THE UNIVERSITY OF CHICAGO

RACE, TELOMERES, AND GENOMIC INSIGHTS INTO PULMONARY FIBROSIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON GENETICS, GENOMICS, AND SYSTEMS BIOLOGY

BY

AYODEJI OLAKUNLE ADEGUNSOYE

CHICAGO, ILLINOIS

AUGUST 2023

Dedication

To my family and my patients.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

5’ UTR	5’ Untranslated Region
AIM-ILD	Autoimmune myositis-associated interstitial lung disease
BMF	Bone marrow failure
BMI	Body mass index
CHP	Chronic hypersensitivity pneumonitis
CNV	Copy Number Variant
CTD-ILD	Connective tissue disease associated-ILD
DLCO	Diffusing capacity of the lungs for carbon monoxide
dsDNA	Double-stranded DNA
EMV	External multicenter validation
eQTL	Expression quantitative trait loci
fHP	Fibrotic hypersensitivity pneumonitis
FIP	Familial interstitial pneumonia
FPF	Familial pulmonary fibrosis
FVC	Forced vital capacity
GAP	Gender-Age-Physiology Index
GATK	Genome Analysis Toolkit
gDNA	Genomic deoxyribonucleic acid
gRNA	Guide RNA

GWAS	Genome-wide association study
HLA	Human leukocyte antigen
HP	Hypersensitivity pneumonitis
HR	Hazard ratio
IIP	Idiopathic interstitial pneumonia
ILA	Interstitial lung abnormality
ILD	Interstitial lung disease
IPAF	Interstitial pneumonia with autoimmune features
IPF	Idiopathic pulmonary fibrosis
iPSC	Induced pluripotent stem cell
LCLs	Lymphoblastoid cell lines
LTL	Leukocyte telomere length
MAF	Minor allele frequency
NA	Not applicable
NAC	N-acetylcysteine
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PF	Pulmonary fibrosis
PFFR	Pulmonary Fibrosis Foundation Registry
RA	Rheumatoid arthritis
RB-ILD	Respiratory bronchiolitis-associated ILD

SNP	Single nucleotide polymorphism
SSc-ILD	Systemic sclerosis-associated interstitial lung disease
TBD	Telomere biology disorders
TL	Telomere length
TNF	Tumor necrosis factor
UIP	Usual interstitial pneumonia
UTR	Untranslated region.
WES	Whole exome sequencing
WGS	Whole genome sequencing

Acknowledgements

Foremost, my deepest appreciation extends to my advisor, Dr. Yoav Gilad, whose unrelenting mentorship has guided me through my training and transition into a faculty role at UChicago. The multiple doors his support has unlocked for me have been instrumental in my journey, and I have striven to seize each opportunity with unyielding dedication. Dr. Gilad's incomparable expertise and his audacious challenges of conventional paradigms in genetics and genomics have profoundly shaped my scientific approach and professional development.

My heartfelt gratitude also goes to my patients and global colleagues, whose generous donations of samples, time, patience, and intellectual capital have propelled this research forward. The critical perspective of my colleagues and inquisitiveness of our patients illuminates the chasms in our knowledge, keeping our focus firmly on the research that carries meaningful implications for their lives and those of their families.

I am indebted to my thesis committee: Drs. Marcelo Nobrega, Yang Li, Gokhan Mutlu, and Mary Strek. Their generosity of time and thoughtful counsel during this journey have been invaluable. I especially cherish the foundational education imparted by Dr. Jonathan Burnett, who tutored me on the generation, maintenance, and characterization of patient-derived induced pluripotent stem cells, and to Dr. Brandon Pierce, for his valuable tutelage on telomere biology and epidemiology. This knowledge has laid the groundwork for the inception of my research program, informing future endeavors and sparking fruitful collaborations beyond the boundaries of Chicago. My deepest thanks

also to Dr. Anne Sperling, a mentor whose astute advice at various stages of my early career has been instrumental.

A profound sense of gratitude is owed to my external mentors, notably Drs. Imre Noth, Matthew Churpek, and Mary Armanios. Their generosity and collaborative spirit has provided me with invaluable knowledge that will resonate throughout my scientific journey.

I also extend my thanks to the wonderful members of the Interstitial Lung Disease group at the University of Chicago, particularly Dr. Mary Strek, whose extraordinary mentorship over the past decade has molded and refined me in ways that I am eternally grateful for, to Drs., Aliya Husain, Rekha Vij, Renea Jabolonski, Iazsmin Ventura, Katie Lee, Kavitha Selvan, and Ms. Spring Maleckar. Their impact and attention to detail in improving the care of patients with pulmonary fibrosis is unquantifiable and continues to influence my work.

My collaborations and friendships with esteemed colleagues dedicated to translational work in the domain of pulmonary fibrosis has been a source of enrichment and personal growth. This includes Drs. Justin Oldham, Chad Newton, Brett Ley, Jonathan Chung, John Kim, Stephen Humphries, Richard Allen, Simone Feurstein, Robert Guzy, Jane Churpek. Our collective resilience during the COVID-19 pandemic has ensured the continuation of our work, underlining the robustness and adaptability of our partnerships.

The Gilad group, both current and past members, deserve special thanks. Each of you, including Dr. Brianna Mittleman, Dr. Anthony Hung, Austin Reilly, Dr. Ben Umans, Dr. Benjamin Fair, Claudia Chavarria, Duhman Mutaseb, Wenhe Lin, Erik McIntre, Dr. Genevieve Housman, Dr. Jonathan Burnett, Dr. Kenneth Barr, Dr. Mellissa DeMille, Dr. Katie Rhodes, Dr. Natalia Gonzales, Dr. Olivia Allen, and Jasmine Aylward have been indispensable companions during my training.

My deepest appreciation goes out to my friends, colleagues, and classmates within the ecosystem of Genetics, Genomics and System Biology. I am deeply appreciative to Dr. Gokhan Mutlu for championing my idea of pursuing rigorous doctoral training during my clinical instructorship in Pulmonary and Critical Care medicine. Dr. Michael Drazer deserves a special place of gratitude for his inspiration, guidance and peer mentorship along the path of this rigorous training. Equally, my profound gratitude extends to Sue Levison, Tamiko Charley, and Candice Lewis, who have exhibited exceptional adaptability in the face of the administrative complexities arising from a doctoral candidate who is concurrently a physician. I extend my heartfelt gratitude to my esteemed colleagues in the Section of Pulmonary and Critical Care Medicine for their unwavering support, constructive feedback, and invaluable contributions toward the research endeavors that culminated in this dissertation.

I am forever thankful to the foundations that support physician-scientists, particularly those at the dawn of our careers. They are acutely aware of the unique sacrifices demanded by this work and are indispensable in mitigating some of the

challenges we face. I have reaped extensive benefits from the Pulmonary Fibrosis Foundation, the American College of Chest Physicians, the Institute for Translational Medicine, the Patient-Centered Outcomes Research Institute, the American Society for Clinical Investigation, the University of Chicago Interstitial Lung Disease center, and the National Institutes of Health.

A special word of thanks goes out to Dr. Mengjie Chen, for her phenomenal computational skills, patience and support through my multiple genomic projects, and to Dr. Yildiz Koca, the first member of my research group, whose adaptability, patience, and resilience in mastering new techniques in genomic modeling, bioinformatics, and predictive algorithms have been extraordinary, particularly during the challenges imposed by the COVID-19 pandemic. My heartfelt gratitude is also due to the trainees who have spent time in my group during the early stages of my faculty position, including Nicole Garcia, Karina Mak, Tinuola Owolabi, Muwafaq (Moe) Ibrahim, Tinuola Owolabi, Dr. Dylan Douglas, Dr. Layne Keating, Dr. Christopher Carlton, Shaili Smith, and Rachel Hodge.

Finally, no words can fully convey the depth of my gratitude for Adekonla, Tobi, Toni, and Tade. Each of you understands the sacrifices we have collectively made to enable me to tread this path of a physician-scientist, and for that, my gratitude is boundless.

Abstract

Pulmonary fibrosis pathophysiology is increasingly recognized to be influenced by the biology of telomeres, a field that has now grown in prominence with the widespread adoption of affordable next-generation sequencing over the past decade, leading to an improved understanding of telomere-related mechanisms underlying pulmonary fibrosis. Despite these advancements, notable gaps in knowledge persist. Firstly, individuals with telomere biology disorders or shortening face a heightened risk of developing conditions such as fibrotic interstitial lung diseases, but the genetic and molecular mechanisms catalyzing this process are yet to be fully elucidated. Secondly, the field lacks comprehensive natural history studies focusing on telomere-related mechanisms underlying the progression of pulmonary fibrosis to death or lung transplantation across diverse racial groups. Thirdly, the current ability to model telomere-related pulmonary fibrosis is constrained due to the absence of high-fidelity models. Lastly, and perhaps most significantly, there is an evident paucity in our understanding of the implications of racial diversity in genetic and genomic research, especially as it impacts disease progression in pulmonary fibrosis. Numerous individuals displaying pulmonary fibrosis phenotypes post-genomic sequencing do not exhibit commonly identifiable genetic markers, suggesting the potential discovery of novel, racially diverse telomere-related risk loci for pulmonary fibrosis syndromes is on the horizon.

