Biomarker of Clinically Relevant Endpoints in Progressive Pulmonary Fibrosis

by

Justin Oldham

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Sciences)
in the University of Michigan
2025

Doctoral Committee:

Associate Professor Jennifer A. Smith, Chair
Associate Professor Alison Mondul
Professor Bethany B. Moore
Professor Susan Murray

Justin Oldham
oldhamj@umich.edu
ORCID iD: 0000-0003-4957-8869

© Justin Oldham 2025

Dedication

This dissertation is dedicated to all those with pulmonary fibrosis, who understand the importance of this work better than anyone.

Acknowledgements

I would like to extend my deepest gratitude to the individuals who have supported me throughout this journey. I thank Dr. Carlos Mendez de Leon, who facilitated my acceptance into the PhD program, and Dr. Jennifer Smith, who adopted me after Carlos departed for Georgetown. Without Jen's sustained support and excellent mentorship, none of this work would have been possible. I am also grateful to members of my dissertation committee - Susan Murray, Beth Moore and Alison Mondul – who provided invaluable guidance as I developed and pursued the ideas presented here.

I also want to thank those who helped set me on this career trajectory. Drs. Imre Noth and Mary Strek introduced me to biomedical research over a decade ago and were the driving force behind my decision to pursue a career in academic medicine. As chief of the Division of Pulmonary and Critical Care Medicine at UM, Dr. MeiLan Han provided the financial investment and protected time necessary for me to complete this PhD. Dr. Fernando J. Martinez facilitated my rise in the field, serving as an invaluable career sponsor and mentor. By recognizing the value of this Epidemiological skillset, Dr. Martinez has allowed me to push the envelope in a field that desperately needs new ideas and innovative approaches to longstanding knowledge gaps. Finally, I want to thank my wife Courtney, son Harrison and daughter Violet, who consider this my "hobby" but have provided laughs and love throughout.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables vii

List of Figures ix

Abstract xi

Chapter 1 Introduction 1

1.1 Overview 1

1.2 Background 2

1.3 Knowledge Gap 7

1.4 Cohorts 8

1.5 Proteomic Data Generation 8

1.6 Specific Aims and Hypotheses 9

Chapter 2 Proteomic Biomarkers of Survival in non-IPF Interstitial Lung Disease 10

2.1 Abstract 10

2.2 Introduction 10

2.3 Material and Methods 12

2.3.1 Cohorts 12

2.3.2 Measures 12

2.3.3 Statistical analysis 12

2.3.4 Overlap with IPF Biomarkers 13

2.3.5 Pathway Analysis 13

2.3.6 Power calculation 14
2.4 Results 14
2.4.1 Cohorts, baseline characteristics and outcomes 14
2.4.2 Protein Biomarkers of non-IPF ILD Survival 15
2.4.3 Overlap with IPF Prognostic Protein Biomarkers 15
2.4.4 Pathway analysis 16
2.5 Discussion 16
2.6 Conclusion 20
Chapter 3 Untangling the Role of Prognostic Biomarkers in Progressive Pulmonary Fibrosis 31
3.1 Abstract 31
3.2 Introduction 31
3.3 Material and Methods 33
3.3.1 Causal Framework 33
3.3.2 Cohorts 34
3.3.3 Measures 34
3.3.4 Statistical analysis 35
3.3.5 Protein Interaction Network Analyses 36
3.4 Results 36
3.4.1 Cohort Characteristics 36
3.4.2 Outcomes 37
3.4.3 Causal Mediation Analysis 37
3.4.4 Pathway and Interaction Network Analysis 38
3.5 Discussion 39
3.6 Conclusion 43
Chapter 4 Independence and Additive Performance of Interstitial Lung Disease Biomarkers 61

4.1 Abstract 61
4.2 Introduction 62
4.3 Material and Methods 63
4.3.1 Cohorts and Phenotypic Measures 63
4.3.2 Biomarker Measures 63
4.3.3 Outcome Measures 64
4.3.4 Statistical analysis 64
4.4 Results 65
4.4.1 Cohort Characteristics 65
4.4.2 Survival Analysis 66
4.4.3 Progression Analysis 67
4.5 Discussion 67
4.6 Conclusion 69
Chapter 5 Conclusions 81
5.1 Summary and Implications of Main Findings 81
5.1.1 Summary and Implications of Chapter 2 Findings 81
5.1.2 Summary and Implications of Chapter 3 Findings 82
5.1.3 Summary and Implications of Chapter 4 Findings 82
5.2 Strengths and Limitations 83
5.3 Future Directions 84
5.4 Conclusion 86
Appendix 87
Bibliography 94

List of Tables

Table 2.1. Baseline characteristics and outcomes for discovery and validation cohorts 21

Table 2.2. Selection of discovery cohort results ranked by FDR p-value..... 22

Table 2.3. Validated protein biomarkers of three-year TFS in discovery and validation cohorts 24

Table 2.4. Proteins nominally associated with TFS IPF and non-IPF cohorts stratified by direction of effect..... 25

Table 3.1. Baseline characteristics of IPF and non-IPF ILD cohorts..... 44

Table 3.2. Mean one-year relative decline in FVC and unadjusted association with restricted mean survival time and transplant-free survival..... 45

Table 3.3. Proteins classified as potentially causal of PPF based on mediation by declining FVC..... 46

Table 3.4. Natural direct effect for mediated proteins when using different stratification thresholds..... 47

Table 3.5. E-values for proteins found to be potentially causal of PPF in IPF cohort..... 48

Table 3.6. Results of traditional association study testing the relationship between mediated proteins and categorical FVC decline. 49

Table 3.7. Causal mediation analysis results for selection of mediated IPF proteins when tested in non-IPF ILD cohort. 50

Table 3.8. Causal mediation analysis results for proteins specific to non-IPF ILD survival when tested in non-IPF ILD cohort. 51

Table 3.9. E-values for proteins found to be potentially causal of PPF in non-IPF ILD cohort.. 52

Table 4.1. Biomarker rank by study..... 70

Table 4.2. Baseline clinical characteristics and biomarker measures 71

Table 4.3. Iterative adjusted modeling results to identify independent biomarkers of TFS 72

Table 4.4. Association between prognostic biomarkers and three-year TFS according to time from high resolution computed tomography (HRCT) to blood draw..... 73

Table 4.5. TFS model building results 74

Table 4.6. Condition Index based collinearity diagnostics and proportion of variance. 75

Table 4.7. Iterative adjusted modeling results to identify independent biomarkers of one-year FVC decline 76

Table 4.8. FVC decline model building results 77

List of Figures

Figure 2.1. Survival in discovery and validation cohorts (a) and for ILD subtypes in the discovery (b) and validation (c) cohorts. 26

Figure 2.2. Survival in those with connective tissue disease-associated ILD (a), fibrotic hypersensitivity pneumonitis (b) and non-IPF idiopathic interstitial pneumonia (c). 27

Figure 2.3. Volcano plot displaying the association between protein biomarkers and three-year transplant-free survival in a non-IPF ILD discovery cohort. 28

Figure 2.4. Plot of p-values for test heterogeneity in outcome association for validated biomarkers across non-IPF ILD subtypes for validated prognostic biomarkers across discovery and validation cohorts. The horizontal and vertical dotted lines correspond to unadjusted p-value threshold. 29

Figure 2.5. Molecular pathways unique to non-IPF ILD (A), IPF (B) and shared across conditions (C), along with a Venn diagram (D) depicting the number of shared and unique biomarkers for these conditions. The vertical dotted line corresponds to false discovery rate corrected p-value threshold. 30

Figure 3.1. Causal framework depicting an association between a prognostic protein biomarker and ILD survival through a causal, PPF-related pathway (a), causal, PPF-unrelated pathway (b) and non-causal pathway due to residual confounding (c). Within this framework, there exists no direct effect of a protein in death/transplant. Abbreviation: U_M = unmeasured mediator; U_L = unmeasured confounder. 53

Figure 3.2. Directed acyclic graph showing the relationship between protein biomarker (exposure) and death/transplant (outcome) through FVC decline (mediator). Abbreviations: L_{AY} = confounders of the exposure-outcome relationship; L_{AM} = confounders of the exposure-mediator relationship; L_{MY} = confounders of the mediator-outcome relationship; L_{AMY} = confounders of the mediator-outcome relationship caused by exposure. 54

Figure 3.3. Case finding diagram for chapter 3 cohort. Abbreviations: FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; ILD = interstitial lung disease. 55

Figure 3.4. Pathway analysis identifying molecular pathways overrepresented by proteins classified as causally related to PPF in an IPF cohort. 56

Figure 3.5. Inference-based protein network analyses. Red proteins exert a direct effect on declining FVC, while blue proteins are upstream of red proteins. 57

Figure 3.6. Results of inference-based protein-protein interaction network analysis in IPF cohort. Red fields indicate that the effect on FVC decline of the exposure protein (x-axis) is completely mediated by the mediator protein (y-axis). Blue fields indicate that the effect on FVC decline of the exposure protein (x-axis) is not mediated by the mediator protein (y-axis). Partial mediation is not shown. 58

Figure 3.7. Biology-based protein network analyses identified through String software at confidence level of 0.25. 59

Figure 3.8. Results of inference-based protein-protein interaction network analysis in non-IPF ILD cohort (a) and resulting direct causal network (b). Red fields indicate that the effect on FVC decline of the exposure protein (x-axis) is completely mediated by the mediator protein (y-axis). Blue fields indicate that the effect on FVC decline of the exposure protein (x-axis) is not mediated by the mediator protein (y-axis). Partial mediation is not shown. Red proteins exert a direct effect on declining FVC, while blue proteins are upstream of red proteins. 60

Figure 4.1. Case finding diagram for chapter 4 cohort. 78

Figure 4.2. Correlation matrix for continuous clinical variables and biomarkers. Size of circles and numerical values represent correlation coefficient. Blank cells denote non-significant correlation ($p<0.05$). 79

Figure 4.3. Variable importance plot showing the relative contribution to TFS discrimination for variables comprising the final model. Larger numerical decrease in concordance index denotes more importance for survival risk discrimination. 80

Abstract

Interstitial lung disease (ILD) comprises a diverse group of rare diffuse pulmonary conditions characterized by varying degrees of inflammation and scarring. Progressive pulmonary fibrosis (PPF) is a common and deadly manifestation of ILD leads to relentless decline in lung function and premature death. With a median survival of 3-5 years, PPF mortality outpaces most cancers. Variability in the incidence of PPF has led to reactive treatment approaches, with most patients needing to display evidence of progression before qualifying for therapy proven to slow lung function decline. This reality misses a valuable opportunity to prevent progression but is currently necessary given our inability to predict a progressive phenotype. The objective of this dissertation is to begin addressing this knowledge gap through dedicated proteomic investigation in patients with idiopathic pulmonary fibrosis (IPF) and other forms of non-IPF ILD that are at high risk for developing PPF. Our central hypothesis was that proteomic investigation will uncover novel prognostic biomarkers of PPF with mechanistic relevance that are independent of genomic and imaging biomarkers. This hypothesis was informed by my intimate understanding of the disease process and longstanding care I have provided to patients with PPF and their families.

In Aim 1 we conducted a high-throughput proteomic investigation of protein biomarkers of non-IPF ILD survival. Of 73 proteins associated with transplant-free survival in a discovery cohort (n=676), 44 showed sustained survival association in an independent validation cohort (n=616). Pathway analysis identified several molecular pathways shared with IPF, along with three molecular pathways unique to non-IPF ILD.

In Aim 2, we performed causal mediation analysis to identify prognostic protein biomarkers are associated with survival through a causal pathway mediated by declining forced vital capacity (FVC), a cardinal feature of PPF. In an IPF cohort comprising 904 individuals, we identified 28 with survival effect that was mediated by declining lung function, suggesting a potentially causal relationship with progressive pulmonary fibrosis. Of these 28 proteins, 7 were also mediated by

declining FVC in a non-IPF ILD cohort consisting of 974 individuals, suggesting shared PPF pathobiology across ILD subtypes. A traditional association study was then performed and demonstrated an association between 26/28 mediated proteins and ≥10% categorical decline in FVC, a well validated measure of ILD progression and commonly used clinical endpoint.

In Aim 3, we analyzed genomic, proteomic and radiomic data for 524 individuals to determine whether seven known prognostic biomarkers (telomere length, five proteins and quantitative fibrosis extent on chest imaging) were independent in multi-marker modeling. While all biomarkers were associated with TFS when adjusting for clinical covariates, only three (integrin beta 6 protein (ITGB6), collagen 24a1 protein (COL24A1) and quantitative fibrosis extent) maintained survival association in multi-marker modeling. A multi-modal prediction model comprised of ITGB6, COL24A1, quantitative fibrosis extent and diffusion capacity better discriminated survival than unimodal models; however, less complex, bimodal models performed similarly as the best trimodal model.

These findings advance our understanding of proteins and pathobiology driving PPF and underscore the potential for multi-marker modeling to improve our ability to discriminate clinically relevant outcomes. This work lays the foundation for future studies aimed at therapeutic targeting of proteins found to be causally related to PPF and developing protein and multi-modal tools to predict progression in this population.

Chapter 1 Introduction

1.1 ***Overview***

Fibrotic interstitial lung disease (ILD) comprises a diverse group of rare but deadly diffuse parenchymal disorders characterized by varying degrees of pulmonary inflammation and scarring.(Travis et al., 2013) Progressive pulmonary fibrosis is a feared manifestation of ILD, leading to declining lung function and premature death.(Pugashetti et al., 2023) Idiopathic pulmonary fibrosis (IPF) has long been regarded as the prototypical PPF (du Bois, 2012; Ley et al., 2011) and for decades has been the focus of most of ILD-related publications, clinical trials and research funding. While this focus greatly advanced our understanding of IPF pathobiology and has resulted in therapeutic breakthroughs,(King et al., 2014; Richeldi et al., 2014) it largely neglected other forms of fibrotic ILD, including connective tissue disease-associated ILD, fibrotic hypersensitivity pneumonitis, and non-IPF idiopathic interstitial pneumonia.

Non-IPF ILDs collectively account for the majority of ILD cases(Adegunsoye et al., 2019; Cottin et al., 2018; Duchemann et al., 2017; Patterson et al., 2017) and though numerous studies have shown that non-IPF ILD survival is favorable compared to IPF, we recently found that those with non-IPF ILD who develop PPF exhibit a natural history that is largely indistinguishable from IPF.(Pugashetti et al., 2023) Moreover, the cumulative incidence of disease progression among patients with non-IPF ILD was found to be roughly similar to the cumulative incidence of newly diagnosed IPF.(Pugashetti et al., 2023) These findings suggest that approximately half of patients with a fatal form of ILD have been understudied, a research disparity that was highlighted during a recent Pulmonary Fibrosis Stakeholder Summit held by the National Heart, Lung and Blood Institute, Three Lakes Foundation and Pulmonary Fibrosis Foundation.(Montesi et al., 2024)

Recent work has begun to address this critical research disparity. Notably, the a recent phase III clinical trial showed that nintedanib slows lung function decline in those with non-IPF PPF with similar effect as those with IPF.(Flaherty et al., 2019) We subsequently identified a number of

novel circulating biomarkers associated with one-year progression in patients with non-IPF ILD, which like the trial above, showed similar effect association irrespective of non-IPF ILD subtype.(Bowman et al., 2022; Wells et al., 2020) While our prior study shed important light on the pathobiology underpinning non-IPF ILD, it was limited by a relatively small sample size and diversity of protein biomarkers tested, along with limited follow-up time. To better understand the molecular underpinnings of PPF, larger studies of disease relevant proteins are needed.

This dissertation aimed to utilize new high-throughput proteomic technology to begin addressing this critical knowledge gap. First, we identified and validate circulating plasma protein biomarkers of three-year transplant-free survival in patients with non-IPF ILD. Next, we performed causal mediation analysis to determine which proteins associate with survival through a causal pathway characterized by declining lung function, a cardinal feature of PPF. Finally, we determined whether established prognostic biomarkers in ILD are independent and augment outcome discrimination in multi-modal prediction modeling. Together these studies improved our understanding of the pathobiology of PPF and the relative importance of prognostic biomarkers.

1.2 ***Background***

Fibrotic ILD comprises a diverse group of rare but deadly diffuse parenchymal disorders characterized by varying degrees of pulmonary inflammation and scarring.(Travis et al., 2013) After years of pre-clinical disease, patients typically present with new onset shortness of breath on exertion and chronic cough. Upon evaluation, fibrotic ILD is typically diagnosed by chest imaging, specifically high-resolution computed tomography, but a surgical lung biopsy is required in a subset of patients to establish a histopathological diagnosis when chest imaging is non-diagnostic.(Travis et al., 2013) Though fibrotic ILDs display overlapping clinical, morphological, and histopathological features, disease classification is performed according to etiology, with idiopathic forms being most common.

The most frequently encountered fibrotic ILD is idiopathic pulmonary fibrosis (IPF), which accounts for roughly 35% of ILD cases evaluated through expert centers. IPF was the first fibrotic ILD described over forty years ago based on key histopathological features.(Liebow AA, 1969) For this reason, ILDs are frequently categorized as being IPF or a non-IPF ILD (Figure 1-1),(Cottin

et al., 2018) which has prognostic implications and impacts treatment approach (see below). The most common forms of non-IPF ILD include connective tissue disease-associated ILD, which accounts for roughly 30% of cases, fibrotic hypersensitivity pneumonitis, which accounts for roughly 15% of cases and non-IPF idiopathic interstitial pneumonia, which accounts for roughly 15% of cases.(Adegunsoye et al., 2019; Cottin et al., 2018; Duchemann et al., 2017; Patterson et al., 2017) Many additional non-IPF ILDs exist but are exceedingly rare, comprising less than 1% of the caseload evaluated through most ILD programs, which in aggregate comprise ~5% of all ILD cases evaluated. Common ILD subtypes are described in further detail below.

Idiopathic Pulmonary Fibrosis (IPF) is defined as defined as a specific form of chronic, progressive fibrosing interstitial pneumonia of unknown cause, occurring primarily in older adults, and limited to the lungs.(Raghu, Collard, Egan, Martinez, Behr, Brown, Colby, Cordier, Flaherty, Lasky, Lynch, Ryu, Swigris, Wells, Ancochea, Bouros, Carvalho, Costabel, Ebina, Hansell, Johkoh, Kim, King, Kondoh, Myers, Muller, Nicholson, Richeldi, Selman, Dudden, Griss, Protzko, Schunemann, et al., 2011) Despite being the most common form of fibrotic ILD, IPF remains relatively rare, with an estimated prevalence of 14-63 per 100,000 persons and incidence of 7-17 cases per 100,000 person years depending on the case definition used.(Fernández Pérez et al., 2010; Raghu et al., 2006) Considered a disease of aging, IPF typically develops in the seventh decade of life or later and has a 3:1 male to female predominance.(Oldham & Noth, 2014) Among the strongest risk factors for IPF is a history of cigarette smoking, with approximately 60-70% of individuals with IPF being former cigarette smokers.(Baumgartner et al., 1997; Wang et al., 2020) Family history is also a strong risk factor, as up to 20% of IPF cases are considered familial and presence of familial disease was recently shown to predict worse survival in IPF and other fibrosing ILDs.(Cutting et al., 2021; Kropski, 2020)

Nearly all patients with IPF develop progressive pulmonary fibrosis (PPF), resulting in high mortality and a median survival of 3-5 years following diagnosis, which outpaces most cancers.(du Bois, 2012; Ley et al., 2011) Consequently, IPF has long been considered the prototypical PPF. Numerous interventional clinical trials have been performed in patients with IPF to date.(Adegunsoye et al., 2019; Patterson et al., 2017; Raghu, 2017) These trials have resulted in the approval of two anti-fibrotic therapies, pirfenidone and nintedanib, for the treatment of IPF

after phase III trials demonstrated efficacy in slowing lung function decline when compared to placebo.(King et al., 2001; Richeldi et al., 2014) Nintedanib is a tyrosine kinase inhibitor that selectively inhibits receptors for platelet derived growth factor, fibroblast growth factor, and vascular endothelial growth factor, which are known to play critical roles in fibrogenesis.(Wollin et al., 2015) The precise mechanism of action for pirfenidone is unknown, but this agent has been shown to block transforming growth factor beta, a central driver of pulmonary fibrosis onset and progression.(Ruwanpura et al., 2020) While approval of these drugs to treat IPF represent a major advance for the field, neither drug stops IPF progression and their effect on survival appears modest.(Dempsey et al., 2019; Suissa & Suissa, 2022) Lung transplant remains the only curative intervention for IPF and other progressive ILDs.(Bjoraker et al., 1998; Raghu et al., 2014; Raghu, Collard, Egan, Martinez, Behr, Brown, Colby, Cordier, Flaherty, Lasky, Lynch, Ryu, Swigris, Wells, Ancochea, Bouros, Carvalho, Costabel, Ebina, Hansell, Johkoh, Kim, King, Kondoh, Myers, Muller, Nicholson, Richeldi, Selman, Dudden, Griss, Protzko, & Schunemann, 2011; Riddell et al., 2020) However, donor lungs remain scarce and many patients die awaiting transplant.(Bennett et al., 2015; George et al., 2011)

Connective Tissue Disease-associated ILD (CTD-ILD) affects 5-8% of the US adult population and is characterized by autoantibody production and immune-mediated tissue destruction. ILD complicates several common forms of CTD, including rheumatoid arthritis, systemic sclerosis and idiopathic inflammatory myopathy, (Figure 1-2).(Fathi et al., 2004; Gabbay et al., 1997; Walker et al., 2007) ILD can also develop in Sjogren's syndrome, systemic lupus erythematosus and mixed connective tissue disease, but is less common (Figure 1-2).(Castelino & Varga, 2010; Flament et al., 2016; Reiseter et al., 2018) The true prevalence of ILD in these conditions remains unknown due to the lack of systematic screening, but prospective studies are currently ongoing to confirm prior estimates. As opposed to IPF, CTD-ILD predominantly affects women and younger individuals.(Fischer & du Bois, 2012)

It is thought chronic systemic inflammation characteristic of most CTDs leads to pulmonary inflammation and subsequent fibrosis.(Fischer & du Bois, 2012) Given the inflammatory nature of CTD, immunosuppressant therapy is considered first line therapy.(Fischer & du Bois, 2012; Wijsenbeek & Cottin, 2020) Randomized controlled trials assessing immunosuppressant therapy

efficacy have been conducted in patients with scleroderma-associated ILD, establishing efficacy for cyclophosphamide in stabilizing lung function along with non-inferiority for mycophenolate mofetil and rituximab.(Maher, Tudor, et al., 2023; Tashkin et al., 2006; Tashkin et al., 2016) Retrospective CTD-ILD case series have suggested similar efficacy for immunosuppressant therapy in CTD-ILDs other than scleroderma.(Fischer et al., 2013; Oldham et al., 2016) Most immunosuppressant drugs currently used to treat CTD-ILD and other inflammatory ILDs exert an effect on the adaptive immune system through suppression of B and T cells,(Allison, 2005; Cerny et al., 2002) and inflammation-producing cytokines.(Okuda, 2008) Because some patients will manifest quiescent disease, the decision to begin immunosuppressant therapy is an individualized one without consensus guidance.(Wijsenbeek & Cottin, 2020) While immunosuppressant therapy may lead to disease stability in a majority of patients, a significant subset still develops PPF.(Adegunsoye et al., 2017; Duarte et al., 2019; Fischer et al., 2013; Huapaya et al., 2019; Keir et al., 2012; Morisset et al., 2017; Oldham et al., 2016; Sharma et al., 2017; Tashkin et al., 2006; Tashkin et al., 2016; Witt et al., 2016) Such patients, whose disease is likely to be fibrotic predominant, now qualify for anti-fibrotic therapy with nintedanib, which was recently showed to slow lung function decline in patients with diverse forms of PPF, including CTD-ILD.(Flaherty et al., 2017; Flaherty et al., 2019) Nintedanib has weak anti-inflammatory properties so is typically reserved for those who first fail anti-inflammatory therapy aimed at suppressing the immune system.

Fibrotic Hypersensitivity Pneumonitis (fHP) results from an inappropriate immune response. But rather than an autoantibody-mediated process, fHP appears to result from an overexuberant immune reaction to repeated inhalation of an organic antigen, usually of avian or mold origin.(Selman et al., 2012) This is supported by genetics investigation showing that polymorphisms in genes comprising the major histocompatibility complex, which is a critical mediator of the immune response, are associated with increased fHP susceptibility.(Aquino-Galvez et al., 2008; Camarena et al., 2001; Joshi et al., 2009) Once alveolar injury occurs, a number of cytokines and other inflammatory proteins lead to parenchymal inflammation, including a classic granulomatous reaction.(Barrera et al., 2008; Blanchet et al., 2011; Bustos et al., 2007; Joshi et al., 2009) Sustained injury and granulomatous inflammation eventually leads to fibrotic remodeling in a subset of patients.(Garcia de Alba et al., 2015; Jimenez-Alvarez et al., 2011; Pardo

et al., 2000; Pardo et al., 2001; Salisbury et al., 2019; Selman et al., 1988) Like CTD-ILD, epidemiological data are sparse for fHP, but a recent study suggested that fHP affects 2-3 per 100,000 persons, which increases to 11 per 100,000 persons in individuals over the age of 65.(Fernandez Perez et al., 2018) In this study an annual cumulative incidence rate of 1-2 per 100,000 persons was observed.

No randomized controlled trials have been performed in patients with fHP, though corticosteroid therapy was shown in a randomized controlled trial to improve time to resolution in those with acute HP, which can occur in the setting of high antigen burden. Retrospective fHP case series have shown conflicting data for the role of immunosuppressant therapy in patients with fHP, though immunosuppression remains first line therapy given the underlying inflammatory changes characteristic of fHP. Like other non-IPF ILDs, those with fHP who develop PPF despite immunosuppression now qualify for anti-fibrotic therapy.(Flaherty et al., 2017; Flaherty et al., 2019) Whether anti-fibrotic therapy may serve as a better first-line therapy for fHP is unclear.

Non-IPF idiopathic interstitial pneumonia comprises a number of rare idiopathic ILDs with unclear epidemiology. (Travis et al., 2013; Travis et al., 2008) The most common form of non-IPF idiopathic interstitial pneumonia is unclassifiable ILD, which captures those for whom a confident ILD diagnosis cannot be established.(Ryerson et al., 2017; Ryerson et al., 2013) This may occur due to several reasons, including 1) non-diagnostic chest imaging in a patient unable (due to advanced disease) or unwilling to undergo surgical lung biopsy to confirm the diagnosis, 2) two competing diagnoses (such as connective tissue disease-associated ILD and fibrotic hypersensitivity pneumonitis) that are felt to be equally likely and 3) a non-diagnostic surgical lung biopsy. The optimal treatment approach for those with unclassifiable ILD remains unclear. Some patients will receive immunosuppressant therapy, while others will be observed and placed on anti-fibrotic therapy if progression is noted.(Wijsenbeek & Cottin, 2020) Two randomized controlled trials have been performed for those with unclassifiable ILD with PPF, one focused on unclassifiable ILD and another that included PPF of any ILD subtype. The unclassifiable ILD focused trial suggested that pirfenidone is likely effective at slowing lung function decline in progressive unclassifiable ILD but failed to meet the primary endpoint.(Maher et al., 2018) The second, trial showed that nintedanib effectively slowed lung function decline in patients with

unclassifiable ILD and other forms of PPF, irrespective of underlying ILD subtype.(Flaherty et al., 2019)

Idiopathic non-specific interstitial pneumonia (NSIP) is another commonly encountered non-IPF idiopathic interstitial pneumonia but remains quite rare outside of expert centers.(Travis et al., 2008) This classification is reserved for those with classic chest imaging and histopathological features without a clear etiology for these imaging and histopathological features. (Travis et al., 2008) Given the strong link between imaging and histopathological features of NSIP and a diagnosis of connective tissue disease, some have proposed that idiopathic NSIP be considered an undifferentiated connective tissue disease or an interstitial pneumonia with autoimmune features when other clinical and serological features of connective tissue disease are present.(Fischer et al., 2015) Given the strong link between NSIP and connective tissue disease, immunosuppressant therapy is considered first line for those with idiopathic NSIP.(Wijsenbeek & Cottin, 2020) A recent randomized controlled trial supported this approach, showing that both mycophenolate mofetil and rituximab stabilized lung function in those with idiopathic NSIP.(Mankikian et al., 2023) As with other conditions, those with progressive idiopathic NSIP qualify for anti-fibrotic therapy, though subgroup analysis suggested that those with NSIP experienced less treatment effect than others in this trial. (Brown et al., 2020; Wijsenbeek & Cottin, 2020)

1.3 ***Knowledge Gap***

IPF-focused investigations account for the overwhelming majority of ILD-related publications to date and the National Institutes of Health has allocated more research dollars for IPF than all other ILDs combined. While this approach acknowledges the deadly nature of IPF, it fails to appreciate the societal importance of progressive non-IPF ILD, which has similar incidence and natural history as IPF (Figure 1-3).(Pugashetti et al., 2023) Our collective focus on IPF has largely ignored roughly half of patients with a fatal form of PPF over the last 40 years since ILD was first described.

While non-IPF ILDs have different causes, recent studies have shown that lung function decline and survival are similar once progressive.(Pugashetti et al., 2023; Wells et al., 2020) Moreover, prior molecular investigations have shown that circulating prognostic biomarkers similarly

discriminate outcomes for connective tissue disease, fibrotic hypersensitivity pneumonitis and non-IPF idiopathic interstitial pneumonia,(Alqalyoobi et al., 2020; Bowman et al., 2022) suggesting that shared pathobiology underpins PPF, irrespective of disease subtype. The objective of this dissertation is to begin addressing this research disparity and knowledge gap through dedicated molecular investigation of PPF due to both IPF and non-IPF ILD.

1.4 ***Cohorts***

Consecutive patients with ILD, including IPF, connective tissue disease, fibrotic hypersensitivity pneumonitis, and non-IPF idiopathic interstitial pneumonia and available survival data who provided a research blood draw as part of prospective cohort studies at the University of California-Davis (UCD) (recruitment dates May 2016 – February 2021), University of California-San Francisco (UCSF) (recruitment dates October 2004 – February 2019) and University of Texas Southwestern (UTSW) (recruitment dates March 2007 – August 2019) and the PRECISIONS multi-omics study were included.(Podolanczuk et al., 2022) The PRECISIONS multi-omic cohort consisted of similarly identified patients participating in the Pulmonary Fibrosis Foundation Patient Registry (PFFPR) (recruitment dates March 2016 – June 2018) and site registries administered through the University of Chicago (UChicago) (recruitment dates March 2007 – July 2017) and University of Virginia (UVA) (recruitment dates September 2018 – November 2021). Site-specific ethics protocols were approved at UC-Davis (protocol #875917), UCSF (protocol #10-01592) UTSW (protocol #092017-007 and 082010-127), UChicago (protocol #13-1180) and UVA (protocol #20937), while the protocol for PFF Patient Registry participation was approved by institutional review boards at each participating site. Patients from UCSF, UTSW, UChicago and UVA who participated in the PFFPF were excluded to ensure each dataset comprised unique patients. The University of California-Davis did not contribute patients to the PFF Patient Registry.

1.5 ***Proteomic Data Generation***

Peripheral blood was collected in consenting patients at each participating center and plasma isolated according to center-specific protocols. After isolation, plasma was aliquoted and stored at -80 degree Celsius without thawing. All discovery and validation plasma samples were then shipped to Olink® (Uppsala, Sweden) US headquarters in Boston, MA for processing as a single

batch. Samples were randomized at the time of plating according to center, age, sex and race to mitigate batch effects. The Explore 3072 panel was used to generate semi-quantitative proteomic data for 2921 analytes. This platform uses proximity extension assays, which require two matched antibodies labelled with unique DNA oligonucleotides to bind the target protein and then hybridize as part of a DNA polymerase-dependent extension step. This step creates a double-stranded DNA unique to the target protein, which is subsequently amplified and indexed in a manner that is quantitatively proportional to the plasma concentration of the target protein. Amplified targets are then quantified with Next Generation Sequencing.(Assarsson et al., 2014) Intensity normalization is used to aggregate data across plates. Normalized protein values are then $\log_2$ transformed to ensure normality and returned via a download link. Proteins below the lower limit of detection were imputed to the lowest observed value for each respective protein.

1.6 ***Specific Aims and Hypotheses***

The objective of this prospectus is to begin addressing key knowledge gaps in patients with PPF. Using proteomic data generated through our involvement in several ongoing cohort studies, we pursued three critical projects to improve our understanding of PPF pathobiology. Our central hypothesis is that proteomic investigation will uncover novel ILD biomarkers of ILD survival with mechanistic relevance and that are independent of known genomic and imaging biomarkers. To address our central hypothesis, we pursued the following specific aims:

Aim 1. Identify proteomic biomarkers of non-IPF ILD survival. We hypothesized that unbiased proteomic investigation would identify novel protein biomarkers of non-IPF ILD survival and that the survival association for validated biomarkers would be robust to ILD diagnosis.

Aim 2. Identify proteins that play a potentially causal role in PPF. We hypothesized that causal mediation analysis will identify circulating plasma biomarkers associated with ILD survival through a causal, PPF-related pathway.

Aim 3. Compare test performance between prognostic proteomic, genomic, and radiologic biomarkers of PPF. We hypothesized that multi-modal biomarkers of ILD survival will be independent of one another and provide additive outcome discrimination.

Chapter 2 Proteomic Biomarkers of Survival in non-IPF Interstitial Lung Disease

2.1 ***Abstract***

While idiopathic pulmonary fibrosis (IPF) has been widely studied, progressive non-IPF interstitial lung disease (ILD) remains poorly understood. Here we aimed to identify and validate proteomic biomarkers of non-IPF ILD survival. High-throughput proteomic data were generated using plasma collected as part of prospective registries at the Universities of California and Texas (discovery cohort, n=676) and PRECISIONS multi-omic study (validation cohort, n=616). Proteins associated with three-year transplant-free survival (TFS) were identified using multivariable Cox proportional hazards regression, and those associated with TFS after adjustment for false discovery were advanced for validation cohort testing. Pathway analysis was performed to identify molecular pathways unique to non-IPF ILD and shared with IPF. Of 2925 proteins tested in the discovery cohort, 73 were associated with TFS, with 44 showing sustained TFS association in the validation cohort. The top TFS-associated proteins were amphiregulin (HR 2.51, 95% CI 2.07-3.04), integrin subunit beta 6 (HR 2.46; 95% CI 1.95-3.10) and keratin 19 (HR 1.70, 95% CI 1.47-1.98). All but one validated biomarkers showed consistent TFS association across non-IPF ILD subtypes. Pathway analysis identified several molecular pathways shared with IPF, along with three pathways unique to non-IPF ILD. Here we identified and validated novel prognostic protein biomarkers in non-IPF ILD, most of which showed consistent association across non-IPF ILD subtypes. While most biomarkers and molecular pathways identified were previously linked to IPF, several were unique to non-IPF ILD, suggesting that unique biology may contribute to progressive non-IPF ILD.

2.2 ***Introduction***

Fibrotic interstitial lung disease (ILD) comprises a diverse group of rare but deadly diffuse parenchymal disorders characterized by varying degrees of pulmonary inflammation and scarring.(Travis et al., 2013) Idiopathic pulmonary fibrosis (IPF) is widely regarded as the most common and deadly fibrotic ILD(du Bois, 2012; Ley et al., 2011) and for decades has been the

focus of most of ILD-related publications, clinical trials and research funding. While this focus greatly advanced our understanding of IPF pathobiology and has resulted in therapeutic breakthroughs,(King et al., 2014; Richeldi et al., 2014) it largely neglected other forms of fibrotic ILD, including connective tissue disease-associated ILD, fibrotic hypersensitivity pneumonitis, and non-IPF idiopathic interstitial pneumonia.

Non-IPF ILDs collectively account for the majority of ILD cases(Adegunsoye et al., 2019; Cottin et al., 2018; Duchemann et al., 2017; Patterson et al., 2017) and though numerous studies have shown that non-IPF ILD survival is favorable compared to IPF, we recently found that those with non-IPF ILD who develop progressive pulmonary fibrosis (PPF) exhibit a natural history that is largely indistinguishable from IPF.(Pugashetti et al., 2023) Moreover, the cumulative incidence of PPF among patients with non-IPF ILD was found to be roughly similar to the cumulative incidence of newly diagnosed IPF.(Pugashetti et al., 2023) These findings suggest that approximately half of patients with PPF have been understudied, a research disparity that was highlighted during a recent Pulmonary Fibrosis Stakeholder Summit held by the National Heart, Lung and Blood Institute, Three Lakes Foundation and Pulmonary Fibrosis Foundation.(Montesi et al., 2024)

Recent work has begun to address this critical research disparity. Notably, the INBUILD trial showed that nintedanib slows lung function decline in those with PPF with similar effect as those with IPF.(Flaherty et al., 2019) We subsequently identified a number of novel circulating biomarkers associated with one-year progression in patients with non-IPF ILD, which like INBUILD, showed similar effect association irrespective of non-IPF ILD subtype.(Bowman et al., 2022; Wells et al., 2020) While our prior study shed important light on the pathobiology underpinning PPF in those with non-IPF ILD, it was limited by a relatively small sample size and diversity of protein biomarkers tested, along with limited follow-up time. To better understand the molecular underpinnings of PPF, larger studies of disease relevant proteins are needed. Here, we utilized new high-throughput proteomic technology to identify and validate circulating plasma protein biomarkers of three-year transplant-free survival in patients with non-IPF ILD. We hypothesized that unbiased proteomic investigation would identify novel biomarkers of PPF in those with non-IPF ILD and that the survival association for validated biomarkers would be robust to ILD diagnosis. We then compared our findings to a similarly designed and recently published

study in IPF (Oldham et al., 2023) to identify prognostic biomarkers and molecular pathways unique to non-IPF ILD and those shared with IPF.

2.3 ***Material and Methods***

2.3.1 ***Cohorts***

Cohorts are described above. Those with non-IPF ILD from the Universities of California (Davis and San Francisco) and Texas comprised the discovery cohort (n=676). Those with non-IPF ILD from the PRECISIONS multi-omic study comprised the validation cohort (n=616).

2.3.2 Measures

Baseline demographics (age, sex, race), ILD subtype (connective tissue disease-associated ILD, fibrotic hypersensitivity pneumonitis and non-IPF idiopathic interstitial pneumonia), smoking history (ever vs. never), percent predicted forced vital capacity (FVC), percent predicted diffusion capacity of the lung for carbon monoxide (DLCO), and binary exposure to one or more immunosuppressive medications (mycophenolate mofetil, azathioprine, rituximab, cyclophosphamide, or prednisone of any dose) at the time of blood draw was collected. DLCO was imputed to the lowest quartile mean of 25% predicted when missing due to the inability to perform the maneuver (n=14). Those with missing DLCO data due to lack to attempt were excluded from the discovery (n=2) and validation (n=8) cohorts. Complete data were otherwise available for all covariates included in modeling.

2.3.3 Statistical analysis

The primary endpoint assessed was three-year transplant-free survival (TFS), defined as the time from blood draw to death, lung transplant or censoring. Censoring was performed at 36 months or the date of last verifiable healthcare interaction when lost to follow-up before 36-months. Biomarkers were modeled continuously, with TFS association reported for each unit change in biomarker concentration. TFS was modeled using a multivariable Cox proportional hazards regression model adjusted for potential confounders of the biomarker-TFS association. These included ILD subtype (connective tissue disease-associated ILD vs. other)(Pugashetti et al., 2023), age, sex, race, smoking history, percent predicted FVC and DLCO, and exposure to one or more immunosuppressive medications at the time of blood draw. Nintedanib exposure was not

included as a covariate in the model, as nearly all patients were recruited prior to its approval for the treatment of progressive non-IPF ILD.

Biomarkers associated with TFS in the discovery cohort at false discovery (FDR) $p<0.05$ were advanced for testing in the validation cohort. A volcano plot was used to display results graphically. Biomarkers maintaining TFS association with consistent direction of effect in the validation cohort at FDR $p<0.05$ were considered validated. FDR p-values were generated using the Benjamini-Hochberg procedure and were based on 2925 tests in the discovery cohort (all proteins) and 73 tests in the validation cohort (advanced proteins only). To test our primary hypothesis that biomarkers of survival are robust to non-IPF ILD diagnosis, we then fit a Cox proportional hazards model with the aforementioned clinical covariates (age, sex, race, smoking history, FVC %, DLCO %, and treatment exposure) along with a biomarker-by-diagnosis interaction term, with ILD diagnosis modeled categorically and connective tissue disease-associated ILD treated as the reference group. A Chi-square test of heterogeneity(Kreuter & Beddo, 2004) was performed to assess whether there was a difference in biomarker-outcome association across any of the diagnostic subgroups tested. Biomarkers with test of heterogeneity $p<0.05$ across both discovery and validation cohorts were considered to have differential effect association across non-IPF ILD subtypes. Tests of heterogeneity were also performed to evaluate whether treatment exposure modified the biomarker-TFS association. All analyses were performed using R (version 4.2.2) and STATA (StataCorp. 2023. Release 18. College Station, TX).

2.3.4 Overlap with IPF Biomarkers

To identify prognostic biomarkers unique to non-IPF ILD and those shared with IPF, we then used the model above to identify biomarkers associated with three-year TFS at a nominal statistical threshold of $p<0.05$. Those associated with TFS using this threshold and with consistent direction of effect across non-IPF ILD discovery and validation cohorts were cross-referenced with those from our recent study of biomarkers of IPF survival identified using the same approach.(Oldham et al., 2023)

2.3.5 Pathway Analysis

Pathway analysis was then performed using Ingenuity Pathway Analysis (QIAGEN Inc., Redwood City, CA) to identify molecular pathways unique to each condition and those that overlap across conditions. Overrepresented pathways were identified using a right-tailed Fisher's exact test adjusted for multiple testing with the Benjamini-Hochberg procedure, taking into account the number of overlapping proteins and the total number of proteins known in each pathway. We report the top five pathways in each IPF, non-IPF, and shared protein category with FDR $p<0.05$.

2.3.6 Power calculation

Based on a three-year event probability of 0.35 and assumed mean biomarker standard deviation of 0.8, we estimated that 670 individuals provide >80% power to detect biomarkers associated with TFS with Cox model coefficient >0.42 (HR >1.5) using a conservative alpha of 0.000017, which Bonferroni adjusts for multiple testing of 2925 analytes. The similarly sized validation cohort was estimated to provide >90% power to detect biomarkers of similar effect size using an alpha of 0.00017, which Bonferroni adjusts for multiple testing of up to 300 analytes advanced from discovery analysis.

2.4 ***Results***

2.4.1 Cohorts, baseline characteristics and outcomes

The discovery and validation cohorts were comprised of 676 and 616 patients, respectively (Table 2-1). Among discovery cohort patients, 480 were included in a recently published(Bowman et al., 2022) medium throughput proteomics investigation of non-IPF ILD. Mean age was 63-64 years in each cohort, with females and whites predominating in both. Connective tissue disease-associated ILD was the most common diagnosis in both cohorts, followed by fibrotic hypersensitivity pneumonitis in the discovery cohort and non-IPF idiopathic interstitial pneumonia in the validation cohort. Both cohorts demonstrated similar reductions in percent predicted FVC and DLCO. Nearly half of patients in each cohort were exposed to immunosuppressant therapy at the time of blood draw. The discovery cohort had lower three-year TFS compared to the validation cohort, with patients with connective tissue disease-associated ILD showing significantly better overall survival compared to those with fibrotic hypersensitivity pneumonitis and non-IPF idiopathic interstitial pneumonia in both cohorts

(Figure 2-1). Differences in discovery and validation cohort survival was driven predominantly by differences in survival for non-IPF idiopathic interstitial pneumonia subgroups (Figure 2-2). Complete survival data through 36-months was available for 638 (94.4%) and 557 (90.4%) individuals in the discovery and validation cohort, respectively.

2.4.2 Protein Biomarkers of non-IPF ILD Survival

Of 2925 proteins assessed, 73 were associated with three-year TFS in the discovery cohort after adjustment for false discovery (Figure 2-3, Table 2-2). Of the 73 proteins identified in discovery cohort analysis, 62 were associated with increased risk of death or lung transplant, while 11 were associated with reduced risk (Table 2-2). The top TFS-associated proteins based on FDR p-value were periplakin (PPL; HR 1.75 per unit increase, 95% CI 1.43-2.14; FDR $p=1.32 \times 10^{-4}$), Interleukin-1 receptor-like 1 (IL1RL1; HR 1.49 per unit increase, 95% CI 1.29-1.73, FDR $p=1.80 \times 10^{-4}$) and 5'(3')-deoxyribonucleotidase, cytosolic type (NT5C1A) (HR 1.53 per unit increase, 95% CI. 1.30-1.80; FDR $p=2.29 \times 10^{-4}$) (Table 2-3).

Of 73 discovery cohort proteins advanced for further testing, 44 maintained TFS association with consistent direction of effect in the validation cohort (Table 2-3). Of these 44 validated proteins, 41 were associated with increased risk of death or lung transplant, while 3 were associated with reduced risk (Table 2-3). The top TFS-associated proteins based on FDR p-value were amphiregulin (AREG; HR 2.51 per one-unit increase, 95% CI 2.07-3.04, FDR $p=4.71 \times 10^{-19}$), integrin subunit beta 6 (ITGB6) (HR 2.46 per one-unit increase; 95% CI 1.95-3.10; FDR $p=9.22 \times 10^{-13}$) and keratin 19 (KRT19) (HR 1.70, 95% CI 1.47-1.98; FDR $p=4.17 \times 10^{-11}$). Of the 44 validated proteins identified in this analysis, 27 were previously linked to IPF survival,(Oldham et al., 2023) while 17 were unique to non-IPF ILD survival. Effect modification by ILD subtype on the biomarker-survival association was not consistently observed across cohorts for any validated biomarker except immunoglobulin-like domains protein 3 (LRIG3) (Figure 2-4).

2.4.3 Overlap with IPF Prognostic Protein Biomarkers

Using a nominal statistical significance threshold of $p<0.05$, we identified 130 proteins associated with TFS with consistent direction of effect across non-IPF ILD discovery and

validation cohorts. Using the same methodology, we identified 326 proteins associated with TFS with consistent direction of effect across IPF discovery and validation cohorts from our recent study of proteomic biomarkers of IPF survival.(Oldham et al., 2023) Of these proteins, 47 were unique to non-IPF ILD, 243 were unique to IPF and 83 were shared across conditions (Table 2-4).

2.4.4 Pathway analysis

When conducting pathway analysis of each category, the antimicrobial peptides, sheddase signaling and tissue factor in cancer pathways were unique to non-IPF ILD (Figure 2-5). Those shared between IPF and non-IPF ILD included the hepatic fibrosis, pathogen induced cytokine storm signaling pathway, wound healing signaling pathway, TNFs bind their physiological receptors and role of JAK family kinases in IL-6-type Cytokine Signaling pathways (Figure 2-5). Pathways unique to IPF included post-translational protein phosphorylation, regulation of insulin-like growth factor transport, role of osteoblasts in rheumatoid arthritis signaling, granulocyte adhesion and diapedesis and airway pathology in chronic obstructive pulmonary disease pathways (Figure 2-5).

2.5 ***Discussion***

In this investigation we identified and validated 44 circulating protein biomarkers of three-year TFS in patients with non-IPF ILD, which generally showed consistent measures of association across key non-IPF ILD subtypes and treatment exposure strata. In an expanded analysis of biomarkers nominally associated with TFS, we identified prognostic protein biomarkers unique to non-IPF ILD and those shared with IPF. Pathway analysis then revealed novel molecular pathways of non-IPF ILD survival and established shared molecular pathways for IPF and non-IPF ILD. This study is among the largest molecular analysis of non-IPF ILD conducted to date and is the first to use high-throughput proteomic technology to glean biological insights of PPF in those with non-IPF ILD.

These findings build on prior protein-based studies of non-IPF ILD by our group. The first, a targeted quantitative analysis, showed that several known prognostic biomarkers of IPF survival also predicted this outcome in non-IPF ILD.(Alqalyoobi et al., 2020) The second, a targeted

proteomic analysis, identified 17 novel proteomic biomarkers of one-year ILD progression in this population.(Bowman et al., 2022) Only two biomarkers (ITGB6 and KRT19) identified through our first proteomic investigation showed sustained TFS association in the current study. This discordance likely reflects the statistical penalty required to test thousands of proteins in this study but also highlights the importance of confounder adjustment when conducting biomarker investigation. Many previously reported prognostic biomarkers of ILD survival are also strongly associated with disease severity,(Bowman et al., 2021; Bowman et al., 2022) which likely confounds the relationship between many of these biomarkers and subsequent survival. Failure to account for confounding likely results in spurious biomarker-outcome associations through non-causal pathways.

Of the 44 validated prognostic protein biomarkers identified in this study, 27 were previously linked to IPF survival.(Oldham et al., 2023) KRT19 and ITGB6, which we previously showed to predict one-year non-IPF ILD progression(Bowman et al., 2022) and found to have a strong association with three-year TFS in the current study, are also validated prognostic biomarkers in IPF, suggesting an important role in fibrogenesis irrespective of ILD subtype.(Molyneaux et al., 2022; Oldham et al., 2023) Importantly, ITGB6 comprises the beta-subunit of αVβ6, selective inhibition of which is now being evaluated as part of an ongoing phase IIb IPF clinical trial. Growth/differentiation factor 15 (GDF15) has also been linked to outcomes across a number of conditions and recent studies suggest that this growth factor could mediate the association chronological age and interstitial lung abnormalities(Sanders et al., 2021) and between chronological age and survival in those with fibrotic ILD.(Pugashetti et al., 2024) Several proteins identified in this analysis were previously selected when using latent class analysis to establish molecular endotypes of IPF, which had differential survival and treatment response.(Maddali et al., 2024) Together these studies provide a growing evidence base that biomarkers identified through unbiased proteomic investigation play an important role in PPF and reflect shared pathobiology that transcends ILD subtype. These findings also provide a biological basis for clinical trials employing a basket approach to PPF classification,(Behr et al., 2021; Flaherty et al., 2019) and warrant new studies to evaluate whether recently identified endotypes in IPF(Maddali et al., 2024) have the potential to similarly discriminate immunosuppressant response in patients with non-IPF ILD.

Beyond prognostic biomarkers shared across IPF and non-IPF ILD, 17 novel biomarkers of TFS in those with non-IPF ILD were also identified. Of these, five were previously linked to fibrosis or inflammation. Sialic acid-binding Ig-like lectin 8 (SIGLEC8) is a cell surface receptor expressed on human mast cells and eosinophils involved in the inflammatory response.(Schanin et al., 2021) A monoclonal antibody targeting SIGLEC8 is currently in development and has been shown to reduce pulmonary inflammation and fibrosis in bleomycin mouse models.(Schanin et al., 2021) Yes1 associated transcriptional regulator (YAP1) has been shown to cause transcription of inflammatory mediators in response to tumor necrosis factor alpha stimulation(Chen et al., 2023) and a recent study showed that YAP1 leads to organ fibrosis through injury-mediated activation of myofibroblasts.(He et al., 2022) Endothelin-1 (EDN1) has been previously linked to IPF pathogenesis through fibroblast activation and proliferation,(Swigris & Brown, 2010) but blockade of EDN1 did not impact IPF survival in a large phase III clinical trial(King et al., 2011) and this protein was not associated with TFS in our recent study of IPF.(Oldham et al., 2023) TGF-beta receptor type-1 (TGFBR1) serves as the receptor for TGF-β and contributes to pulmonary fibrogenesis through persistent expression.(Wu et al., 2023) While their role in pulmonary inflammation and fibrosis remains unclear, autoantibodies directed at 5'(3')-deoxyribonucleotidase, cytosolic type (NT5C1A) were recently linked to several forms of myositis.(Lee et al., 2023)

Little is known about the role the remaining 12 novel prognostic biomarkers of non-IPF ILD survival may play in pulmonary inflammation and fibrosis characteristic of most non-IPF ILDs, which represents an area ripe for investigation. Why fewer protein biomarkers of TFS were identified in non-IPF ILD relative to IPF remains unclear, but could reflect better overall survival in non-IPF ILD relative to IPF, which has been long documented,(Bjoraker et al., 1998; Pugashetti et al., 2023) resulting in fewer events and biomarkers associated with these events. Another explanation may stem from biological differences in non-IPF ILD subtypes, as each condition can exhibit a variable degree of inflammation and fibrosis. Recent studies(Huang et al., 2024; van der Sar et al., 2023) suggest that clinical diagnoses may soon give way to molecular classification, which may help discriminate fibrotic-predominant conditions from those with concurrent inflammation.

Beyond individual biomarkers identified in this study, pathway analysis sheds important light on the molecular pathways shared between IPF and non-IPF ILD and those unique to each condition. Among the relatively few prognostic protein biomarkers unique to non-IPF ILD, the anti-microbial peptides pathway and sheddase signaling pathways were identified. Anti-microbial peptides (AMPs), also known as host defense peptides, comprise a number of small peptides with potent inhibitory effects against diverse types of microorganisms, playing a vital role in the innate immune system. Given their complexity and pivotal role in modulating the immune response, AMP dysregulation has been proposed as a potential driver of tissue-specific inflammation. AMPs with prominent pro-inflammatory effects have been implicated in the development of several autoimmune conditions, including Sjogren's syndrome and rheumatoid arthritis,(Liang & Diana, 2020) while those with anti-inflammatory effects have been proposed as novel therapies to treat autoimmune disease.(Liang & Diana, 2020) Given prior studies linking the regulatory effects of AMPs on the gut microbiome(Gubatan et al., 2021) and the relationship between the airway microbiome composition and ILD diagnosis and outcomes,(Han et al., 2014; Invernizzi et al., 2020; Molyneaux et al., 2017; Molyneaux et al., 2014; O'Dwyer et al., 2024) the interplay between AMPs and the airway microbiome in patients with ILD represents a potentially important knowledge gap. This also suggests that AMPs may play an important role in antecedent parenchymal inflammation and could one day serve as a therapeutic target for ILDs with a prominent inflammatory component.

Sheddase signaling involves the cleavage of proteins from the cell surface, with A Disintegrin and Metalloproteinases (ADAMs) among the most prominent sheddases.(Miller et al., 2017) ADAM10 has been shown to promote myofibroblast activation following induction by transforming growth factor-β1.(Lagares et al., 2017) While ADAM10 was not among the analytes included on the Olink array, the array did include ADAM12, which was among the top proteins TFS associated proteins identified. This protein has been linked to dermal fibrosis, leading to interstitial collagen accumulation. Importantly, knockdown of ADAM12 reduced fibrogenesis and downstream collagen production.(Dulauroy et al., 2012)

2.6 ***Conclusion***

Here we identified and validated novel prognostic protein biomarkers in non-IPF ILD, which showed consistent effect across non-IPF ILD subtypes, supporting a basket approach to studying PPF in those with non-IPF ILD. While pathway analysis of prognostic biomarkers identified several biological pathways shared with IPF, this approach also identified several pathways that were unique to non-IPF ILD. These findings support additional investigation of these pathways and how they may drive non-IPF ILD progression independent of those shared across the spectrum of fibrotic ILD.

Table 2.1. Baseline characteristics and outcomes for discovery and validation cohorts		
Characteristic	**Discovery Cohort (n=676)**	**Validation Cohort (n=616)**
Age, mean (±SD)	63.1 (12.6)	63.7 (11.9)
Female sex, n (%)	416 (61.5)	341 (55.4)
Race/ethnicity, n (%)		
White	474 (70.1)	503 (81.7)
Black	48 (7.1)	79 (12.8)
Hispanic	92 (13.6)	5 (0.8)
Asian	50 (7.4)	17 (2.8)
Other/Unknown	12 (1.8)	12 (2.0)
ILD Diagnosis, n (%)		
Connective tissue disease associated ILD	311 (46.0)	246 (39.9)
Fibrotic hypersensitivity pneumonitis	196 (29.0)	142 (23.1)
Non-IPF idiopathic interstitial pneumonia	169 (25.0)	228 (37.0)
Ever smoker, n (%)	254 (37.6)	320 (52.0)
FVC %, mean (±SD)	65.8 (19.1)	67.9 (19.5)
DLCO %, mean (±SD)	45.3 (17.9)	46.2 (18.8)
Immunosuppressant exposure , n (%)	325 (48.1)	309 (50.2)
Abbreviations: ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis; FVC = forced vital capacity; DLCO = diffusion capacity of the lung for carbon monoxide		

Table 2.2. Selection of discovery cohort results ranked by FDR p-value

Gene	UniProt ID	% below	HR	95% CI	P-value	FDR P-value	Advance
PPL	O60437	0.00	1.753	1.434-2.144	4.500E-08	1.316E-04	yes
IL1RL1	Q01638	0.01	1.492	1.287-1.731	1.230E-07	1.804E-04	yes
CHCHD6	Q9BRQ6	0.01	1.911	1.492-2.446	2.810E-07	2.289E-04	yes
NT5C1A	Q9BXI3	0.01	1.528	1.299-1.798	3.130E-07	2.289E-04	yes
AREG	P15514	0.01	1.482	1.271-1.729	5.120E-07	2.497E-04	yes
INSR	P06213	0.54	3.084	1.989-4.78	4.780E-07	2.497E-04	yes
GDF15	Q99988	0.01	1.488	1.259-1.757	2.910E-06	1.063E-03	yes
SERPINA5	P05154	0.00	0.586	0.469-0.732	2.590E-06	1.063E-03	yes
RBFOX3	A6NFN3	0.00	1.399	1.214-1.613	3.610E-06	1.173E-03	yes
ACTA2	P62736	0.01	1.498	1.259-1.783	5.250E-06	1.277E-03	yes
CEACAM6	P40199	0.00	1.390	1.207-1.601	4.810E-06	1.277E-03	yes
LTBP2	Q14767	0.01	1.592	1.302-1.946	5.670E-06	1.277E-03	yes
SERPINA3	P01011	0.00	2.792	1.794-4.344	5.340E-06	1.277E-03	yes
ADAMTSL4	Q6UY14	0.00	1.725	1.361-2.188	6.770E-06	1.354E-03	yes
LMOD1	P29536	0.00	1.367	1.193-1.567	6.940E-06	1.354E-03	yes
S100P	P25815	0.02	1.253	1.132-1.387	1.310E-05	2.388E-03	yes
THBS2	P35442	0.01	1.481	1.239-1.77	1.630E-05	2.812E-03	yes
FABP3	P05413	0.00	1.213	1.11-1.325	2.010E-05	2.981E-03	yes
GIPC2	Q8TF65	0.01	1.742	1.351-2.246	1.910E-05	2.981E-03	yes
GSN	P06396	0.00	0.608	0.483-0.765	2.140E-05	2.981E-03	yes
IFI30	P13284	0.00	1.392	1.195-1.621	2.100E-05	2.981E-03	yes
BCHE	P06276	0.00	0.728	0.628-0.844	2.500E-05	3.272E-03	yes
FGF23	Q9GZV9	0.00	1.204	1.104-1.312	2.570E-05	3.272E-03	yes
CKB	P12277	0.00	1.367	1.178-1.585	3.600E-05	4.224E-03	yes
MRC1	P22897	0.00	1.764	1.347-2.309	3.610E-05	4.224E-03	yes
LRIG3	Q6UXM1	0.89	2.024	1.445-2.834	4.100E-05	4.521E-03	yes
SIGLEC8	Q9NYZ4	0.00	1.507	1.238-1.833	4.170E-05	4.521E-03	yes
LRG1	P02750	0.01	1.588	1.266-1.992	6.390E-05	6.677E-03	yes
ADAM12	O43184	0.02	1.397	1.186-1.647	6.670E-05	6.726E-03	yes
EGFLAM	Q63HQ2	0.00	1.767	1.334-2.341	7.230E-05	7.046E-03	yes
THSD1	Q9NS62	0.76	1.629	1.278-2.077	8.100E-05	7.641E-03	yes
VSIG2	Q96IQ7	0.02	1.272	1.128-1.433	8.460E-05	7.737E-03	yes
KRT19	P08727	0.01	1.245	1.116-1.39	9.190E-05	8.142E-03	yes
NCR3LG1	Q68D85	0.01	1.515	1.229-1.867	9.690E-05	8.142E-03	yes
RNASE1	P07998	0.00	1.410	1.186-1.677	9.740E-05	8.142E-03	yes
COL24A1	Q17RW2	0.01	1.372	1.17-1.61	1.038E-04	8.433E-03	yes
CLEC3B	P05452	0.00	0.690	0.571-0.833	1.211E-04	9.577E-03	yes
ORM1	P02763	0.00	2.107	1.438-3.089	1.324E-04	1.019E-02	yes
NTPROBNP	P16860	0.01	1.140	1.066-1.22	1.436E-04	1.077E-02	yes
BCAT1	P54687	0.00	1.533	1.229-1.913	1.511E-04	1.105E-02	yes
YAP1	P46937	0.00	1.538	1.229-1.926	1.691E-04	1.206E-02	yes
FCRLB	Q6BAA4	0.01	1.257	1.115-1.417	1.881E-04	1.234E-02	yes
MAMDC2	Q7Z304	0.00	1.465	1.199-1.79	1.848E-04	1.234E-02	yes
SDC1	P18827	0.01	1.304	1.134-1.499	1.898E-04	1.234E-02	yes
SHISA5	Q8N114	0.00	1.345	1.152-1.57	1.785E-04	1.234E-02	yes
CEACAM16	Q2WEN9	0.00	0.793	0.701-0.897	2.150E-04	1.349E-02	yes
TNFRSF10B	O14763	0.01	1.322	1.14-1.533	2.167E-04	1.349E-02	yes
SFTPA1	Q8IWL2	0.28	1.430	1.182-1.731	2.343E-04	1.428E-02	yes
ADAM9	Q13443	0.00	1.511	1.207-1.891	3.096E-04	1.771E-02	yes

Table 2.2. Selection of discovery cohort results ranked by FDR p-value

Gene	UniProt ID	% below	HR	95% CI	P-value	FDR P-value	Advance
CHCHD10	Q8WYQ3	0.00	1.273	1.116-1.452	3.148E-04	1.771E-02	yes
CTHRC1	Q96CG8	0.00	1.683	1.269-2.231	3.031E-04	1.771E-02	yes
EDN1	P05305	0.00	1.528	1.215-1.923	2.970E-04	1.771E-02	yes
AHSG	P02765	0.00	0.723	0.606-0.863	3.315E-04	1.829E-02	yes
EDA2R	Q9HAV5	0.01	1.321	1.133-1.539	3.652E-04	1.978E-02	yes
CHAD	O15335	0.01	0.800	0.707-0.906	4.383E-04	2.331E-02	yes
PPP1R14D	Q9NXH3	0.10	1.306	1.125-1.516	4.633E-04	2.420E-02	yes
HRG	P04196	0.00	0.771	0.666-0.892	4.837E-04	2.422E-02	yes
OSMR	Q99650	0.01	2.288	1.437-3.641	4.827E-04	2.422E-02	yes
WFDC2	Q14508	0.01	1.381	1.152-1.655	4.886E-04	2.422E-02	yes
NPL	Q9BXD5	0.00	1.355	1.142-1.609	5.131E-04	2.502E-02	yes
MSTN	O14793	0.07	0.782	0.68-0.899	5.267E-04	2.526E-02	yes
GALNT5	Q7Z7M9	0.02	1.436	1.169-1.765	5.654E-04	2.667E-02	yes
SERPINA1	P01009	0.25	18.903	3.547-100.723	5.746E-04	2.668E-02	yes
EVPL	Q92817	0.41	1.445	1.171-1.783	5.912E-04	2.702E-02	yes
PSAP	P07602	0.00	1.573	1.213-2.039	6.397E-04	2.879E-02	yes
CA8	P35219	0.99	1.592	1.217-2.081	6.780E-04	3.005E-02	yes
TGFBR1	P36897	0.85	1.508	1.183-1.923	9.091E-04	3.969E-02	yes
ITGB6	P18564	0.01	1.461	1.167-1.829	9.269E-04	3.987E-02	yes
KLK7	P49862	0.00	0.749	0.631-0.89	1.046E-03	4.436E-02	yes
SEL1L	Q9UBV2	0.00	1.362	1.132-1.638	1.078E-03	4.503E-02	yes
IGFBP1	P08833	0.01	1.150	1.057-1.25	1.110E-03	4.573E-02	yes
TFF2	Q03403	0.01	1.227	1.085-1.388	1.166E-03	4.739E-02	yes
WFIKKN1	Q96NZ8	0.00	0.761	0.645-0.898	1.204E-03	4.822E-02	yes
CPXM2	Q8N436	0.01	1.491	1.169-1.901	1.287E-03	5.086E-02	no
CFP	P27918	0.00	0.603	0.442-0.822	1.371E-03	5.315E-02	no
GHR	P10912	0.00	0.647	0.496-0.845	1.381E-03	5.315E-02	no
CGN	Q9P2M7	0.46	1.316	1.111-1.559	1.454E-03	5.524E-02	no
PALM2	Q8IXS6	0.00	1.280	1.099-1.492	1.509E-03	5.634E-02	no
SEPTIN8	Q92599	0.00	1.351	1.122-1.626	1.522E-03	5.634E-02	no
GPR37	O15354	0.01	1.247	1.087-1.43	1.616E-03	5.907E-02	no

Table 2.3. Validated protein biomarkers of three-year TFS in discovery and validation cohorts						
Gene	**Discovery Cohort (n=676)**			**Validation Cohort (n=616)**		
	HR	**95% CI**	**FDR P**	**HR**	**95% CI**	**FDR P**
AREG	1.482	1.271-1.729	2.497E-04	2.508	2.069-3.039	4.711E-19
ITGB6	1.461	1.167-1.829	3.987E-02	2.458	1.950-3.098	9.224E-13
KRT19	1.245	1.116-1.39	8.142E-03	1.704	1.469-1.976	4.168E-11
LTBP2	1.592	1.302-1.946	1.277E-03	2.497	1.917-3.251	2.066E-10
SDC1	1.304	1.134-1.499	1.234E-02	1.644	1.389-1.946	1.053E-07
COL24A1	1.372	1.17-1.61	8.433E-03	1.825	1.484-2.244	1.433E-07
GDF15	1.488	1.259-1.757	1.063E-03	1.667	1.370-2.030	3.122E-06
BCAT1	1.533	1.229-1.913	1.105E-02	1.919	1.495-2.462	3.160E-06
TNFRSF10B	1.322	1.14-1.533	1.349E-02	2.018	1.539-2.648	3.167E-06
WFDC2	1.381	1.152-1.655	2.422E-02	1.790	1.404-2.283	1.966E-05
THBS2	1.481	1.239-1.77	2.812E-03	1.689	1.353-2.107	2.323E-05
MAMDC2	1.465	1.199-1.79	1.234E-02	1.804	1.396-2.331	3.673E-05
CTHRC1	1.683	1.269-2.231	1.771E-02	2.150	1.543-2.996	3.773E-05
EDA2R	1.321	1.133-1.539	1.978E-02	1.753	1.371-2.242	3.960E-05
SERPINA3	2.792	1.794-4.344	1.277E-03	4.071	2.189-7.571	4.473E-05
EDN1	1.528	1.215-1.923	1.771E-02	1.965	1.450-2.664	6.105E-05
IL1RL1	1.492	1.287-1.731	1.804E-04	1.553	1.269-1.900	8.361E-05
SIGLEC8	1.507	1.238-1.833	4.521E-03	1.626	1.296-2.039	1.050E-04
PPL	1.753	1.434-2.144	1.316E-04	1.672	1.308-2.138	1.579E-04
ADAMTSL4	1.725	1.361-2.188	1.354E-03	1.987	1.420-2.780	2.277E-04
GALNT5	1.436	1.169-1.765	2.667E-02	1.767	1.329-2.351	3.188E-04
NTPROBNP	1.140	1.066-1.22	1.077E-02	1.192	1.088-1.306	5.207E-04
LMOD1	1.367	1.193-1.567	1.354E-03	1.519	1.219-1.892	6.212E-04
MRC1	1.764	1.347-2.309	4.224E-03	1.936	1.261-2.755	7.215E-04
YAP1	1.538	1.229-1.926	1.206E-02	1.876	1.333-2.639	8.760E-04
CEACAM6	1.390	1.207-1.601	1.277E-03	1.309	1.130-1.516	9.094E-04
TGFBR1	1.508	1.183-1.923	3.969E-02	1.872	1.191-2.941	1.766E-03
SFTPA1	1.430	1.182-1.731	1.428E-02	1.293	1.112-1.504	2.075E-03
RBFOX3	1.399	1.214-1.613	1.173E-03	1.423	1.157-1.749	2.148E-03
CKB	1.367	1.178-1.585	4.224E-03	1.423	1.153-1.757	2.506E-03
CHAD	0.800	0.707-0.906	2.331E-02	0.781	0.664-0.918	6.570E-03
LRG1	1.588	1.266-1.992	6.677E-03	1.731	1.195-2.507	8.418E-03
ACTA2	1.498	1.259-1.783	1.277E-03	1.380	1.098-1.735	1.283E-02
NPL	1.355	1.142-1.609	2.502E-02	1.267	1.065-1.507	1.644E-02
CLEC3B	0.690	0.571-0.833	9.577E-03	0.526	0.326-0.847	1.681E-02
OSMR	2.288	1.437-3.641	2.422E-02	2.153	1.220-3.801	1.707E-02
NCR3LG1	1.515	1.229-1.867	8.142E-03	1.499	1.099-2.044	2.015E-02
TFF2	1.227	1.085-1.388	4.739E-02	1.278	1.060-1.541	2.032E-02
ADAM12	1.397	1.186-1.647	6.726E-03	1.389	1.074-1.796	2.246E-02
VSIG2	1.272	1.128-1.433	7.737E-03	1.217	1.043-1.421	2.373E-02
IGFBP1	1.150	1.057-1.25	4.573E-02	1.132	1.024-1.252	2.746E-02
MSTN	0.782	0.68-0.899	2.526E-02	0.801	0.666-0.963	3.176E-02
NT5C1A	1.528	1.299-1.798	2.289E-04	1.234	1.036-1.471	3.183E-02
LRIG3	2.024	1.445-2.834	4.521E-03	1.785	1.073-2.971	4.269E-02

Table 2.4. Proteins nominally associated with TFS IPF and non-IPF cohorts stratified by direction of effect

IPF Only					Non-IPF ILD Only		Shared		
Protective		Harmful			Protective	Harmful	Protective	Harmful	
ADAM22	GUSB	ADA2	IL5RA	PTX3	HMGCS1	ACTA2	ASAH2	ACE2	MAMDC2
ADAMTS13	HJV	ADAMTS1	IL6	PVR	CHAD	ACVRL1	CD6	ADAMTS16	MRC1
ADH4	HPGDS	ADGRD1	ITIH3	ROBO1	ITGB7	ADAM12	CFP	ADAMTSL4	NCAM2
AFM	ICOSLG	ANGPTL2	ITIH4	RSPO1	BDNF	ANGPTL1	CLEC3B	ADIPOQ	NCR3LG1
AGR3	IDI2	BMP4	KIRREL2	RSPO3	F7	CA8	CNTN5	AGR2	NPTX2
APOA4	IDS	BPIFB1	KLK4	SCGB1A1	BPIFA2	CALCB	FASLG	AREG	NRP2
ASS1	IFNL2	C9	KRT18	SCGB3A1		CEACAM5	GHR	BCAT1	NTproBNP
BAG3	IGDCC3	CCL17	LAMP3	SDK2		CLEC1A	HMCN2	BRK1	OCLN
BHMT2	IGSF9	CCL18	LAYN	SEL1L		CLIC5	KLKB1	C7	OSMR
BTD	IL13RA1	CCL20	LCP1	SELE		CXADR	MSTN	CAPS	PIGR
C19orf12	IL2RB	CCN1	LEG1	SEPTIN3		EFHD1	NTRK2	CCL23	PPL
CA14	IL31RA	CD274	LEO1	SERPINA11		EFNA1	PI16	CEACAM6	PRSS22
CA4	IL7R	CD276	LIF	SEZ6L2		FGF23	SERPINA5	CKB	PTK7
CA6	ITGAM	CD80	LY96	SLAMF7		GP2	SERPINA7	COL24A1	RBFOX3
CAPN3	KLK7	CDCP1	MATN2	SPOCK1		GRPEL1	TNFSF11	COL4A1	ROBO2
CBX2	KLK8	CDH3	MATN3	SPON1		IGFBP7		CSF2	SCGB3A2
CD1C	KLRB1	CHGB	MBL2	SPON2		IL36G		CSPG4	SDC1
CD209	KLRF1	CHI3L1	MDK	SPP1		LCN15		CTHRC1	SERPINA3
CDAN1	KLRK1	CHL1	MERTK	ST6GAL1		LRIG3		EBI3_IL27	SFTPA1
CDH5	L1CAM	CHRDL1	MILR1	STC1		LTA4H		EDA2R	SFTPA2
CDHR1	LAMP2	CHRDL2	MMP10	TFPI		NPC2		EDN1	SIGLEC8
CEACAM16	LEP	CLEC11A	MMP7	TIMD4		NPL		ELN	SPINK1
CES2	LRRN1	CLEC4G	MSLN	TIMP4		NT5C1A		EVPL	SPINT1
CLEC4M	MAMDC4	CLUL1	MUC16	TK1		PCARE		FDX1	TFF3
CNDP1	MAN2B2	COL18A1	MZB1	TMEM132A		PLXDC2		GALNT5	THBS2
CNTN3	MELTF	COL6A3	NBL1	TNFRSF10C		POF1B		GDF15	TNC
COL9A1	MEP1B	CRIM1	NDUFS6	TNFRSF11B		POSTN		ICAM5	TNFRSF10B
CPM	MPRIP	CSF1	NECTIN2	TNFRSF13B		PRSS2		IGFBP1	TNFRSF1A
CRTAC1	NFATC3	CST1	NECTIN4	TNFRSF17		REG3A		IGFBP2	VCAN
DLK1	NPY	CST5	NEFL	TNFRSF6B		RNASE6		IGFBP4	VEGFD
DSCAM	PLXDC1	CTSL	NELL2	TNFRSF8		SCARA5		IL1RL1	WFDC2
DSG3	PRND	CXCL13	NOS3	TNFSF13B		SCARF2		ITGB6	YAP1
DSG4	PTGR1	CXCL17	NPDC1	TXNDC15		SRPX		KRT19	
DUSP13	RET	ENAH	NPPB	VWA1		TACSTD2		LMOD1	
ECE1	RTN4R	EPO	NRP1	WNT9A		TFF1		LRG1	
ENO3	SELPLG	ESM1	NUCB2	ZNF75D		TFF2		LTBP2	
ENPP6	SERPINF1	FAM20A	OGT			TGFA			
EPHA1	SSC4D	FCGR2A	ORM1			TGFBR1			
ESPL1	TCL1A	FCRL5	PAEP			TNFRSF12A			
F13B	TF	FCRLB	PALM2			VSIG2			
FAP	THBS4	FGL1	PGF			VSIG4			
FCRL6	TMPRSS15	FLT1	PLA2G10						
FGFBP2	TNFSF10	FOLR1	PLAUR						
FNDC1	TNXB	FSTL3	PRSS8						
GALNT2	TPPP3	GH1	PSG1						
GALNT7	TREH	GOLM2	PTN						
GDF2	TTR	GPHA2	PTPRN2						
GFRA3	UPB1	HAVCR1	PTPRZ1						
GFRAL	WFDC12	ICAM1	IL15						
GPC1	XG	ICAM3	IL17C						
GSN	XPNPEP2	IGFBPL1	IL17RB						
GSTA3		IGLC2	IL2RA						

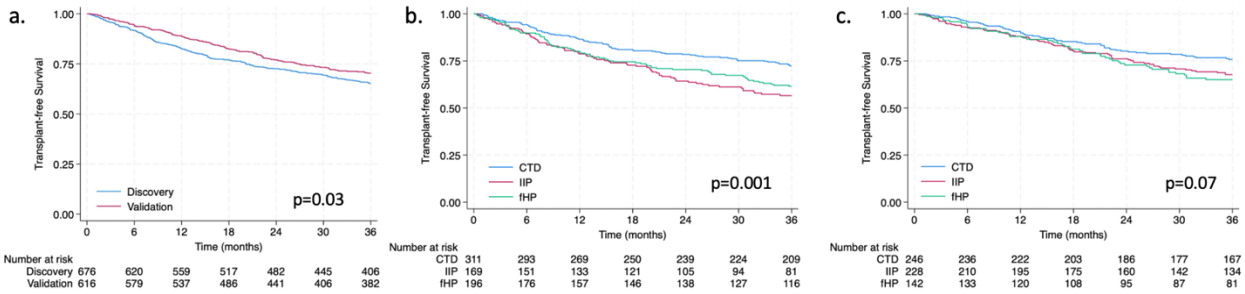


Figure 2.1. Survival in discovery and validation cohorts (a) and for ILD subtypes in the discovery (b) and validation (c) cohorts.

