An Integrated Systems Approach to Infer Protein Networks Associated with Chronic Obstructive Pulmonary Disease

by

Katarina M. Di Lillo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor Kelly B. Arnold, Chair
Associate Professor Brendon M. Baker
Professor Jeffrey L. Curtis
Professor Bethany B. Moore

Katarina M. Di Lillo

kdilillo@umich.edu

ORCID iD: 0000-0002-2425-676X

Dedication

This thesis is dedicated to the individuals who participated in the SPIROMICS and COPDGene clinical cohorts, without whom this research would not be possible. Your presence in this research has given a human dimension to the scientific process. Your impact resonates throughout these pages.

Acknowledgements

Completing this thesis has been a rollercoaster of a journey that I could not have undertaken alone, and I am deeply grateful for the support, guidance, and encouragement that kept me going throughout the past five years.

To my advisor, Dr. Kelly Arnold, thank you for your guidance, insight, and patience. Your mentorship has not only shaped this research but has also been a cornerstone in my growth as a scientist.

To my mentor Dr. Jeffrey Curtis, thank you for your constant support through all the ups and downs of this work. Your insights and guidance have been instrumental in shaping my understanding of COPD, and I am deeply appreciative of your generosity in sharing your time and expertise.

To the members of my thesis committee, Dr. Jeffrey Curtis, Dr. Beth Moore, and Dr Brendon Baker. Thank you for your invaluable feedback and constructive critiques. Your collective wisdom has shaped this work in profound ways.

A special shoutout to my lab mates – Christina, Katy, Melissa, Suzie, and Robby. You all have turned our lab into a place of not just work, but shared discoveries and endless conversations (both educational and entertaining). You all made those long workdays much more bearable, and I can't wait to see the incredible things each of you will achieve next!

To my amazing family, Mom, Dad, Joe, and Vic - your unwavering love and support have been my anchor throughout this academic journey. Thank you for always keeping me

grounded and being my biggest cheerleaders. I love you.

Last but definitely not least, a huge thank you to my chosen support system, my friends. Whether you've been by my side for years or just joined my life in this latest chapter in Ann Arbor, thank you for keeping me sane and smiling throughout the highs and lows of this adventure. Your friendship means everything, and I'm beyond lucky to be surrounded by such intelligent and compassionate humans. Thank you again for all the support and unforgettable memories you've provided over the years. You all are the best!

To everyone who played a part in this journey, no matter how big or small – thanks for being a part of my story!

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Chronic obstructive pulmonary disease (COPD) is a widespread and fatal respiratory condition that affects millions globally. Despite its prevalence, effective interventions to halt or reverse disease progression remain limited due to an incomplete characterization of COPD's underlying mechanisms. Recent strides in understanding the pathobiology of COPD highlight its heterogeneous nature, encompassing diverse respiratory and systemic manifestations that likely stem from pathological changes across multiple biological pathways.

To navigate the intricate landscape of COPD pathogenesis, this thesis employs a network-based approach utilizing data-driven modeling techniques. These methodologies can help unravel complex pathological networks by facilitating the modeling of interactions among multiple biological factors, thus providing novel insights into the systems-level mechanisms underlying disease states. Furthermore, data-driven techniques permit the integration of data from diverse tissue compartments, offering a more holistic depiction of disease pathology.

In this thesis, we applied data-driven techniques to high-throughput protein measurements from human blood and lung samples. Overall, this approach revealed proteomic signatures that differentiate distinct COPD subpopulations and offered novel insights into disease progression and its underlying molecular drivers. Our findings specifically introduce new perspectives into mechanisms governing rapid spirometric decline, age-dependent emphysema development, and immune cell communication networks within COPD. Mechanistically, our investigations suggest a critical role for early modifications in the complement cascade in lung

function decline, highlight a unique contribution of inflammatory and apoptotic pathways in early emphysema development, and elucidate deficiencies in the adaptive immune responses of individuals with COPD.

In summary, this work provides novel insights into potential mechanisms driving COPD pathogenesis and progression, while also establishing a framework for exploring other heterogeneous diseases. Our findings underscore the importance of adopting a network-level approach in studying COPD and of integrating data from diverse sources to uncover new molecular insights. Furthermore, this thesis highlights the difficulties of studying complex diseases, and as a result, the methods and insights presented here carry broader significance for investigating other progressive and heterogeneous conditions extending beyond COPD.

Chapter 1 Introduction

Chronic obstructive pulmonary disease (COPD) is a highly prevalent respiratory condition with no cure and limited disease-modifying therapies[1]. In COPD, chronic exposure to inhaled oxidants (ex., cigarette smoke) causes inflammation and remodeling in the lungs, resulting in airflow limitation that is incompletely reversible. Although the mechanisms underlying COPD remain incompletely characterized, recent strides in understanding COPD pathobiology emphasize that it is a heterogenous condition with distinct subpopulations ("phenotypes") whose variable disease courses and health outcomes are likely driven by alterations in multiple biological pathways or pathway constituents. Network-level approaches hold promise in aiding the unraveling of these complex pathological networks by enabling the modeling of interactions among multiple biological factors across various tissue compartments. This thesis seeks to utilize network-level approaches, particularly data-driven modeling, to assess changes in protein networks and generate insights into the systems-level mechanisms underpinning COPD pathogenesis and progression. This chapter introduces the thesis by first discussing the background and context, followed by the research problem, the thesis aims, and their significance.

1.1 COPD diagnosis and treatment

COPD is a fatal respiratory condition that is the a leading cause of death and disability world-wide[2]. In the United States (US) alone, COPD affects an estimated 15 million Americans and contributes to more than 150,000 deaths[3] and 30 billion dollars in direct health expenditures each year[4]. Currently, the diagnosis of COPD requires patient presentation with (1) spirometric

airflow obstruction ($FEV_1/FVC < 0.7$) that is not fully reversible after bronchodilator administration, (2) relevant respiratory symptoms (e.g., dyspnea, chronic cough, or sputum production) that have no alternative explanation, and (3) appropriate clinical context, such a history of cigarette smoking[5,6]. Cigarette smoking is the primary risk factor for COPD in industrialized nations, with an overwhelming 75% of COPD diagnoses in the US related to tobacco smoke[7]. However, in the developing world, pollution, both due to indoor use of biomass fuel and outdoor, is a significant cause[8]. Importantly, not all smokers develop COPD, and other risk factors include genetic predisposition, workplace or environmental exposures, and suboptimal lung growth[9–11]. Although much of the current understanding of COPD pathophysiology is rooted in studies examining the effects of cigarette smoke, the development of COPD is believed to arise from the repeated inhalation of harmful particles and gases. These particles, frequently present in noxious irritants, trigger persistent injury to lung tissue, ultimately leading to incompletely reversible airflow obstruction, which arises from a combination of conditions including, but not limited to, parenchymal destruction (emphysema), mucous hypersecretion (chronic bronchitis), and small airways disease (obstructive bronchiolitis)[5].

Progressive declines in spirometric lung function often mark the course of COPD. Consequently, the Global Initiative for Obstructive Lung Disease (GOLD) commonly characterizes disease severity into four grades based on the extent of airflow obstruction, as measured by post-bronchodilator forced expiratory volume in one second (FEV_1): GOLD 1 (Mild: $FEV_1 \geq 80\%$ predicted), GOLD 2 (Moderate: $50\% \leq FEV_1 < 80\%$ predicted), GOLD 3 (Severe: $30\% \leq FEV_1 < 30\%$ predicted), and GOLD 4 (Very severe: $FEV_1 < 30\%$ predicted)[1]. FEV_1% predicted is a derived value that reflects the percentage of an individual's observed FEV_1 compared to the average FEV_1 in persons of similar age, sex, and body composition, especially

height.

In recent years, new methodologies have emerged as complementary tools to traditional spirometry, providing enhanced granularity in the characterization of airflow obstruction in COPD. Parametric response mapping (PRM) is one such method that is gaining popularity. PRM is a computed tomography (CT) voxel-based imaging technique that uses dynamic image registration and separate density thresholds for inspiratory and expiratory voxel measurements. This voxel-mapping process allows regions of "normal" lung to be distinguished from "functional small airways disease" (PRM^{fSAD}) or "emphysema" (PRM^{Emph})[12]. PRM metrics correlate with spirometry[13,14] and are measurable even when spirometrically-defined airflow obstruction is absent[15,16], emphasizing its heightening sensitivity and potential value in studies of early airway abnormalities.

Although spirometry-based grading classifications continue to be the conventional method used for patient stratification, there is a growing appreciation for the discrepancy between the degree of spirometric airflow obstruction and symptom burden experienced by individuals with COPD. This acknowledgment comes in the wake of much work evidencing that symptom presentation in smokers, even in the absence of airflow obstruction, is associated with adverse health outcomes[17–21]. As a result of these findings, up-to-date pharmaceutical treatment regimens are guided by a combination symptom burden, as measured by the modified Medical Research Council (mMRC) dyspnea scale[22] and COPD Assessment Test (CAT) scores[23,24], and exacerbation history[1]. Exacerbations are episodes or flare-ups of worsening respiratory symptoms; a history of prior exacerbations is the most important risk factor for future exacerbations[25].

Guided by these metrics, patients are categorized into one of three treatment groups

(**Table 1.1**): Group A, Group B, or Group E. The primary pharmacotherapy options for patients include beta-agonists (BA) and muscarinic agnostics (MA), which contribute to airway widening by promoting bronchodilation or inhibiting bronchoconstriction, respectively. Single-acting bronchodilators (BA *or* MA) are used to treat Group A, whereas Groups B and E are treated with dual-acting (BA *and* MA) regimens. In cases where patients have a history of severe exacerbations (i.e., those requiring hospitalization) or blood eosinophil counts ≥ 300/μL, treatments may also incorporate the use of inhaled corticosteroids (ICS). Although these pharmacotherapy options effectively improve patients' symptom burden, they cannot arrest or reverse disease progression. These current limitations in therapeutic efficacy stem from an incomplete understanding of the mechanisms that underpin COPD pathobiology[26,27].

Table 1.1. GOLD ABE assessment tool and pharmacologic treatments[1]

	mMRC	CAT	Exacerbation History	Treatment
Group A	0-1	< 10	0-1 moderate*	bronchodilator
Group B	≥ 2	≥ 10	0-1 moderate	dual-acting bronchodilator
Group E	Any	Any	≥ 2 moderate or ≥ 1 severe†	dual-acting bronchodilator‡

* Moderate exacerbation: exacerbation not leading to hospital admission
† Severe exacerbation: exacerbation leading to hospitalization
‡ Consider dual-acting bronchodilator + ICS if blood eosinophil counts ≥ 300/μL

1.2 Molecular mechanisms and markers of COPD

The mechanisms contributing to COPD pathogenesis and progression are likely a combination of genetic predisposition modified by oxidative stress, protease activity, and inflammation[28,29]. Given the capacity of immune cells to modulate the production of reactive oxygen species (ROS)[30,31], proteases[32], and inflammatory agents[33], the immune system is believed to play a substantial role in the progression of COPD. The central involvement of

immune networks in COPD pathogenesis is evidenced by a myriad of studies implicating nearly all immune cell types (innate and adaptive) in COPD[34–37]. Specifically, macrophage, neutrophil, and CD8+ T-cell counts are consistently reported to increase in COPD patients' lungs, including bronchoalveolar lavage (BAL) and sputum samples, and positively correlate with progression status[38,39].

In addition to the increased recruitment of inflammatory cells to the lungs, immune cells from patients with COPD often display modifications in cellular phenotype and function[40–44]. Notably, alveolar macrophages isolated from individuals with COPD demonstrate impaired phagocytic capabilities, resulting in deficient clearance of apoptotic airway epithelial cells[20,45]. This deficiency could suggest an inability of COPD patients to effectively resolve lung epithelial damage. Furthermore, conventional dendritic cells (cDCs) from individuals with COPD exhibit diminished expression of CCR5, a chemokine receptor that plays a crucial role in the uptake and processing of microbial antigens, which may impede the ability of mDCs to effectively interact with pathogens[46].

Alterations in receptor expression on immune cells can also influence their secretion of inflammatory proteins ("cytokines"). This concept is exemplified by a study that showcases how lung CD8+ T cells from COPD patients not only display elevated levels of toll-like receptors (TLRs) 2/1 but also respond to co-stimulation through TLR2/1 with amplified production of pro-inflammatory cytokines, including TNF-α and IFN-γ[47]. In line with this observation, heightened expression of additional cytokines, including IL-1β, IL-6, IL-8, IL-17A, IL-18, IL-32, and TSLP, and growth factors, such as TGF-β, have been reported in the lower airways and lungs of patients with stable COPD[35,48]. Additionally, elevated levels of other immune-cell modulated secretions, such as proteases involved in parenchymal destruction (e.g., matrix metalloproteinases (MMPs)

and neutrophil elastase), have also been shown to be increased in COPD lungs[34].

A substantial portion of the cytokines observed in COPD-affected lungs are thought to originate from structural and inflammatory cells within the respiratory tract. This release occurs in response to chronic exposure to inhaled irritants and initiates a cascade of downstream responses. For instance, in response to cigarette smoke, airway epithelial cells and alveolar macrophages release pro-inflammatory factors (such as TNF-α and IL-1β) and chemokines (such as IL-8) that promote tissue damage and neutrophil recruitment, respectively[49,50]. Additionally, activated alveolar macrophages release various factors, including CXCL9 and CXCL10, which initiate chemotaxis of T cells into the lung. Following infiltration into the lung tissue, T cells generate further pro-inflammatory factors like IFN-γ and IL-17, thus perpetuating a cyclic pattern of inflammation[51–53]. Although initially triggered by inhaled irritants, prolonged exposure to harmful stimuli (ex., cigarette smoke) permanently reshapes the immune response of COPD patients into a pathogenic state. This shift is exemplified by the persistent inflammation observed in individuals with COPD even after smoking cessation[34].

Although the pulmonary microenvironment represents the immediate location of airway injury and repair, several inflammatory proteins have also been reportedly elevated in the peripheral blood of patients with COPD, many of which correlate with phenotypes of disease progression[54]. Whether the systemic inflammation observed in COPD results from pulmonary inflammation which has "spilled" into peripheral circulation or reflects an underlying predisposition to COPD or its comorbidome remains unanswered[55]. However, the possibility of local small airway remodeling having systemic implications suggests that circulating proteins could serve as minimally invasive biomarkers.

To date, the most extensively studied systemic biomarkers of COPD include TNF-α,

fibrinogen, C-reactive protein (CRP), IL-6, IL-8, soluble receptor for advanced glycation end-products (sRAGE), MMP-9, and club cell secretory protein 16 (CC16)[48,56–61]. Unfortunately, therapies targeting these proteins have not been widely successful[48]. For example, despite the elevated levels of TNF-α in COPD patient sputum and serum and its recognized role as a major instigator of the inflammatory response, a 6-month treatment course with infliximab, a TNF-α inhibitor, yielded no discernible clinical benefit. Instead, this approach displayed a trend towards a heightened risk of cancer and pneumonia[62]. Likewise, endeavors to target IL-8 via anti-IL-8 antibodies and CXCR2 (IL-8 receptor) agonists failed to demonstrate efficacy in large-scale clinical trials, despite the consistent elevation of IL-8 levels in COPD and its established role in the chemotactic attraction of neutrophils and monocytes[63,64]. More recently, in clinical trials, benralizumab, a monoclonal antibody targeting the alpha chain of the IL-5 receptor expressed by eosinophils, demonstrated no significant reduction in the annual rate of COPD exacerbations compared to a placebo[65], even in COPD patients with elevated eosinophil counts[66]. Several possibilities exist for why these biomarkers may not have proven robust in practice, such as the targets acting downstream of a key node or redundancy in the signaling network. However, on a broader scale, it is plausible that a single protein may not be capable of addressing the inherent heterogeneity among COPD patients[58,67].

1.3 The impact of disease heterogeneity on COPD outcomes and clinical investigations

COPD is an immensely heterogeneous condition. Individuals with COPD have evidence of various respiratory conditions, including chronic bronchitis and emphysema (**Figure 1.1**)[55]. The relative contributions of each condition to overall airflow limitation vary on an individual basis[55]. This variability is reflected in diverse clinical presentations (phenotypes), which are further complicated by differing underlying pathologies (endotypes). The acknowledgment of

COPD heterogeneity has propelled the creation of extensively phenotyped cohorts aimed at identifying patient subgroups with shared demographic and molecular traits to enhance the prediction of health outcomes and therapeutic responses. One such cohort is the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS)[68].

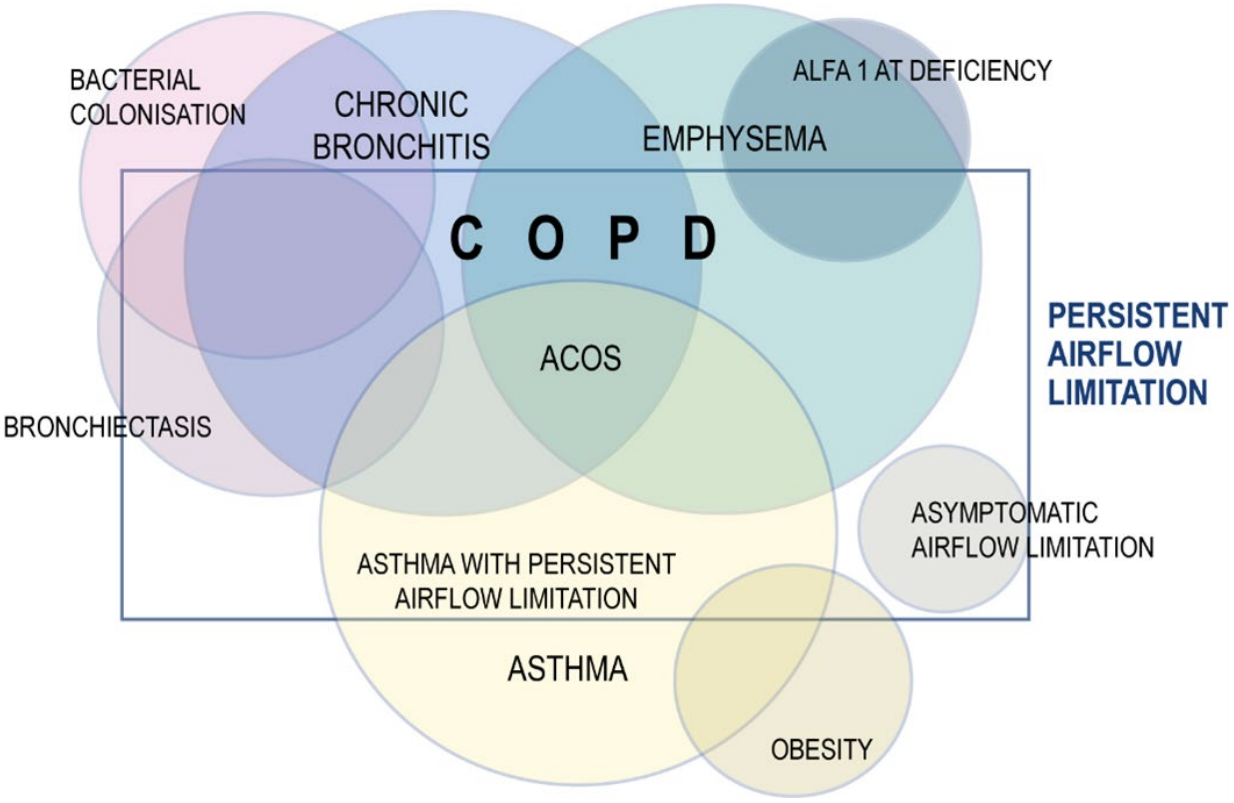


Figure 1.1. Venn diagram summarizing the diverse respiratory conditions that can contribute to airflow limitation in COPD patients[55].

SPIROMICS is a multi-center longitudinal study that enrolled nearly 3,000 participants, including smokers with and without COPD as well as never-smoking controls[68,69]. This large-scale project aims to comprehensively understand the underlying mechanisms, natural progression, and various subtypes of COPD. SPIROMICS has undertaken extensive data collection to support these initiatives, encompassing clinical assessments, lung function tests, radiographic imaging, genetic profiles, blood samples, and lung-derived specimens. This diverse

assortment of biospecimens permits the generation and integration of a broad spectrum of omics datasets (i.e., genomics, proteomics, and metabolomics), enabling comprehensive comparisons between COPD and control participants for both pulmonary and systemic factors. In fact, no COPD cohort worldwide has the detailed phenotyping and breadth of biospecimens found within SPIROMICS. As a result, the data derived from this cohort presents a distinct opportunity for researchers to explore the intricate interplay between genetic variations and molecular pathways that contribute to COPD susceptibility, severity, and progression.

Collective efforts from SPIROMICS and other large multi-center cohorts have suggested a variety of distinct COPD subpopulations, marked by taxonomic annotations (i.e., emphysema predominant vs. airway predominant disease), comorbid conditions, and transcriptomic inflammatory profiles[70–72]. Although these findings represent great strides toward creating more targeted and effective therapeutics, they also underscore that complex biological networks contribute to COPD pathology. Thus, methods that allow for network-level inference could provide valuable new insights into the systems-level mechanisms that underly COPD pathobiology.

1.4 The promise of systems-focused approaches in COPD research

Systems-biology approaches, including data-driven modeling, offer a promising avenue for generating novel, network-level insights. Data-driven modeling is a subset of systems-biology tools that use high-throughput data to model the relationships between variables without requiring knowledge of the underlying system behavior (ex., biological mechanisms). Additionally, data-driven models can accommodate the integration of datasets derived from multiple tissue compartments or anatomic levels and therefore hold the promise of generating integrative (i.e., multi-level or cross-tissue compartment) insights into disease pathobiology.

Data-driven tools consist of unsupervised and supervised algorithms. Unsupervised algorithms use clustering and dimensionality reduction approaches to identify global patterns in a dataset and may aid in identifying novel subgroups within a patient population[73]. Conversely, supervised approaches allow the identification of small combinations ("signatures") of key molecular elements that covary with each other and are associated with clinically relevant outcomes. Once identified, these signatures can be linked to mechanisms using knowledge-based bioinformatics databases and experimental follow-up and validation. The identification of critical players in the network and their linkage to mechanisms provide a starting point for generating insight into disease-specific biology, potential diagnostic or prognostic criteria, and possible targets for combinatorial therapeutic intervention[74].

In COPD, unsupervised data-driven modeling approaches have successfully been employed across a range of omics datasets to aid in identifying distinct patient populations ("phenotypes") with unique molecular fingerprints[75–78]. For instance, researchers have identified four distinct COPD subpopulations with unique peripheral blood gene expression patterns. Each subgroup was characterized by varying degrees of lung function impairment, emphysema, and respiratory symptoms. These subgroups also exhibited differences in biological pathway enrichments, blood cell composition, and inflammatory protein biomarker levels (CRP, fibrinogen, IL-6, and CC16)[75]. Comparable success has been achieved at the proteomic level, where unsupervised clustering of 57 serum protein measurements revealed distinct subpopulations with unique clinical characteristics and comorbidities[77]. In both genomic and proteomic analyses, the enriched pathways that differentiated the identified subgroups were mainly associated with wound healing and inflammatory processes.

Several other studies took an alternative approach by directly applying clustering-based

methods to clinical datasets to identify COPD subgroups with unique clinical presentations[79–82]. In a notably successful study validated in a secondary cohort, Castaldi et al. employed a *k*-means clustering algorithm on quantitative chest CT, spirometric data, and clinical measurements from smokers participating in the COPDGene study. This effort revealed four patient clusters exhibiting unique radiographic patterns of airway disease and emphysema. Two of the clusters, one featuring individuals with mild upper zone emphysema and the other with severe emphysema, showed significant correlations with single nucleotide polymorphisms (SNPs) previously associated with COPD[83].

Although clustering-based approaches have demonstrated significant success in independent investigations, the reproducibility of identified subtypes across different studies is often modest[84,85]. Furthermore, despite dedicated attempts to classify subtypes based on various clinical features, studies have suggested that COPD heterogeneity is more accurately depicted by continuous traits (like airflow limitation or quantitative emphysema) that exist in different degrees within the same individual rather than by distinct and mutually exclusive COPD subtypes or phenotypes[84]. These inherent limitations imply that a more practical approach could involve defining group boundaries and COPD subpopulations based on meaningful clinical outcomes like exacerbation risk, mortality, or FEV_1 decline. Such an alternative strategy can be realized by applying supervised data-driven algorithms.

Numerous studies have highlighted the effectiveness of supervised approaches in identifying predictive signatures associated with predetermined clinical outcomes in COPD. Specifically, researchers have utilized these methods to create polygenic[86], transcriptomic[87], and metabolomic[88] scores (referred to as "risk scores"), enabling assessment of an individual's susceptibility to COPD development. Moreover, supervised techniques have been applied to

transcriptomic data from bronchial epithelial brushings to formulate signatures predicting COPD endotypes. These transcriptomic signatures differentiate COPD patients into T2 or T17-like inflammatory endotypes[71,72]. The importance of these endotype signatures lies in their potential to assist in predicting patient responsiveness to pharmacotherapeutic interventions, including inhaled corticosteroids (ICS).

Despite the promise of these findings, a limitation in current approaches is that they tend to emphasize only the additive significance of each protein in differentiating clinical groups rather than co-variance, which may improve classification ability and can better assist with network inference[89]. Moreover, these studies lack the use of data integrated across multiple physiological compartments. Although sample availability has likely limited previous studies to evaluating molecular markers from a single tissue compartment, studies indicate that data integration across multiple compartments holds promise for improving the prediction of COPD outcomes[58]. In cross-sectional studies, integrating datasets from various molecular levels or anatomical locations improved the ability to classify individuals with COPD from controls[90–92]. Accordingly, we may obtain a deeper understanding of the natural history of COPD by exploring network connectivity at a systems level[57,93].

1.5 Structure of thesis

Given the lack of research regarding network-level alterations in the proteome associated with COPD pathobiology, this thesis aims to identify and evaluate pathogenic changes in protein signaling pathways associated with key COPD subpopulations. We decided to focus our analysis on proteins because these are biologically active factors that directly reflect the current state of patients. When possible, we use data-driven modeling approaches to integrate data across tissue compartments (i.e., lung-derived samples and blood). We approached this challenge with the

following aims:

Aim 1: Elucidate systems-level insights into cross-tissue compartment changes that predict accelerated spirometric progression.

Aim 2: Utilize network-based approaches to identify blood signatures associated with unique COPD subpopulations, including:

Aim 2a: Individuals with accelerated spirometric progression.

Aim 2b: Emphysema progression in younger vs. older individuals.

Aim 3: Identify dysregulated nodes of immune cell-cell communication networks in peripheral blood that are associated with COPD disease state.

The findings generated from this work will provide novel insights into proteomic network alterations associated with COPD pathogenesis and progression. Results will inform the generation of hypotheses about important pathway dysregulations contributing to COPD pathology that can be further investigated in validation cohorts and follow-up murine studies.

Completion of these aims will be presented in the following format: **Chapter 2** presents published work in which data-driven approaches were applied to integrated blood and BAL proteins to identify cross-tissue compartment signatures capable of predicting COPD patients' risk of accelerated spirometric progression (Aim 1). The related supplemental materials for this work are presented in **Appendix A**. Validation of the findings in Chapter 2 presented are presented in **Chapter 3** (Aim 2a). **Chapter 4** presents unpublished work that uses data-driven approaches to uncover differential mechanisms driving the progression of emphysema in younger and older individuals using plasma proteomics samples from ever-smokers (Aim 2b). Supplemental material for Chapter 4 is in **Appendix B**. **Chapter 5** includes an analysis of secretion profiles from patient-derived immune cells to explore cytokine-chemokine network

alterations associated with COPD (Aim 3), with supplemental material in **Appendix C**.

Chapter 2 A Blood and Bronchoalveolar Lavage Protein Signature of Rapid FEV_1 Decline in Smoking-Associated COPD

2.1 Original publication information

This chapter was originally published as a scientific article:

DiLillo KM, Norman KC, Freeman CM, Christenson SA, Alexis NE, Anderson WH, Barjaktarevic IZ, Barr RG, Comellas AP, Bleecker ER, Boucher RC, Couper DJ, Criner GJ, Doerschuk CM, Wells JM, Han MK, Hoffman EA, Hansel NN, Hastie AT, Kaner RJ, Krishnan JA, Labaki WW, Martinez FJ, Meyers DA, O'Neal WK, Ortega VE, Paine R, Peters SP, Woodruff PG, Cooper CB, Bowler RP, Curtis JL, Arnold KD. (2023). A blood and bronchoalveolar lavage protein signature of rapid FEV_1 decline in smoking-associated COPD. *Scientific Reports*, 13(1):8228. https://doi.org/10.1038/s41598-023-32216-0

Changes made to the original document are mainly cosmetic to adhere to the format of this document.

2.2 Abstract

Accelerated progression of chronic obstructive pulmonary disease (COPD) is associated with increased risks of hospitalization and death. Prognostic insights into mechanisms and markers of progression could facilitate development of disease-modifying therapies. Although individual biomarkers exhibit some predictive value, performance is modest, and their univariate nature limits network-level insights. To overcome these limitations and gain insights into early

pathways associated with rapid progression, we measured 1305 peripheral blood and 48 bronchoalveolar lavage proteins in individuals with COPD [n=45, mean initial forced expiratory volume in one second (FEV_1) 75.6±17.4% predicted]. We applied a data-driven analysis pipeline, which enabled identification of protein signatures that predicted individuals at-risk for accelerated lung function decline (FEV_1 decline ≥70 mL/year) ~6 years later, with high accuracy. Progression signatures suggested that early dysregulation in elements of the complement cascade is associated with accelerated decline. Our results propose potential biomarkers and early aberrant signaling mechanisms driving rapid progression in COPD.

2.3 Introduction

Chronic obstructive pulmonary disease (COPD), a leading cause of death in the United States[94], accounts annually for >600,000 hospitalizations[95] and $30 billion in direct health expenditures[4]. The course of COPD is heterogeneous. Such heterogeneity is exemplified, at least in part, by the highly variable rates of annualized decline in forced expiratory volume in one second (FEV_1) observed in prospective observational cohort studies[96–99]. This variability between rates of lung function decline causes some individuals to experience relatively stable courses, while in others, accelerated loss of function leads to severe breathlessness and increased risk of hospitalization and death[98]. Because phenotypic diversity is underpinned by biological heterogeneity, there is growing interest in identifying molecular biomarkers to predict multiple aspects of COPD progression. Such molecular markers could help explain phenotypic heterogeneity and facilitate the early detection of individuals at risk for accelerated lung function decline, enabling personalized management to arrest disease progression. Additionally, biomarkers could direct research into underlying pathogenic mechanisms, uncovering novel therapeutic targets.

To date, accelerated FEV_1 decline has been associated with individual blood proteins, including club cell secretory protein 16 (CC16)[58,59], soluble receptor for advanced glycation end-products (sRAGE)[58], fibrinogen[58], C-reactive protein (CRP)[58,61], and IL-6[60], although contradictory reports exist[61,100–102]. The ratio of leptin to adiponectin in plasma also demonstrated predictive value for rapid lung function decline but showed only moderate sensitivity (63.5%) and specificity (65.1%)[103]. While valuable, univariate analyses of candidate biomarkers have provided limited insights into underlying mechanisms, a shortcoming that could be complemented by network-level analysis of large numbers of proteins.

Previous studies have also been limited to measuring biomarkers in a single tissue compartment due to sample availability. However, data indicate that combinations of proteins appear to be better predictors of multiple COPD outcomes (including FEV_1 decline) than individual factors[8], especially when derived from multiple compartments. For example, in cross-sectional studies, integrating datasets across multiple molecular levels, anatomical locations, or both, improved the ability to classify individuals with a smoking history by COPD status and to uncover novel disease-associated pathways[90–92]. Accordingly, analyses of lung function decline may benefit from evaluating systems-level integrated networks, as they are more likely to capture the diverse biology driving airflow obstruction[57,93].

Data-driven modeling is one approach that allows inference of network-level relation-ships driving progression. By permitting data integration across multiple tissue compartments, data-driven modeling generates systemic networks (“signatures”) of co-varying biological factors associated with disease phenotypes. Identified signatures can be linked to pathogenic mechanisms, providing insight into potential targets for follow-up experiments or as biomarkers for therapeutic intervention. Previously, we successfully used these approaches to identify blood

and bronchoalveolar (BAL) protein signatures associated with disease state and progression in idiopathic pulmonary fibrosis[89,104]. We have also used the approach to integrate blood and sputum proteins to define signatures that differentiated stable and exacerbated COPD states[105].

Here, to gain insights into cross-compartment mechanisms associated with a greater lung function decline in COPD, we applied an integrative data-driven modeling pipeline to proteins recovered from matched blood and BAL samples from participants in the bronchoscopy sub-study[106,107] of the SubPopulations and InteRmediate Outcome Measures In COPD Study (SPIROMICS)[68]. Our results suggest that proteomic signatures can effectively detect individuals at increased risk of accelerated lung function decline. They also provide insights into COPD progression mechanisms that can be further investigated in validation cohorts and follow-up murine studies.

2.4 Results

2.4.1 Participant characteristics

We initially analyzed participants of the SPIROMICS bronchoscopy sub-study who had COPD, paired baseline (V1) and final (V5) spirometry, plus matched proteomic measurements from plasma samples and BAL samples (n=45) (**Supplemental Figure A.1**). Participants' mean (±SD) age at V1 was 63±7.7 years; they had a follow-up time of 6.3±0.9 years (**Table 2.1**). To characterize rapid progression, we dichotomized participants based on their annualized FEV_1 decline (ΔFEV_1): greater decliners ($<30^{th}$ percentile) (n=14) versus lesser decliners ($\geq 30^{th}$ percentile) (n=31) (**Figure 2.1a**). This threshold equaled a ΔFEV_1 of –70 mL/year. Greater decliners trended non-significantly to be male (85.7% vs. 54.8%) but were well-matched for other demographic criteria. Despite significantly higher FEV_1% predicted (p=0.017) and absolute FEV_1 (p=0.015) at V1, they experienced a 3.6-fold greater ΔFEV_1 compared to their lesser

decliner counterparts (-104.6 ±32.0 vs. -28.8 ±21.5 mL/year). The observed association between faster decline and higher baseline lung function is in line with previous reports[96,108].

Table 2.1. Baseline characteristics of COPD cases

	All (N=45)	Greater Decliners‡ (N=14)	Lesser Decliners (N=31)	P-Value†
Age*	63.4 (± 7.75)	64.2 (± 6.24)	63.1 (± 8.41)	0.65
Currently Smoking*	15 (33.3%)	5 (35.7%)	10 (32.3%)	>0.99
BMI*	27.8 (± 4.91)	27.9 (± 3.67)	27.8 (± 5.43)	0.93
Sex (Male)	29 (64.4%)	12 (85.7%)	17 (54.8%)	0.09
Race (White / Other)	37/8 (82.2%)	12/2 (85.7%)	25/6 (80.7%)	>0.99
ICS use* (yes)	17 (37.8%)	3 (21.4%)	14 (45.2%)	0.18
FEV_1* (% predicted)	75.6 (± 17.4)	**84.2 (± 13.1)**	**71.1 (± 17.7)**	**0.017**
FEV_1/FVC*	0.58 (± 0.09)	0.60 (± 0.08)	0.57(± 0.10)	0.27
FEV_1* (L)	2.27 (± 0.68)	**2.63 (± 0.60)**	**2.11 (± 0.66)**	**0.015**
Visit 5 FEV_1 (L)	1.94 (± 0.67)	1.97 (± 0.67)	1.93 (± 0.68)	0.85
Time from baseline to Visit 5 (yrs.)	6.31 (±0.86)	6.25 (± 0.76)	6.33 (± 0.91)	0.79
Time from baseline to bronchoscopy (months)	20.3 (±11.6)	20.4 (± 10.0)	20.3 (12.5)	0.96
ΔFEV_1 (mL/yr.)	-52.4 (± 43.3)	**-104.6 (± 32.0)**	**-28.8 (±21.5)**	

Two-sample, two-tailed t-test or Fisher's exact test were used to determine significant differences. Bold values denote significant differences between greater and lesser decliners.

* Demographic information from baseline visit (Visit 1)

† P-values are associated with differences between greater decliner and lesser decliner groups

‡ Decline in FEV_1 (mL/yr.) ≥70 mL/year (see pg. 34)

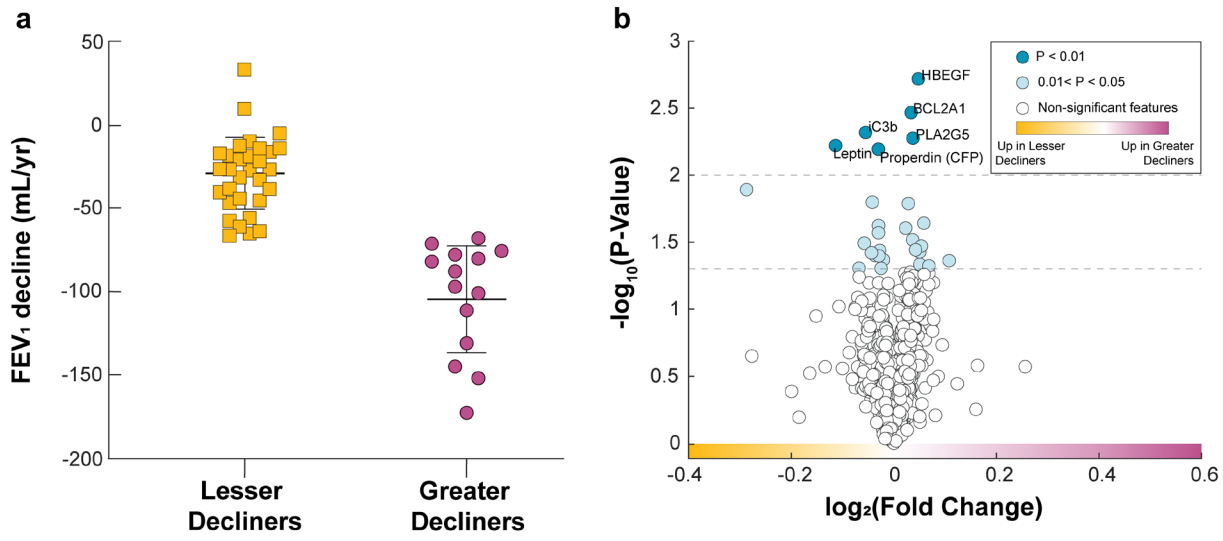


Figure 2.1. Individual blood and BAL proteins cannot discriminate between annualized greater versus lesser rates of FEV_1 decline in COPD.
(a) Comparison of annualized post-bronchodilator FEV_1 decline from V1 to V5. Decline was calculated as (V5 FEV_1 – V1 FEV_1)/ time, where time is the duration in years between V1 and V5 for each participant. (b) Volcano plot of blood and BAL proteins. Light and dark blue protein markers have a p-value < 0.05 and < 0.01, respectively, after a two-sampled two-tailed t-test. All depicted p-values are before correction for multiple comparisons. No proteins remained significant after applying the Benjamini-Hochberg false discovery rate (FDR) correction for multiple comparisons ($\alpha = 0.05$).

2.4.2 Individual blood and BAL proteins cannot discriminate between rates of longitudinal lung function decline.

We first determined whether differences existed between greater decliners and lesser decliners in multi-compartment protein expression measured early in the study, using the concentrations of 1305 blood and 25 BAL proteins measured with SOMAScan and Luminex technology, respectively. To reduce biases associated with the unequal distribution of women across classes (14.3% vs. 45.2%), we first removed proteins (n=8) that exhibited significant associations with sex (**Supplemental Figure A.2**). Across the remaining 1322 proteins, 28 (2.1%) had a mean concentration that differed significantly ($p<0.05$) between groups (**Figure 2.1b**). Of these, 13 were increased in greater decliners [$\log_2$ fold change (FC)>0] and 15 were increased in lesser decliners (FC<0). The top six most significantly different proteins ($p<0.01$)

were all identified in blood: Heparin-binding EGF-like growth factor (HBEGF; FC=0.05), BCL2-related protein A1 (BCL2A1; FC=0.03), inactivated Complement C3b (iC3b; FC=-0.05), Calcium-Dependent Phospholipase A2 (PLA2G5; FC=0.03), Leptin (FC=-0.11), and Properdin (CFP; FC=-0.03). None of the 28 proteins remained significant after Benjamini-Hochberg adjustment.

2.4.3 Data-driven modeling based on protein measurements identifies a novel multi-compartment signature that prospectively differentiates progression phenotypes.

We next used data-driven modeling approaches to identify a multi-compartment signature of co-varying proteins associated with greater FEV_1 decline. Here, we combined all 1322 protein measurements (1297 blood and 25 BAL) into a single dataset, then applied elastic net (EN) in tandem with partial-least squares discriminant analysis (PLSDA). First, EN regularization was applied iteratively to 2000 subsets of randomly resampled data. Then, based on their selection frequency throughout the iterations, proteins were ranked (most to least frequent) and fed stepwise into the PLSDA algorithm. We evaluated PLSDA model performance at each step using 6-fold cross-validation (CV) and selected the model with the highest CV accuracy as the optimal signature.

Using this feature selection pipeline, we identified a signature of 52 proteins (51 blood and 1 BAL) that distinguished greater decliners (FEV_1 decline ≥70 mL/year), along the latent variable 1 (LV1) axis, with 98.4% calibration and CV accuracy, 100% sensitivity, and 96.7% specificity (**Figure 2.2a-c**). Permutation tests performed on participant scores across the first two principal components of PCA models generated with the 52-feature signature, show no significant influence of baseline ICS use ($p = 0.35$) or smoking status ($p = 0.57$) on participant classification (**Supplemental Figure A.3**). To confirm the accuracy of the selected features, we

compared its cross-validated accuracy to 1000 random signatures, generated by selecting iterative groups of 52 random proteins from the original dataset. None of the random signatures outperformed the optimal model ($p<0.0001$) (**Supplemental Figure A.4**). We also observed a significant, albeit moderate, Pearson correlation between LV1 scores and plasma concentrations of CRP ($r_p=0.33$, $p=0.03$), a protein previously associated with FEV_1 decline[58,61]. However, LV1 scores exhibited no significant correlations with other reported blood markers of spirometric decline, including IL-6[60] ($r_p=0.18$, $p=0.29$), fibrinogen[58] ($r_p=-0.01$, $p=0.93$), and matrix metalloproteinase 9 (MMP-9)[61] ($r_p=-0.27$, $p=0.08$).

Lack of a formal definition of "rapid progression" in COPD has led to literature variability, so we explored whether the identified 52-feature signature maintained significance across alternative characterization approaches. Recently, using the entire SPIROMICS dataset, Anderson et al. proposed a threshold-based definition, classifying progression into three groups based on annualized FEV_1 declines: rapid decliners (>100 mL/year), decliners (20-100 mL/year), and stable/ improvers (< 20 mL/year)[109]. The limited number of participants in our bronchoscopy sub-study exhibiting such extreme decline hindered direct exploration of this definition. However, an exploratory PCA using the 52-feature multi-compartment signature demonstrated significant signature enrichment in rapid decliners thus defined (**Supplemental Figure A.5**), suggesting our model reliably extends to this more rigorous definition. A regression-based analysis also found that participant scores on LV1 correlated highly with annualized declines in FEV_1 ($r_p=0.758$, $p<0.0001$), even after adjustment for age, race, height, sex, baseline FEV_1% predicted, smoking status, pack-years, and ICS use ($p<0.0001$) (**Figure 2.2d**). Alternative estimations of FEV_1 decline using all available longitudinal spirometry (rather than just V1 and V5) produced similar results, as did classifications using FEV_1% predicted in

lieu of absolute FEV_1 volumes (**Supplemental Figure A.6**). These results suggest that our approach captures complex progression trends and is largely not skewed by demographic factors influencing lung capacity. For the remainder of our analysis, we use the -70 mL/year definition.

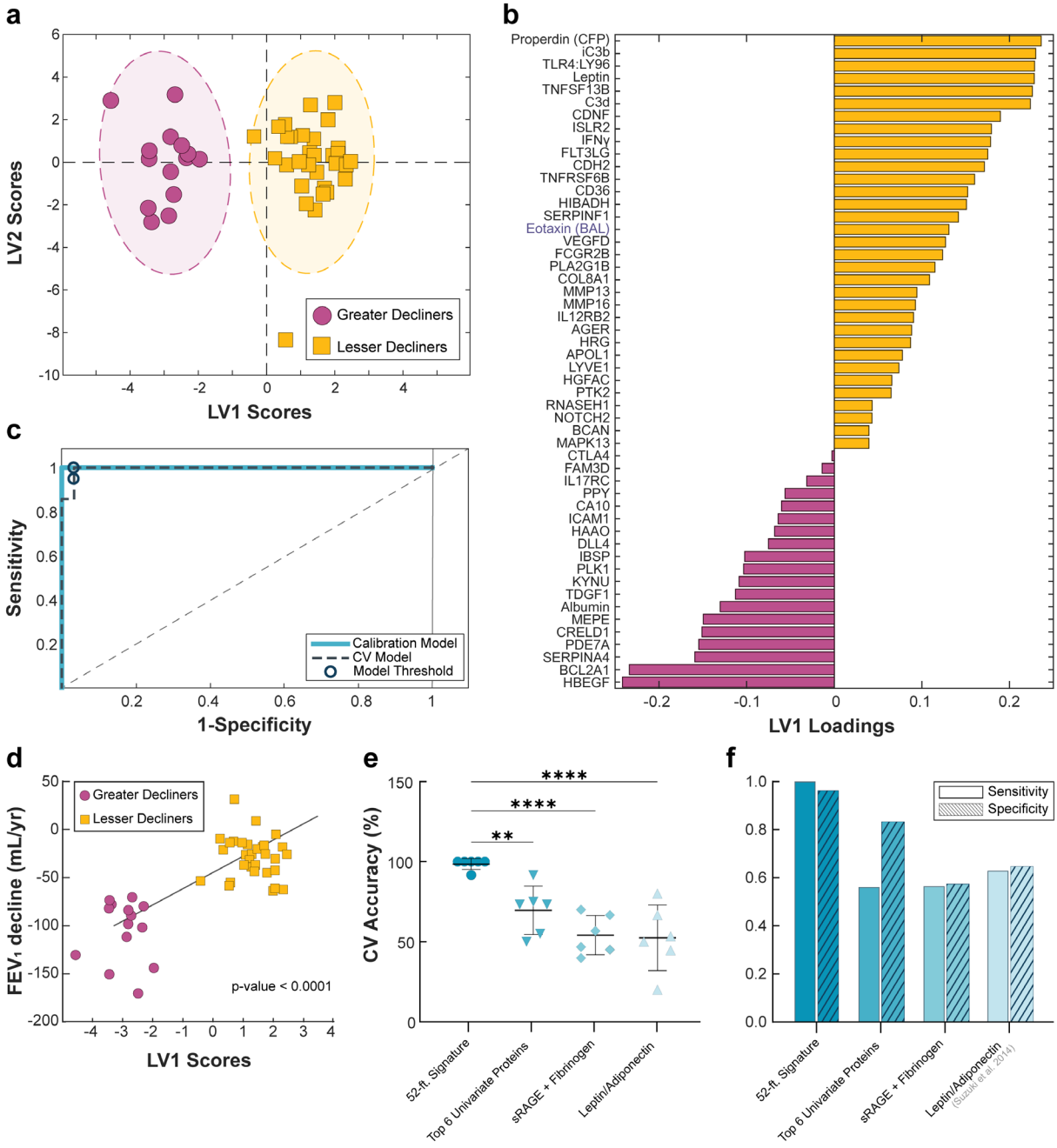


Figure 2.2. A 52-feature Elastic Net (EN) signature identified individuals at-risk for FEV_1 decline ≥70 mL/year with high accuracy.
(a) PLSDA scores plot highlighting strong differentiation between greater decliners (magenta) and lesser decliners (yellow), separating the two groups with 98.4% cross-validation (CV) and calibration accuracy. (b) Loadings on

latent variable 1 (LV1) (with negatively loaded proteins being comparatively increased in greater decliners and positively loaded proteins being comparatively reduced) captured 11.9% of the total variance in the data. (c) ROC curve of 52-feature signature suggests greater decliners classification with 100% sensitivity and 96.8% specificity in the cross-validated model. (d) LV1 scores were associated with annualized FEV_1 decline (mL/yr.). P-values and fit line shown for linear models adjusted for age, race, height, sex, baseline FEV_1% predicted, smoking status, pack-years, and inhaled corticosteroids (ICS) use within three months of baseline visit. (e-f) Comparison of (e) 6-fold CV accuracies, (f) sensitivities, and specificities between the 52-feature EN signature, a collection of the six top proteins identified in Figure 2.1, and literature-based models. All reported values are from cross-validated PLSDA models, unless otherwise noted. One-way ANOVA with Dunnett's post hoc test; **p<0.01, ****p < 0.0001.

2.4.4 The progression signature is enriched for proteins involved in the complement system

Having generated a high-performing progression signature, we sought to understand the biological implications of its components. Unsupervised hierarchical clustering identified greater decliners with 88% accuracy (**Figure 2.3a**). A Metascape analysis found 20 significantly enriched ontology clusters (**Figure 2.3b**), with only three, related to aging and phosphorylation-dependent signal transduction, shared between groups. Lesser decliners displayed unique enrichment of 16 clusters related primarily to inflammation and immune functions. Notably, the only uniquely enriched cluster in greater decliners was associated with the complement system (q=4.10e-04) (**Figure 2.3c-d**). Proteins in this cluster included blood albumin, bone sialoprotein 2, intracellular adhesion molecule 1, interferon-gamma, kynureninase, and three complement proteins (iC3b, C3d, Properdin). These three complement proteins were involved in 10 of 20 total clusters and represented three of the top eight loaded proteins in the PLSDA, indicating a potentially significant impact of complement processes in COPD progression.

2.4.5 Patterns of dysregulation in complement-associated proteins precede accelerated FEV_1 decline

These enrichment data, coupled with the known importance of the complement cascade to immunity, suggest that complement proteins may be more globally altered in greater decliners than was captured in the original signature. To explore this possibility, we performed PCA using the concentration of 22 complement proteins measured in plasma. Results suggested that our

groups differed significantly in complement profiles as measured by scores across PC1 ($p = 0.0045$) (**Supplemental Figure A.9**). We selected PC1 scores as the PC of interest, as only they correlated significantly with annualized FEV_1 decline ($p < 0.001$) (**Supplemental Figure A.10a**).

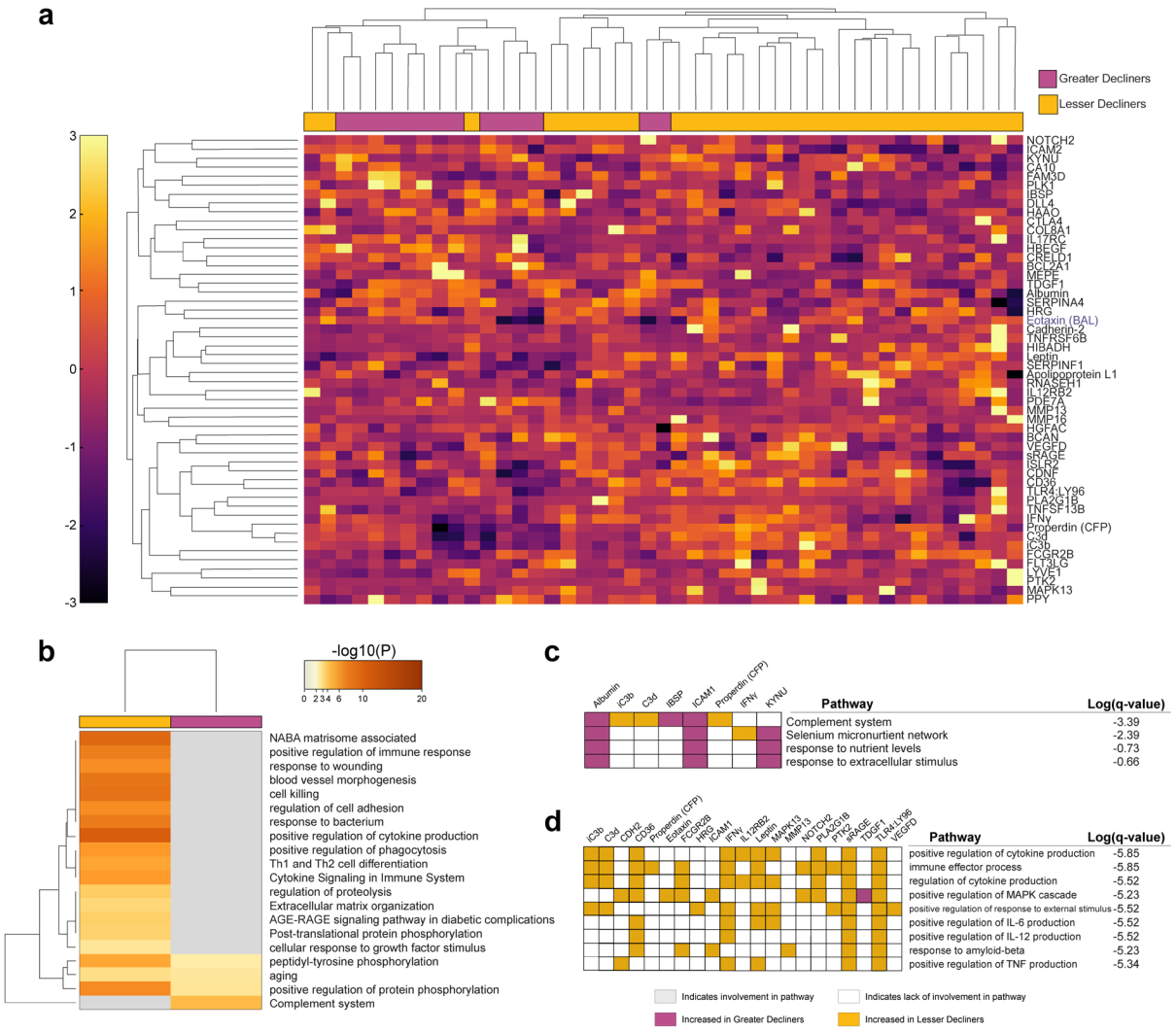


Figure 2.3. Clustering of COPD participants by the EN-identified signature highlights distinct regulation of immune-associated processes.
(a) Hierarchical clustering of the 52-feature signature highlights distinct clustering of greater decliners (magenta) and lesser decliners (yellow). Only 5 out of the 45 participants were misclassified (Sensitivity: 85.7%, Specificity: 90.3%). BAL proteins denoted by blue text. (b) Significantly enriched ontology clusters by Metascape analysis. (c, d) Pathways encompassed in the (c) complement cascade cluster and in the (d) positive regulation of cytokine production cluster are included pop-out table. Hatched squares indicate protein involvement in a particular pathway, colorations of magenta or yellow represent a relative elevation of the protein concentration in greater decliners or lesser decliners, respectively.

To explore whether the observed patterns were specific to the greater decliners rather than COPD more generally, we extended our analysis to include the reference group of TEPPS (n=38). Interestingly, the baseline profiles of TEPPS were similar to that of lesser decliners. In contrast, complement profiles from greater decliners differed from those of lesser decliners and TEPPS, both as visualized using PCA (**Figure 2.4a**) and via direct comparison of PC1 scores (**Figure 2.4b**). The variance in complement profiles, as captured by PC1, correlated with FEV_1 decline (p=0.004) (**Supplemental Figure A.10b**), a relationship that remained significant after adjustment for age, race, height, sex, FEV_1% predicted, smoking status, pack-years, and ICS use (p=0.012) (**Supplemental Figure A.10c**). In univariate comparisons, among the 22 complement proteins, only C1r, iC3b, C3d, Properdin, and C4 reached statistical significance (**Supplemental Figure A.11**). Performing permutation tests with participant scores across the first two principal components labeled by key clinical variables, we show no significant influence of baseline ICS use (p = 0.36) or current smoking status (p = 0. 70) on observed complement profiles (**Supplemental Figure A.12**). Collectively, these findings reinforce the Metascape results, suggesting that early patterns of complement dysregulation are specific to a more rapidly progressing phenotype.

2.4.6 Alternative minimal signatures highlight a small number of proteins that maintain high predictive power

Although our model's large size proved advantageous in exploring functional enrichments associated with accelerated FEV_1 decline, it is complex and hence costly if clinically implemented as a prediction tool. Therefore, we next explored whether smaller sub-models ("minimal signatures") maintained predictive value. Using insights from our step forward PLSDA pipeline, we visually analyzed the trade-off between model size and performance. Two

minimal signatures, with 6 and 11 features, respectively, had CV accuracies similar to our optimized model (CV accuracy: 6-feature, 81.6%; 11-feature, 88.4%; 52-feature, 98.4%) (**Supplemental Figure A.13a**). The 6-feature model consisted of BAL Eotaxin and blood iC3b, Leptin, Cadherin-2, Heparin-binding EGF-like growth factor (HBEGF), and Teratocarcinoma-derived growth factor 1 (TDGF1). The 11-feature model comprised those six, plus blood-derived Low affinity immunoglobulin gamma Fc region receptor IIb (FCGR2B), Interferon gamma (IFN-γ), Carbonic anhydrase-related protein 10 (CA10), Apolipoprotein L1, and Lymphatic vessel endothelial hyaluronic acid receptor 1 (LYVE1). ROC curves created for the 11- and 6-feature models resulted in areas under the curve (AUC) of roughly 0.982 and 0.947 for the calibration models and 0.935 and 0.878 for the cross-validated models, respectively (**Figure 2.5a**; **Supplemental Figure A.14**). Intriguingly, each minimal signature included at least one protein from both tissue compartments.

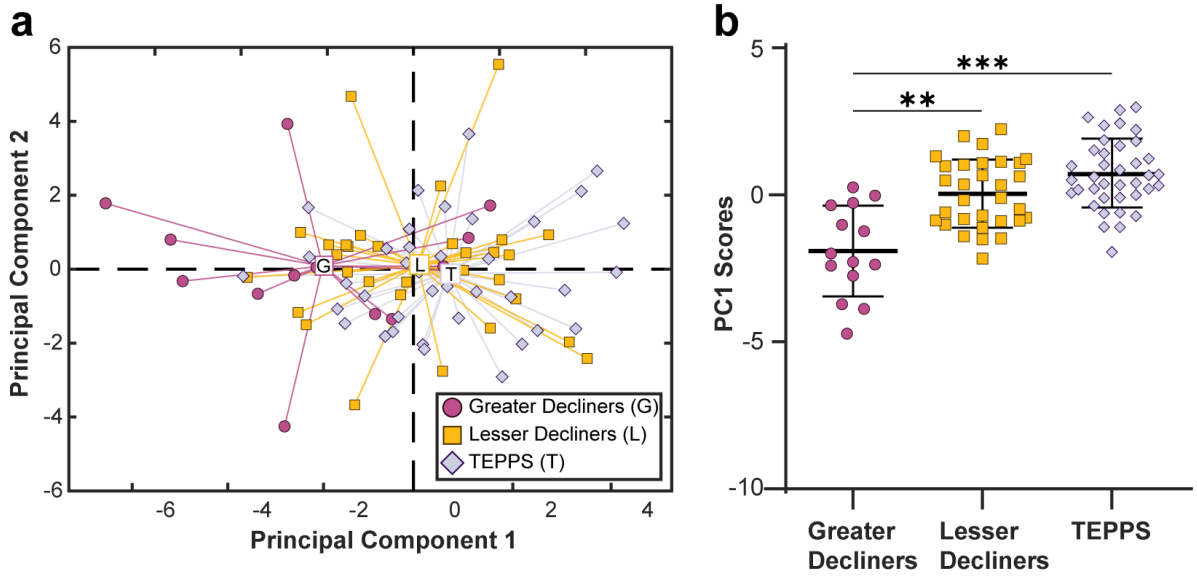


Figure 2.4. Complement profiles in COPD lesser decliners behave more similarly to TEPPS than COPD greater decliners.
(a) PCA completed using all complement proteins measured in plasma (C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B, Factor D, Properdin) of greater decliners (circles), lesser decliners (squares), and a reference group of tobacco-exposed people with preserved spirometry (TEPPS) (diamonds). First two principal components (PCs) capture 33.9% of the variance in the dataset. (b) Comparison of scores on PC1 (one-way ANOVA with Tukey's post-hoc test; **p<0.01, ***p<0.001).

We tested proteins from both blood and alveolar compartments to maximize the chances of understanding underlying biology in COPD, specifically the link between distal airways and systemic events. However, due to the invasiveness of bronchoscopy, biomarkers obtained solely from the blood would be preferable. Accordingly, we evaluated a minimal blood signature by applying our step forward EN/PLSDA algorithm exclusively to the 1297 blood proteins measured in the same COPD participants (n=45). Analysis identified 5- and 10-feature signatures that maintained strong cross-validated performance (CV accuracy: 5-feature, 81.6%; 10-feature, 86.8%) (**Supplemental Figure A.13b**). The 5-feature model included blood-derived iC3b, Cadherin-2, Leptin, HBEGF, and TDGF1; the 10-feature model comprised those five, plus CA10, IFN-γ, FCGR2B, LYVE1, and Apolipoprotein L1. Both minimal blood signatures displayed a slight drop in performance compared to their multi-compartment counterparts of nearest sizes (10- and 5-feature signatures, AUC 0.947, 0.901 for calibration models and 0.912, 0.862 for CV models, respectively). However, comparisons using both AUCs and 6-fold cross-validation found no significant difference in the performance of any of the minimal signatures and the optimal 52-feature model (**Figure 2.5a**; **Supplemental Figure A.15**).

Finally, we subjected the minimal signatures to cross-validated analyses relative to multivariate signatures identified through our univariate analysis or published literature. Comparisons using 6-fold CV accuracy indicated that overall performance was largely sustained, with the 6- and 11-feature multi-compartment signatures and the 10-feature blood signature significantly outperformed literature-based models and trended towards outperforming a signature based on the top 6 univariate proteins (**Figure 2.5b-c**). The 5-feature blood signature did not reach statistical significance in any comparison, though its performance over literature-based models was substantially improved. All four signatures showed a >20% increase in

a

Model Name	Signature Type	Proteins in Signature	AUC Calibration	AUC CV	p-value*
52-ft Signature	Cross-compartment	---	1.00	0.9954	
11-ft Signature	Cross-compartment	iC3b, Leptin, CDH2, HBEGF, Eotaxin (BAL), TDGF1, FCGR2B, IFNγ, CA10, ApoL1, LYVE1	0.9816	0.9355	0.469
6-ft Signature	Cross-compartment	iC3b, Leptin, CDH2, HBEGF, Eotaxin (BAL), TDGF1	0.9470	0.8779	0.215
10-ft Signature	Blood	iC3b, CDH2,Leptin, HBEGF, TDGF1, CA10, IFNγ, FCGR2B, LYVE1, ApoL1	0.9470	0.9124	0.215
5-ft Signature	Blood	iC3b, CDH2,Leptin, HBEGF, TDGF1	0.9009	0.8618	0.085

*p-values based on comparisons to 52-ft model using calibration models

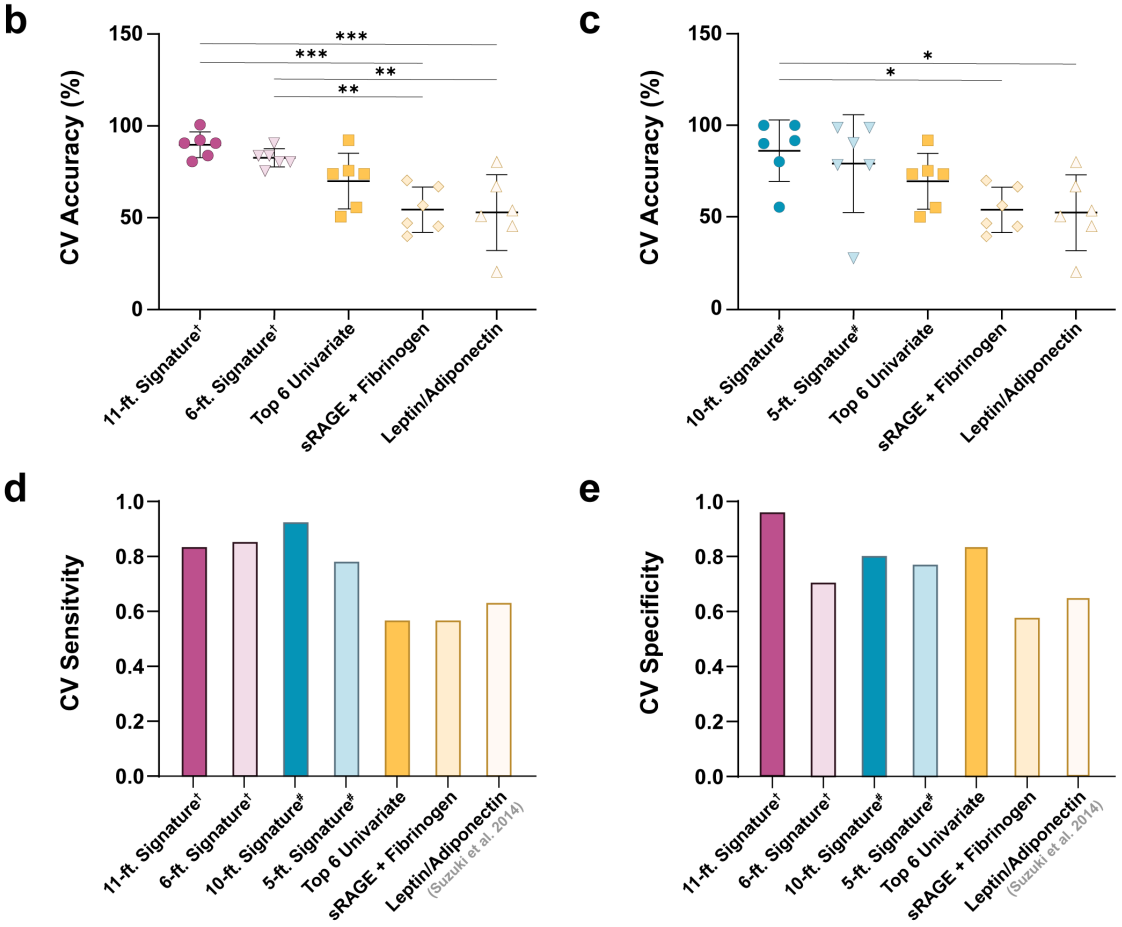


Figure 2.5. Select subgroup of signature proteins retain high predictive value for accelerated FEV_1 decline.
(a) Table denote AUCs for both ROC curves generated from calibration and cross-validation PLSDA models. AUCs of all calibration models were compared to that of the optimal (52-feature model) using the Hanley and McNeil method[110]. (b) Comparisons of 6-fold CV accuracies of multi-compartment and (c) blood-only biomarker models a collection of the 6 top proteins identified in Fig. 1, and literature models, as determined by ANOVA with Bonferroni's post hoc test (*p<0.05, **p < 0.01, ***p<0.001). (d) Sensitivity and specificity of signatures. All reported values are from cross-validated PLSDA models, unless otherwise noted. (†: multi-compartment model, #: blood-only model).
CV: cross-validated; AUC: area under curve.

sensitivity, with lesser though considerably improved specificity versus other models (**Figure 2.5d-e**). Collectively, these findings imply that small, proteomic signatures can differentiate individuals with tobacco smoking-associated COPD at risk for rapid lung function decline with high accuracy.

2.5 Discussion

This study used two complementary datasets from COPD participants in the SPIROMICS bronchoscopy sub-study to generate insights into early, aberrant signaling mechanisms associated with accelerated lung function decline. Using a systems analysis, we identified a multivariate signature of early blood and BAL proteins that predicted individuals at-risk for greater FEV_1 decline ≥70 mL/year ("greater decliners") with >98% accuracy. Investigation of this signature disclosed that differences in longitudinal FEV_1 decline are associated with variability in host immune and defense responses, with greater decliners uniquely exhibiting early dysregulated patterns of complement protein expression. Finally, refinement of the signature identified a minimal model with 10 blood proteins that, if validated, may serve as a clinically feasible prognostic tool. This work complements previous predictions of COPD progression[58–61,100–102] by starting from a data-driven approach, rather than prior knowledge, to obtain unbiased insights into cross-compartment proteins and pathways driving accelerated airflow obstruction.

To our knowledge, this is the first longitudinal study of COPD progression to use integrated proteomic datasets derived from blood and BAL samples. Our tandem EN and PLSDA approach identified concise proteomic signatures from thousands of proteins measured across multiple tissue compartments. This framework accurately differentiated individuals with COPD who sustained declines in lung function ≥70 mL/year based on proteins measured early

after participant enrollment. We specifically sought to explore integrated lung and systemic compartment models because COPD has coupled local and peripheral manifestations[54]. Although several proteins were measured in lung and blood tissue compartments, classifying signatures did not select any matched BAL/blood proteins, which may be partly due to differences in measurement platforms. Moreover, that only one BAL protein was identified in the 52-feature signature is likely in part a consequence of the marked difference in numbers of analytes in blood and BAL (1297 vs. 25). Currently, the SOMAmer technology used in our plasma analyses has not been validated for use in BAL samples. However, the importance of multi-compartment representation is exemplified by the over 15% improvement in calibration model sensitivity on adding BAL Eotaxin to the 5-feature blood signature.

Our multi-compartment 52-protein signature was enriched for inflammation and immune responses processes, consistent with the central role of immune dysregulation in COPD pathogenesis[34,111,112]. Chief among these processes was the complement system. By providing evidence that complement alterations precede accelerated FEV_1 decline, we extend previous associations between the levels of blood-derived complement proteins (C3[113,114], C4[115], C4b[116], C5a[117], C9[118], Factor B[116]) and COPD status (case vs. control), cross-sectional analyses of FEV_1% predicted[119,120], and emphysema severity[121]. How such aberrations might contribute to airflow limitation is unknown. In a murine model, C3 cleavage contributed to smoking-induced emphysema via an influx of conventional dendritic cells[122], a cell type that can initiate both innate and adaptive immune responses. Our findings regarding involvement of complement proteins are specific to blood. Although our chosen assay system could not analyze complement components in BAL, complement dysregulation might extend into the lung, as suggested by altered levels of C5a in the sputum in COPD[117,123] and airway C3 deposits in lungs in smoking-

associated emphysema[122]. Airway epithelial cells have also been shown to secrete and store C3, suggesting the presence of a localized C3 supply that may aid host defense[124]. The observed global patterns of complement dysregulation were robustly associated with FEV_1 decline, but SomaLogic aptamers cannot reliably distinguish between complement cleavage products and their parent proteins. Hence, we cannot explore the relationship between global complement levels and pathway activation. However, our observation of complement dysregulation before FEV_1 decline provides compelling temporal evidence that the complement pathway contributes to accelerated progression.

Focusing on eventual clinical feasibility, we identified an alternative parsimonious 10-protein signature derived using only peripheral blood; its potential prognostic value, if validated, is suggested by its superior performance to reported multivariate biomarkers of FEV_1 decline[58,103]. To our knowledge, except for leptin[103], these proteins have not been associated with accelerated loss of lung function. Consistent with previous studies[58–61], the proteins in this signature are primarily related to immune and inflammation responses (leptin, iC3b, IFN-γ, FCGR2B, APOL1). However, we also observed notable contributions from endothelial-mesenchymal transition (EMT) proteins (Cadherin-2 and HBEGF). EMT is active in both large and small airways of COPD and relates to airflow obstruction[125–127]. Cadherin-2 is increased in epithelial cells from COPD participants as compared to healthy controls[128]. Similarly, serum and sputum HBEGF levels have been positively associated with COPD severity measures, including FEV_1% predicited[129] and CAT score[130]. The final proteins involved in the signature (TDGF1, LYVE1, CA10) have not previously been associated with COPD. However, in lung cancer, which is thought to share overlapping etiologic features, TDGF1 (increased in greater decliners) predicts poor progression-free survival[131], while LYVE1 (increased in lesser decliners) is

associated with reduced metastasis and mortality[132]. Even in this parsimonious signature, we observed diverse biological enrichment. These findings emphasize the value of multivariate signatures in evaluating heterogeneous conditions such as COPD, where abnormalities in several pathways or pathway constituents likely drive a singular clinical outcome.

Our study has limitations. Chief among these is the lack of a validation cohort as, to our knowledge, none currently exists with both BAL protein measurements and longitudinal follow-up. Missing data also limited our sample size. BAL was not collected successfully on all participants, chiefly due to airway collapse during the procedures, and not all participants completed V5. Hence, we have relatively small numbers of participants, disproportionately non-Hispanic whites. To mitigate the influence of individual participants, we applied an iterative bootstrapping framework during model generation. Still, the generalizability of our findings remains unclear, given the issues of limited heterogeneity based on sample size and demographics. We recognize that we cannot definitively conclude whether the identified signatures precede lung function decline, as decline may have been ongoing prior to baseline sample acquisition. However, the modestly higher baseline FEV_1 observed in greater decliners support this possibility. Additionally, the cross-sectional nature of our proteomic data limits any insight into the temporal stability of our identified signatures. Because there is no universally agreed-upon definition of rapid progression, we used a percentile cut-off in FEV_1 decline, as used by others[97,103]; resulting in a cut-off (≥70 mL/year) similar to previous reports. However, other threshold-based definitions (i.e., FEV_1 decline >100 mL/year) have been proposed[109]. Few participants (n=6) experienced declines >100 mL/year in our data, providing insufficient power to investigate this definition accurately. However, in an exploratory PCA, we show that our 52-feature multi-compartment is enriched significantly in participants with declines >100 mL/year,

suggesting our signature extends to this more stringent characterization. Moreover, while we acknowledge that a fixed cut-off definition may favor an overrepresentation of males as greater decliners due to physiologic differences in lung function measurements between sexes, we are underpowered to explore sex-stratified. Still, alternative estimations of FEV_1 decline using FEV_1% predicted, which accounts for age, sex, and body composition, in lieu of absolute FEV_1 volumes, produced similar results, suggesting a minimal impact of sex on group affiliation in this study. Lastly, a longitudinal decline in FEV_1 is only one parameter that can evaluate progression and is less sensitive to capturing changes in small airway loss than other clinical measures, like parametric response mapping. However, it is worth noting that elements of our signature may reflect potential markers of small airway damage, as our signature exhibits enrichment of several processes that contribute to small airway damage, such as the response to wounding and ECM organization[133]. Nonetheless, future studies are needed to explore the relevance of the identified signature in assessing other outcomes.

In summary, data-driven modeling approaches identified early cross-tissue compartment proteomic signatures and provided insight into potential mechanisms associated with accelerated disease progression in COPD. This work highlights the ability of quantitative, systems-focused analytical techniques to accomplish both these goals. Data-driven modeling approaches could be applied to integrate spatiotemporal data in clinical samples from other diseases with a progressive or heterogeneous population.

2.6 Methods

2.6.1 Human participants

SPIROMICS (ClinicalTrials.gov Identifier: NCT01969344) is an ongoing multicenter, prospective observational study designed to identify new COPD subgroups and intermediate

biomarkers of disease progression[68]. Briefly, we enrolled participants aged 40-80 years at entry with a history of cigarette smoking (≥20 pack-years), either with COPD by the fixed ratio definition (post-bronchodilator $FEV_1/FVC < 0.7$), or without COPD; as controls, we recruited healthy individuals without smoking history. SPIROMICS participants (n=2,974) underwent a baseline examination (V1) followed by yearly visits for up to three years and a final follow-up visit (V5) approximately 5-8 years after V1. The first participant entered on November 10, 2010, and we censored all data on July 31, 2021. The study was conducted according to the principles of the Declaration of Helsinki. The human study protocol was approved by the institutional review board of all participating centers and methods were carried out in accordance with the relevant guidelines and regulations (Columbia University, New York, NY, United States; Johns Hopkins University, Baltimore, MD, United States; National Jewish Health, Denver, CO, United States; Temple University, Philadelphia, PA, United States; University of Alabama at Birmingham, Birmingham, AL, United States; University of California Los Angeles, Los Angeles, CA, United States; University of California San Francisco, San Francisco, CA, United States; University of Illinois at Chicago, Chicago, IL, United States; University of Iowa, Iowa City, IA, United States; University of Michigan, Ann Arbor, MI, United States; University of North Carolina at Chapel Hill, Chapel Hill, NC, United States; University of Utah, Salt Lake City, UT, United States; Wake Forest University, Winston Salem, NC, United States). All participants were aware of the study's intent and provided written informed consent before any procedures.

Some SPIROMICS participants (n=215) from all groups except those with severe (GOLD 4) COPD participated in a bronchoscopic sub-study[106,107], which included BAL of the right middle lobe and lingula. Participants (n=149) with a history of smoking who had available

blood and BAL samples were considered for inclusion in the initial EN-analysis (**Supplemental Figure A.1**). This analysis was restricted to participants (n=85) of that sub-study who had a history of smoking, available spirometry that did not improve at V5, and full biospecimens, which were plasma samples at V1 and BAL cytokine analysis; their baseline characteristics are shown in **Table 2.1**. They comprised two study groups: COPD cases (n=45) and a reference group (n=40) of TEPPS who had no airflow obstruction at both V1 and V5 (**Supplemental Table A.1**). The demographics of our study participants (n = 85) did not differ significantly from the entire SPIROMICS bronchoscopy cohort (**Supplemental Table A.2**).

Based on the magnitude of annual change in FEV_1, we dichotomized the COPD cases into greater decliners ($<30^{th}$ percentile; n=14) versus lesser decliners ($\geq 30^{th}$ percentile; n=31). FEV_1 decline was calculated using the two-point slope equation: [*V5 FEV_1 - V1 FEV_1*]/*time*, where *time* is the duration, in years, from V1 to V5 for each participant. Time calculations assumed a fixed-length year equal to 365.2425 days.

2.6.2 Sample preparation & datasets

Blood dataset: Fresh plasma samples collected at V1 were frozen in either an EDTA collection tube or a P100 tube with K2EDTA[134,135]. SOMAmer© (slow off-rate modified aptamer) technology[136] (SomaLogic, Boulder, CO) was used to measure 1305 proteins from participants in the SPIROMICS bronchoscopy sub-study.

BAL dataset: We measured the concentration of 48 proteins including cytokines, chemokines, and growth factors (HCYTA-60K-PX48, Milliplex, EMD Millipore Corporation) in BAL aliquots from a subset of participants in the SPIROMICS bronchoscopy sub-study (n=184) using Luminex FlexMAP 3D (Luminex Corporation, Austin, TX) technology. Any results above the upper limit of detection were set to the maximum detectable concentration of that analyte.

We set samples below the lower limit of detection to be half the lowest minimum detectable concentration across the standard curves of all analytes. We removed the 23 proteins in which ≥50% of measurements were below the lower limit of detection across all samples, yielding 25 analyzable BAL proteins. Before analysis, we normalized all BAL protein concentrations to total BAL protein concentration as quantified by a Pierce BCA Protein Assay Kit (Pierce Protein Biology, Rockford, IL).

Multi-compartment dataset: We removed eight proteins that were associated with sex in standard two-tailed, two-sample t-test after correction for multiple comparisons using Benjamini-Hochberg. The final dataset consisted of 1322 proteins (1297 blood and 25 BAL). All analytes were log-transformed for normality before analysis.

2.6.3 Derivation of data-driven progression signature(s)

Relative fold-changes in the expression levels of individual proteins from the blood and BAL were calculated by dividing the average concentration of each protein in COPD greater decliners by the average concentration in lesser decliners.

Based on proteomic measurements from the COPD participants, we generated optimal progression signatures using EN in tandem with PLSDA for feature selection in the: (a) combined blood and BAL and (b) blood-only datasets. First, the data were randomly sampled without replacement to generate 2000 subsets. To correct for effects of class size imbalances during regularization, we completed resampling at the size of the smallest class. We then performed EN regularization on each of the 2000 subsets. Once regularization was complete, the proteins were subsequently reordered based on their selection frequency throughout the EN iterations and fed in a step-forward manner into the PLSDA algorithm (starting with the protein with the highest selection frequency).

Model performance was evaluated at each step using *k*-fold cross-validation ($k = 6$). The model with the lowest resultant cross-validated error was selected as the optimal classification signature. Alternative minimal signatures were identified as signatures with <15 features with cross-validated accuracies >80%. ROC curves were generated based on the classification ability of each PLSDA model. All models were orthogonalized to improve interpretability.

2.6.4 Comparison of progression signature performance parameters

Random variants: To explore the meaningfulness of the optimized signature, we compared its cross-validated performance to that of 2000 random variants. Variant signatures were generated by randomly selecting 52 features from the original dataset. The 6-fold

cross-validated accuracy was calculated for each random signature. Performance across all variants was compared to the identified signature using a two-tailed, two-sample t-test.

Cross-validated accuracies: For quantitative comparisons of cross-validation accuracy across multiple models of interest, we split the data into six groups, iteratively excluded random subsets of 6-7 samples during model calibration, and later used them to test model predictions. The percentage of excluded samples correctly classified in each of the six iterations was used to statistically compare alternative models to the 52-protein signature. We determined statistical significance using a standard one-way ANOVA.

ROC curves: To explore the diagnostic ability of binary classifiers, ROC curves were generated from PLSDA models and resultant AUCs were statistically compared using the method outlined by Hanley and McNeil to account for correlation between curves generated from the same cohort[110]. Standard errors were calculated using the Wilcoxon statistic. All reported sensitivities and specificities are generated based on PLSDA model performance, except for the leptin/adiponectin signature, which had metrics stated in the original text[103].

Signature enrichment in alternative definitions of progression: PCA was applied to the 52 proteins identified in the optimal multi-compartment signature (**Figure 2.2**). Participants (n=45) were labeled using the alternative progression definitions proposed by Anderson et al.[109]: rapid decliners (>100 mL/year), decliners (20 – 100 mL/year), stable/improvers (< 20 mL/year). All data were mean-centered and variance-scaled prior to analysis. One-way ANOVA with Holm-Šídák's post hoc test compared participant scores on the first principal component across the groups. Significance was defined as a p-value < 0.05 for all analyses.

2.6.5 Bioinformatic analysis

Clustering: Hierarchical clustering of the 52-feature signature based on blood and BAL proteins was generated with supervised average linkage clustering using Spearman's correlation coefficient as the distance metric. Samples were colored by progression status.

Metascape analysis: Metascape[137] [https://metascape.org] was used to identify biological processes that were significant and differentially enriched between greater decliners and lesser decliners based on the identified 52-feature signature. PLSDA loadings on LV1 were used to dichotomize proteins between cohorts, such that proteins with positive or negative loadings were increased in lesser decliners or greater decliners, respectively.

Complement profiles: PCA was applied to a subset of 22 complement proteins (C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B, Factor D, Properdin) measured in the original plasma SOMAscan dataset (outlined above) from greater decliners (n=14), lesser decliners (n=31), and a TEPPS reference group (n=40). Participants were identified as outliers and removed from model if they had a Hotelling's Reduced T^2 statistic value > 2, determined via PCA (n=2). All data were mean-centered and variance scaled prior to analysis. One-way ANOVA with Tukey's post hoc test compared

participant scores across the first principal component. Significance was defined as $p < 0.05$.

2.6.6 Software summary

Volcano plots and hierarchical clustering were completed using MATLAB (v2017b, MathWorks, Natick, MA). Elastic net was implemented using Glmnet package in MATLAB[138]. We generated PCA and PLSDA models and ROC curves using the PLS toolbox available in MATLAB (v8.2.1, Eigenvector, Mason, WA). All statistics were performed using Prism version 9 (GraphPad Software, San Diego, CA).

Chapter 3 Validation of Systemic Complement Signatures in COPD Progression

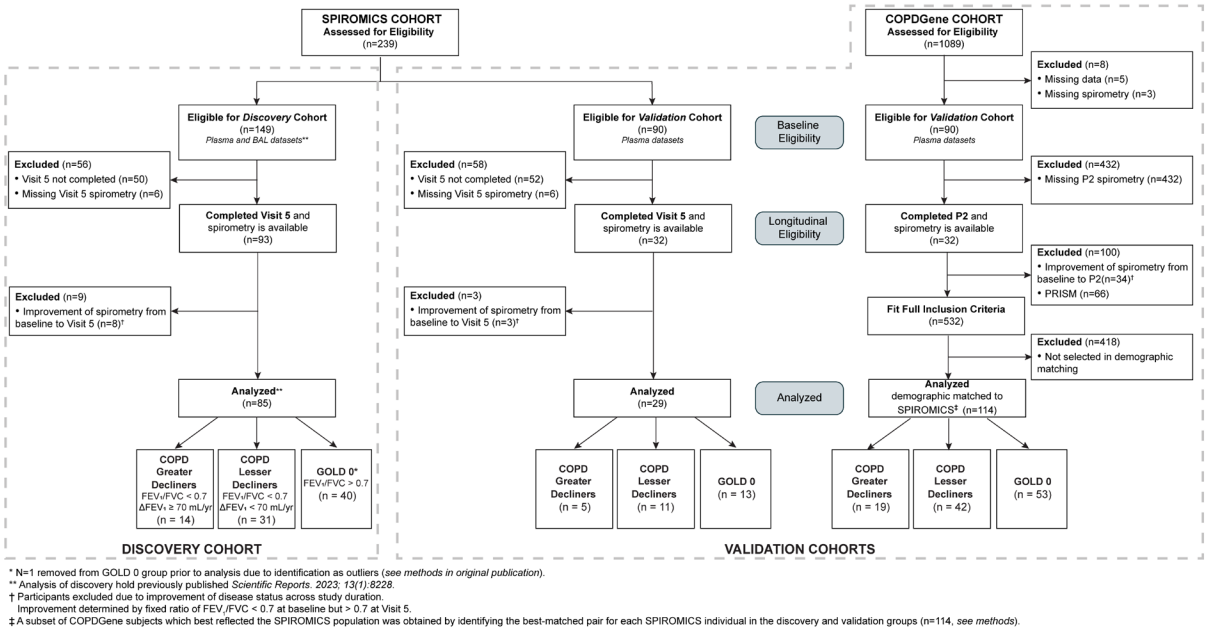
3.1 Introduction

The rates of lung function decline experienced by individuals with chronic obstructive pulmonary disease (COPD) are highly variable[99]. Such heterogeneity yields a subgroup of rapidly progressing individuals who endure persistent losses of pulmonary function and poorer prognostic outcomes, including increased risk of hospitalization and mortality[98]. A better understanding of the mechanisms underlying progression is needed to inform new treatment strategies for arresting accelerated lung function decline.

Using samples collected at enrollment from the SPIROMICS cohort, we previously identified a multivariate signature of 52 proteins that predicted individuals at risk for accelerated FEV_1 decline (≥70 mL/year) with high cross-validation accuracy[139]. Differences in spirometric progression rates were uniquely associated with baseline alterations in the expression of blood-derived complement proteins ("complement profiles"). Further refinement of the multivariate model identified two smaller signatures with 5- and 10-blood proteins that showed promise as potential prognostic biomarkers, however, a lack of external validation limited insights into their clinical utility. Here, we used independent plasma samples from two prospective cohorts to examine the reproducibility of our previously identified biomarker signatures and baseline complement expression profiles in predicting longitudinal FEV_1 decline.

3.2 Results

3.2.1 Characteristics of validation cohorts



* N=1 removed from GOLD 0 group prior to analysis due to identification as outliers (*see methods in original publication*).
** Analysis of discovery hold previously published *Scientific Reports. 2023; 13(1):8228.*
† Participants excluded due to improvement of disease status across study duration.
Improvement determined by fixed ratio of $FEV_1/FVC < 0.7$ at baseline but > 0.7 at Visit 5.
‡ A subset of COPDGene subjects which best reflected the SPIROMICS population was obtained by identifying the best-matched pair for each SPIROMICS individual in the discovery and validation groups (n=114, *see methods*).

Figure 3.1. Schematic illustrating the participants in the discovery and validation cohorts.

We analyzed a subset of participants from the SPIROMICS and COPDGene cohorts who had paired baseline and 5-year follow-up spirometry, proteomic measurements from plasma samples, and a history of cigarette smoking. Eligible participants were classified into three groups based on their annualized rate of FEV_1 decline (ΔFEV_1) and presence of airflow obstruction: COPD greater decliners ($\Delta FEV_1 \geq 70$ mL/year and $FEV_1/FVC < 0.7$), COPD lesser decliners ($\Delta FEV_1 < 70$ mL/year and $FEV_1/FVC < 0.7$), and ever-smoking controls without airflow obstruction (GOLD 0; $FEV_1/FVC > 0.7$). This classification scheme is identical to our discovery cohort (**Chapter 2**) and all validation samples are independent of those analyzed in our previous work[139]. The final SPIROMICS and COPDGene validation groups consisted of 29 (greater decliners, n=5; lesser decliners, n=11; GOLD 0, n=13) and 114 individuals (greater decliners, n=19; lesser decliners, n=42; GOLD 0, n=53), respectively (**Figure 3.1**). To ensure concordance of population demographics across studies, the COPDGene validation cohort reflects a subset of participants matched to the SPIROMICS participants (discovery and

validation; n=114) based on sex and specified ranges of age, FEV_1/FVC, and FEV_1. Final validation groups were well matched for all demographics (**Table 3.1**) except ICS use; however, we previously identified no impact of ICS use on our analyses[139]. The time between the baseline and follow-up visits in the SPIROMICS and COPDGene analysis groups was also significantly different (6.2 ± 0.9 years vs. 5.4 ± 0.7 years). This discrepancy is not expected to impact our models, as FEV_1 declines were normalized to the exact follow-up time between visits for each patient.

Table 3.1. Baseline participant demographics

	SPIROMICS (N=114)	COPDGene (N=114)	P-Value*
Age	61 (± 8.6)	62 (± 8.7)	0.79
Currently Smoking	74 (65%)	78 (68%)	0.67
BMI	28 (± 5.2)	29 (± 5.2)	0.61
Sex (Male)	51 (45%)	51 (45%)	1
Race (White / Other)	**85/29 (75%)**	**102/12 (89%)**	**0.006**
ICS use (yes)	**27 (24%)**	**4 (4%)**	**<0.001**
FEV_1 (% predicted)	86 (± 21)	83 (± 22)	0.30
FEV_1/FVC	0.67 (± 0.13)	0.67(± 0.13)	0.86
FEV_1 (L)	2.5 (± 0.78)	2.5 (± 0.89)	0.85
Time from baseline to follow-up (yrs.)	**6.2 (± 0.94)**	**5.4 (± 0.71)**	**<0.001**
ΔFEV_1 (mL/yr.)	-40 (± 47)	-39 (± 42)	0.87

*Two-sample, two-tailed t-test or Fisher's exact test were used to determine significant differences. Bold rows denote a statically significant difference between validation groups ($p<0.05$).

3.2.2 Previously identified blood biomarkers signatures do not predict COPD progression in an independent validation group

We first determined whether our previously identified prognostic biomarker signatures[139] accurately distinguish individuals at risk for accelerated progression (≥ 70 mL/yr.) in independent samples from SPIROMICS and COPDGene. In the SPIROMICS validation group, the 5-protein signature (iC3b, Cadherin-2, Leptin, HBEGF, and TDGF1), measured by a partial-least squares discriminant model (PLSDA), was not reproducible, resulting in an AUC of 0.6 which was substantially lower than the cross-validated AUC of 0.86 observed in the discovery analysis (**Table 3.2, Figure 3.2a**). The 10-protein signature (iC3b, Cadherin-2, Leptin, HBEGF, TDGF1, CA10, IFN-γ, FCGR2B, Apolipoprotein L1) also performed poorly, with an AUC below 0.5. Outcomes were similar in the COPDGene validation group, with the 5- and 10-protein signatures performing only modestly better than random assignment (**Figure 3.2b**; 5-protein signature, AUC = 0.56; 10-protein signatures, AUC = 0.58). For thoroughness, we assessed our optimal signature's performance (from **Chapter 2**). As one of the 52 proteins in this model was from BAL, we tested a 51-protein blood-only model. Unfortunately, the larger signature did not improve performance, achieving ~56% classification accuracy in both groups.

Table 3.2. Summary of biomarker signature(s) performance differentiating greater and lesser decliners in independent samples from the SPIROMICS and COPDGene validation groups

Validation Group	Protein Signature	AUC	Accuracy	Sensitivity	Specificity
SPIROMICS	5-protein*	0.60	50%	0.20	0.82
SPIROMICS	10-protein†	0.14	36%	0.00	0.73
COPDGene	5-protein	0.56	52%	0.10	0.95
COPDGene	10-protein	0.58	56%	0.21	0.90

* Proteins in 5-feature signature: iC3b, Cadherin-2, Leptin, HBEGF, and TDGF1

† Proteins in 10-feature signature: iC3b, Cadherin-2, Leptin, HBEGF, TDGF1, CA10, IFN-γ, FCGR2B, Apolipoprotein L1

3.2.3 Previously identified complement profiles are significantly associated with COPD progression in an independent validation group

In contrast to the previously identified blood biomarker signature, the previously established complement profiles, generated through a principal component analysis (PCA) model encompassing 22 baseline complement proteins[139], exhibited notable differentiation between COPD individuals with greater and lesser rates of decline within the SPIROMICS validation group (p=0.045; **Figure 3.2c**; **Figure 3.2e**). The differential expression associated with COPD greater decliners remained after introducing a GOLD 0 reference group to the PCA model (p=0.039). In the COPDGene validation group, systemic complement profiles were also successfully validated, with COPD greater decliners exhibiting significantly different profiles from COPD lesser decliners (p=0.005; **Figure 3.2d**, **Figure 3.2f**) and a GOLD 0 reference group (p=0.029). All validation metrics were obtained by evaluating the performance of our previously identified models on these two independent external datasets[139].

There is currently no formal definition of "rapid progression" in COPD, so we determined whether the identified complement profiles maintained significance when an alternative classification of "progression" was used. Recently, Anderson et al. proposed a threshold-based definition, classifying progression into three groups based on annualized FEV_1 declines: rapid decliners (>100 mL/year), decliners (20-100 mL/year), and stable/ improvers (< 20 mL/year)[109]. Alternative classification of individuals from the COPDGene validation group demonstrated significant profile enrichment in rapid decliners thus defined (**Figure 3.3**), suggesting our model reliably extends to this more rigorous definition. This secondary analysis could not be applied to the SPIROMICS validation group due to its minimal sample size.

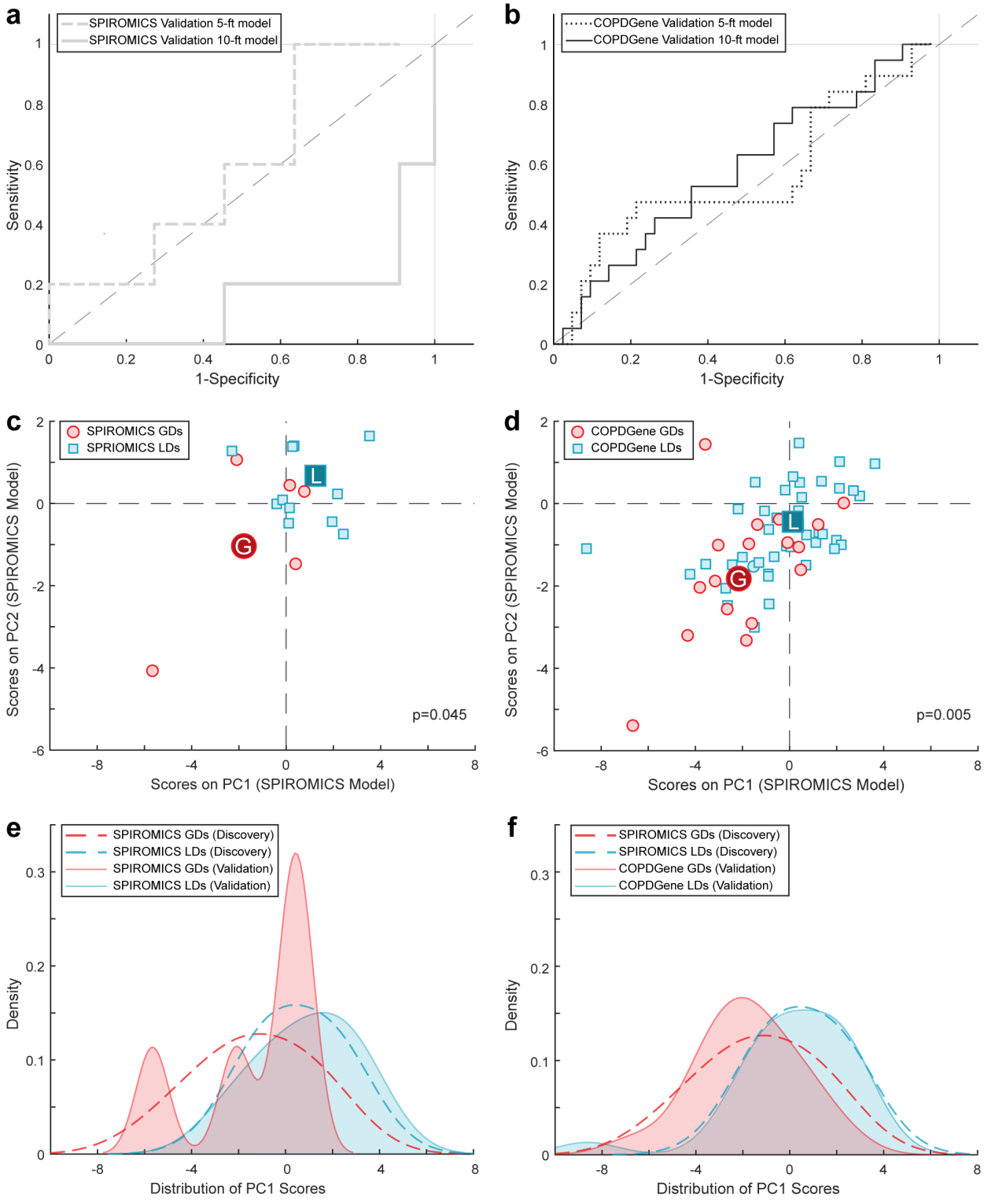


Figure 3.2. Systemic complement profiles validate in independent samples from two independent analysis groups. (a-b) ROC curves from COPD patients in the (a) SPIROMICS and (b) COPDGene validations groups fit to 5-protein (dashed lines) and 10-protein (solid lines) PLSDA models of FEV_1 decline identified in the discovery group[139]. Validation performed poorly with AUCs ≈ 0.5 for most models. (c-d) Measurements from 22 complement proteins (C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B,

Factor D, Properdin) measured in plasma samples from (c) greater decliners (GD; n=5; red circles) and lesser decliners (LD; n=11; blue squares) from the SPIROMICS validation group and (d) greater decliners (GD; n=19; red circles) and lesser decliners (LD; n=31; blue squares) from the COPDGene validation group. Complement measurements from each validation group were fit to the PCA models previously generated on the SPIROMICS discovery analysis[139]. All p-values are reported from a permutation test (n = 2000 permutations) between groups' mean scores across PC1 and PC2. Group means from the validation data are denoted by the labels on the plots. (e-f) Kernel density plots visualizing the distribution of the scores across the first principal component (PC1) from greater and lesser decliners in the (e) SPIROMICS and (f) COPDGene validation groups, as compared to scores from the original PCA model generated on the discovery group (dashed line)[139].

3.3 Discussion

Our results suggest that the early alterations in blood complement levels we previously identified reliably associate with accelerated spirometric progression in two independent validation groups, while the minimum biomarker signatures do not. These findings suggest that COPD progression results from disparate pathway alterations and confirms previous studies highlighting the value of multivariate models in exploring COPD outcomes[58].

To our knowledge, this study is the first to provide reproducible, longitudinal evidence that early alterations in the complement pathway measured at baseline are associated with a risk for accelerated FEV_1 decline ~5-6 years later. By providing evidence that complement alterations precede accelerated FEV_1 decline, we extend previously reported associations between the levels of blood-derived complement proteins and cross-sectional COPD outcomes, such as FEV_1% predicted[119,120] and emphysema severity[121]. Although our findings reflect systemic profiles, alterations in the levels of various complement components have been reported in COPD lungs (including C3[140], C4b[141], C5a[117], and MASP-2[142]), suggesting dysregulation might extend into the pulmonary compartment, which we could not test. How such aberrations in the blood or tissue might contribute to airflow limitation is unknown. However, murine studies have implicated a potential role for C1q[143,144] and C3[122] in cigarette-smoke-associated emphysema development through modifying APC-directed Th17 inflammation.

Although pathway-related validation across validation analyses was successful, small

prognostic signatures replicated poorly. Poor biomarker replication has been reported in other proteomics investigations of large COPD cohorts[85] and may suggest that no single or even combination of proteins can predict progression in a heterogenous disease such as COPD. Challenges in identifying succinct subsets of biomarkers may further suggest that alterations in several pathways or pathway constituents drive a singular phenotypic outcome in patients with COPD. Together, our results highlight the potential importance of shifting focus from biomarker identification towards pathway dysregulation in future analyses of spirometric progression and other complex COPD outcomes.

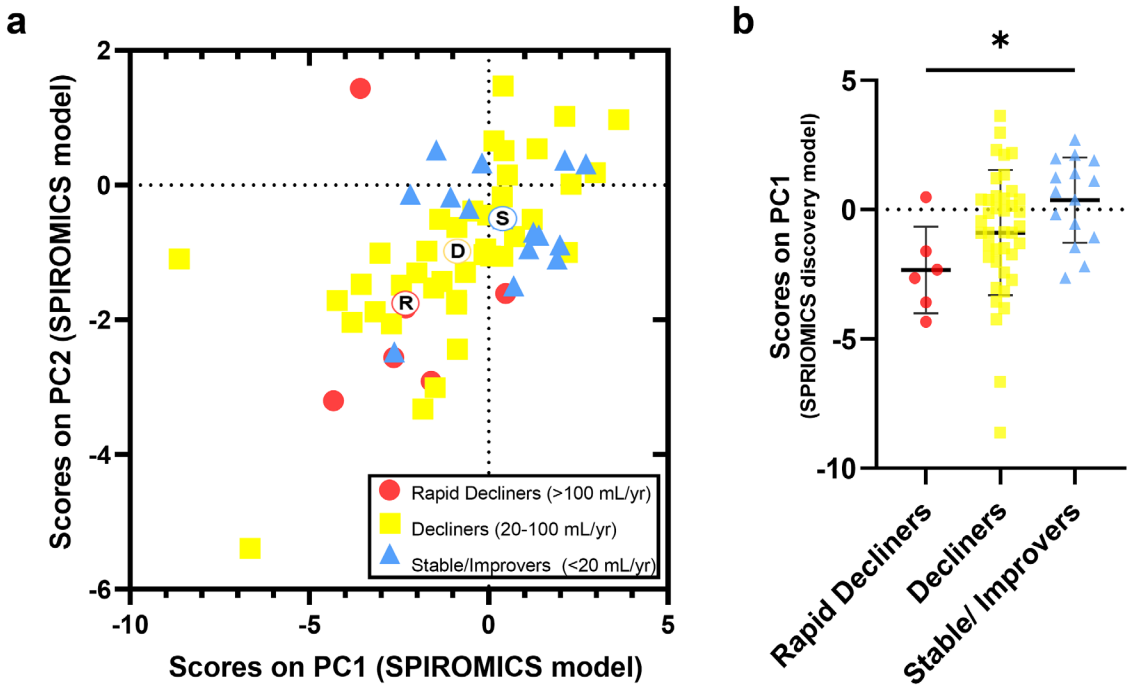


Figure 3.3. Complement profiles from a principal component model generated with data from the SPIROMICS discovery group validate in a demographic matched COPDGene validation group.
(a) Measurements from 22 complement proteins (C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B, Factor D, Properdin) measured in plasma samples from the COPDGene validation cohort were fit to the PCA models previously generated on the SPIROMICS discovery analysis. Validation participant scores (n=61) were alternatively labeled by the definitions proposed by Anderson et.al[109]: rapid decliners (n=6; red circles), decliners (n=40; blue squares), and stable/ improvers (n=15; yellow triangles) from the COPDGene validation group. The circular labels on plot represent the centroids for each class.
(b) Statistical comparison of COPDGene participant scores across PC1 (one-way ANOVA with Holm-Šídák's multiple comparisons test; *$p < 0.05$).

Our study has several limitations. Chief among these is our inability to assess the activity of complement components, as SomaLogic aptamers cannot reliably distinguish between complement cleavage products and their parent proteins. Comprehensive demographic matching also limited the size of the COPDGene validation group. However, without matching, key demographic characteristics, including all baseline lung function parameters, were significantly different across validation groups. Previous COPD studies with unmatched cohorts report poor replication of systemic plasma proteins across studies[85], emphasizing the importance of population matching when validating prognostic models.

In summary, our results suggest that the alterations in blood complement levels precede and are reliably associated with accelerated spirometric decline. These findings indicate that studies of COPD may benefit from targeted analyses of the complement pathway to help deconvolve its link to spirometric progression.

3.4 Methods

3.4.1 Human participants

Validation Cohorts: SPIROMICS (ClinicalTrials.gov Identifier: NCT01969344) is an ongoing multicenter, prospective observational study designed to identify new COPD subgroups and intermediate biomarkers of disease progression[68]. Briefly, SPIROMICS enrolled participants aged 40-80 years at entry with a history of cigarette smoking (≥20 pack-years), either with COPD by the fixed ratio definition (post-bronchodilator $FEV_1/FVC < 0.7$), or without COPD. SPIROMICS participants (n=2,974) underwent a baseline examination (V1) followed by yearly visits for up to three years and a final follow-up visit (V5) approximately 5-8 years after V1. This analysis was restricted to participants (n=114) of the bronchoscopy sub-study who had a history of smoking, available spirometry that did not improve at V5, and plasma samples at V1.

Individuals previously included in the original discovery analysis were excluded (n=85)[139]. Participants in the final validation groups comprised three study groups: COPD cases (n=16) and a reference group (n=13) of GOLD 0 who had no airflow obstruction at both V1 and V5 (**Figure 3.1**).

Based on the distribution of annual change in FEV_1 among individuals in our previous analysis, we dichotomized the COPD cases into greater decliners (≥ 70 mL/yr.; n=5) versus lesser decliners (< 70 mL/yr.; n=11). Accordingly, we retained this definition in this analysis, while recognizing that it may not be an optimal cut-point for the entire SPIROMICS cohort. We calculated FEV_1 decline using the two-point slope equation: *[V5 FEV_1 - V1 FEV_1]/time*, where time is the duration, in years, from V1 to V5 for each participant. Time calculations assumed a fixed-length year equal to 365.2425 days. Two-point slope equations were used to maximize the size of the final study population, as there was significant missingness in spirometry data from intermediate participant visits.

COPDGene (ClinicalTrials.gov Identifier: NCT00608764) is an ongoing multicenter, prospective observational study designed to identify genetic factors associated with COPD[145]. Briefly, COPDGene enrolled participants aged 45-80 years at entry with a history of cigarette smoking (≥10 pack-years), either with COPD by the fixed ratio definition (post-bronchodilator $FEV_1/FVC < 0.7$) or without COPD. Participants underwent a baseline phase 1examination (P1) and a follow-up phase 2 visit (P2) approximately 5 years later.

To ensure concordance of population demographics across studies, the final COPDGene validation grops used in this analysis reflects a subset of participants matched to the SPIROMICS participants from both the discovery and validation groups, based on sex and specified ranges of age, FEV_1/FVC, and FEV_1 (n=114). The exact ranges used for matching were

age (± 10 years), FEV_1/FVC (± 0.07), and FEV_1% predicted (± 18% for COPD, ± 25) for GOLD 0). When possible, participants were also matched on race and FEV_1 decline ± 25 mL/year, but those criteria were not required. Using the same classification scheme outlined for SPIROMICS, the matched COPDGene validation cohort (n=114) contained 19 COPD greater decliners, 42 COPD lesser decliners, and 53 GOLD 0 participants.

Institutional review boards at participating institutions approved the study, and participants provided written informed consent for both studies.

3.4.2 Plasma dataset

Fresh plasma samples collected at baseline were frozen. SOMAmer© (slow off-rate modified aptamer) technology[136] (SomaLogic, Boulder, CO) was used to measure 1305 proteins. Proteomic data were quantified at National Jewish Health.

3.4.3 Evaluation of data-driven models

Biomarker Signatures: We evaluated the ability of the 5-protein (iC3b, Cadherin-2, Leptin, HBEGF, and TDGF1) and 10-protein (iC3b, Cadherin-2, Leptin, HBEGF, TDGF1, CA10, IFN-γ, FCGR2B, Apolipoprotein L1) biomarker signatures to predict accelerated progression in the SPIROMICS and COPDGene validation groups through PLSDA model(s) identified in the discovery analysis (**Figure 2.5**)[139] on the validation datasets using the PLS toolbox in MATLAB (v8.2.1, Eigenvector, Mason, WA). Performance metrics, including ROC curves, AUC, sensitivity, and specificity, were generated based on the predictions from the PLSDA model.

Complement profiles: The 22 complement proteins (C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B, Factor D,

Properdin) measured in the validation SOMAscan datasets from SPIROMICS and COPDGene participants were fit to the complement PCA model(s) generated on the discovery participants (**Figure 2.4**), using PLS toolbox in MATLAB. Resultant scores were plotted, and profiles were compared using a permutation test (2000 permutations) to evaluate whether the mean scores across principal component (PC) 1 and PC2 were significantly different across groups. Significance was defined as $p < 0.05$. All data were mean-centered and variance-scaled prior to analysis.

Signature enrichment in alternative definitions of progression: The PC scores generated from evaluating COPDGene dataset on the discovery PCA model (n=61) were labeled using the alternative progression definitions proposed by Anderson et al.[109]: rapid decliners (>100 mL/year), decliners (20 – 100 mL/year), stable/improvers (< 20 mL/year). Significance, defined as a p-value < 0.05, was determined by a one-way ANOVA with Holm-Šídák's post hoc test to compare participant scores on the first principal component across the groups. Univariate statistics were performed using Prism version 9 (GraphPad Software, San Diego, CA).

Chapter 4 Multivariate Proteomic Signatures Reveal Age-Dependent Mechanisms Contributing to Progression in Smoking-Associated Emphysema

4.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a highly prevalent respiratory disorder characterized by progressive airflow limitation that is often linked to a history of chronic cigarette smoking. Historically, COPD has been considered a disease of the elderly[146]. However, recent studies highlight a shift in this perspective, revealing that symptoms indicative of COPD can manifest in smokers as early as 30-50 years of age and strongly correlate with lung function impairment later in life[17,112,147–151]. This younger demographic has traditionally been underrepresented in large, longitudinal COPD studies, which have tended to focus on older individuals (≥ 60 years old) with mild disease (GOLD 0-2)[112]. Additional studies focused on early pathogenesis of COPD in younger individuals could provide new insight into COPD pathogenesis and progression.

Recent work by Martinez et al. has introduced the term "early COPD" to categorize individuals under the age of 50 with 10 or more pack-years smoking history who present with airway abnormalities[112]. Given the early stages of obstruction expected in this population, new clinical biomarkers capable of measuring airway changes before significant spirometric decline are needed. One promising option is parametric response mapping (PRM). PRM is a CT voxel-based imaging biomarker that employs dynamic image registration and separate density thresholds for inspiratory and expiratory voxel measurements. PRM distinguishes regions of "normal" lung from "functional small airway disease" (PRM^{fSAD}) and "emphysema"

(PRM^{Emph})[12]. PRM measurements correlate with clinical and functional COPD parameters[13,14], as well as histological evidence of airway narrowing and loss of alveolar attachments[152]. Importantly, changes in PRM are measurable in the absence of spirometrically-defined airflow obstruction[15,16], highlighting its value for evaluating early airway and parenchymal abnormalities.

Combinations of proteins provide improved predictive performance compared to individual biomarkers in studies of stable COPD[58]. Accordingly, analyses of early COPD may also benefit from evaluating multi-factor networks, as they are more likely to capture the diverse biology driving disease progression[57,93]. Data-driven modeling is one approach that allows inference of these network-level relationships. By identifying signatures of co-varying biological factors, data-driven approaches are valuable tools for gaining novel insights into mechanisms of action based on protein interaction pathways rather than individual factors. Once validated, signatures may be used for diagnostic or prognostic purposes or to generate new hypotheses for future experimental work. We have previously used these approaches to successfully identify prognostic protein signatures associated with progression in patients with COPD[139] and idiopathic pulmonary fibrosis[89].

Here, to gain insights into age-dependent pathways contributing to COPD progression, we applied a data-driven modeling pipeline to proteins recovered from blood samples from participants of the COPDGene cohort[145]. Our results suggest that proteomic signatures can effectively detect individuals at increased risk of accelerated development of emphysema and small airway disease. They also provide insights into age-dependent mechanisms of emphysema progression that can be further investigated in newly enrolling early COPD cohorts.

4.2 Results

4.2.1 Participant characteristics

We analyzed a subset of participants in the COPDGene study ages > 65 or ≤ 55 years with a smoking history and mild (GOLD 0-2), who had paired baseline (P1) and 5-year (P2) parametric response mapping (PRM) measurements and baseline protein measurements from plasma samples (n=158). Eligible participants were dichotomized into older (> 65 years old) and younger groups (≤ 55 years old) based on their age at enrollment to generate subpopulations that reflect traditional COPD populations and early COPD populations, respectively. We then used patient-specific changes in PRM (ΔPRM) from P1 to P2 to define subclasses ("phenotypes") within each age group who experienced relatively fast or slow progression. Progression phenotypes were selected using percentiles, with fast and slow/stable progression defined as participants exhibiting ΔPRM > 75th percentile (%tile) and < 25th %tile, respectively (**Figure 4.1**). Using this classification scheme, we generated two unique analyses using ΔPRM^{Emph} and ΔPRM^{fSAD} to characterize the age-dependent progression of emphysema and small airways disease (SAD), respectively (**Table 4.1**, **Supplemental Table B.1**).

In the ΔPRM^{Emph} analysis (**Table 4.1**), participants in the older subpopulation were on average 21 years older than their younger counterparts and had significantly worse baseline lung function, as measured by spirometry ($p < 0.001$) and PRM^{Emph} ($p < 0.001$). However, the younger subpopulation was about 30% more likely to be current smokers. Additionally, older participants in the fast-progressing phenotype experienced larger declines in PRM^{Emph} than younger fast- progressors (4.4% ± 3.5 vs. 2.4% ± 3.0; $p = 0.007$). The opposite trend was observed in the slow/stable groups, with older individuals exhibiting a ΔPRM^{Emph} of -1.8% ± 1.6 compared to the -0.57% ± 1.6 observed in the younger participants ($p = 0.002$). Similar trends

were observed in the analysis group used for the ΔPRM^{fSAD} analysis, with older participants exhibiting worse baseline lung function and experiencing more severe progression than the younger participants, despite decreased current smoking rates (**Supplemental Table B.1**).

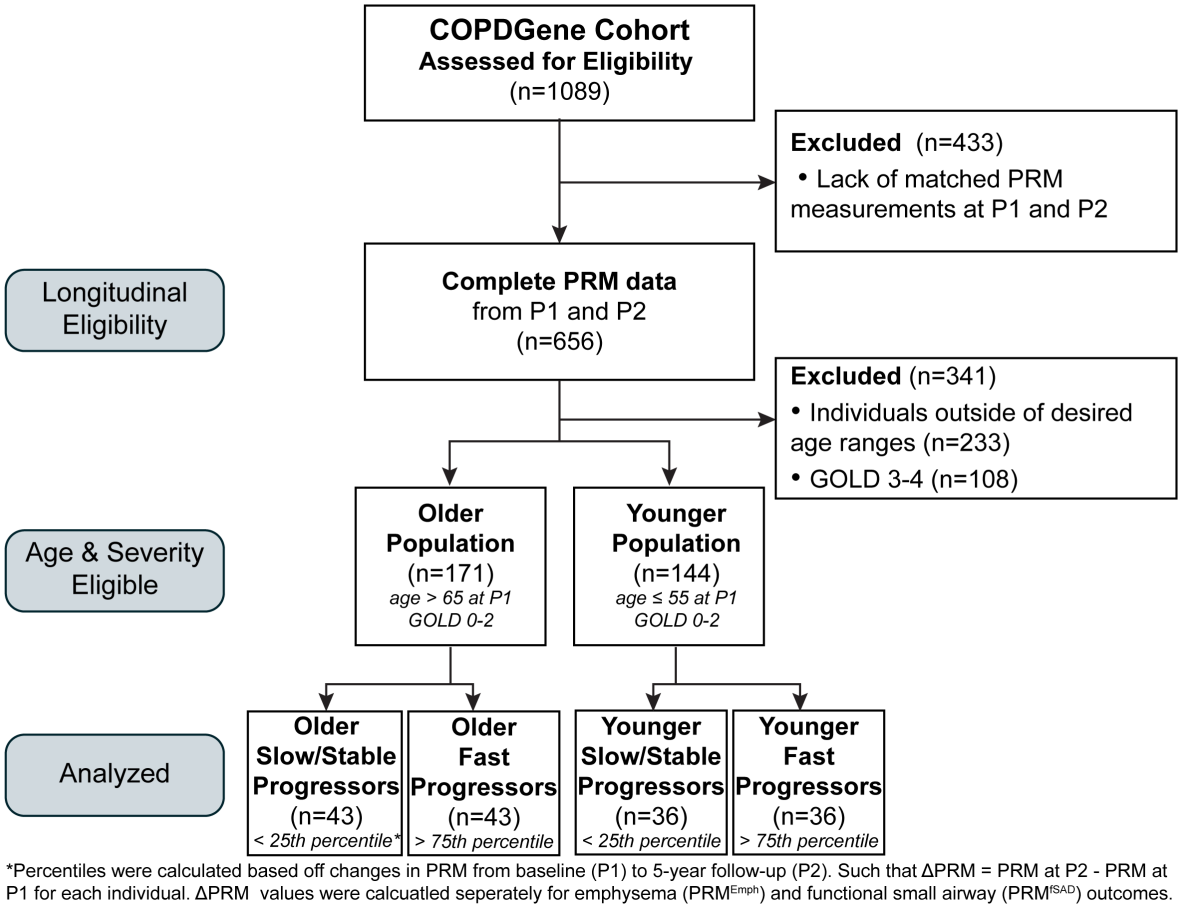


Figure 4.1. Schematic illustrating the inclusion criteria and patient breakdown of GOLD 0-2 participants from COPDGene included in analysis.

4.2.2 Data-driven modeling identifies baseline proteomics signatures capable of differentiating GOLD 0-2 participants based on age and magnitude of emphysema progression

We first used data-driven modeling approaches to determine whether systemic signatures could differentiate GOLD 0-2 ever-smokers based on their age and relative magnitude of emphysema or SAD progression. To do this, we applied EN regularization in tandem with partial-least squares discriminant analysis (PLSDA) to the n=1305 plasma proteins measured

Table 4.1 Baseline demographics for ΔPRM^{Emph} model

	Fast		Slow/Stable		Overall*	
	Older (N=43)	Younger (N=36)	Older (N=43)	Younger (N=36)	Older (N=86)	Younger (N=72)
Age	**72 (± 4)**	**51 (± 3)**	**70 (± 4)**	**49 (± 3)**	**71 (± 4)**	**50 (± 3)**
Currently Smoking	**14 (33%)**	**26 (72%)**	**4 (9%)**	**17 (47%)**	**18 (21%)**	**43 (60%)**
BMI	28 (± 5)	28 (± 6)	28 (± 5)	29 (± 8)	28 (± 5)	29 (± 7)
Sex (Male)	28 (65%)	20 (56%)	21 (49%)	14 (39%)	49 (57%)	30 (47%)
Race (White / Other)	41/2 (95%)	30/6 (83%)	**43/0 (100%)**	**26/10 (72%)**	**84/2 (98%)**	**56/16 (79%)**
ICS use (yes)	2 (5%)	1 (3%)	1 (2%)	3 (8%)	3 (3%)	4 (6%)
FEV_1 (% predicted)	**77 (±20)**	**87 (± 15)**	**86 (±18)**	**96 (± 11)**	**82 (±19)**	**91 (± 14)**
FEV_1/FVC	**0.6 (± 0.1)**	**0.7 (± 0.1)**	**0.7 (± 0.1)**	**0.8 (± 0.1)**	**0.6 (± 0.1)**	**0.7 (± 0.1)**
FEV_1 (L)	**2.1 (± 0.6)**	**3.0 (± 0.8)**	**2.4 (± 0.7)**	**3.1 (± 0.7)**	**2.2 (± 0.6)**	**3.0 (± 0.8)**
PRM^{Emph} (%)	**6.1 (± 7)**	**1.8 (± 4)**	**4.9 (± 6.1)**	**1.1 (± 4)**	**5.5 (± 6)**	**1.5 (± 4)**
ΔPRM^{Emph} (%)	**4.4 (± 3)**	**2.4 (± 3)**	-1.8 (± 2)	**-0.57 (± 2)**	1.3 (± 4)	0.92 (± 3)

*Two-sample, two-tailed t-test or Fisher's exact test were used to determine significant differences. Individual statistical tests were evaluated for younger vs older comparisons in fast, slow/stable, and overall columns. Bold values denote a statically significant difference between younger and older groups ($p<0.05$).

using SOMAscan technology. First, EN regularization was applied iteratively to 2000 subsets of proteins were ranked (most to least frequent) and fed stepwise into the PLSDA algorithm. We evaluated PLSDA model performance at each step using 10-fold cross-validation (CV). To control the size of the final signature, we selected the smallest model that performed with a CV accuracy within 0.03 of the optimal model.

Using this feature selection pipeline, we identified two unique signatures describing age-related progression associated with emphysema (ΔPRM^{Emph}) and small airway disease (ΔPRM^{fSAD}). The ΔPRM^{Emph} signatures contained 20 plasma proteins, performed with 83.4% calibration accuracy and 81.6% cross-validation accuracy, and successfully distinguished between the influence of age and magnitude of emphysema progression across latent variable 1 (LV1) and LV2 axes, respectively (**Figure 4.2**). Supplemental investigations suggested current smoking status and FEV_1% predicted had no influence on the final model, despite their significantly different distributions across groups (**Supplemental Figure B.1**). Receiver operating characteristic (ROC) curves generated for the cross-validated PLSDA model had associated AUCs above 0.8 for all phenotypes (**Supplemental Figure B.2**; older slow/stable: AUC = 0.9, older fast; AUC = 0.9; younger slow/stable, AUC = 0.8; younger fast, AUC = 0.86).

The identified ΔPRM^{fSAD} signature consisted of 35 proteins and performed with 81.7% calibration accuracy and 79.4% cross-validation accuracy. However, contrary to the emphysema signature, the ΔPRM^{fSAD} signature lacked the ability to clearly differentiate older from younger individuals based on their relative magnitude of SAD progression (**Supplemental Figure B.3**). This finding is similarly reflected by the ROC curves, with AUCs from the older participants averaging 0.76, compared to the average AUC of 0.92 observed in the younger subclasses (**Supplemental Figure B.4**). Given the success of the ΔPRM^{Emph} model in differentiating

participants based on both the parameters of age and progression, for the remainder of this analysis, we focused on the emphysema model.

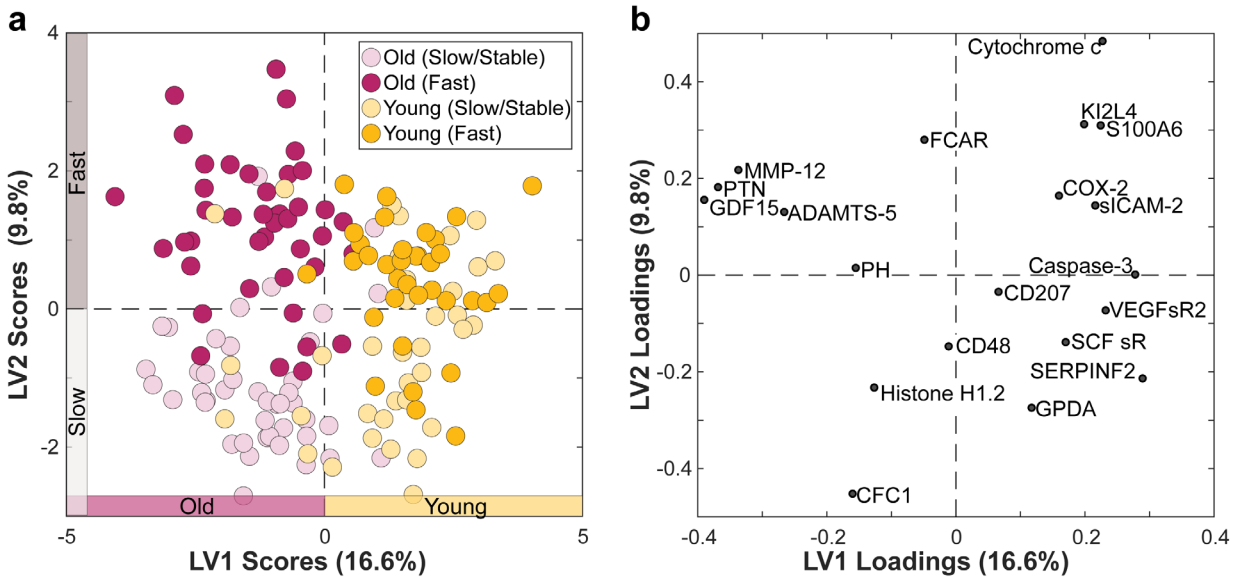


Figure 4.2. Elastic net and PLSDA identify 20 plasma proteins measured at a baseline visit that can uniquely discriminate ever-smokers based on their age and relative magnitude of emphysema progression after 5-years. (a) PLSDA scores plot highlighting 20-protein signature selected by bootstrap elastic net. Signature differentiates between younger (yellow) and older (pink) ever-smokers with relative fast (darker) or slow (lighter) progression with 83.4% calibration and 81.6% cross-validation (CV) accuracy. (b) The two latent variable (LV) model captures 26.4% of the total variance in the dataset. LV1 separates ever-smokers based on age (with negatively loaded proteins being comparatively increased in older ever-smokers and positively loaded proteins being comparatively reduced), while LV2 separates individuals based on relative magnitudes of progression.

4.2.3 Proteomic signatures suggest age-dependent pathways associated with emphysema progression in COPD

Having generated a high-performing age-dependent emphysema signature, we sought to understand the biological implication of its components. Unsupervised hierarchical clustering differentiated younger from older ever-smokers with 88% sensitivity and 84.8% specificity (**Figure 4.3**). A subsequent Metascape analysis found six significantly enriched ontology clusters (**Figure 4.4**). Interestingly, the dendrogram associated with the hierarchal cluster suggests that proteins increased in the younger population most often shared pathway enrichments with the fast-progressing phenotype. In contrast, enrichments in the older population and slow/stable

phenotype were more similar. Four of the six ontology clusters were associated with immune responses (GO:0032103 and R-HSA-1280218) or extracellular matrix organization (R-HSA-1474224 and M5885).

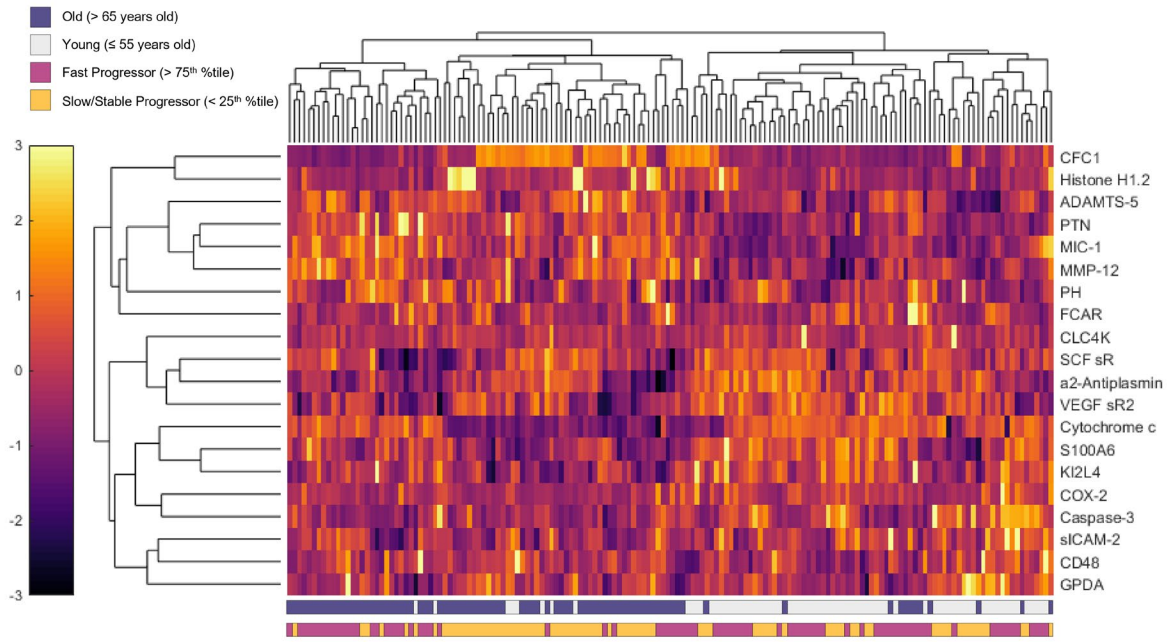


Figure 4.3. Hierarchical clustering of the 20-feature signature highlights distinct age-dependent clustering of ever-smokers.
Unsupervised clustering allowed for identification of younger (grey) ever-smokers from older (purple) ever-smokers with 88.8% sensitivity and 84.8% specificity. Smaller clusters exist within each age population that begin to cluster individuals based on their relative magnitude of progression.

By further examining the specific proteins increased in the fast- progressing phenotypes, we observed that younger, fast progressors exhibit increased levels of proteins with inflammatory and apoptotic functions, including killer cell immunoglobulin-like receptor 2DL3 (KI2L4), intercellular adhesion molecule 2 (sICAM-2), prostaglandin-endoperoxide synthase 2 (COX-2), cytochrome c, and caspase-3. Conversely, the older fast progressing phenotype was associated with increases in matrix remodeling proteins, such as matrix metallopeptidase 12 (MMP-12), pleiotrophin (PTN), growth differentiation factor 15 (GDF-15), and ADAM metallopeptidase with thrombospondin type 1 motif 5 (ADAMTS-5). Collectively, these findings

suggest that the biological processes contributing to accelerated progression in smoking-associated emphysema may be age-dependent.

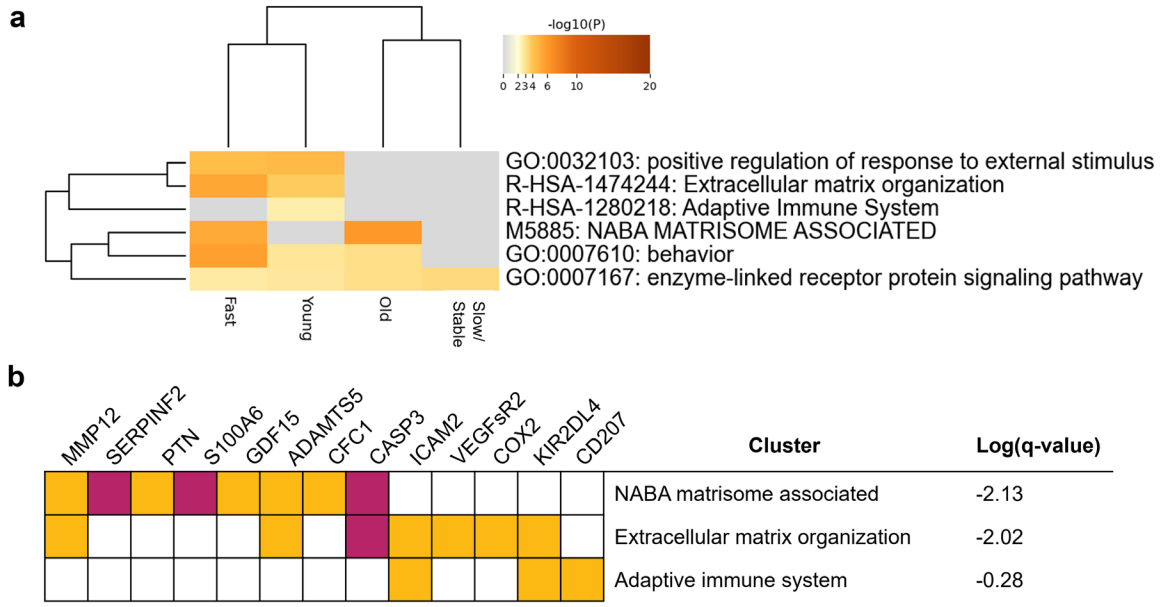


Figure 4.4. Pathway analysis identifies unique biological process associated with age-dependent emphysema progression.
(a) Significantly enriched ontology clusters by Metascape analysis. (b) Pathways encompassed in the clusters of interest are included table. Colored squares indicate protein involvement in a particular pathway, colorations of magenta or yellow represent a relative elevation of the protein concentration in older or younger populations, respectively.

4.3 Discussion

This study used plasma protein measurements from participants in the COPDGene cohort to generate insights into early signaling pathways associated with the age-dependent progression of emphysema and SAD. Using a data-driven approach, we identified two unique signatures of baseline plasma proteins that differentiated the age-dependent risk of accelerated ΔPRM^{Emph} and ΔPRM^{fSAD} in GOLD 0-2 ever-smokers with 81% and 79% cross-validation accuracy, respectively. The ΔPRM^{Emph} signature successfully differentiated participants based on age and magnitude of emphysema progression. In contrast, the ΔPRM^{fSAD} signature captured only age-

dependent differences. Investigation of the ΔPRM^{Emph} signature suggested that age dictates disparate contributions of apoptosis, inflammation, and matrix remodeling pathways to emphysema progression. This work provides novel evidence into the age-dependent mechanisms driving smoking-associated emphysema development and emphasizes the importance of studying COPD in younger populations ("early COPD").

To our knowledge, this is the first longitudinal study of ever-smokers to explore age-dependent mechanisms associated with the progression of emphysema. Our tandem EN and PLSDA approach identified concise proteomic signatures from thousands of plasma proteins. This framework accurately differentiated the 5-year progression of emphysema experienced by GOLD 0-2 ever-smokers using proteins measured early after participant enrollment. The identified emphysema signature comprised 20 plasma proteins enriched in inflammation, apoptotic, and matrix remodeling functions. Pathway enrichment occurred in an age-dependent manner, with proteins increased in younger (≤ 55 years old) and oldere (> 65 years old) fast progressors ($\Delta PRM^{Emph} > 75^{th}$ %tile) enriched for inflammatory/ apoptotic and matrix remodeling functions, respectively.

Younger participants who experienced higher ΔPRM^{Emph} exhibited increased levels of proteins with inflammatory (KI2L4, sICAM-2, CD207, COX-2) and apoptotic functions (cytochrome C, and caspase-3). To our knowledge, except for COX-2[153], none of these inflammatory proteins have been associated with COPD. However, cytochrome C and caspase-3, critical proteins in the apoptotic pathway, are reportedly elevated in COPD patients[154], and in murine models, emphysema development can be induced or prevented by administering caspases or their inhibitors, respectively[155,156]. These aforementioned findings align with multiple studies that report close associations between the increased rates of alveolar epithelial and endothelial

apoptosis and emphysema in COPD lungs[157–159]. The causes of apoptosis in COPD are incompletely understood but may involve oxidative stress and reductions in endothelial cell survival factors (i.e., VEGF)[159,160]. However, inflammatory factors and immune cells may also play a vital role. In COPD, increased neutrophils and macrophages contribute to excess proteases and oxidative stress, while adaptive immune cells, such as CD8+ T cells and natural killer cells, mediate apoptosis of alveolar epithelial cells by releasing granzymes and perforin[42,161–163]. Although apoptosis of structural cells is often the focus of COPD literature, increased apoptosis of T cells has been reported in the blood and BAL of COPD patients compared to controls[160,164], which may lead to an inadequate immune response to infective organisms, contributing to the high frequency of infections seen in COPD.

In contrast, proteins involved in matrix remodeling processes (MMP-12, PTN, GDF-15, ADAMTS-5) were enriched in older participants with accelerated emphysema progression. Structural remodeling is a well-documented phenomenon in COPD, contributing to emphysema through lung elasticity loss. Structural changes are thought to result from enzymatic secretions produced by infiltrating immune cells, which destroy and remodel the ECM in the parenchyma[165]. Matrix metalloproteinases (MMPs), including MMP-12, are a central family of enzymes responsible for ECM remodeling[166]. MMP-12, released exclusively by macrophages, breaks down elastin[167] and is elevated in the sputum of COPD patients in association with emphysema severity[168–170]. ADAMs are additional proteinases involved in ECM remodeling and have putative roles in airway diseases, including COPD. While ADAMTS5, the protein selected in our signature, has no documented association with COPD, deficiencies in ADAM17 protect against emphysema in mice[171]. PTN and GDF-15, the final proteins identified in this phenotype, both exhibit high affinity binding to glycosaminoglycans domains in ECM[172,173]. Therefore, it is

plausible that the increased levels observed in our signature results from their proteolytic cleavage during ECM remodeling. This possibility is consistent with previously reported associations between increased PTN and GDF-15 levels and reductions in DLCO[174] and FEV_1[175]. Taken together, our results extend previous studies highlighting the importance of apoptotic, inflammatory, and matrix remodeling processes in emphysema, by providing evidence that they act age-dependently.

Our study has limitations. Chief among these is the lack of a validation cohort and paired lung measurements, which limits insights into the clinical utility of our findings. However, alterations in the identified pathways have been extensively linked to emphysema, increasing confidence in the relevance of our findings. Additionally, our analysis groups disproportionately consist of non-Hispanic white individuals[176]. To mitigate the influence of individual participants, we applied an iterative bootstrapping framework during model generation. Still, the generalizability of our findings remains unclear. Because there is no universally agreed-upon definition of accelerated emphysema progression, we used a percentile cut-off. Although this classification scheme effectively captured subclasses with different magnitudes of progression, we acknowledge that this approach may lead to discrepancies between future studies. However, percentile approaches have been commonly used to define other COPD outcomes, including FEV_1 decline[97,103]. Lastly, the age of participants in our analysis were limited by the inclusion criteria of COPDGene, which enrolled individuals ages 45-80. Because studies suggest that alterations in lung function as early as 30 years of age indicate later lung function impairments, this work requires exploration in younger cohorts, such as the newly enrolling SPIROMICS Study of Early COPD Progression (SOURCE)[177], which are better powered to explore progression in a younger participants.

In summary, data-driven modeling approaches identified early plasma proteomic signatures and provided insight into potential mechanisms associated with age-dependent progression of emphysema in ever-smokers. This work highlights the ability of quantitative, systems-focused analytical techniques to deconvolve the influence of various clinical and demographic variables in human samples.

4.4 Methods

4.4.1 Human participants

COPDGene (ClinicalTrials.gov Identifier: NCT00608764) is an ongoing multicenter, prospective observational study designed to identify genetic factors associated with COPD[145]. Briefly, COPDGene enrolled participants aged 45-80 years at entry with a history of cigarette smoking (≥10 pack-years), either with COPD by the fixed ratio definition (post-bronchodilator $FEV_1/FVC < 0.7$) or without COPD; as controls, healthy individuals with no history of smoking were also enrolled. COPDGene participants (n=10,371) underwent a baseline examination (P1) and follow-up visit (P2), which occurred after approximately 5 years. The study was conducted according to the principles of the Declaration of Helsinki. The institutional review board approved the human study protocol of all participating centers, and methods were carried out in accordance with the relevant guidelines and regulations. All participants were aware of the study's intent and provided written informed consent before any procedures.

SOMAscan data (1.3k version) was available for a subset of COPDGene participants at P1 with a smoking history (n=1089). Participants with available PRM measurements at P1 and P2 (n=656) were considered for inclusion in this study (**Figure 4.1**). The final analysis was restricted to individuals > 65 or ≤ 55 years old with a smoking history ≥ 10 pack-years and GOLD 0-2 disease (n=315). Our progression metric, longitudinal change in PRM (ΔPRM), was

calculated using the following equation for each individual: [*PRM at P2 – PRM at P1*]. We calculated ΔPRM metrics for emphysema (PRM^{Emph}) and small airway disease (PRM^{fSAD}). Participant classification was performed independently for both outcomes. Using participant age and magnitude of ΔPRM, we dichotomized the GOLD 0-2 participants into four groups: older slow/stable progressors (> 65 years old and < 25th %tile ΔPRM; n=43), older fast progressors (> 65 years old and > 75th %tile ΔPRM; n=43), younger slow/stable progressors (≤ 55 years old and < 25th %tile ΔPRM; n=36), younger fast progressors (≤ 55 years old and > 75th %tile ΔPRM; n=36). The baseline characteristics of the participants involved in the final analyses associated with ΔPRM^{Emph} and ΔPRM^{fSAD} outcomes are summarized in **Table 4.1** and **Supplemental Table B.1**, respectively.

4.4.2 Plasma dataset

Fresh plasma samples collected at P1 were frozen in an 8.5 mL P100 tube (Becton Dickinson). SOMAmer© (slow off-rate modified aptamer) technology[136] (SomaLogic, Boulder, CO) was used to measure n=1305 proteins.

4.4.3 Derivation of age-related progression protein signature(s)

Based on proteomic measurements from eligible GOLD 0-2 participants (**Figure 4.1**), we used Elastic Net (EN) in tandem with PLSDA for feature selection to select optimal age-related signatures associated with the progression of (a) emphysema (ΔPRM^{Emph}) and (b) small airways disease (ΔPRM^{fSAD}). First, the data were randomly sampled without replacement to generate 2000 subsets. To correct the effects of class size imbalances during regularization, we completed resampling at the smallest class size (n=36). We then performed EN regularization on each of the 2000 subsets. Once regularization was complete, the proteins were reordered based on their

selection frequency throughout the EN iterations and fed in a step-forward manner into the PLSDA algorithm (starting with the protein with the highest selection frequency).

Model performance was evaluated at each step using *k*-fold cross-validation ($k = 10$). The smallest model with a cross-validated error within 0.03 of the minimal error was selected as the final classification signature. ROC curves were generated based on the classification ability of each PLSDA model. In all models, orthogonalization was employed on the protein measurements (X-block) to generate orthogonal components, thereby ensuring their lack of correlation with the class labels (Y-block). This procedure involves modifying the X-variables to eliminate the portion of their variance unrelated to class discrimination. Its purpose is to enhance model interpretability by isolating the variation most pertinent for class differentiation.

4.4.4 Bioinformatic analysis

Clustering: Hierarchical clustering of the 20-feature ΔPRM^{Emph} signature was generated with supervised average linkage clustering using Spearman's correlation coefficient as the distance metric. Samples were colored by age and progression phenotype.

Metascape analysis: Metascape[137] [https://metascape.org] was used to identify biological processes that were significant and differentially enriched in association with age and emphysema progression, using the identified ΔPRM^{Emph} signature (**Figure 4.2**). PLSDA loadings on LV1 and LV2 were used to segregate proteins which were increased in association with differences in age and magnitude of emphysema progression phenotypes, respectively.

4.4.5 Software summary

Hierarchical clustering was completed using MATLAB (v2017b, MathWorks, Natick, MA). Elastic net was implemented using Glmnet package in MATLAB[138]. We generated

PLSDA models and ROC curves using the PLS toolbox available in MATLAB (v8.2.1, Eigenvector, Mason, WA).

Chapter 5 Cytokine-Chemokine Network Changes in the Immune Responses of Individuals with COPD

5.1 Introduction

Chronic obstructive pulmonary disease (COPD) is a common respiratory condition with high global morbidity and mortality[178] whose pathogenesis is marked, at least in part, by a dysregulated immune and inflammatory response[34]. The central involvement of inflammation in COPD pathogenesis is supported by a myriad of studies that implicate nearly all immune cells (innate and adaptive) in COPD[34–39]. In addition to an increased burden of immune cell infiltrates in the lungs of COPD patients[38,39], many immune cells from individuals with COPD also exhibit unique cellular phenotypes[40–42], characterized by changes in proteolytic enzyme secretions[43], phagocytic capabilities[20,45], and receptor expression[44,46,47]. These observed alterations pair with modified cellular production of several inflammatory cytokines (ex., TNF-α, IFN-γ, IL-1β, IL-6, IL-8, IL-10) in the lungs and blood of individuals with stable COPD[35,48,56,60]. Yet, despite the significant immune alterations present in COPD, the relative importance of each of these factors contributing to the overall pathologic intercellular communication networks has been challenging to ascertain.

To date, studies of intercellular immune communication tend to measure cytokine secretions at a singular timepoint from an individual cell type in isolation, therefore lacking valuable information about the broader network of diverse cellular interactions that collectively drive immune system behavior. Consequently, we may generate unique insights by instead examining secreted factors from a heterogeneous population of cells in response to activation

with well-defined immune challenges. The promise of this approach is exemplified by a multivariate analysis of cytokine measurements from stimulated peripheral blood mononuclear cells (PBMCs) of HIV+ donors, which demonstrated that stimulating a heterogeneous milieu of patient-derived immune cells could yield novel perspectives into altered cell-cell communication networks specific to disease state[179]. Although blood is not the tissue directly involved in COPD pathology, the role of peripheral circulation in trafficking immune cells to the lung, in conjunction with its non-invasive nature of sample acquisition, underscores the advantages of its use in inflammatory research. In this work we hypothesize that multivariate analysis of temporal cytokine secretions from stimulated peripheral immune cells may help uncover novel changes in paracrine signaling networks associated with COPD pathobiology.

Using multivariate cytokine profiles as robust indicators of the overall state of the cellular network, we tested whether integrative analysis of multivariate proteomic signatures could identify critical differences in cytokine network behavior between peripheral blood mononuclear cells (PBMCs) from individuals with ≥20 pack-years smoking history with COPD or with normal spirometry (GOLD 0) after activation with an adaptive stimulus or innate stimuli. Our results suggest that the paracrine signaling networks generated by PBMCs from COPD donors, after immune challenge, significantly differ from GOLD 0 controls. Most notably, in response to adaptive stimulation, PBMCs from individuals with COPD exhibit deficient secretion of early M-CSF and IL-13 signals which drive divergent downstream responses. Together, our approach identified the most influential secreted cytokines that underlie complex immune cell communication networks and elucidated possible cellular actors responsible for these behaviors.

5.2 Results

5.2.1 *PLSDA-identified profiles of PBMC-secreted cytokines overcome clinical variability and*

identify unique proteomic profiles associated with innate or adaptive immune stimuli

To investigate changes in paracrine signaling networks that are altered in COPD, we utilized a data-driven methodology that would allow the identification of critical relationships between cytokine network behavior and disease state[179]. Specifically, we stimulated PBMCs from participants in the SPIROMICS cohort with well-defined immune challenges to examine diverse signaling pathways that activate a variety of unique cell types and receptors: beads coated with anti-CD3 and anti-CD28 antibodies (anti-CD3/CD28 beads) to stimulate polyclonal T cells, LPS to activate TLR4 primarily on innate immune cells, and the synthetic imidazoquinoline compound R848 to activate TLR7 and TLR8 in innate immune cells. LPS and R848 mimic (Gram-negative) bacterial and viral stimulation, respectively. For each donor and stimulus combination, we obtained a set of 96 cytokine measurements (48 cytokines measured 6 and 72 hours after PBMC stimulation) by Luminex assays. To help control participant-to-participant variability, we normalized each stimulation condition to an unstimulated control from the same donor. Despite normalization, none of the 96 individual cytokine measurements could independently overcome clinical variability and differentiate all three stimulus conditions with statistical significance (**Figure 5.1a-d**).

In contrast, a multivariate analysis using partial least squares discriminant analysis (PLSDA) identified distinct cytokine profiles associated with each stimulus. PLSDA is a method used for supervised pattern recognition and dimensionality reduction to identify weighted linear combinations of individual cytokine measurements (termed “latent variables”) that most effectively classify a dependent variable of interest (i.e., donor class or stimulus types). PLSDA is a useful analytical method for gaining biological insight because latent variables identified by the model are often biologically meaningful[179–181]. Variable importance in projection (VIP)

scores can be used in tandem with PLSDA analysis to assess the importance of each variable (i.e., cytokine) in discriminating between different donor classes. Here, we used VIP scores to eliminate variables that did not contribute to classification (see Methods).

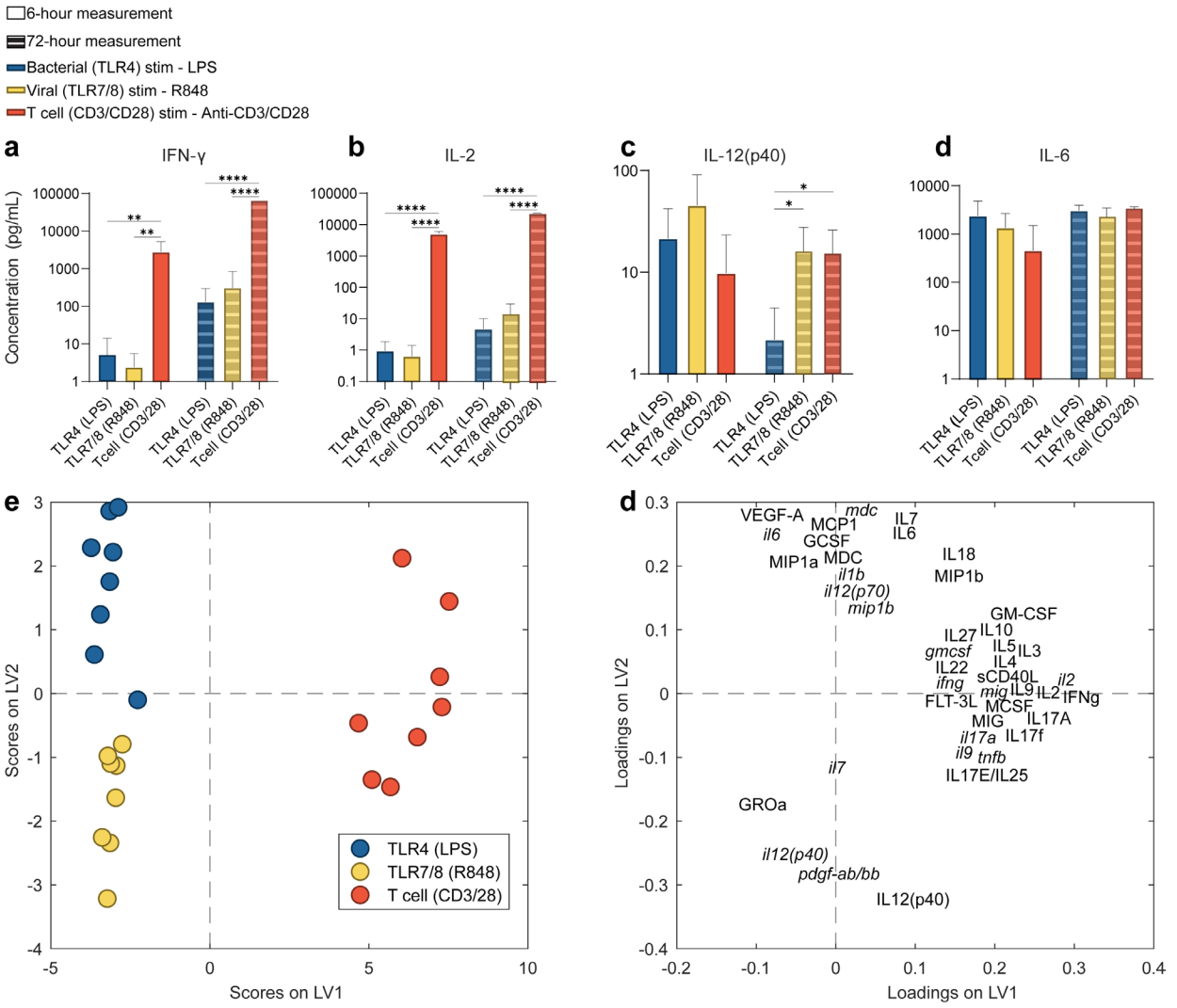


Figure 5.1. Dimensionality reduction by PLSDA of 48 cytokines measured at 6 and 72 hours after the stimulation of PBMCs from healthy donors.
(a-c) Representative individual cytokine measurements from PBMCs of healthy, never-smoking donors (n=8) 6-hours (solid) and 72-hours (dotted) after stimulation with TLR-4 (LPS; blue), TLR-7/8 (R848; yellow), or anti-CD3/CD28 stimulus (red). One-way ANOVA with Tukey's post-hoc test; *p<0.05, **p<0.01, ***p<0.001. (d) PLSDA scores and (e) loadings plots after VIP selection reveal cytokine signatures that differentiate stimulus profiles with 97.9% calibration accuracy and 91.7% cross-validation accuracy. A model with two latent variables (LVs) captured 56.4% of the total variance in the dataset. 6- and 72-hour loadings are indicated by lower and uppercase labels, respectively. LV1 separates cytokine secretion associated with adaptive simulation (positive loadings on LV1) from innate stimulation (negative loadings on LV1), and LV2 further separates bacterial innate stimulation (positive loadings on LV2) from viral innate stimulation (negative loadings on LV2).

In an analysis with healthy never-smokers (n=8), VIP selection identified 43 cytokine measurements that distinguished stimulus-specific secretions between anti-CD3/CD28–, R848-, and LPS-induced PBMC secretion profiles with 97.9% calibration accuracy and 91.7% cross-validation accuracy across two latent variables (**Figure 5.1e-f**). Latent variable 1 (LV1) differentiated innate and adaptive stimulus responses, while LV2 further discriminated between innate stimuli (LPS vs. R848). The scores (**Figure 5.1e**) and loadings (**Figure 5.1f**) plots, indicated that the adaptive stimulus profile, summarized by cytokines with positive loadings across LV1, was composed of IFN-γ, IL-2, IL-9, MIG, IL-17A, and GM-CSF at both 6 and 72 hours, as well as secretion of IL-3, IL-4, IL-5, IL-10, IL-17F, IL-17E/IL-25, IL-22, IL-27 at 72 hours. These cytokines we similar to those previously measured in PBMCs of HIV-infected individuals after an adaptive stimulus[179]. The LPS profile (2nd quadrant) had substantial contributions from IL-6 (6 and 72 hours), MDC (6 and 72 hours), IL-1β (6 hours), IL-12p70 (6 hours), MCP-1 (72 hours), G-CSF (72 hours), and MIP-1a (72 hours), whereas the R848 profile (3rd quadrant) contained IL-12p40 (6 and 72 hours), PDGF-AB/BB (6 hours), and GROα (72 hours). Taken together, cytokine profiles were better capable of overcoming clinical variability and discriminating between immune responses than individually evaluated cytokines.

5.2.2 Multivariate cytokine profiles reveal altered immune responses in PBMCs from donors with COPD compared to ever-smoking controls

Once we established the ability of multivariate models to differentiate immune responses derived from *in vitro* stimulation of PBMC from never-smoking controls, we applied our analytical approach to understand critical alterations in paracrine signaling networks associated with COPD. A dysregulated immune response marks COPD pathology[34]. However, the relative importance of immune factors (i.e., cytokines and cells) in these intercellular communication

networks has been difficult to ascertain. Because immune system behavior arises from cell-cell interactions, unique insights may be derived from examining the overall system output of heterogeneous immune cell populations in response to well-defined immune challenges. To test this hypothesis, we compared multivariate cytokine responses from PBMCs after adaptive and innate immune stimulation from two patient groups: individuals with a smoking history with COPD (n=13) and without COPD (GOLD 0; n = 14). The presence of COPD was defined by the fixed ratio definition ($FEV_1/FVC < 0.7$). GOLD 0 and COPD participants were well-matched for all demographic criteria (**Table 5.1**).

Table 5.1. Demographics for GOLD 0 and COPD participants from SPIROMICS at visit 5

	GOLD 0 (N=14)	COPD (N=13)	P-Value*
Age	61 (± 9.2)	67 (± 7.8)	0.06
Currently Smoking	4 (28%)	4 (30%)	> 0.99
BMI	29 (± 5.5)	26 (± 4.8)	0.19
Sex (Male)	9 (64%)	8 (61%)	> 0.99
Race (White / Other)	5/8 (38%)	9/4 (69%)	0.24
FEV_1 (% predicted)	**100 (±22)**	**79 (± 19)**	**0.01**
FEV_1/FVC	**0.76 (± 0.03)**	**0.59(± 0.13)**	**< 0.0001**

*Two-sample, two-tailed t-test or Fisher's exact test were used to determine significant differences. Bold values denote significant differences between GOLD 0 and COPD participants.

To determine multivariate differences in cytokine profiles between GOLD 0 and COPD participants, we generated three separate feature-selected PLSDA models (one for each stimulus type) using VIP scores. All cytokine data used in these models were normalized to a donor-

matched unstimulated control and adjusted for batch and smoking status (see **Methods**). In all three models, the differentiation of profiles from GOLD 0 and COPD participants was captured by the first latent variable (LV1). The multivariate cytokine profiles produced by PBMCs from COPD donors stimulated with anti-CD3/CD28 beads were characterized by a notable reduction in early (6-hour) cytokine signals (**Figure 5.2a-b**). Secretions from COPD donors were enhanced for cytokines involved in a Type 1 (Th1) immune response, such as IL-2 and IL-12, whereas PBMCs from GOLD 0 donors produced factors more traditionally associated with a Type 2 (Th2) immune response (i.e., IL-13, IL-4, and IL-5). In contrast to the adaptive stimulus response, after innate stimulation PBMCs from COPD donors were characterized by increases in the concentration of early (6-hour) cytokine secretions (**Figure 5.2d-f**). However, similar to the patterns observed in the adaptive stimulus model, late induction of Th2-associated factors, including IL-13, IL-5, IL-4, and IL-10, were also present and elevated in PBMC responses from GOLD 0 donors in both innate models (LPS and R848). Unsupervised hierarchical clustering with the VIP-selected cytokines profiles identified COPD patients with 100% accuracy in the anti-CD3/CD28 and R848 models and 60% accuracy in the LPS model (**Supplemental Figure C.1**).

We next explored whether differences in the percentages of immune cell types between donor classes might have been the cause for the observed alterations in cytokines profiles. Specifically, we compared the percentages of pan (CD3+) T cells, CD8+ T cells, CD4+ T cells, B cells, NK cells, dendritic cells, and monocytes among COPD and GOLD 0 donors. COPD donors had significantly decreased percentages of CD8+ T cells at 6 and 72 hours and significantly increased proportions of CD4+ T cells at 72 hours (**Figure 5.3**). However, after adjustment for current smoking status using logistic regression models, only the differences in

the 72-hour CD8+ T cells percentages remained significant ($p = 0.015$). Because CD8+ T cells secrete IL-2, IFN-γ, and other T cell-derived cytokines in response to anti-CD3/CD28 stimulation, this change could have contributed to variances in the production of cytokines between PBMCs from GOLD 0 and COPD donors in response to adaptive stimulation. However, the observed differences in cellular distribution were measured in the absence of stimulation and were only present 72 hours after incubation, which may instead reflect donor-specific cellular turnover. While some late (72-hour) cytokine secretions may result from differences in CD8+ T cell distributions, it was difficult to assess changes directly for all 92 cytokines that differentiated the GOLD 0 and COPD profiles across the three independent stimuli models (**Figure 5.3**). Therefore, we used random forest algorithms to determine the most influential cytokines for differentiating cytokine profiles from stimulated PBMCs from COPD and GOLD 0 donors.

5.2.3 The altered response of PBMCs from COPD donors to adaptive immune stimuli is driven by reduced early M-CSF and IL-13 secretion

Proteomic profiles identified by PLSDA suggested that PBMCs from COPD donors maintained the ability to secrete cytokines in response to all stimulus challenges. However, the composition of the responses from COPD donors were largely altered compared to those from GOLD 0 donors. Using a random forest (RF) algorithm we identified the most influential cytokines within each stimulus profile that differentiated PBMC responses across donor classes. RFs are a complementary ensemble learning technique that emphasizes contingent, rather than independent, contributions of cytokines to classification. We independently applied RF algorithms to the VIP-selected datasets from the three stimuli-specific models. The resulting RF models classified PBMC responses from COPD donors to anti-CD3/CD28, LPS, and R848 stimulation with 85%, 54%, and 62% accuracy, respectively (**Figure 5.4a**).

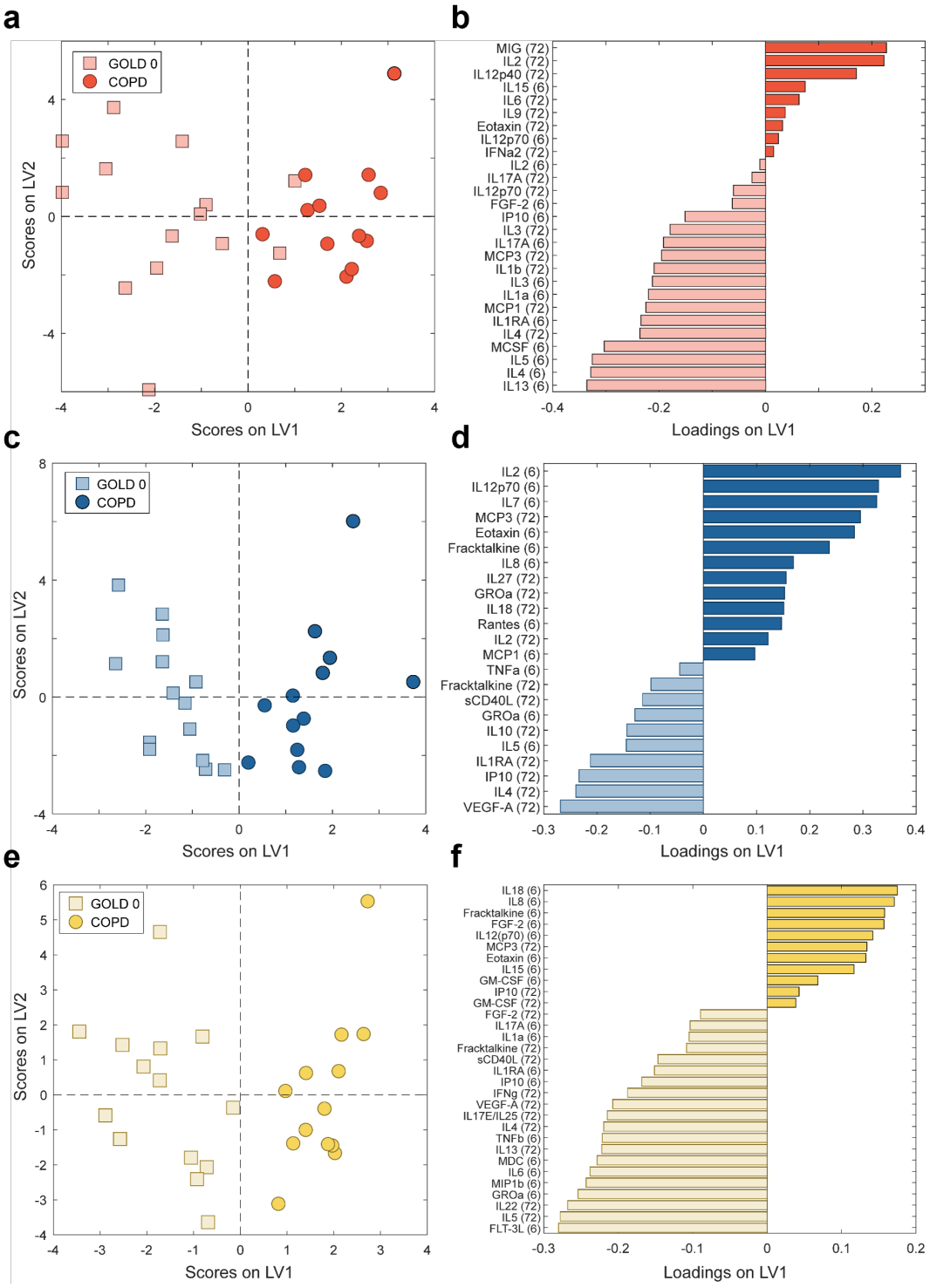


Figure 5.2. Comparison of multivariate cytokine profiles from PBMCs of individuals with COPD and ever-smoking controls without COPD (GOLD 0).
(a-b) Analysis of the cytokine profiles of the PBMCs derived from GOLD 0 (squares) and COPD (circles) donors from the SPIROMICS cohort, in response to stimulation with anti-CD3/CD28 beads. A model with one latent variable captured 26% of variance in the cytokines dataset and performed with 89% calibration accuracy and 81% cross-validation (CV) accuracy. (c-f) Analysis of the cytokine profiles of the indicated sources of PBMCs in

response to stimulation with LPS (c-d) and R848 (e-f). The LPS model (c-d) captured 15% of the variance in the cytokine dataset and performed with 85% calibration accuracy and 78% CV accuracy. The LPS model (e-f) captured 14.4% of the variance in the cytokine dataset and performed with 85% calibration and 70% CV accuracy. All PLSDA models underwent feature selection with VIP scores.

The cytokine(s) with the most significant impact on differentiating the secretion profiles observed in response to adaptive stimulation were early (6-hour) M-CSF and IL-13 (**Figure 5.4b**). Together these two cytokines accounted for ~15% of the overall classification accuracy of the anti-CD3/CD28 RF models. RF algorithms were not as helpful in identifying individual cytokine drivers of secretion profiles resulting from innate immune challenge; a result echoed by the modest classification accuracies associated with LPS and R848 RF models (**Supplemental Figure C.2**). Given the success in pinpointing early secretions in response to anti-CD3/CD28 stimulus that differentiated PBMC responses from GOLD 0 and COPD donors, we next investigated potential mechanisms contributing to the divergent paracrine signaling events observed after T cell activation.

To explore the potential cellular sources of early M-CSF and IL-13 secretion, we performed correlations between immune cell subsets and 6-hour M-CSF and IL-13 levels within each donor type (COPD or GOLD 0). In GOLD 0 donors, early IL-13 concentrations correlated with 6-hour CD4+ T cell percentages ($r = 0.68$, $p = 0.006$; **Figure 5.4c**, **Supplemental Figure C.3**). Interestingly, early M-CSF levels also trended towards correlating with the percentage of CD4+ T cells in the GOLD 0 donors ($r = 0.41$, $p = 0.11$). Conversely, there were no significant associations between cell populations and 6-hour M-CSF or IL-13 secretions in COPD donors (**Figure 5.4d**). Notably, PBMCs from COPD donors had nominally higher percentages of CD4+ T cells than PBMCs from GOLD 0 donors (**Figure 5.3**). Hence, the lessened secretion of M-CSF and IL-13 from COPD samples are not simply explained by reduced CD4+ T cell populations. Instead, the observed deficits may be linked to altered phenotypic or functional behaviors of T

cells in individuals with COPD.

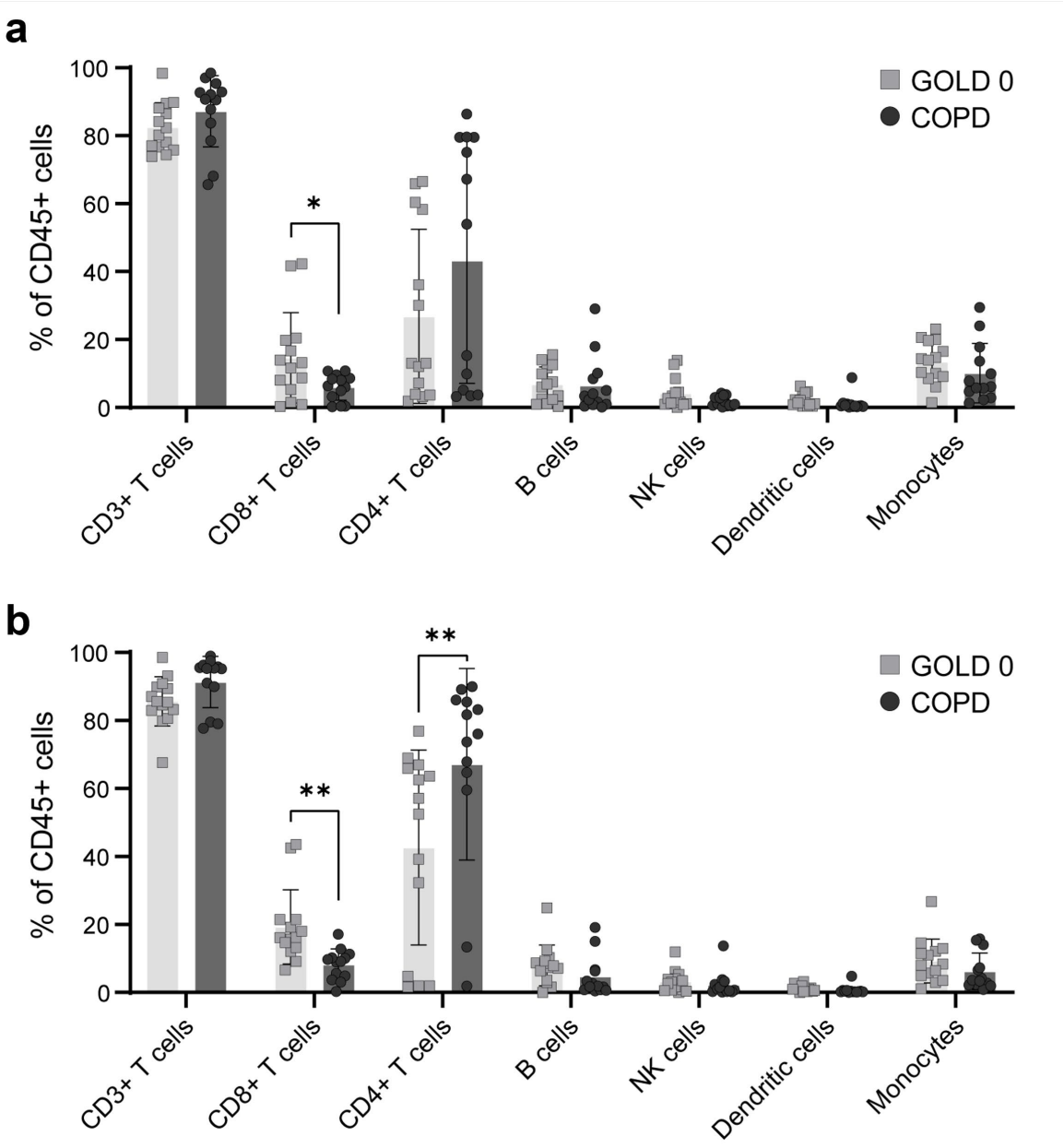


Figure 5.3. Analysis of the cellular composition of PBMCs from various donors.
(a-b) Cellular distribution across PBMC samples from GOLD 0 (grey squares) and COPD (black circles) donors (a) 6- and (b) 72-hours after incubation with R10 media. Significance determined by unpaired, two-tailed t-tests (*$p<0.05$, **$p<0.01$). After adjustment for smoking in logistic regression models, the significant comparisons were reduced to: CD8+ T cells at 6-hours ($p = 0.1$), CD8+ T cells at 72-hours ($p = 0.015$), and CD4+ T cells at 72-hours ($p = 0.05$). Cellular subtypes were measured by the following surface markers: Pan T cells (CD3), CD8+ T cells (CD8), CD4 T cells (CD4), B cells (CD20), NK cells (CD56), dendritic cells (CD11c+ HLA-DR), and monocytes (CD14).

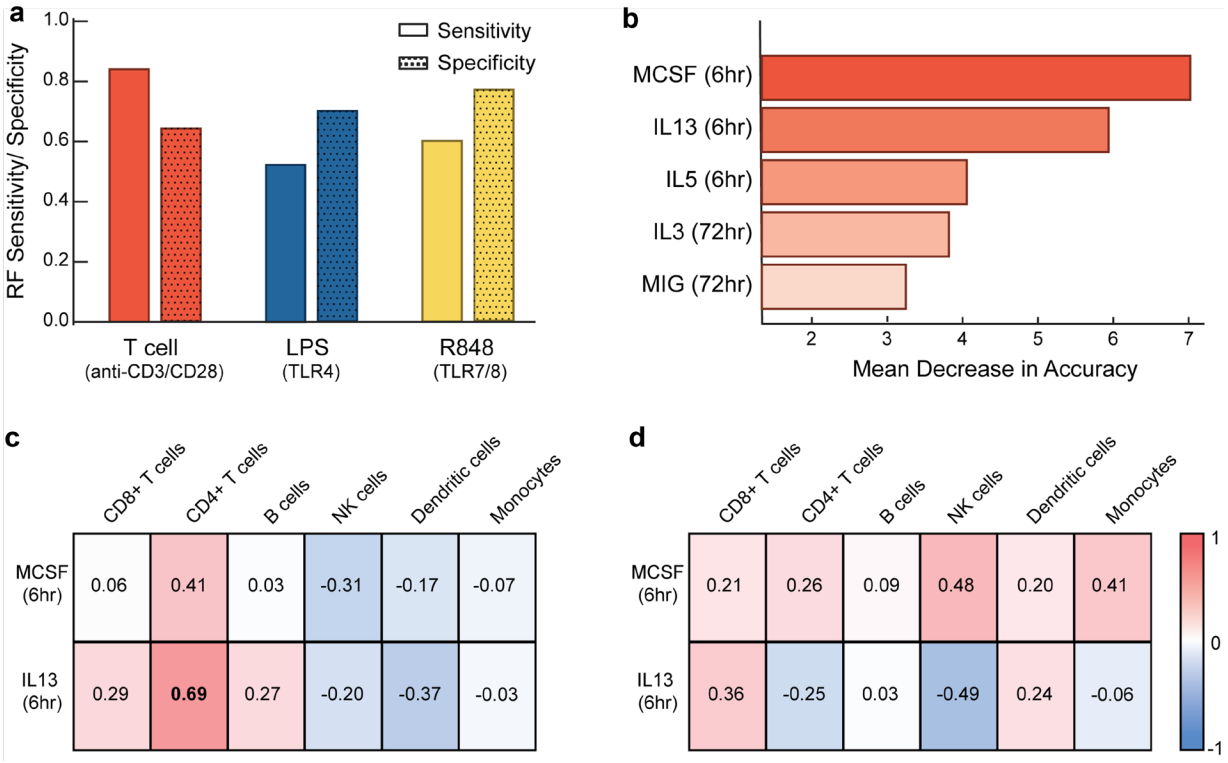


Figure 5.4. Random forest analysis reveals the hierarchy of importance of cytokine secretion events in distinguishing PBMCs responses from GOLD 0 and COPD donors.
(a) Sensitivity (solid bars) and specificity (dotted bars) of random forest (RF) classification models generated from VIP-selected cytokines profiles of PBMCs from GOLD 0 and COPD donors after stimulation with anti-CD3/CD28 beads (red), LPS (blue), and R848 (yellow). (b) Plot of the top 5 most important features in the anti-CD3/CD28 RF model contributing to classification of cytokines profiles from GOLD0 and COPD donors. Importance is determined by the percent accuracy each protein contributes to the overall classification accuracy of the RF model. X-axis denotes the percent accuracy lost by the overall model if the cytokine in question is removed. (c-d) Pearson correlation coefficients between RF identified proteins, M-CSF (6hr) and IL-13 (6hr) and cellular distributions at 6 hours in GOLD 0 (c) and COPD (d) donors. Bolded values are associated with significant p-values ($p < 0.05$).

We next created correlation networks of the VIP-selected cytokine profiles in response to anti-CD3/CD28 stimulation to further explore how variable levels of M-CSF and IL-13 altered the behavior of protein-protein connectivity between donor classes. Using pairwise correlations between the 28 VIP-selected cytokines in the anti-CD3/CD28 model, we built independent networks for GOLD 0 and COPD donors. The cytokine network based on PBMC responses to adaptive stimulation from GOLD 0 donors contained two modules connected with 6-hour M-CSF (**Figure 5.5a**). Interestingly, in the COPD network, the connections linking the two modules

[M-CSF (6 hours) with MCP-1 (72 hours) and M-CSF (6 hours) with MCP-3 (72 hours)], are lost, resulting in two independent networks (**Figure 5.5b**). Additionally, while M-CSF (6 hours) and IL-13 (6 hours) are nodes with a high degree of connectivity in both networks, in COPD, these cytokines lack nearly all connections with late (72-hour) cytokine signals (**Figure 5.5c**). We speculate that this behavior may lead to divergent downstream signaling events between donor classes.

5.2.4 Multivariate models predict that reduced secretion of M-CSF and IL-13 by PBMCs from COPD donors early in response to adaptive stimuli markedly influences later cytokine profiles

RF models indicated that a vital difference in the cytokine profiles generated by PBMCs from COPD donors was the reduced secretion of M-CSF and IL-13 early in the response to adaptive stimulation. Although it is unclear with the present data which cell type is definitively responsible for these secretions, we chose to continue to explore this result using the current dataset. Specifically, we tested whether the divergence in the 72-hour cytokine profiles of the differently sourced PBMCs after anti-CD3/CD28 stimulation could be related to the early loss of M-CSF or IL-13, suggesting that individual cellular and molecular effects propagate through the cell-cell communication network to achieve broader effects.

To explore this, we generated a new PLSDA model for PBMC responses to anti-CD3/CD28 stimulation using only the 72-hour cytokine measurements to differentiate between the responses of PBMCs from GOLD 0 and COPD donors. VIP scores were used to eliminate cytokines that did not strongly contribute to class differentiation. We found that 16 of the 72-hour cytokine measurements (IL-12(p40), IL-2, MIG/ CXCL9, IL-9, IL-6, IL-17A, Eotaxin, IFNα2, PDGF-AA, MCP-3, IL-10, IL-3, MCP-1, IL-1RA, IL-1β, IL-4) distinguished the responses of PBMCs from COPD and GOLD 0 donors with 93% calibration accuracy and 85%

cross-validation accuracy (**Figure 5.6a**). The first two latent variables cumulatively captured 40% of the variance in the cytokine dataset. LV1 best separated the responses from COPD and GOLD 0 donors (**Figure 5.6b**). To determine the magnitude and direction of the relationship between the differences in the 72-hour profiles and deficiencies in the secretion of M-CSF and IL-13 at 6 hours, we computed Pearson correlation coefficients for the levels of M-CSF and IL-13 at 6 hours and the corresponding score on LV1 from the 72-hour anti-CD3/CD28 PLSDA model. We found a statistically significant inverse relationship between M-CSF ($r = -0.42$, $p = 0.03$) and IL-13 ($r = -0.52$, $p = 0.005$) with LV1 scores, confirming correspondence between the loss of early M-CSF and IL-13 and the divergence in the cytokine profiles at 72 hours. These associations remained for M-CSF ($p = 0.02$) and IL-13 ($p = 0.005$) even after adjustment for sex, age, current smoking status, and experimental batch (**Figure 5.6c-d**). Except for IL-5 and IL-4, which are involved in Type 2 inflammation similar to IL-13, no significant correlation existed between any other 6-hour cytokine measurements and the scores on LV1 from the 72-hour model (**Supplemental Figure C.4**).

We lastly aimed to better characterize which specific downstream (72-hour) signals, following PBMC challenge with an adaptive stimulus, were most influenced by alterations in early M-CSF or IL-13 secretion and how these associations varied across donor classes (COPD vs. GOLD 0). We performed correlations between M-CSF or IL-13 at 6 hours, and cytokines measurements at 72 hours for GOLD 0 and COPD donors separately (**Figure 5.6e-h**, **Supplemental Figure C.5**). Early M-CSF secretion from GOLD 0 donors positively correlated with later secretion of chemotactic factors (MCP-1: $r = 0.85$, $p < 0.001$; MCP-3: $r = 0.71$, $p = 0.005$; M-CSF: $r = 0.62$, $p = 0.02$), cytokines related to Type 17 (Th17) inflammation (IL-17A: $r = 0.64$, $p = 0.01$; IL-17F: $r = 0.68$, $p = 0.008$) and IL-10 ($r = 0.59$, $p = 0.03$; **Figure 5.6e**).

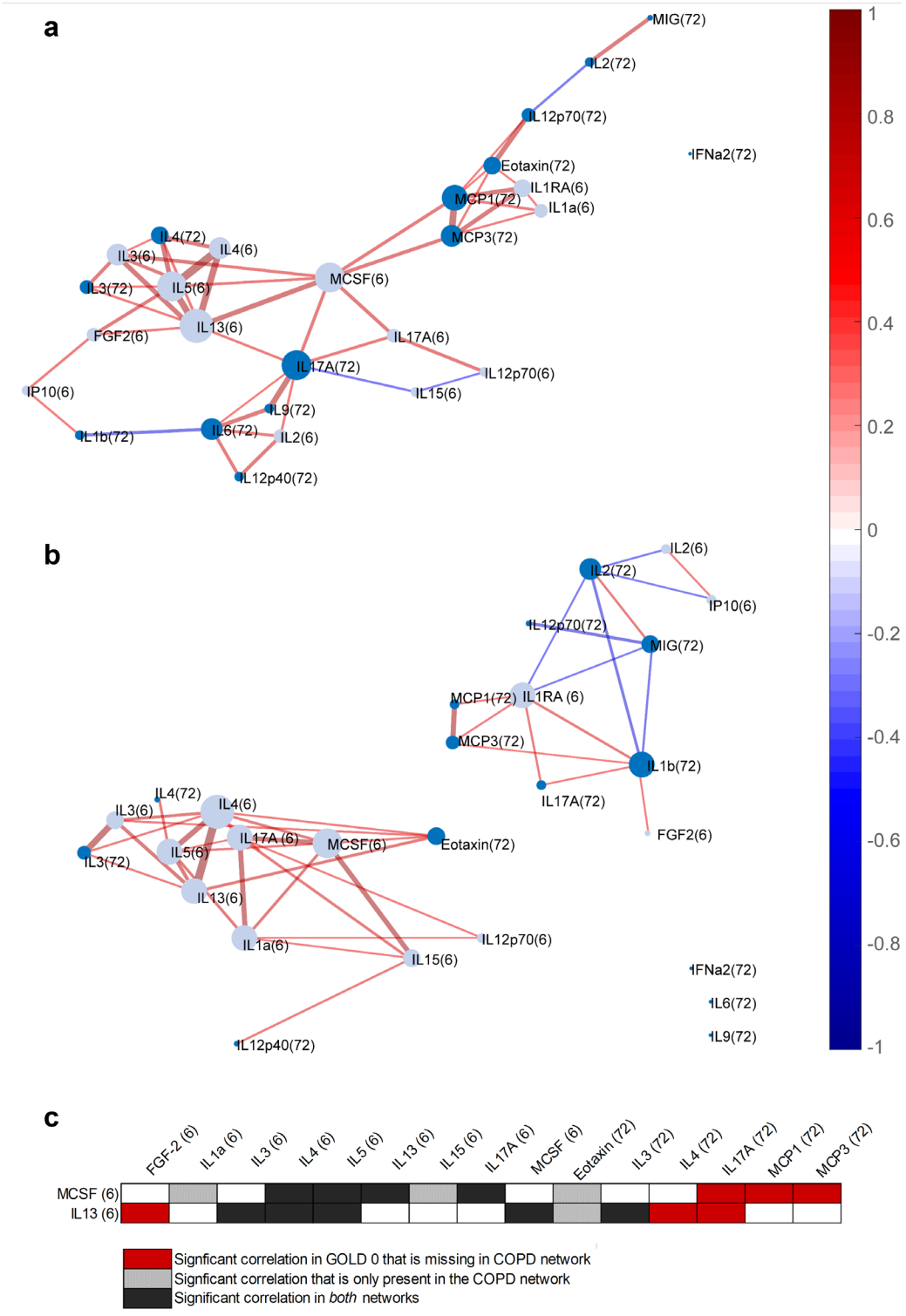


Figure 5.5. Correlation networks of VIP-selected cytokines from anti-CD3/CD28 model highlight key differences in cytokine-chemokine network connectivity.
(a-b) Protein correlation networks of the VIP-identified cytokine profiles from PBMCs of GOLD 0 (a) and COPD (b) donors in response to anti-CD3/CD28 stimulation. A line connecting two proteins indicates the presence of a significant ($p < 0.05$) correlation, as calculated by Pearson's correlation coefficient. Brighter and thicker lines

indicate stronger, more significant correlations, respectively. The value of the correlation coefficient for both networks is displayed in the color bar scale on the right, with red indicating a positive relationship and blue a negative relationship. Light blue nodes indicate protein measurement at 6-hours while dark blue nodes indicate 72-hour measurements. (c) Table summarizing the significant correlations between 6-hour M-CSF and IL-13 in GOLD 0 and COPD networks.

Conversely, early M-CSF produced by PBMCs from COPD donors lacked all those downstream associations and instead displayed moderate significant correlations with Eotaxin ($r = 0.63$, $p = 0.02$) and IL-3 ($r = 0.58$, $p = 0.04$) at 72 hours (**Figure 5.6f**). This lack of shared behavior between temporal signals across donor classes was similarly observed in IL-13 correlations. In

fact, except for 72-hour IL-3 (COPD: $r = 0.58$, $p = 0.01$; GOLD 0 $r = 0.68$; $p = 0.009$), early IL-13 secretion from PBMCs of GOLD 0 and COPD donors shared no significant correlations (**Figure 5.6g-h**). Again, early IL-13 levels from PBMCs of COPD donors exhibited limited correlations with later signaling events, with IL-13 levels at 6 hours from COPD donors correlating significantly with only three cytokines (Eotaxin: $r = 0.63$, $p = 0.01$; IFNα2: $r = 0.42$, $p = 0.02$; IL-12p70: $r = 0.36$, $p = 0.01$; **Figure 5.6h**). Conversely, early IL-13 levels from anti-CD3/CD28 stimulated PBMCs of GOLD 0 donors correlated with later increases in Th2 (IL-4: $r = 0.61$, $p = 0.02$, IL-5: $r = 0.69$, $p = 0.008$; IL-10: $r = 0.56$, $p = 0.04$) and Th17-related factors (IL-17A: : $r = 0.69$, $p = 0.03$; IL-17F: $r = 0.69$, $p = 0.008$;), as well as G-CSF ($r = 0.54$, $p = 0.04$), IL-7 ($r = 0.61$, $p = 0.02$) and MCP-1 ($r = 0.64$, $p = 0.02$; **Figure 5.6g**). These findings highlight a significant loss of downstream signaling diversity in immune cell networks from individuals with COPD due to reduced capacity to secrete M-CSF and IL-13 early in the response to adaptive stimulation. Together, these data suggest that early impaired M-CSF and IL-13 secretion from COPD participants may critically affect the shaping of the immune response to adaptive stimuli.

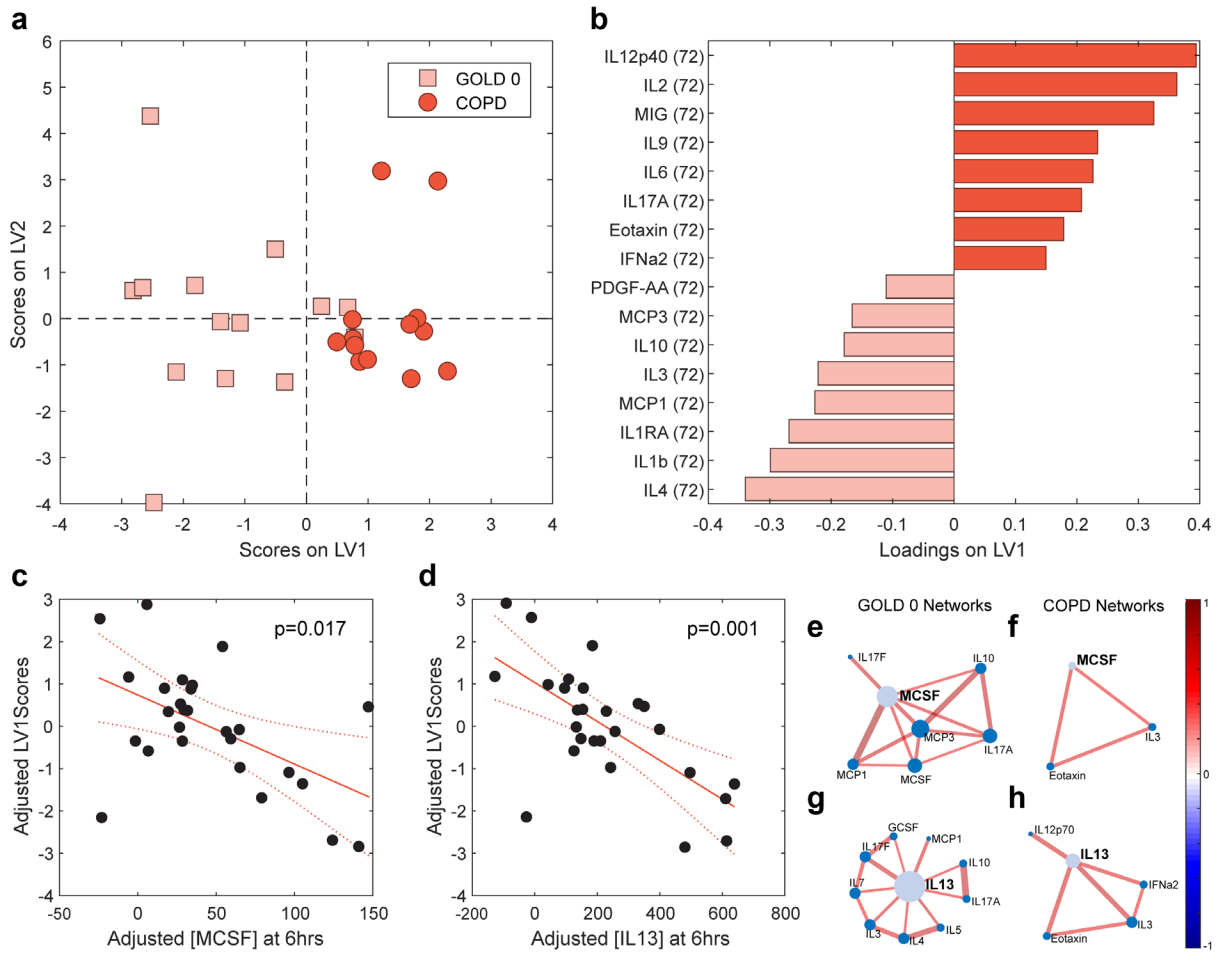


Figure 5.6. The magnitude of IL-13 and M-CSF secretion 6 hours after anti-CD3/CD28 stimulation is associated with the divergence of cytokine profiles at 72 hours.
(a) Scores and (b) loadings plots from PLSDA depicting VIP-selected cytokine profiles 72 hours after the stimulation of PBMCs from GOLD 0 and COPD donors with anti-CD3/CD28. This analysis distinguished the responses of PBMCs COPD donors (n = 13; red circles) from those of GOLD 0 donors (n = 13; pink squares) with 85% cross-validation accuracy and 93% calibration accuracy. A model with two latent variables captured 40% of the variance in the cytokine dataset. (c) The magnitude of M-CSF and (d) IL-13 secretion 6 hours after stimulation was significantly correlated with late cytokine responses (or score on LV1) after adjustment for sex, age, current smoking status and experimental batch by linear regression. Plots illustrate data after covariate adjustment. (e-f) Significant correlations between 6-hour M-CSF levels and 72-hour cytokine secretions from PBMCs (e) GOLD 0 and (f) COPD donors in response to anti-CD3/CD28 stimulus. (g-h) Significant correlations between 6-hour IL-13 levels and 72-hour cytokine secretions from PBMCs (g) GOLD 0 and (h) COPD donors in response to anti-CD3/CD28 stimulus. A line connecting two proteins indicates the presence of a significant ($p < 0.05$) correlation, as calculated by Spearman's correlation coefficient. All non-significant correlations were removed from the network graphic. Brighter and thicker lines indicate stronger, more significant correlations, respectively. The value of the correlation coefficient for both networks is displayed in the color bar scale on the right, with red indicating a positive relationship and blue a negative relationship. Light blue nodes indicate protein measurement at 6-hours while dark blue nodes indicate 72-hour measurements.

5.3 Discussion

This study aimed to investigate how cytokine secretion profiles of PBMCs from COPD donors differed from ever-smoking control donors (GOLD 0) after exposure to diverse immune stimuli (both innate and adaptive). To achieve this, we employed PLSDA with VIP feature selection which facilitated the identification of temporal cytokine signatures that distinguished between stimuli-specific and donor-specific responses. These signatures provided valuable information into the cytokine subsets that best characterized each type of response (i.e., stimulus and donor combination). Subsequent RF models defined the most influential cytokine events in differentiating donor classes. Follow-up investigations highlighted likely cell types responsible for these critical events and explored the implications of early aberrant responses to adaptive stimuli in driving downstream signaling behavior.

The PBMC responses from COPD donors showed significant differences compared to GOLD 0 controls when stimulated with adaptive (anti-CD3/CD28) and innate (LPS and R848) immune challenges. Notably, across all stimulus types, PBMCs from GOLD 0 donors exhibited higher secretion of CD4+ T cell-associated factors, such as IL-4, IL-5, and IL-13, compared to PBMCs from COPD donors. RF models revealed that early secretion of M-CSF and IL-13 played a vital role in distinguishing between the adaptive immune responses of PBMCs from GOLD 0 and COPD donors. Correlation analyses with matched flow cytometry data indicated that, in GOLD 0 donors, early M-CSF and IL-13 are likely generated by CD4+ T cells upon activation. However, in PBMCs from COPD donors, despite nominally higher percentages of CD4+ T cells, this relationship appeared to be diminished. A final network analysis predicted that early secretion of M-CSF acts as an integral node guiding downstream cytokine profiles after adaptive stimulation. These findings suggest a critical deficit in the adaptive immune

response in individuals with COPD.

The disparities observed between the PBMC responses from COPD and GOLD 0 donors could have resulted from synergistic cell-cell interactions, both direct and indirect, within the heterogeneous PBMC population. These findings align with previous reports of diminished cytokine responses in T cells from individuals with COPD[182]. One potential reason for the loss of these critical T cell signals may be partially explained by phenotypic alterations in CD4+ T cell populations in COPD. For example, previous studies have reported down-regulation of the surface receptor CD28, an essential costimulatory molecule required for T cell activation, in CD4+ T cells from individuals with COPD[183,184]. Our adaptive stimulus system activated T cells *in vitro* by providing dual signaling through CD3 and CD28. Considering the reduced CD28 receptors previously observed in T cells from COPD donors, it is plausible that this deficiency may have resulted in standalone signaling through CD3, potentially promoting T cell apoptosis or functional alteration into an inactivated state, reducing overall cytokine secretion. Another potential reason for the observed disparities in cytokine profiles between donor classes could be a genetic or epigenetic predisposition of T cells from COPD donors to elicit an altered T-helper (Th) cell phenotype. Genetic or epigenetic modifications in T cells may influence their responsiveness to immune stimuli, leading to distinct cytokine expression patterns. Our analysis suggests that T cells from COPD patients may be predisposed to elicit a Type 1 (Th1) response (i.e., IL-2, IL-12, IFN-γ) over a Type 2 (Th2) inflammatory response (i.e., IL-4, IL-5, IL-13). Similar increases in this Th1:Th2 response ratio have been reported previously in COPD[185,186]. However, contrasting findings are present in other studies of epithelial brushings, which report enhanced expression of Th2-related genes in a subset of COPD patients[72]. These discrepancies are likely due to differing sample sources (blood immune cells vs. bronchial epithelial cells).

Nevertheless, the crucial interplay between T cells and other immune cell populations in shaping the overall immune response in COPD[187] underscores the significance of our findings, which highlight a potential role for T cells as critical regulators in disease pathology.

To our knowledge, this is the first study to identify critical early regulators of immune signaling networks in response to adaptive stimuli that influence the impaired immune responses observed in COPD. This work builds upon previous *in vitro* studies[188] by providing a novel, multivariate analysis of temporal peripheral immune cell secretions from individuals with COPD after diverse immune stimulation. This approach enables the recapitulation of complex paracrine signaling networks similar to those present *in vivo*, adding valuable insights to our understanding of the immune responses in COPD. However, our study has limitations, with the primary concern being the use of frozen PBMC samples. It is conceivable that the freezing procedure altered the cellular processes or cell-cell interactions that led to the observed disparities between donor classes. However, this technical limitation is broadly applicable given the nature of cell sampling and the cryopreservation of PBMCs throughout the field. Additionally, groups have reported that cytokine profiles measured with fresh (never frozen) PBMCs isolated from healthy donors are similar to those measured from thawed PBMCs[179]. Furthermore, while our results suggest an essential role for M-CSF in dictating adaptive immune responses, follow-up intracellular flow cytometry and blocking experiments are required to confirm the cell types responsible for these critical secretions and validate the significance of this signaling event in governing later responses. Lastly, it is important to note that while peripheral circulation plays a vital role in trafficking immune cells to the lung, our current findings are specific to the blood. To fully comprehend how the observed behavior relates to local alterations in the lung, further investigations in murine models or lung-based specimens are warranted. Such studies will

provide valuable insights into the intricate interplay between systemic and local immune responses in COPD.

In summary, our findings implicate early M-CSF as a critical communication node lost in adaptive immune responses from COPD patients, possibly due to alterations in CD4+ T cell function or phenotype. Because T cells can modulate the innate immune response, and adaptive modulation is essential in tempering early innate immune responses[189,190], dysregulation of T cell behavior is a significant concern for immune defense, such as defense against respiratory infections that can exacerbate COPD symptoms. This study builds upon previous work reporting an impaired adaptive response in COPD[191] and emphasizes the importance of considering adaptive immune behavior in COPD progression and treatment. Our early results propose that targeting M-CSF signaling via CD4+ T cells could be a pivotal intervention in COPD. However, further research is required to understand better the role of M-CSF and how its loss contributes to pathogenic behaviors. Broadly, throughout this work, we illustrate using an integrative approach to unravel the significance of multiple disease-related cell alterations in immune responses at the cellular network level.

5.4 Methods

5.4.1 Isolation of PBMCs from human donors and cryopreservation

Fresh whole blood specimens were obtained from Visit 5 of SPIROMICS (ClinicalTrials.gov Identifier: NCT01969344), a longitudinal prospective cohort study[68]. Peripheral blood mononuclear cells (PBMCs) from 27 ever-smokers, with (n=13) and without (n=14) COPD, and 8 never-smokers enrolled in the bronchoscopy substudy[106,107] were isolated from whole blood by density centrifugation. Briefly, equal volumes of whole blood and wash buffer (phosphate-buffered saline (PBS), 2% fetal bovine serum (FBS), and 1 mM EDTA) were

combined, layered on Ficoll-Paque PLUS (Cytiva), and centrifuged at 2500 rpm for 30 minutes at room temperature in a swinging bucket rotor. PBMCs were collected from the buffy layer, washed with PBS, then counted with Trypan blue before resuspension in a cryopreservation medium (90% FBS and 10% DMSO). Cells were stored in a -80°C freezer until downstream experimentation. This study was conducted according to the principles of the Declaration of Helsinki. The institutional review board approved the protocols of all participating centers, and methods were carried out in accordance with the relevant guidelines and regulations. All participants were aware of the study's intent and provided written informed consent before any procedures.

5.4.2 Immune cell stimulation

PBMCs were thawed and cultured in 96-well U-bottom plates in R10 medium [RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum (FBS), HEPES, L-glutamine, and penicillin-streptomycin] at a density of 1 million cells/ well. After resting overnight at 37˚C, cell cultures were stimulated with R10 media alone or with anti-CD3/CD28 beads (1 bead:2 cells, Gibco), R848 (1 μg/mL, Invivogen), or LPS (1 μg/mL, Invivogen). Supernatant was collected from the stimulated cultures after 6 and 72 hours; independent cells were plated at each timepoint. Cells from participants with and without COPD were randomized across plates to reduce batch effects.

5.4.3 Luminex assays to quantify secreted cytokines

48-plex magnetic bead assays were purchased from Millipore kit HCYTA-60K-PX48 (included analytes: sCD40L EGF Eotaxin FGF-2 Flt-3 ligand Fractalkine G-CSF GM-CSF GROα IFNα2 IFNγ IL-1α IL-1β IL-1ra IL-2 IL-3 IL-4 IL-5 IL-6 IL-7 IL-8 IL-9 IL-10 IL-12

(p40) IL-12 (p70) IL-13 IL-15 IL-17A IL-17E/IL-25 IL-17F IL-18 IL-22 IL-27 IP-10 MCP-1 MCP-3 M-CSF MDC (CCL22) MIG MIP-1α MIP-1β PDGF-AA PDGF-AB/BB RANTES TGFα TNFα TNFβ VEGF-A) and were performed using Luminex FlexMAP 3D (Luminex Corporation, Austin, TX) technology according to the manufacturer's instructions. Any measurements above the upper limit of detection were set to the maximum detectable concentration of that analyte. Samples below the lower limit of detection were set to half of the lowest minimum detectable concentration across the standard curves of all analytes.

5.4.4 Flow cytometry

Supernatants from PBMCs cultured in R10 media (as outline in section 5.4.2) or 6 and 72 hours were harvested, then the cells were washed and stained for extracellular surface markers with fluorescently conjugated antibodies specific for CD45, CD3, CD4, CD8, CD56, CD20, CD14, CD16, HLA-DR, and CD11c (BioLegend) and a live/dead marker. Appropriate isotype-matched controls were used in all experiments. Cells were incubated in the dark with primary antibodies for 25 minutes at room temperature, followed by washing and fixation in 2% paraformaldehyde. Cells were analyzed on a BD FACSymphony A3 flow cytometer. Data were then analyzed with FlowJo v10.7.2.

5.4.5 Data-driven analyses

Partial least squares discriminant analysis (PLSDA): PLSDA was used to determine multivariate cytokine profiles that best distinguished between the responses of PBMCs to anti-CD3/CD28 beads, R848, and LPS[179]. PLSDA decomposes explanatory variables (i.e., cytokine measurements) into orthogonal linear combinations (PLS components) while simultaneously maximizing the covariance with the outcome variable (i.e., stimuli or cohort class). Each sample

is assigned a score and projected into the latent variable space (score plots). We used latent variable loadings (loadings plots) to identify cytokine profiles associated with different outcome variables of interest. All data were normalized with mean centering and variance scaling before analysis. *K*-fold cross-validation (k=5) was performed by iteratively excluding random subsets of data during model calibration and then using excluded data samples to test model predictions. An orthogonal signal correction was used to improve model interpretability (as outline in section 4.4.3).

We next performed feature reduction by calculating variable importance projection (VIP) scores for each cytokine and removing from the model cytokines with VIP scores < 1. Before the PLSDA analyses, we also removed any proteins with ≥ 70% of measurements below the lower detection limit across all samples. All remaining cytokine concentrations were normalized relative to their background secretion, defined by their paired negative control (R10 media) condition. This background-subtracted data were further adjusted for smoking status (current/ former) and experimental batch by fitting multiple linear regression models for each cytokine variable as the outcome and the two covariates (smoking status and batch) as the predictors. The residuals from these models were used as the adjusted dataset for the PLS analyses[192,193].

Random forest (RF) models: We used RF algorithms to determine the hierarchical importance of cytokine secretion events in classifying the responses of PBMCs from GOLD 0 and COPD participants. A classification RF algorithm was used to predict each class as a function of the VIP-selected cytokine measurements made from the corresponding PLSDA model (**Figure 5.2**). RF algorithms generate a set of classification trees based on a continual sampling of the experimental units and compounds. Then each observation is classified based on the majority votes from all the classification trees[194]. Gini's diversity index was chosen as the

split criterion. RF models were built with n=500 trees. Out-of-bag-evaluation errors for the resultant models were used to determine classification accuracy.

5.4.6 Bioinformatic analysis

Clustering: Hierarchical clustering of the VIP-selected signatures was generated with supervised complete linkage clustering using Pearson's correlation coefficient as the distance metric. Samples were colored by disease state (GOLD 0 vs. COPD).

Correlation networks: Protein correlation networks were constructed separately for GOLD 0 and COPD participants using pairwise Pearson's correlation coefficients between protein expression in the VIP-selected signatures (**Figure 5.2**). Edge color and thickness correspond to coefficient value and statistical significance, respectively. Only significant correlations ($p < 0.05$) are shown. Node size is proportional to its degree of connectedness.

5.4.7 Software summary

Hierarchical clustering and correlation networks were completed using MATLAB (v2017b, MathWorks, Natick, MA). We generated PLSDA models and ROC curves using the PLS toolbox available in MATLAB (v8.2.1, Eigenvector, Mason, WA). Random forest analyses were completed in R using the randomForest package (v4.1.2, R Core Team, Vienna, Austria). All standard statistics were performed using Prism version 9 (GraphPad Software, San Diego, CA).

Chapter 6 Discussion

In this thesis, we leveraged network-based modeling approaches to improve our understanding of COPD and overcome challenges associated with traditional single-target methodologies. Specifically, we presented novel insights into potential biological mechanisms contributing to rapid spirometric progression (**Chapter 2** and **Chapter 3**), age-dependent emphysema development (**Chapter 4**), and altered immune cell communication networks (**Chapter 5**) in COPD. A discussion of the scientific contributions, broader impacts, and future work for these three investigations are detailed in the sections below, along with comments on the overall work's limitations, future directions, and conclusions.

6.1 Early complement pathway changes predict rapid FEV_1 decline in COPD; combined biomarker signatures lack the same predictive accuracy (Chapters 2-3)

6.1.1 Summary of findings

For the first time, data-driven modeling approaches were applied to integrated baseline proteomic datasets from the blood and lung to investigate early cross-tissue compartment mechanisms contributing to accelerated spirometric progression, as measured by annualized declines in FEV_1. This work identified an integrated signature with 52 proteins capable of predicting the risk of a COPD patient sustaining declines in lung function ≥70 mL/year with high accuracy (> 90% accuracy). Further investigation of this signature identified potential prognostic biomarkers and biological pathways contributing to accelerated FEV_1 decline. In particular, we found that individuals who experience losses in lung function ≥70 mL/year exhibit unique

alterations in elements of the complement cascade ("complement profiles") at baseline. Using an independent cohort, we successfully replicated the complement profiles but could not validate the biomarker signatures.

6.1.2 Scientific contributions

By providing evidence that alterations in complement cascade proteins precede accelerated FEV_1 decline, we extend previous associations between this complex and evolutionarily conserved system and cross-sectional outcomes in COPD[113–121]. Our findings suggest for the first time that early complement-related alterations may be a key contributor to severe lung function decline. Moreover, our inability to validate prognostic biomarker signatures may suggest that alterations in several pathways or pathway constituents can drive a singular phenotypic outcome in patients with COPD. Moreover, various factors, such as mucus blockage, small airway problems, and emphysema, can contribute to airflow obstruction. It's improbable that a single pathway, like the complement pathway, is equally implicated in all these processes. This underscores the diverse nature of factors influencing COPD outcomes. Together, our results highlight the potential importance of shifting focus from biomarker identification towards pathway dysregulation in future analyses of spirometric progression and other complex COPD outcomes.

6.1.3 The complement system and its role in COPD progression

One key outcome of this work is the link identified between early complement alterations and later lung function decline. The precise role of the complement system in driving COPD progression remains uncertain; however, it is plausible that genetic variants mediate complement protein levels. Sun et al. demonstrated within the SPIROMICS and COPDGene cohorts that the

decreased C3 levels observed in individuals with COPD and emphysema can be attributed to protein quantitative trait loci (pQTL) and expression QTL (eQTL) single nucleotide polymorphisms (SNPs) [119]. In murine models, complement proteins specifically appear to exert a pivotal influence on smoking-induced emphysema development by modulating both innate and adaptive immune responses. Mice deficient in C3, the central component of the complement system, exhibit protection against emphysema development[122]. Further investigations suggested that C3 contributes to emphysema by modifying dendritic cell-directed Th17 inflammation[122]. Complementary studies with C1q, a subcomponent of the C1 complex of the classical complement pathway, support the importance of complement proteins in directing pro-emphysema Th17 inflammatory profiles. These studies demonstrate that cigarette smoke-mediated loss of C1q contributes to emphysema development through simultaneous induction of Th17 cells and reduced differentiation of peripheral tolerance (T regulatory cells)[143].

The paradoxical responses that C3 and C1q elicit in mice help to explain the diverse pathogenic and protective roles which have been suggested for complement components in human COPD studies. For instance, increases in C4a:C4 ratios (a marker of lectin pathway activation)[142] and reductions in complement regulator CD46[195] are observed in COPD patients compared to controls, suggesting that overactive complement activation may be pathogenic. Conversely, localized C3 reserves in airway epithelial cells offer protective effects by aiding host defense against stress-associated death[124]. Some discrepancies in reported trends may be associated with specimen type and study demographics, as complement proteins measured from blood and lung specimens often have inverse associations with COPD outcomes[196,197], and their concentrations are significantly impacted by sex and age[198]. However, it is also plausible that, like many complex biological mechanisms, the contributions of the complement pathway to

disease pathology are dynamic, varying over time and space.

The complex involvement of complement pathways in COPD progression supports our choice of a multivariate approach to model complement profiles. Multivariate models can capture global trends in numerous pathway constituents and help overcome a lack of sensitivity or specificity afforded by an individual protein. Hence, future studies investigating the complement system in COPD may benefit from implementing more dynamic modeling approaches to help deconvolve the role of complement components in progression.

6.1.4 The complexity of COPD patients may limit the availability of succinct prognostic biomarkers

The second notable outcome of this study was our inability to validate the biomarker signatures we identified in our discovery cohort. Although unfortunate, this finding draws attention to larger challenges facing biomarker identification in COPD, including limited replication and clinical interpretation. Predictive biomarkers of FEV_1 decline in COPD historically struggle with a lack of replication, even when cohorts share many common features (ex. SPIROMICS and COPDGene)[60]. Additionally, when replication is successful, individual proteins often only add minor prognostic value over easily obtainable clinical variables[93]. For example, of the top prognostic biomarkers studied in COPD (CC16, SP-D, sRAGE, CRP, fibrinogen, and IL-6) each marker explained only 1-8% of the variance in FEV_1 decline, compared to the 34% explained by clinical variables[59,60,93,199]. This observation stresses the importance for future studies to focus more on the predictive value added by a molecular biomarker (R^2) rather than just a statistically significant association (P value).

Similar issues have been reported in studies of alternative COPD outcomes, including exacerbations. A study evaluating the potential of n=90 protein biomarkers to predict

exacerbation frequency in COPDGene and SPIROMICS reported successful identification of significant markers in each cohort but poor replication between studies[85]. Issues with lack of replication in this and other COPD studies likely have to do with the heterogenous clinical and pathophysiological nature of COPD as well as variability in the definitions of clinical outcomes across studies. To date, fibrinogen is the only proteomic marker successfully replicated and approved by the FDA in COPD, yet not yet the target of any specific therapeutic interventions. A combined analysis of five cohorts showed combining plasma fibrinogen with a clinical history of exacerbations increased the ability to predict the occurrence of future exacerbation events[200]. Still, fibrinogen lacks the sensitivity and specificity to be used on an individual basis[93].

The use of multiple biomarkers may improve predictive performance. A combination of CC16, SP-D, CRP, sRAGE, and fibrinogen increased the sensitivity of predicting cross-sectional FEV_1 from 2-10% for an individual marker to 24%[58]. In analyses with FEV_1 decline, however, improvements were less pronounced, with combinatorial biomarkers accounting for only 7% of variance[58]. This trend suggests that no single protein, or even their combination may be capable of predicting progression in a heterogenous disease like COPD. As such, it may be prudent for future studies either to focus analyses on analyzing more targeted COPD phenotypes or to shift investigation approaches toward uncovering dysregulated pathways as a basis for biomarker identification instead of focusing on individual elements.

6.1.5 Future work

Subsequent work requires follow-up investigations that better detail the role of the complement system in the lung. Because our study utilizes complement protein measurements from plasma samples, exploring the identified complement profiles in lung-derived specimens is essential. Although the complement profiles in the lung may not be identical to the systemic

profiles we reported, we are confident that future studies can successfully evaluate complement profiles in the lung, as complement proteins have been quantified in BAL, sputum, and lung tissue specimens from individuals with COPD[122,142,196]. To our knowledge, BAL samples collected during the SPIROMICS bronchoscopy sub-studies still exist. It could be especially useful to measure complement proteins from these BAL specimens to explore, in matched samples, how the identified blood complement profiles extrapolate into the lung compartment.

Lung-derived complement profiles would ideally be explored on the same platform; however, SomaLogic assays are not validated for use in sputum or BAL samples. Hence, follow-up studies may require alternative assays. Luminex assays have complement-specific kits available that allow for the measurement of a limited number of complement proteins and their cleavage products (kits: HCMP1MAG-19K and HCMP2MAG-19K). Although, alternative functional assays may be particularly advantageous to generate more granular insights into the patterns of complement activation, including pathway-specific responses (classical, lectin, or alternative). Together, these follow-up experiments would provide a clearer view of complement pathway dysregulations driving progression in the tissue of interest, the lung, and whether measurements from the blood accurately reflect the local behavior.

6.2 Distinct age-dependent biological processes contribute to COPD progression (Chapter 4)

6.2.1 Summary of findings

A data-driven investigation of the age-dependent progression of CT-indicated airway abnormalities uncovered unique proteomic signatures associated with the progression of emphysema and small airway disease in younger (≤ 55 year-old) and older (> 65 year-old) GOLD 0-2 ever-smokers. Investigation of the signatures suggested that age significantly impacts

the biological pathways contributing to smoking-associated emphysema progression. Specifically, our results suggested that inflammatory and apoptotic pathways primarily contribute to accelerated emphysema development in ever-smokers ≤ 55 years old. In contrast, matrix remodeling is the primary mechanism driving emphysema in older individuals (> 65).

6.2.2 Scientific contributions

Recent advancements in understanding COPD have hypothesized that the field may gain unique insights into early disease pathogenesis ("early COPD") by studying younger populations. In support of this hypothesis, this is the first investigation to present evidence that different protein signatures are associated with COPD progression depending on an individual's age. More importantly, these signatures implicate differential age-dependent pathologic mechanisms contributing to emphysema development in early (younger) vs late (older) onset disease. If validated, these signatures can inform novel therapeutic targets of early COPD.

6.2.3 Investigations of younger ever-smokers can inform novel mechanisms in early COPD

Results from our preliminary evaluations suggest that progression in GOLD 0-2 ever-smokers results from age-dependent pathologic mechanisms. These findings have significant implications for justifying the study of early COPD. The importance of studying early COPD resulted from a collection of studies evidencing that in ever-smokers with ≥ 10 pack-year smoking history, respiratory symptoms originating by 30-50 years of age were associated with accelerated lung function decline evident as early as 43 years of age[147–150]. Because most COPD studies have focused on individuals ≥ 60 years old with mild disease (GOLD 0-1)[146], current literature lacks robust characterizations of early COPD pathogenesis, which, as past studies suggest[112], likely occurs much younger.

Early COPD patients may define a subpopulation with unique molecular alterations that drive disease progression toward more severe illness than older individuals with mild-slowly progressing disease. Though limited, studies suggest that age impacts both molecular and clinical profiles in COPD[201,202]. Moreover, individuals who fit the criteria of early COPD[112] are at increased risk of acute respiratory hospitalizations and early death[203]. These findings highlight potentially distinct underlying mechanisms between early COPD and later-onset disease and expose gaps in our understanding of initial disease pathogenesis within existing cohorts. Taken together, these insights may offer explanations for the limited success of targeted therapies, including biologics targeting IL-5 and IL-33, in COPD clinical trials[66,204]. Our work supports this emerging hypothesis.

Although it requires extensive validation, our results suggest that apoptotic pathways may be a key driver in early emphysema development. This finding is supported by evidence that alveolar matrix destruction by the combination of inflammation and excessive proteolysis has failed to account fully for the mechanisms behind eradicating septal structures[205]. Hence, apoptotic pathways may be a promising target to halt early emphysema development in patients with COPD. Pharmacological inhibitors of apoptosis exist and are used to treat Parkinson's disease, cancer, Rheumatoid arthritis, and Crohn's disease[206]. However, currently, there are no FDA-approved uses for pulmonary conditions.

Looking forward, investigations of younger populations with less severe airflow limitation hinge on developing clinical biomarkers more sensitive to early airway abnormalities than traditional spirometry. In COPD, small airway abnormalities are commonly thought to precede emphysema development[14,207]. However, small airways contribute little to overall airflow resistance, so extensive loss of small airways may occur before the appearance of any

symptoms[208,209]. Additionally, spirometry maneuvers require participant cooperation, which can introduce substantial variability to measrements[210,211]. Promising alternative biomarkers for early COPD studies which afford higher reproducibility and sensitivity include parametric response mapping (PRM)[12], as used in our analysis (detailed in **Chapter 4**), and impulse oscillometry (IOS).

IOS is a noninvasive method that uses sound waves to measure respiratory mechanics. By generating small pulses of air pressure (impulses), IOS creates pressure oscillations within the lungs. As the impulses travel through the airways, they encounter resistance from various parts of the respiratory system due to narrowed or inflamed airways, mucus build-up, or other obstructions. An oscilloscope then measures the response of the respiratory system by recording changes in pressure and flow caused by the impulses and analyzing the resulting waveforms. From these waveforms, IOS calculates resistance (characterizes airway constriction or obstruction) and reactance (characterizes airway elasticity and stiffness) metrics[212]. One advantage of IOS is that testing is performed during quiet tidal breathing, which contrasts other measures of spirometry, which are effort dependent. Because the diagnosis of COPD relies on spirometric definitions, it is challenging to compare IOS. However, since IOS is better for detecting small airway disease in asthma and post-environmental exposures, even where spirometry is normal, the presumption is that IOS should be more sensitive in identifying early COPD[211]. In support of this hypothesis, in COPD, IOS has been reported to better correlate with small airway structures[213–215] and symptoms burden[216] when compared to spirometry. In the future, studying younger populations with more sensitive clinical metrics may aid in identifying early mechanisms contributing to COPD development that can inform novel preventative treatment strategies.

6.2.4 Future work

Although the findings from this study were encouraging, we were limited to using available data from the previously enrolled longitudinal cohort, COPDGene. Because COPDGene's inclusion criteria included individuals ages 45-80, our "young" population for this analysis (mean age: 50 ± 3) was not optimally enriched to study early COPD as per the definition proposed by Martinez et al.[112] (ages 30-50). As such, exploring similar trends in newly enrolling cohorts such as SOURCE[177], which enrolls individuals ages 30-55, is essential.

Additionally, our approach was more successful in identifying age-dependent processes related to emphysema progression than for fSAD. This might be because emphysema represents a final pathway characterized by irreversible lung damage, while fSAD can develop de novo but also progress into emphysema. This dual nature makes it more challenging to model definitively. Therefore, future work may have more success by employing more sophisticated voxel-based calculations of progression, as proposed by Labaki et al.[207], rather than the simple delta used throughout this study.

Looking forward, it will also be imperative to explore proteomic signatures in the lung, as the current profiles are from plasma samples. Finally, universal definitions for early COPD and rapid progression are needed to ensure consistency across future investigations.

6.3 Analyzing cytokine-chemokine networks from stimulated PBMCs reveals novel immune network changes in COPD (Chapter 5)

6.3.1 Summary of findings

To better recapitulate the diverse cell-cell interactions that collectively drive immune system behavior, we stimulated peripheral blood mononuclear cells (PBMCs) from COPD and

ever-smoking control (GOLD 0) donors with three well-defined adaptive and innate immune challenges: anti-CD3/CD28 to stimulate T cells polyclonally, and to activate innate immune cells via TLR4 and TLR7/TLR8 using LPS and R848, respectively. Multivariate explorations of temporal cytokine production following PBMC stimulation highlighted a critical deficiency in the adaptive immune response in PBMCs from individuals with COPD. Using data-driven tools, including random forest algorithms and correlation networks, we identified that a loss of early M-CSF and IL-13 production, likely by CD4+ T cells, led to a significant loss of downstream signaling diversity in immune cell networks from PBMCs of COPD donors. Additionally, PBMCs from individuals with COPD displayed a diminished production of Type 2 (T2) related cytokines in response to both innate and adaptive stimuli. Collectively, our findings indicate a modified adaptive immune response in individuals with COPD, which may be the result of altered T cell phenotypes in these patients.

6.3.2 Scientific contributions

By proposing critical early regulators of immune signaling networks in response to adaptive stimuli that influence the impaired immune responses observed in COPD, this study builds upon previous work reporting functional alterations in immune cells and their cytokine production in patients with COPD[20,43–47]. We specifically highlight a novel role for CD4+ T cells as critical regulators in COPD pathology through their loss of production of M-CSF and IL-13. Our results emphasize the importance of considering adaptive immune behavior in COPD progression and treatment and, if validated, propose that targeting CD4+ T cell signaling via M-CSF or IL-13 could be a possible intervention in COPD.

6.3.3 The adaptive immune response is modified in individuals with COPD

Significant disparities in PBMC secretion profiles between COPD and GOLD 0 donors following immune stimulation underscore the broad-reaching impact of COPD and its associated airway obstruction on the immune system. This impact is evidenced by discernible alterations occurring in various immune components, including cells and cytokines, even in peripheral areas beyond the lungs. Notably, our findings indicate that an early deficiency of M-CSF secretion from PBMCs following T cell activation plays a significant role in shaping an overall modified adaptive immune response in COPD. Biologically, M-CSF facilitates the differentiation and activation of monocytes and macrophages. A network analysis of our *in vitro* system revealed critical connections (i.e., correlations) between early M-CSF signaling and later induction of MCP-1 and MCP-3 within GOLD 0 participants. These connections were entirely lost in networks modeled from secretions of PBMCs from COPD donors. Because MCP-1 and MCP-3 play pivotal roles in driving the migration of monocytes and other immune cells to sites of inflammation and injury, the breakdown of this connectivity observed in COPD may collectively suggest the potential for impaired pathogen clearance due to altered migration or activation of peripheral monocytes.

Endothelial and stromal cells are the primary sources of M-CSF during homeostasis, but under inflammatory conditions, M-CSF is also produced by activated monocytes and macrophages[217,218]. Intriguingly, our exploration hints at M-CSF production by CD4+ T-cells. Although previous studies have reported M-CSF expression in CD4+ T cells cultured *in vitro*[219–221], T cells are generally not considered to be a biologically relevant source of M-CSF[218]. Given the uncertainty of T cells as reliable producers of M-CSF upon anti-CD3/C28 stimulation, it may instead be plausible that other known T cell derivatives, like IL-13, are responsible for driving distinct cellular secretions in monocytes and other innate immune cells within the PBMC culture

system. For instance, monocyte exposure to T1 (e.g., IFN-γ) signals can incite M1 polarization characterized by proinflammatory secretions, including IL-1β, IL-6, TNF-α, and IL-12 production. In contrast, exposure to other immune factors, such as IL-13 (a second critical cytokine identified in our analysis, decreased in COPD participants), favors an alternative activation profile[222]. Collectively, these variations in the function and phenotype of adaptive and innate immune cells shape the immune response.

Our findings suggest an inclination of the adaptive immune response in COPD patients toward a Th1-dominant phenotype. This finding is consistent with earlier studies noting an elevated proportion of CD4+ IFN-γ+ T cells in the circulation of COPD patients compared to controls[188]. Along with enhanced production of Th1-related factors, our results suggest that COPD patients experience notable reduction (though not complete elimination) of cytokines associated with a Th2 response. This phenomenon has been documented previously, wherein CD4+ T cells from individuals with COPD show an increased production of Th1-related cytokine IFN-γ and a decreased generation of Th2-related cytokine IL-4[185,223]. Beyond CD4+ T cells, CD8+ T cells from individuals with COPD exhibit altered cytokine production[188,224–232], which, consistent with our findings, predominantly aligns with a Type 1 immune response, marked by increased IFN-γ and TNF-α production[188,223,225–227]. However, it is worth noting that other inflammatory profiles, including Th2 and Th17 responses, have also been reported in COPD and are often characteristic of patients with distinct disease endotypes[71,72].

The origins of the observed differences in inflammatory profiles, whether stemming from altered activation due to changes in receptor expression or genetic/ epigenetic modifications in immune cells from COPD patients, remain uncertain. However, given that homeostasis hinges on finely balanced immune responses, skewed Th1/Th2 responses, as delineated in our study, can

lead to less effective or potentially harmful T cell responses[233,234]. Consequently, it is important to work towards understanding the specific alterations that drive imbalanced responses in COPD. Our research, while necessitating further investigation, represents a significant stride toward this goal by identifying early signaling events that propel pathogenic immune behaviors in COPD.

6.3.4 Future work

The outcomes of this study underscore the value of adopting a multivariate approach to biological research to help overcome clinical variability and uncover disruptive cell-cell signaling patterns in diseases involving chronic inflammation, such as COPD. Results from this investigation implicate M-CSF and IL-13 as pivotal communication hubs in the adaptive immune response that are compromised in COPD. To formulate hypotheses regarding the cell types responsible for these secretions we employed straightforward correlations between cell subsets and cytokine concentrations. Although these assessments implicate CD4+ T cells as the potential cellular source of both M-CSF and IL-13, follow-up experiments utilizing intracellular flow cytometry are required to conclusively identify the cell types responsible for these critical secretion events. Regrettably, the original experiment did not entail the preservation of cell aliquots, necessitating new, unmatched PBMC samples for subsequent studies. While not ideal, we believe that the randomized sample generation and strong class representation used in our in the original workflow should have allowed for the identification of robust signatures which are applicable in new patient populations.

Once the cellular source has been pinpointed, further experiments aimed at validating the proposed protein networks should be conducted *in vitro*. One possible option to explore this question could involve implementing combinatorial antibody treatments or media supplementation to modulate the concentration of key cytokines in PBMC culture system to

specified ranges identified our original study. After target factor alteration, subsequent cytokine measurements and flow cytometry can be used to characterize the ensuing phenotype. For example, based on the findings in our original work, PBMC cultures from COPD donors can be individually supplemented with M-CSF or IL-13 to explore whether adaptive immune signaling profiles can be restored closer to those of the control group (GOLD 0).

In the event that supplementation alone fails to reinstate cytokine secretion profiles from COPD participants, alternative investigations could be explored. Specifically, given the skewedness observed in inflammatory profiles across COPD and GOLD 0 donors (Th1 vs. Th2), it could prove insightful to delve into potential epigenetic modifications within immune cells from COPD patients that may be favoring this Th1 response. Alterations in chromatin accessibility can reflect epigenetic modifications that influence patterns of gene regulation. Techniques such as assay for transposase-accessible chromatin with sequencing (ATAC-Seq) can be employed to examine how these epigenetic modifications impact the expression patterns of genes, helping to draw conclusions about the sources of the proteomic modifications observed in our system.

6.4 Limitations

The limitations of the work presented in this thesis primarily relate to data availability and constraints inherent to our data-driven modeling framework. The cohorts used throughout this analysis were comprised of predominantly non-Hispanic white (NHW) participants (SPIROMICS: 80% NHW, COPDGene: 70% NHW)[235,236]. Given the substantial racial-specific differences in COPD susceptibility[237], our lack of participant diversity limits our insights into the utility of our findings in other populations, particularly African-Americans, who compared to NHWs have worse outcomes[238,239]. NHW populations are often overrepresented in COPD

research, underscoring the necessity for more diverse cohorts and analyses that recognize race as a determinant of pharmacological benefit[240,241].

Our sample sizes and ability for model validation were often restricted due to the availability of matched molecular datasets and longitudinal clinical variables. For instance, in the SPIROMICS cohort, 40% of participants with available protein measurements lacked longitudinal spirometry metrics required for classification. Additionally, in several instances, no external cohort was available for validation; however, we took steps to improve the generalizability of our models through resampling and cross-validation. Validating machine-learning models with an independent cohort is essential to evaluating model reproducibility, especially for biomarker signatures. Hence, in our analyses that lack proper validation, the protein signatures presented are likely better suited for pathway insights than indications of clinical biomarkers.

Percentiles determined the outcomes for our progression models (in **Chapter 2** - **Chapter 4**); while this approach has been used previously, it is susceptible to variability depending on the demographics of the base cohort. Past studies of spirometric progression using percentile cut-offs to define patterns FEV_1 decline have reported wide ranges of mean FEV_1 declines in "rapid progressors" varying from annualized rates of -50 to -100 mL[97,103]. Consistency across future studies requires establishing definitions for high-priority outcomes, including rapid progression (as measured by spirometry and other CT-based outcomes) and early COPD. Definitions for rapid FEV_1 decline[109] and early COPD[112] have been proposed but are not yet universally implemented.

Lastly, while our data-driven approaches allow for insights into co-varying factors associated with disease phenotypes, our models, like many other statistical techniques, are

limited to providing insights into factors that are correlated with outcomes of interest. Because correlation does not always reflect causation, there is no guarantee that the identified associations are biologically meaningful or directly contribute to progression. However, our models still offer valuable insights and novel integration of multi-compartment data.

6.5 Future directions for systems-focused research in COPD

Improving on three key elements can help to improve future systems-level investigations into COPD pathobiology: data integration, population selection, and biological interpretation. Integration of datasets generated from different molecular levels and tissue compartments is the first step in improving network analyses in COPD. These global alterations spanning multiple biological regions can improve predictive performance[57,93] and uncover intricate mechanistic insights into the lung's disease-related biology. Network findings may also improve our understanding of how local pathologic changes affect systemic behavior and inform biomarkers or pathways measurable in the blood that can reliably reflect pulmonary alterations.

The selection of the model population is pivotal in establishing robust models that yield diverse clinical insights. Notably, future analyses should strive to model distinct subpopulations ("phenotypes") within COPD, incorporating individuals with demographic diversity. As discussed, the heterogeneous nature of COPD patients has posed challenges to achieving consistent clinical insights, particularly in terms of reproducibility. However, within this thesis, significant success was achieved in generating high-performing models when exploring protein associations within well-defined subgroups (e.g., age-dependent emphysema progression). Consequently, in future models, insights can likely be improved by focusing evaluations on participants exhibiting specific COPD phenotypes or endotypes. However, definitions for phenotypes are first needed to ensure consistency and reproducibility across studies. In addition,

because data-driven models, like most mathematical or statistical approaches, are only capable of identifying trends in the datasets from which they learn, future studies need to make efforts to model populations with diverse ethnic and sex representations or create independent analyses based on these important demographic factors. Overall, eliminating disparities in medicine and research require enrolling more diverse COPD cohorts (based on race, ethnicity, sex, gender, etc.) and explicitly discussing the demographic limitations of insights generated from analyses of skewed populations.

Lastly, our current data-driven modeling approach identifies correlations without considering underlying biological connectivity between molecular elements (i.e., functional networks). In future work, integrating information about network connectivity can help ensure that model insights have a basis for biological interpretation that can better inform therapeutic targets. Methodologies with these capabilities, such as Elastic-Net based Prognosis Prediction (ENCAPP)[242], have been developed but have yet to be used in COPD. ENCAPP provides insights beyond the correlation of co-varying proteins by mapping protein measurements to known protein-protein interaction networks from prior knowledge databases (ex., STRING). The weighted contributions of individual proteins to small protein networks are reflected as "module scores," which are then used as input to predictive models. This mapping process creates an inherent functional basis for the model input, thereby improving interpretability. This approach has been used in predicting outcomes in other complex immunological conditions, including organ rejection and cancer[242,243], and as such, may hold great promise for helping to deconvolve COPD pathobiology.

6.6 Conclusions

In summary, our study has successfully revealed proteomic signatures capable of

distinguishing various subpopulations within COPD. These identified signatures hold biological significance and serve as initial reference points for generating novel hypotheses concerning disease mechanisms. They also serve as catalysts for exploring novel scientific directions in follow-up experiments. Our data-driven methods for signature identification hold potential for identifying disrupted signaling pathways linked to COPD outcomes. If robustly validated across distinct cohorts, these findings could offer significant clinical value. Moving forward, we aim to further validate our data-driven findings in human or murine COPD models, and ultimately construct mechanistic models of critical pathogenic pathways and binding events. This would enable quantitative investigations of system perturbations and mechanistic theories *in silico*, thus enhancing our understanding of COPD pathology.

Appendices

Appendix A: Supplement to: A Blood and Bronchoalveolar Lavage Protein Signature of FEV_1 Decline in Smoking-Associated COPD

This data provided in Appendix A was originally published as a supplementary data file accompanying scientific article:

> DiLillo KM, Norman KC, Freeman CM, Christenson SA, Alexis NE, Anderson WH, Barjaktarevic IZ, Barr RG, Comellas AP, Bleecker ER, Boucher RC, Couper DJ, Criner GJ, Doerschuk CM, Wells JM, Han MK, Hoffman EA, Hansel NN, Hastie AT, Kaner RJ, Krishnan JA, Labaki WW, Martinez FJ, Meyers DA, O'Neal WK, Ortega VE, Paine R, Peters SP, Woodruff PG, Cooper CB, Bowler RP, Curtis JL, Arnold KD. (2023). A blood and bronchoalveolar lavage protein signature of rapid FEV_1 decline in smoking-associated COPD. *Scientific Reports*, 13(1):8228. https://doi.org/10.1038/s41598-023-32216-0

Changes made to the original document are mainly cosmetic to adhere to the format of this document.

Supplemental Table A.1. Baseline demographics of COPD cases and TEPPS reference group

	TEPPS[§] (N=40)	Greater Decliners[**] (N=14)	Lesser Decliners (N=31)	P-Value
Age*	59.3 (± 8.67)	64.2 (± 6.24)	63.1 (± 8.41)	0.068
Currently Smoking*	13 (32.5%)	5 (35.7%)	10 (32.3%)	0.97
BMI*	29.3 (± 5.01)	27.9 (± 3.67)	27.8 (± 5.43)	0.41
Sex (Male)	**17 (42.5%)**	**12 (85.7%)**	**17 (54.8%)**	**0.02**†
Race (White/Other)	29/11 (72.5%)	12/2 (85.7%)	25/6 (80.6%)	0.52
ICS use* (yes)	**2 (5.0%)**	**3 (21.4%)**	**14 (45.2%)**	**0.0003**‡
FEV_1* (% predicted)	**100.4 (± 13.1)**	**84.2 (± 13.1)**	**71.1 (± 17.7)**	**< 0.0001**†‡
FEV_1/FVC*	**0.78 (± 0.04)**	**0.60 (± 0.08)**	**0.57(± 0.10)**	**< 0.0001**†‡
FEV_1* (L)	**2.89 (± 0.69)**	**2.63 (± 0.60)**	**2.11 (± 0.66)**	**< 0.0001**‡
Visit 5 FEV_1 (L)	**2.74 (± 0.77)**	**1.97 (± 0.67)**	**1.93 (± 0.68)**	**< 0.0001**†‡
Time from baseline to Visit 5 (yrs.)	6.18 (± 0.93)	6.25 (± 0.76)	6.33 (± 0.91)	0.77
Time from baseline to bronchoscopy (months)	**12.5 (±11.1)**	**20.4 (± 10.0)**	**20.3 (± 12.5)**	**0.0107**‡
ΔFEV_1 (mL/yr.)	**-25.7 (± 47.5)**	**-104.6 (± 32.0)**	**-28.8 (±21.5)**	**< 0.0001**†

One-way ANOVA with Tukey's post hoc test or chi-squared test were used to determine significance. Bold values denote significant differences between groups ($p < 0.05$).

* Demographic information from baseline visit (Visit 1)

† Significant difference between TEPPS and greater decliners

‡ Significant difference between TEPPS and lesser decliners

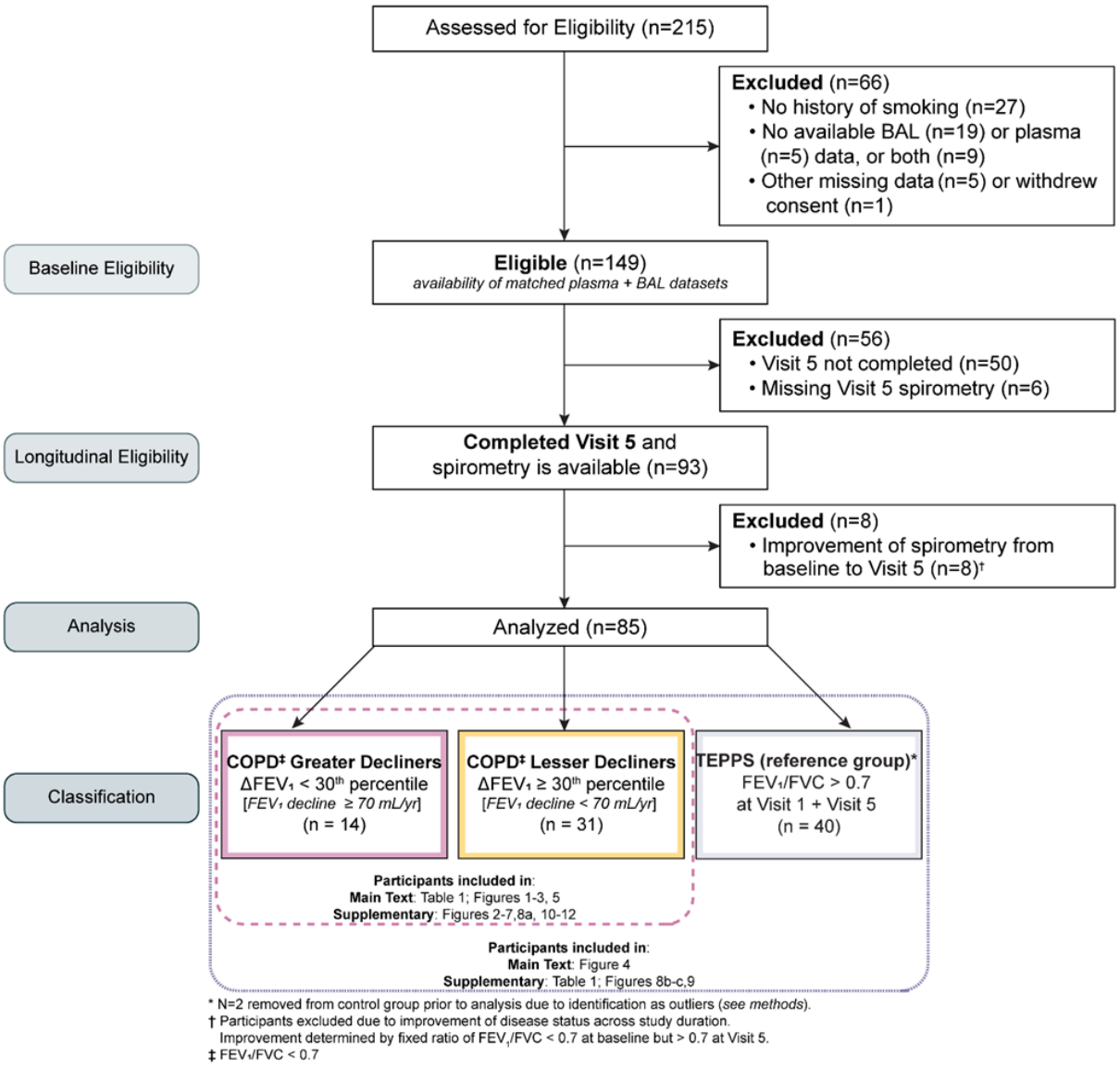
§ Tobacco-exposed people with preserved spirometry (TEPPS); n=2 removed prior to analysis due to identification as outliers (see pg. 34)

** Defined as annualized decline in $FEV_1 \geq 70$ mL/year (see pg. 34)

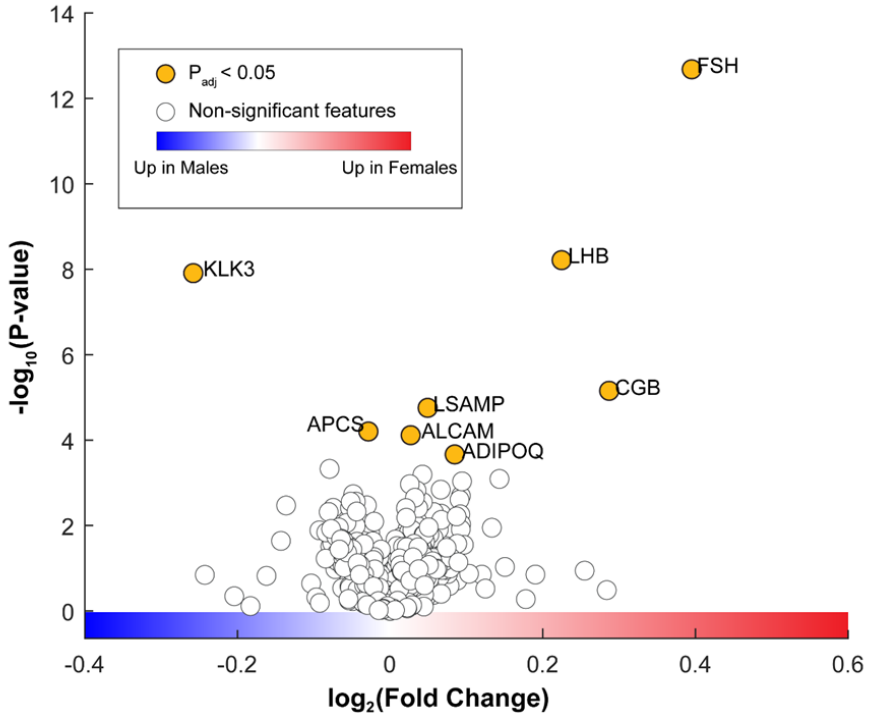
Supplemental Table A.2. Demographics of study group compared to full SPIROMICS bronchoscopy sub-study cohort

	Study Group (N=85)	Full Bronchoscopy Cohort (N=188)	P-Value
Age*	61.5 (± 8.41)	60.1 (± 8.89)	0.24
Currently Smoking*	28 (32.9%)	81 (43.1%)	0.11
BMI*	28.5 (± 5.01)	28.5 (± 4.97)	0.99
Sex (Male)	46 (54.1%)	104 (55.3%)	0.85
Race (White/Other)	66/19 (77.6%)	131/57 (69.7%)	0.17
ICS use* (yes)	19 (22.4%)	37 (19.7%)	0.61
FEV_1* (% predicted)	87.0 (± 19.9)	88.1 (± 19.1)	0.67
FEV_1/FVC*	0.67 (± 0.12)	0.68 (± 0.12)	0.75
FEV_1* (L)	2.56 (± 0.75)	2.59 (± 0.76)	0.75
Visit 5 FEV_1 (L)	2.32 (± 0.82)	2.33 (± 0.80)	0.87
Time from baseline to Visit 5 (yrs.)	6.25 (± 0.89)	6.28 (± 0.97)	0.81
Time from baseline to bronchoscopy (months)	16.6 (± 12.0)	15.2 (± 12.4)	0.37
ΔFEV_1 (mL/yr.)	-39.8 (± 47.0)	-34.4 (± 51.7)	0.46

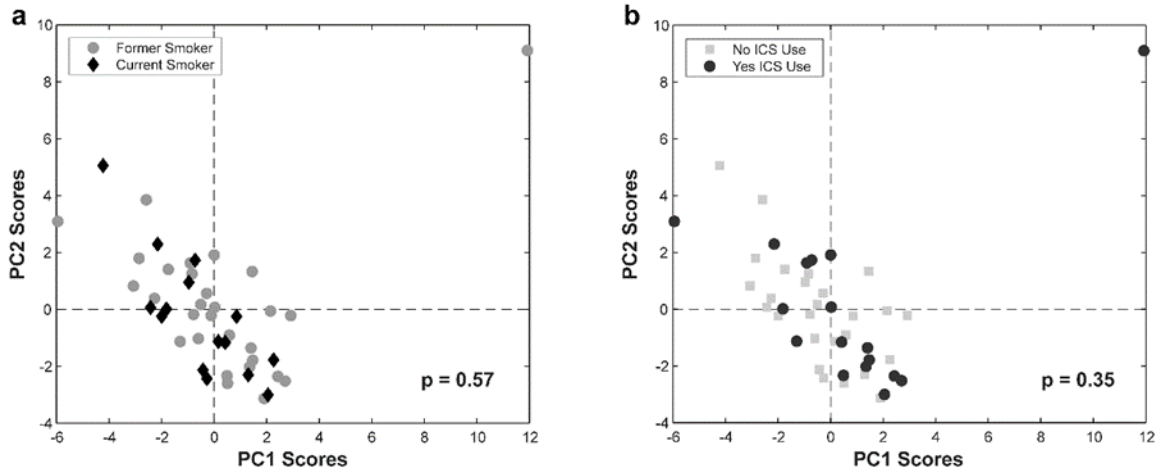
Unpaired two-tailed t-test or chi-squared test were used to determine significance.



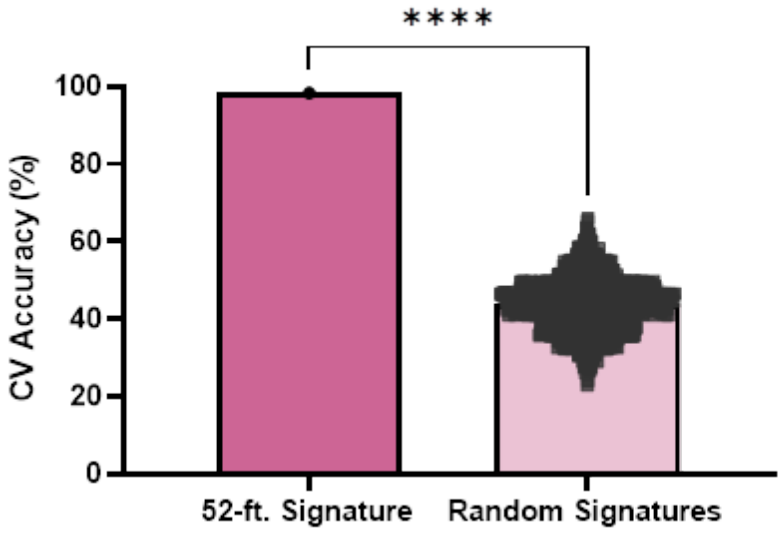
Supplemental Figure A.1. CONSORT diagram
Schematic illustrating the inclusion criteria and patient breakdown of participants with a history of smoking from SPIROMICS I bronchoscopy sub-study (n=215) included in analysis.



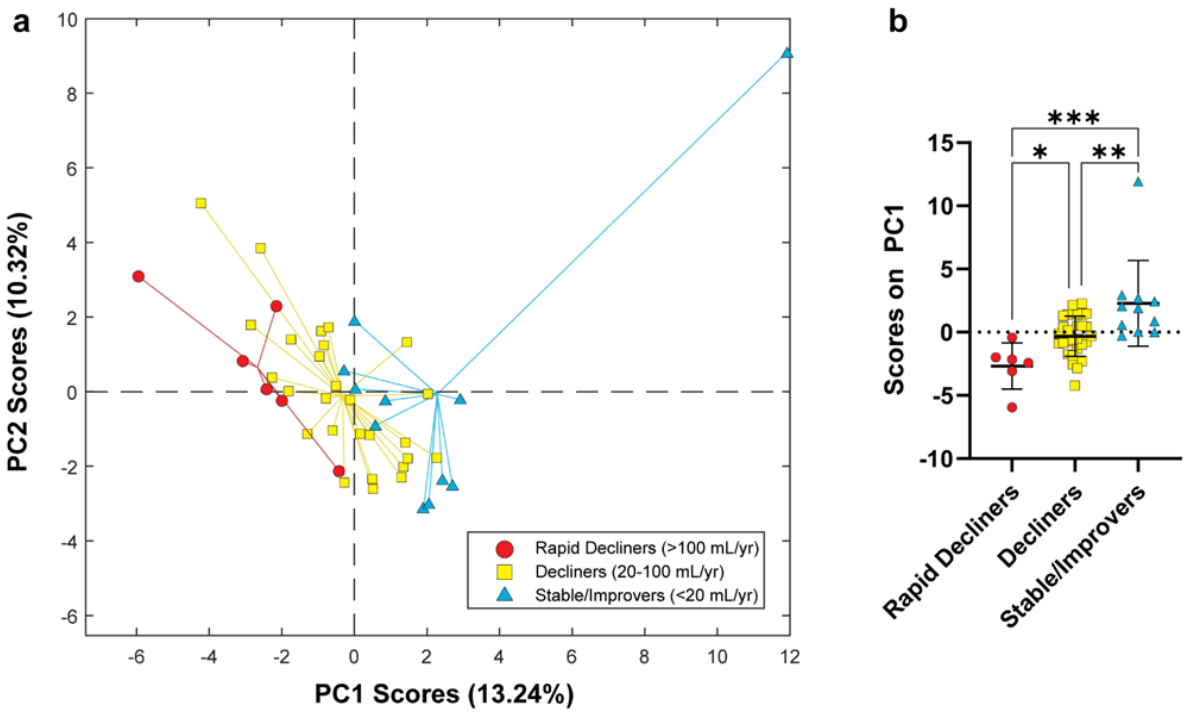
Supplemental Figure A.2. Differential expression of cross-compartment proteins between sexes.
Volcano plot of blood and BAL proteins (two-sampled, two tailed t-test). Yellow markers represent proteins which have p-value < 0.05 after correction for multiple comparison using Benjamini-Hochberg false discovery rate (FDR) ($\alpha = 0.05$).



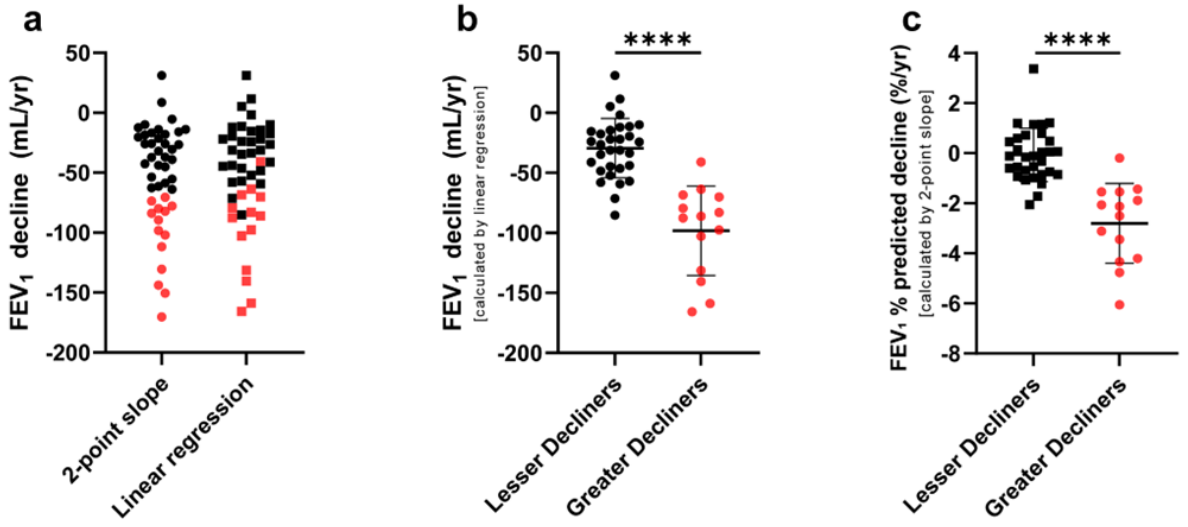
Supplemental Figure A.3. 52-ft cross-compartment signature is not influenced by participants' smoking status or ICS use.
PCA was completed with the 52 (blood and BAL) proteins identified in the multi-compartment progression signature (Figure 2.2; n=45). Participants (n=45) are classified into two groups using self-reported (a) baseline smoking status (current/ former) or (b) ICS use within three months of the baseline visit (yes/no). P-values reported from a permutation test (n = 2000 permutations) between groups' mean scores across PC1 and PC2.



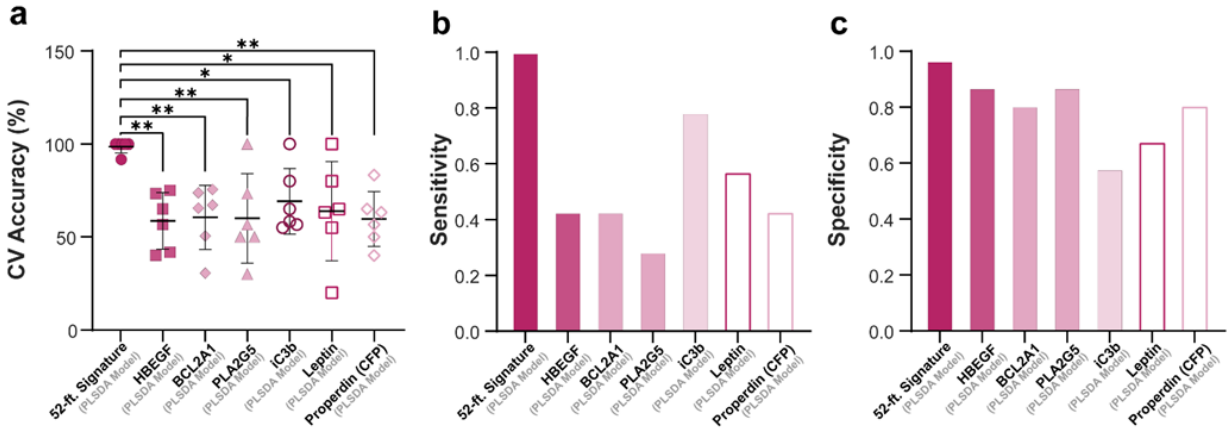
Supplemental Figure A.4. The 52-feature elastic net-identified signature outperforms random variants of equal size. Comparison of optimal model CV accuracy (98.4%) to average performance of 1000 random variant signatures, generated by selecting randomized feature sets (proteins) from the original dataset (1322 proteins) at a size equal to our optimal signature (52 features; two-tailed two-sample t-test; ****$p < 0.0001$).



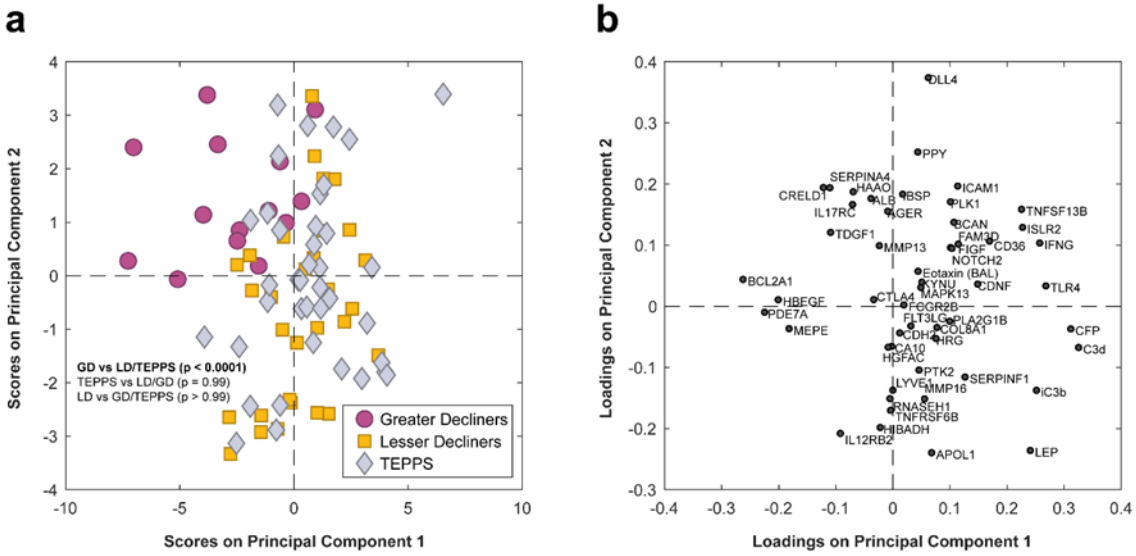
Supplemental Figure A.5. Multi-compartment progression signature is significantly enriched in an alternative, more stringent definition of rapid lung function decline.
(a) PCA completed with the 52 (blood and BAL) proteins identified in the multi-compartment progression signature (Figure 2.2). Participants (n=45) are classified into 3 groups using the alternative progression definitions proposed by Anderson et al.[109]: rapid decliners (circles; >100 mL/year), decliners (squares; 20 – 100 mL/year), stable/improvers (triangles; < 20 mL/year). (b) Comparisons of participant scores across PC1 show significant enrichment of signature in rapid decliners by this definition, compared to decliners and stable/improvers (one-way ANOVA with Holm-Šídák's multiple comparisons test; *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).



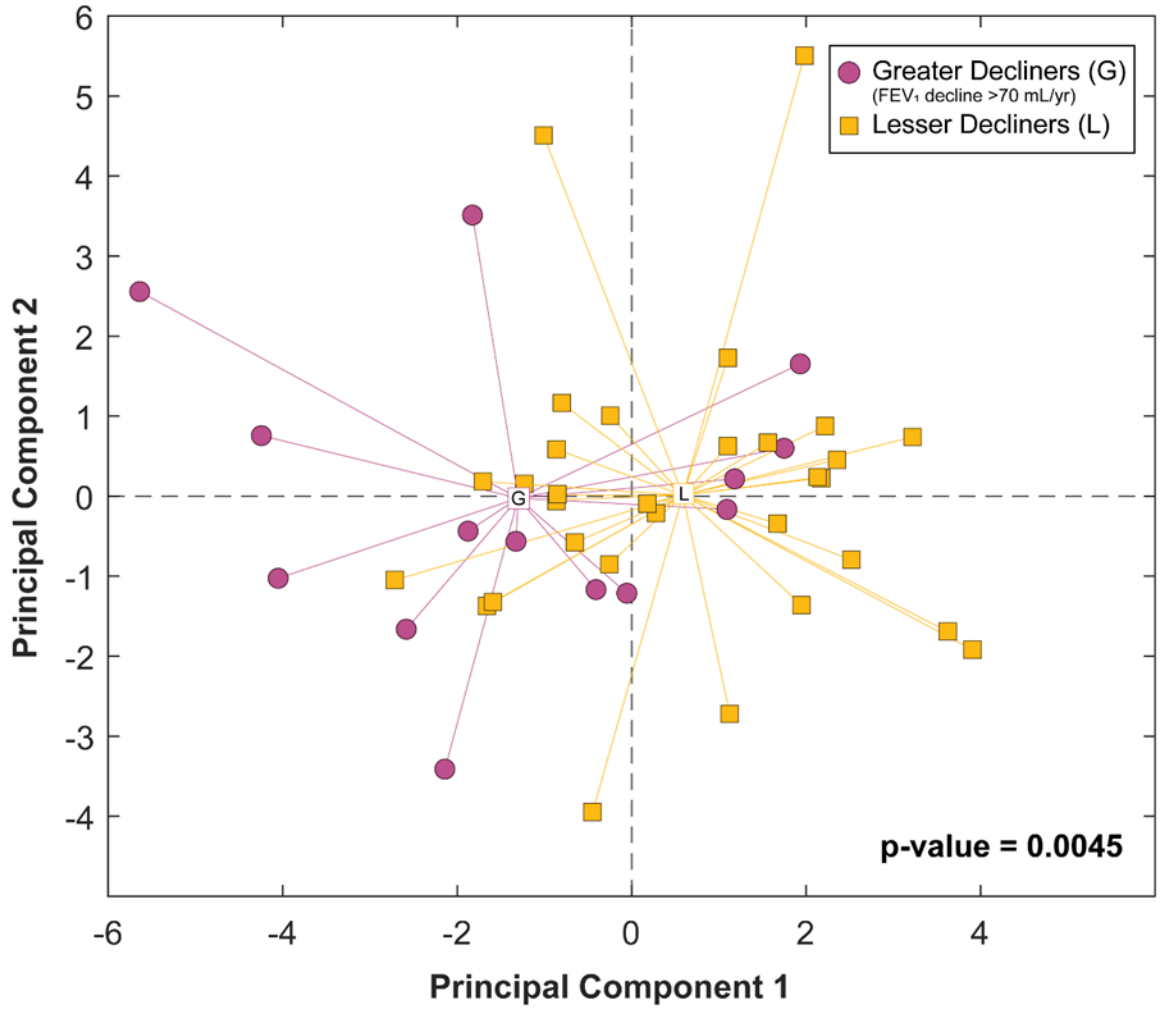
Supplemental Figure A.6. Classifications are not largely affected by alternative calculations of spirometric decline. (a) Visualization of the impact of calculating FEV_1 decline using 2 data points (Visit 1 and Visit 5) versus all available longitudinal spirometry. FEV_1 declines using 2 data points (left) were calculated using the 2-point slope equation: $\frac{V5\ FEV_1 - V1\ FEV_1}{Time\ from\ V5\ to\ V1}$. The multiple point estimate (right) was defined as the slope (β_1) of the linear regression equation $FEV_1 = \beta_0 + \beta_1(Time\ since\ V1)$ for each participant. (b) Plot of FEV_1 (mL/yr.) values generated from the linear regression approach, grouped into greater or lesser decliners based on their 2-point slope evaluations scheme. A two-sample t-test suggests that using only Visit 1 and 5 data is sufficient to capture more complex progression trends. (c) Plot of lung function decline calculated using FEV_1 % predicted at Visit 1 and 5 exhibit similar distributions as the original calculations, which use absolute FEV_1 (two-sample t-test; ****$p < 0.0001$). Red points denote participants classified as greater decliners in all plots as per the 2-point evaluation calculations completed with absolute FEV_1. For all estimates, the change in time was calculated using the visit dates for each participant; calculations assumed a fixed-length year equal to 365.2425 days.



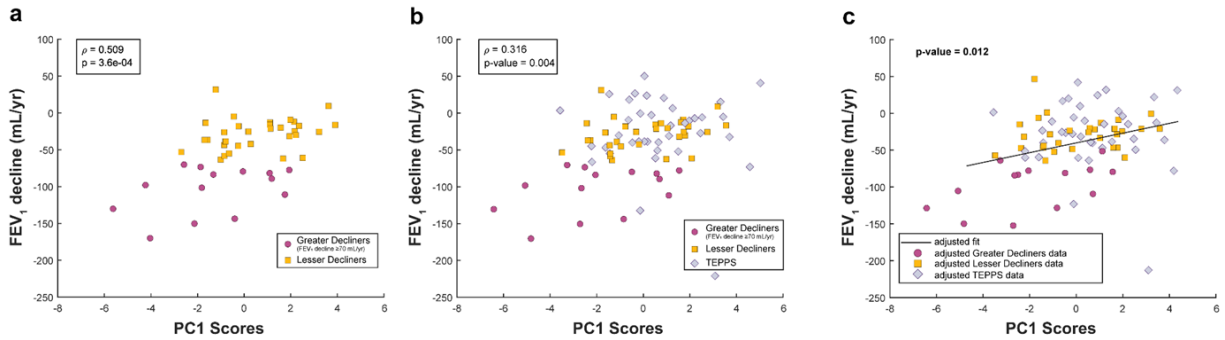
Supplemental Figure A.7. The 52-feature signature significantly outperforms analyses based on individual proteins. (a) Comparison of 6-fold cross-validation (CV) accuracies, (b) sensitivities, and (c) specificities between optimized data-driven signature and the top six individually identified proteins in Fig. 1b. One-way ANOVA with Dunnett's post hoc test; *p<0.05, **p<0.01.



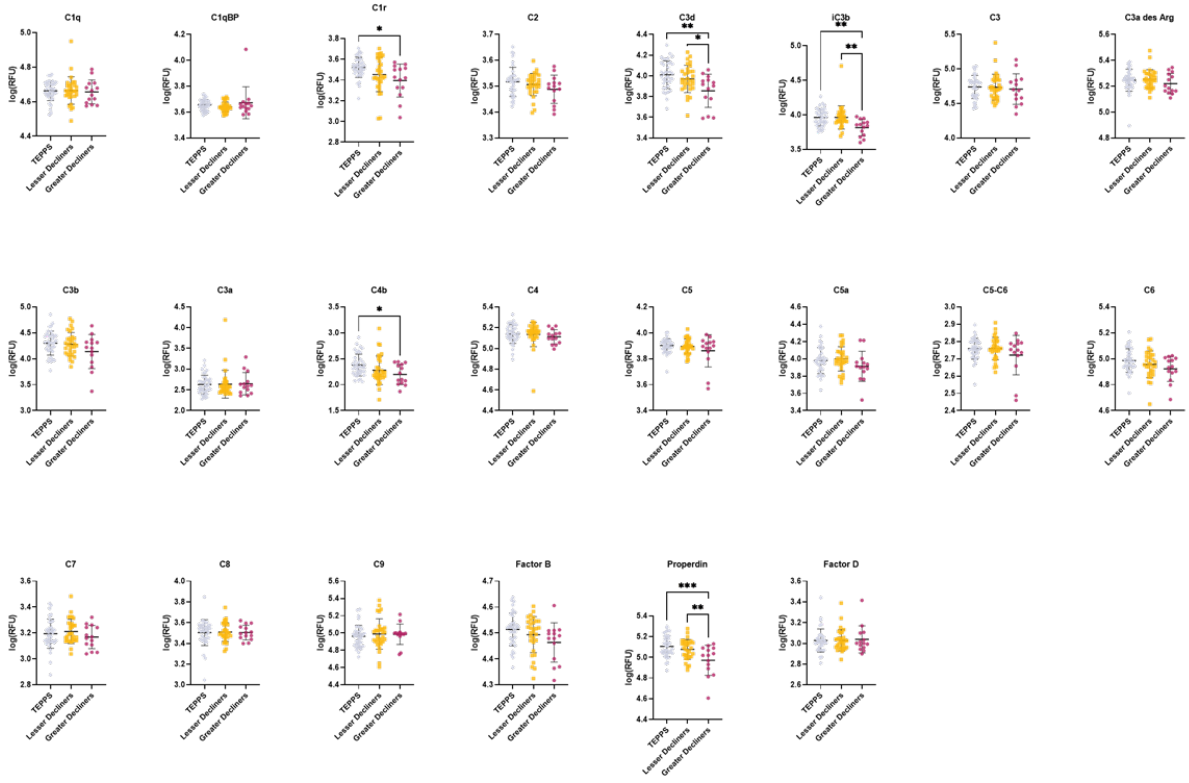
Supplemental Figure A.8. 52-feature cross-compartment signature can differentiate TEPPS from greater decliners but not lesser decliners.
(a) PCA scores plot generated with the 52 signature-identified cross-compartment proteins of greater decliners (circles), lesser decliners (squares), and a reference group of tobacco-exposed people with preserved spirometry (TEPPS) (diamonds). First two principal components (PCs) capture 17% of the variance in the dataset. P-values reported from a permutation test (n = 2000 permutations) between groups' mean scores across PC1 and PC2. (b) Protein loadings across PC1 and PC2.



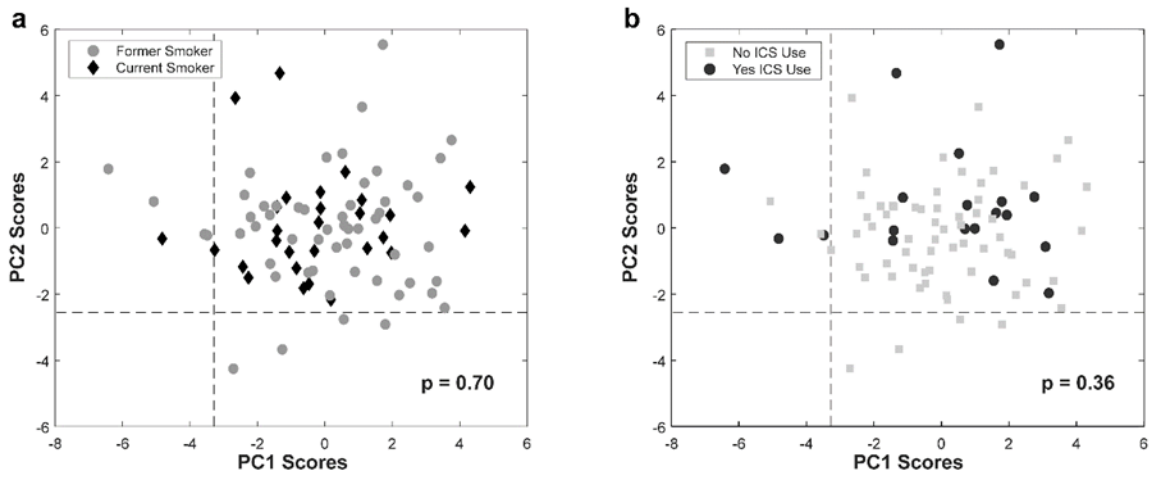
Supplemental Figure A.9. Complement profiles from greater decliners significantly differ from lesser decliners. PCA of all complement proteins measured in plasma (including C1q, C1qBP, C1r, C2, C3d, C3b, C3, C3a, iC3b, C3a des Arg, C4, C4b, C5, C5a, C5-6, C6, C7, C8, C9, Factor B, Factor D, Properdin). The first two principal components collectively capture 34.8% of total variance. P-value shown for two sample, two-tailed t-test of PC1 scores across groups.



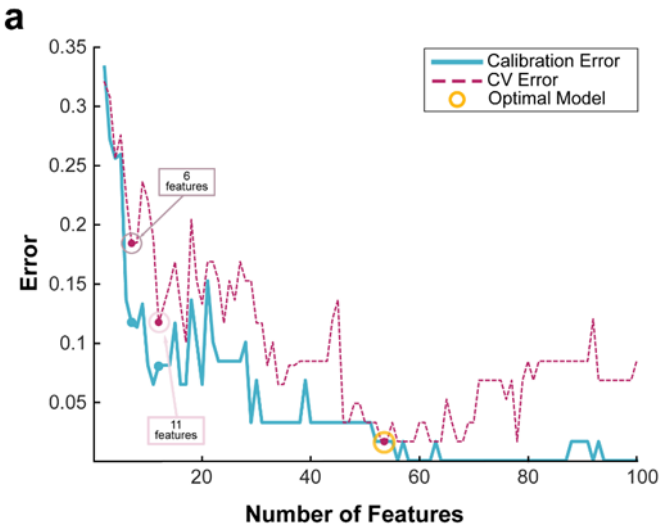
Supplemental Figure A.10. PC1 scores from the complement-specific PCA highly correlate with decline in FEV_1 (mL/yr.).
This analysis uses the PC1 scores from Supplemental Figure A.6. (a) Pearson correlation of PC1 scores and FEV_1 mL/ year in a model with greater decliners and lesser decliners alone and (b) with the addition of a reference group of tobacco smoke-exposed persons with preserved spirometry (TEPPS). (c) Observed relationship between complement protein profiles on PC1 and FEV_1 decline remains significant after adjusting for clinical covariates (p-value shown for linear regression adjusted for age, race, height, sex, baseline FEV_1% predicted, smoking status, ICS use within three months of baseline visit, and pack-years).



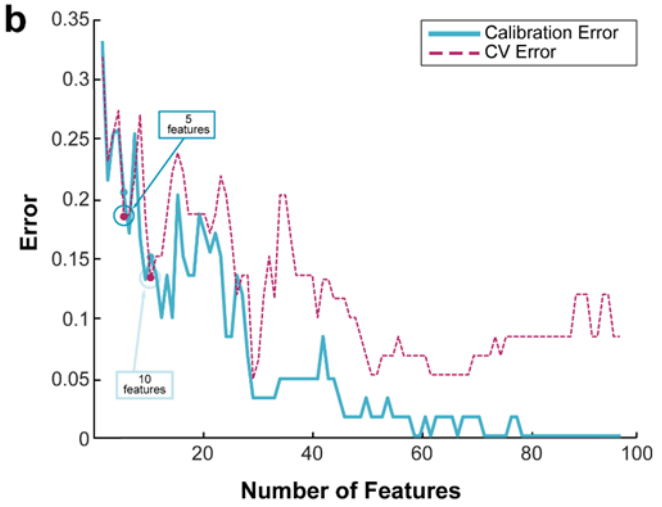
Supplemental Figure A.11. Univariate comparison of complement cascade proteins in blood.
Participant groups are as in the legend to Supplementary Figure 7. Individual levels of complement-associated proteins in greater decliners, lesser decliners, and TEPPS reference group (one-way ANOVA with Tukey's post hoc test; *p<0.05, **p<0.01, ***p < 0.001).



Supplemental Figure A.12. Complement profiles are not influenced by participants' smoking status or ICS use. Complement PCA profile (Figure 2.4; n=83) classified into two groups using self-reported (a) baseline smoking status (current/ former) or (b) ICS use within three months of the baseline visit (yes/no). P-values reported from a permutation test (n = 2000 permutations) between groups' mean scores across PC1 and PC2.

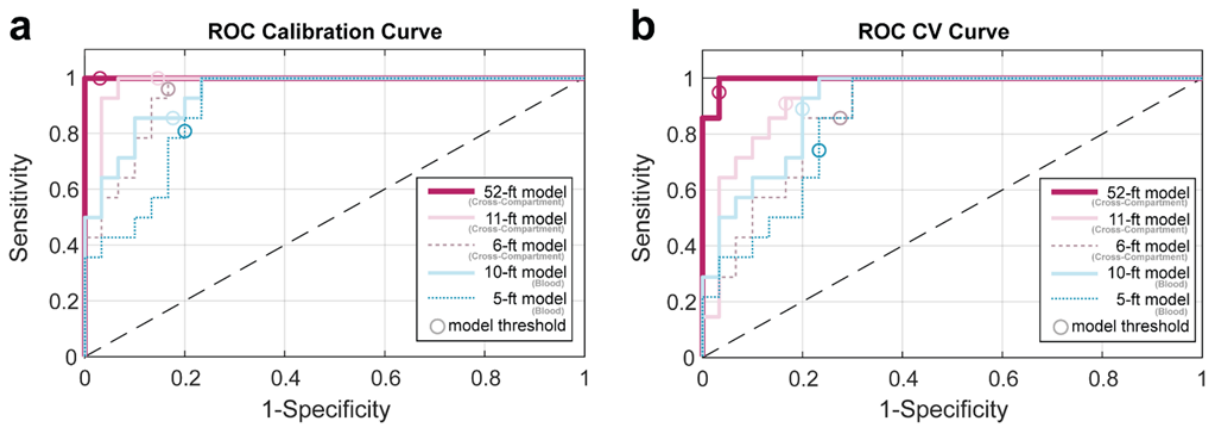


Protein Rank	Protein
1	iC3b
2	Leptin
3	Cadherin-2
4	HBEGF
5	Eotaxin (BAL)
6	TDGF1
7	FCGR2B
8	IFNγ
9	CA10
10	Apolipoprotein L1
11	LYVE1
12	ISLR2
13	CRELD1
14	PDE7A
15	TNFSF13B
16	BCL2A1
17	Properdin (CFP)
18	MAPK13
19	PTK2
20	CDNF

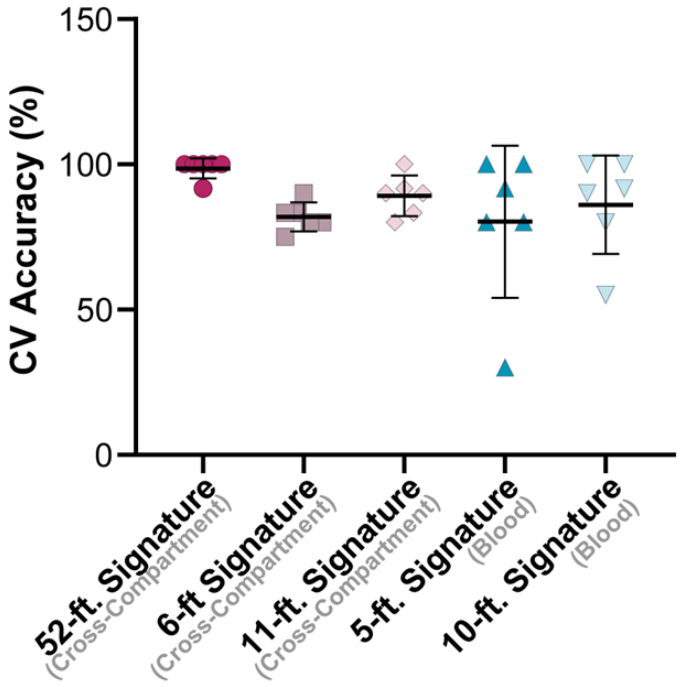


Protein Rank	Protein
1	iC3b
2	Cadherin-2
3	Leptin
4	HBEGF
5	TDGF1
6	CA10
7	IFNγ
8	FCGR2B
9	LYVE1
10	Apolipoprotein L1
11	BCL2A1
12	CRELD1
13	Properdin (CFP)
14	ISLR2
15	PDE7A
16	TNFSF13B
17	MAPK13
18	PTK2
19	CDNF
20	IBSP

Supplemental Figure A.13. Stepwise PLSDA identified minimal signatures with strong cross-validated performance. Plot of calibration and CV errors from stepwise PLSDA models for (a) cross-compartment and (b) blood-only datasets. Smaller models that separated groups with statistically comparable CV accuracies to the 52-feature models are highlighted with call out boxes (11-feature signature: 88.4%; 6-feature signature: 81.6%; 10-feature signature: 86.8%; 5-feature signature: 81.6%). Accompanying lists on the right depict the order of proteins added in models based on Elastic Net resampling selection frequency of proteins (1 = most frequently selected). Highlighted proteins denote those in the identified minimal signatures displayed in the plots.



Supplemental Figure A.14. ROC curves from PLSDA models generated from minimal signatures.
ROC curves from cross-compartment and blood minimal signatures generated from (a) calibration and (b) CV PLSDA models.



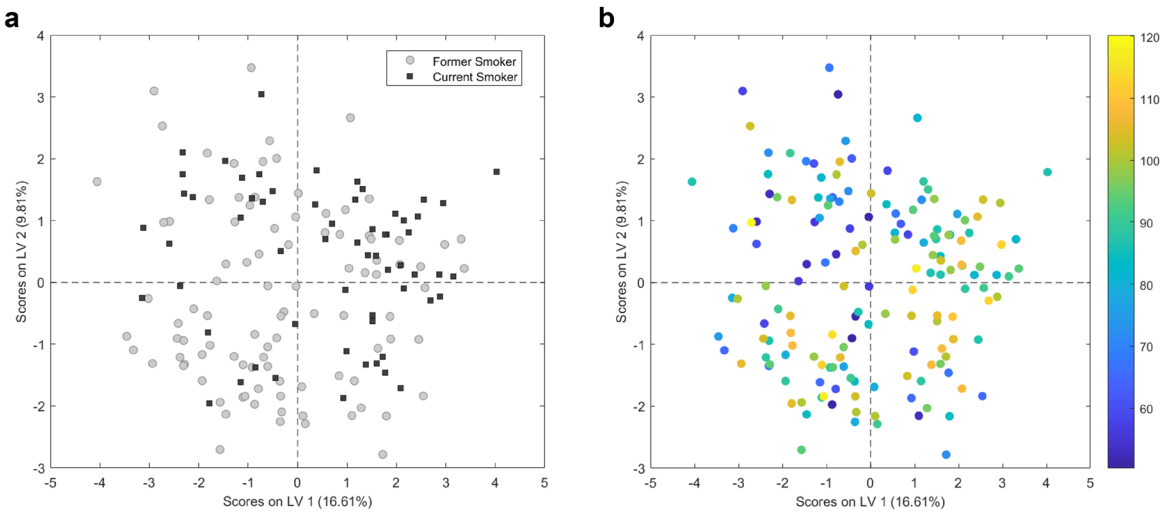
Supplemental Figure A.15. Cross-validation accuracies do not significantly vary between optimal model and smaller variants.
Comparisons of 6-fold CV accuracies between models (one-way ANOVA with Dunnett's post hoc test).

Appendix B: Supplement to: Multivariate Proteomic Signatures Reveal Age-Dependent Mechanisms Contributing to Progression in Smoking-Associated Emphysema

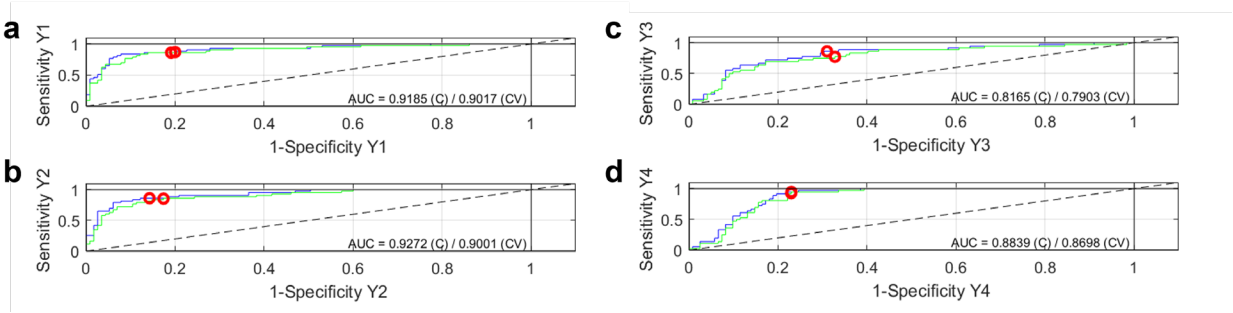
Supplemental Table B.1 Baseline demographics for ΔPRM^{fSAD} models

	Fast		Slow/Stable		Overall	
	Older (N=43)	Younger (N=36)	Older (N=43)	Younger (N=36)	Older (N=86)	Younger (N=72)
Age	**71 (± 4)**	**52 (± 3)**	**70 (± 3)**	**50 (± 3)**	**71 (± 4)**	**62 (± 9)**
Currently Smoking	**16 (37%)**	**25 (69%)**	**4 (9%)**	**17 (47%)**	**20 (23%)**	**42 (58%)**
BMI	28 (± 5)	28 (± 6)	29 (± 5)	29 (± 9)	29 (± 5)	29 (± 7)
Sex (Male)	19 (44%)	18 (50%)	23 (53%)	12 (33%)	42 (49%)	30 (42%)
Race (White / Other)	41/2 (95%)	29/7 (81%)	**43/0 (100%)**	**28/8 (78%)**	**84/2 (98%)**	**57/1 (79%)**
ICS use (yes)	1 (2%)	0 (0%)	0 (0%)	2 (6%)	1 (1%)	2 (3%)
FEV_1 (% predicted)	85 (±21)	92 (± 19)	**85 (±18)**	**94 (± 11)**	85 (±20)	93 (± 22)
FEV_1/FVC	**0.66 (± 0.1)**	**0.71 (± 0.1)**	**0.68 (± 0.1)**	**0.79 (± 0.1)**	**0.67 (± 0.1)**	**0.75 (± 0.1)**
FEV_1 (L)	**2.1 (± 0.55)**	**2.9 (± 0.61)**	**2.4 (± 0.64)**	**3.0 (± 0.75)**	**2.2 (± 0.6)**	**3.0 (± 0.7)**
PRM^{fSAD} (%)	**14 (± 8.5)**	**8.0 (± 6.0)**	**21 (± 8.9)**	**7.5 (± 5.1)**	**18 (± 9)**	**7.7 (± 5)**
ΔPRM^{fSAD} (%)	11 (± 12)	9. 7 (± 9)	**-6.6 (± 6)**	**-2.4 (± 4)**	2.2 (± 13)	3.6 (± 9)

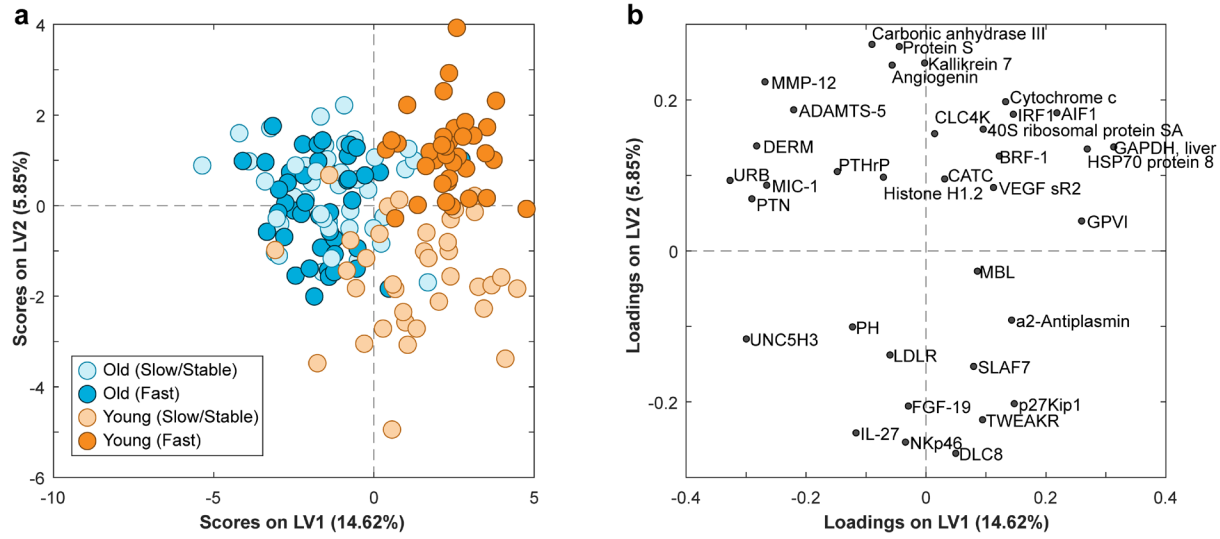
Two-sample, two-tailed t-test or Fisher's exact test were used to determine significant differences. Individual statistical tests were evaluated for younger vs older comparisons in fast, slow/stable, and overall columns. Bold values denote a statically significant difference between younger and older groups ($p<0.05$).



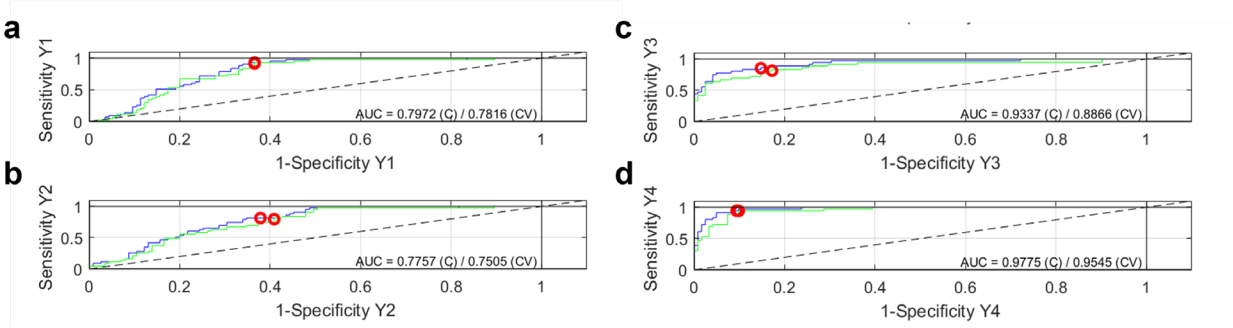
Supplemental Figure B.1. ΔPRM^{Emph} signatures is not influenced by participants' smoking status or baseline FEV_1% predicted.
PLSDA profiles (Figure 4.2; n=158) alternatively colored by self-reported (a) baseline smoking status (current/ former) or (b) baseline FEV1% predicted after bronchodilator administration. Color bar denotes FEV_1% predicted at baseline.



Supplemental Figure B.2. ROC curves for ΔPRM^{Emph} signature.
Receiver operating characteristic curves for ΔPRM^{Emph} PLSDA model (Figure 4.2). Blue lines and green lines represent ROC curves for the calibration and cross-validation models, respectively. The red circles denote the decision thresholds for PLSDA classification. Each plot summarizes the classification ability of one class used in the model: (a) older slow/stable (b) older fast (c) younger slow/stable (d) younger fast.

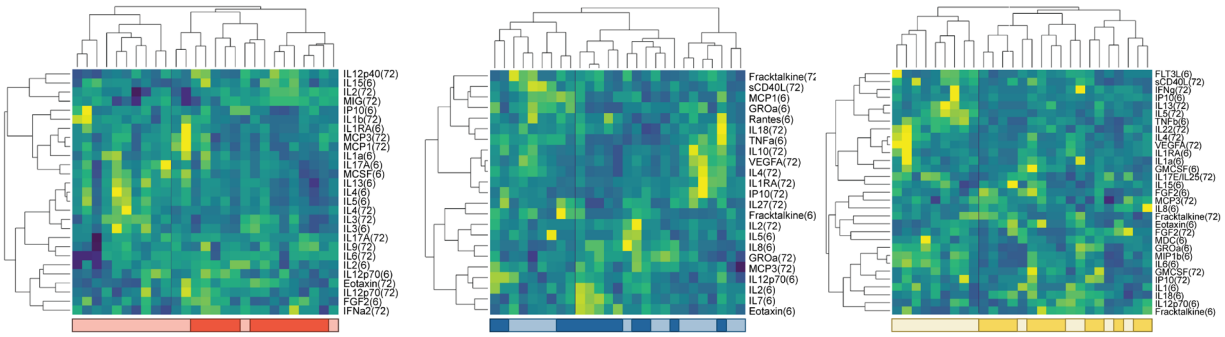


Supplemental Figure B.3. Elastic net and PLSDA identify 35 plasma proteins that can moderately differentiate ever-smokers based on their age and relative magnitudes of functional small airways disease progression after 5-years. (a) PLSDA scores plot highlighting 35-protein signature selected by bootstrap elastic net. Signature differentiates between younger (orange) and older (blue) ever-smokers with relative fast (darker) or slow (lighter) with 81.7% calibration and 79.4% cross-validation (CV) accuracy. (b) The two latent variable (LV) model captures 20.4% of the total variance in the dataset. LV1 separates ever-smokers based on age (with negatively loaded proteins being comparatively increased in older ever-smokers and positively loaded proteins being comparatively reduced). LV2 only successfully separates individuals based on their relative rates of progression in the younger population.

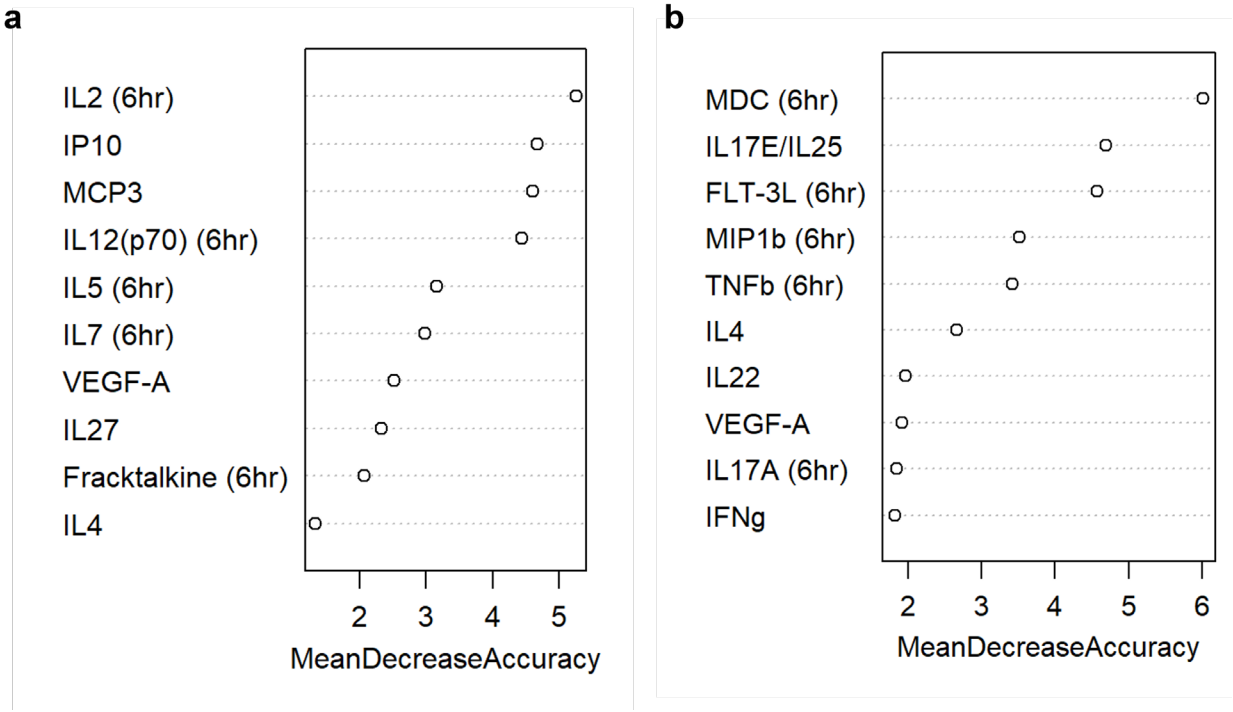


Supplemental Figure B.4. ROC curves for ΔPRM^{fSAD} signature.
Receiver operating characteristic curves for ΔPRM^{fSAD} PLSDA model (Supplemental Figure B.3). Blue lines and green lines represent ROC curves for the calibration and cross-validation models, respectively. The red circles denote the decision thresholds for PLSDA classification. Each plot summarizes the classification ability of one class used in the model: (a) older slow/stable (b) older fast (c) younger slow/stable (d) younger fast.

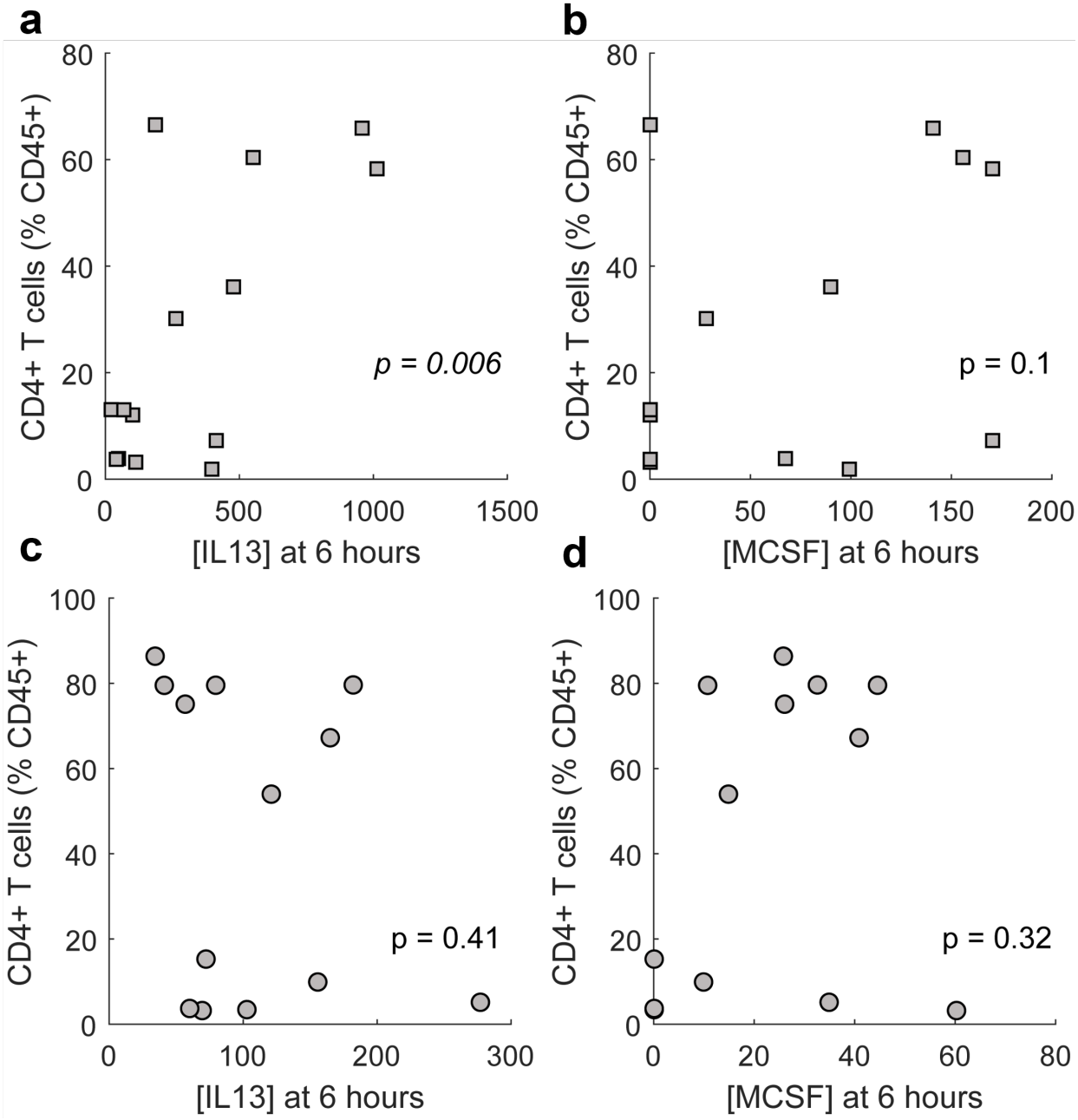
Appendix C: Supplement to: Cytokine-Chemokine Network Changes in the Immune Responses of Individuals with COPD



Supplemental Figure C.1. Hierarchal clusters generated from VIP-selected protein signatures differentiating GOLD 0 and COPD paracrine secretions.
Clustering of PBMCs stimulated with (a) anti-CD3/CD28 (b) LPS/ TLR-4, and (c) R848/ TLR7/8 stimuli for 6- and 72-hours. Clusters were performed with complete linkage and Pearson correlations. Clustering identified COPD participants with (a) 100% sensitivity and 71% specificity (b) 69% sensitivity and 35% specificity, and (c) 100% sensitivity and 64% specificity. Color bars denote participant identity (GOLD 0 or COPD). In each cluster the darker and lighter shades denote COPD and GOLD 0, respectively. Time of cytokine collection is indicated in paratheses to the right of the protein name.

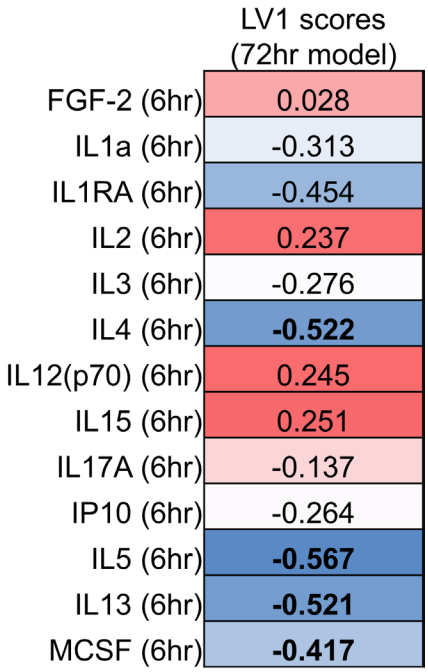


Supplemental Figure C.2. Plot of the top 5 most important features in LPS and R848 random forest models. Most important features in the (a) LPS and (b) R848 random forest models contributing to differentiation of GOLD 0 and COPD cytokine profiles. Importance was determined by the percent accuracy each protein contributes to the overall classification accuracy of the model. X-axis denotes the percent accuracy lost by the overall model if the cytokine in question is removed. Measurements obtained at 6 hours are labels as such, all other protein measurements were collected at 72 hours.

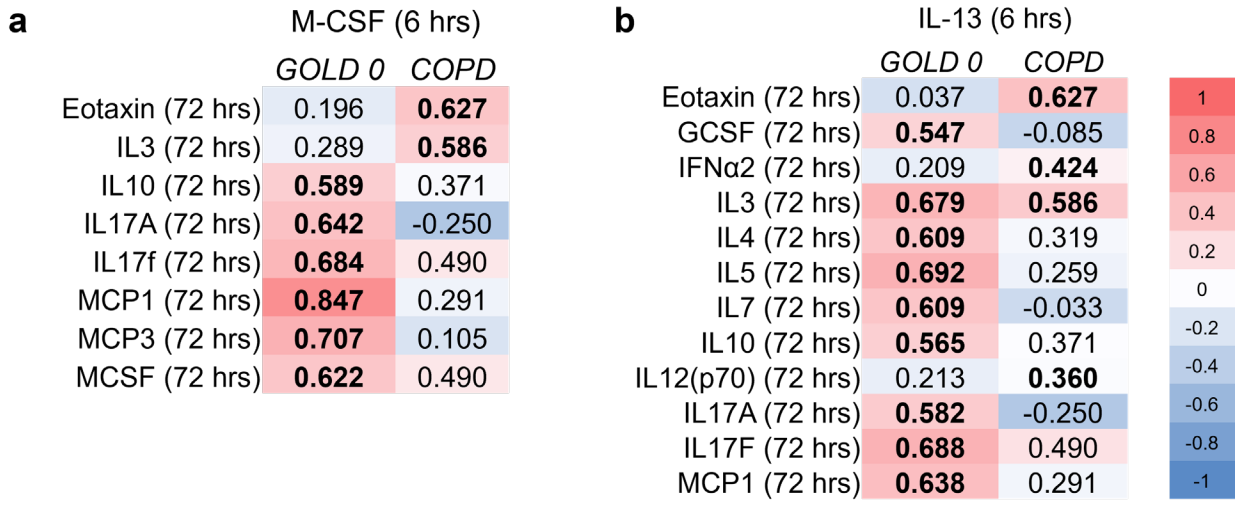


Supplemental Figure C.3. Correlations between CD4+ T cell percentages and IL-13 and M-CSF concentrations at 6 hours from PBMCs

PBMCs from (a-b) GOLD 0 and (c-d) COPD donors after stimulation with anti-CD3/CD28. M-CSF and IL-13 concentrations are background-adjusted. P-values are reported from Pearson correlations between cytokine concentrations and 6-hour CD4+ T cells percentages.



Supplemental Figure C.4. Pearson correlations between LV1 scores and early proteins in the longitudinal anti-CD3/CD28 model.
LV1 scores used in this analysis were generated from the VIP-selected PLSDA model built with 72-hour cytokine secretions after anti-CD3/CD28 stimulation (Figure 5.6a). Bolded correlation coefficient denote correlation with significant p-values ($p < 0.05$).



Supplemental Figure C.5. Spearman correlation coefficients between early M-CSF and IL-13 and late cytokines.
Correlation coefficients between early (a) M-CSF and (b) IL-13 secretions (at 6 hours) with 72-hour protein measurements. Only 72-hour proteins which were significant in either the GOLD 0 or COPD classes were included in the figure. Bolded values are associated with significant p-values ($p<0.05$). Color bar indicates correlation coefficient. Color bar represents the value of the correlation coefficient.

Bibliography

1. Agustí, A. *et al.* Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. *Eur. Respir. J.* **61**, 2300239 (2023).
2. Soriano, J. B. *et al.* Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet. Respir. Med.* **8**, 585 (2020).
3. Frieden, T. R. *et al.* Employment and activity limitations among adults with chronic obstructive pulmonary disease - United States, 2013. *Morb. Mortal. Wkly. Rep.* **64**, 289 (2015).
4. Ford, E. S. *et al.* Total and state-specific medical and absenteeism costs of COPD among adults aged ≥ 18 years in the United States for 2010 and projections through 2020. *Chest* **147**, 31–45 (2015).
5. Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease: 2019 Report*. www.goldcopd.org (2019).
6. Rennard, S. *et al.* Introducing the COPD foundation guide for diagnosis and management of COPD, recommendations of the COPD foundation. *COPD* **10**, 378–89 (2013).
7. Riley, C. M. & Sciurba, F. C. Diagnosis and outpatient management of chronic pbstructive pulmonary disease: A review. *JAMA - J. Am. Med. Assoc.* **321**, 745–746 (2019).
8. Salvi, S. S. & Barnes, P. J. Chronic obstructive pulmonary disease in non-smokers. *Lancet* **374**, 733–743 (2009).
9. Duffy, S. P. & Criner, G. J. Chronic obstructive pulmonary disease: Evaluation and management. *Med. Clin. North Am.* **103**, 453–461 (2019).
10. Agustí, A. & Hogg, J. C. Update on the pathogenesis of chronic obstructive pulmonary disease. *N. Engl. J. Med.* **381**, 1248–1256 (2019).
11. Lange, P. *et al.* Lung-function trajectories leading to chronic obstructive pulmonary disease. *N. Engl. J. Med.* **373**, 111–122 (2015).
12. Galbán, C. J. *et al.* Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat. Med.* **18**, 1711–1715 (2012).
13. Pompe, E. *et al.* Parametric response mapping on chest computed tomography associates with clinical and functional parameters in chronic obstructive pulmonary disease. *Respir. Med.* **123**, 48–55 (2017).
14. Boes, J. L. *et al.* Parametric response mapping monitors temporal changes on lung CT scans in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *Acad. Radiol.* **22**, 186–194 (2015).
15. Martinez, C. H. *et al.* Age and small airway imaging abnormalities in subjects with and without airflow obstruction in spiromics. *Am. J. Respir. Crit. Care Med.* **195**, 464–472 (2017).
16. Boudewijn, I. M. *et al.* Effects of ageing and smoking on pulmonary computed

tomography scans using parametric response mapping. *Eur. Respir. J.* **46**, 1193–1196 (2015).
17. Woodruff, P. G. *et al.* Clinical significance of symptoms in smokers with preserved pulmonary function. *N. Engl. J. Med.* **374**, 1811–1821 (2016).
18. Regan, E. A. *et al.* Clinical and radiologic disease in smokers with normal spirometry. *JAMA Intern. Med.* **175**, 1539–1549 (2015).
19. Fortis, S. *et al.* Increased mortality associated with frequent exacerbations in COPD patients with mild-to-moderate lung function impairment, and smokers with normal spirometry. *Respir. Med. X* **3**, (2021).
20. Hodge, S. *et al.* Smoking alters alveolar macrophage recognition and phagocytic ability: Implications in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **37**, 748–755 (2007).
21. Garudadri, S. *et al.* Systemic markers of inflammation in smokers with symptoms despite preserved spirometry in SPIROMICS. *Chest* **155**, 908 (2019).
22. Bestall, J. C. *et al.* Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax* **54**, 581–586 (1999).
23. Jones, P. W. *et al.* Development and first validation of the COPD Assessment Test. *Eur. Respir. J.* **34**, 648–654 (2009).
24. Tsiligianni, I. G. *et al.* Assessing health status in COPD. A head-to-head comparison between the COPD assessment test (CAT) and the clinical COPD questionnaire (CCQ). *BMC Pulm. Med.* **12**, (2012).
25. Hurst, J. R. *et al.* Prognostic risk factors for moderate-to-severe exacerbations in patients with chronic obstructive pulmonary disease: a systematic literature review. *Respir. Res.* **23**, (2022).
26. Agustí, A. & Celli, B. Natural history of COPD: gaps and opportunities. *ERJ Open Res.* **3**, 00117–02017 (2017).
27. Dransfield, M., Stolz, D. & Kleinert, S. Towards eradication of chronic obstructive pulmonary disease: a Lancet Commission. *Lancet* **393**, 1786–1788 (2019).
28. Fischer, B. M., Pavlisko, E. & Voynow, J. A. Pathogenic triad in COPD: Oxidative stress, protease-antiprotease imbalance, and inflammation. *Int. J. COPD* **6**, 413–421 (2011).
29. Christenson, S. A., Smith, B. M., Bafadhel, M. & Putcha, N. Chronic obstructive pulmonary disease. *Lancet* **399**, 2227–2242 (2022).
30. Yang, Y., Bazhin, A. V., Werner, J. & Karakhanova, S. Reactive oxygen species in the immune system. *Int. Rev. Immunol.* **32**, 249–270 (2013).
31. Bassoy, E. Y., Walch, M. & Martinvalet, D. Reactive oxygen species: Do they play a role in adaptive immunity? *Front. Immunol.* **12**, 755856 (2021).
32. Heutinck, K. M., ten Berge, I. J. M., Hack, C. E., Hamann, J. & Rowshani, A. T. Serine proteases of the human immune system in health and disease. *Mol. Immunol.* **47**, 1943–1955 (2010).
33. Commins, S. P., Borish, L. & Steinke, J. W. Immunologic messenger molecules: cytokines, interferons, and chemokines. *J. Allergy Clin. Immunol.* **125**, (2010).
34. Barnes, P. J. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* **138**, 16–27 (2016).
35. Caramori, G. *et al.* COPD immunopathology. *Semin. Immunopathol.* **38**, 497–515 (2016).
36. Curtis, J. L., Freeman, C. M. & Hogg, J. C. The immunopathogenesis of chronic

obstructive pulmonary disease: Insights from recent research. *Proc. Am. Thorac. Soc.* **4**, 512–521 (2007).
37. Freeman, C. M. & Curtis, J. L. Lung dendritic cells: Shaping immune responses throughout chronic obstructive pulmonary disease progression. *Am. J. Respir. Cell Mol. Biol.* **56**, 152–159 (2017).
38. Bagdonas, E., Raudoniute, J., Bruzauskaite, I. & Aldonyte, R. Novel aspects of pathogenesis and regeneration mechanisms in COPD. *Int. J. COPD* **10**, 995–1013 (2015).
39. Wang, Y., Xu, J., Meng, Y., Adcock, I. M. & Yao, X. Role of inflammatory cells in airway remodeling in COPD. *Int. J. COPD* **13**, 3341–3348 (2018).
40. Rovina, N., Koutsoukou, A. & Koulouris, N. G. Inflammation and immune response in COPD: Where do we stand? *Mediators Inflamm.* **2013**, 1–9 (2013).
41. Martinez, C. H. *et al.* Alveolar eosinophilia in current smokers with chronic obstructive pulmonary disease in the SPIROMICS cohort. *J. Allergy Clin. Immunol.* **141**, 429–432 (2018).
42. Freeman, C. M. *et al.* Human CD56+ cytotoxic lung lymphocytes kill autologous lung cells in chronic obstructive pulmonary disease. *PLoS One* **9**, e103840 (2014).
43. Overbeek, S. A. *et al.* Cigarette smoke-induced collagen destruction; key to chronic neutrophilic airway inflammation? *PLoS One* **8**, (2013).
44. Freeman, C. M. *et al.* Lung dendritic cell expression of maturation molecules increases with worsening chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **180**, 1179–1188 (2009).
45. Hodge, S., Hodge, G., Scicchitano, R., Reynolds, P. N. & Holmes, M. Alveolar macrophages from subjects with chronic obstructive pulmonary disease are deficient in their ability to phagocytose apoptotic airway epithelial cells. *Immunol. Cell Biol.* **81**, 289–296 (2003).
46. Stoll, P. *et al.* Impact of smoking on dendritic cell phenotypes in the airway lumen of patients with COPD. *Respir. Res.* **15**, (2014).
47. Freeman, C. M. *et al.* Lung CD8+ T cells in COPD have increased expression of bacterial TLRs. *Respir. Res.* **14**, 13 (2013).
48. Caramori, G., Adcock, I. M., Di Stefano, A. & Chung, K. F. Cytokine inhibition in the treatment of COPD. *Int. J. COPD* **9**, 397–412 (2014).
49. Chung, K. F. & Adcock, I. M. Multifaceted mechanisms in COPD: Inflammation, immunity, and tissue repair and destruction. *Eur. Respir. J.* **31**, 1334–1356 (2008).
50. Caramori, G. *et al.* Chemokines and chemokine receptors blockers as new drugs for the treatment of chronic obstructive pulmonary disease. *Curr. Med. Chem.* **20**, 4317–4349 (2013).
51. Jones, C. E. & Chan, K. Interleukin-17 stimulates the expression of interleukin-8, growth-related oncogene-alpha, and granulocyte-colony-stimulating factor by human airway epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **26**, 748–753 (2002).
52. Kawayama, T. *et al.* Interleukin-18 in pulmonary inflammatory diseases. *J. Interferon Cytokine Res.* **32**, 443–449 (2012).
53. Smith, D. E. The biological paths of IL-1 family members IL-18 and IL-33. *J. Leukoc. Biol.* **89**, 383–392 (2011).
54. Regan, E. A. *et al.* Omics and the search for blood biomarkers in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **61**, 143–149 (2019).
55. Vanfleteren, L. E. G. W. The Complexity of a Respiratory Patient. in *Textbook of*

Pulmonary Rehabilitation 37–43 (Springer International Publishing, 2018). doi:10.1007/978-3-319-65888-9_4.
56. Moon, J. Y., Leitao Filho, F. S., Shahangian, K., Takiguchi, H. & Sin, D. D. Blood and sputum protein biomarkers for chronic obstructive pulmonary disease (COPD). *Expert Rev. Proteomics* **15**, 923–935 (2018).
57. Mannino, D. M. Biomarkers for chronic obstructive pulmonary disease diagnosis and progression: Insights, disappointments and promise. *Curr. Opin. Pulm. Med.* **25**, 144–149 (2019).
58. Zemans, R. L. *et al.* Multiple biomarkers predict disease severity, progression and mortality in COPD. *Respir. Res.* **18**, (2017).
59. Park, H. Y. *et al.* Club cell protein 16 and disease progression in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **188**, 1413–1419 (2013).
60. Bradford, E. *et al.* The value of blood cytokines and chemokines in assessing COPD. *Respir. Res.* **18**, 180 (2017).
61. Higashimoto, Y. *et al.* Serum biomarkers as predictors of lung function decline in chronic obstructive pulmonary disease. *Respir. Med.* **103**, 1231–1238 (2009).
62. Rennard, S. I. *et al.* The safety and efficacy of infliximab in moderate to severe chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **175**, 926–934 (2007).
63. Mahler, D. A., Huang, S., Tabrizi, M. & Bell, G. M. Efficacy and safety of a monoclonal antibody recognizing interleukin-8 in COPD: a pilot study. *Chest* **126**, 926–934 (2004).
64. Nair, P. *et al.* Safety and efficacy of a CXCR2 antagonist in patients with severe asthma and sputum neutrophils: a randomized, placebo-controlled clinical trial. *Clin. Exp. Allergy* **42**, 1097–1103 (2012).
65. Brightling, C. E. *et al.* Benralizumab for chronic obstructive pulmonary disease and sputum eosinophilia: a randomised, double-blind, placebo-controlled, phase 2a study. *Lancet. Respir. Med.* **2**, 891–901 (2014).
66. Criner, G. J. *et al.* Benralizumab for the prevention of COPD exacerbations. *N. Engl. J. Med.* **381**, 1023–1034 (2019).
67. Moll, M. *et al.* Chronic obstructive pulmonary disease and related phenotypes: polygenic risk scores in population-based and case-control cohorts. *Lancet. Respir. Med.* **8**, 696 (2020).
68. Couper, D. *et al.* Design of the subpopulations and intermediate outcomes in COPD study (SPIROMICS). *Thorax* **69**, 491–494 (2014).
69. Ejike, C. O. *et al.* Contribution of individual and neighborhood factors to racial disparities in respiratory outcomes. *Am. J. Respir. Crit. Care Med.* **203**, 987–997 (2021).
70. Segal, L. N. & Martinez, F. J. Chronic obstructive pulmonary disease subpopulations and phenotyping. *J. Allergy Clin. Immunol.* **141**, 1961–1971 (2018).
71. Christenson, S. A. *et al.* An airway epithelial IL-17A response signature identifies a steroid-unresponsive COPD patient subgroup. *J. Clin. Invest.* **129**, (2019).
72. Christenson, S. A. *et al.* Asthma-COPD overlap: Clinical relevance of genomic signatures of type 2 inflammation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **191**, 758–766 (2015).
73. Norman, K. C., Moore, B. B., Arnold, K. B. & O'Dwyer, D. N. Proteomics: Clinical and research applications in respiratory diseases. *Respirology* **23**, 993 (2018).
74. Benedict, K. F. & Lauffenburger, D. A. Insights into proteomic immune cell signaling and communication via data-driven modeling. *Curr. Top. Microbiol. Immunol.* **363**, 201–233

(2013).
75. Chang, Y. *et al.* COPD subtypes identified by network-based clustering of blood gene expression. *Genomics* **107**, 51–58 (2016).
76. Shi, W. J. *et al.* Unsupervised discovery of phenotype-specific multi-omics networks. *Bioinformatics* **35**, 4336–4343 (2019).
77. Vanfleteren, L. E. G. W. *et al.* Biomarker-based clustering of patients with chronic obstructive pulmonary disease. *ERJ Open Res.* **9**, (2023).
78. Morrow, J. D. *et al.* Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. *BMC Med. Genomics* **8**, (2015).
79. Burgel, P. R. *et al.* Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur. Respir. J.* **36**, 531–539 (2010).
80. Burgel, P. R. *et al.* A simple algorithm for the identification of clinical COPD phenotypes. *Eur. Respir. J.* **50**, (2017).
81. Pikoula, M. *et al.* Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. *BMC Med. Inform. Decis. Mak.* **19**, 1–14 (2019).
82. Raherison, C. *et al.* Comorbidities and COPD severity in a clinic-based cohort. *BMC Pulm. Med.* **18**, 1–10 (2018).
83. Castaldi, P. J. *et al.* Cluster analysis in the COPDGene study identifies subtypes of smokers with distinct patterns of airway disease and emphysema. *Thorax* **69**, 416–423 (2014).
84. Castaldi, P. J. *et al.* Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. *Thorax* **72**, 998–1006 (2017).
85. Keene, J. D. *et al.* Biomarkers predictive of exacerbations in the SPIROMICS and COPDGene cohorts. *Am. J. Respir. Crit. Care Med.* **195**, 473–481 (2017).
86. Zhang, J. *et al.* A polygenic risk score and age of diagnosis of chronic obstructive pulmonary disease. *Eur. Respir. J.* **60**, (2022).
87. Moll, M. *et al.* Development of a blood-based transcriptional risk score for chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **205**, 161–170 (2022).
88. Godbole, S. *et al.* A metabolomic severity score for airflow obstruction and emphysema. *Metabolites* **12**, (2022).
89. Norman, K. C. *et al.* Identification of a unique temporal signature in blood and BAL associated with IPF progression. *Sci. Rep.* **10**, 12049 (2020).
90. Liu, Z. *et al.* Identification of potential COPD genes based on multi-omics data at the functional level. *Mol. Biosyst.* **12**, 191–204 (2016).
91. Zhao, P. *et al.* Integration of transcriptomics, proteomics, metabolomics and systems pharmacology data to reveal the therapeutic mechanism underlying Chinese herbal Bufei Yishen formula for the treatment of chronic obstructive pulmonary disease. *Mol. Med. Rep.* **17**, 5247–5257 (2018).
92. Li, C. X., Wheelock, C. E., Sköld, M. C. & Wheelock, Å. M. Integration of multi-omics datasets enables molecular classification of COPD. *Eur. Respir. J.* **51**, (2018).
93. Stockley, R. A., Halpin, D. M. G., Celli, B. R. & Singh, D. Chronic obstructive pulmonary disease biomarkers and their interpretation. *Am. J. Respir. Crit. Care Med.* **199**, 1195–1204 (2019).
94. Kochanek, K. D., Xu, J. & Arias, E. *Mortality in the United States, 2019*. *NCHS Data Brief, no 395* (2020).

95. Akinbami, L. J. & Liu, X. *Chronic obstructive pulmonary disease among adults aged 18 and over in the United States, 1998-2009. NCHS Data Brief, no 63* (2011).
96. Casanova, C. *et al.* The progression of chronic obstructive pulmonary disease is heterogeneous: The experience of the BODE cohort. *Am. J. Respir. Crit. Care Med.* **184**, 1015–1021 (2011).
97. Nishimura, M. *et al.* Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **185**, 44–52 (2012).
98. Mannino, D. M., Reichert, M. M. & Davis, K. J. Lung function decline and outcomes in an adult population. *Am. J. Respir. Crit. Care Med.* **173**, 985–990 (2006).
99. Vestbo, J. *et al.* Changes in forced expiratory volume in 1 second over time in COPD. *N. Engl. J. Med.* **365**, 1184–1192 (2011).
100. Celli, B. R. *et al.* Serum biomarkers and outcomes in patients with moderate COPD: A substudy of the randomised SUMMIT trial. *BMJ Open Respir. Res.* **6**, 431 (2019).
101. Ahmadi-Abhari, S., Kaptoge, S., Luben, R. N., Wareham, N. J. & Khaw, K. T. Longitudinal association of C-reactive protein and lung function over 13 years. *Am. J. Epidemiol.* **179**, 48–56 (2014).
102. Fogarty, A. W., Jones, S., Britton, J. R., Lewis, S. A. & McKeever, T. M. Systemic inflammation and decline in lung function in a general population: A prospective study. *Thorax* **62**, 515–520 (2007).
103. Suzuki, M. *et al.* Lower leptin/adiponectin ratio and risk of rapid lung function decline in chronic obstructive pulmonary disease. *Ann. Am. Thorac. Soc.* **11**, 1511–1519 (2014).
104. O'Dwyer, D. N. *et al.* The peripheral blood proteome signature of idiopathic pulmonary fibrosis is distinct from normal and is associated with novel immunological processes. *Sci. Rep.* **7**, 1–13 (2017).
105. Norman, K. C. *et al.* Inference of cellular immune environments in sputum and peripheral blood associated with acute exacerbations of COPD. *Cell. Mol. Bioeng.* **12**, 165–177 (2019).
106. Wells, M. J. *et al.* Safety and tolerability of comprehensive research bronchoscopy in chronic obstructive pulmonary disease results from the SPIROMICS bronchoscopy substudy. *Ann. Am. Thorac. Soc.* **16**, 439–446 (2019).
107. Freeman, C. M. *et al.* Design of a multi-center immunophenotyping analysis of peripheral blood, sputum and bronchoalveolar lavage fluid in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *J. Transl. Med.* **13**, (2015).
108. Perez-Padilla, R. *et al.* Lung function decline in subjects with and without COPD in a population-based cohort in Latin-America. *PLoS One* **12**, e0177032 (2017).
109. Anderson, W. *et al.* Classification of lung function trajectories in SPIROMICS [abstract]. *Eur. Respir. J.* **58**, Suppl 65, PA1858 (2021).
110. Hanley, J. A. & McNeil, B. J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, 839–843 (1983).
111. Oudijk, E. J. D., Lammers, J. W. J. & Koenderman, L. Systemic inflammation in chronic obstructive pulmonary disease. *Eur. Respir. Journal, Suppl.* **22**, 5s-13s (2003).
112. Martinez, F. J. *et al.* At the root: Defining and halting progression of early chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **197**, 1540–1551 (2018).
113. Miller, R. D., Kueppers, F. & Offord K P. Serum concentrations of C3 and C4 of the complement system in patients with chronic obstructive pulmonary disease. *J. Lab. Clin. Med.* **95**, 266–271 (1980).

114. Chauhan, S., Gupta, M., Goyal, A. & Dasgupta, D. Alterations in immunogolublin & complement levels in chronic obstructive pulmonary disease. *Indian J. Med. Res.* **92**, 241–245 (1990).
115. Kosmas, E. N., Zorpidou, D., Vassilareas, V., Roussou, T. & Michaelides, S. Decreased C4 complement component serum levels correlate with the degree of emphysema in patients with chronic bronchitis. *Chest* **112**, 341–347 (1997).
116. Baralla, A. *et al.* Plasma proteomic signatures in early chronic obstructive pulmonary disease. *Proteomics. Clin. Appl.* **12**, 1700088 (2018).
117. Marc, M. M. *et al.* Complement factor C5a in acute exacerbation of chronic obstructive pulmonary disease. *Scand. J. Immunol.* **71**, 386–391 (2010).
118. Tan, D. B. A. *et al.* Protein network analysis identifies changes in the level of proteins involved in platelet degranulation, proteolysis and cholesterol metabolism pathways in AECOPD patients. *COPD J. Chronic Obstr. Pulm. Dis.* **17**, 29–33 (2020).
119. Sun, W. *et al.* Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet.* **12**, e1006011 (2016).
120. Mastej, E. *et al.* Identifying protein–metabolite networks associated with COPD phenotypes. *Metabolites* **10**, 124 (2020).
121. Mahesh, M., Yalamudi, M. & Lokesh, S. Complement levels in chronic obstructive pulmonary disease: Correlation with pulmonary function and radiological emphysema score. *Int. J. Sci. Study* **3**, 280–284 (2016).
122. Yuan, X. *et al.* Activation of C3a receptor is required in cigarette smoke-mediated emphysema. *Mucosal Immunol.* **8**, 874–885 (2015).
123. Marc, M. M. *et al.* Complement factors C3a, C4a, and C5a in chronic obstructive pulmonary disease and asthma. *Am. J. Respir. Cell Mol. Biol.* **31**, 216–219 (2004).
124. Kulkarni, H. S. *et al.* Intracellular C3 protects human airway epithelial cells from stress-associated cell death. *Am. J. Respir. Cell Mol. Biol.* **60**, 144–157 (2019).
125. Sohal, S. S. *et al.* Reticular basement membrane fragmentation and potential epithelial mesenchymal transition is exaggerated in the airways of smokers with chronic obstructive pulmonary disease. *Respirology* **15**, 930–938 (2010).
126. Mahmood, M. Q. *et al.* Epithelial mesenchymal transition in smokers: large versus small airways and relation to airflow obstruction. *Int. J. Chron. Obstruct. Pulmon. Dis.* **10**, 1515–1524 (2015).
127. Milara, J., Peiró, T., Serrano, A. & Cortijo, J. Epithelial to mesenchymal transition is increased in patients with COPD and induced by cigarette smoke. *Thorax* **68**, 410–420 (2013).
128. Nishioka, M. *et al.* Fibroblast-epithelial cell interactions drive epithelial-mesenchymal transition differently in cells from normal and COPD patients. *Respir. Res.* **16**, (2015).
129. Cockayne, D. *et al.* Systemic biomarkers of neutrophilic inflammation, tissue injury and repair in COPD patients with differing levels of disease severity. *PLoS One* **7**, (2012).
130. Lai, T. *et al.* Heparin-binding epidermal growth factor contributes to COPD disease severity by modulating airway fibrosis and pulmonary epithelial-mesenchymal transition. *Lab. Invest.* **98**, 1159–1169 (2018).
131. Wei, Y. *et al.* Prognostic value of cripto-1 expression in non-small-cell lung cancer patients: a systematic review and meta-analysis. *Biomark. Med.* **14**, 317–329 (2020).
132. Nunomiya, K. *et al.* Relationship between serum level of lymphatic vessel endothelial hyaluronan receptor-1 and prognosis in patients with lung cancer. *J. Cancer* **5**, 242 (2014).

133. Higham, A., Quinn, A. M., Cançado, J. E. D. & Singh, D. The pathology of small airways disease in COPD: historical aspects and future directions. *Respir. Res.* **20**, 49 (2019).
134. Raffield, L. M. *et al.* Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics* **20**, 1900278 (2020).
135. O'Neal, W. K. *et al.* Comparison of serum, EDTA plasma and P100 plasma for luminex-based biomarker multiplex assays in patients with chronic obstructive pulmonary disease in the SPIROMICS study. *J. Transl. Med.* **12**, (2014).
136. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).
137. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, (2019).
138. Qian, J., Hastie, T., Friedman, J., Tibshirani, R. & Simon, N. Glmnet for MATLAB. http://www.stanford.edu/~hastie/glmnet_matlab/ (2013).
139. DiLillo, K. M. *et al.* A blood and bronchoalveolar lavage protein signature of rapid FEV1 decline in smoking-associated COPD. *Sci. Rep.* **13**, (2023).
140. Reis, E. S., Falcão, D. A. & Isaac, L. Clinical aspects and molecular basis of primary deficiencies of complement component C3 and its regulatory proteins factor I and factor H. *Scand. J. Immunol.* **63**, 155–168 (2006).
141. Polverino, F., Seys, L. J. M., Bracke, K. R. & Owen, C. A. B cells in chronic obstructive pulmonary disease: moving to center stage. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **311**, L687–L695 (2016).
142. Serban, K. A. *et al.* Lectin complement pathway in emphysema. *Am. J. Respir. Crit. Care Med.* **199**, 659–661 (2019).
143. Yuan, X. *et al.* Cigarette smoke-induced reduction of C1q promotes emphysema. *JCI insight* **5**, (2019).
144. McCubbrey, A. L. & Curtis, J. L. Efferocytosis and lung disease. *Chest* **143**, 1750–1757 (2013).
145. Regan, E. A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD J. Chronic Obstr. Pulm. Dis.* **7**, 32–43 (2010).
146. Rennard, S. I. & Drummond, M. B. Early chronic obstructive pulmonary disease: Definition, assessment, and prevention. *Lancet* **385**, 1778–1788 (2015).
147. Allinson, J. P. *et al.* The presence of chronic mucus hypersecretion across adult life in relation to chronic obstructive pulmonary disease development. *Am. J. Respir. Crit. Care Med.* **193**, 662–672 (2016).
148. Allinson, J. P. *et al.* Combined impact of smoking and early-life exposures on adult lung function trajectories. *Am. J. Respir. Crit. Care Med.* **196**, 1021–1030 (2017).
149. Chen, S. *et al.* Risk factors for FEV1 decline in mild COPD and high-risk populations. *Int. J. Chron. Obstruct. Pulmon. Dis.* **12**, 435–442 (2017).
150. Kohansal, R. *et al.* The natural history of chronic airflow obstruction revisited: An analysis of the Framingham Offspring Cohort. *Am. J. Respir. Crit. Care Med.* **180**, 3–10 (2009).
151. Jetmalani, K. *et al.* Peripheral airway dysfunction and relationship with symptoms in smokers with preserved spirometry. *Respirology* **23**, 512–518 (2018).
152. Vasilescu, D. M. *et al.* Noninvasive imaging biomarker identifies small airway damage in severe chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **200**, 575–581 (2019).

153. Verma, A. K. *et al.* Increased serum levels of matrix-metalloproteinase-9, cyclo-oxygenase-2 and prostaglandin E-2 in patients with chronic obstructive pulmonary disease (COPD). *Indian J. Clin. Biochem.* **37**, 169–177 (2022).
154. Zhang, X., Li, D., Wang, H., Wu, Y. & Wen, F. Elevated plasma cytochrome c levels in patients with chronic obstructive pulmonary disease. *Curr. Sci. Assoc.* **110**, 1532–1535 (2016).
155. Kasahara, Y. *et al.* Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J. Clin. Invest.* **106**, 1311–1319 (2000).
156. Tuder, R. M. *et al.* Oxidative stress and apoptosis interact and cause emphysema due to vascular endothelial growth factor receptor blockade. *Am. J. Respir. Cell Mol. Biol.* **29**, 88–97 (2003).
157. Kasahara, Y. *et al.* Endothelial cell death and decreased expression of vascular endothelial growth factor and vascular endothelial growth factor receptor 2 in emphysema. *Am. J. Respir. Crit. Care Med.* **163**, 737–744 (2001).
158. Imai, K., Mercer, B. A., Schulman, L. L., Sonett, J. R. & D'Armiento, J. M. Correlation of lung surface area to apoptosis and proliferation in human emphysema. *Eur. Respir. J.* **25**, 250–258 (2005).
159. Demedts, I. K., Demoor, T., Bracke, K. R., Joos, G. F. & Brusselle, G. G. Role of apoptosis in the pathogenesis of COPD and pulmonary emphysema. *Respir. Res.* **7**, (2006).
160. Plataki, M. *et al.* Apoptotic mechanisms in the pathogenesis of COPD. *Int. J. Chron. Obstruct. Pulmon. Dis.* **1**, 161–171 (2006).
161. Majo, J., Ghezzo, H. & Cosio, M. G. Lymphocyte population and apoptosis in the lungs of smokers and their relation to emphysema. *Eur. Respir. J.* **17**, 946–953 (2001).
162. Finch, D. K. *et al.* Lung dendritic cells drive natural killer cytotoxicity in chronic obstructive pulmonary disease via IL-15Ra. *Am. J. Respir. Crit. Care Med.* **198**, 1140–1150 (2018).
163. Pallazola, A. M. *et al.* Human lung cDC1 drive increased perforin-mediated NK cytotoxicity in chronic obstructive pulmonary disease. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **321**, L1183–L1193 (2021).
164. Hodge, S. J., Hodge, G. L., Reynolds, P. N., Scicchitano, R. & Holmes, M. Increased production of TGF-beta and apoptosis of T lymphocytes isolated from peripheral blood in COPD. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **285**, (2003).
165. Karakioulaki, M., Papakonstantinou, E. & Stolz, D. Extracellular matrix remodelling in COPD. *Eur. Respir. Rev.* **29**, 1–13 (2020).
166. Hautamaki, R. D., Kobayashi, D. K., Senior, R. M. & Shapiro, S. D. Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice. *Science (80-.).* **277**, 2002–2004 (1997).
167. Shapiro, S. D. The macrophage in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **160**, 29–32 (1999).
168. Genovese, F., Manresa, A. A., Leeming, D. J., Karsdal, M. A. & Boor, P. The extracellular matrix in the kidney: a source of novel non-invasive biomarkers of kidney fibrosis? *Fibrogenesis Tissue Repair* **7**, (2014).
169. Karsdal, M. A. *et al.* Review article: the efficacy of biomarkers in chronic fibroproliferative diseases - early diagnosis and prognosis, with liver fibrosis as an exemplar. *Aliment. Pharmacol. Ther.* **40**, 233–249 (2014).

170. Karsdal, M. A. *et al.* Novel insights into the function and dynamics of extracellular matrix in liver fibrosis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **308**, G807–G830 (2015).
171. Saad, M. I. *et al.* ADAM17 deficiency protects against pulmonary emphysema. *Am. J. Respir. Cell Mol. Biol.* **64**, 183–195 (2021).
172. Wischhusen, J., Melero, I. & Fridman, W. H. Growth/Differentiation Factor-15 (GDF-15): From biomarker to novel targetable immune checkpoint. *Front. Immunol.* **11**, 951 (2020).
173. Wang, X. Pleiotrophin: Activity and mechanism. *Adv. Clin. Chem.* **98**, 51 (2020).
174. Serban, K. A. *et al.* Unique and shared systemic biomarkers for emphysema in Alpha-1 Antitrypsin deficiency and chronic obstructive pulmonary disease. *eBioMedicine* **84**, 104262 (2022).
175. Husebø, G. R. *et al.* Growth differentiation factor-15 is a predictor of important disease outcomes in patients with COPD. *Eur. Respir. J.* **49**, (2017).
176. Flanagin, A., Frey, T. & Christiansen, S. L. Updated guidance on the reporting of race and ethnicity in medical and science journals. *JAMA* **326**, 621–627 (2021).
177. SPIROMICS Study of Early COPD Progression (SOURCE). ClinicalTrials.gov identifier: NCT05033990. https://classic.clinicaltrials.gov/ct2/show/NCT05033990 (2023).
178. Barnes, P. J. *et al.* Chronic obstructive pulmonary disease. *Nat. Rev. Dis. Prim. 2015 11* **1**, 1–21 (2015).
179. Arnold, K. B., Szeto, G. L., Alter, G., Irvine, D. J. & Lauffenburger, D. A. CD4+ T cell-dependent and CD4+ T cell-independent cytokine-chemokine network changes in the immune responses of HIV-infected individuals. *Sci. Signal.* **8**, ra104 (2015).
180. Lau, K. S. *et al.* In vivo systems analysis identifies spatial and temporal aspects of the modulation of TNF-α-induced apoptosis and proliferation by MAPKs. *Sci. Signal.* **4**, (2011).
181. Lau, K. S. *et al.* Multi-scale in vivo systems analysis reveals the influence of immune cells on TNF-α-induced apoptosis in the intestinal epithelium. *PLoS Biol.* **10**, (2012).
182. Roberts, M. E. P. *et al.* CD4+ T-cell profiles and peripheral blood ex-vivo responses to T-cell directed stimulation delineate COPD phenotypes. *Chronic Obstr. Pulm. Dis. J. COPD Found.* **2**, 268–280 (2015).
183. Gadgil, A., Zhu, X., Sciurba, F. C. & Duncan, S. R. Altered T-cell phenotypes in chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **3**, 487–488 (2006).
184. Lambers, C. *et al.* T cell senescence and contraction of T cell repertoire diversity in patients with chronic obstructive pulmonary disease. *Clin. Exp. Immunol.* **155**, 466–475 (2009).
185. Lemaire, F. *et al.* The elastin peptide VGVAPG increases CD4+ T-cell IL-4 production in patients with chronic obstructive pulmonary disease. *Respir. Res.* **22**, (2021).
186. Tang, W. *et al.* The correlation between a Th1/Th2 cytokines imbalance and vitamin D level in patients with early chronic obstructive pulmonary disease (COPD), based on screening results. *Front. Physiol.* **14**, 1032786 (2023).
187. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4+T cells: Differentiation and functions. *Clin. Dev. Immunol.* **2012**, 12 (2012).
188. Paats, M. S., Bergen, I. M., Hoogsteden, H. C., Van Der Eerden, M. M. & Hendriks, R. W. Systemic CD4+ and CD8+ T-cell cytokine profiles correlate with GOLD stage in stable COPD. *Eur. Respir. J.* **40**, 330–337 (2012).
189. Guarda, G. *et al.* T cells dampen innate immune responses through inhibition of NLRP1 and NLRP3 inflammasomes. *Nat. 2009 4607252* **460**, 269–273 (2009).

190. Dong Kim, K. *et al.* Adaptive immune cells temper initial innate responses. *Nat. Med.* **13**, 1248–1252 (2007).
191. Knobloch, J. *et al.* The T-helper cell type 1 immune response to gram-negative bacterial infections is impaired in COPD. *Am. J. Respir. Crit. Care Med.* **183**, 204–214 (2011).
192. Cunningham, E., Ciampi, A., Joober, R. & Labbe, A. Estimating and correcting optimism bias in multivariate PLS regression: Application to the study of the association between single nucleotide polymorphisms and multivariate traits in attention deficit hyperactivity disorder. in *The Multiple Facets of Partial Least Squares and Related Methods* (eds. Abdi, H., Vinzi Esposito, V., Russolillo, G., Saporta, G. & Trinchera, L.) 103–114 (Springer International Publishing, 2014).
193. Aadland, E., Kvalheim, O. M., Anderssen, S. A., Resaland, G. K. & Andersen, L. B. The multivariate physical activity signature associated with metabolic health in children. *Int. J. Behav. Nutr. Phys. Act.* **15**, (2018).
194. Chen, T. *et al.* Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. *Evidence-based Complement. Altern. Med.* **2013**, (2013).
195. Grumelli, S., Lu, B., Peterson, L., Maeno, T. & Gerard, C. CD46 protects against chronic obstructive pulmonary disease. *PLoS One* **6**, e18785 (2011).
196. Westwood, J. P. *et al.* The role of complement activation in COPD exacerbation recovery. *ERJ Open Research* vol. 2 (2016).
197. Zhao, P. *et al.* The imbalance in the complement system and its possible physiological mechanisms in patients with lung cancer. *BMC Cancer* **19**, 1–11 (2019).
198. Da Costa, M. G. *et al.* Age and sex-associated changes of complement activity and complement levels in a healthy caucasian population. *Front. Immunol.* **9**, 2664 (2018).
199. Dickens, J. A. *et al.* COPD association and repeatability of blood biomarkers in the ECLIPSE cohort. *Respir. Res.* **12**, (2011).
200. Miller, B. E. *et al.* Plasma fibrinogen qualification as a drug development tool in chronic obstructive pulmonary disease. Perspective of the chronic obstructive pulmonary disease biomarker qualification consortium. *Am. J. Respir. Crit. Care Med.* **193**, 607–613 (2016).
201. Martinez, C. H. *et al.* Age-related differences in health-related quality of life in COPD: An analysis of the COPDGene and SPIROMICS cohorts. *Chest* **149**, 927 (2016).
202. Maremanda, K. P., Sundar, I. K., Li, D. & Rahman, I. Age-dependent assessment of genes involved in cellular senescence, telomere, and mitochondrial pathways in human lung tissue of smokers, COPD, and IPF: associations with SARS-CoV-2 COVID-19 ACE2-TMPRSS2-Furin-DPP4 axis. *Front. Pharmacol.* **11**, 584637 (2020).
203. Çolak, Y., Afzal, S., Nordestgaard, B. G., Vestbo, J. & Lange, P. Prevalence, characteristics, and prognosis of early chronic obstructive pulmonary disease the Copenhagen general population study. *Am. J. Respir. Crit. Care Med.* **201**, 671–680 (2020).
204. Rabe, K. F. *et al.* Safety and efficacy of itepekimab in patients with moderate-to-severe COPD: a genetic association study and randomised, double-blind, phase 2a trial. *Lancet Respir. Med.* **9**, 1288–1298 (2021).
205. Shapiro, S. D. Vascular atrophy and VEGFR-2 signaling: old theories of pulmonary emphysema meet new data. *J. Clin. Invest.* **106**, 1309 (2000).
206. Green, D. R. & Kroemer, G. Pharmacological manipulation of cell death: clinical applications in sight? *J. Clin. Invest.* **115**, 2610–2617 (2005).
207. Labaki, W. W. *et al.* Reprint of: Voxel-wise longitudinal parametric response mapping

analysis of chest computed tomography in smokers. *Acad. Radiol.* **26**, 306–312 (2019).
208. Hogg, J. C., McDonough, J. E. & Suzuki, M. Small airway obstruction in COPD: New insights based on micro-CT imaging and MRI imaging. *Chest* **143**, 1436 (2013).
209. Koo, H. K. *et al.* Small airways disease in mild and moderate chronic obstructive pulmonary disease: a cross-sectional study. *Lancet. Respir. Med.* **6**, 591–602 (2018).
210. Konstantinos Katsoulis, K., Kostikas, K. & Kontakiotis, T. Techniques for assessing small airways function: Possible applications in asthma and COPD. *Respir. Med.* **119**, e2–e9 (2016).
211. Contoli, M. *et al.* The small airways and distal lung compartment in asthma and COPD: a time for reappraisal. *Allergy* **65**, 141–151 (2010).
212. Desiraju, K. & Agrawal, A. Impulse oscillometry: The state-of-art for lung function testing. *Lung India* **33**, 410 (2016).
213. Skloot, G. *et al.* Respiratory symptoms and physiologic assessment of ironworkers at the World Trade Center disaster site. *Chest* **125**, 1248–1255 (2004).
214. Oppenheimer, B. W. *et al.* Distal airway function in symptomatic subjects with normal spirometry following world trade center dust exposure. *Chest* **132**, 1275–1282 (2007).
215. Su, Z. Q. *et al.* Significances of spirometry and impulse oscillometry for detecting small airway disorders assessed with endobronchial optical coherence tomography in COPD. *Int. J. Chron. Obstruct. Pulmon. Dis.* **13**, 3031–3044 (2018).
216. Li, L. Y. *et al.* Impulse oscillometry for detection of small airway dysfunction in subjects with chronic respiratory symptoms and preserved pulmonary function. *Respir. Res.* **22**, 1–10 (2021).
217. Pixley, F. J. & Stanley, E. R. CSF-1 regulation of the wandering macrophage: complexity in action. *Trends Cell Biol.* **14**, 628–638 (2004).
218. Fontana, M. F. *et al.* Macrophage colony stimulating factor derived from CD4+ T cells contributes to control of a blood-borne infection. *PLoS Pathog.* **12**, (2016).
219. Cerdan, C. *et al.* Activated but not resting T cells or thymocytes express colony-stimulating factor 1 mRNA without co-expressing c-fms mRNA. *Eur. J. Immunol.* **20**, 331–335 (1990).
220. Hallet, M. *et al.* Macrophage colony-stimulating factor (CSF-1) gene expression in human T-lymphocyte clones. *Blood* **77**, 780–786 (1991).
221. Zisman, E., Waisman, A., Ben-Yair, E. & Tartakovsky, B. Production of colony-stimulating factor 1 by T cells: possible involvement in their interaction with antigen-presenting cells. *Cytokine* **5**, 309–318 (1993).
222. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep.* **6**, (2014).
223. Hodge, G., Nairn, J., Holmes, M., Reynolds, P. N. & Hodge, S. Increased intracellular T helper 1 proinflammatory cytokine production in peripheral blood, bronchoalveolar lavage and intraepithelial T cells of COPD subjects. *Clin. Exp. Immunol.* **150**, 22–29 (2007).
224. Chang, Y. *et al.* CD8 positive T cells express IL-17 in patients with chronic obstructive pulmonary disease. *Respir. Res.* **12**, (2011).
225. Xu, W. H., Hu, X. L., Liu, X. F., Bai, P. & Sun, Y. C. Peripheral Tc17 and Tc17/interferon-γ cells are increased and associated with lung function in patients with chronic obstructive pulmonary disease. *Chin. Med. J. (Engl).* **129**, 909–916 (2016).
226. Zhu, X. *et al.* Peripheral T cell functions correlate with the severity of chronic obstructive pulmonary disease. *J. Immunol.* **182**, 3270–3277 (2009).

227. Shirai, T., Suda, T., Inui, N. & Chida, K. Correlation between peripheral blood T-cell profiles and clinical and inflammatory parameters in stable COPD. *Allergol. Int.* **59**, 75–82 (2010).
228. Nadigel, J. *et al.* Cigarette smoke increases TLR4 and TLR9 expression and induces cytokine production from CD8(+) T cells in chronic obstructive pulmonary disease. *Respir. Res.* **12**, (2011).
229. Barceló, B. *et al.* Intracellular cytokine profile of T lymphocytes in patients with chronic obstructive pulmonary disease. *Clin. Exp. Immunol.* **145**, 474–479 (2006).
230. Barczyk, A. *et al.* Cytokine production by bronchoalveolar lavage T lymphocytes in chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* **117**, 1484–1492 (2006).
231. Eppert, B. L., Wortham, B. W., Flury, J. L. & Borchers, M. T. Functional characterization of T cell populations in a mouse model of chronic obstructive pulmonary disease. *J. Immunol.* **190**, 1331–1340 (2013).
232. Lethbridge, M. W. *et al.* A novel technique to explore the functions of bronchial mucosal T cells in chronic obstructive pulmonary disease: application to cytotoxicity and cytokine immunoreactivity. *Clin. Exp. Immunol.* **161**, 560–569 (2010).
233. Walzl, G., Tafuro, S., Moss, P., Openshaw, P. J. M. & Hussell, T. Influenza virus lung infection protects from respiratory syncytial virus-induced immunopathology. *J. Exp. Med.* **192**, 1317–1326 (2000).
234. Yewdell, J. W. & Bennink, J. R. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.* **17**, 51–88 (1999).
235. EC, O. *et al.* A genetic risk score associated with chronic obstructive pulmonary disease susceptibility and lung structure on computed tomography. *Am. J. Respir. Crit. Care Med.* **200**, 721–731 (2019).
236. Wan, E. S. *et al.* Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respir. Res.* **15**, 1–13 (2014).
237. Gim, J. *et al.* A between ethnicities comparison of chronic obstructive pulmonary disease genetic risk. *Front. Genet.* **11**, 438794 (2020).
238. Han, M. L. K. *et al.* Racial differences in quality of life in patients with COPD. *Chest* **140**, 1169–1176 (2011).
239. Dransfield, M. T. & Bailey, W. C. COPD: Racial disparities in susceptibility, treatment, and outcomes. *Clin. Chest Med.* **27**, 463–471 (2006).
240. Ferdinand, K. C. Recommendations for the management of special populations: racial and ethnic populations. *Am. J. Hypertens.* **16**, 50S-54S (2003).
241. Nelson, H. S., Weiss, S. T., Bleecker, E. K., Yancey, S. W. & Dorinsky, P. M. The Salmeterol Multicenter Asthma Research Trial: a comparison of usual pharmacotherapy for asthma or usual pharmacotherapy plus salmeterol. *Chest* **129**, 15–26 (2006).
242. Ningappa, M. *et al.* A network-based approach to identify expression modules underlying rejection in pediatric liver transplantation. *Cell Reports Med.* **3**, 100605 (2022).
243. Das, J., Gayvert, K. M., Bunea, F., Wegkamp, M. H. & Yu, H. ENCAPP: Elastic-net-based prognosis prediction and biomarker discovery for human cancers. *BMC Genomics* **16**, 1–13 (2015).