In my doctoral research, I have addressed pivotal knowledge gaps by conducting an unprecedented large-scale natural history study, exploring the progression of pulmonary fibrosis across diverse racial groups, leading to either death or lung transplantation. I collaborated with several institutions to investigate telomere length's predictive capacity in racially diverse pulmonary fibrosis cohorts, uncovering stark racial disparities, especially among Black patients. I found that peripheral leukocyte telomere length consistently correlates with chronological age and serves as a predictive mortality biomarker in pulmonary fibrosis across all racial groups. Further, I delved into the nuances of telomere biology in peripheral blood mononuclear cell subsets and induced pluripotent stem cells (iPSCs) derived from multiracial cohorts of patients with pulmonary fibrosis. My research revealed the prognostic impact of telomere length in patients with fibrotic hypersensitivity pneumonitis receiving immunomodulatory therapy, marking a pioneering study in this domain. My findings identified significant mutations in eight genes—PDE4DIPP, ZNF683, SFRP5, MIR6077, RPSAP72, WASIR2, GAPDHP27, and CNTNAP3P2—as critical drivers of honeycomb fibrosis in multiethnic populations with pulmonary fibrosis. This study unveiled unique transcriptomic differences in key variants across Black and White patients. Through a comprehensive integration of whole genome sequencing and expression quantitative trait loci analyses, I have revealed striking differences between self-identified race and genetic ancestry and uncovered novel host defense and cell senescence gene variants contributing to disease risk across a spectrum of pulmonary fibrosis subtypes. In sum, my findings regarding telomere biology and genomic risk variants in pulmonary fibrosis have yielded new insights into disease

progression mechanisms, enhanced theragnostic modeling, fostered international and institutional data sharing and collaboration, and generated patient-derived cellular models for future research into disease mechanisms and innovative therapeutic strategies specific to pulmonary fibrosis.

Chapter I

Introduction

The majority of this chapter is adapted from several foundational clinical studies in pulmonary fibrosis that I led, and review articles detailing the role of genetics and genomics in pulmonary fibrosis and iPSC-derived genomic models of pulmonary disease phenotypes that I have written in collaboration with Drs. Yoav Gilad, Natalia Gonzales, Rekha Vij, and Imre Noth. These articles have been published in *JAMA Network Open, European Respiratory Journal, Chest, Annals of American Thoracic Society,* and accepted for publication in the *Annual Review of Genetics.*

Adegunsoye A, Freiheit E, White EN, Kaul B, Newton CA, Oldham JM, Lee CT, Chung J, Garcia N, Ghodrati S, Vij R, Jablonski R, Flaherty KR, Wolters PJ, Garcia CK, Strek ME. Evaluation of Pulmonary Fibrosis Outcomes by Race and Ethnicity in US Adults. *JAMA Netw Open*. 2023 Mar 1;6(3): e232427. PMID: 36897590.

Newton CA, Oldham JM, Ley B, Anand V, **Adegunsoye A**, Liu G, Batra K, Torrealba J, Kozlitina J, Glazer C, Strek ME, Wolters PJ, Noth I, Garcia CK. Telomere Length and Genetic Variant Associations with Interstitial Lung Disease Progression and Survival*. Eur Respir J.* 2019 Jan 11. PMID: 30635297.

Adegunsoye A, Oldham JM, Bellam SK, Montner S, Churpek MM, Noth I, Vij R, Strek ME, Chung JH. CT Honeycombing Identifies a Progressive Fibrotic Phenotype with Increased Mortality Across Diverse Interstitial Lung Diseases. *Ann Am Thorac Soc.* 2019 Jan 17. PMID: 30653927

Adegunsoye A, Vij R, Noth I. Integrating Genomics into Management of Fibrotic Interstitial Lung Disease. *CHEST* 2019 Jan 17. pii: S0012-3692(19)30001-7. PMID: 30660786

Adegunsoye A, Gonzales NM, Gilad Y. Induced Pluripotent Stem Cells in Disease Biology and the Evidence for Their In Vitro Utility. *Annu. Rev. Genet*. 2023;57. doi.org/10.1146/annurev-genet-022123-090319

I have adapted tables, figures, and texts from several of these articles and reviews for this chapter. Additionally, I have also adapted other tables, figures, and texts from published papers for which I served as an author or co-author:

(i) Data presented in the section of this chapter, "Pulmonary fibrosis outcomes by race and ethnicity" have been published in: **Adegunsoye A**, Oldham JM, Bellam SK, Chung JH, Chung PA, Biblowitz KM, Montner SM, Lee C, Hsu S, Husain AN, Vij R, Mutlu G, Noth I, Churpek MM, Strek ME. African-American Race and Mortality in Interstitial Lung Disease: A Multicenter Propensity-Matched Analysis. *Eur Respir J.* 2018. PMID: 29724923.

(ii) Data presented in the section of this chapter, "Telomere length and gene variant association with survival in pulmonary fibrosis" have been published in:

Allen RJ, Oldham JM, Jenkins DA, Leavy OC, Guillen-Guio B, Melbourne CA, Ma SF, Jou J, Kim JS; CleanUP-IPF Investigators of the Pulmonary Trials Cooperative; Fahy WA, Oballa E, Hubbard RB, Navaratnam V, Braybrooke R, Saini G, Roach KM, Tobin MD, Hirani N, Whyte MKB, Kaminski N, Zhang Y, Martinez FJ, Linderholm AL,

Adegunsoye A, Strek ME, Maher TM, Molyneaux PL, Flores C, Noth I, Gisli Jenkins R, Wain LV. Longitudinal lung function and gas transfer in individuals with idiopathic pulmonary fibrosis: a genome-wide association study. Lancet Respir Med. 2023 Jan;11(1):65-73. PMID: 35985358; PMCID: PMC10077113.

Kim JS, Manichaikul AW, Hoffman EA, Balte P, Anderson MR, Bernstein EJ, Madahar P, Oelsner EC, Kawut SM, Wysoczanski A, Laine AF, **Adegunsoye A**, Ma JZ, Taub MA, Mathias RA, Rich SS, Rotter JI, Noth I, Garcia CK, Barr RG, Podolanczuk AJ. MUC5B, telomere length and longitudinal quantitative interstitial lung changes: the MESA Lung Study. *Thorax*. 2022 Aug 5:thoraxjnl-2021-218139. PMID: 36690926; PMCID: PMC9899287.

Oldham JM, Allen RJ, Lorenzo-Salazar JM, Molyneaux PL, Ma SF, Joseph C, Kim JS, Guillen-Guio B, Hernández-Beeftink T, Kropski JA, Huang Y, Lee CT, **Adegunsoye A**, Pugashetti JV, Linderholm AL, Vo V, Strek ME, Jou J, Muñoz-Barrera A, Rubio-Rodriguez LA, Hubbard R, Hirani N, Whyte MKB, Hart S, Nicholson AG, Lancaster L, Parfrey H, Rassl D, Wallace W, Valenzi E, Zhang Y, Mychaleckyj J, Stockwell A, Kaminski N, Wolters PJ, Molina-Molina M, Banovich NE, Fahy WA, Martinez FJ, Hall IP, Tobin MD, Maher TM, Blackwell TS, Yaspan BL, Jenkins RG, Flores C, Wain LV, Noth I. PCSK6 and Survival in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2023 Feb 13. PMID: 36780644.

Allen RJ, Stockwell A, Oldham JM, Guillen-Guio B, Schwartz DA, Maher TM, Flores C, Noth I, Yaspan BL, Jenkins RG, Wain LV; **International IPF Genetics Consortium**. Genome-wide association study across five cohorts identifies five novel loci

associated with idiopathic pulmonary fibrosis. *Thorax*. 2022 Aug;77(8):829-833. PMID: 35688625

Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, Kraven LM, Obeidat M, Li X, Ng M, Braybrooke R, Molina-Molina M, Hobbs BD, Putman RK, Sakornsakolpat P, Booth HL, Fahy WA, Hart SP, Hill MR, Hirani N, Hubbard RB, McAnulty RJ, Millar AB, Navaratnam V, Oballa E, Parfrey H, Saini G, Whyte MKB, Zhang Y, Kaminski N, **Adegunsoye A**, Strek ME, Neighbors M, Sheng XR, Gudmundsson G, Gudnason V, Hatabu H, Lederer DJ, Manichaikul A, Newell JD Jr, O'Connor GT, Ortega VE, Xu H, Fingerlin TE, Bossé Y, Hao K, Joubert P, Nickle DC, Sin DD, Timens W, Furniss D, Morris AP, Zondervan KT, Hall IP, Sayers I, Tobin MD, Maher TM, Cho MH, Hunninghake GM, Schwartz DA, Yaspan BL, Molyneaux PL, Flores C, Noth I, Jenkins RG, Wain LV. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. 2020. PMID: 31710517.

Pulmonary fibrosis outcomes by race and ethnicity

The complexities of the genetic landscape within the interstitial lung diseases (ILDs) extends beyond their clinical, radiographic, and pathological characteristics, extending into the realm of racial and ethnic diversity [1-3]. ILDs, representing a diverse collection of diffuse parenchymal lung disorders, often culminate in devastating pulmonary fibrosis, leading to significant impairment of lung function and elevated mortality rates [4]. With the aging population worldwide, the incidence of ILDs and their mortality rates have seen an upward trajectory[5]. The influence of racial disparities on disease risk and mortality, however, remains an underexplored domain.

African-Americans, who constitute the largest minority race in the United States, have been associated with higher rates of respiratory impairment and worsened survival outcomes in numerous pulmonary diseases[6-8]. This population exhibits earlier onset of chronic obstructive pulmonary disease, a steeper decline in lung function, and increased mortality risk associated with lung cancer compared to their European-American counterparts. Similarly, the severity of connective tissue diseases and associated risk for pulmonary involvement is greater among African-Americans[3]. Yet, despite these disparities, the majority of studies characterizing patients with ILD and assessing survival outcomes have been conducted predominantly in White populations. Epidemiological evidence suggests that patients of African descent are less likely to develop idiopathic pulmonary fibrosis (IPF), the most severe form of ILD, but face a younger age of death from IPF compared to other races[9]. As with other pulmonary disorders, earlier onset of ILD and worsened outcomes among African-Americans could necessitate earlier

therapeutic intervention. Independent of access to care, it appears that African-American race is linked with a younger age at ILD diagnosis and an increased mortality rate[10].

Recently, death rates from chronic respiratory diseases have seen a significant upsurge, largely driven by the escalating burden of ILDs, leading to a twofold increase in mortality rates over the past four decades[5]. Pulmonary fibrosis (PF), a form of ILD characterized by lung tissue destruction, accounts for the most substantial increase in mortality rates. This disproportionate impact is primarily due to the increasing burden of ILD on PF-related outcomes such as respiratory-related deaths, disease severity, and an increasingly aging population.

Racial and ethnic minority populations bear the brunt of the health disparities and preexisting socioeconomic inequities, which have a significant impact on morbidity and mortality rates. Black patients have high rates of respiratory impairment, are thrice as likely to die from obstructive lung diseases such as asthma, and may have differential survival outcomes in ILD compared to White individuals[8]. Despite these disparities, our understanding of the relationship between health disparities and racial differences in outcomes among patients with PF is limited due to poor enrollment of racial and ethnic minority individuals in ILD registries and clinical trials. Additionally, our understanding of the age at which clinically relevant outcomes occur in racial and ethnic minority populations with PF is also far from complete. Hence, there is an urgent need to examine the age at PF-related outcomes and the heterogeneity in survival patterns among White, Black, and Hispanic participants with PF in a nationally acquired US registry. The

validation of these findings with data from geographically diverse tertiary care centers with PF expertise is equally crucial.

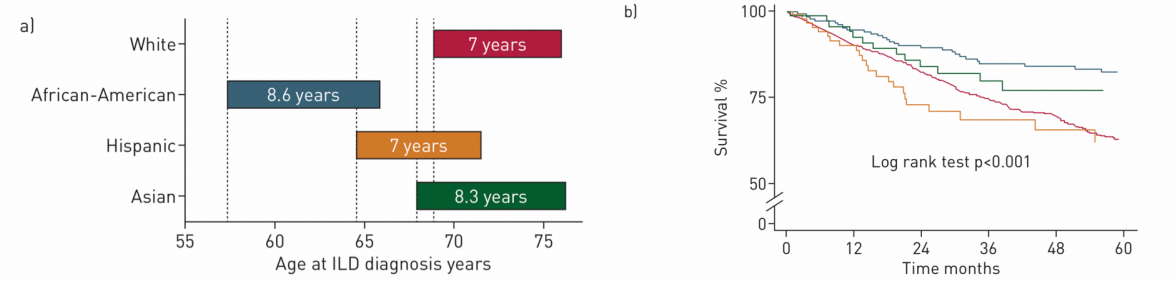
While substantial advancements have been made in understanding the genetic and molecular underpinnings of ILDs, there remains a paucity of studies examining the impact of racial and ethnic diversity in these diseases. More extensive research is needed to close the gap in our knowledge regarding racial disparities in ILD and PF, with a particular focus on the inclusion of minorities in genetic and genomic studies. This will ensure a comprehensive understanding of the disease and inform targeted strategies to address healthcare disparities, ultimately improving patient outcomes across all races and ethnicities.

Black race is associated with younger age at diagnosis of interstitial lung disease

One of the initial retrospective investigations into the influence of race and ethnicity in ILD [9], put forth the intriguing finding that African-Americans were less prone to develop IPF, the most severe form of ILD. However, they encountered a notably earlier age of mortality due to IPF [9]. In the light of these insights, I postulated that beyond IPF, and independent of healthcare access, African-American individuals would likely present with ILD diagnosis at a younger age compared to other racial groups. My expectation, mirroring patterns in other pulmonary disorders, was that a premature onset of ILD and subsequent worsened outcomes in African-Americans could necessitate earlier therapeutic measures. This hypothesis was substantiated by my subsequent research involving a large, extensively characterized multi-center cohort with a range of ILD types,

conclusively demonstrating a correlation between African-American race and an earlier age at diagnosis of ILD (**Figure 1.1**) [3].

Figure 1.1 Heterogeneity in pulmonary fibrosis survival time by racial group. a) Survival time from interstitial lung disease (ILD) diagnosis stratified by race/ ethnicity; b) survival stratified by race among patients with ILD.

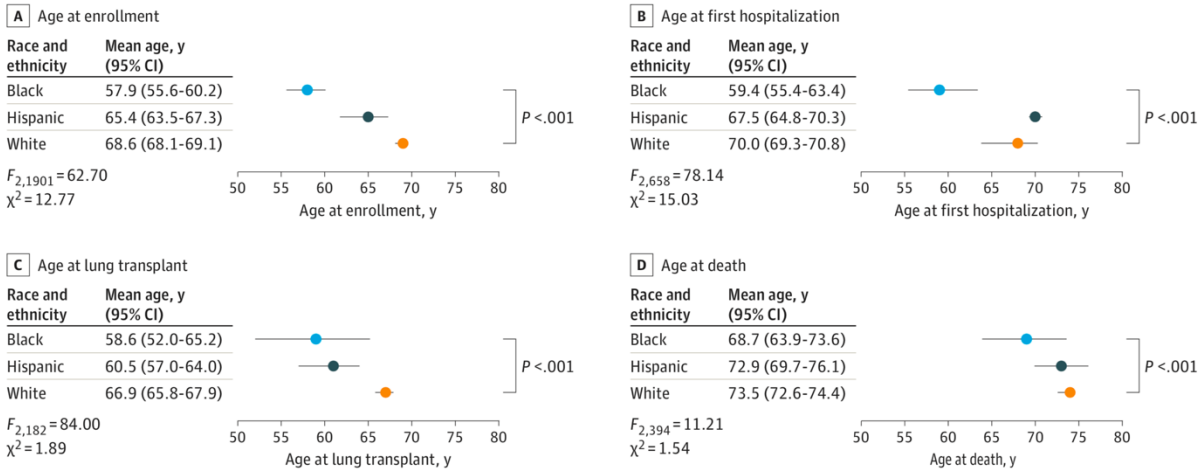


Racial and ethnic disparities are present in pulmonary fibrosis-related outcomes

The significance of race and ethnicity as influential variables in the progression and outcomes of pulmonary fibrosis remains an area of exploration yet to be fully tapped. Given the limited understanding of the age at which pulmonary fibrosis-related outcomes occur in racial and ethnic minority populations, I hypothesized the existence of racial or ethnic disparities in the age at these outcomes among patients with pulmonary fibrosis. Consequently, I aimed to assess the age at clinically relevant pulmonary fibrosis-related outcomes (diagnosis, hospitalization, lung transplant, and death) and evaluate the heterogeneity in survival patterns among White, Black, and Hispanic participants with pulmonary fibrosis. For these purposes, I employed a nationally acquired US registry and corroborated these findings using data from four geographically distinct tertiary care centers with pulmonary fibrosis expertise [11].

This research uncovered considerable racial and ethnic disparities in pulmonary fibrosis outcomes throughout the lifespan of affected individuals from these minority groups. Remarkably, Black patients with pulmonary fibrosis experienced diagnosis, hospitalization, lung transplantation, and death at younger ages compared to Hispanic and White patients (**Figure 1.2**), and these findings remained consistent across various age groups (**Figure 1.3**)[11]. While mortality rates seemed lower in Black individuals than in Hispanic and White individuals, the slight increase in life expectancy did not compensate for the significant age differences, given that the diagnosis of pulmonary fibrosis in Black patients generally occurred a decade earlier compared to White patients. pulmonary fibrosis-induced debilitation typically manifests around middle age, with the median age at pulmonary fibrosis diagnosis occurring between 60 and 70 years [11]. Consequently, the earlier onset of the disease likely has substantial repercussions on the quality of life, frequency of hospitalization, and functional capacity of the affected individuals.

Figure 1.2 Mean age at pulmonary fibrosis outcomes stratified by race and ethnicity. *P* values are for analysis of variance comparing all 3 racial groups.



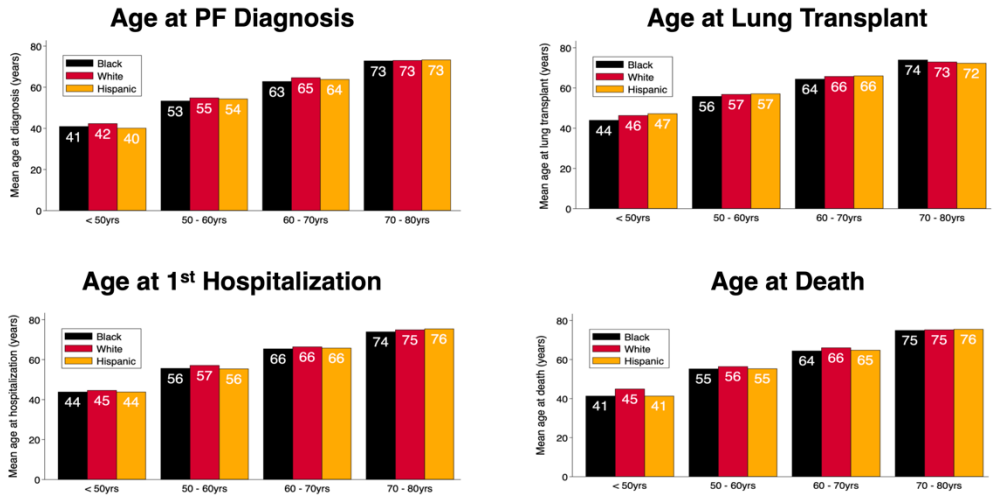
Interestingly, Black individuals reportedly encounter delays in disease diagnosis[10]. This observation raises concerns that even the seemingly earlier diagnosis of pulmonary fibrosis in Black patients might still be delayed. Consistent with current trends in national epidemiological data, our study demonstrated lower lung transplant rates and disproportionately higher hospitalization rates among Black and Hispanic patients with pulmonary fibrosis, compared to White patients. These disparities underscore the profound impact of healthcare disparities on racial minority populations.

Our findings also elucidated intriguing demographic patterns across racial and ethnic groups. Aligning with the male predominance in the overall study population, the majority of White patients with pulmonary fibrosis were men. This contrasted with the female predominance observed in the Black subpopulation. Accompanying this divergent sex distribution, pulmonary fibrosis diagnostic subcategories varied in their distribution patterns across racial and ethnic groups. While idiopathic pulmonary fibrosis (IPF) was the prevalent diagnosis in White patients, fibrotic hypersensitivity pneumonitis (fHP) was most commonly diagnosed in Hispanic patients. Autoimmune related interstitial lung disease, alternatively known as Connective tissue disease-interstitial lung disease (CTD-ILD), was three to four times more common in Black patients compared to White patients. Considering the higher reports of autoimmune disease in Black individuals and the frequent presence of pet birds in Hispanic households, it's plausible that confirmation bias in the medical diagnosis process contributes to these observed diagnostic disparities across racial and ethnic groups.

For instance, a study examining US decedents with IPF within the National Center for Health Statistics database indicated that Black decedents were less likely to be coded with IPF than White decedents [9]. Conversely, Hispanic decedents were more likely to receive an IPF code at death. This anomaly underscores the need for clinical data from geographically diverse ILD centers and meticulously curated datasets prioritizing diagnostic ascertainment when assessing outcome disparities across racial subgroups.

This study illuminated the substantial racial and ethnic disparities present in pulmonary fibrosis-related outcomes [11]. It underscores the need for continued exploration and a deeper understanding of these disparities to inform equitable healthcare policies and interventions. A more inclusive approach to research and clinical care will not only enhance our understanding of the disease but also drive us towards tailored therapeutic strategies that address these disparities, leading to improved patient outcomes across all races and ethnicities.

Figure 1.3 Consistent disparities in pulmonary fibrosis outcomes across age groups.



Telomere length and gene variant association with survival in pulmonary fibrosis

Our understanding of the interplay between genetics and the pathogenesis of interstitial lung diseases (ILDs), including pulmonary fibrosis (PF), has expanded considerably in recent years. Evidence now suggests that common genetic variants and short telomere lengths contribute significantly to the risk of developing either sporadic or familial ILD [12-14]. Telomeres are crucial noncoding parts of chromosomes, vital for maintaining genomic stability [15]. They consist of repeated 6 nucleotide sequences (TTAGGG), which shrink with each cellular division, triggering apoptosis when a critical shortening threshold is reached. The complex molecular machinery involved in telomere structure and regulation has fascinating implications in a variety of inherited diseases, many of which result in ILD or bone marrow dysfunction.

It is worth delving into the intricacies of telomere structure: telomeres exhibit a unique "D-loop-T-loop" structure, featuring an elongated single guanine-rich strand that invades a double-strand region of the telomere, displacing one strand to form the displaced loop (D-loop). The accompanying double-strand region, known as the T-loop or telomere loop, and the telomere loop's elongated single-strand guanine-rich area, stabilize this unique D-loop-T-loop structure. Key proteins, such as Ku, involved in DNA double-strand break repair, play a role in telomere maintenance. Notably, anti-Ku antibodies have been reported in systemic sclerosis and myositis overlap syndromes that frequently manifest with significant associated ILD [16]. Moreover, telomerase, comprising telomere reverse transcriptase (TERT) and the telomere RNA template (TERC), is integral in the generation of telomeric repeats [15]. Additional proteins also

contribute to telomere replication, regulation, and maintenance. Mutations in these telomere-related genes have a clear association with the development of severe ILD. However, beyond these conditions, telomere length in itself presents an exciting potential biomarker. Shortened age-adjusted leukocyte telomere length (LTL) has been linked to worse survival in patients with idiopathic pulmonary fibrosis (IPF) and chronic hypersensitivity pneumonitis (CHP) [17, 18]. However, the pathogenic implications of shortened telomere length across various ILD subtypes remain underexplored.

Figure 1.4 Telomere-related gene variants across clinical phenotypes of pulmonary fibrosis

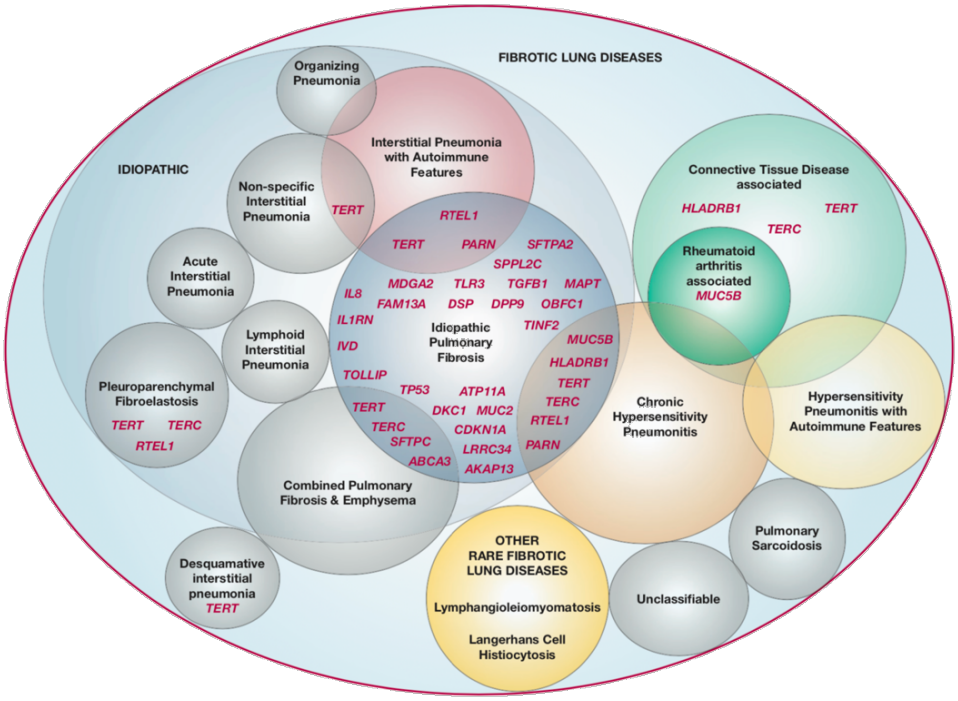


Table 1.1. Gene function of common and rare gene variants associated with pulmonary fibrosis.

COMMON GENE VARIANTS (risk allele)	Chromosome	Gene function	Phenotype
MUC5B *(rs35705950)*[+]	11p15.5	Affects airway mucus properties, mucociliary transport, & airway defense	IPF, CHP, RA-ILD
TOLLIP *(rs111521887, rs5743894, rs2743890)*	11p15.5, 1q32.2	Regulates TLR-mediated immune responses, & TGF-β signaling pathway	IPF
IL1RN *(rs408392, rs419598)*	2q14.1	Inhibits pro-inflammatory effect of IL-1α and IL-1β	IPF
IL8 *(rs4073, rs2227307)*	4q13	Pro-inflammatory cytokine	IPF
FAM13A *(rs2609255)*[+]	4q22	Affects signal transduction	IPF
TLR3 *(rs3775291)*	4q35	Influences recognition of pathogens and activation of innate immunity	IPF
TERT *(rs2736100)*	5p15.33	Telomerase complex enzyme that maintains telomere length	IPF, CPFE
HLA-DRB1 *(rs2395655)*	6p21.2	Major histocompatibility complex, critical to immune system	IPF, CHP, CTD-ILD
DSP *(rs2076295)*[+]	6p24	Cell adhesion, integrity, and mechanotransduction	IPF
OBFC1 *(rs11191865)*[+]	10q25.1	Stimulates DNA polymerase-alpha-primase activity	IPF
MUC2 *(rs7934606)*	11p15.5	Mucin production	IPF
ATP11A *(1278769)*[+]	13q34	Phospholipid translocation	IPF
MDGA2 *(rs7144383)*	14q21	Cell-cell interaction	IPF
MAPT *(rs1981997)*	17q21.3	Promotes microtubule assembly and stability	IPF
SPPL2C *(rs17690703)*	17q21.3	Protein cleavage	IPF
DPP9 *(rs12610495)*[+]	19p13.3	Cell adhesion, integrity, and mechanotransduction	IPF
TGFB1 *(rs1800470)*	19q13.2	Controls proliferation and differentiation of diverse cell types	IPF
TP53 *(rs12951053, rs12602273)*	17p13	Production of tumor suppressor protein p53	IPF
CDKN1A *(rs2395655)*	6p21.2	Critical in cell cycle and response to DNA damage	IPF
IVD *(rs2034650)*	15q15	Production of isovaleryl-CoA dehydrogenase mitochondrial matrix enzyme	IPF
LRRC34 *(rs6793295)*	3q26.2	Production of leucine-rich repeat protein for pluripotent stem cells	IPF
AKAP13 *(rs62025270)*	15q25	Regulates activation of profibrotic signaling pathways	IPF

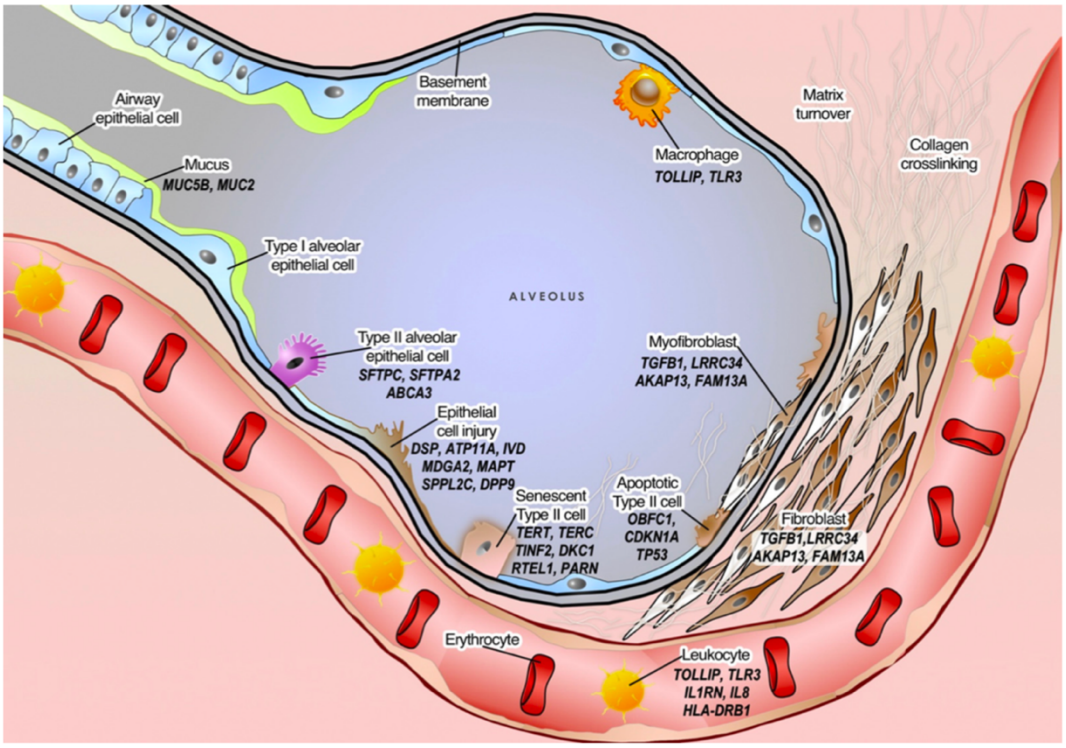
RARE GENE VARIANTS (risk allele)	Chromosome	Gene function	Phenotype
SFTPC[##]	8p21.3	Protein C component of surfactant fluid	IPF, CPFE
SFTPA2[#]	10q22.3	Pulmonary surfactant-associated protein A2 affects immunomodulation	IPF
ABCA3[#]	16p13.3	Lipid transport across plasma membrane	IPF, CPFE
TERT[#]	5p15.33	Telomerase complex reverse transcriptase enzyme	IPF, IPAF, CHP, CTD-ILD, PPFE, NSIP, DIP, CPFE
TERC[#]	3q26.2	Maintenance of telomerase RNA complex template	IPF, CHP, CTD-ILD, PPFE, CPFE
DKC1[#]	Xq28	Telomerase complex template stabilization	IPF
TINF2[#]	14q12	Maintenance of telomeres	IPF
RTEL1[#]	20q13.33	DNA helicase activity	IPF, IPAF, CHP, PPFE
PARN[#]	16p13.12	Stability of mRNA	IPF, IPAF, CHP

Gene data obtained from U.S. National Library of Medicine (https://ghr.nlm.nih.gov/gene/ and https://www.ncbi.nlm.nih.gov/genome/gdv/).
CHP 1⁄4 chronic hypersensitivity pneumonitis; CPFE=combined pulmonary fibrosis and emphysema; CTD-ILD 1⁄4 connective tissue disease-associated interstitial lung disease; DIP=desquamative interstitial pneumonia; IPAF=interstitial pneumonia with autoimmune features; coA= coenzyme A; IPF=idiopathic pulmonary fibrosis; mRNA=messenger RNA; NSIP=nonspecific interstitial pneumonia; PPFE=pleuroparenchymal fibroelastosis; RA-ILD=rheumatoid arthritis-associated interstitial lung disease; TGF-β=transforming growth factor-b; TLR=toll-like receptor.
[a]Minor allele frequency > 5%.
[b]Minor allele frequency 2% to 25%.
[c]Minor allele frequency < 1%.

To fill this gap in knowledge, I embarked on a comprehensive review of current literature, identifying studies examining common and rare gene variants impacting telomere biology in individuals across the fibrotic lung disease spectrum (**Figure 1.4**)(**Table 1.1**) [19]. I noted that about 15% of familial kindreds with ILD harbor rare coding mutations in *TERT*, the gene locus that encodes the protein component of telomerase [20-22]. Other rare mutations with large effects were found in genes encoding the human telomerase RNA component (*TERC*), surfactant proteins (*SFTPC* and *SFTPA2*), and dyskerin (*DKC1*). Interestingly, gene burden analysis of 99 unrelated patients with familial pulmonary fibrosis (FPF) revealed that an exoribonuclease, *PARN*, shared five new heterozygous mutations among all affected relatives. Furthermore, damaging and missense variants were found in the *RTEL1* gene encoding a DNA helicase critical for telomere stability, protection, and elongation [22]. Whole exome sequencing (WES) of genomic DNA obtained from 25 familial interstitial pneumonia (FIP) kindreds registered in an ongoing study and validated through further sequencing of affected individuals from another 163 kindreds authenticated the existence of candidate rare variants [23]. Peripheral blood mononuclear cells (PBMCs) from affected individuals as well as from unaffected carriers of identified rare variants exhibited telomere lengths falling below the 10th percentile for age. This phenomenon suggested compromised functionality of *RTEL1*, a DNA helicase vital for unwinding the T-loop structure of telomeres at chromosomal ends. Thus, an impairment of *RTEL1* results in the release of circular telomeric DNA fragments (T-circles) and a progressive shortening of telomeres

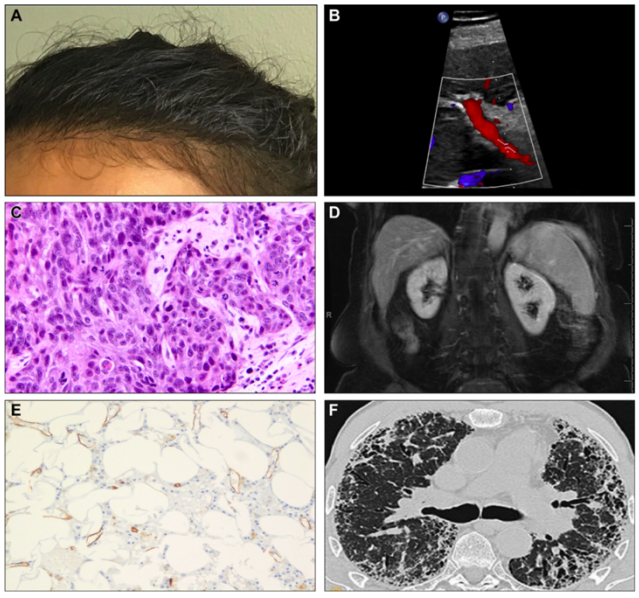
with successive rounds of DNA replication. In addition to inducing genomic instability, the loss of *RTEL1* functionality also contributes to global genome replication defects, culminating in a disrupted cell cycle progression (**Figure 1.5**). Similarly, telomere length measurement, classical burden test evaluation of excess-risk mutations, and WES of nine FPF-linked genes in a cohort of 101 well-phenotyped cases of rheumatoid arthritis-associated interstitial lung disease (RA-ILD) in comparison to 1,010 control subjects without known autoimmune, inflammatory, or pulmonary disease showed that 12% of patients with RA-ILD had heterozygous mutations in the coding regions for *TERT*, *RTEL1*, *PARN*, or *SFTPC* gene loci [24]. These mutations were present at a significantly higher frequency among patients with RA-ILD compared to healthy control subjects, underscoring the role of these FPF-linked genes in RA-ILD susceptibility.

Figure 1.5 Conceptual fibrosis model of altered gene function and affected cells.



Currently, we do not have concrete guidelines for undertaking genetic testing in FIP or for integrating test results into patient care. However, existing recommendations suggest updating the family history of IPF patients at every visit since up to 10% of bloodline relatives subsequently receive a diagnosis of idiopathic interstitial pneumonia. Indications of short telomeres, such as a family history of neonatal respiratory distress, childhood ILD, aplastic anemia, cryptogenic cirrhosis, or premature graying before 25 years of age, could guide such an evaluation [25] (**Figure 1.6**). A positive genetic diagnosis may not only be prognostic, but also facilitate risk stratification in lung transplant candidates and assist in estimating risk for close relatives. Therefore, genetic testing could be beneficial in patients with FIP and high likelihood of a positive genetic diagnosis.

Figure 1.6 Telomere shortening and observed clinical phenotypes. A, Early onset gray hair. B, Coronal MRI image depicting splenomegaly in non-cirrhotic portal hypertension. C, Hepatic Doppler ultrasonographic image depicting enlarged caliber of portal vein in non-cirrhotic portal hypertension. D, CD34 immunostaining histologic image of bone marrow aspirate depicting occasional mast cells and plasma cells in a patient with aplastic anemia. E, High-resolution hematoxylin and eosin stain depicting squamous cell carcinoma of the tonsil. F, Chest high-resolution CT axial image depicting usual interstitial pneumonia in a patient with pulmonary fibrosis.

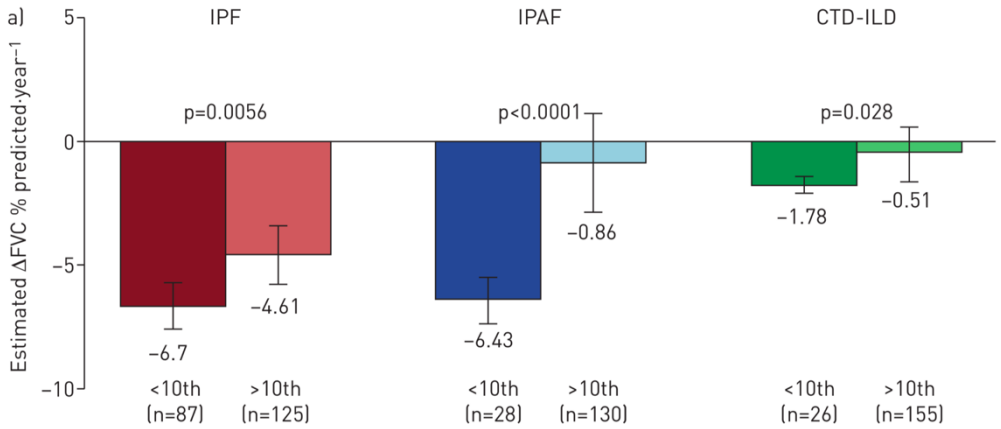


In collaboration with colleagues from the University of Texas Southwestern, and the University of California San Francisco, I also investigated LTL, and *MUC5B rs35705950* and *TOLLIP rs5743890*, which had all been previously linked to IPF [26]. Our study aimed to assess the associations between these genomic markers and survival outcomes as well as the rate of disease progression in patients with interstitial pneumonia with autoimmune features (IPAF), and connective tissue disease-associated interstitial lung disease (CTD-ILD). IPF was used as a comparator. Our findings illustrated that IPAF and CTD-ILD patients have a longer LTL compared to IPF patients. Among IPAF patients, LTL below the 10th percentile correlated with a faster lung function decline and worse transplant-free survival (**Figure 1.7**). The *MUC5B rs35705950* minor allele frequency (MAF) was higher in IPAF patients and was linked to worse transplant-free survival. Interestingly, RA-ILD had a shorter LTL and higher MUC5B MAF compared to non-RA CTD-ILD. This study emphasized the differential associations of LTL and MUC5B MAF with lung function progression and survival for the autoimmune-related forms of ILD (IPAF and CTD-ILD) and underscores the burgeoning recognition of the role of genetic variants and telomere biology in the pathogenesis, progression, and prognosis of ILD. This area of research promises to bring forth novel diagnostic and therapeutic strategies to address these devastating diseases.

Radiologic honeycombing and increased mortality in pulmonary fibrosis

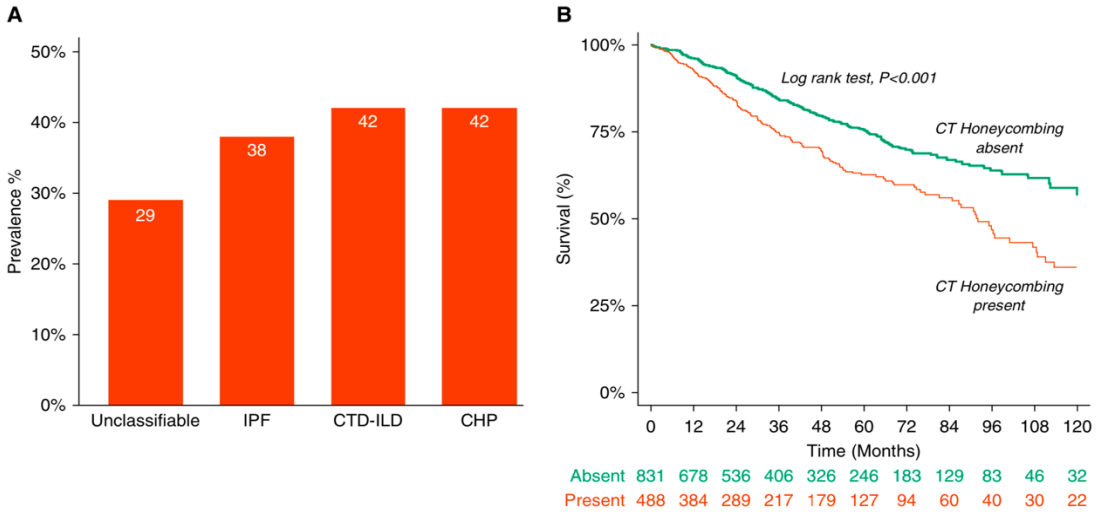
A gamut of pulmonary fibrosis features can be identified on high-resolution computed tomography (CT) chest imaging, notably reticulation, bronchiectasis or bronchiolectasis, and honeycombing [27, 28]. Honeycombing, defined as clustered cystic airspaces of similar diameters (~3–10 mm), typically subpleural with thick, well-defined walls, is a critical CT characteristic in IPF, the most lethal form of pulmonary fibrosis [28]. IPF, which is diagnosable in individuals with no discernable cause, often presents as a "usual interstitial pneumonia" (UIP) pattern characterized by predominant lower lobe and peripheral bilateral reticulation and honeycombing on high-resolution CT imaging. Recent consensus guidelines by the American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Association stipulate that the 'extent' of CT honeycombing independently forecasts mortality in individuals with IPF [28]. Consequently, the extent of honeycombing is frequently assessed on chest CT scans during IPF evaluation.

Figure 1.7 Progression of pulmonary fibrosis by leukocyte telomere length.



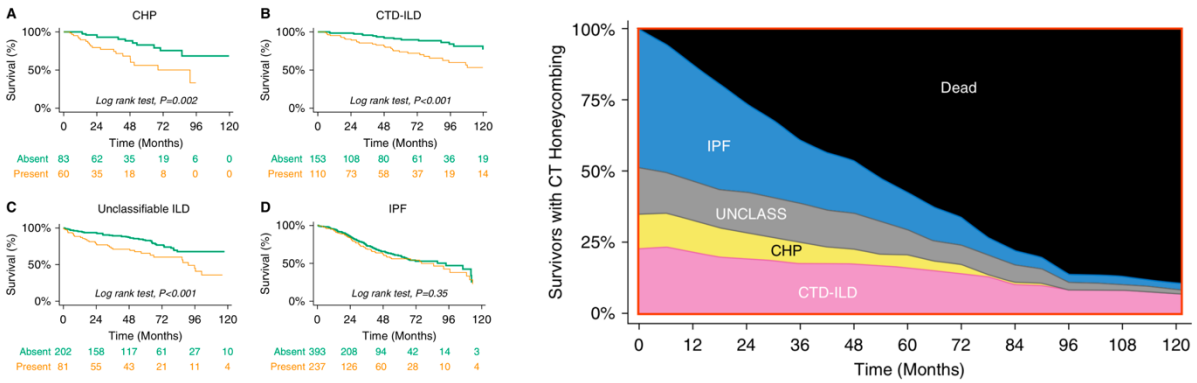
However, despite the critical role honeycombing plays in IPF, its prevalence, prognostic impact, and associated mortality patterns across the diverse spectrum of ILDs are not well understood. The frequent diagnostic reclassification to alternate ILD subtypes in patients initially classified as IPF and the shared phenotypic similarities amongst different progressive fibrotic ILDs prompted me to hypothesize that the binary presence of honeycombing on chest CT scans (radiologic honeycombing) might be prevalent and indicative of reduced survival across various ILD subtypes, independent of the ILD classification or clinical disease severity [29]. I thus sought to ascertain if CT honeycombing was associated with mortality and delineate the mortality patterns in study participants with CT honeycombing [30].

Figure 1.8 Prevalence of honeycomb pulmonary fibrosis and survival patterns. (A) Prevalence across diverse forms of interstitial lung disease (ILD) subtypes; (B) Kaplan-Meier analysis demonstrates worsened 10-year survival in participants with ILD and CT honeycombing.



To bridge this knowledge gap, I orchestrated an observational cohort study involving adult participants with multidisciplinary or adjudicated ILD diagnosis and available chest CT imaging documentation at the index diagnosis [30]. The sample spanned five U.S. hospitals (including one tertiary and four non-tertiary medical centers) and comprised 1,330 participants, evaluated over 4,831 person-years of follow-up. We observed the prevalence of CT honeycombing as 42.0%, 41.9%, 37.6%, and 28.6% in fibrotic HP, CTD-ILD, IPF, and unclassifiable/other ILDs, respectively (**Figure 1.8**). Among participants with CT honeycombing, cumulative mortality hazards were consistent across ILD subtypes, except for CTD-ILD, which exhibited a lower mortality hazard. Overall, the mean survival time was notably shorter among those participants with CT honeycombing (**Figure 1.8-1.9**).

Figure 1.9 Variation in mortality from honeycombing across pulmonary fibrosis subtypes.



After adjusting for center, sex, age, forced vital capacity, diffusing capacity, ILD subtype, and use of immunosuppressive therapy, CT honeycombing was associated with an increased mortality rate within non-IPF ILD subgroups (**Table 1.2**). Interestingly, in IPF,

mortality rates were comparable between those with and without CT honeycombing. This pivotal study underscored that CT honeycombing is remarkably prevalent in diverse ILD forms and identifies a progressive fibrotic ILD phenotype with a high mortality rate akin to IPF. Furthermore, this investigation elucidated that CT honeycombing is associated with a greater mortality rate when compared to those without honeycombing [30].

Table 1.2 Associations between computed tomography (CT) honeycombing and mortality.

Characteristics	CT Honeycombing Absent	CT Honeycombing Present	*P* Value
Participants	*n* = 838	*n* = 492	
Survival time, mo, mean (95% CI)	161 (147–174)	107 (92–122)	<0.001
Number of deaths, *n* (%)	170 (20.3)	160 (32.5)	<0.001
Crude mortality rate (events per 100 person-years)	5.0	9.0	<0.001
Unadjusted hazard ratio* (95% CI)	—	1.72 (1.38–2.14)	<0.001
Adjusted hazard ratio*† (95% CI)	—	1.62 (1.29–2.02)	<0.001
Lung transplantation, *n* (%)	23 (2.7)	19 (3.9)	0.26

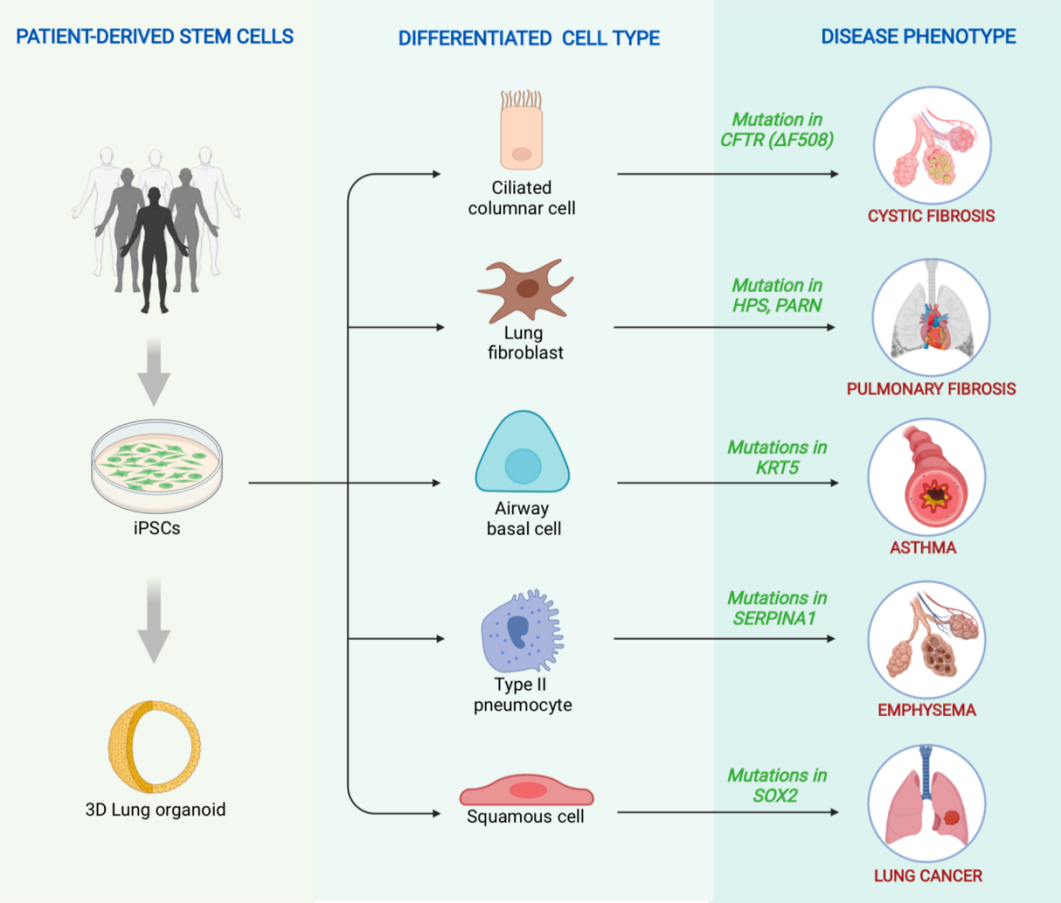
Definition of abbreviations: CI = confidence interval; CT = computed tomography.
*Computed using Cox proportional hazard models.
†Adjusted for hospital center, sex, age, forced vital capacity, diffusing capacity of the lungs for carbon monoxide, interstitial lung disease subtype, and immunosuppressive therapy. Unadjusted model, *n* = 1,330; adjusted model, *n* = 1,232.

Utility of induced pluripotent stem cells in pulmonary fibrosis pathobiology.

Over the last two decades, we have seen a surge in the identification of genomic loci linked to human traits and diseases, thanks to the proliferation of genome-wide association studies [31]. Despite this progress, our understanding of the genetic underpinnings of diseases is still unfolding, largely due to the fact that most disease-associated genetic variants are noncoding. These variants mostly influence gene regulation rather than the function of proteins. This recognition catalyzed my interest in exploring tools that enable us to delve into gene regulation in contexts that are most pertinent to the disease. Among the arsenal of models available, induced pluripotent stem

cells (iPSCs) and differentiated cell types stand out for their ability to simulate disease conditions [32, 33]. I thus undertook an extensive review of the literature on the utility of iPSCs during my doctoral studies, drawing from my own research and that of others [34]. In this review, I underlined the potential of iPSCs as a promising model for pulmonary fibrosis, helping us illuminate disease biology and augment our understanding of human genetic diseases. Historically, attempts to understand the causes and consequences of human disease have relied on functional data from frozen adult tissues obtained postmortem or via invasive surgery [35, 36]. Despite the enormous dataset generated by the Genotype-Tissue Expression project (GTEx), they were often limited in their ability to investigate the genetic influence on gene regulation in cell types and tissues that are most relevant to many complex diseases. This limitation sparked my hypothesis that we need to extend quantitative trait loci (QTL) mapping studies to more relevant cell types, tissues, and states to better understand the genetic regulation of diseases. My research led me to recognize iPSCs as a powerful model for studying gene regulation in disease-relevant contexts. iPSCs can be reprogrammed from a variety of accessible somatic cell types such as blood cells and fibroblasts [37], and they can be differentiated into virtually any cell type [38-41]. Crucially, iPSCs retain disease-specific DNA mutations, making them suitable for modeling both Mendelian and complex diseases at a population level. They provide us a dynamic in vitro environment that closely mimics the in vivo condition, making them a superior model over primary cell lines. Furthermore, they enable us to study regulatory variation during development, an aspect critical to disease progression that isn't easily studied in humans in vivo.

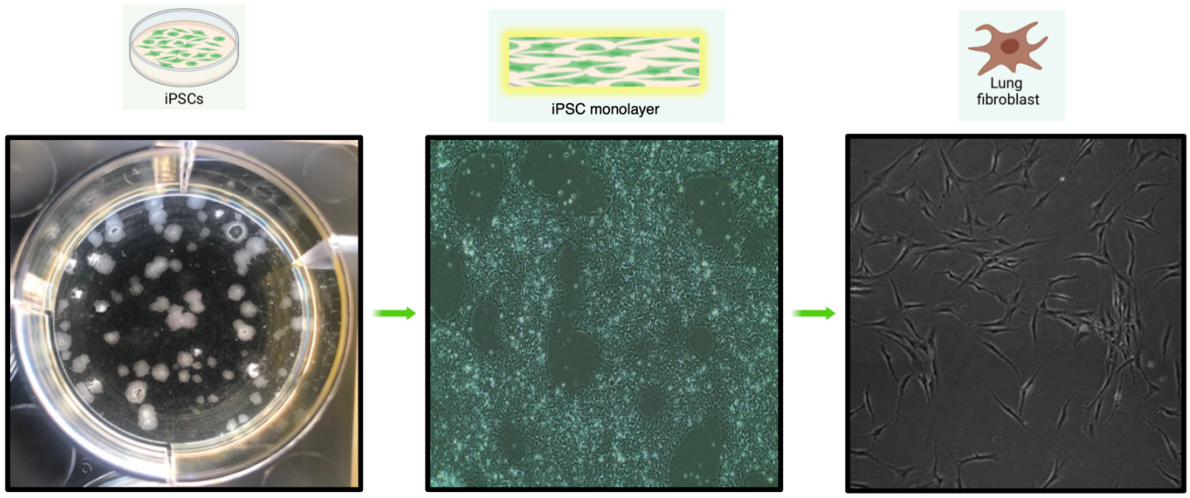
Figure 1.10 Genomic models of pulmonary disease phenotypes derived from iPSCs



In the realm of pulmonary diseases, the inaccessibility of lung cell types poses a significant challenge. Alveolar epithelial cells (AECs) and lung fibroblasts play crucial roles in lung function and the pathogenesis of pulmonary diseases [42, 43]. Investigative studies which I performed during my doctoral studies focused on leveraging iPSCs to model these cell types to explore the pathobiology of pulmonary fibrosis and other lung diseases. One of the promising aspects of iPSC-derived cells is that their gene expression profiles appear to closely resemble those seen in human fetal pulmonary cells. Through these patient-specific preclinical platforms, my ongoing research, in concert with that of

my peers, has leveraged the generation of iPSC-derived pulmonary cells with the potential to study a range of pulmonary disorders, including monogenic pulmonary disorders and fibrotic ILD (**Figure 1.10**). These studies provided me with critical insights into iPSC-derived pulmonary cells have enabled patient-specific preclinical platforms for elucidating epithelial-intrinsic dysfunction that occurs during the pathogenesis of ILD and provide assurance that iPSCs will continue to deepen our understanding of the pathobiology of pulmonary fibrosis and the influence of genetic variants. Following on knowledge gained from this endeavor during the course of my doctoral studies, I generated seventy (n=70) pulmonary fibrosis PBMC lines into lymphoblastoid cell lines (LCLs of which twenty (n=20) have been further converted into iPSCs, all of which have had QC indexing, transcriptomic analyses and WGS performed (**Figure 1.11**). These cell lines will provide an enduring platform for characterizing gene regulatory differences between LCLs, iPSCs, and differentiated cell types such as fibroblasts and AECs. They will also provide a platform to study gene regulation throughout fibroblast development using iPSCs and recently developed genomic techniques. Leveraging this resource, we can better study how regulation is encoded in the genome and the mechanisms by which regulation is achieved, and these cell lines will enable me and others to also explore new questions while assaying many different molecular phenotypes with the aid of additional genomic techniques such as ATAC-seq, ChIP-seq, and TWAS.

Figure 1.11 Digital images of PBMC-derived iPSCs in pulmonary fibrosis



Chapter II

Telomere length associates with chronological age across racially diverse pulmonary fibrosis cohorts

The data in this chapter are adapted from the following manuscript, which was published during my doctoral studies: **Adegunsoye A**, Newton CA, Oldham JM, Ley B, Lee CT, Linderholm AL, Chung JH, Garcia N, Zhang D, Vij R, Guzy R, Jablonski R, Bag R, Voogt RS, Ma SF, Sperling AI, Raghu G, Martinez FJ, Strek ME, Wolters PJ, Garcia CK, Pierce BL, Noth I. Telomere length associates with chronological age and mortality across racially diverse pulmonary fibrosis cohorts. Nature Communications 2023 March 17;14(1):1489. PMID: 36932145; PMCID: PMC10023792.

In this study, I designed and performed the experiments, analyzed the data, and wrote the manuscript.

Additional data in this chapter are adapted from the following publication in Thorax: Kim JS, Manichaikul AW, Hoffman EA, Balte P, Anderson MR, Bernstein EJ, Madahar P, Oelsner EC, Kawut SM, Wysoczanski A, Laine AF, **Adegunsoye A**, Ma JZ, Taub MA, Mathias RA, Rich SS, Rotter JI, Noth I, Garcia CK, Barr RG, Podolanczuk AJ. MUC5B, telomere length and longitudinal quantitative interstitial lung changes: the MESA Lung Study. Thorax. 2022 Aug 5: thoraxjnl-2021-218139. PMID: 36690926; PMCID: PMC9899287.

Introduction

Despite the significant advancements made in understanding the pathophysiology of pulmonary fibrosis, gaps still remain in our knowledge of the prognostic implications of this disease across different racial and ethnic groups[10]. In an effort to enhance the precision of pharmacotherapeutic measures and inform targeted management strategies for diverse patients, I embarked on an extensive study of pulmonary fibrosis biomarkers, with a particular focus on their implications across varying racial and ethnic groups [44]. A genomic biomarker that has recently caught the attention of the scientific community is the short leukocyte telomere length found in peripheral blood, which serves as an indicator of cellular senescence [45]. It is associated with germline mutations in both common and rare gene variants involved in telomere biology and linked to the incidence as well as mortality of pulmonary fibrosis [26]. The relationship between short leukocyte telomere length, age, and pulmonary fibrosis is well documented [17, 22, 26]. However, the significance of these biomarkers across different racial and ethnic populations is largely obscured by the underrepresentation of racial and ethnic minorities in most genetic studies to date. My research focuses on filling this void and answering the pivotal question of whether telomere shortening is consistently associated with increasing age in pulmonary fibrosis across diverse racial and ethnic groups. Interestingly, the scientific community has barely scratched the surface when it comes to understanding racial differences in gene polymorphisms among patients with pulmonary fibrosis. Most existing research has excluded African-American patients, which could be due to the reported higher prevalence of pulmonary fibrosis in European Americans.

While an increasing amount of work has been dedicated to understanding pulmonary fibrosis, a striking gap remains in our understanding of racial differences in gene polymorphisms among patients. