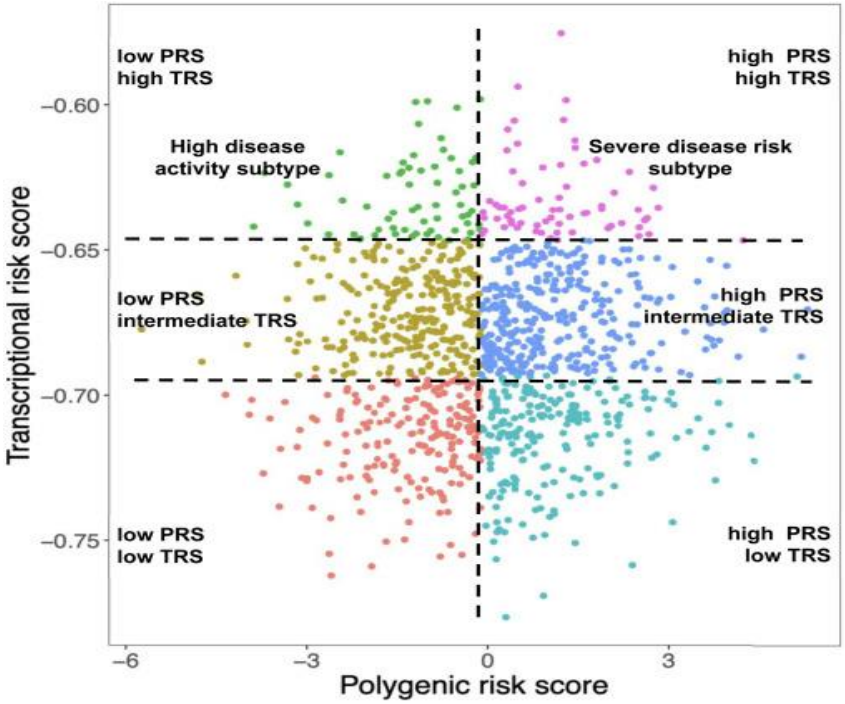

Figure 2. Placement of omics-defined groups and subtypes on the plane defined by polygenic risk score (PRS; horizontal axis) and transcriptional risk score (TRS; vertical axis) in the COPDGene testing set (n = 1432).

Polygenic and transcriptional risk scores identify "high disease activity" and "severe disease risk" subtypes

Marked differences emerged between the two high-risk categories that shared elevated TRS levels (**Table 2**). Both were benchmarked against a low-risk reference group defined by the lowest quantiles of both scores ("low PRS, low TRS"). In the validation datasets, the high-TRS subgroups ("low PRS, high TRS" and "high PRS, high TRS") were characterized by reduced BMI, worse lung function, increased emphysematous destruction, and more pronounced airway wall thickening than the reference (**Table 2).** After adjustment, prospective FEV1 decline accelerated to a similar degree in both high-risk groups within COPDGene; however, only the "low PRS, high TRS" group showed consistently greater decline in ECLIPSE (−40 mL/year).

Table 2. Description of omics-defined subtypes within the COPDGene testing set and ECLIPSE.

Variable	COPDGene NHW Validation Cohort			COPDGene AA Validation Cohort			ECLIPSE Cohort		
	Reference (Low PRS, Low TRS)	Low PRS, High TRS	High PRS, High TRS	Reference (Low PRS, Low TRS)	Low PRS, High TRS	High PRS, High TRS	Reference (Low PRS, Low TRS)	Low PRS, High TRS	High PRS, High TRS
Participants (n)	196	65	68	61	16	11	57	63	56
Age, years (mean ± SD)	67.85 ± 8.52	67.16 ± 6.46	69.04 ± 8.11	61.26 ± 7.55	60.58 ± 7.24	65.95 ± 8.05	63.00 ± 5.10	65.43 ± 6.65	66.59 ± 5.95
Female participants, n (%)	114 (58.2)	19 (29.2)	24 (35.3)	36 (59.0)	5 (31.2)	8 (72.7)	30 (52.6)	10 (15.9)	11 (19.6)
Smoking history, pack-years (mean ± SD)	36.51 ± 22.27	60.07 ± 26.49	57.28 ± 23.35	37.01 ± 20.85	43.42 ± 16.14	61.36 ± 38.87	45.49 ± 26.24	55.29 ± 35.25	54.29 ± 21.10
Current smokers, n (%)	20 (10.2)	36 (55.4)	36 (52.9)	31 (50.8)	15 (93.8)	8 (72.7)	6 (10.5)	7 (11.1)	12 (21.4)
Body mass index (kg/m²) (mean ± SD)	30.00 ± 6.84	26.51 ± 5.81	25.70 ± 5.58	30.61 ± 6.36	29.14 ± 8.52	25.26 ± 6.57	27.54 ± 5.06	26.70 ± 6.22	25.49 ± 4.76
Predicted FEV1 (%) (mean ± SD)	90.15 ± 18.13	66.50 ± 27.07	60.55 ± 24.11	90.83 ± 17.41	80.61 ± 25.43	71.08 ± 10.67	50.25 ± 13.42	43.41 ± 15.65	37.44 ± 13.44
Predicted FVC (%) (mean ± SD)	90.44 ± 15.27	85.25 ± 20.18	83.25 ± 16.65	91.65 ± 14.37	85.12 ± 20.95	91.99 ± 13.77	90.53 ± 20.17	79.60 ± 18.68	75.25 ± 22.51
FEV1/FVC ratio (mean ± SD)	0.75 ± 0.09	0.57 ± 0.15	0.53 ± 0.15	0.77 ± 0.10	0.73 ± 0.14	0.61 ± 0.12	45.84 ± 11.13	43.16 ± 12.79	39.71 ± 10.29
Low attenuation area (% LAA < −950 HU) (mean ± SD)	2.78 ± 4.71	8.50 ± 10.39	11.00 ± 11.98	1.71 ± 4.00	3.51 ± 8.08	5.49 ± 7.30	16.01 ± 11.46	20.88 ± 13.50	22.44 ± 13.22
Perc15 (mean ± SD)	−913.62 ± 20.43	−927.23 ± 27.21	−931.05 ± 31.55	−895.53 ± 26.89	−901.19 ± 29.94	−913.62 ± 34.95	−943.75 ± 50.13	−960.00 ± 49.44	−971.11 ± 45.84
Pi10 (mean ± SD)	2.01 ± 0.50	2.55 ± 0.65	2.60 ± 0.51	2.11 ± 0.48	2.56 ± 0.59	2.42 ± 0.44	4.38 ± 0.20	4.42 ± 0.23	4.43 ± 0.19
Wall area percentage (WA%) (mean ± SD)	46.97 ± 8.11	54.46 ± 8.35	54.61 ± 7.41	47.48 ± 7.76	54.86 ± 10.52	51.62 ± 4.06	64.35 ± 2.79	65.05 ± 3.58	67.22 ± 3.44
Prospective FEV1 decline (mL/year) (mean ± SD)	−32.11 ± 48.07	−51.21 ± 74.41	−66.45 ± 61.58	−42.28 ± 36.59	−87.86 ± 156.98	−56.26 ± 76.62	−32.21 ± 59.54	−40.13 ± 75.31	−15.22 ± 67.46

The low-risk reference category corresponded to the "Low PRS, Low TRS" group. BMI, body-mass index. LAA, low attenuation area. HU, Hounsfield units. Perc15, 15th percentile of lung density histogram on inspiratory CT scans. WA %, wall area percent. Pi10, square root of wall area of a hypothetical internal perimeter of 10 mm. ACO, asthma-COPD overlap. Consult **Table 1** legend for further abbreviations. NHW, non-Hispanic white. AA, African American.

We then applied linear regression modeling to selected COPD-relevant endpoints. We compared anthropometric indices, pulmonary function parameters, imaging findings, and additional clinical measures between the COPDGene and ECLIPSE cohorts (**Table 3**). Relative to the reference category, both high-risk groups showed inferior spirometric values, more extensive emphysema, and greater airway wall thickening (**Table 3**). Although estimates of FEV1 decline after adjustment did not attain statistical significance, the "low PRS, high TRS" group showed concordant point estimates across studies (−30 mL/yr in COPDGene and −24 mL/year in ECLIPSE; P = 0.15 and P = 0.11, respectively). This pattern was notable given the limited number of FEV1 assessments in COPDGene (two measurements) versus the more frequent sampling in ECLIPSE (up to six). On account of the clinical significance of faster lung function deterioration in this category, the "low PRS, high TRS" group was reclassified as the "high disease activity" subtype. The "high PRS, high TRS" group, despite absence of replicated FEV1 decline, demonstrated the

poorest baseline lung function and most pronounced emphysema; hence, it was termed the "severe disease risk" subtype.

Distribution of COPD-associated traits across omics-defined subtypes was largely preserved when examined separately by sex. Within NHW individuals in COPDGene, the "high disease activity" subtype displayed a non-significant trend for accelerated BMI loss relative to the reference (β = −0.154 [95% CI: −0.333 to 0.0253], P = 0.094). Interaction testing revealed no significant multiplicative effects between PRS and TRS (all cross-product P-values > 0.05) for either spirometric or CT-based outcomes. Restricting the sample to participants with airflow obstruction (FEV1/FVC < 0.7) at any visit yielded comparable patterns. All ECLIPSE subjects fulfilled COPD diagnostic criteria. Findings restricted to COPDGene participants alone mirrored the main results concerning subtype profiles.

Table 3. Multivariable-adjusted linear regression results in the COPDGene testing set and ECLIPSE.

	COPDGene NHW Validation Cohort				COPDGene AA Validation Cohort				ECLIPSE Cohort			
COPD Outcome	**Low PRS, High TRSEstimate (95% CI)**	**p-value**	**High PRS, High TRSEstimate (95% CI)**	**p-value**	**Low PRS, High TRSEstimate (95% CI)**	**p-value**	**High PRS, High TRSEstimate (95% CI)**	**p-value**	**Low PRS, High TRSEstimate (95% CI)**	**p-value**	**High PRS, High TRSEstimate (95% CI)**	**p-value**
Predicted FEV1 (%)	−22.3 (−29.5 to −15.2)	4.3E-09	−31.5 (−38.2 to −24.8)	1E-17	−4.11 (−16.8 to 8.59)	0.53	−19.6 (−32.7 to −6.51)	0.0048	−7.8 (−13.6 to −1.95)	0.01	−15.5 (−21.4 to −9.61)	1.3E-06
Predicted FVC (%)	−7.24 (−13 to −1.51)	0.014	−11.9 (−17.1 to −6.63)	1.3E-05	−4.07 (−14.6 to 6.45)	0.45	−3.77 (−15.5 to 7.91)	0.53	−6.89 (−14.9 to 1.15)	0.096	−12.4 (−22.2 to −2.51)	0.016
FEV1/FVC ratio	−0.147 (−0.184 to −0.11)	3E-13	−0.197 (−0.233 to −0.161)	4.8E-22	−0.00368 (−0.0731 to 0.0657)	0.92	−0.118 (−0.193 to −0.043)	0.0032	−4.5 (−9.07 to 0.0772)	0.057	−8.83 (−13.5 to −4.16)	0.00035
LAA < −950 HU (%)	3.1 (1.8–5.2)	3.30E-05	3.4 (2.1–5.6)	2.30E-06	1.3 (0.44–3.7)	0.66	0.94 (0.23–3.9)	0.93	2.33 (−3.34 to 8)	0.42	6.97 (−0.0748 to 14)	0.058
Perc15	−11.2 (−18.6 to −3.72)	0.0037	−15.3 (−22.8 to −7.89)	7.6E-05	−0.305 (−17.1 to 16.5)	0.97	−4.66 (−26.2 to 16.9)	0.67	−0.994 (−24.1 to 22.1)	0.93	−19.2 (−49.3 to 10.8)	0.21
Pi10	0.498 (0.298–0.698)	2.10E-06	0.605 (0.429–0.78)	1.40E-10	0.346 (−0.032 to 0.724)	0.079	0.475 (0.00489–0.946)	0.054	−0.0786 (−0.147 to −0.00988)	0.029	−0.0753 (−0.163 to 0.0124)	0.099
Wall area (%)	5.88 (3.19–8.58)	2.90E-05	7.17 (4.63–9.71)	8.90E-08	4.55 (−1.6 to 10.7)	0.15	7.48 (−0.0534 to 15)	0.058	0.516 (−0.972 to 2)	0.5	2.18 (0.376–3.98)	0.021
Annual FEV1 change (mL/year)	−30 (−70.6 to 10.7)	0.15	−43.9 (−91.7 to 3.94)	0.077	−103 (−294 to 87.5)	0.32	−60.5 (−105 to −15.6)	0.033	−23.7 (−52.9 to 5.47)	0.11	6.97 (−25.8 to 39.8)	0.68

Models accounted for age, sex, current smoking, pack-years, and genetic ancestry principal components. CT

imaging variables were additionally adjusted for scanner type. Abbreviations are explained in the legends of **Tables 1 and 2**.

Biological characterization and drug repurposing analyses of subtypes

After documenting clinical distinctions between the high-risk categories, we investigated their underlying biological features in greater depth. We assessed differential transcript and protein expression by comparing each high-risk subtype with the reference in the COPDGene NHW testing sample. Fourteen proteins were differentially expressed in the "high disease activity" subtype and two in the "severe disease risk" subtype.

Proteins showing differential abundance for each high-risk subtype in the COPDGene testing set were projected onto the human protein–protein interactome [32] and used as seed nodes to construct STRING-based PPI networks (**Figures 3 and 4**). These networks incorporated associated MCL clusters and underwent pathway enrichment testing. For the low PRS/high TRS group, we generated separate networks for up- and down-regulated proteins, revealing additional involvement of Asparagine N-linked glycosylation, NCAM1, and RAF/MAP kinase cascade pathways. We identified potential subtype-tailored drug repurposing candidates by querying the MAGMA Drugs and Diseases database39 with the same seed proteins. Proteomic signatures in both subtypes indicated possible responsiveness to ACE inhibitors, thyroid agents, carvedilol, bromocriptine, and lovastatin. The "high disease activity" subtype further showed enrichment signals for 5-lipoxygenase inhibitors, fomepizole, and galantamine, while the "severe disease risk" subtype highlighted atypical antipsychotics.

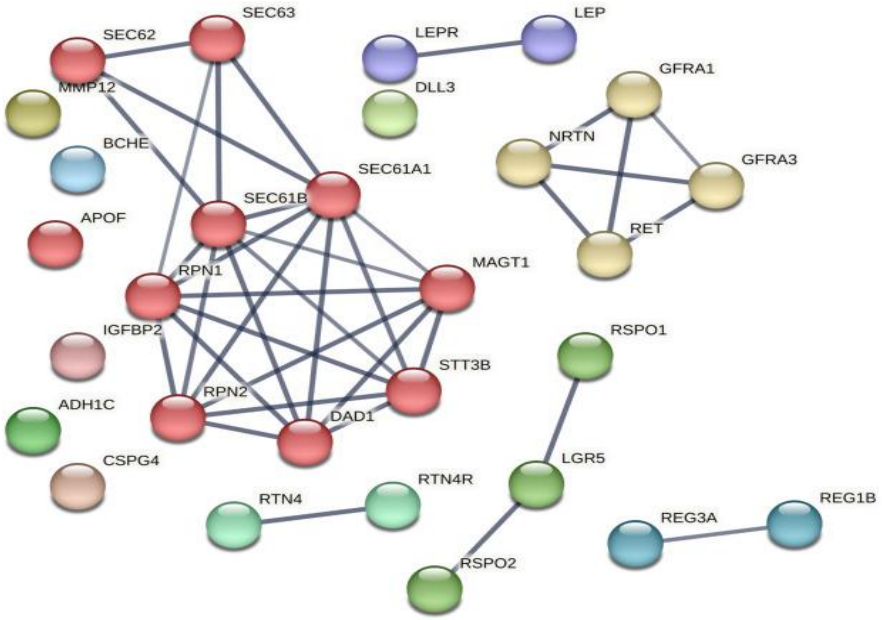

Figure 3. STRING protein–protein interaction networks for the high disease activity ("low PRS, high TRS") subtype, constructed using differentially expressed proteins from the omics-defined groups (subtypes) in the COPDGene testing set as seeds. Up to 10 first-shell and 5 second-shell interactors were permitted. Only high-confidence interactions (score ≥ 0.7) were retained; edge thickness reflects interaction confidence. Differentially expressed proteins were identified relative to the reference group. Colors indicate MCL (Markov) clusters.

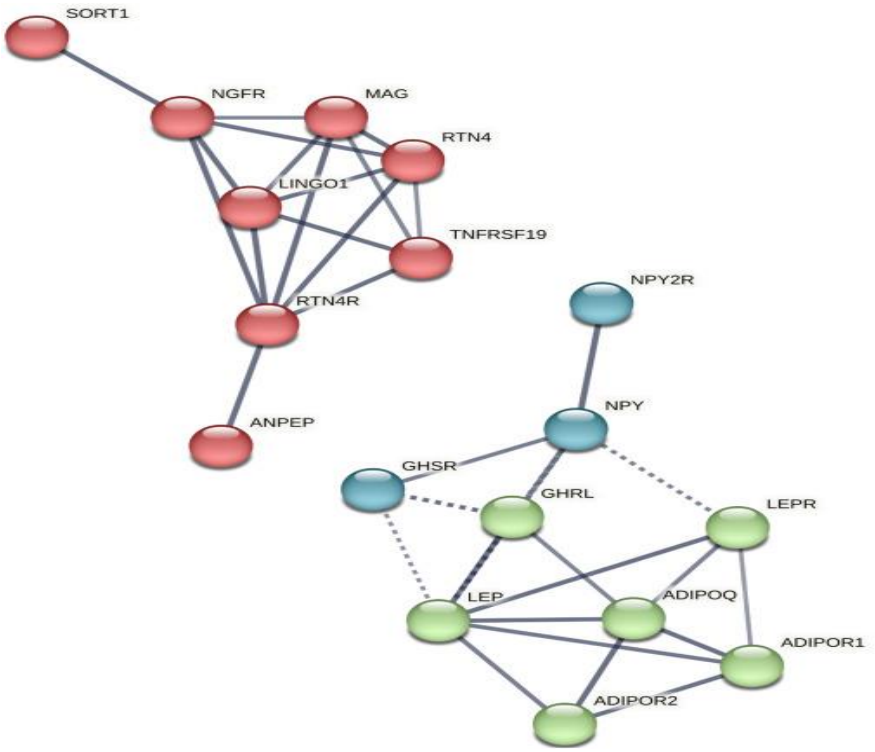

Figure 4. STRING protein–protein interaction networks for the severe disease risk ("high PRS, high TRS") subtype, constructed using differentially expressed proteins from the omics-defined groups (subtypes) in the COPDGene testing set as seeds. Up to 10 first-shell and 5 second-shell interactors were permitted. Only high-confidence interactions (score ≥ 0.7) were retained; edge thickness reflects interaction confidence. We identified differentially expressed proteins relative to the reference group. Colors indicate MCL (Markov) clusters.

Results and Discussion

In this analysis of 3,274 ever-smokers from two separate cohorts, we applied blood-derived polygenic (PRS) and transcriptional (TRS) risk scores to uncover heterogeneity among individuals at elevated risk, delineating "high disease activity" and "severe disease risCOPD subtypes. Relative to a low-risk reference group, both subtypes exhibited reduced average BMI and perturbations in metabolic, growth factor, and immune signaling pathways. Participants classified as "high disease activity" demonstrated prospective FEV1 decline in both validation cohorts, although the association was not statistically significant (yet directionally consistent) in models adjusted for baseline FEV1. Individuals in the "severe disease risk" subtype presented with markedly impaired spirometry, extensive emphysema, and prominent airway wall thickening. We identified biological pathways and potential drug-repurposing opportunities for each subtype, including agents previously evaluated in COPD trials. Overall, these findings illustrate how omics-based risk scores can reveal COPD subtypes possessing distinct clinical and molecular features that may inform targeted therapeutic strategies.

Characterizing the pathobiological underpinnings of omics-defined high-risk populations is an evolving research domain, and this work extends it to respiratory disease. In schizophrenia research, for instance, PRS-stratified high-risk groups have been profiled using weighted gene co-expression network methods [41]. Alternative strategies have integrated gene expression data into PRSs to enhance predictive performance and illuminate mechanistic pathways [42-44]. The current study used genetic and transcriptomic information to define subtypes and then used proteomic differences to elucidate their biology. A key observation is that distinct omics risk scores are not interchangeable; elevated scores do not uniformly associate with identical clinical outcomes. Participants in the highest transcriptomic risk stratum showed divergent clinical and biological traits depending on their concurrent polygenic burden.

Although regression models for COPD-related endpoints showed no significant statistical interactions between PRS and TRS, the quantile-based subtyping strategy identified individuals with advanced disease and heterogeneous presentations—heterogeneity that conventional interaction testing failed to detect.

Furthermore, while participants occupied a continuous spectrum of COPD risk without discrete clusters, the resulting omics-defined subgroups exhibited meaningful and stable clinical and biological differences. The two high-risk categories ("low PRS, high TRS" and "high PRS, high TRS") maintained consistent phenotypic profiles across various partitioning schemes. These observations indicate that genetic and transcriptomic measures offer complementary perspectives on pulmonary biology.

The identified high-risk groups displayed notable clinical distinctions, including varying degrees of airflow limitation, emphysema severity, airway pathology, and rates of lung function deterioration. From a prognostic standpoint, distinguishing patients with advanced disease or rapid progression is particularly valuable. Relying solely on blood-based omics profiling, we could isolate a subtype characterized by very advanced disease (high PRS, high TRS) and another predisposed to accelerated FEV1 decline (low PRS, high TRS).

Clinically, the combination of reduced spirometry, increased emphysema, and replicated FEV1 decline points to the "high disease activity" subtype as a potentially modifiable trait. This risk profile may be amenable to existing therapies, including 5-lipoxygenase inhibitors, angiotensin-converting enzyme (ACE) inhibitors, fomepizole, and galantamine. To our knowledge, fomepizole has not previously been proposed as a repurposing candidate for COPD. Galantamine is known to induce bronchospasm, and enrichment of its target proteins implies likely harm in this patient subset. Although a randomized trial of the 5-lipoxygenase inhibitor zileuton failed to shorten hospital stay or reduce treatment failure in acute COPD exacerbations, the study was probably underpowered [45] and did not assess longer-term effects. ACE inhibitors and angiotensin receptor blockers (ARBs) have been suggested as repurposing candidates in COPD. A recent trial found that losartan did not slow emphysema progression [46]. Our analyses specifically implicated captopril—not the broader class of ACE inhibitors or ARBs—in the "high disease activity" but not the "severe disease risk" subtype.

In contrast, atypical antipsychotics emerged as potential candidates for the "severe disease risk" group. The frequent co-occurrence of schizophrenia and COPD is commonly ascribed to smoking. Yet, phenome-wide and polygenic risk investigations indicate possible shared genetic architecture between schizophrenia and obstructive lung disorders [47]. Although models adjusted for self-reported smoking, such measures are imperfect, precluding definitive conclusions regarding shared mechanisms independent of tobacco exposure. A broader implication arising from the drug repurposing results, which warrants prospective validation, is that earlier trial failures may have stemmed from unaccounted patient heterogeneity that omics-guided stratification could help resolve.

The cachectic phenotype in COPD, often linked to heightened exacerbation risk, is a well-recognized clinical presentation [48-51]. Consistent with this, both the "high disease activity" and "severe disease risk" subtypes showed lower average BMI values than other groups. Although pulmonary cachexia has traditionally been regarded as a feature confined to advanced COPD [52], our findings reveal further heterogeneity among ever-smokers with reduced BMI and delineate a "high disease activity" subtype characterized by comparatively milder airflow obstruction yet unique molecular features. This subtype may share characteristics with a previously reported cachexia-associated comorbidity cluster [53]. In the present work, we uncovered corresponding molecular signatures that could help direct therapeutic decisions.

Regarding body composition, leptin—an adipokine—showed differential expression in both high-risk subtypes. This observation admits several non-exclusive explanations. Leptin functions as a hormone that suppresses appetite and regulates fat storage [54] while simultaneously acting as a proinflammatory cytokine critical for immune defense [55, 56]. Leptin's appetite-regulating role may be particularly pertinent to the "severe disease risk" subtype, which showed the strongest reduction in leptin levels (−0.71 log-fold change). Such decreased circulating leptin could represent either a compensatory mechanism or a shared causal pathway with pulmonary cachexia.

Furthermore, reduced leptin concentrations, commonly observed in states of malnutrition, are associated with impaired cellular immunity and increased vulnerability to infection [55, 57]. In COPD populations, some investigations have reported elevated leptin in plasma and airways [58-62], while others have not [63]; leptin

has also been proposed as a marker of emphysema progression [64, 65]. Our data suggest that individuals in the "high disease activity" or "severe disease risk" subtypes may display immune and cytokine patterns resembling those in malnourished states, potentially increasing susceptibility to infections. It remains uncertain whether the detected blood proteomic changes reflect downstream consequences of disease activity or contribute causally to a self-reinforcing cycle that accelerates progression. Because the biomarkers examined here derive from peripheral blood, establishing direct mechanistic links to lung tissue processes is challenging. Nevertheless, blood-based markers offer clear practical advantages for clinical use, and future studies correlating these changes with lung-specific pathophysiology could help translate predictive models into precision treatments.

This investigation integrated multiple layers of omics information through well-validated and replicated risk scores to define COPD subtypes across two cohorts of ever-smokers. Although transcript and protein levels often show weak correspondence, complicating joint analyses, we leveraged this limited correlation to gain insight into protein-level alterations within subtypes originally defined by genetic and transcriptomic data. The proteome, used here for protein–protein interaction mapping, offers distinct benefits for subtype discovery: proteins are the primary effectors of cellular function, participate directly in pathways and interactions, and often serve as accessible biomarkers. Moreover, because most approved drugs are small molecules directed against proteins, proteomic data are especially advantageous for identifying therapeutic targets and repurposing opportunities.

A central methodological challenge involved partitioning subjects from continuous, standardized risk distributions. We addressed this by systematically testing different quantile combinations, selecting those that maximized detection of differentially expressed proteins, and then deriving raw score thresholds to classify every participant in the COPDGene and ECLIPSE test sets. Other valid strategies for applying omics risk scores undoubtedly exist. Both COPDGene and ECLIPSE primarily enrolled heavy smokers; therefore, this approach applies to populations meeting these inclusion criteria.

Participants with PRS values near the chosen cut-offs may lie along a continuum between the "high disease activity" and "severe disease risk" subtypes and could potentially shift categories over time. Nevertheless, establishing explicit thresholds represents an essential step toward clinical implementation. Our focus remained on the highest-risk groups. Although the high PRS/low TRS combination was theoretically interesting, individuals in this category exhibited relatively mild disease compared with the other high-risk groups. The conventional FEV1/FVC threshold of 0.7 for defining airflow obstruction has faced increasing scrutiny. Recent ATS/ERS spirometry recommendations [66] acknowledge preserved ratio impaired spirometry (PRISm) as a category associated with respiratory symptoms, emphysema, airway abnormalities, elevated mortality, and exacerbation risk [40, 67-69]. At the same time, mortality risk rises sharply once FEV1/FVC falls below 0.7 [70]. Accordingly, we repeated the primary analyses restricted to participants with FEV1/FVC < 0.7. We obtained comparable findings, indicating that this omics-driven subtyping strategy is relevant for both established COPD and smokers at high risk of developing the disease.

Comparable patterns, including the association between the "high disease activity" subtype and FEV1 decline, were observed among African American participants in COPDGene. We have previously shown that the PRS, trained primarily in European-ancestry populations, performs less well in individuals of non-European descent [6]. Although we successfully tested subtypes in African Americans, additional research is needed to improve genetic risk prediction across diverse ancestries. While COPDGene served as the discovery cohort and ECLIPSE as the validation cohort, important differences between the studies should be acknowledged. COPDGene contained more participants without COPD, whereas ECLIPSE enrolled a higher proportion with severe disease. ECLIPSE also employed lower-dose CT protocols. Prior work has documented systematic differences in emphysema and airway wall measurements between these cohorts even after propensity score matching [71], likely attributable to imaging protocol variations. Drug repurposing candidates were derived from proteomic enrichment analyses that did not incorporate directionality of changes; therefore, further mechanistic and clinical validation is necessary. Finally, although we identified two high-risk subtypes, they do not capture the full spectrum of COPD heterogeneity, and additional clinically relevant subtypes likely exist.

Conclusion

In conclusion, the combination of polygenic and transcriptional risk scores, both derived from spirometric measures, successfully delineated “high disease activity” and “severe disease risk” COPD subtypes with unique clinical and molecular profiles. Proteomic profiling and drug repurposing analyses revealed subtype-specific therapeutic enrichments, including several agents previously considered for COPD treatment.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Safiri S, Carson-Chahhoud K, Noori M, Nejadghaderi SA, Sullman MJM, Kolahi AA, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019. BMJ. 2022;378:e069679.
2. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med. 2010;363(12):1128-38.
3. Wedzicha JA. The heterogeneity of chronic obstructive pulmonary disease. Thorax. 2000;55(8):631-2.
4. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. Circulation. 2017;135(22):2091-101.
5. Rodon J, Soria JC, Berger R, Miller WH, Rubin E, Kugel A, et al. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. Nat Med. 2019;25(5):751-8.
6. 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.
7. Moll M, Boueiz A, Ghosh AJ, Saferali A, Lee S, Xu H, et al. Development of a blood-based transcriptional risk score for chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2022;205(2):161-70.
8. Zhang J, Xu H, Qiao D, Sun J, Shrine N, Li X, et al. A polygenic risk score and age of diagnosis of chronic obstructive pulmonary disease. Eur Respir J. 2022;2101954.
9. Sun J, Wang Y, Folkersen L, Borné Y, Amlien IK, Buil A, et al. Translating polygenic risk scores for clinical use by estimating the confidence bounds of risk prediction. Nat Commun. 2021;12(1):5276.
10. Castaldi PJ, Dy J, Ross J, Chang Y, Washko GR, Curran-Everett D, et al. Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. Thorax. 2014;69(5):415-22.
11. Martinez FJ, Agusti A, Celli BR, Han MK, Allinson JP, Bhatt SP, et al. Treatment trials in young patients with chronic obstructive pulmonary disease and pre-chronic obstructive pulmonary disease patients: time to move forward. Am J Respir Crit Care Med. 2022;205(3):275-87.
12. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. COPD. 2010;7(1):32-43.
13. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype reference consortium panel. Nat Genet. 2016;48(11):1443-8.
14. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD longitudinally to identify predictive surrogate endpoints (ECLIPSE). Eur Respir J. 2008;31(4):869-73.
15. Global Initiative for Chronic Obstructive Lung Disease - GOLD. 2024 GOLD report [Internet]. [cited 2024 Mar 11]. Available from: https://goldcopd.org/2024-gold-report/
16. Parker MM, Chase RP, Lamb A, Reyes A, Saferali A, Yun JH, et al. Correction to: RNA sequencing identifies novel non-coding RNA and exon-specific effects associated with cigarette smoking. BMC Med Genom. 2019;12(1):166.

17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl. 2010;26(1):139-40.
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
19. Serban KA, Pratte KA, Strange C, Barker AF, Bilyk S, Criner GJ, et al. Unique and shared systemic biomarkers for emphysema in Alpha-1 Antitrypsin deficiency and chronic obstructive pulmonary disease. eBioMedicine. 2022;84:104262.
20. 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.
21. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genet Epidemiol. 2017;41(6):469-80.
22. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B. 1996;4196-249.
23. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Methodol. 2001;63(2):411-23.
24. Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic prediction of weight and obesity trajectories from birth to adulthood. Cell. 2019;177(3):587-596.e9.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289-300.
26. Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, et al. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. Radiology. 2011;261(1):274-82.
27. Parr DG, Sevenoaks M, Deng CQ, Stoel BC, Stockley RA. Detection of emphysema progression in alpha 1-antitrypsin deficiency using CT densitometry; Methodological advances. Respir Res. 2008;9:1-8.
28. Van Tho N, Ogawa E, Trang LTH, Ryujin Y, Kanao R, Nakano Y, et al. A mixed phenotype of airway wall thickening and emphysema is associated with dyspnea and hospitalization for chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2015;12(7):988-96.
29. Sakornsakolpat P, Prokopenko D, Lamontagne M, DeMeo DL, Hunninghake GM, McGeachie MJ, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet. 2019;51(3):494-505.
30. Vestbo J, Edwards LD, Scanlon PD, Yates JC, Agusti A, Bakke P, et al. Changes in forced expiratory volume in 1 second over time in COPD. N Engl J Med. 2011;365(13):1184-92.
31. Foreman MG, Zhang L, Murphy J, Hansel NN, Make B, Hokanson JE, et al. Early-onset chronic obstructive pulmonary disease is associated with female sex, maternal factors, and African American race in the COPDGene study. Am J Respir Crit Care Med. 2011;184(4):414-20.
32. Wang RS, Loscalzo J. Network module-based drug repositioning for pulmonary arterial hypertension. CPT Pharmacomet Syst Pharmacol. 2021;10(9):994-1005.
33. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinf. 2017;18(1):142.
34. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. 2013;14:128.
35. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90-7.
36. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with Enrichr. Curr Protoc. 2021;1(3):e90.
37. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37(Database issue):D412-6.
38. Satuluri V, Parthasarathy S, Ucar D. Markov clustering of protein interaction networks with improved balance and scalability. In: Proceedings of

the first ACM international conference on bioinformatics and computational biology. New York, NY, USA: Association for Computing Machinery; 2010. p. 247-56. (BCB '10). doi:10.1145/1854776.1854812
39. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.
40. Lowe KE, Regan EA, Anzueto A, Austin E, Austin JH, Beaty TH, et al. COPDGene® 2019: redefining the diagnosis of chronic obstructive pulmonary disease. Chronic Obstr Pulm Dis Miami Fla. 2019;6(5):384-99.
41. Radulescu E, Jaffe AE, Straub RE, Chen Q, Shin JH, Hyde TM, et al. Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. Mol Psychiatry. 2020;25(4):791-804.
42. Hu X, Qiao D, Kim W, Moll M, Hokanson JE, Crapo JD, 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):857-70.
43. Mareckova K, Hawco C, Dos Santos FC, Cahn W, Maat A, van Haren NEM, et al. Novel polygenic risk score as a translational tool linking depression-related changes in the corticolimbic transcriptome with neural face processing and anhedonic symptoms. Transl Psychiatry. 2020;10(1):1-10.
44. Bandres-Ciga S, Saez-Atienzar S, Kim JJ, Makarious MB, Faghri F, Diez-Fairen M, et al. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in Parkinson disease. Acta Neuropathol. 2020;140(3):341-58.
45. Woodruff PG, Albert RK, Bailey WC, Casaburi R, Connett JE, Cooper JA Jr, et al. Randomized trial of zileuton for treatment of COPD exacerbations requiring hospitalization. COPD. 2011;8(1):21-9.
46. Wise RA, Holbrook JT, Brown RH, Criner GJ, Dransfield MT, He J, et al. Clinical trial of losartan for pulmonary emphysema: pulmonary trials cooperative losartan effects on emphysema progression clinical trial. Am J Respir Crit Care Med. 2022;206(7):838-45.
47. Zhang R, Sjölander A, Ploner A, Lu D, Bulik CM, Bergen SE. Novel disease associations with schizophrenia genetic risk revealed in ~400 000 UK Biobank participants. Mol Psychiatry. 2022;27(3):1448-54.
48. McDonald MLN, Wouters EFM, Rutten E, Casaburi R, Rennard SI, Lomas DA, et al. It's more than low BMI: prevalence of cachexia and associated mortality in COPD. Respir Res. 2019;20(1):100.
49. Grigsby MR, Siddharthan T, Pollard SL, Chowdhury M, Rubinstein A, Miranda JJ, et al. Low body mass index is associated with higher odds of COPD and lower lung function in low- and middle-income countries. COPD. 2019;16(1):58-65.
50. Wada H, Ikeda A, Maruyama K, Yamagishi K, Umesawa M, Tanaka S, et al. Low BMI and weight loss aggravate COPD mortality in men, findings from a large prospective cohort: the JACC study. Sci Rep. 2021;11(1):1531.
51. Mason SE, Moreta-Martinez R, Labaki WW, Strand M, Regan E, Bon J, et al. Respiratory exacerbations are associated with muscle loss in current and former smokers. Thorax. 2021;76(6):554-60.
52. Eriksson B, Backman H, Bossios A, Engström G, Hedman L, Nilsson U, et al. Only severe COPD is associated with being underweight: results from a population survey. ERJ Open Res. 2016;2(3):51-2015.
53. Vanfleteren LEEGW, Spruit MA, Groenen M, Gaffron S, van Empel VPM, Bruijnzeel PLB, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013;187(7):728-35.
54. Collins S, Kuhn CM, Petro AE, Swick AG, Chrunyk BA, Surwit RS. Role of leptin in fat regulation. Nature. 1996;380(6576):677.
55. La Cava A. Leptin in inflammation and autoimmunity. Cytokine. 2017;98:51-8.
56. La Cava A, Matarese G. The weight of leptin in immunity. Nat Rev Immunol. 2004;4(5):371-9.
57. Matarese G, La Cava A, Sanna V, Lord GM, Lechler RI, Fontana S, et al. Balancing susceptibility to infection and autoimmunity: a role for leptin? Trends Immunol. 2002;23(4):182-7.
58. Suzuki M, Makita H, Östling J, Thomsen LH, Schirmer H, Bakke P, et al. Lower leptin/adiponectin ratio and risk of rapid lung function decline in chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2014;11(10):1511-9.

59. Takabatake N, Nakamura H, Abe S, Inoue S, Hino T, Saito H, et al. Circulating leptin in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1999;159(4 Pt 1):1215-9.
60. Hansel NN, Gao L, Rafaels NM, Mathias RA, Neptune ER, Tankersley C, et al. Leptin receptor polymorphisms and lung function decline in COPD. Eur Respir J. 2009;34(1):103-10.
61. Bruno A, Chanez P, Chiappara G, Siena L, Giammanco S, Gjomarkaj M, et al. Does leptin play a cytokine-like role within the airways of COPD patients? Eur Respir J. 2005;26(3):398-405.
62. Vernooy JHJ, Drummen NEA, van Suylen RJ, Cloots RHE, Moller GM, Bracke KR, et al. Enhanced pulmonary leptin expression in patients with severe COPD and asymptomatic smokers. Thorax. 2009;64(1):26-32.
63. Sueblinvong V, Liangpunsakul S. Relationship between serum leptin and chronic obstructive pulmonary disease in US adults: results from the third National Health and Nutrition Examination Survey. J Investig Med. 2014;62(7):934-7.
64. Curtis JL. Queens beat one-eyed jacks, but nobody's played the ace yet. Adipokines as chronic obstructive pulmonary disease biomarkers. Ann Am Thorac Soc. 2015;12(7):971-3.
65. Oh YM, Jeong BH, Woo SY, Kim SY, Lee JH, Lee SM, et al. Association of plasma adipokines with chronic obstructive pulmonary disease severity and progression. Ann Am Thorac Soc. 2015;12(7):1005-12.
66. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J. 2022;60(1):2101499.
67. Wan ES, DeMeo DL, Hersh CP, Shapiro SD, Rosiello RA, Sama SR, et al. Clinical predictors of frequent exacerbations in subjects with severe chronic obstructive pulmonary disease (COPD). Respir Med. 2011;105(4):588-94.
68. Wan ES, Fortis S, Regan EA, Hokanson J, Han MK, Casaburi R, et al. Longitudinal phenotypes and mortality in preserved ratio impaired spirometry in the COPDGene study. Am J Respir Crit Care Med. 2018;198(11):1397-405.
69. Wan ES, Castaldi PJ, Cho MH, Hokanson JE, Regan EA, Make BJ, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. Respir Res. 2014;15:89.
70. Bhatt SP, Balte PP, Schwartz JE, Cassano PA, Couper D, Jacobs DR Jr, et al. Discriminative accuracy of FEV1:FVC thresholds for COPD-related hospitalization and mortality. JAMA. 2019;321(24):2438.
71. Moll M, Qiao D, Regan EA, Hunninghake GM, Make BJ, Tal-Singer R, et al. Machine learning and prediction of all-cause mortality in COPD. Chest. 2020;158(3):952-64.