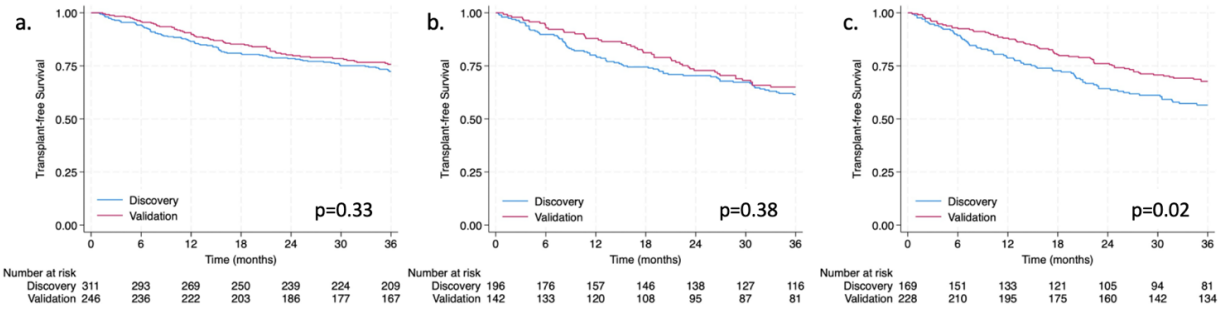


Figure 2.2. Survival in those with connective tissue disease-associated ILD (a), fibrotic hypersensitivity pneumonitis (b) and non-IPF idiopathic interstitial pneumonia (c).

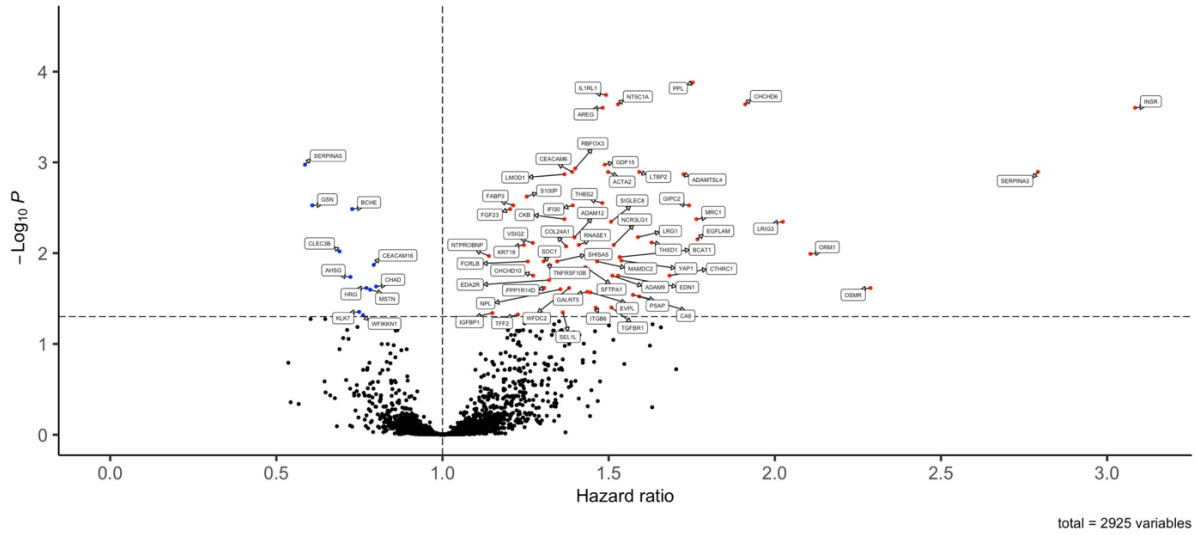


Figure 2.3. Volcano plot displaying the association between protein biomarkers and three-year transplant-free survival in a non-IPF ILD discovery cohort.

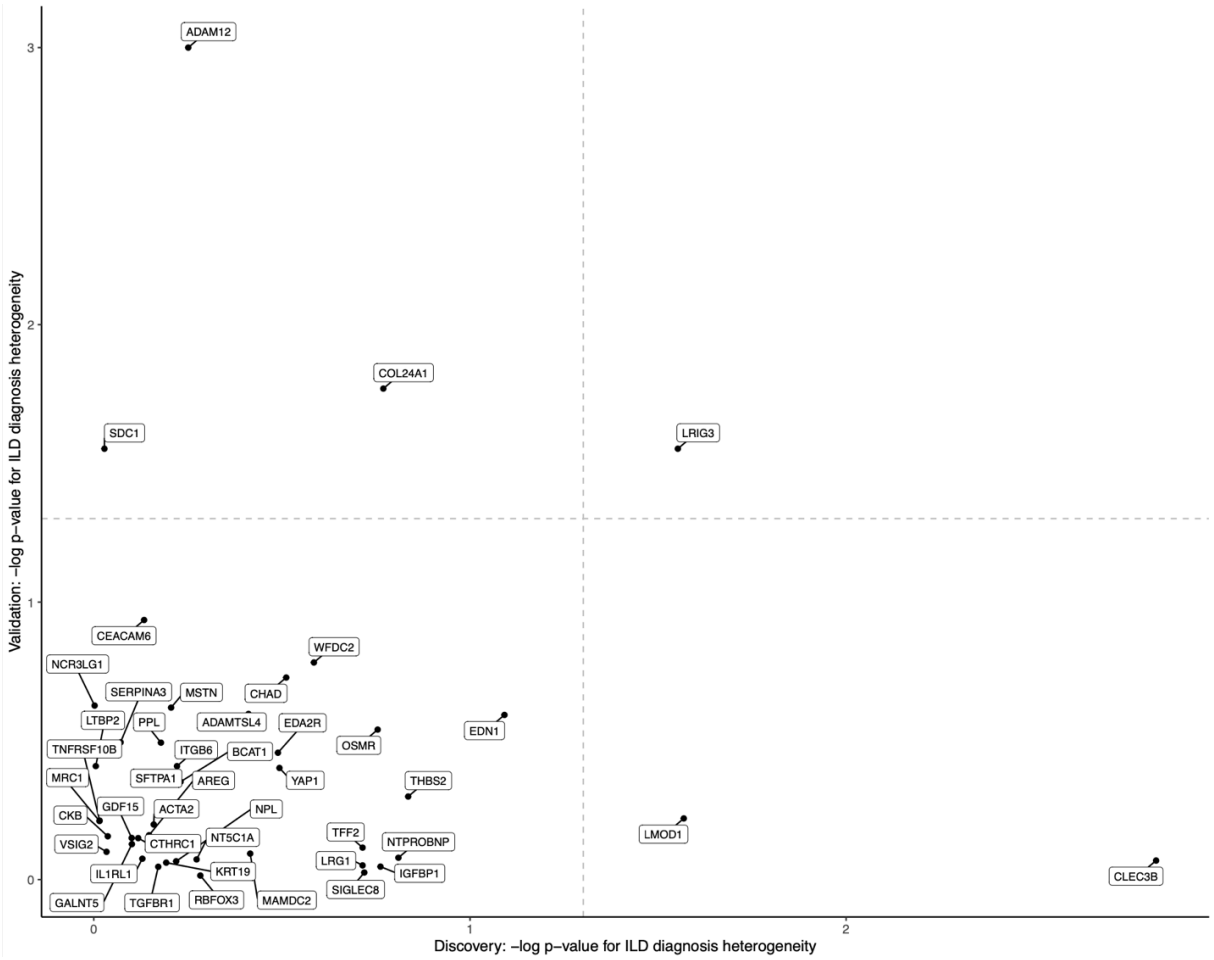


Figure 2.4. Plot of p-values for test heterogeneity in outcome association for validated biomarkers across non-IPF ILD subtypes for validated prognostic biomarkers across discovery and validation cohorts. The horizontal and vertical dotted lines correspond to unadjusted p-value threshold.

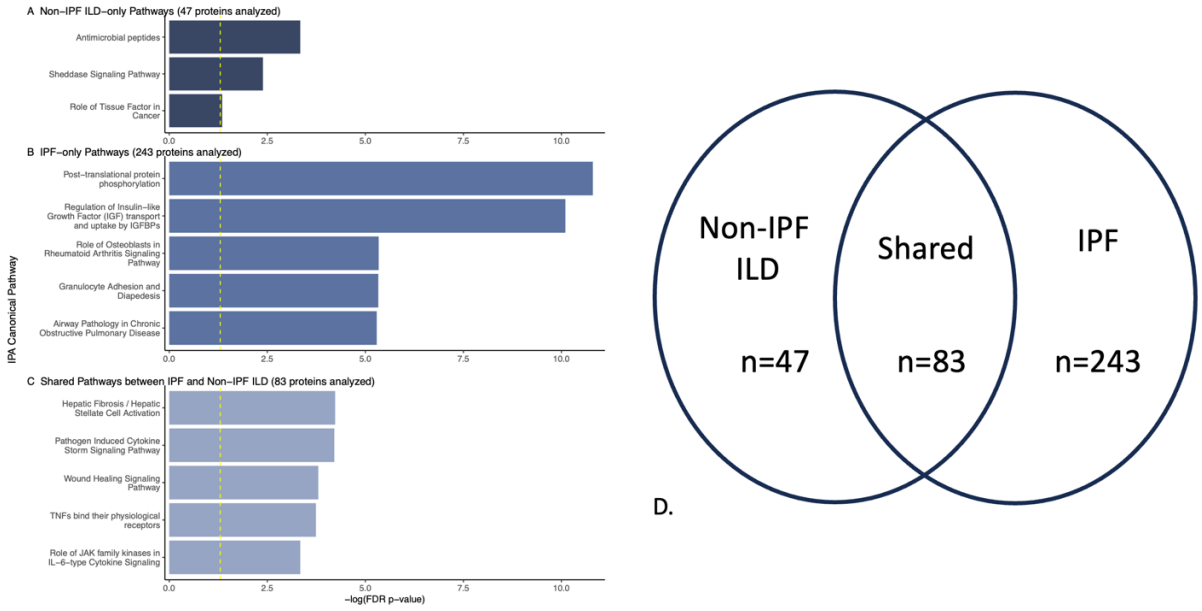


Figure 2.5. Molecular pathways unique to non-IPF ILD (A), IPF (B) and shared across conditions (C), along with a Venn diagram (D) depicting the number of shared and unique biomarkers for these conditions. The vertical dotted line corresponds to false discovery rate corrected p-value threshold.

Chapter 3 Untangling the Role of Prognostic Biomarkers in Progressive Pulmonary Fibrosis

3.1 ***Abstract***

Progressive pulmonary fibrosis (PPF) leads to declining lung function and death. While prognostic protein biomarkers have been identified for interstitial lung disease (ILD), it remains unclear which are causally related to PPF. Here we aimed to identify prognostic protein biomarkers that associate with idiopathic pulmonary fibrosis (IPF) and non-IPF ILD survival through a PPF-mediated causal pathway. Serially acquired lung function was used to estimate one-year change in forced vital capacity (FVC) in multi-center IPF (n=904) and non-IPF ILD (n=974) cohorts. Causal mediation analysis was performed to estimate the effect of previously reported prognostic protein biomarkers (exposure) on two-year restricted mean survival time (outcome) through PPF (mediator), defined by declining FVC. Proteins for which the survival effect was mediated by declining FVC were classified as potentially causal of PPF. Declining FVC mediated the IPF survival effect for 28 proteins, suggesting potential causality. Seven of these proteins were also mediated by declining FVC when tested in the non-IPF ILD cohort, including amphiregulin (AREG), integrin beta 6 (ITGB6), Kunitz-type protease inhibitor 1 (SPINT1), carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6), transforming growth factor beta binding protein 2 (LTBP2), Keratin, type I cytoskeletal 19 (KRT19), and intracellular adhesion molecule 5 (ICAM5). An inference-based protein interaction network suggested that most mediated proteins were upstream of AREG and ITGB6 and a biology-based network confirmed potential biological interaction for these proteins. By identifying proteins potentially causal of PPF, these findings help prioritize individual proteins and associated molecular pathways for therapeutic consideration.

3.2 ***Introduction***

Progressive pulmonary fibrosis (PPF) is a common and deadly manifestation of interstitial lung disease (ILD) that leads to declining lung function and premature death.(Pugashetti et al., 2023;

Raghu et al., 2022) While idiopathic pulmonary fibrosis (IPF) is considered the prototypical PPF, those with non-IPF ILD, including connective tissue disease-associated ILD, fibrotic hypersensitivity pneumonitis, and non-IPF idiopathic interstitial pneumonia also commonly develop PPF.(Cottin et al., 2018) Once PPF develops, survival is similarly poor irrespective of ILD subtype.(Pugashetti et al., 2023) While nintedanib has been shown to slow lung function decline in those with IPF(Richeldi et al., 2014) and non-IPF PPF,(Flaherty et al., 2019) the survival benefit appears modest,(Dempsey et al., 2019) highlighting the urgent need for new therapies.

Biomarker research is commonly used to identify candidate molecules for therapeutic targeting, with prognostic biomarkers of particular interest given their association with clinically relevant outcomes. While high-throughput proteomic investigations have led to a rapid expansion in prognostic biomarkers for ILD, these traditional association studies are limited by the inability to establish causation, leaving it unclear which proteins should be prioritized for therapeutic targeting. This stems from the multiple pathways through which a prognostic protein biomarker can associate with mortality. The first involves a disease-related causal pathway (Figure 3-1A). An example of this pathway is a protein that causes lung function decline, culminating in death. Alternately, a disease-unrelated causal pathway (Figure 3-1B) can result, whereby a protein leads to a competing cause of death such as heart disease or cancer. Finally, a non-causal pathway that results from residual confounding is also possible (Figure 3-1C). This pathway could result when an unmeasured factor (e.g., environmental pollution) leads to an elevation in protein measures and increases the risk of death in a manner unrelated to the disease of interest.

Causal mediation analysis is an advanced epidemiological method that employs the potential outcomes framework to estimate the causal effect of an exposure on an outcome through a third variable called a mediator, producing causal interpretations when core assumptions are satisfied.(VanderWeele, 2016) We recently used this approach to show that chronological age has no direct effect on ILD survival, but is instead mediated by several protein biomarkers of biological aging.(Pugashetti et al., 2024) A similar approach could potentially discriminate which prognostic protein biomarkers are likely to be causal of PPF and which are likely to represent an epiphenomenon, thereby helping to prioritize proteins for therapeutic targeting.

In this study, we performed causal mediation analysis to identify prognostic protein biomarkers that effect ILD survival through a PPF-related causal pathway, defined by declining forced vital capacity (FVC).(Pugashetti et al., 2023; Raghu et al., 2022) To do this, we leveraged protein, lung function and survival data from patients with IPF and non-IPF ILD who underwent proteomic profiling as part of two recently published investigations aimed at identifying prognostic protein biomarkers for these populations.(Alqalyoobi et al., 2025; Oldham et al., 2023) We hypothesized that causal mediation analysis would discriminate protein biomarkers that associate with ILD survival through a PPF-related causal pathway from those that associate with this outcome through an alternate pathway. Proteins with survival effect that was mediated by declining lung function were classified as potentially causal of PPF. Pathway and network analyses were then performed to explore potential protein-protein interactions.

3.3 ***Material and Methods***

3.3.1 ***Causal Framework***

Based on the causal framework depicted in Figure 3-1, there exists no direct effect of a protein biomarker on death without intermediary clinical progression. We therefore sought to discriminate proteins associated with transplant-free survival through a PPF-related causal pathway (Figure 3-1A), defined by declining FVC, from those associated with this outcome through an alternate pathway (Figures 3-1B-C). A directed acyclic graph was constructed to address core assumptions of causal mediation analysis, specifically that there exists no confounding of the 1) exposure-outcome relationship, 2) the exposure-mediator relationship, 3) the mediator-outcome relationship, and 4) the mediator-outcome relationship caused by the exposure (Figure 3-2).(VanderWeele, 2016)

Adjustment for age, sex, race, ILD diagnosis, smoking status, body mass index, baseline percent predicted FVC and diffusion capacity of the lung for carbon monoxide, and dichotomous exposure to immunosuppressant (mycophenolate mofetil, azathioprine, cyclophosphamide, rituximab or prednisone of any dose) and anti-fibrotic (nintedanib or pirfenidone) therapy at blood draw was expected to satisfy assumptions 1 and 2, while these covariates plus new immunosuppressant or anti-fibrotic therapy following blood draw were expected to satisfy assumption 3. To satisfy assumption 4, we attempted to adjust for competing causes of

death(Thenappan et al., 2024), including ischemic heart disease, heart failure (including due to pulmonary hypertension), and cancer, which could plausibly result from one or more biomarkers tested and also confound the mediator-outcome relationship. Because these conditions were not systematically assessed in these cohorts, we instead adjusted our mediator model for known clinical risk factors and protein biomarkers of each. Ischemic heart disease covariates included age, sex, body mass index and smoking history.(Safiri et al., 2021) Heart failure (including pulmonary hypertension) covariates included age, sex, body mass index, CTD diagnosis, percent predicted DLCO and circulating N-terminal prohormone of brain natriuretic peptide.(Chen et al., 2016; He et al., 2001; Zou et al., 2020) Cancer covariates included age, body mass index, smoking history and circulating cancer biomarkers, including alpha fetoprotein (liver, ovary), carcinoembryonic antigen-related cell adhesion molecule 5 (colon), gamma-enolase (lung), receptor tyrosine-protein kinase erbB-3 (breast, lung, others), gastrin-releasing peptide (gastric, lung), kallikrein-1 (colon, lung, thyroid, others), kallikrein-3 (prostate), spondin-2 (prostate, colon, others), and cellular tumor antigen p53 (multiple).(Arem & Loftfield, 2018; Das et al., 2023; Haikala & Jänne, 2021; Isgrò et al., 2015; Planque et al., 2008; Tailor et al., 2018; Zhang et al., 2023; Zong et al., 2014)

3.3.2 *Cohorts*

Patients with fibrotic ILD due to IPF and non-IPF ILD who underwent high-throughput proteomic profiling (described in 1.3) were eligible for inclusion. Those without serial forced vital capacity (FVC) performed at blood draw (allowable range -6 months to +3 months) and 3-18 months following blood draw were excluded, as were those with missing data for covariates included in mediation models described below.

3.3.3 *Measures*

Detailed methods for proteomic data generation have been described previously.(Oldham et al., 2023) Briefly, proteomic data were determined using the Explore 3072 panel (Olink®, Uppsala, Sweden), which uses proximity extension assays to estimate the relative abundance of circulating plasma proteins.(Lundberg et al., 2011) Data were generated as a single batch in the IPF cohort and in two batches for the non-IPF ILD cohort. To pool non-IPF ILD batches, quantile normalization was performed, with each protein categorized according to quartile of relative

abundance. Batch identity was also included as covariate in all non-IPF ILD modeling to account for potential batch effects.

Exposure was defined categorically according to quartile of protein relative abundance, with first and fourth quartiles compared. Mediator was defined continuously according to one-year decline in FVC relative to baseline. One-year change in FVC was first estimated for each individual by fitting a linear regression model with available FVC measures and treating time from blood draw as the independent variable. A linear spline was then placed at zero percent change. Values of FVC decline above the 95th percentile were winsorized(Yang et al., 2024) to mitigate influential outliers. Outcome was defined as two-year restricted mean survival time (RMST),(Royston & Parmar, 2013) which converts survival time to a continuous pseudo-observation for generalized linear modeling. To do this, we first estimated two-year transplant-free survival, defined as the time from blood draw to death, lung transplant or censoring at 24-months or date prior to 24 months if lost to follow-up. This measure was then converted to RMST using the stpmean package in STATA (version 18, College Station, TX). RMST values were normalized to range from 0-1 to allow mediation coefficients to be interpreted as absolute percent change in two-year RMST.

3.3.4 *Statistical analysis*

Causal mediation analysis was performed using the mediate package in STATA. This approach separates the direct and indirect effects of an exposure on an outcome, with the latter occurring through a third variable called a mediator.(VanderWeele, 2016) For this analysis, the natural indirect effect estimated the RMST effect mediated by declining FVC (Figure 3-1A), while the natural direct effect reflected the RMST effect through an alternate causal pathway or residual confounding (Figures 3-1B-C). The total effect reflected the sum of the natural direct and indirect effects. Because causal mediation analysis requires an exposure-outcome relationship, we restricted the analysis to proteins with total effect $p<0.05$ after adjustment for false discovery rate (FDR) using the Benjamini Hochberg procedure.(Benjamini & Hochberg, 1995) Those with natural indirect effect at FDR $p<0.05$ were considered potentially causal of PPF. Exposure-mediator interaction was allowed in all analyses, and robust standard errors were used when estimating causal effects to account for heteroskedasticity that could arise in the exposure-

mediator, exposure-outcome and mediator-outcome relationships. Sensitivity analyses were performed to evaluate the impact of protein categorization and assess the likelihood that our findings were the result of residual confounding. To corroborate results of the mediation analysis, a traditional association study was performed by fitting a log binomial model with covariates used to satisfy mediation assumptions 1 and 2 described above and treating FVC decline as a categorical measure at a threshold of ≥10%, which is among the most reliable measures of PPF given the strong and reproducible association with death. (Durheim et al., 2015; Ley et al., 2011; Pugashetti et al., 2023)

3.3.5 *Protein Interaction Network Analyses*

Ingenuity Pathway Analysis (QIAGEN Inc., Redwood City, CA) was performed to identify molecular pathways identified by proteins classified as potentially causal of PPF. Overrepresented pathways were identified using an FDR adjusted, right-tailed Fisher's exact test, which accounted for the number of overlapping proteins and the total number of known proteins in each pathway. To explore potential protein-protein interactions, we constructed an inference-based interaction network by treating each protein as a mediator of the relationship between other proteins (exposure) and continuous FVC decline (outcome). Proteins with outcome association that was completely mediated by another protein were considered upstream of the mediator protein on the causal pathway. Proteins displaying the strongest mediation were selected when evidence of bidirectional mediation was observed. Results are displayed using a heat map and directed causal network. A biology-based protein interaction network was constructed using String software(Szklarczyk et al., 2019) at a 25% confidence threshold to assess for potential biological interactions between proteins identified through our inference-based interaction network.

3.4 ***Results***

3.4.1 *Cohort Characteristics*

Of 1299 and 1394 individuals with IPF and non-IPF ILD who underwent proteomic profiling, 904 and 974 had serial FVC measures and complete covariate data and were included in the final analysis (Figure 3-3). As expected, there were notable differences between demographics and treatment exposure for IPF and non-IPF ILD cohorts, with the IPF cohort being older and with

higher proportions of ever smokers and anti-fibrotic exposed individuals (Table 3-1). The non-IPF ILD cohort was more racially diverse and had a higher proportion of women and immunosuppressant exposed individuals (Table 3-1).

3.4.2 *Outcomes*

During the 24-month observation period, death or transplant occurred in 183 (20.2%) individuals with IPF (death n=127, transplant n=56). Mean one-year change in FVC was -4.4% (±12.6%) and each percent decline in FVC was associated with a 1% absolute decrease in RMST (coefficient -1.05; 95% CI -1.18, -0.92%; p<0.0001). This corresponded with a 10% increased hazard of death or lung transplant when using a Cox proportional hazards model (HR 1.10; 95% CI 1.09, 1.12; p<0.0001) (Table 3-2). Among those with non-IPF ILD, death or transplant occurred in 120 (12.3%) individuals (death n=79, transplant n=41). Mean one-year change in FVC for the overall cohort was -1.8% (±16.4%), with fibrotic hypersensitivity pneumonitis (-3.6% ±16.9%) displaying the largest numerical one-year decline in FVC, followed by non-IPF idiopathic interstitial pneumonia (-2.1% ±17.5%) and connective tissue disease-associated ILD (-0.5% ±15.5%) (Table 3-2). Each percent decline in FVC was associated with a similar hazard of death or lung transplant as observed in IPF, but lower absolute change in RMST, which reflected the lower cumulative incidence of non-IPF ILD outcomes (Table 3-2).

3.4.3 *Causal Mediation Analysis*

Of the 140 proteins previously associated with three-year IPF TFS, 64 were associated with differential two-year RMST at total effect FDR p<0.05 and advanced for causal mediation analysis. Of these 64 proteins, 28 had RMST effect that was mediated by declining FVC, suggesting potential causality. Proteins with the largest mediated RMST effect mediated were integrin beta 6 (ITGB6), amphiregulin (AREG), Kunitz-type protease inhibitor 1 (SPINT1), latent transforming growth factor beta binding protein 2 (LTBP2) and intracellular adhesion molecule 5 (ICAM5) (Table 3-3). RMST was shorter (negative absolute change) for the 4th versus 1st quartile of biomarker concentration for 23/38 proteins and longer (positive absolute change) for 5/28 proteins (Table 3-3).

Complete mediation, defined as natural direct effect FDR p>0.05, was observed for 22/28 proteins, suggesting that nearly all RMST effect occurred through the PPF-related pathway. Sensitivity analyses showed similar findings when stratifying proteins according to tertiles, quintiles and deciles, with magnitude of effect generally increasing as the number of strata was increased, supporting a dose-dependent effect (Table 3-4). Confounding sensitivity analysis also suggested that results were moderately robust to unmeasured confounding, with e-values ≥1.5 for 23/28 proteins (Table 3-5), indicating that a unmeasured confounder with risk ratio greater than this e-value would be required to attenuate the observed mediation.(VanderWeele & Ding, 2017) A traditional association study corroborated these results, showing that 26/28 proteins were associated with a categorical decline in FVC of ≥10% (Table 3-6).

Of the 44 proteins previously associated with differential TFS in those with non-IPF ILD, 13 were associated with two-year RMST at FDR p<0.05 and advanced for causal mediation analysis (Table 3-7). Of these 13 proteins, five had RMST effect that was mediated by declining FVC, suggesting potential causality. These five mediated proteins were AREG, ITGB6, LTBP2, carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6), Keratin, type I cytoskeletal 19 (KRT19), all of which were also mediated by declining FVC in the IPF cohort, suggesting shared pathobiology across fibrotic ILD subtypes. When testing other mediated proteins identified through the IPF analysis, ICAM5 and SPINT1 were also mediated by declining FVC in the non-IPF ILD cohort (Table 3-8). Confounding sensitivity analysis suggested that these results were also moderately robust to unmeasured confounding, with e-values ≥1.5 for 7/7 mediated proteins (Tables 3-9).

3.4.4 *Pathway and Interaction Network Analysis*

Ingenuity pathway analysis of proteins classified as potentially causal of PPF identified eleven biological pathways that were overrepresented by these proteins. Top pathways included the IL-33 signaling, FXR/RXR activation, O-linked glycosylation, immunoregulatory interactions and glycation signaling (Figure 3-4). An inference-based protein interaction network (Figure 3-5) developed through protein-protein mediation analysis (Figure 3-6) showed that 17/28 proteins were upstream of at least one other protein on the causal pathway, with most mediated by either ITGB6 and/or AREG. A biology-driven protein interaction network confirmed known or

potential biological interactions for most proteins identified through this inference-based approach (Figure 3-7). Direct ITGB6 interactions were observed for CEACAM6, KRT19 and LTBP2, while direct AREG interactions were observed for CEACAM6 and KRT19. Indirect interactions for most other proteins were observed through ICAM1, syndecan 1 (SDC1) and Killer cell lectin-like receptor subfamily B member 1 (KLRB1). This biology-based approach confirmed minimal protein-protein interaction for GSTA3, ASAH2 and BCAT1. It also suggested more interaction for KLRB1 than was detected in the inference-based approach. Similar findings were observed when repeating this analysis in the non-IPF ILD cohort, with most upstream proteins having FVC decline effect that was mediated through AREG (Figure 3-8). In this analysis, ITGB6 was classified as upstream of AREG, CEACAM6 and ICAM5, while LTBP2 and SPINT1 remained upstream of ITGB6 (Figure 3-8).

3.5 ***Discussion***

In this study, we identified 28 prognostic protein biomarkers with survival effect that was mediated by declining FVC, suggesting a potentially causal relationship with PPF. Seven of these 28 proteins showed consistent findings in a similarly defined non-IPF ILD cohort, suggesting shared PPF pathobiology across ILD subtypes. Sensitivity analyses confirmed that results were robust to protein categorization strategies and would require one or more unmeasured confounders with at least moderate effect size to attenuate results. Pathway analysis confirmed that key fibrogenic pathways were overrepresented among these proteins. Finally, we used inference-based interaction network analysis to establish a directed causal network to characterize protein-protein interaction. A biology-based interaction network confirming potential biological relationships for most protein-protein interactions identified through this inference-based approach. To our knowledge, this study is the first to use causal mediation analysis to identify protein that play a potentially causal role in PPF. Amid the rapidly growing number of prognostic biomarkers being linked to survival in patients with ILD, this helps clarify which protein biomarkers identified through association studies are most likely to contribute to disease outcomes, thereby prioritizing these proteins and their associated molecular pathways for therapeutic consideration.

Among the 28 proteins identified as potentially causal of PPF in our IPF cohort, ITGB6, AREG, KRT19, LTBP2, SPINT1, ICAM5, and CEACAM6 were similarly mediated by declining FVC in a large non-IPF ILD cohort. ITGB6 makes up the β6 subunit of integrin αvβ6, which activates latent transforming growth factor beta 1 (TGF-β1), leading to fibroblast-to-myofibroblast transition and collagen deposition in the lungs and elsewhere.(Saini et al., 2015) Though a strong candidate for therapeutic intervention, αvβ6 blockade has proven challenging in patients with IPF, with two recent clinical trials stopped due to safety concerns. AREG is an epidermal growth factor receptor ligand, which like ITGB6, activates TGF-β.(Liu et al., 2016) A recent study showed that alveolar AREG can lead to fibrotic remodeling through EGFR-mediated fibroblast activation.(Zhao et al., 2024) AREG blockade has been shown to attenuate pulmonary fibrosis in mouse models(Lindzen et al., 2021; Zhao et al., 2024) and warrants consideration as a potential therapeutic intervention for PPF.

LTBP2 supports the structural integrity of the extracellular matrix and leads to fibroblast-to-myofibroblast differentiation through NF-κB.(Zou et al., 2021) Importantly, LTBP2 silencing was shown to suppress fibroblast-to-myofibroblast differentiation, suggesting therapeutic potential. KRT19, long known to represent a marker of epithelial injury, also regulates several pathways critical to pulmonary fibrosis, including Wnt and Notch.(Saha et al., 2019) KRT19 silencing has shown promise for treating colon cancer and represents a potential target for pulmonary fibrosis.(Saha et al., 2019) SPINT1, also known as HAI-1, is increased in IPF fibroblasts and has been shown to prevent the activation of hepatocyte growth factor, which regulates fibroblast activity, potentially protecting against pulmonary fibrosis.(Marchand-Adam et al., 2006) SPINT1/HAI-1 blockade is being explored for several types of malignancy. Less is known about role ICAM5 and CEACAM6 may play in PPF; however, ICAM5 was highly correlated with ICAM1 in this analysis (ρ=0.73), which was also mediated by declining FVC in the IPF cohort and is known to cause fibrosis through inflammatory cytokine-mediated activation of TGF-β1.(Hamaguchi et al., 2002)

While our inference-based protein interaction network suggests that each of the aforementioned proteins exerted an effect on FVC decline through ITGB6 and AREG-related pathways, several proteins appeared to exert an effect through another pathway, including MZB1, CCL17, GSTA3,

BCAT1 and KLRB1. MZB1 localizes to antibody secreting B cells and expression is widespread throughout fibrotic skin and lung tissue relative to healthy control specimens.(Schiller et al., 2017) MZB1 deficient mice were recently shown to be protected from bleomycin-induced pulmonary fibrosis.(Bueno et al.) CCL17 has been shown to promote fibroblast activation through the TGF-β/Smad pathway, which can be attenuated through CCL17 antibody blockade.(Wang et al., 2023) GSTA3 was one of two protective proteins identified through our analysis and supports in vitro data showing that this protein attenuates renal fibrosis through inhibition of TGF-β.(Xiao et al., 2016) BCAT1 is a cytosolic protein that cause reversible transamination of branched-chain amino acids and has been shown to facilitate hypoxia-induced autophagy,(Xin et al., 2020) which has been suggested as a potential driver of pulmonary fibrogenesis.(Hill & Wang, 2022) KLRB1, also known as CD161, is a cell surface marker expressed primarily in natural killer cells and T cell subsets, which have been implicated in systemic sclerosis-related fibrotic lung disease.(Padilla et al., 2024)

A notable lack of mediation was also observed for several proteins, suggesting that their association with TFS represents an epiphenomenon. The first is MUC16, also called CA125, which is a well described marker of epithelial injury and has been proposed as a biomarker of ILD activity given prior studies establishing an association with ILD survival.(Alqalyoobi et al., 2020; Maher et al., 2020) The lack of mediation for MUC16/CA125 serves as an important negative control for this study, as this is among the best examples of a reactive, rather than causal, biomarker.(Charkhchi et al., 2020) Our findings for MUC16/CA125 stand in stark contrast to those for KRT19, which appears to reflect both epithelial injury and potentially contribute to PPF. Mediation was also absent for MMP7, which was unexpected given its longstanding link to fibrotic remodeling.(Craig et al., 2015) While MMP7 is strongly associated with survival, our results corroborate other studies showing inconsistent association with declining FVC, leaving it unclear whether MMP7 targeting will impact PPF.(Bauer et al., 2017; Neighbors et al., 2018; Raghu et al., 2018) Like BCAT1, FASLG has been implicated in lung fibrogenesis through its role in apoptosis. Studies have shown the FAS/FASLG pathway facilitates myofibroblast apoptosis in the IPF lungs. FASLG deficient myofibroblasts are able to avoid apoptosis, preventing normal wound repair. Our data support the protective role of FASLG, but suggest that FASLG deficiency is a byproduct, rather than driver of fibrosis.

Beyond the individual proteins identified in our study, the pathway analysis of these proteins supports a strong immunological contribution to fibrogenesis. IL33 signaling was the top pathway identified by our list of mediated proteins, which has been linked to fibrogenesis through interaction with TGF-β1.(Di Carmine et al., 2022) Other pathways identified include those relevant to lymphoid signaling, macrophages, and tumor necrosis factor. While immunosuppression leads to worse outcomes in patients with IPF,(Idiopathic Pulmonary Fibrosis Clinical Research et al., 2012) our results suggest that targeted immunomodulation of fibrogenic cytokines and other inflammatory markers still has potential to slow PPF. Beyond immunological pathways, the farnesoid X receptor (FXR) has been linked to liver fibrosis and FXR agonists have been shown to have anti-fibrotic effects in the liver, suggesting a potential role in lung fibrosis.(Zhou et al., 2020)

While mediation analysis cannot definitively establish causation between an exposure and outcome, causal interpretations can be gleaned when key assumptions are satisfied.(VanderWeele, 2016) For our results to be valid, one must first accept that there exists a causal relationship between declining lung function and death. While this cannot be ethically confirmed in humans through controlled studies, lung transplantation has been shown to prolong survival in patients with disease characterized by declining lung function,(Vock et al., 2017) providing indirect evidence of a causal relationship. One must also accept that the proteins evaluated in this study could plausibly cause lung function decline through a pathological mechanism. Pathway analysis showed that several key fibrogenic pathways, were overrepresented among proteins identified through our mediation analysis, providing indirect evidence of causality through the laboratory-based studies informing these pathways. Residual confounding must also be minimal for these assumptions to hold. Our sensitivity analysis produced e-values of >1.5 for most mediated proteins, suggesting that unmeasured confounders with this effect size would be required to attenuate results. While we attempted to adjust for known factors that could plausibly impact exposure-mediator, exposure-outcome and mediator-outcome relationships, several recently recognized social determinants of health(DeDent et al., 2022; Goobie et al., 2022) were not measured in these cohorts.

3.6 Conclusion

Here we employed causal mediation analysis to glean novel insights into PPF pathobiology. While most biomarkers identified through this analysis have an established role in fibrogenesis, other represent novel findings. Additional research will be needed to better understand why these novel protein biomarkers may be causally linked to PPF. By identifying proteins potentially causal of PPF, this work may help inform future responder analysis in clinical trials and prioritize proteins that warrant consideration for therapeutic development.

Table 3.1. Baseline characteristics of IPF and non-IPF ILD cohorts		
Characteristic	**IPF (n=904)**	**non-IPF ILD (n=974)**
Age, mean (±SD)	71.3 (7.6)	62.7 (12.5)
Female sex, n (%)	684 (75.7)	397 (40.8)
Race/ethnicity, n (%)		
White	841 (93.0)	741 (76.1)
Black	13 (1.4)	94 (9.7)
Hispanic	9 (1.0)	68 (7.0)
Asian	28 (3.1)	55 (5.7)
Other/Unknown	13 (1.4)	16 (1.6)
ILD Diagnosis, n (%)		
IPF	904 (100)	0 (0)
Connective tissue disease-associated ILD	0 (0)	426 (43.7)
Fibrotic hypersensitivity pneumonitis	0 (0)	264 (27.1)
Non-IPF idiopathic interstitial pneumonia	0 (0)	284 (29.2)
Ever smoker, n (%)	561 (62.1)	407 (41.8)
FVC %, mean (±SD)	70.8 (16.9)	68.2 (19.0)
DLCO %, mean (±SD)	44.1 (16.2)	48.1 (18.0)
Anti-fibrotic exposure, n (%)	731 (80.9)	126 (12.9)
At blood draw, n (%)	483 (53.4)	40 (4.1)
Following blood draw, n (%)	248 (27.4)	86 (8.8)
Immunosuppressant exposure , n (%)	102 (11.3)	669 (68.7)
At blood draw, n (%)	77 (8.5)	482 (49.5)
Following blood draw, n (%)	25 (2.8)	187 (19.2)
Abbreviations: ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis; FVC = forced vital capacity; DLCO = diffusion capacity of the lung for carbon monoxide		

Table 3.2. Mean one-year relative decline in FVC and unadjusted association with restricted mean survival time and transplant-free survival

Cohort	Mean annual percent decline in FVC (SD)	RMST (Linear Model)				TFS (Cox Model)			
		% change*	95% CI Low	95% CI High	p-value	HR**	95% CI Low	95% CI High	p-value
IPF	-4.4% (12.6%)	-1.10%	-1.18	-0.92	<0.0001	1.10	1.09	1.12	<0.0001
Non-IPF	-1.8% (16.4%)	-0.66%	-0.74	-0.57	<0.0001	1.11	1.09	1.13	<0.0001
CTD-ILD	-0.5% (15.5%)	-0.58%	-0.72	-0.44	<0.0001	1.11	1.08	1.13	<0.0001
fHP	-3.6% (16.9%)	-0.75%	-0.93	-0.58	<0.0001	1.11	1.09	1.15	<0.0001
non-IPF IIP	-2.1% (17.5%)	-0.65%	-0.8	-0.5	<0.0001	1.12	1.08	1.15	<0.0001

* Relative change in restricted mean survival time per 1% decline in FVC

* Absolute change in per hazard of death or transplant 1% decline in FVC

Abbreviations: IPF = idiopathic pulmonary fibrosis; ILD = interstitial lung disease; CTD = connective tissue disease; fHP = fibrotic hypersensitivity pneumonitis; IIP = idiopathic interstitial pneumonia: HR = hazard ratio; CI = confidence interval; RMST = restricted mean survival time; TFS = transplant-free survival

Table 3.3. Proteins classified as potentially causal of PPF based on mediation by declining FVC

Protein	Natural Indirect Effect				Natural Direct Effect			
	% RMST Change	CI Low	CI High	FDR P	% RMST Change	CI Low	CI High	FDR P
ADAMTS13	1.84%	0.48%	3.21%	2.34E-02	3.22%	0.05%	6.39%	1.43E-01
ADAMTS16	-3.48%	-5.49%	-1.47%	8.95E-03	-1.61%	-4.38%	1.16%	2.84E-01
AREG*	-5.31%	-7.90%	-2.73%	1.79E-03	-2.80%	-5.74%	0.13%	1.31E-01
ASAH2	2.21%	0.56%	3.87%	2.45E-02	1.94%	-1.21%	5.10%	2.76E-01
BCAT1	-4.30%	-6.76%	-1.83%	1.00E-02	-1.32%	-4.00%	1.36%	3.45E-01
CCL17	-3.29%	-5.48%	-1.11%	1.69E-02	-1.67%	-4.29%	0.96%	2.85E-01
CEACAM6*	-4.57%	-7.25%	-1.90%	7.39E-03	-3.56%	-6.44%	-0.67%	6.40E-02
COL24A1	-4.16%	-6.57%	-1.75%	7.73E-03	-4.50%	-7.30%	-1.69%	1.40E-02
CSF2	-3.02%	-5.06%	-0.99%	1.72E-02	-4.05%	-6.74%	-1.37%	1.68E-02
CTHRC1	-3.31%	-5.60%	-1.02%	1.99E-02	-2.02%	-4.98%	0.94%	2.83E-01
EDA2R	-2.92%	-5.21%	-0.62%	3.14E-02	-1.94%	-5.02%	1.13%	2.75E-01
FCRL5	-2.65%	-4.49%	-0.81%	1.92E-02	-2.01%	-5.01%	1.00%	2.80E-01
GALNT5	-2.98%	-5.28%	-0.67%	3.04E-02	-4.04%	-6.78%	-1.30%	1.87E-02
GDF15	-2.70%	-4.69%	-0.71%	2.38E-02	-1.94%	-4.92%	1.04%	2.83E-01
GSTA3	2.05%	0.60%	3.51%	2.11E-02	3.25%	-0.19%	6.69%	1.28E-01
ICAM1	-3.05%	-5.22%	-0.88%	2.06E-02	-2.52%	-5.31%	0.28%	1.46E-01
ICAM5*	-4.07%	-6.47%	-1.68%	6.95E-03	-5.01%	-7.92%	-2.11%	2.80E-02
IGLC2	-2.34%	-4.06%	-0.63%	2.36E-02	-2.16%	-4.93%	0.61%	2.22E-01
ITGB6*	-6.12%	-8.92%	-3.32%	1.20E-03	-3.42%	-6.48%	-0.36%	1.02E-01
KLRB1	2.26%	0.79%	3.74%	1.52E-02	3.16%	-0.02%	6.34%	1.43E-01
KRT19*	-3.22%	-5.51%	-0.93%	1.95E-02	-1.52%	-4.37%	1.33%	3.19E-01
LTBP2*	-4.72%	-7.50%	-1.93%	6.40E-03	-2.93%	-5.94%	0.07%	1.31E-01
MZB1	-3.59%	-5.59%	-1.59%	9.21E-03	-2.80%	-5.63%	0.03%	1.35E-01
PRSS22	-3.10%	-5.18%	-1.01%	1.65E-02	-1.74%	-4.56%	1.08%	2.65E-01
SDC1	2.78%	-4.96%	-0.60%	3.18E-02	-1.27%	-4.25%	1.72%	4.06E-01
SPINT1*	-4.28%	-6.82%	-1.74%	6.19E-03	-4.52%	-7.37%	-1.68%	1.87E-02
SPON1	-2.54%	-4.69%	-0.38%	4.77E-02	-2.06%	-5.00%	0.89%	2.82E-01
TNFSF13B	-2.75%	-4.93%	-0.58%	3.11E-02	-4.57%	-7.32%	-1.81%	1.40E-02

The natural indirect effect represents the amount of total RMST effect mediated by declining FVC. The natural direct effect represents the remaining amount of total RMST effect that occurs through other pathways.

* Also mediated by declining FVC in non-IPF ILD cohort

Abbreviations: PPF = progressive pulmonary fibrosis; RMST = restricted mean survival time; CI = confidence interval; FDR = false discovery rate.

Table 3.4. Natural direct effect for mediated proteins when using different stratification thresholds

Protein	Tertiles		Quintiles		Deciles	
	% RMST Change	FDR P	% RMST Change	FDR P	% RMST Change	FDR P
ADAMTS13	1.79%	6.78E-03	2.08%	5.78E-03	2.23%	9.62E-03
ADAMTS16	-3.79%	2.03E-03	-4.36%	1.91E-04	-5.27%	8.25E-04
AREG	-4.29%	1.29E-03	-5.81%	5.60E-05	-7.25%	2.28E-04
ASAH2	1.88%	1.45E-02	2.53%	3.61E-03	2.66%	1.00E-02
BCAT1	-3.70%	2.80E-03	-5.22%	2.07E-04	-6.23%	1.10E-03
CCL17	-2.72%	1.16E-02	-3.03%	1.17E-02	-4.21%	9.54E-03
CEACAM6	-3.69%	3.88E-03	-6.03%	1.45E-04	-7.22%	1.30E-03
COL24A1	-4.05%	1.46E-03	-4.92%	2.33E-04	-6.01%	9.52E-04
CSF2	-2.37%	1.22E-02	-3.45%	1.94E-03	-4.01%	3.44E-03
CTHRC1	-2.17%	2.70E-02	-3.09%	1.24E-02	-3.80%	1.01E-02
EDA2R	-2.98%	7.07E-03	-2.78%	2.51E-02	-4.06%	9.85E-03
FCRL5	-2.16%	1.32E-02	-2.77%	4.11E-03	-3.15%	9.22E-03
GALNT5	-2.84%	9.27E-03	-2.63%	3.00E-02	-3.62%	1.44E-02
GDF15	-2.97%	6.73E-03	-3.44%	5.61E-03	-4.49%	3.77E-03
GSTA3	2.05%	7.91E-03	2.29%	5.11E-03	2.39%	1.01E-02
ICAM1	-3.85%	2.01E-03	-4.08%	1.28E-03	-5.14%	2.61E-03
ICAM5*	-3.33%	3.85E-03	-3.79%	4.11E-03	-4.96%	3.50E-03
IGLC2	-1.49%	5.43E-02	-2.01%	2.93E-02	-2.82%	1.03E-02
ITGB6*	-5.48%	2.76E-04	-6.71%	1.90E-05	-8.61%	1.50E-04
KLRB1	1.28%	6.15E-02	1.92%	8.54E-03	2.18%	1.17E-02
KRT19*	-2.61%	1.26E-02	-3.60%	4.11E-03	-4.41%	5.23E-03
LTBP2*	-4.59%	1.07E-03	-4.81%	1.24E-03	-6.45%	1.50E-03
MZB1	-2.99%	3.12E-03	-3.71%	6.26E-04	-4.43%	1.03E-03
PRSS22	-2.63%	9.23E-03	-2.87%	1.31E-02	-3.71%	8.68E-03
SDC1	-2.60%	8.67E-03	-2.84%	1.35E-02	-3.76%	1.02E-02
SPINT1*	-3.79%	3.04E-03	-4.86%	6.50E-04	-6.15%	1.18E-03
SPON1	-2.27%	2.22E-02	-2.68%	2.38E-02	-3.37%	1.83E-02
TNFSF13B	-2.50%	1.43E-02	-3.03%	1.28E-02	-3.59%	1.23E-02

Table 3.5. E-values for proteins found to be potentially causal of PPF in IPF cohort				
Protein	**NIE Coefficient**	**Standardized NIE**	**NIE_~RR**	**E-value**
ADAMTS13	0.018	0.095	1.090	1.404
ADAMTS16	-0.035	0.179	1.177	1.634
AREG	-0.053	0.274	1.283	1.886
ASAH2	0.022	0.114	1.109	1.458
BCAT1	-0.043	0.222	1.223	1.746
CCL17	-0.033	0.170	1.167	1.608
CEACAM6	-0.046	0.236	1.239	1.784
COL24A1	-0.042	0.214	1.215	1.727
CSF2	-0.030	0.156	1.152	1.572
CTHRC1	-0.033	0.170	1.168	1.611
EDA2R	-0.029	0.150	1.147	1.557
FCRL5	-0.026	0.136	1.132	1.519
GALNT5	-0.030	0.153	1.150	1.565
GDF15	-0.027	0.139	1.135	1.527
GSTA3	0.021	0.106	1.101	1.435
ICAM1	-0.031	0.157	1.154	1.575
ICAM5	-0.041	0.210	1.211	1.715
IGLC2	-0.023	0.121	1.116	1.476
ITGB6	-0.061	0.315	1.332	1.998
KLRB1	0.023	0.117	1.112	1.465
KRT19	-0.032	0.166	1.163	1.599
LTBP2	-0.047	0.243	1.248	1.804
MZB1	-0.036	0.185	1.183	1.649
PRSS22	-0.031	0.160	1.156	1.581
SDC1	0.028	0.143	1.139	1.538
SPINT1	-0.043	0.221	1.222	1.743
SPON1	-0.025	0.131	1.126	1.503
TNFSF13B	-0.028	0.142	1.138	1.534
Abbreviations: NIE = natural indirect effect; ~RR = approximated risk ratio using NIE coefficient				
Standardized NIE = NIE coefficient/RMST standard deviation of 0.194				
NIE_~RR =0.91*exp(Standardized NIE)				
E-value = NIE_~RR + sqrt|NIE_~RR*(NIE_~RR-1)|				

Table 3.6. Results of traditional association study testing the relationship between mediated proteins and categorical FVC decline.

Protein	RR	CI Low	CI High	p-value	FDR p-value
ITGB6	1.637	1.364	1.963	1.10E-07	3.15E-06
AREG	1.334	1.178	1.510	5.50E-06	7.66E-05
ADAMTS16	1.345	1.179	1.535	1.00E-05	9.63E-05
LTBP2	1.481	1.227	1.787	4.40E-05	3.06E-04
COL24A1	1.333	1.158	1.535	6.50E-05	3.65E-04
SPINT1	1.493	1.210	1.840	1.79E-04	8.38E-04
MZB1	1.306	1.127	1.514	3.86E-04	1.55E-03
ICAM1	1.502	1.194	1.891	5.19E-04	1.82E-03
BCAT1	1.302	1.111	1.526	1.11E-03	3.46E-03
ASAH2	0.816	0.721	0.924	1.32E-03	3.71E-03
GDF15	1.288	1.101	1.506	1.54E-03	3.92E-03
FCRL5	1.206	1.071	1.358	2.04E-03	4.76E-03
ICAM5	1.300	1.095	1.543	2.72E-03	5.86E-03
CSF2	1.276	1.085	1.501	3.27E-03	6.53E-03
SDC1	1.252	1.072	1.462	4.55E-03	8.49E-03
KRT19	1.175	1.048	1.319	5.84E-03	1.02E-02
IGLC2	1.445	1.101	1.896	7.98E-03	1.31E-02
ADAMTS13	0.729	0.577	0.922	8.45E-03	1.32E-02
GALNT5	1.290	1.054	1.580	1.36E-02	2.01E-02
CTHRC1	1.282	1.045	1.572	1.73E-02	2.43E-02
CCL17	1.086	1.014	1.164	1.87E-02	2.49E-02
CEACAM6	1.145	1.019	1.287	2.30E-02	2.92E-02
EDA2R	1.195	1.023	1.397	2.50E-02	3.04E-02
SPON1	1.301	1.032	1.638	2.57E-02	3.00E-02
GSTA3	0.899	0.817	0.989	2.93E-02	3.29E-02
TNFSF13B	1.211	1.008	1.454	4.03E-02	4.34E-02
PRSS22	1.216	1.001	1.477	4.83E-02	5.01E-02
KLRB1	0.856	0.708	1.034	1.07E-01	1.07E-01

Abbreviations: RR = risk ratio; CI = confidence interval. Abbreviation list in vii describes proteins

Table 3.7. Causal mediation analysis results for selection of mediated IPF proteins when tested in non-IPF ILD cohort.

Protein	Natural Indirect Effect				Natural Direct Effect			
	RMST Change	95% CI Low	95% CI High	FDR p-value	RMST Change	95% CI Low	95% CI High	FDR p-value
AREG	-2.94%	-4.78%	-1.10%	1.30E-02	-3.55%	-5.47%	-1.63%	1.91E-03
CEACAM6	-2.94%	-4.71%	-1.18%	1.30E-02	-0.30%	-2.15%	1.54%	7.49E-01
ITGB6	-2.11%	-3.68%	-0.53%	2.34E-02	-1.01%	-2.88%	0.86%	1.43E-02
KRT19	-2.16%	-3.70%	-0.63%	1.95E-02	-2.33%	-4.07%	-0.58%	5.75E-03
LTBP2	-2.48%	-4.23%	-0.73%	2.17E-02	-3.07%	-5.29%	-0.85%	5.85E-03
COL24A1	-0.33%	-1.58%	0.91%	7.09E-01	-2.70%	-4.76%	-0.65%	3.13E-01
GDF15	-1.36%	-3.12%	0.40%	1.69E-01	-3.13%	-5.14%	-1.13%	1.43E-02
IL1RL1	-1.12%	-2.44%	0.20%	1.80E-01	-3.01%	-4.97%	-1.04%	1.25E-02
SERPINA3	-1.38%	-2.81%	0.06%	1.32E-01	-2.12%	-4.40%	0.15%	7.95E-02
TGFBR1	0.02%	-1.13%	1.17%	9.72E-01	-3.49%	-5.57%	-1.41%	4.34E-03
THBS2	-1.29%	-2.84%	0.25%	1.63E-01	-3.70%	-5.47%	-1.94%	5.15E-04
TNFRSF10B	-0.18%	-2.00%	1.63%	9.12E-01	-3.58%	-5.77%	-1.38%	4.62E-03
VSIG2	-1.08%	-2.45%	0.29%	1.76E-01	-2.55%	-4.51%	-0.58%	1.44E-02

The natural indirect effect represents the amount of total RMST effect mediated by declining FVC.

The natural direct effect represents the remaining amount of total RMST effect that occurs through other pathways.

Highlighted FDR p-values indicate statistical significance.

Table 3.8. Causal mediation analysis results for proteins specific to non-IPF ILD survival when tested in non-IPF ILD cohort.

Protein	Natural Indirect Effect				Natural Direct Effect			
	RMST Change	95% CI Low	95% CI High	FDR p-value	RMST Change	95% CI Low	95% CI High	FDR p-value
ICAM5	-2.55%	-4.15%	-0.95%	1.20E-02	-3.25%	-5.33%	-1.17%	1.20E-02
SPINT1	-2.25%	-3.76%	-0.74%	9.00E-03	-1.06%	-2.96%	0.83%	2.72E-01
ADAMTS16	-0.90%	-2.11%	0.32%	2.22E-01	-2.24%	-4.24%	-0.24%	5.60E-02
CSF2	-0.63%	-1.84%	0.59%	3.13E-01	-2.86%	-4.84%	-0.88%	1.50E-02
ICAM1	-0.98%	-2.48%	0.52%	2.39E-01	-1.97%	-3.86%	-0.08%	4.92E-02
PRSS22	-1.08%	-2.26%	0.10%	1.48E-01	-1.84%	-3.51%	-0.18%	4.50E-02
ADAMTS13	NA	NA	NA	NA	NA	NA	NA	NA
ASAH2	NA	NA	NA	NA	NA	NA	NA	NA
CCL17	NA	NA	NA	NA	NA	NA	NA	NA
FCRL5	NA	NA	NA	NA	NA	NA	NA	NA
GSTA3	NA	NA	NA	NA	NA	NA	NA	NA
IGLC2	NA	NA	NA	NA	NA	NA	NA	NA
KLRB1	NA	NA	NA	NA	NA	NA	NA	NA
MZB1	NA	NA	NA	NA	NA	NA	NA	NA
SPON1	NA	NA	NA	NA	NA	NA	NA	NA
TNFSF13B	NA	NA	NA	NA	NA	NA	NA	NA

The natural indirect effect represents the amount of total RMST effect mediated by declining FVC. NA indicates that protein did not pass total effect threshold.

The natural direct effect represents the remaining amount of total RMST effect that occurs through other pathways.

Highlighted FDR p-values indicate statistical significance.

Table 3.9. E-values for proteins found to be potentially causal of PPF in non-IPF ILD cohort				
Protein	**NIE Coefficient**	**Standardized NIE**	**NIE_~RR**	**E-value**
AREG	-0.029	0.095	1.090	1.404
CEACAM6	-0.029	0.179	1.177	1.634
ITGB6	-0.021	0.274	1.283	1.886
KRT19	-0.022	0.114	1.109	1.458
LTBP2	-0.025	0.222	1.223	1.746
ICAM5	-0.025	0.170	1.167	1.608
SPINT1	-0.023	0.236	1.239	1.784
Abbreviations: NIE = natural indirect effect; ~RR = approximated risk ratio using NIE coefficient				
Standardized NIE = NIE coefficient/RMST standard deviation of 0.133				
NIE_~RR =0.91*exp(Standardized NIE)				
E-value = NIE_~RR + sqrt|NIE_~RR*(NIE_~RR-1)|				

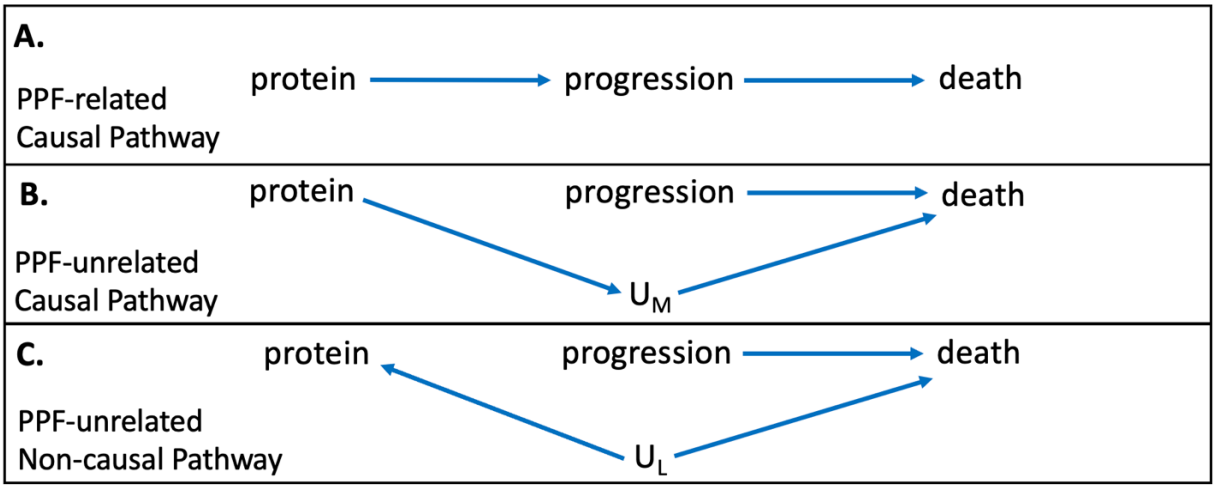


Figure 3.1. Causal framework depicting an association between a prognostic protein biomarker and ILD survival through a causal, PPF-related pathway (a), causal, PPF-unrelated pathway (b) and non-causal pathway due to residual confounding (c). Within this framework, there exists no direct effect of a protein in death/transplant. Abbreviation: U_M = unmeasured mediator; U_L = unmeasured confounder.

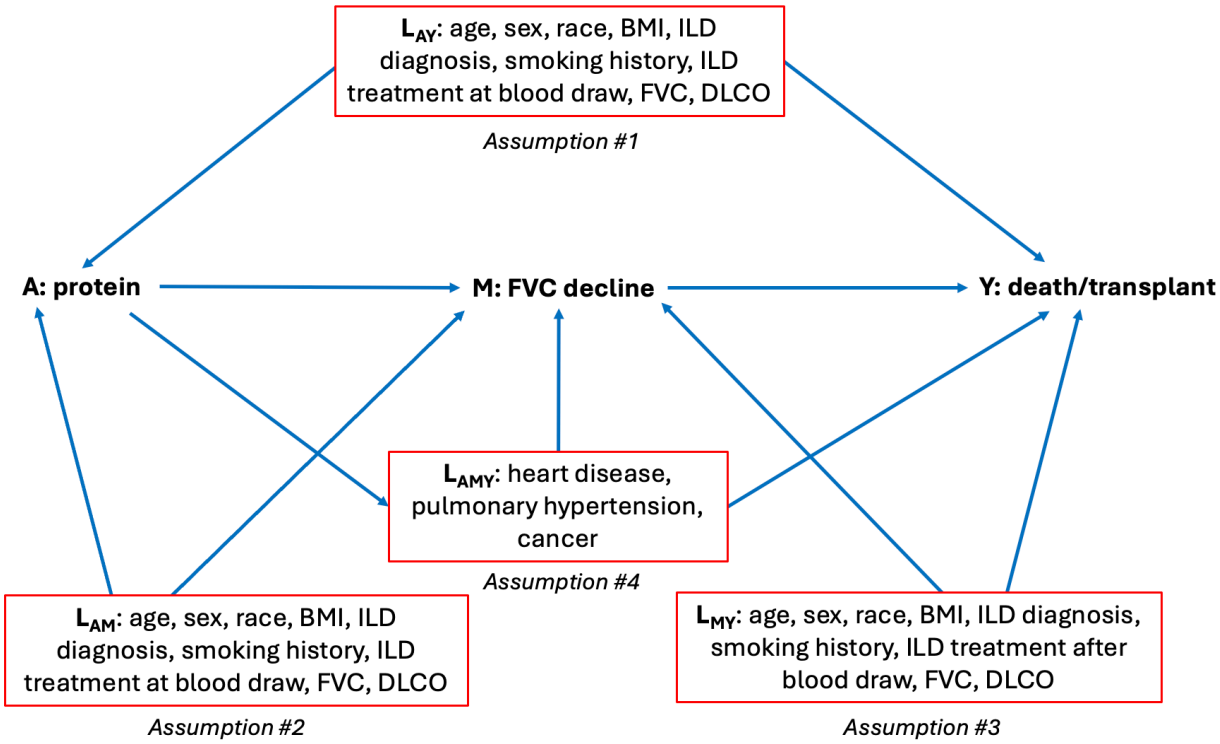


Figure 3.2. Directed acyclic graph showing the relationship between protein biomarker (exposure) and death/transplant (outcome) through FVC decline (mediator). Abbreviations: L_{AY} = confounders of the exposure-outcome relationship; L_{AM} = confounders of the exposure-mediator relationship; L_{MY} = confounders of the mediator-outcome relationship; L_{AMY} = confounders of the mediator-outcome relationship caused by exposure.

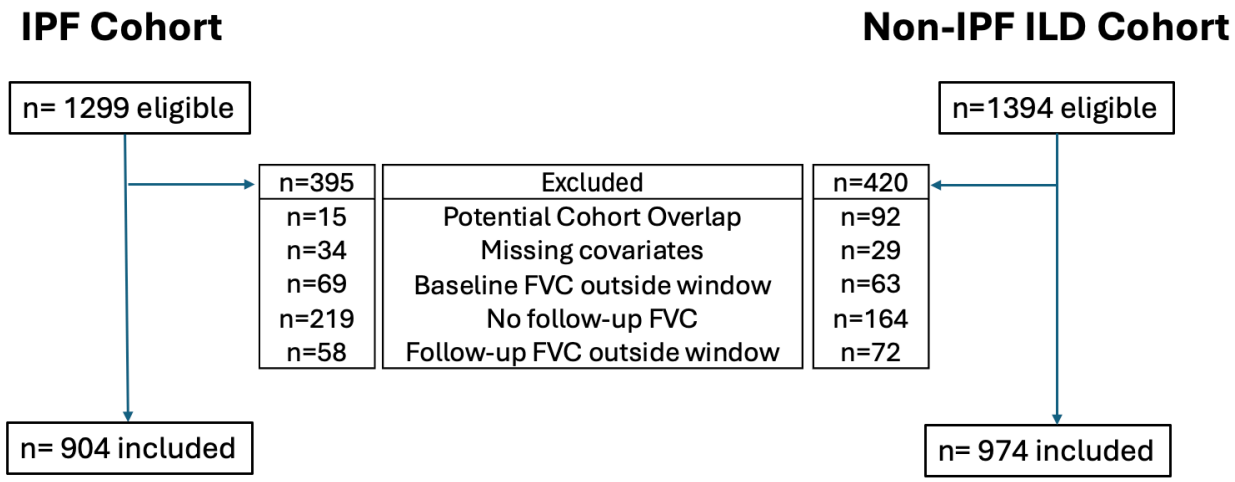


Figure 3.3. Case finding diagram for chapter 3 cohort. Abbreviations: FVC = forced vital capacity; IPF = idiopathic pulmonary fibrosis; ILD = interstitial lung disease.

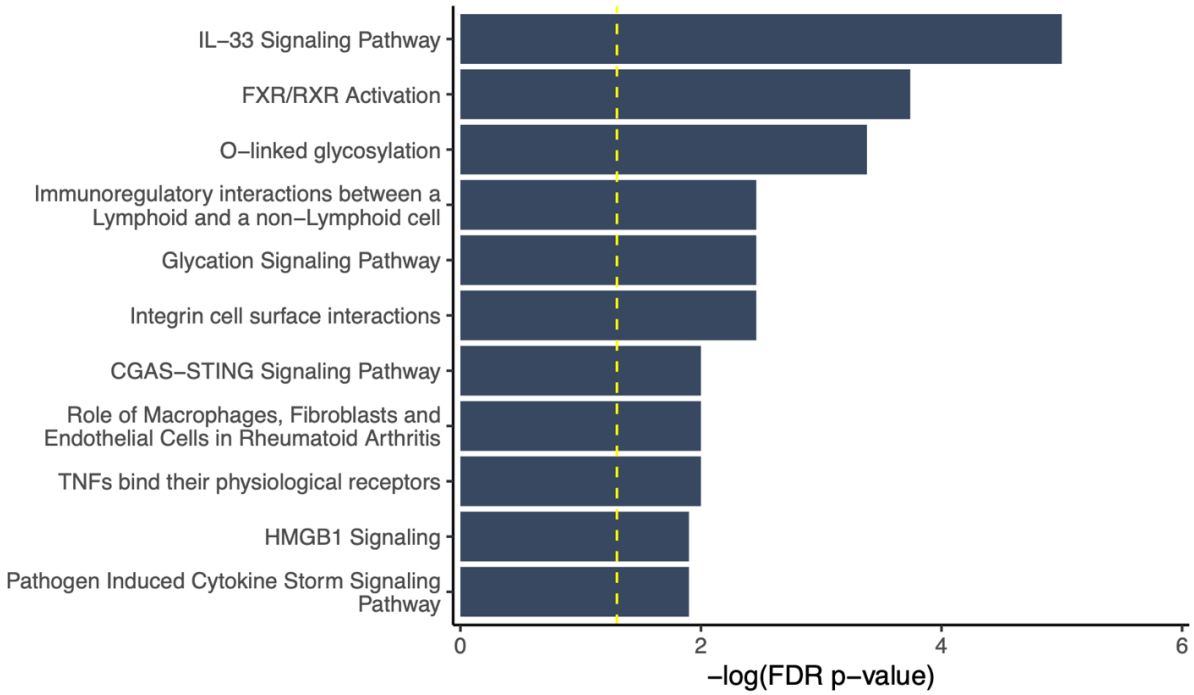


Figure 3.4. Pathway analysis identifying molecular pathways overrepresented by proteins classified as causally related to PPF in an IPF cohort.

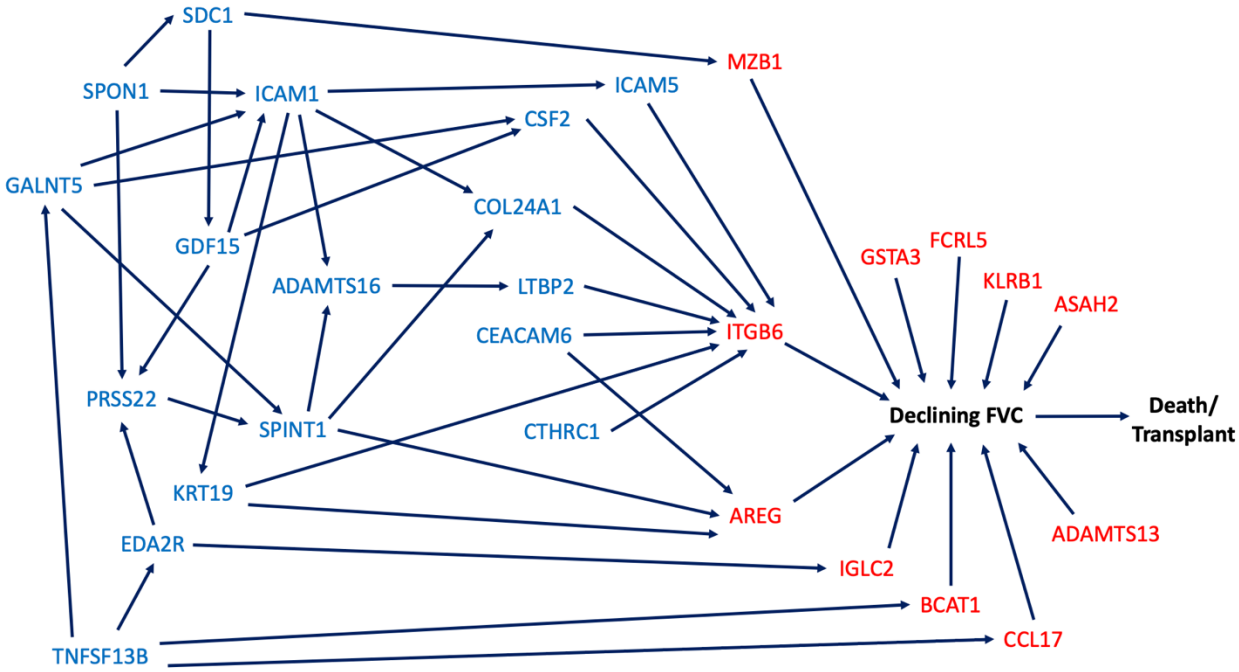


Figure 3.5. Inference-based protein network analyses. Red proteins exert a direct effect on declining FVC, while blue proteins are upstream of red proteins.

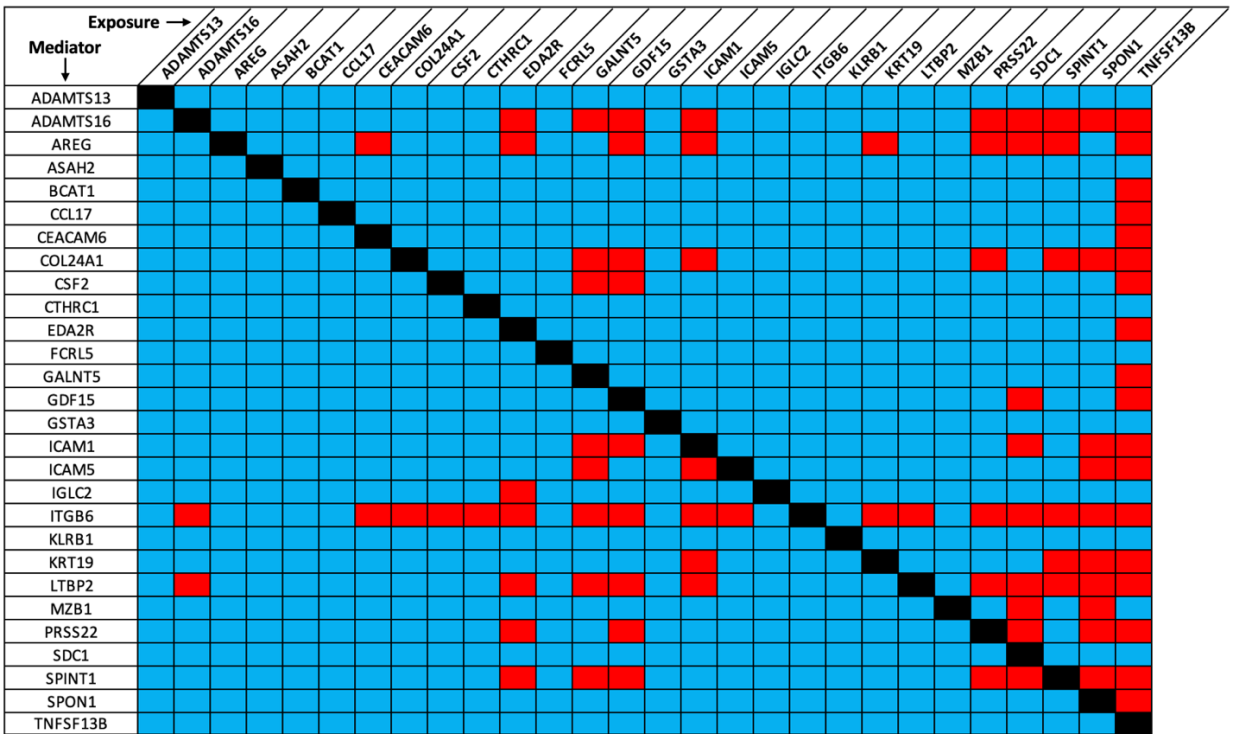


Figure 3.6. Results of inference-based protein-protein interaction network analysis in IPF cohort. Red fields indicate that the effect on FVC decline of the exposure protein (x-axis) is completely mediated by the mediator protein (y-axis). Blue fields indicate that the effect on FVC decline of the exposure protein (x-axis) is not mediated by the mediator protein (y-axis). Partial mediation is not shown.

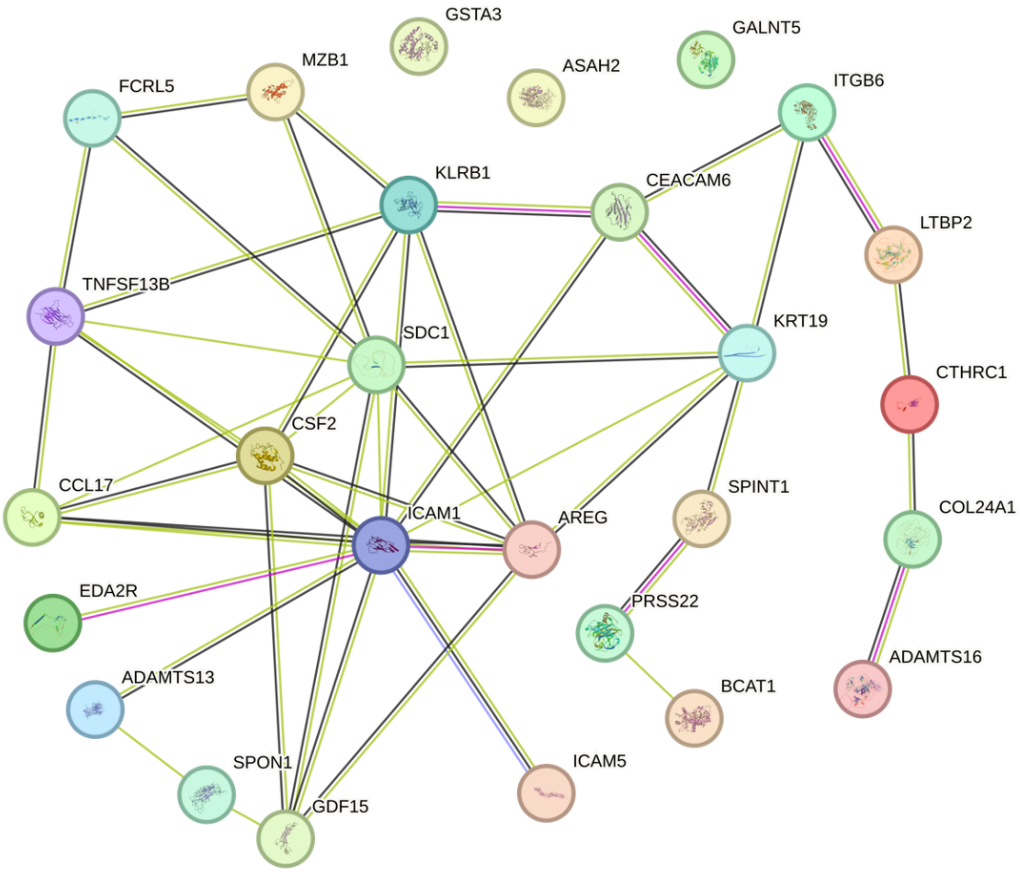


Figure 3.7. Biology-based protein network analyses identified through String software at confidence level of 0.25.

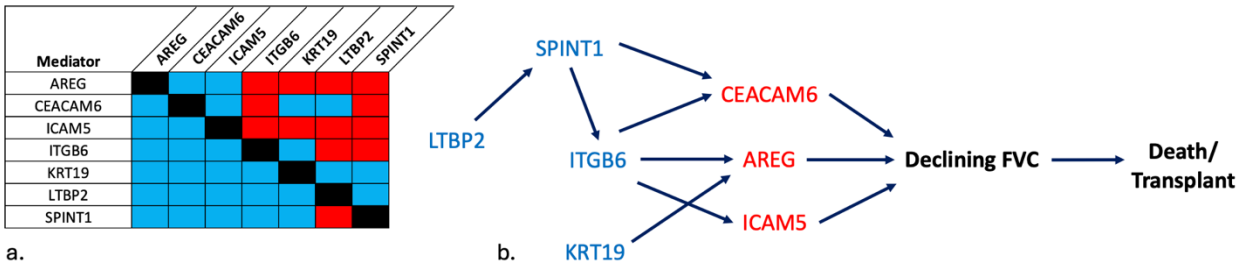


Figure 3.8. Results of inference-based protein-protein interaction network analysis in non-IPF ILD cohort (a) and resulting direct causal network (b). Red fields indicate that the effect on FVC decline of the exposure protein (x-axis) is completely mediated by the mediator protein (y-axis). Blue fields indicate that the effect on FVC decline of the exposure protein (x-axis) is not mediated by the mediator protein (y-axis). Partial mediation is not shown. Red proteins exert a direct effect on declining FVC, while blue proteins are upstream of red proteins.

Chapter 4 Independence and Additive Performance of Interstitial Lung Disease Biomarkers

4.1 ***Abstract***

A host of prognostic biomarkers have been identified in patients with fibrotic interstitial lung disease (ILD). Whether these biomarkers are independent of one another and provide additive outcome discrimination remains unclear. Here we aimed to determine whether genomic, proteomic and imaging biomarkers are independently associated with clinically relevant ILD outcomes and provide additive outcome discrimination as multimodal biomarkers. Participants in the Pulmonary Fibrosis Foundation Patient Registry with whole genome sequencing (WGS) estimated telomere length, plasma protein measures and quantitative fibrosis extent on chest imaging were included. The association between biomarkers and three-year transplant-free survival (TFS) was estimated to determine whether biomarkers were independently associated with TFS after adjustment for potential confounders. A secondary analysis was then performed in a subset of individuals with serial lung function data to evaluate the association between these biomarkers and ≥10% categorical decline in forced vital capacity (FVC). Model building was performed to assess whether multimodal composite biomarkers better discriminated outcomes when compared to unimodal biomarkers. Of 1460 eligible individuals, 524 had biomarker data. All biomarkers were associated with TFS when adjusting for clinical covariates, but only integrin beta six (ITGB6), collagen 24a1 (COL24A1) and quantitative fibrosis extent maintained TFS association in biomarker-adjusted modeling. ITGB6 was the only biomarker associated with ≥10% FVC decline. A trimodal biomarker model comprised of ITGB6, COL24A1, quantitative fibrosis extent and diffusion capacity outperformed the best unimodal model with C-statistic of 0.79. A less complex, bimodal model displayed similar discriminatory performance (C=0.78) as the best performing trimodal model. These findings highlight substantial collinearity between previously reported prognostic biomarkers but also the potential for multimodal biomarkers to improve outcome discrimination

4.2 ***Introduction***

Progressive pulmonary fibrosis (PPF) leads to declining lung function and premature death. A host of prognostic biomarkers have been linked to increased mortality risk in patients with ILD, including phenotypic traits,(Ley et al., 2012; Ryerson et al., 2014; Wells et al., 2003) peripheral blood telomere length,(Ley et al., 2020; Ley et al., 2019; Newton et al., 2016; Newton et al., 2019; Stuart et al., 2014) circulating plasma proteins(Bowman et al., 2021; Bowman et al., 2022) and quantitative extent of fibrosis on high resolution computed tomography (HRCT).(Camiciottoli et al., 2007; Distler et al., 2020; Humphries et al.; Humphries et al., 2018; Jacob et al., 2017; Jacob et al., 2019; Salisbury et al., 2018) While each biomarker has shown a reliable association with ILD survival, and has the potential to improve risk stratification, it remains unclear whether these biomarkers are independent of one another and provide discriminant value when modeled together.

While idiopathic pulmonary fibrosis (IPF) is the prototypical PPF, those with non-IPF ILD also commonly develop PPF, which leads to a natural history that closely resembles IPF.(Pugashetti et al., 2023) The optimal definition of PPF remains controversial; however, a categorical decline in forced vital capacity (FVC) of ≥10% relative to baseline is among strongest clinical predictors of reduced TFS and is a commonly used inclusion criterion in clinical trials targeting progressive non-IPF ILD.(Flaherty et al., 2019; Maher, Assassi, et al., 2023) Interestingly, while many prognostic biomarkers have been linked to differential survival, few have shown a reliable association with FVC decline or accurately discriminate this near-term, clinically relevant outcome.

In this study, we leveraged phenotypic, genomic, proteomic and quantitative imaging data collected from participants in the Pulmonary Fibrosis Foundation Patient Registry to determine whether previously reported prognostic biomarkers represent independent predictors of ILD outcomes or are collinear. We evaluated telomere length estimated through whole genome sequencing (WGS),(Pugashetti et al., 2024) five recently described prognostic protein biomarkers(Oldham et al., 2023) and quantitative fibrosis extent on high resolution computed tomography estimated using data-driven textural analysis (DTA).(Oh et al., 2024) We then evaluated whether these biomarkers provided additive discriminant value beyond clinical covariates. Analyses were repeated in a subset of individuals with serially acquired lung function

to determine whether biomarkers were similarly associated with a relative decline in FVC of ≥10% and effectively discriminated this outcome. We hypothesized that each biomarker evaluated would independently predict key ILD outcomes and would provide additive discriminant value beyond unimodal biomarkers.

4.3 ***Material and Methods***

4.3.1 ***Cohorts and Phenotypic Measures***

Participants in the Pulmonary Fibrosis Foundation Patient Registry(Wang et al., 2020) with fibrotic ILD due to IPF, connective tissue disease-associated ILD, fibrotic hypersensitivity pneumonitis and non-IPF idiopathic interstitial pneumonia who underwent high-throughput genomic and proteomic profiling as part of the PRECISIONS multi-omic study(Podolanczuk et al., 2022) and DTA data generation as part of an unrelated study(Oh et al., 2024) were eligible for inclusion. Those without complete data for all biomarkers were excluded. Phenotypic data collected included age, sex, smoking history, percent predicted FVC, and percent predicted diffusion capacity (DLCO), immunosuppressant (mycophenolate mofetil, azathioprine, cyclophosphamide, rituximab or prednisone of any dose) exposure and anti-fibrotic (nintedanib or pirfenidone) exposure, which was complete for all patients included. Serial lung function measures were also collected and used for the progression analysis described below in a subset of individuals for whom these data were available.

4.3.2 ***Biomarker Measures***

Detailed methods for genomic, proteomic and imaging data generation have been previously described previously.(Oh et al., 2024; Oldham et al., 2023; Pugashetti et al., 2024) Briefly, WGS-estimated telomere length was determined using TelSeq software (version 0.0.2) and is reported in kilo-base pairs. Proteomic data were generated in using the Explore 3072 panel (Olink®, Uppsala, Sweden), which uses proximity extension assays to estimate the relative abundance of circulating plasma proteins.(Lundberg et al., 2011) To select proteins for this analysis, we averaged the rank of the top 10 proteins identified in our recent studies of proteomic biomarkers of TFS in IPF(Oldham et al., 2023) and non-IPF ILD (chapter 2 above) and selected the top 5 for this study (Table 4-1), as these were most likely to be prognostically relevant across key ILD subtypes. Quantitative fibrosis extent was generated using DTA software(Humphries et

al., 2018) and is reported in percent fibrosis. Telomere length and protein data were measured using blood drawn the same day, which occurred upon registry entry. DTA fibrosis extent was determined using clinically acquired chest imaging, which was performed up to 3 years prior to blood draw and up to 90 days after blood draw. Those with chest imaging performed outside that range were excluded. Sensitivity analyses were performed to evaluate the impact of the time from chest imaging to blood draw on TFS association.

4.3.3 *Outcome Measures*

Transplant-free survival was the primary outcome measure of interest and was defined as the time from blood draw to death, lung transplant or censoring at 36-months or date lost to follow-up if prior to 36 months. To mitigate immortal time bias that could arise due to chest imaging performed after blood draw, we excluded patients with survival time of <90 days. Categorical decline in FVC of ≥10% was performed as a secondary analysis. This analysis was restricted to a subset of individuals for whom this measure of one-year FVC decline status could be ascertained at one-year following blood draw. FVC decline was considered present in those with ≥10% relative FVC decline within 15 months of blood draw and absent in those with <10% FVC decline 9-15 months following blood draw. Those for whom FVC decline status could not be ascertained were excluded from this secondary analysis.

4.3.4 *Statistical analysis*

Continuous measures are reported according to the mean and standard deviation, while categorical measures are reported according to count and percentage of the overall cohort. Pairwise Pearson's correlation was assessed between all continuous phenotypic and molecular variables. The phenotype adjusted association between each biomarker and TFS was estimated using a multi-variable Cox proportional hazards regression model adjusted for age, sex, race, smoking status (ever vs never), baseline percent predicted FVC and diffusion capacity of the lung for carbon monoxide (DLCO), and dichotomous exposure to immunosuppressant and anti-fibrotic therapy at the time of blood draw. The TFS association was then estimated for all biomarkers simultaneously, along with phenotypic covariates above. To simplify effect estimate comparisons, we multiplied the WGS-estimated TL values by -1 given its inverse relationship with survival. To identify biomarkers with independent TFS association, those with the highest

p-value were iteratively removed until all remaining biomarkers had TFS association at Wald $p<0.05$. The association between biomarkers and categorical FVC decline was estimated using a multivariable logistic regression model that included the same clinical covariates as described above. The same process was used to identify biomarkeres independently associated with FVC decline.

Model building was then performed to identify biomarkers and combinations of biomarkers with the best discriminant performance for TFS and FVC decline. This was done by iteratively fitting models with combinations of clinical, molecular and imaging variables and assessing discriminant performance using Harrel's C-statistic for TFS and area under the curve for categorical FVC decline. A variable importance plot was used to display the relative contribution of each of the retained covariates to survival risk discrimination by calculating the decrease in Harrel's C-statistic after iteratively dropping individual variables selected for the final model.

To assess potential collinearity, multivariable linear regression models were constructed using Martingale residuals from the Cox models for the independent variable and combinations of the clinical, molecular, and imaging variables as the dependent variables to generate the condition index (CI). Briefly, the CI is derived from singular value decomposition of a covariate matrix, whereby the square root of the ratio of the largest eigenvalue to each individual eigenvalue provides a measure of potential collinearity. Moderate or severe collinearity are suggested by a CI of 10-30 or >30, respectively. Collinear variables were defined as those with the highest proportion of variance accounting for CI >10. All statistical analyses were performed using STATA (version 18, College Station, TX). Statistical significance was considered present at $p<0.05$.

4.4 ***Results***

4.4.1 ***Cohort Characteristics***

Of 1232 eligible individuals with molecular data, 524 also had complete DTA fibrosis extent data and were included in the primary analysis (Figure 4-1). The mean age of the cohort was 67.9 years (± 10.4 years) and a majority were male (67.2%) and white (89.9%) and ever smokers (61.6%) (Table 4-2). IPF was the predominant diagnosis (71.4%), followed by CTD-ILD

(16.2%) and other fibrotic ILDs (12.4%). Mean percent predicted FVC and DLCO was 69.4% and 42.7%, respectively. A large minority (44.9%) was treated with anti-fibrotic therapy and a smaller proportion (27.7%) was treated with immunosuppressant therapy. Mean biomarker values are shown in Table 1. Nearly all clinical and biomarker variables were weakly correlated (Figure 4-2); however moderate correlation was observed between AREG and KRT19 (Pearson's R=0.67, p=3.1e-68) and between ITGB6 and LTBP2 (R=0.61, p=5.8e-55). Cumulative TFS over the 36-month observation period was 66.7%, with complete 36-month TFS data available for 91% (477/524) of the cohort.

4.4.2 *Survival Analysis*

All biomarkers evaluated were associated with three-year TFS after adjustment for phenotypic traits, with WGS-estimated telomere length showing a weaker effect estimate than the other biomarkers (Table 4-3). When adjusting for phenotypic traits and other biomarkers, only COL24A1, ITGB6, and DTA fibrosis extent remained independently associated with TFS (Table 4-3). In sensitivity analysis, time from chest imaging to blood draw did not meaningfully impact results (Table 4-4), with TFS association for COL24A1 weakening, ITGB6 strengthening and DTA fibrosis extent remaining unchanged as the timeframe narrowed. When adding clinical diagnosis to the final independent model, neither IPF (HR 0.99; 95% CI 0.55-1.76; p=0.97), nor connective tissue disease-associated ILD (HR 0.64; 95% CI 0.30-1.39; p=0.26) diagnostic label showed an association with TFS.

Model building was then performed to identify variable sets that best discriminated TFS (Table 4-5). A unimodal clinical model that included age, sex, race, smoking history and baseline lung function discriminated TFS with a C-statistic of 0.73; however, the max CI suggests moderate collinearity primarily among primarily FVC % and age (Table 4-6). A trimodal biomarker model that included all proteins (AREG, COL24A1, ITGB6, KRT19, LTBP2), plus WGS-estimated TL and DTA fibrosis extent discriminated TFS with C-statistic of 0.78, with moderate collinearity (max CI 16.8) explained by KRT19, LTBP2, and WGS-estimated TL. A fully saturated trimodal model that included all clinical variables and biomarkers provided only modest improvement (C=0.79) over the trimodal biomarker model but exhibited severe collinearity (maximum CI 38.6) explained by age, WGS-estimated TL, and KRT19 (Tables 4-5, 4-6).

Through iterative modeling, we identified a three-variable trimodal model (COL24A1, ITGB6, DTA fibrosis extent and DLCO percent predicted) that performed as well as the fully saturated trimodal model, with a C-statistic of 0.79. In this model, DTA fibrosis extent explained the largest amount TFS risk prediction, followed by ITGB6, DLCO% and COL24A1 (Figure 4-3). This trimodal model (C=0.79) outperformed the best performing unimodal models, including the clinical model (C=0.73), DTA model (C=0.74) and protein model (C=0.74). Given the complexity of a trimodal model, we also explored less complex bimodal models. We found that a protein + imaging model (AREG + COL24A1 + ITGB6 + DTA fibrosis extent) performed similarly to the final trimodal model, yielding a C-statistic of 0.78. Bimodal models for clinical + protein data and clinical + imaging data yielded C-statistics of 0.77 and 0.76, respectively (Table 4-5).

4.4.3 *Progression Analysis*

Of the 524 individuals included in the primary analysis, one-year FVC decline status could be ascertained for 340. Of these 340, a ≥10% relative decline in FVC occurred in 97 (28.5%). After adjustment for phenotypic traits, AREG, ITGB6 and LTBP2 were associated with one-year FVC decline (Table 4-7). No association with this endpoint was observed for telomere length for DTA fibrosis extent. When adjusting for phenotypic traits and other biomarkers, only ITGB6, remained independently associated with categorical FVC decline. Model building showed that a clinical-protein model comprising three variables (age + DLCO % + ITGB6) discriminated one-year FVC decline as well as a fully saturated model with all clinical variables and biomarkers; however, discrimination was modest for both models with an area under the curve of 0.68 (Table 4-8).

4.5 ***Discussion***

In this study, we showed that seven previously known prognostic biomarkers were independently associated with TFS when adjusted for key phenotypic traits, but only a subset maintained independent TFS association without collinearity in a biomarker adjusted model. Independent biomarkers included COL24A1, ITGB6 and DTA fibrosis extent on chest imaging, while WGS-estimated telomere length, AREG, KRT19 and LTBP2 were either collinear with the other markers or no longer associated with TFS after adjustment for other biomarkers. In developing a

prediction model for TFS, we found that a multimodal approach that incorporated clinical, molecular and imaging data best discriminated TFS (C-statistic 0.79), while less complex bimodal models that incorporated two of these three data sources performed nearly as well (C-statistics 0.76-0.78) and did not suffer from collinearity. Despite promising results when modeling TFS, we also found that no biomarkers beyond ITGB6 were associated with one-year FVC decline. Consequently, no biomarker combinations accurately discriminated this outcome. This study highlights the potential value of multi-dimensional outcome prediction, but also the analytic considerations that arise with increasingly complex models that require multiple data sources.

While multi-modal biomarker studies shed important light on the relationship between different biomarker classes and their prognostic performance, they also pose unique challenges. We identified substantial collinearity between several important clinical predictors, including age and lung function, and molecular markers, along with collinearity between the molecular markers themselves. Collinearity between variables leads to dispersion of causal effects across other collinear variables and inflates standard errors. Consequently, the high degree of collinearity between several of the biomarkers tested likely obscured independent contributions to outcome risk.(Vatcheva et al., 2016)

Mediation can also impact interpretations of multi-modal biomarker modeling, as adjustment for a mediator will attenuate the outcome association for variables upstream of the mediator on the causal pathway.(Etminan et al., 2021) As such, our study indicates that some biomarkers may lie upstream of others on the same causal pathway. For example, if short telomere length leads to protein changes that drive outcomes, adjustment for this protein dysregulation would attenuate the outcome association for telomere length. We demonstrated this phenomenon through protein-protein mediation analysis in chapter 2 of this dissertation, whereby many prognostic protein biomarkers, including KRT19 and LTBP2, were found to be upstream of ITGB6 on a causal pathway that included declining lung function and death. That does not mean that these proteins are not relevant to disease, as mechanistic studies have clearly demonstrated the opposite.(Liu et al., 2016; Saha et al., 2019; Zhao et al., 2024; Zou et al., 2021) Instead, attenuation of the

outcome association by these proteins helps characterize their relationship with other disease relevant proteins that function along the same causal pathway.

While several of the biomarkers tested were independently associated with TFS in this analysis, only ITGB6 was associated with categorical decline in FVC. These findings stand in contrast to prior studies showing that telomere length(Newton et al., 2019) and KRT19(Bowman et al., 2022) are associated with this endpoint and could reflect our use of a single measure of progression rather than composite approaches that also incorporated death in these prior studies. Beyond inference, no biomarker model accurately discriminated categorical decline in FVC, with the best performing model having area under the curve of 0.68. While predicting death and need for transplant is of inarguable value, accurately predicting upstream progression has higher likelihood of impacting patient care, as FVC trends are often used to guide monitoring strategies and management decisions.

4.6 ***Conclusion***

Identifying independent prognostic biomarkers and gauging their ability to independently improve prediction over standard clinical variables is an important requirement to advance prognostic biomarkers into the clinical domain. Here we found that the addition of protein and imaging biomarkers improved upon easily acquired clinical data, highlighting the need to consider additional biomarker domains to optimize prognostication. Although multiple biomarkers have shown reliable association with fILD survival and most have the potential to improve risk stratification, we found that few of these prognostic biomarkers are independent. While circulating proteins and quantitative fibrosis extent are independent biomarkers that conveyed good survival discrimination, they were less reliable for discriminating FVC decline, highlighting a knowledge gap that requires urgent attention.

Table 4.1. Biomarker rank by study				
Protein	**Non-IPF ILD Rank**	**IPF Rank**	**Average Rank**	**Overall Rank**
LTBP2	4	1	2.5	1
ITGB6	2	4	3	2
KRT19	3	3	3	2
COL24A1	6	2	4	4
AREG	1	9	5	5

Table 4.2. Baseline clinical characteristics and biomarker measures	
Age, mean (±SD)	67.9 (10.4)
Male sex, n (%)	352 (67.2)
Race/ethnicity, n (%)	
White	471 (89.9)
Black	21 (4.0)
Asian	17 (3.2)
Other/Unknown	15 (2.9)
ILD Diagnosis, n (%)	
IPF	374 (71.4)
Connective tissue disease-associated ILD	85 (16.2)
Other ILD	65 (12.4)
Ever smoker, n (%)	323 (61.6)
FVC %, mean (±SD)	69.4 (17.8)
DLCO %, mean (±SD)	42.7 (16.1)
Anti-fibrotic exposure, n (%)	235 (44.9)
Immunosuppressant exposure , n (%)	145 (27.7)
AREG, mean (±SD)	1.46 (0.73)
COL24A1, mean (±SD)	1.57 (0.72)
ITGB6, mean (±SD)	1.09 (0.60)
KRT19, mean (±SD)	2.63 (0.97)
LTBP2, mean (±SD)	1.69 (0.62)
Telomere length (kilo base pairs), mean (±SD)	2.70 (0.77)
DTA fibrosis extent, mean percent (±SD)	28.46 (16.15)
Abbreviations: ILD = interstitial lung disease; IPF = idiopathic pulmonary fibrosis; FVC = forced vital capacity; DLCO = diffusion capacity of the lung for carbon monoxide	
Protein measures are semi-quantitative and unitless	

Table 4.3. Iterative adjusted modeling results to identify independent biomarkers of TFS						
Protein	**Fully Saturated Model**			**Final Independent Model**		
	HR	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
AREG	1.09	0.81-1.48	0.562	Removed		
COL24A1	1.49	1.13-1.97	0.005	1.5	1.14-1.98	0.004
ITGB6	1.4	0.92-2.10	0.114	1.68	1.20-2.34	0.002
KRT19	1.07	0.84-1.36	0.59	Removed		
LTBP2	1.13	0.79-1.61	0.513	Removed		
Telomere length	1.16	0.91-1.48	0.24	Removed		
DTA fibrosis extent	1.02	1.01-1.04	<0.001	1.03	1.01-1.04	<0.001
Abbreviations: HR = hazard ratio; CI = confidence interval						
All models adjusted for age, sex, race, smoking history, percent predicted FVC, percent predicted DLCO, anti-fibrotic exposure and immunosuppressant exposure						
HR reflects change in outcome risk per-unit decrease in telomere length and per-unit increase for all other biomarkers						

Table 4.4. Association between prognostic biomarkers and three-year TFS according to time from high resolution computed tomography (HRCT) to blood draw

Protein	HRCT to Draw <36 months (n=524)			HRCT to Draw <24 months (n=491)			HRCT to Draw <12 months (n=400)		
	HR	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
COL24A1	1.50	1.14-1.98	0.004	1.39	1.04-1.85	0.026	1.32	0.95-1.84	0.095
ITGB6	1.68	1.20-2.34	0.002	1.85	1.29-2.64	0.001	1.84	1.23-2.74	0.003
DTA fibrosis extent	1.03	1.01-1.04	<0.001	1.03	1.01-1.04	<0.001	1.03	1.02-1.04	<0.001

Abbreviations: HRCT = high resolution computed tomography; HR = hazard ratio; CI = confidence interval

HR reflects change in outcome risk per-unit decrease in telomere length and per-unit increase for all other biomarkers

* Adjusted for age, sex, race, smoking history, percent predicted FVC, percent predicted DLCO, anti-fibrotic exposure and immunosuppressant exposure

Protein measures are semi-quantitative and unitless

Table 4.5. TFS model building results		
Model type and variables	**Variables**	**C-Statistic**
Unimodal		
Clinical model (age + sex + race + smoking + FVC % + DLCO %)	6	0.73
Genomic model (telomere length)	1	0.54
Imaging model (DTA)	1	0.74
Protein model (AREG + COL24A1 + ITGB6 + KRT19 + LTBP2)	5	0.74
Bimodal		
Best bimodal clinical-imaging model (sex + smoking + DLCO % + DTA)	4	0.76
Best bimodal clinical-protein model (DLCO % + AREG + COL24A1 + ITGB6)	4	0.77
Best bimodal protein-imaging model (AREG + COL24A1 + ITGB6 + DTA)	4	0.78
Trimodal		
Fully saturated model (all clinical, imaging and molecular variables)	13	0.79
Biomarker model (AREG + COL24A1 + ITGB6 + KRT19 + LTBP2 + telomere length + DTA)	7	0.78
Best trimodal model (COL24A1 + ITGB6 + K DTA fibrosis extent + DLCO %)	8	0.79

Table 4.6. Condition Index based collinearity diagnostics and proportion of variance.

		Proportion of Variance											
Collinearity	**Condition Index**	**Intercept**	**Age**	**Male**	**FVC**	**DLCO**	**AREG**	**COL24A1**	**ITGB6**	**KRT19**	**LTBP2**	**WGS-TL**	**DTA**
Clinical Model (age, sex, smoking history, FVC %, DLCO %)													
Minimal	1.0	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	4.0	--	--	0.878	--	--	--	--	--	--	--	--	--
Minimal	7.3	--	--	--	--	0.638	--	--	--	--	--	--	--
Moderate	11.6	--	--	--	0.99	--	--	--	--	--	--	--	--
Moderate	20.7	0.923	0.886	--	--	--	--	--	--	--	--	--	--
Multi-marker (AREG, COL24A1, ITGB6, KRT19, LTBP2, WGS-TL, DTA)													
Minimal	1.0	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	5.6	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	6.5	--	--	--	--	--	--	--	--	--	--	--	0.83
Minimal	7.4	--	--	--	--	--	--	0.381	--	--	--	--	--
Minimal	9.5	--	--	--	--	--	--	0.545	--	--	--	--	--
Moderate	11.2	--	--	--	--	--	--	--	0.56	--	0.676	--	--
Moderate	12.2	--	--	--	--	--	0.465	--	--	0.846	--	--	--
Moderate	16.8	0.968	--	--	--	--	--	--	--	--	--	0.648	--
Fully Saturated Model (age, sex, smoking history, FVC %, DLCO %, AREG, COL24A1, ITGB6, KRT10, LTBP2, WGS-TL, DTA)													
Minimal	1.0	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	4.7	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	5.9	--	--	0.882	--	--	--	--	--	--	--	--	--
Minimal	7.7	--	--	--	--	--	--	--	--	--	--	--	0.384
Minimal	8.9	--	--	--	--	--	--	0.383	--	--	--	--	--
Moderate	11.4	--	--	--	--	--	--	0.487	--	--	--	--	--
Moderate	11.9	--	--	--	--	--	--	--	--	--	--	0.568	--
Moderate	13.2	--	--	--	--	--	--	--	0.488	--	0.386	--	--
Moderate	14.8	--	--	--	--	--	0.462	--	--	0.836	--	--	--
Moderate	16.3	--	--	--	0.469	0.437	--	--	--	--	--	--	--
Moderate	22.6	--	0.394	--	0.463	--	--	--	--	--	--	--	--
Severe	38.6	0.93	0.567	--	--	--	--	--	--	--	--	0.31	--
Final Multimodal Model (COL24A1, ITGB6, DLCO %, DTA)													
Minimal	1.0	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	3.7	--	--	--	--	--	--	--	--	--	--	--	--
Minimal	5.3	--	--	--	--	--	--	--	0.363	--	--	--	0.465
Minimal	7.1	--	--	--	--	--	--	0.892	0.54	--	--	--	--
Moderate	13.8	0.984	--	--	--	0.848	--	--	--	--	--	--	0.392

Table 4.7. Iterative adjusted modeling results to identify independent biomarkers of one-year FVC decline						
Protein	**Fully Saturated Model**			**Final Independent Model**		
	OR	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
AREG	1.31	0.80-2.12	0.28	Removed		
COL24A1	0.85	0.55-1.31	0.458	Removed		
ITGB6	2.26	1.15-4.43	0.018	2.38	1.46-3.89	0.001
KRT19	0.87	0.59-1.29	0.495	Removed		
LTBP2	1.38	0.80-2.41	0.249	Removed		
Telomere length	0.84	0.59-1.20	0.342	Removed		
DTA fibrosis extent	1	0.98-1.02	0.925	Removed		
Abbreviations: OR = odds ratio; CI = confidence interval						
All models adjusted for age, sex, race, smoking history, percent predicted FVC, percent predicted DLCO, anti-fibrotic exposure and immunosuppressant exposure						
OR reflects change in outcome risk per-unit decrease in telomere length and per-unit increase for all other biomarkers						

Table 4.8. FVC decline model building results		
Model type and variables	**Variables, n**	**AUC**
Unimodal		
Clinical model (age + sex + race + smoking + FVC % + DLCO %)	6	0.64
Genomic model (telomere length)	1	0.50
Imaging model (DTA fibrosis extent)	1	0.56
Protein model (AREG + COL24A1 + ITGB6 + KRT19 + LTBP2)	5	0.64
Bimodal		
Best bimodal clinical-imaging model (age + DLCO % + DTA)	3	0.64
Best clinical + protein model (age + DLCO % + ITGB6)	3	0.68
Best bimodal protein-imaging model (AREG + COL24A1 + ITGB6 + SPINT1+ DTA)	5	0.65
Trimodal		
Fully saturated model (all variables above)	13	0.68
Biomarker model (AREG + COL24A1 + ITGB6 + KRT19 + LTBP2 + telomere length + DTA)	7	0.65
Best trimodal model (age + DLCO% +ITGB6 + telomere length)	4	0.68

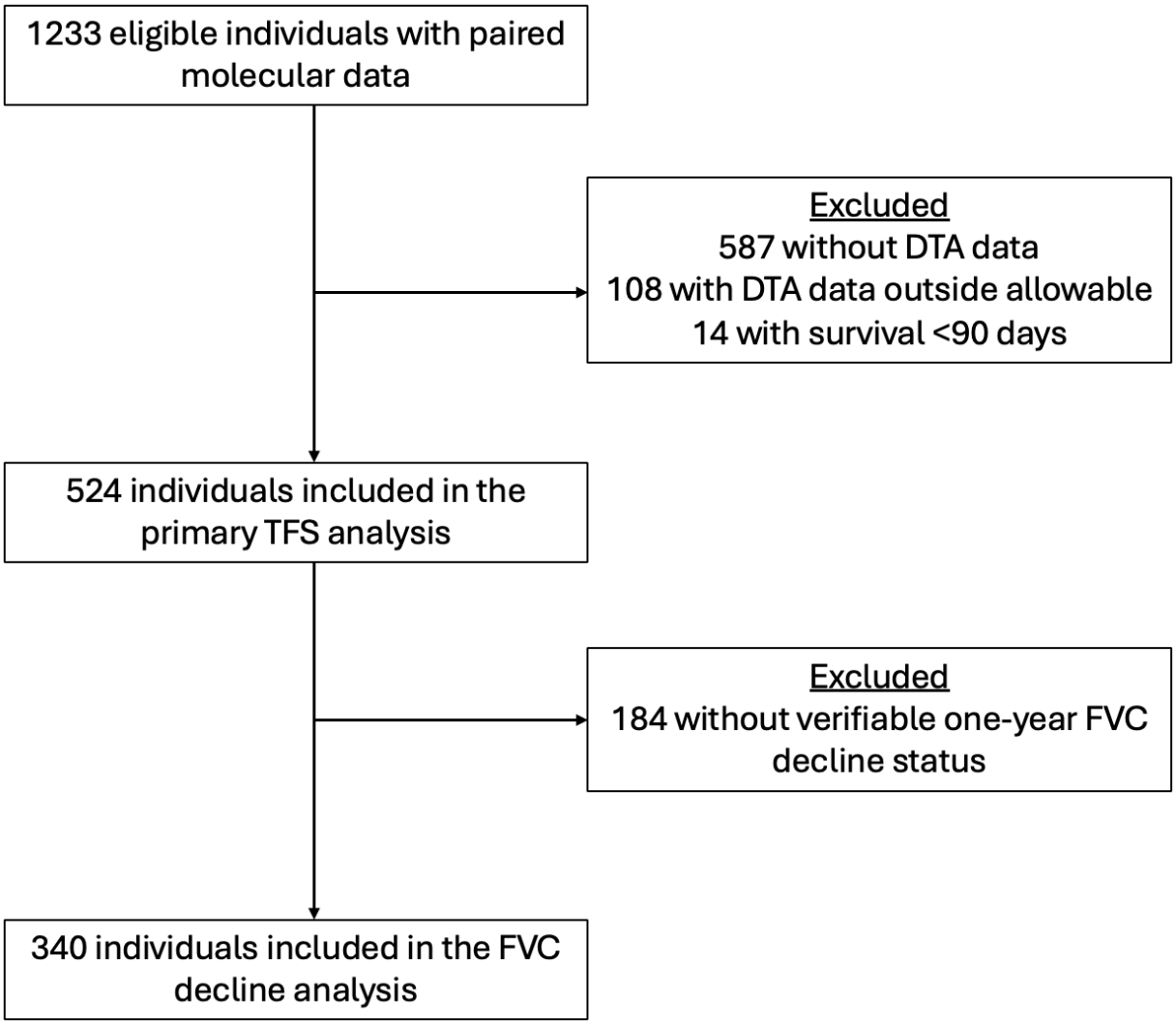


Figure 4.1. Case finding diagram for chapter 4 cohort.

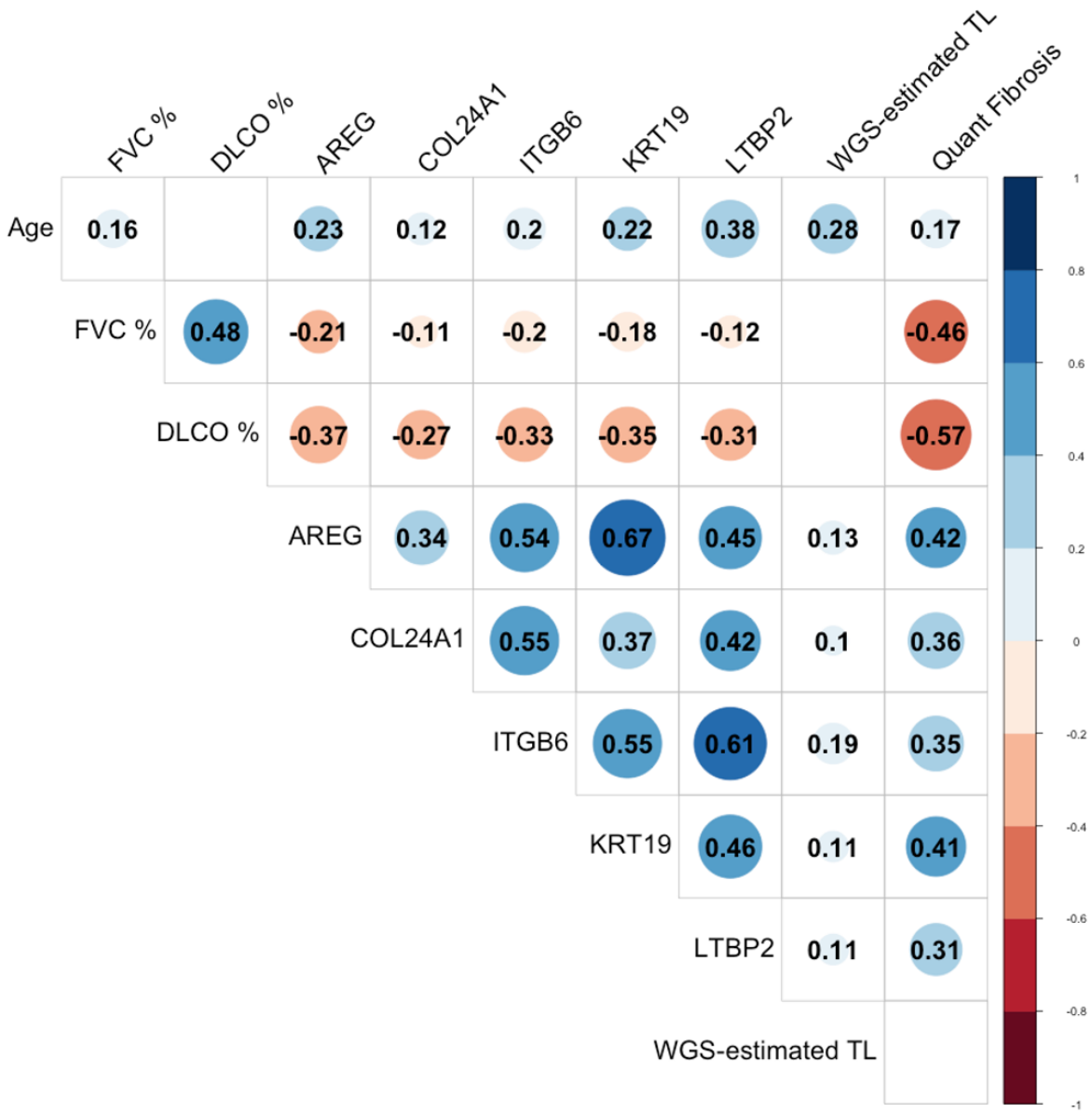


Figure 4.2. Correlation matrix for continuous clinical variables and biomarkers. Size of circles and numerical values represent correlation coefficient. Blank cells denote non-significant correlation ($p<0.05$).

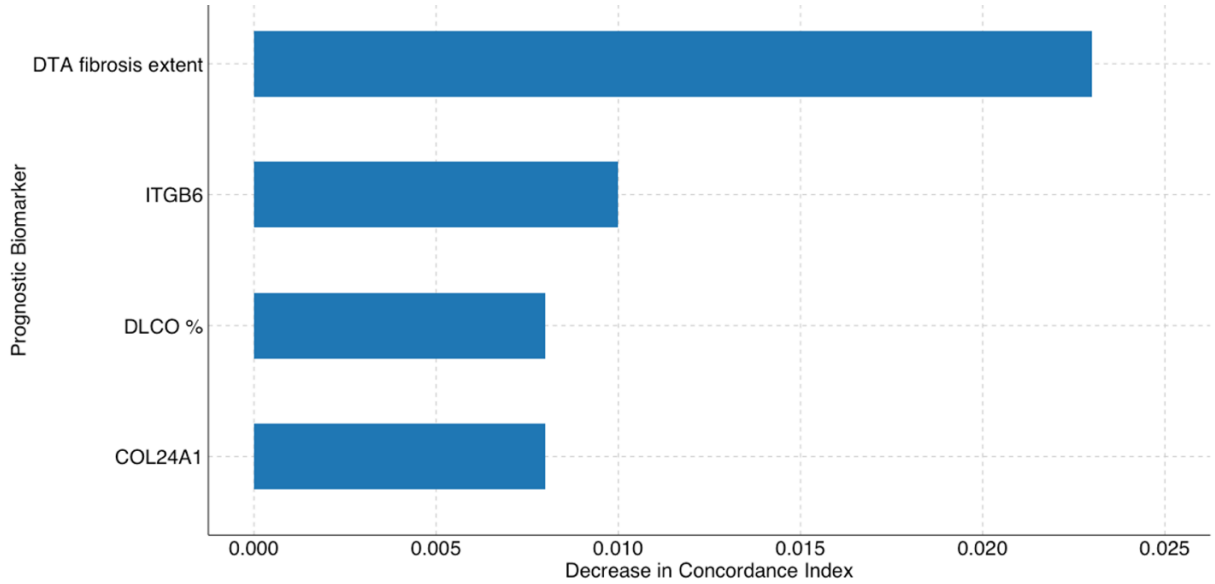


Figure 4.3. Variable importance plot showing the relative contribution to TFS discrimination for variables comprising the final model. Larger numerical decrease in concordance index denotes more importance for survival risk discrimination.

Chapter 5 Conclusions

5.1 ***Summary and Implications of Main Findings***

The primary goals of this dissertation were to 1) identify novel proteomic biomarkers of non-IPF ILD survival, 2) identify proteins that may play a potentially causal role in PPF and 3) determine whether known prognostic proteomic, genomic and radiological biomarkers were independently associated with clinically relevant outcomes and provided additive discriminant value. By accomplishing these goals, we advanced our understanding of PPF and helped address a critical knowledge gap for patients with non-IPF ILD. These findings also help prioritize proteins and molecular pathways for therapeutic targeting, which could produce new effective drugs for ILD.

5.1.1 ***Summary and Implications of Chapter 2 Findings***

In chapter 2, we identified 44 protein biomarkers associated with differential transplant-free survival across independent non-IPF ILD cohorts. Results were robust to non-IPF ILD clinical diagnosis, suggesting that shared pathobiology underpins ILD progression for diverse forms of ILD. Pathway analysis of these proteins then revealed novel molecular pathways unique to non-IPF ILD and established shared molecular pathways for IPF and non-IPF ILD. These findings build on a similarly conducted proteomic investigation in patients with IPF,(Bowman et al., 2021) along with a previously published, medium throughput proteomic investigation in patients with non-IPF ILD.(Bowman et al., 2022) This study represents the largest molecular investigation conducted in non-IPF ILD to date, including the first to use a high-throughput proteomic platform to investigate non-IPF ILD outcomes. By cross referencing the findings of this study with those from a recent study of IPF,(Bowman et al., 2021) we were able to identify novel molecular pathways unique to non-IPF ILD. Among these was the anti-microbial peptides pathway, which is critical for host defense and highlights the potential role that infection and immune dysregulation plays in non-IPF ILD progression. Together these findings help provide foundational knowledge for non-IPF ILD and set the stage for proteomic-based prognostication for this population

5.1.2 *Summary and Implications of Chapter 3 Findings*

In chapter 3, we identified 28 prognostic protein biomarkers with survival effect that was mediated by declining FVC, suggesting a potentially causal relationship with PPF. Seven of these 28 proteins identified in an IPF cohort showed consistent findings in a similarly defined non-IPF ILD cohort, supporting findings from chapter 2 that indicate shared PPF pathobiology across ILD subtypes. Pathway analysis confirmed that key fibrogenic pathways were overrepresented among proteins mediated by PPF and an inference-based interaction network analysis helped characterize directional relationships between proteins classified as potentially causal of PPF. The potential for biological interaction between these proteins was then corroborated through biology-based interaction network analysis. This study is the first to use causal mediation analysis to identify protein that play a potentially causal role in PPF. Amid the rapidly growing number of prognostic biomarkers being linked to survival in patients with ILD, this study help identify protein biomarkers most likely to contribute to disease outcomes, allowing prioritization of proteins and their associated molecular pathways for therapeutic consideration. These results are particularly timely given failure of two randomized controlled trials evaluating αvβ6 blockade. By identifying a number of proteins that potentially function upstream of αvβ6, our results suggest that indirect targeting of this critically important molecule remains possible. Beyond ITGB6, our results may help inform future responder analyses in clinical trials and prioritize new proteins that warrant consideration for therapeutic development.

5.1.3 *Summary and Implications of Chapter 4 Findings*

In chapter 4, we showed that seven previously known prognostic biomarkers were independently associated with TFS when adjusted for key phenotypic traits, but only three maintained independent TFS association in a biomarker adjusted model. Other biomarkers assessed, including WGS-estimated telomere length and three key protein biomarkers (AREG, KRT19 and LTBP2) were either collinear with the other biomarkers or no longer associated with TFS after adjustment for other biomarkers. We then explored multimodal prediction models for TFS, finding that a trimodal model that incorporated clinical, molecular and imaging data best discriminated TFS, while less complex bimodal models that incorporated only two of these data sources performed nearly as well. Despite promising results when modeling TFS, we found that no biomarkers beyond ITGB6 were associated with one-year FVC decline and no model

effectively discriminated this endpoint. This study is among the first to evaluate independence among previously reported prognostic protein biomarkers and demonstrates the potential value of prognostication through multimodal biomarkers. However, it also highlighted the analytical and logistical considerations that arise with increasingly complex models that require multiple data sources. While predicting death and need for transplant is of inarguable value, accurately predicting upstream progression has higher likelihood of impacting patient care, as FVC trends are often used to guide monitoring strategies and management decisions. These findings underscore the urgent need to develop biomarkers to accurately predict upstream progression.

5.2 ***Strengths and Limitations***

Major strengths of chapter 2 include: 1) the relatively large sample size for a rare disease, 2) use of independent discovery and validation cohort when identifying novel protein biomarkers of non-IPF ILD survival and 3) use of a confounder adjusted model in all analyses, which was not performed in many prior prognostic biomarker studies. Major limitations include 1) the inability to assess the effect of immunosuppressant therapy beyond binary exposure, 2) inability to assess anti-fibrotic therapy effect, as nearly all patients were recruited prior to the approval of nintedanib for progressive non-IPF ILD(Flaherty et al., 2019) and 3) lack of clarity whether blood-based findings reflect lung pathobiology. There was potential for selection bias, as the cohort was restricted to individuals seeking care at tertiary referral centers with ILD expertise. There was also likely some degree of measurement error given the lack of standardized diagnosis across centers and use of a proteomic array that could result in cross-reactivity with proteins of similar structure. However, neither of these sources of measurement error were likely to produce information bias.

Major strengths of chapter 3 include: 1) innovative methods to increase confidence in a causal relationship between protein biomarkers and clinically relevant ILD outcomes, 2) use of a well phenotyped cohort with outcome data for both death (or transplant) and antecedent progression and 3) the use of biology-based interaction network analysis to corroborate inference-based findings. Major limitations include likely selection bias and measurement error. The study was restricted to individuals seeking care at tertiary medical centers with ILD expertise, which could impact generalizability of these findings. Moreover, restricting the analysis to those with serial

FVC measures likely produced some degree of selective survival bias, as those who died prior to undergoing another FVC measure were excluded. Measurement error was possible for exposure, mediator and outcome. While proximity extension assays bind target proteins with high specificity, cross-reactivity with other proteins of similar structure remains possible. While declining FVC is a cardinal feature of PPF,(Cottin et al., 2018; Pugashetti et al., 2023; Raghu et al., 2022; Wijsenbeek et al., 2022) declining DLCO, symptomatic worsening and increasing extent of fibrosis on chest imaging can also indicate PPF but are more difficult to measure.(Raghu et al., 2022) While death and lung transplant are among the most clinically relevant outcomes, we previously showed that a large minority of patients with IPF die from conditions other than IPF.(Thenappan et al., 2024) Each of these sources of measurement error could have attenuated results, but were unlikely to be differential by one of the other key measures, making information bias unlikely.

Major strengths of chapter 4 include: 1) generation of multimodal biomarker data, which has not been done previously and 2) significance of this work, which is the first study to characterize the relationship between previously reported prognostic biomarkers and evaluate the additive value of a multimodal biomarker approach. Major limitations include 1) the relatively small sample size, which may have resulted in the misclassification of some biomarkers as non-independent, 2) collinearity between several of the proteins selected, which chapter 3 showed to be on the same causal pathway and 3) use WGS-estimated TL, which has higher variance and less consistent outcome association compared to directly measured peripheral blood leukocyte telomere length.(Ley et al., 2020; Ley et al., 2019; Newton et al., 2016; Newton et al., 2019; Stuart et al., 2014) Other limitations stems from the possibility of selection bias and measurement error for the same as reasons described for chapters 2 and 3 above. However, as in other chapters, none of the sources of exposure and outcome measurement error were likely to produce information bias, as they were unlikely to be differential by outcome or exposure, respectively.

5.3 ***Future Directions***

The findings of this dissertation lay the foundation for new lines of research in those with ILD. By identifying protein biomarkers associated with TFS in patients with non-IPF ILD, Chapter 2

highlights the potential for these biomarkers to discriminate this and other clinically relevant endpoints.(Alqalyoobi et al., 2025) Importantly, many of the proteins identified in Chapter 2 overlapped with those identified in a similar study of patients with IPF,(Oldham et al., 2023) suggesting that shared pathobiology underpins progressive ILD irrespective of ILD clinical diagnosis. Chapter 4 builds on Chapter 2, showing that multi-modal biomarker modeling in a pooled ILD cohort comprised of individuals with IPF and non-IPF ILD augments discriminant value beyond what can be achieved using a single data source (e.g., clinical data, protein data, imaging data) alone. A follow-up study has now showed that bimodal model comprised of two clinical features and twenty proteins accurately predicted 18-month survival in transplant-eligible patients with IPF.(Vu Pugashetti et al., 2024) Importantly, this classifier also performed well when applied in a non-IPF ILD cohort, supporting that an individual's molecular makeup may transcend ILD clinical diagnosis when prognosticating outcomes.

While survival represents the ultimate outcome in patients with ILD, the ability to predict upstream progression remains elusive, as underscored in Chapter 4, where no model displayed an area under the curve of >0.7. Our findings therefore set the stage for an expansive new project aimed at developing clinically actionable multi-modal biomarkers to discriminate near-term progression in patients with ILD. While the optimal measure of ILD progression remains debated,(Johannson et al., 2022; Vu Pugashetti et al., 2023) change in forced vital capacity is among the best validated measures of progression in patients with ILD and is the most commonly used metric to gauge therapeutic response in ILD clinical trials.(Collard et al., 2003; Gimenez et al., 2018; Goh et al., 2017; Maher, Assassi, et al., 2023; Pugashetti et al., 2022) In our proteomic datasets, we will now use machine learning to identify candidate proteins for predicting FVC decline in patients with IPF and non-IPF ILD. Using custom, multiplex proteomic arrays built as part of two new R01 awards, we will quantify candidate proteins and develop a machine learning-based proteomic signature to predict continuous and categorical FVC decline in this population. A binary classifier will be developed, with sensitivity prioritized, and then applied to several independent ILD cohorts from around the world. We will then determine the longitudinal validity of this biomarker signature and change over time in individual biomarkers comprising this signature using serially collected plasma from prospectively recruited patients.

While Chapter 3 sought to identify proteins most appropriate for therapeutic targeting, we will build upon these results by identifying biomarkers of treatment response in patients with IPF and non-IPF ILD. While IPF is treated with anti-fibrotic therapy,(Lederer & Martinez, 2018) non-IPF ILD is often first treated with immunosuppressant therapy, (Fischer et al., 2013; Fischer & du Bois, 2012) providing two different classes of therapy to investigate. Through independent NIH funding, we have been recruiting a 1000 patient ILD cohort over the last three years at the University of Michigan. At present, over 500 of these individuals have baseline and follow-up plasma samples performed 6-12 months later. Using this valuable resource, we will conduct high-throughput proteomic investigation to identify biomarkers 1) that change in response to anti-fibrotic and immunosuppressant therapy and 2) whose change over time predicts favorable response to these two classes of therapy. We expect this approach to compliment Chapter 3 by confirming that mediated proteins change over time and in response to therapy.

5.4 ***Conclusion***

While rare, progressive ILD represents a deadly condition that leads to worse survival than most cancers. A better understanding of underlying pathobiology is urgently needed to develop new therapies to slow this condition. This dissertation helps characterize this pathobiology and identify proteins for therapeutic consideration. It also highlights the potential for multimodal biomarkers to help better predict ILD progression. Finally, it sets the stage for groundbreaking new molecular Epidemiological work aimed at developing clinically actionable biomarkers for this population.

Appendix

Common Abbreviations	
CTD	Connective tissue disease
DLCO	diffusion capacity of the lung for carbon monoxide
DTA	data-driven textural analysis
FDR	false discovery rate
fHP	Fibrotic hypersensitivity pneumonitis
FVC	Forced vital capacity
HR	hazard ratio
HRCT	high-resolution computed tomography
IIP	Idiopathic interstitial pneumonia
ILD	Interstitial lung disease
INBUILD	INvestigating Nintedanib for the treatment of progressive fibrosing interstitial lung diseases
NSIP	non-specific interstitial pneumonia
PFFPR	Pulmonary Fibrosis Foundation Patient Registry
PRECISIONS	Prospective Treatment Efficacy in IPF Using Genotype for Nac Selection
RMST	restricted mean survival time
TFS	Transplant-free survival
UCD	University of california at Davis
UCSF	University of california at San Francisco
uILD	Unclassifiable ILD
UTSW	University of Texas Southwestern Medical Center
UVA	University of Virginia
WGS	whole genome sequencing
Protein Abbreviations	
ACE2	Angiotensin-converting enzyme 2
ACTA2	Actin, aortic smooth muscle
ADA2	Adenosine deaminase 2
ADAM12	Disintegrin and metalloproteinase domain-containing protein 12
ADAM22	Disintegrin and metalloproteinase domain-containing protein 22
ADAM9	Disintegrin and metalloproteinase domain-containing protein 9
ADAMs	A Disintegrin and Metalloproteinases
ADAMTS1	A disintegrin and metalloproteinase with thrombospondin motifs 1
ADAMTS13	A disintegrin and metalloproteinase with thrombospondin motifs 13
ADAMTS16	A disintegrin and metalloproteinase with thrombospondin motifs 16
ADAMTSL4	ADAMTS-like protein 4
ADGRD1	Adhesion G-protein coupled receptor D1
ADH4	All-trans-retinol dehydrogenase [NAD(+)] ADH4
ADIPOQ	Adiponectin
AFM	Afamin
AGR2	Anterior gradient protein 2 homolog
AGR3	Anterior gradient protein 3
AHSG	Alpha-2-HS-glycoprotein
AMPs	Antimicrobial peptides

ANGPTL2	Angiopoietin-related protein 2
APOA4	Apolipoprotein A-IV
AREG	Amphiregulin
ASAH2	Neutral ceramidase
ASS1	Argininosuccinate synthase
BAG3	BAG family molecular chaperone regulator 3
BCAT1	Branched-chain-amino-acid aminotransferase, cytosolic
BCHE	Cholinesterase
BHMT2	S-methylmethionine--homocysteine S-methyltransferase BHMT2
BMP4	Bone morphogenetic protein 4
BPIFB1	BPI fold-containing family B member 1
BRK1	Protein BRICK1
BTD	Biotinidase
C19ORF12	Protein C19orf12
C7	Complement component C7
C9	Complement component C9
CA14	Carbonic anhydrase 14
CA4	Carbonic anhydrase 4
CA6	Carbonic anhydrase 6
CA8	Carbonic anhydrase-related protein
CAPN3	Calpain-3
CAPS	Calcyphosin
CBX2	Chromobox protein homolog 2
CCL17	C-C motif chemokine 17
CCL18	C-C motif chemokine 18
CCL20	C-C motif chemokine 20
CCL23	C-C motif chemokine 23
CCN1	CCN family member 1
CD1C	T-cell surface glycoprotein CD1c
CD209	CD209 antigen
CD274	Programmed cell death 1 ligand 1
CD276	CD276 antigen
CD6	T-cell differentiation antigen CD6
CD80	T-lymphocyte activation antigen CD80
CDAN1	Codanin-1
CDCP1	CUB domain-containing protein 1
CDH3	Cadherin-3
CDH5	Cadherin-5
CDHR1	Cadherin-related family member 1
CEACAM16	Carcinoembryonic antigen-related cell adhesion molecule 16
CEACAM6	Carcinoembryonic antigen-related cell adhesion molecule 6
CES2	Cocaine esterase
CFP	Properdin
CGN	Cingulin
CHAD	Chondroadherin
CHCHD10	Coiled-coil-helix-coiled-coil-helix domain-containing protein 10, mitochondrial
CHCHD6	MICOS complex subunit MIC25
CHGB	Secretogranin-1
CHI3L1	Chitinase-3-like protein 1
CHL1	Neural cell adhesion molecule L1-like protein
CHRDL1	Chordin-like protein 1
CHRDL2	Chordin-like protein 2
CKB	Creatine kinase B-type

CLEC11A	C-type lectin domain family 11 member A
CLEC3B	Tetranectin
CLEC4G	C-type lectin domain family 4 member G
CLEC4M	C-type lectin domain family 4 member M
CLIC5	Chloride intracellular channel protein 5
CLUL1	Clusterin-like protein 1
CNDP1	Beta-Ala-His dipeptidase
CNTN3	Contactin-3
CNTN5	Contactin-5
COL18A1	Collagen alpha-1(XVIII) chain
COL24A1	Collagen alpha-1(XXIV) chain
COL4A1	Collagen alpha-1(IV) chain
COL6A3	Collagen alpha-3(VI) chain
COL9A1	Collagen alpha-1(IX) chain
CPM	Carboxypeptidase M
CPXM2	Inactive carboxypeptidase-like protein X2
CRIM1	Cysteine-rich motor neuron 1 protein
CRTAC1	Cartilage acidic protein 1
CSF1	Macrophage colony-stimulating factor 1
CSF2	Granulocyte-macrophage colony-stimulating factor
CSPG4	Chondroitin sulfate proteoglycan 4
CST1	Cystatin-SN
CST5	Cystatin-D
CTHRC1	Collagen triple helix repeat-containing protein 1
CTSL	Cathepsin L1
CXCL13	C-X-C motif chemokine 13
CXCL17	C-X-C motif chemokine 17
DLK1	Protein delta homolog 1
DSCAM	Down syndrome cell adhesion molecule
DSG3	Desmoglein-3
DSG4	Desmoglein-4
DUSP13	Dual specificity protein phosphatase 13 isoform A
EBI3_IL27	Interleukin-27
ECE1	Endothelin-converting enzyme 1
EDA2R	Tumor necrosis factor receptor superfamily member 27
EDN1	Endothelin-1
EGFLAM	Pikachurin
EGFR	epidermal growth factor receptor
ELN	Elastin
ENAH	Protein enabled homolog
ENO3	Beta-enolase
ENPP6	Glycerophosphocholine cholinephosphodiesterase ENPP6
EPHA1	Ephrin type-A receptor 1
EPO	Erythropoietin
ESM1	Endothelial cell-specific molecule 1
ESPL1	Separin
EVPL	Envoplakin
F13B	Coagulation factor XIII B chain
FABP3	Fatty acid-binding protein, heart
FAM20A	Pseudokinase FAM20A
FAP	Prolyl endopeptidase FAP
FASLG	Tumor necrosis factor ligand superfamily member 6
FCGR2A	Low affinity immunoglobulin gamma Fc region receptor II-a
FCRL5	Fc receptor-like protein 5

FCRL6	Fc receptor-like protein 6
FCRLB	Fc receptor-like B
FDX1	Adrenodoxin, mitochondrial
FGF23	Fibroblast growth factor 23
FGFBP2	Fibroblast growth factor-binding protein 2
FGL1	Fibrinogen-like protein 1
FLT1	Vascular endothelial growth factor receptor 1
FNDC1	Fibronectin type III domain-containing protein 1
FOLR1	Folate receptor alpha
FSTL3	Follistatin-related protein 3
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2
GALNT5	Polypeptide N-acetylgalactosaminyltransferase 5
GALNT7	N-acetylgalactosaminyltransferase 7
GDF15	Growth/differentiation factor 15
GDF2	Growth/differentiation factor 2
GFRA3	GDNF family receptor alpha-3
GFRAL	GDNF family receptor alpha-like
GH1	Somatotropin
GHR	Growth hormone receptor
GIPC2	PDZ domain-containing protein GIPC2
GOLM2	Protein GOLM2
GPC1	Glypican-1
GPHA2	Glycoprotein hormone alpha-2
GPR37	Prosaposin receptor GPR37
GSN	Gelsolin
GSTA3	Glutathione S-transferase A3
GUSB	Beta-glucuronidase
HAVCR1	Hepatitis A virus cellular receptor 1
HJV	Hemojuvelin
HMCN2	Hemicentin-2
HPGDS	Hematopoietic prostaglandin D synthase
HRG	Histidine-rich glycoprotein
ICAM1	Intercellular adhesion molecule 1
ICAM3	Intercellular adhesion molecule 3
ICAM5	Intercellular adhesion molecule 5
ICOSLG	ICOS ligand
IDI2	Isopentenyl-diphosphate delta-isomerase 2
IDS	Iduronate 2-sulfatase
IFI30	Gamma-interferon-inducible lysosomal thiol reductase
IFNL2	Interferon lambda-2
IGDCC3	Immunoglobulin superfamily DCC subclass member 3
IGFBP1	Insulin-like growth factor-binding protein 1
IGFBP2	Insulin-like growth factor-binding protein 2
IGFBP4	Insulin-like growth factor-binding protein 4
IGFBPL1	Insulin-like growth factor-binding protein-like 1
IGLC2	Immunoglobulin lambda constant 2
IGSF9	Protein turtle homolog A
IL13RA1	Interleukin-13 receptor subunit alpha-1
IL15	Interleukin-15
IL17C	Interleukin-17C
IL17RB	Interleukin-17 receptor B
IL1RL1	Interleukin-1 receptor-like 1
IL2RA	Interleukin-2 receptor subunit alpha
IL2RB	Interleukin-2 receptor subunit beta

IL31RA	Interleukin-31 receptor subunit alpha
IL5RA	Interleukin-5 receptor subunit alpha
IL6	Interleukin-6
IL7R	Interleukin-7 receptor subunit alpha
INSR	Insulin receptor
IPF	Idiopathic pulmonary fibrosis
ITGAM	Integrin alpha-M
ITGB6	Integrin beta-6
ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3
ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4
KIRREL2	Kin of IRRE-like protein 2
KLK4	Kallikrein-4
KLK7	Kallikrein-7
KLK8	Kallikrein-8
KLKB1	Plasma kallikrein
KLRB1	Killer cell lectin-like receptor subfamily B member 1
KLRF1	Killer cell lectin-like receptor subfamily F member 1
KLRK1	NKG2-D type II integral membrane protein
KRT18	Keratin, type I cytoskeletal 18
KRT19	Keratin, type I cytoskeletal 19
L1CAM	Neural cell adhesion molecule L1
LAMP2	Lysosome-associated membrane glycoprotein 2
LAMP3	Lysosome-associated membrane glycoprotein 3
LAYN	Layilin
LCP1	Plastin-2
LEG1	Protein LEG1 homolog
LEO1	RNA polymerase-associated protein LEO1
LEP	Leptin
LIF	Leukemia inhibitory factor
LMOD1	Leiomodin-1
LRG1	Leucine-rich alpha-2-glycoprotein
LRIG3	Leucine-rich repeats and immunoglobulin-like domains protein 3
LRRN1	Leucine-rich repeat neuronal protein 1
LTBP2	Latent transforming growth factor beta binding protein 2
LY96	Lymphocyte antigen 96
MAMDC2	MAM domain-containing protein 2
MAMDC4	Apical endosomal glycoprotein
MAN2B2	Epididymis-specific alpha-mannosidase
MATN2	Matrilin-2
MATN3	Matrilin-3
MELTF	Melanotransferrin
MEP1B	Meprin A subunit beta
MPRIP	Myosin phosphatase Rho-interacting protein
MRC1	Macrophage mannose receptor 1
MSTN	Growth/differentiation factor 8
MZB1	Marginal zone B- and B1-cell-specific protein
NCAM2	Neural cell adhesion molecule 2
NCR3LG1	Natural cytotoxicity triggering receptor 3 ligand 1
NFATC3	Nuclear factor of activated T-cells, cytoplasmic 3
NPL	N-acetylneuraminate lyase
NPTX2	Neuronal pentraxin-2
NPY	Pro-neuropeptide Y
NRP2	Neuropilin-2
NT5C1A	Cytosolic 5'-nucleotidase 1A

NTPROBNP	N-terminal prohormone of brain natriuretic peptide
NTRK2	BDNF/NT-3 growth factors receptor
OCLN	Occludin
ORM1	Alpha-1-acid glycoprotein 1
OSMR	Oncostatin-M-specific receptor subunit beta
PALM2	Paralemmin-2
PI16	Peptidase inhibitor 16
PIGR	Polymeric immunoglobulin receptor
PLXDC1	Plexin domain-containing protein 1
PLXDC2	Plexin domain-containing protein 2
PPF	Progressive pulmonary fibrosis
PPL	Periplakin
PPP1R14D	Protein phosphatase 1 regulatory subunit 14D
PRND	Prion-like protein doppel
PRSS22	Brain-specific serine protease 4
PSAP	Prosaposin
PTGR1	Prostaglandin reductase 1
PTK7	Inactive tyrosine-protein kinase 7
RBFOX3	RNA binding protein fox-1 homolog 3
RET	Proto-oncogene tyrosine-protein kinase receptor Ret
RNASE1	Ribonuclease pancreatic
ROBO2	Roundabout homolog 2
RTN4R	Reticulon-4 receptor
S100P	Protein S100-P
SCGB3A2	Secretoglobin family 3A member 2
SDC1	Syndecan-1
SEL1L	Protein sel-1 homolog 1
SELPLG	P-selectin glycoprotein ligand 1
SEPTIN8	Septin-8
SERPINA1	Alpha-1-antitrypsin
SERPINA3	Alpha-1-antichymotrypsin
SERPINA5	Plasma serine protease inhibitor
SERPINA7	Thyroxine-binding globulin
SERPINF1	Pigment epithelium-derived factor
SFTPA1	Pulmonary surfactant-associated protein A1
SFTPA2	Pulmonary surfactant-associated protein A2
SHISA5	Protein shisa-5
SIGLEC8	Sialic acid-binding Ig-like lectin 8
SPINK1	Serine protease inhibitor Kazal-type 1
SPINT1	Kunitz-type protease inhibitor 1
SPON1	Spondin-1
SSC4D	Scavenger receptor cysteine-rich domain-containing group B protein
TCL1A	T-cell leukemia/lymphoma protein 1A
TF	Serotransferrin
TFF2	Trefoil factor 2
TFF3	Trefoil factor 3
TGF-β	transforming growth factor beta
TGFBR1	TGF-beta receptor type-1
THBS2	Thrombospondin-2
THBS4	Thrombospondin-4
THSD1	Thrombospondin type-1 domain-containing protein 1
TMPRSS15	Enteropeptidase
TNC	Tenascin
TNFRSF10B	Tumor necrosis factor receptor superfamily member 10B

TNFRSF1A	Tumor necrosis factor receptor superfamily member 1A
TNFSF10	Tumor necrosis factor ligand superfamily member 10
TNFSF11	Tumor necrosis factor ligand superfamily member 11
TNFSF13B	Tumor necrosis factor ligand superfamily member 13B
TNXB	Tenascin-X
TPPP3	Tubulin polymerization-promoting protein family member 3
TREH	Trehalase
TTR	Transthyretin
UPB1	Beta-ureidopropionase
VCAN	Versican core protein
VEGFD	Vascular endothelial growth factor D
VSIG2	V-set and immunoglobulin domain-containing protein 2
WFDC12	WAP four-disulfide core domain protein 12
WFDC2	WAP four-disulfide core domain protein 2
WFIKKN1	WAP, Kazal, immunoglobulin, Kunitz and NTR domain-containing protein 1
XG	Glycoprotein Xg
XPNPEP2	Xaa-Pro aminopeptidase 2
YAP1	Transcriptional coactivator YAP1

Bibliography

Adegunsoye, A., Oldham, J. M., Bellam, S. K., Montner, S., Churpek, M. M., Noth, I., Vij, R., Strek, M. E., & Chung, J. H. (2019). Computed Tomography Honeycombing Identifies a Progressive Fibrotic Phenotype with Increased Mortality across Diverse Interstitial Lung Diseases. *Ann Am Thorac Soc*, *16*(5), 580-588. https://doi.org/10.1513/AnnalsATS.201807-443OC

Adegunsoye, A., Oldham, J. M., Fernandez Perez, E. R., Hamblin, M., Patel, N., Tener, M., Bhanot, D., Robinson, L., Bullick, S., Chen, L., Hsu, S., Churpek, M., Hedeker, D., Montner, S., Chung, J. H., Husain, A. N., Noth, I., Strek, M. E., & Vij, R. (2017). Outcomes of immunosuppressive therapy in chronic hypersensitivity pneumonitis. *ERJ Open Res*, *3*(3). https://doi.org/10.1183/23120541.00016-2017

Allison, A. C. (2005). Mechanisms of action of mycophenolate mofetil. *Lupus*, *14 Suppl 1*, s2-8. https://doi.org/10.1191/0961203305lu2109oa

Alqalyoobi, S., Adegunsoye, A., Linderholm, A., Hrusch, C., Cutting, C., Ma, S. F., Sperling, A., Noth, I., Strek, M. E., & Oldham, J. M. (2020). Circulating Plasma Biomarkers of Progressive Interstitial Lung Disease. *Am J Respir Crit Care Med*, *201*(2), 250-253. https://doi.org/10.1164/rccm.201907-1343LE

Alqalyoobi, S., Smith, J. A., Maddali, M. V., Pugashetti, J. V., Newton, C. A., Kim, J. S., Linderholm, A. L., Chen, C. H., Ma, S. F., Bose, S., Murray, S., Adegunsoye, A., Strek, M. E., Garcia, C. K., Wolters, P. J., Martinez, F. J., Noth, I., & Oldham, J. M. (2025). Proteomic Biomarkers of Survival in Non-IPF Interstitial Lung Disease. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202407-1506OC

Aquino-Galvez, A., Camarena, A., Montano, M., Juarez, A., Zamora, A. C., Gonzalez-Avila, G., Checa, M., Sandoval-Lopez, G., Vargas-Alarcon, G., Granados, J., Pardo, A., Zuniga, J., & Selman, M. (2008). Transporter associated with antigen processing (TAP) 1 gene polymorphisms in patients with hypersensitivity pneumonitis. *Exp Mol Pathol*, *84*(2), 173-177. https://doi.org/10.1016/j.yexmp.2008.01.002

Arem, H., & Loftfield, E. (2018). Cancer Epidemiology: A Survey of Modifiable Risk Factors for Prevention and Survivorship. *Am J Lifestyle Med*, *12*(3), 200-210. https://doi.org/10.1177/1559827617700600

Assarsson, E., Lundberg, M., Holmquist, G., Bjorkesten, J., Thorsen, S. B., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., Andersson, A. C., Lindstedt, P., Stenvang, J., Gullberg, M., & Fredriksson, S. (2014). Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*, *9*(4), e95192. https://doi.org/10.1371/journal.pone.0095192

Barrera, L., Mendoza, F., Zuniga, J., Estrada, A., Zamora, A. C., Melendro, E. I., Ramirez, R., Pardo, A., & Selman, M. (2008). Functional diversity of T-cell subpopulations in subacute and chronic hypersensitivity pneumonitis. *Am J Respir Crit Care Med*, *177*(1), 44-55. https://doi.org/10.1164/rccm.200701-093OC

Bauer, Y., White, E. S., de Bernard, S., Cornelisse, P., Leconte, I., Morganti, A., Roux, S., & Nayler, O. (2017). MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Res*, *3*(1). https://doi.org/10.1183/23120541.00074-2016

Baumgartner, K. B., Samet, J. M., Stidley, C. A., Colby, T. V., & Waldron, J. A. (1997). Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *155*(1), 242-248. https://doi.org/10.1164/ajrccm.155.1.9001319

Behr, J., Prasse, A., Kreuter, M., Johow, J., Rabe, K. F., Bonella, F., Bonnet, R., Grohe, C., Held, M., Wilkens, H., Hammerl, P., Koschel, D., Blaas, S., Wirtz, H., Ficker, J. H., Neumeister, W., Schonfeld, N., Claussen, M., Kneidinger, N., . . . investigators, R. (2021). Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. *Lancet Respir Med*. https://doi.org/10.1016/S2213-2600(20)30554-3

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300. https://doi.org/https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bennett, D., Fossi, A., Bargagli, E., Refini, R. M., Pieroni, M., Luzzi, L., Ghiribelli, C., Paladini, P., Voltolini, L., & Rottoli, P. (2015). Mortality on the Waiting List for Lung Transplantation in Patients with Idiopathic Pulmonary Fibrosis: A Single-Centre Experience. *Lung*, *193*(5), 677-681. https://doi.org/10.1007/s00408-015-9767-x

Bjoraker, J. A., Ryu, J. H., Edwin, M. K., Myers, J. L., Tazelaar, H. D., Schroeder, D. R., & Offord, K. P. (1998). Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *157*(1), 199-203. https://doi.org/10.1164/ajrccm.157.1.9704130

Blanchet, M. R., Bennett, J. L., Gold, M. J., Levantini, E., Tenen, D. G., Girard, M., Cormier, Y., & McNagny, K. M. (2011). CD34 is required for dendritic cell trafficking and pathology in murine hypersensitivity pneumonitis. *Am J Respir Crit Care Med*, *184*(6), 687-698. https://doi.org/10.1164/rccm.201011-1764OC

Bowman, W. S., Echt, G. A., & Oldham, J. M. (2021). Biomarkers in Progressive Fibrosing Interstitial Lung Disease: Optimizing Diagnosis, Prognosis, and Treatment Response. *Front Med (Lausanne)*, *8*, 680997. https://doi.org/10.3389/fmed.2021.680997

Bowman, W. S., Newton, C. A., Linderholm, A. L., Neely, M. L., Pugashetti, J. V., Kaul, B., Vo, V., Echt, G. A., Leon, W., Shah, R. J., Huang, Y., Garcia, C. K., Wolters, P. J., & Oldham, J. M. (2022). Proteomic biomarkers of progressive fibrosing interstitial lung disease: a multicentre cohort analysis. *Lancet Respir Med*, *10*(6), 593-602. https://doi.org/10.1016/S2213-2600(21)00503-8

Brown, K. K., Martinez, F. J., Walsh, S. L. F., Thannickal, V. J., Prasse, A., Schlenker-Herceg, R., Goeldner, R. G., Clerisme-Beaty, E., Tetzlaff, K., Cottin, V., & Wells, A. U. (2020). The natural history of progressive fibrosing interstitial lung diseases. *Eur Respir J*, *55*(6). https://doi.org/10.1183/13993003.00085-2020

Bueno, M., Suresh, P., Chuan, B., Sembrat, J., Koenigshoff, M., & Eickelberg, O. MZB1+ Plasma Cells as Potential Therapeutic Targets for Cell Depletion in Interstitial Lung Diseases. In *B29. BIOMARKERS AND NEW THERAPEUTICS TARGETS IN LUNG FIBROSIS* (pp. A3227-A3227). https://doi.org/10.1164/ajrccm-conference.2024.209.1_MeetingAbstracts.A3227

Bustos, M. L., Frias, S., Ramos, S., Estrada, A., Arreola, J. L., Mendoza, F., Gaxiola, M., Salcedo, M., Pardo, A., & Selman, M. (2007). Local and circulating microchimerism is associated with hypersensitivity pneumonitis. *Am J Respir Crit Care Med*, *176*(1), 90-95. https://doi.org/10.1164/rccm.200608-1129OC

Camarena, A., Juarez, A., Mejia, M., Estrada, A., Carrillo, G., Falfan, R., Zuniga, J., Navarro, C., Granados, J., & Selman, M. (2001). Major histocompatibility complex and tumor necrosis factor-alpha polymorphisms in pigeon breeder's disease. *Am J Respir Crit Care Med*, *163*(7), 1528-1533. https://doi.org/10.1164/ajrccm.163.7.2004023

Camiciottoli, G., Orlandi, I., Bartolucci, M., Meoni, E., Nacci, F., Diciotti, S., Barcaroli, C., Conforti, M. L., Pistolesi, M., Matucci-Cerinic, M., & Mascalchi, M. (2007). Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *Chest*, *131*(3), 672-681. https://doi.org/10.1378/chest.06-1401

Castelino, F. V., & Varga, J. (2010). Interstitial lung disease in connective tissue diseases: evolving concepts of pathogenesis and management. *Arthritis Res Ther*, *12*(4), 213. https://doi.org/10.1186/ar3097

Cerny, T., Borisch, B., Introna, M., Johnson, P., & Rose, A. L. (2002). Mechanism of action of rituximab. *Anticancer Drugs*, *13 Suppl 2*, S3-10. https://doi.org/10.1097/00001813-200211002-00002

Charkhchi, P., Cybulski, C., Gronwald, J., Wong, F. O., Narod, S. A., & Akbari, M. R. (2020). CA125 and Ovarian Cancer: A Comprehensive Review. *Cancers (Basel)*, *12*(12). https://doi.org/10.3390/cancers12123730

Chen, L., Jin, X., Ma, J., Xiang, B., & Li, X. (2023). YAP at the progression of inflammation. *Front Cell Dev Biol*, *11*, 1204033. https://doi.org/10.3389/fcell.2023.1204033

Chen, Y., Liu, C., Lu, W., Li, M., Hadadi, C., Wang, E. W., Yang, K., Lai, N., Huang, J., Li, S., Zhong, N., Zhang, N., & Wang, J. (2016). Clinical characteristics and risk factors of pulmonary hypertension associated with chronic respiratory diseases: a retrospective study. *J Thorac Dis*, *8*(3), 350-358. https://doi.org/10.21037/jtd.2016.02.58

Collard, H. R., King, T. E., Jr., Bartelson, B. B., Vourlekis, J. S., Schwarz, M. I., & Brown, K. K. (2003). Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *168*(5), 538-542. https://doi.org/10.1164/rccm.200211-1311OC

Cottin, V., Hirani, N. A., Hotchkin, D. L., Nambiar, A. M., Ogura, T., Otaola, M., Skowasch, D., Park, J. S., Poonyagariyagorn, H. K., Wuyts, W., & Wells, A. U. (2018). Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*, *27*(150). https://doi.org/10.1183/16000617.0076-2018

Craig, V. J., Zhang, L., Hagood, J. S., & Owen, C. A. (2015). Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol*, *53*(5), 585-600. https://doi.org/10.1165/rcmb.2015-0020TR

Cutting, C., Bowman, W. S., Dao, N., Pugashetti, J. V., Garcia, C. K., Oldham, J. M., & Newton, C. A. (2021). Family History of Pulmonary Fibrosis Predicts Worse Survival in Patients with Interstitial Lung Disease. *Chest*. https://doi.org/10.1016/j.chest.2021.01.026

Das, S., Dey, M. K., Devireddy, R., & Gartia, M. R. (2023). Biomarkers in Cancer Detection, Diagnosis, and Prognosis. *Sensors (Basel)*, *24*(1). https://doi.org/10.3390/s24010037

DeDent, A. M., Collard, H. R., & Thakur, N. (2022). Disparities in Rural Populations With Idiopathic Pulmonary Fibrosis. *Chest*, *162*(3), 630-634. https://doi.org/10.1016/j.chest.2022.03.042

Dempsey, T. M., Sangaralingham, L. R., Yao, X., Sanghavi, D., Shah, N. D., & Limper, A. H. (2019). Clinical Effectiveness of Antifibrotic Medications for Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *200*(2), 168-174. https://doi.org/10.1164/rccm.201902-0456OC

Di Carmine, S., Scott, M. M., McLean, M. H., & McSorley, H. J. (2022). The role of interleukin-33 in organ fibrosis. *Discovery Immunology*, *1*(1). https://doi.org/10.1093/discim/kyac006

Distler, O., Assassi, S., Cottin, V., Cutolo, M., Danoff, S. K., Denton, C. P., Distler, J. H. W., Hoffmann-Vold, A. M., Johnson, S. R., Muller Ladner, U., Smith, V., Volkmann, E. R., & Maher, T. M. (2020). Predictors of progression in systemic sclerosis patients with interstitial lung disease. *Eur Respir J*, *55*(5). https://doi.org/10.1183/13993003.02026-2019

du Bois, R. M. (2012). An earlier and more confident diagnosis of idiopathic pulmonary fibrosis. *Eur Respir Rev*, *21*(124), 141-146. https://doi.org/10.1183/09059180.00000812

Duarte, A. C., Cordeiro, A., Fernandes, B. M., Bernardes, M., Martins, P., Cordeiro, I., Santiago, T., Seixas, M. I., Ribeiro, A. R., & Santos, M. J. (2019). Rituximab in connective tissue disease-associated interstitial lung disease. *Clin Rheumatol*, *38*(7), 2001-2009. https://doi.org/10.1007/s10067-019-04557-7

Duchemann, B., Annesi-Maesano, I., Jacobe de Naurois, C., Sanyal, S., Brillet, P. Y., Brauner, M., Kambouchner, M., Huynh, S., Naccache, J. M., Borie, R., Piquet, J., Mekinian, A., Virally, J., Uzunhan, Y., Cadranel, J., Crestani, B., Fain, O., Lhote, F., Dhote, R., . . . Nunes, H. (2017). Prevalence and incidence of interstitial lung diseases in a multi-ethnic county of Greater Paris. *Eur Respir J*, *50*(2). https://doi.org/10.1183/13993003.02419-2016

Dulauroy, S., Di Carlo, S. E., Langa, F., Eberl, G., & Peduto, L. (2012). Lineage tracing and genetic ablation of ADAM12(+) perivascular cells identify a major source of profibrotic cells during acute tissue injury. *Nat Med*, *18*(8), 1262-1270. https://doi.org/10.1038/nm.2848

Durheim, M. T., Collard, H. R., Roberts, R. S., Brown, K. K., Flaherty, K. R., King, T. E., Jr., Palmer, S. M., Raghu, G., Snyder, L. D., Anstrom, K. J., Martinez, F. J., & investigators, I. P. (2015). Association of hospital admission and forced vital capacity endpoints with survival in patients with idiopathic pulmonary fibrosis: analysis of a pooled cohort from three clinical trials. *Lancet Respir Med*, *3*(5), 388-396. https://doi.org/10.1016/S2213-2600(15)00093-4

Etminan, M., Brophy, J. M., Collins, G., Nazemipour, M., & Mansournia, M. A. (2021). To Adjust or Not to Adjust: The Role of Different Covariates in Cardiovascular Observational Studies. *American Heart Journal*, *237*, 62-67. https://doi.org/https://doi.org/10.1016/j.ahj.2021.03.008

Fathi, M., Dastmalchi, M., Rasmussen, E., Lundberg, I. E., & Tornling, G. (2004). Interstitial lung disease, a common manifestation of newly diagnosed polymyositis and dermatomyositis. *Ann Rheum Dis*, *63*(3), 297-301. https://doi.org/10.1136/ard.2003.006122

Fernández Pérez, E. R., Daniels, C. E., Schroeder, D. R., St Sauver, J., Hartman, T. E., Bartholmai, B. J., Yi, E. S., & Ryu, J. H. (2010). Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest*, *137*(1), 129-137. https://doi.org/10.1378/chest.09-1002

Fernandez Perez, E. R., Kong, A. M., Raimundo, K., Koelsch, T. L., Kulkarni, R., & Cole, A. L. (2018). Epidemiology of Hypersensitivity Pneumonitis among an Insured Population in the

United States: A Claims-based Cohort Analysis. *Ann Am Thorac Soc*, *15*(4), 460-469. https://doi.org/10.1513/AnnalsATS.201704-288OC

Fischer, A., Antoniou, K. M., Brown, K. K., Cadranel, J., Corte, T. J., du Bois, R. M., Lee, J. S., Leslie, K. O., Lynch, D. A., Matteson, E. L., Mosca, M., Noth, I., Richeldi, L., Strek, M. E., Swigris, J. J., Wells, A. U., West, S. G., Collard, H. R., Cottin, V., & CTD-ILD, E. A. T. F. o. U. F. o. (2015). An official European Respiratory Society/American Thoracic Society research statement: interstitial pneumonia with autoimmune features. *Eur Respir J*, *46*(4), 976-987. https://doi.org/10.1183/13993003.00150-2015

Fischer, A., Brown, K. K., Du Bois, R. M., Frankel, S. K., Cosgrove, G. P., Fernandez-Perez, E. R., Huie, T. J., Krishnamoorthy, M., Meehan, R. T., Olson, A. L., Solomon, J. J., & Swigris, J. J. (2013). Mycophenolate mofetil improves lung function in connective tissue disease-associated interstitial lung disease. *J Rheumatol*, *40*(5), 640-646. https://doi.org/10.3899/jrheum.121043

Fischer, A., & du Bois, R. (2012). Interstitial lung disease in connective tissue disorders. *Lancet*, *380*(9842), 689-698. https://doi.org/10.1016/S0140-6736(12)61079-4

Flaherty, K. R., Brown, K. K., Wells, A. U., Clerisme-Beaty, E., Collard, H. R., Cottin, V., Devaraj, A., Inoue, Y., Le Maulf, F., Richeldi, L., Schmidt, H., Walsh, S., Mezzanotte, W., & Schlenker-Herceg, R. (2017). Design of the PF-ILD trial: a double-blind, randomised, placebo-controlled phase III trial of nintedanib in patients with progressive fibrosing interstitial lung disease. *BMJ Open Respir Res*, *4*(1), e000212. https://doi.org/10.1136/bmjresp-2017-000212

Flaherty, K. R., Wells, A. U., Cottin, V., Devaraj, A., Walsh, S. L. F., Inoue, Y., Richeldi, L., Kolb, M., Tetzlaff, K., Stowasser, S., Coeck, C., Clerisme-Beaty, E., Rosenstock, B., Quaresma, M., Haeufel, T., Goeldner, R. G., Schlenker-Herceg, R., Brown, K. K., & Investigators, I. T. (2019). Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. *N Engl J Med*. https://doi.org/10.1056/NEJMoa1908681

Flament, T., Bigot, A., Chaigne, B., Henique, H., Diot, E., & Marchand-Adam, S. (2016). Pulmonary manifestations of Sjogren's syndrome. *Eur Respir Rev*, *25*(140), 110-123. https://doi.org/10.1183/16000617.0011-2016

Gabbay, E., Tarala, R., Will, R., Carroll, G., Adler, B., Cameron, D., & Lake, F. R. (1997). Interstitial lung disease in recent onset rheumatoid arthritis. *Am J Respir Crit Care Med*, *156*(2 Pt 1), 528-535. https://doi.org/10.1164/ajrccm.156.2.9609016

Garcia de Alba, C., Buendia-Roldan, I., Salgado, A., Becerril, C., Ramirez, R., Gonzalez, Y., Checa, M., Navarro, C., Ruiz, V., Pardo, A., & Selman, M. (2015). Fibrocytes contribute to inflammation and fibrosis in chronic hypersensitivity pneumonitis through paracrine effects. *Am J Respir Crit Care Med*, *191*(4), 427-436. https://doi.org/10.1164/rccm.201407-1334OC

George, T. J., Arnaoutakis, G. J., & Shah, A. S. (2011). Lung transplant in idiopathic pulmonary fibrosis. *Arch Surg*, *146*(10), 1204-1209. https://doi.org/10.1001/archsurg.2011.239

Gimenez, A., Storrer, K., Kuranishi, L., Soares, M. R., Ferreira, R. G., & Pereira, C. A. C. (2018). Change in FVC and survival in chronic fibrotic hypersensitivity pneumonitis. *Thorax*, *73*(4), 391-392. https://doi.org/10.1136/thoraxjnl-2017-210035

Goh, N. S., Hoyles, R. K., Denton, C. P., Hansell, D. M., Renzoni, E. A., Maher, T. M., Nicholson, A. G., & Wells, A. U. (2017). Short-Term Pulmonary Function Trends Are Predictive of Mortality in Interstitial Lung Disease Associated With Systemic Sclerosis. *Arthritis Rheumatol*, *69*(8), 1670-1678. https://doi.org/10.1002/art.40130

Goobie, G. C., Carlsten, C., Johannson, K. A., Khalil, N., Marcoux, V., Assayag, D., Manganas, H., Fisher, J. H., Kolb, M. R. J., Lindell, K. O., Fabisiak, J. P., Chen, X., Gibson, K. F., Zhang, Y., Kass, D. J., Ryerson, C. J., & Nouraie, S. M. (2022). Association of Particulate Matter Exposure With Lung Function and Mortality Among Patients With Fibrotic Interstitial Lung Disease. *JAMA Intern Med*, *182*(12), 1248-1259. https://doi.org/10.1001/jamainternmed.2022.4696

Gubatan, J., Holman, D. R., Puntasecca, C. J., Polevoi, D., Rubin, S. J., & Rogalla, S. (2021). Antimicrobial peptides and the gut microbiome in inflammatory bowel disease. *World J Gastroenterol*, *27*(43), 7402-7422. https://doi.org/10.3748/wjg.v27.i43.7402

Haikala, H. M., & Jänne, P. A. (2021). Thirty Years of HER3: From Basic Biology to Therapeutic Interventions. *Clinical Cancer Research*, *27*(13), 3528-3539. https://doi.org/10.1158/1078-0432.Ccr-20-4465

Hamaguchi, Y., Nishizawa, Y., Yasui, M., Hasegawa, M., Kaburagi, Y., Komura, K., Nagaoka, T., Saito, E., Shimada, Y., Takehara, K., Kadono, T., Steeber, D. A., Tedder, T. F., & Sato, S. (2002). Intercellular adhesion molecule-1 and L-selectin regulate bleomycin-induced lung fibrosis. *Am J Pathol*, *161*(5), 1607-1618. https://doi.org/10.1016/s0002-9440(10)64439-2

Han, M. K., Zhou, Y., Murray, S., Tayob, N., Noth, I., Lama, V. N., Moore, B. B., White, E. S., Flaherty, K. R., Huffnagle, G. B., Martinez, F. J., & Investigators, C. (2014). Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med*, *2*(7), 548-556. https://doi.org/10.1016/S2213-2600(14)70069-4

He, J., Ogden, L. G., Bazzano, L. A., Vupputuri, S., Loria, C., & Whelton, P. K. (2001). Risk factors for congestive heart failure in US men and women: NHANES I epidemiologic follow-up study. *Arch Intern Med*, *161*(7), 996-1002. https://doi.org/10.1001/archinte.161.7.996

He, X., Tolosa, M. F., Zhang, T., Goru, S. K., Ulloa Severino, L., Misra, P. S., McEvoy, C. M., Caldwell, L., Szeto, S. G., Gao, F., Chen, X., Atin, C., Ki, V., Vukosa, N., Hu, C., Zhang, J., Yip, C., Krizova, A., Wrana, J. L., & Yuen, D. A. (2022). Myofibroblast YAP/TAZ activation is a key step in organ fibrogenesis. *JCI Insight*, *7*(4). https://doi.org/10.1172/jci.insight.146243

Hill, C., & Wang, Y. (2022). Autophagy in pulmonary fibrosis: friend or foe? *Genes & Diseases*, *9*(6), 1594-1607. https://doi.org/https://doi.org/10.1016/j.gendis.2021.09.008

Huang, Y., Ma, S. F., Oldham, J. M., Adegunsoye, A., Zhu, D., Murray, S., Kim, J. S., Bonham, C., Strickland, E., Linderholm, A. L., Lee, C. T., Paul, T., Mannem, H., Maher, T. M., Molyneaux, P. L., Strek, M. E., Martinez, F. J., & Noth, I. (2024). Machine Learning of Plasma Proteomics Classifies Diagnosis of Interstitial Lung Disease. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202309-1692OC

Huapaya, J. A., Silhan, L., Pinal-Fernandez, I., Casal-Dominguez, M., Johnson, C., Albayda, J., Paik, J. J., Sanyal, A., Mammen, A. L., Christopher-Stine, L., & Danoff, S. K. (2019). Long-Term Treatment With Azathioprine and Mycophenolate Mofetil for Myositis-Related Interstitial Lung Disease. *Chest*, *156*(5), 896-906. https://doi.org/10.1016/j.chest.2019.05.023

Humphries, S. M., Mackintosh, J. A., Jo, H. E., Walsh, S. L. F., Silva, M., Calandriello, L., Chapman, S., Ellis, S., Glaspole, I., Goh, N., Grainge, C., Hopkins, P. M. A., Keir, G. J., Moodley, Y., Reynolds, P. N., Walters, E. H., Baraghoshi, D., Wells, A. U., Lynch, D. A.,

& Corte, T. J. Quantitative computed tomography predicts outcomes in idiopathic pulmonary fibrosis. *Respirology*, *n/a*(n/a). https://doi.org/https://doi.org/10.1111/resp.14333

Humphries, S. M., Swigris, J. J., Brown, K. K., Strand, M., Gong, Q., Sundy, J. S., Raghu, G., Schwarz, M. I., Flaherty, K. R., Sood, R., O'Riordan, T. G., & Lynch, D. A. (2018). Quantitative high-resolution computed tomography fibrosis score: performance characteristics in idiopathic pulmonary fibrosis. *Eur Respir J*, *52*(3). https://doi.org/10.1183/13993003.01384-2018

Idiopathic Pulmonary Fibrosis Clinical Research, N., Raghu, G., Anstrom, K. J., King, T. E., Jr., Lasky, J. A., & Martinez, F. J. (2012). Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med*, *366*(21), 1968-1977. https://doi.org/10.1056/NEJMoa1113354

Invernizzi, R., Wu, B. G., Barnett, J., Ghai, P., Kingston, S., Hewitt, R. J., Feary, J., Li, Y., Chua, F., Wu, Z., Wells, A. U., Renzoni, E. A., Nicholson, A. G., Rice, A., Devaraj, A., Segal, L. N., Byrne, A. J., Maher, T. M., Lloyd, C. M., & Molyneaux, P. L. (2020). The Respiratory Microbiome in Chronic Hypersensitivity Pneumonitis is Distinct from that of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202002-0460OC

Isgrò, M. A., Bottoni, P., & Scatena, R. (2015). Neuron-Specific Enolase as a Biomarker: Biochemical and Clinical Aspects. *Adv Exp Med Biol*, *867*, 125-143. https://doi.org/10.1007/978-94-017-7215-0_9

Jacob, J., Bartholmai, B. J., Rajagopalan, S., Kokosi, M., Nair, A., Karwoski, R., Walsh, S. L., Wells, A. U., & Hansell, D. M. (2017). Mortality prediction in idiopathic pulmonary fibrosis: evaluation of computer-based CT analysis with conventional severity measures. *Eur Respir J*, *49*(1). https://doi.org/10.1183/13993003.01011-2016

Jacob, J., Hirani, N., van Moorsel, C. H. M., Rajagopalan, S., Murchison, J. T., van Es, H. W., Bartholmai, B. J., van Beek, F. T., Struik, M. H. L., Stewart, G. A., Kokosi, M., Egashira, R., Brun, A. L., Cross, G., Barnett, J., Devaraj, A., Margaritopoulos, G., Karwoski, R., Renzoni, E., . . . Wells, A. U. (2019). Predicting outcomes in rheumatoid arthritis related interstitial lung disease. *Eur Respir J*, *53*(1). https://doi.org/10.1183/13993003.00869-2018

Jimenez-Alvarez, L., Arreola, J. L., Ramirez-Martinez, G., Ortiz-Quintero, B., Gaxiola, M., Reynoso-Robles, R., Avila-Moreno, F., Urrea, F., Pardo, A., Selman, M., & Zuniga, J. (2011). The effect of CTLA-4Ig, a CD28/B7 antagonist, on the lung inflammation and T cell subset profile during murine hypersensitivity pneumonitis. *Exp Mol Pathol*, *91*(3), 718-722. https://doi.org/10.1016/j.yexmp.2011.09.010

Johannson, K. A., Kolb, M., Fisher, J. H., & Walsh, S. L. (2022). Progressive Pulmonary Fibrosis: Putting the Cart Before the Horse. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202206-1166LE

Joshi, A. D., Fong, D. J., Oak, S. R., Trujillo, G., Flaherty, K. R., Martinez, F. J., & Hogaboam, C. M. (2009). Interleukin-17-mediated immunopathogenesis in experimental hypersensitivity pneumonitis. *Am J Respir Crit Care Med*, *179*(8), 705-716. https://doi.org/10.1164/rccm.200811-1700OC

Keir, G. J., Maher, T. M., Hansell, D. M., Denton, C. P., Ong, V. H., Singh, S., Wells, A. U., & Renzoni, E. A. (2012). Severe interstitial lung disease in connective tissue disease:

rituximab as rescue therapy. *Eur Respir J*, *40*(3), 641-648. https://doi.org/10.1183/09031936.00163911

King, T. E., Jr., Bradford, W. Z., Castro-Bernardini, S., Fagan, E. A., Glaspole, I., Glassberg, M. K., Gorina, E., Hopkins, P. M., Kardatzke, D., Lancaster, L., Lederer, D. J., Nathan, S. D., Pereira, C. A., Sahn, S. A., Sussman, R., Swigris, J. J., Noble, P. W., & Group, A. S. (2014). A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*, *370*(22), 2083-2092. https://doi.org/10.1056/NEJMoa1402582

King, T. E., Jr., Brown, K. K., Raghu, G., du Bois, R. M., Lynch, D. A., Martinez, F., Valeyre, D., Leconte, I., Morganti, A., Roux, S., & Behr, J. (2011). BUILD-3: a randomized, controlled trial of bosentan in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *184*(1), 92-99. https://doi.org/10.1164/rccm.201011-1874OC

King, T. E., Jr., Tooze, J. A., Schwarz, M. I., Brown, K. R., & Cherniack, R. M. (2001). Predicting survival in idiopathic pulmonary fibrosis: scoring system and survival model. *Am J Respir Crit Care Med*, *164*(7), 1171-1181. https://doi.org/10.1164/ajrccm.164.7.2003140

Kreuter, F., & Beddo, V. (2004). A Handbook of Statistical Analyses using Stata. *Journal of Statistical Software, Book Reviews*, *11*(3), 1 - 11. https://doi.org/10.18637/jss.v011.b03

Kropski, J. A. (2020). Familial Interstitial Lung Disease. *Semin Respir Crit Care Med*, *41*(2), 229-237. https://doi.org/10.1055/s-0040-1708054

Lagares, D., Ghassemi-Kakroodi, P., Tremblay, C., Santos, A., Probst, C. K., Franklin, A., Santos, D. M., Grasberger, P., Ahluwalia, N., Montesi, S. B., Shea, B. S., Black, K. E., Knipe, R., Blati, M., Baron, M., Wu, B., Fahmi, H., Gandhi, R., Pardo, A., . . . Kapoor, M. (2017). ADAM10-mediated ephrin-B2 shedding promotes myofibroblast activation and organ fibrosis. *Nat Med*, *23*(12), 1405-1415. https://doi.org/10.1038/nm.4419

Lederer, D. J., & Martinez, F. J. (2018). Idiopathic Pulmonary Fibrosis. *N Engl J Med*, *379*(8), 797-798. https://doi.org/10.1056/NEJMc1807508

Lee, S. A., Lee, H. J., Suh, B. C., Shin, H. Y., Kim, S. W., Yoon, B. A., Choi, Y. C., & Park, H. J. (2023). Clinical significance of anti-NT5c1A autoantibody in Korean patients with inflammatory myopathies. *PLoS One*, *18*(4), e0284409. https://doi.org/10.1371/journal.pone.0284409

Ley, B., Collard, H. R., & King, T. E., Jr. (2011). Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *183*(4), 431-440. https://doi.org/10.1164/rccm.201006-0894CI

Ley, B., Liu, S., Elicker, B. M., Henry, T. S., Vittinghoff, E., Golden, J. A., Jones, K. D., & Wolters, P. J. (2020). Telomere length in patients with unclassifiable interstitial lung disease: a cohort study. *Eur Respir J*, *56*(2). https://doi.org/10.1183/13993003.00268-2020

Ley, B., Ryerson, C. J., Vittinghoff, E., Ryu, J. H., Tomassetti, S., Lee, J. S., Poletti, V., Buccioli, M., Elicker, B. M., Jones, K. D., King, T. E., Jr., & Collard, H. R. (2012). A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med*, *156*(10), 684-691. https://doi.org/10.7326/0003-4819-156-10-201205150-00004

Ley, B., Torgerson, D. G., Oldham, J. M., Adegunsoye, A., Liu, S., Li, J., Elicker, B. M., Henry, T. S., Golden, J. A., Jones, K. D., Dressen, A., Yaspan, B. L., Arron, J. R., Noth, I., Hoffmann, T. J., & Wolters, P. J. (2019). Rare Protein-Altering Telomere-related Gene Variants in Patients with Chronic Hypersensitivity Pneumonitis. *Am J Respir Crit Care Med*, *200*(9), 1154-1163. https://doi.org/10.1164/rccm.201902-0360OC

Liang, W., & Diana, J. (2020). The Dual Role of Antimicrobial Peptides in Autoimmunity. *Front Immunol*, *11*, 2077. https://doi.org/10.3389/fimmu.2020.02077

Liebow AA, C. C. (1969). *The interstitial pneumonias, frontiers of pulmonary radiology*. Grune and Stratton.

Lindzen, M., Ghosh, S., Noronha, A., Drago, D., Nataraj, N. B., Leitner, O., Carvalho, S., Zmora, E., Sapoznik, S., Shany, K. B., Levanon, K., Aderka, D., Ramírez, B. S., Dahlhoff, M., McNeish, I., & Yarden, Y. (2021). Targeting autocrine amphiregulin robustly and reproducibly inhibits ovarian cancer in a syngeneic model: roles for wildtype p53. *Oncogene*, *40*(21), 3665-3679. https://doi.org/10.1038/s41388-021-01784-8

Liu, T., De Los Santos, F. G., Ding, L., Wu, Z., & Phan, S. H. (2016). Amphiregulin Promotes Fibroblast Activation in Pulmonary Fibrosis. *The FASEB Journal*, *30*(S1), 50.56-50.56. https://doi.org/https://doi.org/10.1096/fasebj.30.1_supplement.50.6

Lundberg, M., Eriksson, A., Tran, B., Assarsson, E., & Fredriksson, S. (2011). Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res*, *39*(15), e102. https://doi.org/10.1093/nar/gkr424

Maddali, M. V., Moore, A. R., Sinha, P., Newton, C. A., Kim, J. S., Adegunsoye, A., Ma, S. F., Strek, M. E., Chen, C. H., Linderholm, A. L., Zemans, R. L., Moore, B. B., Wolters, P. J., Martinez, F. J., Rogers, A. J., Raj, R., Noth, I., & Oldham, J. M. (2024). Molecular Endotypes of Idiopathic Pulmonary Fibrosis: A Latent Class Analysis of Two Multicenter Observational Cohorts. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202402-0339OC

Maher, T. M., Assassi, S., Azuma, A., Cottin, V., Hoffmann-Vold, A. M., Kreuter, M., Oldham, J. M., Richeldi, L., Valenzuela, C., Wijsenbeek, M. S., Coeck, C., Schlecker, C., Voss, F., Wachtlin, D., & Martinez, F. J. (2023). Design of a phase III, double-blind, randomised, placebo-controlled trial of BI 1015550 in patients with progressive pulmonary fibrosis (FIBRONEER-ILD). *BMJ Open Respir Res*, *10*(1). https://doi.org/10.1136/bmjresp-2022-001580

Maher, T. M., Corte, T. J., Fischer, A., Kreuter, M., Lederer, D. J., Molina-Molina, M., Axmann, J., Kirchgaessler, K. U., & Cottin, V. (2018). Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: design of a double-blind, randomised, placebo-controlled phase II trial. *BMJ Open Respir Res*, *5*(1), e000289. https://doi.org/10.1136/bmjresp-2018-000289

Maher, T. M., Simpson, J. K., Porter, J. C., Wilson, F. J., Chan, R., Eames, R., Cui, Y., Siederer, S., Parry, S., Kenny, J., Slack, R. J., Sahota, J., Paul, L., Saunders, P., Molyneaux, P. L., Lukey, P. T., Rizzo, G., Searle, G. E., Marshall, R. P., . . . Vahdati-Bolouri, M. (2020). A positron emission tomography imaging study to confirm target engagement in the lungs of patients with idiopathic pulmonary fibrosis following a single dose of a novel inhaled alphavbeta6 integrin inhibitor. *Respir Res*, *21*(1), 75. https://doi.org/10.1186/s12931-020-01339-7

Maher, T. M., Tudor, V. A., Saunders, P., Gibbons, M. A., Fletcher, S. V., Denton, C. P., Hoyles, R. K., Parfrey, H., Renzoni, E. A., Kokosi, M., Wells, A. U., Ashby, D., Szigeti, M., & Molyneaux, P. L. (2023). Rituximab versus intravenous cyclophosphamide in patients with connective tissue disease-associated interstitial lung disease in the UK (RECITAL): a double-blind, double-dummy, randomised, controlled, phase 2b trial. *Lancet Respir Med*, *11*(1), 45-54. https://doi.org/10.1016/s2213-2600(22)00359-9

Mankikian, J., Caille, A., Reynaud-Gaubert, M., Agier, M. S., Bermudez, J., Bonniaud, P., Borie, R., Brillet, P. Y., Cadranel, J., Court-Fortune, I., Crestani, B., Debray, M. P., Gomez, E.,

Gondouin, A., Hirschi-Santelmo, S., Israel-Biet, D., Jouneau, S., Juvin, K., Leger, J., . . . Marchand-Adam, S. (2023). Rituximab and mycophenolate mofetil combination in patients with interstitial lung disease (EVER-ILD): a double-blind, randomised, placebo-controlled trial. *Eur Respir J*, *61*(6). https://doi.org/10.1183/13993003.02071-2022

Marchand-Adam, S., Fabre, A., Mailleux, A. A., Marchal, J., Quesnel, C., Kataoka, H., Aubier, M., Dehoux, M., Soler, P., & Crestani, B. (2006). Defect of pro-hepatocyte growth factor activation by fibroblasts in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *174*(1), 58-66. https://doi.org/10.1164/rccm.200507-1074OC

Miller, M. A., Sullivan, R. J., & Lauffenburger, D. A. (2017). Molecular Pathways: Receptor Ectodomain Shedding in Treatment, Resistance, and Monitoring of Cancer. *Clin Cancer Res*, *23*(3), 623-629. https://doi.org/10.1158/1078-0432.Ccr-16-0869

Molyneaux, P. L., Cox, M. J., Wells, A. U., Kim, H. C., Ji, W., Cookson, W. O., Moffatt, M. F., Kim, D. S., & Maher, T. M. (2017). Changes in the respiratory microbiome during acute exacerbations of idiopathic pulmonary fibrosis. *Respir Res*, *18*(1), 29. https://doi.org/10.1186/s12931-017-0511-3

Molyneaux, P. L., Cox, M. J., Willis-Owen, S. A., Mallia, P., Russell, K. E., Russell, A. M., Murphy, E., Johnston, S. L., Schwartz, D. A., Wells, A. U., Cookson, W. O., Maher, T. M., & Moffatt, M. F. (2014). The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *190*(8), 906-913. https://doi.org/10.1164/rccm.201403-0541OC

Molyneaux, P. L., Fahy, W. A., Byrne, A. J., Braybrooke, R., Saunders, P., Toshner, R., Albers, G., Chua, F., Renzoni, E. A., Wells, A. U., Karkera, Y., Oballa, E., Saini, G., Nicholson, A. G., Jenkins, R. G., & Maher, T. M. (2022). CYFRA 21-1 Predicts Progression in Idiopathic Pulmonary Fibrosis: A Prospective Longitudinal Analysis of the PROFILE Cohort. *Am J Respir Crit Care Med*, *205*(12), 1440-1448. https://doi.org/10.1164/rccm.202107-1769OC

Montesi, S. B., Gomez, C. R., Beers, M., Brown, R., Chattopadhyay, I., Flaherty, K. R., Garcia, C. K., Gomperts, B., Hariri, L. P., Hogaboam, C. M., Jenkins, R. G., Kaminski, N., Kim, G. H. J., Königshoff, M., Kolb, M., Kotton, D. N., Kropski, J. A., Lasky, J., Magin, C. M., . . . Martinez, F. J. (2024). Pulmonary Fibrosis Stakeholder Summit: A Joint NHLBI, Three Lakes Foundation, and Pulmonary Fibrosis Foundation Workshop Report. *Am J Respir Crit Care Med*, *209*(4), 362-373. https://doi.org/10.1164/rccm.202307-1154WS

Morisset, J., Johannson, K. A., Vittinghoff, E., Aravena, C., Elicker, B. M., Jones, K. D., Fell, C. D., Manganas, H., Dube, B. P., Wolters, P. J., Collard, H. R., Ryerson, C. J., & Ley, B. (2017). Use of Mycophenolate Mofetil or Azathioprine for the Management of Chronic Hypersensitivity Pneumonitis. *Chest*, *151*(3), 619-625. https://doi.org/10.1016/j.chest.2016.10.029

Neighbors, M., Cabanski, C. R., Ramalingam, T. R., Sheng, X. R., Tew, G. W., Gu, C., Jia, G., Peng, K., Ray, J. M., Ley, B., Wolters, P. J., Collard, H. R., & Arron, J. R. (2018). Prognostic and predictive biomarkers for patients with idiopathic pulmonary fibrosis treated with pirfenidone: post-hoc assessment of the CAPACITY and ASCEND trials. *Lancet Respir Med*, *6*(8), 615-626. https://doi.org/10.1016/S2213-2600(18)30185-1

Newton, C. A., Batra, K., Torrealba, J., Kozlitina, J., Glazer, C. S., Aravena, C., Meyer, K., Raghu, G., Collard, H. R., & Garcia, C. K. (2016). Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. *Eur Respir J*, *48*(6), 1710-1720. https://doi.org/10.1183/13993003.00308-2016

Newton, C. A., Oldham, J. M., Ley, B., Anand, V., Adegunsoye, A., Liu, G., Batra, K., Torrealba, J., Kozlitina, J., Glazer, C., Strek, M. E., Wolters, P. J., Noth, I., & Garcia, C. K. (2019). Telomere length and genetic variant associations with interstitial lung disease progression and survival. *Eur Respir J*, *53*(4). https://doi.org/10.1183/13993003.01641-2018

O'Dwyer, D. N., Kim, J. S., Ma, S. F., Ranjan, P., Das, P., Lipinski, J. H., Metcalf, J. D., Falkowski, N. R., Yow, E., Anstrom, K., Dickson, R. P., Huang, Y., Gilbert, J. A., Martinez, F. J., & Noth, I. (2024). Commensal Oral Microbiota, Disease Severity, and Mortality in Fibrotic Lung Disease. *Am J Respir Crit Care Med*, *209*(9), 1101-1110. https://doi.org/10.1164/rccm.202308-1357OC

Oh, A. S., Lynch, D. A., Swigris, J. J., Baraghoshi, D., Dyer, D. S., Hale, V. A., Koelsch, T. L., Marrocchio, C., Parker, K. N., Teague, S. D., Flaherty, K. R., & Humphries, S. M. (2024). Deep Learning-based Fibrosis Extent on Computed Tomography Predicts Outcome of Fibrosing Interstitial Lung Disease Independent of Visually Assessed Computed Tomography Pattern. *Ann Am Thorac Soc*, *21*(2), 218-227. https://doi.org/10.1513/AnnalsATS.202301-084OC

Okuda, Y. (2008). Review of tocilizumab in the treatment of rheumatoid arthritis. *Biologics*, *2*(1), 75-82. https://doi.org/10.2147/btt.s1828

Oldham, J. M., Huang, Y., Bose, S., Ma, S. F., Kim, J. S., Schwab, A., Ting, C., Mou, K., Lee, C. T., Adegunsoye, A., Ghodrati, S., Vu Pugashetti, J., Nazemi, N., Strek, M. E., Linderholm, A. L., Chen, C. H., Murray, S., Zemans, R. L., Flaherty, K. R., . . . Noth, I. (2023). Proteomic Biomarkers of Survival in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202301-0117OC

Oldham, J. M., Lee, C., Valenzi, E., Witt, L. J., Adegunsoye, A., Hsu, S., Chen, L., Montner, S., Chung, J. H., Noth, I., Vij, R., & Strek, M. E. (2016). Azathioprine response in patients with fibrotic connective tissue disease-associated interstitial lung disease. *Respir Med*, *121*, 117-122. https://doi.org/10.1016/j.rmed.2016.11.007

Oldham, J. M., & Noth, I. (2014). Idiopathic pulmonary fibrosis: early detection and referral. *Respir Med*, *108*(6), 819-829. https://doi.org/10.1016/j.rmed.2014.03.008

Padilla, C. M., Valenzi, E., Tabib, T., Nazari, B., Sembrat, J., Rojas, M., Fuschiotti, P., & Lafyatis, R. (2024). Increased CD8+ tissue resident memory T cells, regulatory T cells and activated natural killer cells in systemic sclerosis lungs. *Rheumatology (Oxford)*, *63*(3), 837-845. https://doi.org/10.1093/rheumatology/kead273

Pardo, A., Barrios, R., Gaxiola, M., Segura-Valdez, L., Carrillo, G., Estrada, A., Mejia, M., & Selman, M. (2000). Increase of lung neutrophils in hypersensitivity pneumonitis is associated with lung fibrosis. *Am J Respir Crit Care Med*, *161*(5), 1698-1704. https://doi.org/10.1164/ajrccm.161.5.9907065

Pardo, A., Smith, K. M., Abrams, J., Coffman, R., Bustos, M., McClanahan, T. K., Grein, J., Murphy, E. E., Zlotnik, A., & Selman, M. (2001). CCL18/DC-CK-1/PARC up-regulation in hypersensitivity pneumonitis. *J Leukoc Biol*, *70*(4), 610-616. https://www.ncbi.nlm.nih.gov/pubmed/11590198

Patterson, K. C., Shah, R. J., Porteous, M. K., Christie, J. D., D'Errico, C. A., Chadwick, M., Triano, M. J., Deshpande, C., Rossman, M. D., Litzky, L. A., Kreider, M., & Miller, W. T., Jr. (2017). Interstitial Lung Disease in the Elderly. *Chest*, *151*(4), 838-844. https://doi.org/10.1016/j.chest.2016.11.003

Planque, C., Li, L., Zheng, Y., Soosaipillai, A., Reckamp, K., Chia, D., Diamandis, E. P., & Goodglick, L. (2008). A multiparametric serum kallikrein panel for diagnosis of non-small

cell lung carcinoma. *Clin Cancer Res*, *14*(5), 1355-1362. https://doi.org/10.1158/1078-0432.Ccr-07-4117

Podolanczuk, A. J., Kim, J. S., Cooper, C. B., Lasky, J. A., Murray, S., Oldham, J. M., Raghu, G., Flaherty, K. R., Spino, C., Noth, I., & Martinez, F. J. (2022). Design and rationale for the prospective treatment efficacy in IPF using genotype for NAC selection (PRECISIONS) clinical trial. *BMC Pulm Med*, *22*(1), 475. https://doi.org/10.1186/s12890-022-02281-8

Pugashetti, J. V., Adegunsoye, A., Wu, Z., Lee, C. T., Srikrishnan, A., Ghodrati, S., Vo, V., Renzoni, E. A., Wells, A. U., Garcia, C. K., Chua, F., Newton, C. A., Molyneaux, P. L., & Oldham, J. M. (2022). Validation of Proposed Criteria for Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202201-0124OC

Pugashetti, J. V., Adegunsoye, A., Wu, Z., Lee, C. T., Srikrishnan, A., Ghodrati, S., Vo, V., Renzoni, E. A., Wells, A. U., Garcia, C. K., Chua, F., Newton, C. A., Molyneaux, P. L., & Oldham, J. M. (2023). Validation of Proposed Criteria for Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *207*(1), 69-76. https://doi.org/10.1164/rccm.202201-0124OC

Pugashetti, J. V., Kim, J. S., Bose, S., Adegunsoye, A., Linderholm, A. L., Chen, C. H., Strek, M. E., Flaherty, K. R., Murray, S., Newton, C. A., Alqalyoobi, S., Ma, S. F., Mychaleckyj, J. C., Bowler, R. P., Han, M. K., Curtis, J. L., Martinez, F. J., Smith, J. A., Noth, I., & Oldham, J. M. (2024). Biological Age, Chronological Age and Survival in Pulmonary Fibrosis: A Causal Mediation Analysis. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202310-1887OC

Raghu, G. (2017). Idiopathic pulmonary fibrosis: lessons from clinical trials over the past 25 years. *Eur Respir J*, *50*(4). https://doi.org/10.1183/13993003.01209-2017

Raghu, G., Chen, S. Y., Yeh, W. S., Maroni, B., Li, Q., Lee, Y. C., & Collard, H. R. (2014). Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001-11. *Lancet Respir Med*, *2*(7), 566-572. https://doi.org/10.1016/S2213-2600(14)70101-8

Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., . . . Schunemann, H. J. (2011). An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management [Practice Guideline

Research Support, Non-U.S. Gov't]. *Am J Respir Crit Care Med*, *183*(6), 788-824. https://doi.org/10.1164/rccm.2009-040GL

Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., . . . Fibrosis, A. E. J. A. C. o. I. P. (2011). An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*, *183*(6), 788-824. https://doi.org/10.1164/rccm.2009-040GL

Raghu, G., Remy-Jardin, M., Richeldi, L., Thomson, C. C., Inoue, Y., Johkoh, T., Kreuter, M., Lynch, D. A., Maher, T. M., Martinez, F. J., Molina-Molina, M., Myers, J. L., Nicholson, A. G., Ryerson, C. J., Strek, M. E., Troy, L. K., Wijsenbeek, M., Mammen, M. J., Hossain, T., . . . Wilson, K. C. (2022). Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*, *205*(9), e18-e47. https://doi.org/10.1164/rccm.202202-0399ST

Raghu, G., Richeldi, L., Jagerschmidt, A., Martin, V., Subramaniam, A., Ozoux, M. L., Esperet, C. A., & Soubrane, C. (2018). Idiopathic Pulmonary Fibrosis: Prospective, Case-Controlled Study of Natural History and Circulating Biomarkers. *Chest*, *154*(6), 1359-1370. https://doi.org/10.1016/j.chest.2018.08.1083

Raghu, G., Weycker, D., Edelsberg, J., Bradford, W. Z., & Oster, G. (2006). Incidence and prevalence of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*, *174*(7), 810-816. https://doi.org/10.1164/rccm.200602-163OC

Reiseter, S., Gunnarsson, R., Mogens Aalokken, T., Lund, M. B., Mynarek, G., Corander, J., Haydon, J., & Molberg, O. (2018). Progression and mortality of interstitial lung disease in mixed connective tissue disease: a long-term observational nationwide cohort study. *Rheumatology (Oxford)*, *57*(2), 255-262. https://doi.org/10.1093/rheumatology/kex077

Richeldi, L., du Bois, R. M., Raghu, G., Azuma, A., Brown, K. K., Costabel, U., Cottin, V., Flaherty, K. R., Hansell, D. M., Inoue, Y., Kim, D. S., Kolb, M., Nicholson, A. G., Noble, P. W., Selman, M., Taniguchi, H., Brun, M., Le Maulf, F., Girard, M., . . . Investigators, I. T. (2014). Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*, *370*(22), 2071-2082. https://doi.org/10.1056/NEJMoa1402584

Riddell, P., Kleinerova, J., Eaton, D., Healy, D. G., Javadpour, H., McCarthy, J. F., Nolke, L., Redmond, K. C., & Egan, J. J. (2020). Meaningful survival benefit for single lung transplantation in idiopathic pulmonary fibrosis patients over 65 years of age. *Eur Respir J*, *56*(1). https://doi.org/10.1183/13993003.02413-2019

Royston, P., & Parmar, M. K. B. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology*, *13*(1), 152. https://doi.org/10.1186/1471-2288-13-152

Ruwanpura, S. M., Thomas, B. J., & Bardin, P. G. (2020). Pirfenidone: Molecular Mechanisms and Potential Clinical Applications in Lung Disease. *Am J Respir Cell Mol Biol*, *62*(4), 413-422. https://doi.org/10.1165/rcmb.2019-0328TR

Ryerson, C. J., Corte, T. J., Lee, J. S., Richeldi, L., Walsh, S. L. F., Myers, J. L., Behr, J., Cottin, V., Danoff, S. K., Flaherty, K. R., Lederer, D. J., Lynch, D. A., Martinez, F. J., Raghu, G., Travis, W. D., Udwadia, Z., Wells, A. U., & Collard, H. R. (2017). A Standardized Diagnostic Ontology for Fibrotic Interstitial Lung Disease. An International Working Group Perspective. *Am J Respir Crit Care Med*, *196*(10), 1249-1254. https://doi.org/10.1164/rccm.201702-0400PP

Ryerson, C. J., Urbania, T. H., Richeldi, L., Mooney, J. J., Lee, J. S., Jones, K. D., Elicker, B. M., Koth, L. L., King, T. E., Jr., Wolters, P. J., & Collard, H. R. (2013). Prevalence and prognosis of unclassifiable interstitial lung disease. *Eur Respir J*, *42*(3), 750-757. https://doi.org/10.1183/09031936.00131912

Ryerson, C. J., Vittinghoff, E., Ley, B., Lee, J. S., Mooney, J. J., Jones, K. D., Elicker, B. M., Wolters, P. J., Koth, L. L., King, T. E., Jr., & Collard, H. R. (2014). Predicting survival across chronic interstitial lung disease: the ILD-GAP model. *Chest*, *145*(4), 723-728. https://doi.org/10.1378/chest.13-1474

Safiri, S., Karamzad, N., Singh, K., Carson-Chahhoud, K., Adams, C., Nejadghaderi, S. A., Almasi-Hashiani, A., Sullman, M. J. M., Mansournia, M. A., Bragazzi, N. L., Kaufman, J. S., Collins, G. S., & Kolahi, A.-A. (2021). Burden of ischemic heart disease and its attributable risk factors in 204 countries and territories, 1990–2019. *European Journal of Preventive Cardiology*, *29*(2), 420-431. https://doi.org/10.1093/eurjpc/zwab213

Saha, S. K., Yin, Y., Chae, H. S., & Cho, S. G. (2019). Opposing Regulation of Cancer Properties via KRT19-Mediated Differential Modulation of Wnt/β-Catenin/Notch Signaling in Breast and Colon Cancers. *Cancers (Basel)*, *11*(1). https://doi.org/10.3390/cancers11010099

Saini, G., Porte, J., Weinreb, P. H., Violette, S. M., Wallace, W. A., McKeever, T. M., & Jenkins, G. (2015). αvβ6 integrin may be a potential prognostic biomarker in interstitial lung disease. *Eur Respir J*, *46*(2), 486-494. https://doi.org/10.1183/09031936.00210414

Salisbury, M. L., Gross, B. H., Chughtai, A., Sayyouh, M., Kazerooni, E. A., Bartholmai, B. J., Xia, M., Murray, S., Myers, J. L., Lagstein, A., Konopka, K. E., Belloli, E. A., Sheth, J. S., White, E. S., Holtze, C., Martinez, F. J., & Flaherty, K. R. (2018). Development and validation of a radiological diagnosis model for hypersensitivity pneumonitis. *Eur Respir J*, *52*(2). https://doi.org/10.1183/13993003.00443-2018

Salisbury, M. L., Gu, T., Murray, S., Gross, B. H., Chughtai, A., Sayyouh, M., Kazerooni, E. A., Myers, J. L., Lagstein, A., Konopka, K. E., Belloli, E. A., Sheth, J. S., White, E. S., Holtze, C., Martinez, F. J., & Flaherty, K. R. (2019). Hypersensitivity Pneumonitis: Radiologic Phenotypes Are Associated With Distinct Survival Time and Pulmonary Function Trajectory. *Chest*, *155*(4), 699-711. https://doi.org/10.1016/j.chest.2018.08.1076

Sanders, J. L., Putman, R. K., Dupuis, J., Xu, H., Murabito, J. M., Araki, T., Nishino, M., Benjamin, E. J., Levy, D., Ramachandran, V. S., Washko, G. R., Curtis, J. L., Freeman, C. M., Bowler, R. P., Hatabu, H., O'Connor, G. T., & Hunninghake, G. M. (2021). The Association of Aging Biomarkers, Interstitial Lung Abnormalities, and Mortality. *Am J Respir Crit Care Med*, *203*(9), 1149-1157. https://doi.org/10.1164/rccm.202007-2993OC

Schanin, J., Gebremeskel, S., Korver, W., Falahati, R., Butuci, M., Haw, T. J., Nair, P. M., Liu, G., Hansbro, N. G., Hansbro, P. M., Evensen, E., Brock, E. C., Xu, A., Wong, A., Leung, J., Bebbington, C., Tomasevic, N., & Youngblood, B. A. (2021). A monoclonal antibody to Siglec-8 suppresses non-allergic airway inflammation and inhibits IgE-independent mast cell activation. *Mucosal Immunol*, *14*(2), 366-376. https://doi.org/10.1038/s41385-020-00336-9

Schiller, H. B., Mayr, C. H., Leuschner, G., Strunz, M., Staab-Weijnitz, C., Preisendörfer, S., Eckes, B., Moinzadeh, P., Krieg, T., Schwartz, D. A., Hatz, R. A., Behr, J., Mann, M., & Eickelberg, O. (2017). Deep Proteome Profiling Reveals Common Prevalence of MZB1-Positive Plasma B Cells in Human Lung and Skin Fibrosis. *Am J Respir Crit Care Med*, *196*(10), 1298-1310. https://doi.org/10.1164/rccm.201611-2263OC

Selman, M., Montano, M., Ramos, C., Chapela, R., Gonzalez, G., & Vadillo, F. (1988). Lung collagen metabolism and the clinical course of hypersensitivity pneumonitis. *Chest*, *94*(2), 347-353. https://doi.org/10.1378/chest.94.2.347

Selman, M., Pardo, A., & King, T. E., Jr. (2012). Hypersensitivity pneumonitis: insights in diagnosis and pathobiology. *Am J Respir Crit Care Med*, *186*(4), 314-324. https://doi.org/10.1164/rccm.201203-0513CI

Sharma, N., Putman, M. S., Vij, R., Strek, M. E., & Dua, A. (2017). Myositis-associated Interstitial Lung Disease: Predictors of Failure of Conventional Treatment and Response to Tacrolimus in a US Cohort. *J Rheumatol*, *44*(11), 1612-1618. https://doi.org/10.3899/jrheum.161217

Stuart, B. D., Lee, J. S., Kozlitina, J., Noth, I., Devine, M. S., Glazer, C. S., Torres, F., Kaza, V., Girod, C. E., Jones, K. D., Elicker, B. M., Ma, S. F., Vij, R., Collard, H. R., Wolters, P. J., & Garcia, C. K. (2014). Effect of telomere length on survival in patients with idiopathic

pulmonary fibrosis: an observational cohort study with independent validation. *Lancet Respir Med*, *2*(7), 557-565. https://doi.org/10.1016/S2213-2600(14)70124-9

Suissa, S., & Suissa, K. (2022). Antifibrotics and Reduced Mortality in Idiopathic Pulmonary Fibrosis: Immortal Time Bias. *Am J Respir Crit Care Med*. https://doi.org/10.1164/rccm.202207-1301LE

Swigris, J. J., & Brown, K. K. (2010). The role of endothelin-1 in the pathogenesis of idiopathic pulmonary fibrosis. *BioDrugs*, *24*(1), 49-54. https://doi.org/10.2165/11319550-000000000-00000

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, *47*(D1), D607-d613. https://doi.org/10.1093/nar/gky1131

Tailor, P. D., Kodeboyina, S. K., Bai, S., Patel, N., Sharma, S., Ratnani, A., Copland, J. A., She, J. X., & Sharma, A. (2018). Diagnostic and prognostic biomarker potential of kallikrein family genes in different cancer types. *Oncotarget*, *9*(25), 17876-17888. https://doi.org/10.18632/oncotarget.24947

Tashkin, D. P., Elashoff, R., Clements, P. J., Goldin, J., Roth, M. D., Furst, D. E., Arriola, E., Silver, R., Strange, C., Bolster, M., Seibold, J. R., Riley, D. J., Hsu, V. M., Varga, J., Schraufnagel, D. E., Theodore, A., Simms, R., Wise, R., Wigley, F., . . . Scleroderma Lung Study Research, G. (2006). Cyclophosphamide versus placebo in scleroderma lung disease. *N Engl J Med*, *354*(25), 2655-2666. https://doi.org/10.1056/NEJMoa055120

Tashkin, D. P., Roth, M. D., Clements, P. J., Furst, D. E., Khanna, D., Kleerup, E. C., Goldin, J., Arriola, E., Volkmann, E. R., Kafaja, S., Silver, R., Steen, V., Strange, C., Wise, R., Wigley, F., Mayes, M., Riley, D. J., Hussain, S., Assassi, S., . . . Sclerodema Lung Study, I. I. I. (2016). Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir Med*, *4*(9), 708-719. https://doi.org/10.1016/S2213-2600(16)30152-7

Thenappan, A., Maher, T. M., Yazbeck, L., Jenkins, R. G., Johnson, S. R., Stewart, I., Oldham, J. M., & Molyneaux, P. L. (2024). Competing Causes of Death in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*, *210*(7), 938-940. https://doi.org/10.1164/rccm.202403-0595RL

Travis, W. D., Costabel, U., Hansell, D. M., King, T. E., Jr., Lynch, D. A., Nicholson, A. G., Ryerson, C. J., Ryu, J. H., Selman, M., Wells, A. U., Behr, J., Bouros, D., Brown, K. K., Colby, T. V., Collard, H. R., Cordeiro, C. R., Cottin, V., Crestani, B., Drent, M., . . . Pneumonias, A. E. C. o. I. I. (2013). An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*, *188*(6), 733-748. https://doi.org/10.1164/rccm.201308-1483ST

Travis, W. D., Hunninghake, G., King, T. E., Jr., Lynch, D. A., Colby, T. V., Galvin, J. R., Brown, K. K., Chung, M. P., Cordier, J. F., du Bois, R. M., Flaherty, K. R., Franks, T. J., Hansell, D. M., Hartman, T. E., Kazerooni, E. A., Kim, D. S., Kitaichi, M., Koyama, T., Martinez, F. J., . . . Wells, A. (2008). Idiopathic nonspecific interstitial pneumonia: report of an American Thoracic Society project. *Am J Respir Crit Care Med*, *177*(12), 1338-1347. https://doi.org/10.1164/rccm.200611-1685OC

van der Sar, I. G., Wijsenbeek, M. S., Braunstahl, G. J., Loekabino, J. O., Dingemans, A. C., In 't Veen, J., & Moor, C. C. (2023). Differentiating interstitial lung diseases from other respiratory diseases using electronic nose technology. *Respir Res*, *24*(1), 271. https://doi.org/10.1186/s12931-023-02575-3

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health*, *37*, 17-32. https://doi.org/10.1146/annurev-publhealth-032315-021402

VanderWeele, T. J., & Ding, P. (2017). Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*, *167*(4), 268-274. https://doi.org/10.7326/m16-2607

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)*, *6*(2). https://doi.org/10.4172/2161-1165.1000227

Vock, D. M., Durheim, M. T., Tsuang, W. M., Finlen Copeland, C. A., Tsiatis, A. A., Davidian, M., Neely, M. L., Lederer, D. J., & Palmer, S. M. (2017). Survival Benefit of Lung Transplantation in the Modern Era of Lung Allocation. *Ann Am Thorac Soc*, *14*(2), 172-181. https://doi.org/10.1513/AnnalsATS.201606-507OC

Vu Pugashetti, J., Kim, J. S., Combs, M. P., Ma, S. F., Adegunsoye, A., Linderholm, A. L., Strek, M. E., Chen, C. H., Dilling, D. F., Whelan, T. P. M., Flaherty, K. R., Martinez, F. J., Noth, I., & Oldham, J. M. (2024). A Multi-dimensional Classifier to Support Lung Transplant Referral in Patients with Pulmonary Fibrosis. *J Heart Lung Transplant*. https://doi.org/10.1016/j.healun.2024.03.018

Vu Pugashetti, J., Newton, C. A., Molyneaux, P. L., & Oldham, J. M. (2023). Reply to Noboa-Sevilla et al. *Am J Respir Crit Care Med*, *207*(3), 369-370. https://doi.org/10.1164/rccm.202209-1807LE

Walker, U. A., Tyndall, A., Czirjak, L., Denton, C., Farge-Bancel, D., Kowal-Bielecka, O., Muller-Ladner, U., Bocelli-Tyndall, C., & Matucci-Cerinic, M. (2007). Clinical risk assessment of organ manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials And Research group database. *Ann Rheum Dis*, *66*(6), 754-763. https://doi.org/10.1136/ard.2006.062901

Wang, B. R., Edwards, R., Freiheit, E. A., Ma, Y., Burg, C., de Andrade, J., Lancaster, L., Lindell, K., Nathan, S. D., Raghu, G., Gibson, K., Gulati, M., Mason, W., Noth, I., Schmidt, B., Spino, C., Staszak, S., Stauffer, J., Wolters, P. J., . . . Flaherty, K. R. (2020). The Pulmonary Fibrosis Foundation Patient Registry. Rationale, Design, and Methods. *Ann Am Thorac Soc*, *17*(12), 1620-1628. https://doi.org/10.1513/AnnalsATS.202001-035SD

Wang, Q. R., Liu, S. S., Min, J. L., Yin, M., Zhang, Y., Zhang, Y., Tang, X. N., Li, X., & Liu, S. S. (2023). CCL17 drives fibroblast activation in the progression of pulmonary fibrosis by enhancing the TGF-β/Smad signaling. *Biochem Pharmacol*, *210*, 115475. https://doi.org/10.1016/j.bcp.2023.115475

Wells, A. U., Desai, S. R., Rubens, M. B., Goh, N. S., Cramer, D., Nicholson, A. G., Colby, T. V., du Bois, R. M., & Hansell, D. M. (2003). Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med*, *167*(7), 962-969. https://doi.org/10.1164/rccm.2111053

Wells, A. U., Flaherty, K. R., Brown, K. K., Inoue, Y., Devaraj, A., Richeldi, L., Moua, T., Crestani, B., Wuyts, W. A., Stowasser, S., Quaresma, M., Goeldner, R. G., Schlenker-Herceg, R., Kolb, M., & investigators, I. t. (2020). Nintedanib in patients with progressive fibrosing interstitial lung diseases-subgroup analyses by interstitial lung disease diagnosis

in the INBUILD trial: a randomised, double-blind, placebo-controlled, parallel-group trial. *Lancet Respir Med*. https://doi.org/10.1016/S2213-2600(20)30036-9

Wijsenbeek, M., & Cottin, V. (2020). Spectrum of Fibrotic Lung Diseases. *N Engl J Med*, *383*(10), 958-968. https://doi.org/10.1056/NEJMra2005230

Wijsenbeek, M., Suzuki, A., & Maher, T. M. (2022). Interstitial lung diseases. *Lancet*, *400*(10354), 769-786. https://doi.org/10.1016/s0140-6736(22)01052-2

Witt, L. J., Demchuk, C., Curran, J. J., & Strek, M. E. (2016). Benefit of adjunctive tacrolimus in connective tissue disease-interstitial lung disease. *Pulm Pharmacol Ther*, *36*, 46-52. https://doi.org/10.1016/j.pupt.2015.12.004

Wollin, L., Wex, E., Pautsch, A., Schnapp, G., Hostettler, K. E., Stowasser, S., & Kolb, M. (2015). Mode of action of nintedanib in the treatment of idiopathic pulmonary fibrosis. *Eur Respir J*, *45*(5), 1434-1445. https://doi.org/10.1183/09031936.00174914

Wu, M., Wang, Z., Shi, X., Zan, D., Chen, H., Yang, S., Ding, F., Yang, L., Tan, P., Ma, R. Z., Wang, J., Ma, L., Ma, Y., & Jin, J. (2023). TGFβ1-RCN3-TGFBR1 loop facilitates pulmonary fibrosis by orchestrating fibroblast activation. *Respir Res*, *24*(1), 222. https://doi.org/10.1186/s12931-023-02533-z

Xiao, Y., Liu, J., Peng, Y., Xiong, X., Huang, L., Yang, H., Zhang, J., & Tao, L. (2016). GSTA3 Attenuates Renal Interstitial Fibrosis by Inhibiting TGF-Beta-Induced Tubular Epithelial-Mesenchymal Transition and Fibronectin Expression. *PLoS One*, *11*(9), e0160855. https://doi.org/10.1371/journal.pone.0160855

Xin, W., Zhang, M., Yu, Y., Li, S., Ma, C., Zhang, J., Jiang, Y., Li, Y., Zheng, X., Zhang, L., Zhao, X., Pei, X., & Zhu, D. (2020). BCAT1 binds the RNA-binding protein ZNF423 to activate autophagy via the IRE1-XBP-1-RIDD axis in hypoxic PASMCs. *Cell Death Dis*, *11*(9), 764. https://doi.org/10.1038/s41419-020-02930-y

Yang, L., Zhang, X., & Chen, J. (2024). Winsorization greatly reduces false positives by popular differential expression methods when analyzing human population samples. *Genome Biol*, *25*(1), 282. https://doi.org/10.1186/s13059-024-03230-w

Zhang, J., Liu, G., Liu, Y., Yang, P., Xie, J., & Wei, X. (2023). The biological functions and related signaling pathways of SPON2. *Front Oncol*, *13*, 1323744. https://doi.org/10.3389/fonc.2023.1323744

Zhao, R., Wang, Z., Wang, G., Geng, J., Wu, H., Liu, X., Bin, E., Sui, J., Dai, H., & Tang, N. (2024). Sustained amphiregulin expression in intermediate alveolar stem cells drives progressive fibrosis. *Cell Stem Cell*, *31*(9), 1344-1358.e1346. https://doi.org/10.1016/j.stem.2024.07.004

Zhou, J., Cui, S., He, Q., Guo, Y., Pan, X., Zhang, P., Huang, N., Ge, C., Wang, G., Gonzalez, F. J., Wang, H., & Hao, H. (2020). SUMOylation inhibitors synergize with FXR agonists in combating liver fibrosis. *Nat Commun*, *11*(1), 240. https://doi.org/10.1038/s41467-019-14138-6

Zong, L., Chen, P., & Wang, D. X. (2014). Death decoy receptor overexpression and increased malignancy risk in colorectal cancer. *World J Gastroenterol*, *20*(15), 4440-4445. https://doi.org/10.3748/wjg.v20.i15.4440

Zou, M., Zou, J., Hu, X., Zheng, W., Zhang, M., & Cheng, Z. (2021). Latent Transforming Growth Factor-β Binding Protein-2 Regulates Lung Fibroblast-to-Myofibroblast Differentiation in Pulmonary Fibrosis via NF-κB Signaling. *Front Pharmacol*, *12*, 788714. https://doi.org/10.3389/fphar.2021.788714

Zou, R. H., Wallace, W. D., Nouraie, S. M., Chan, S. Y., & Risbano, M. G. (2020). Lower DLco% identifies exercise pulmonary hypertension in patients with parenchymal lung disease referred for dyspnea. *Pulm Circ*, *10*(1), 2045894019891912. https://doi.org/10.1177/2045894019891912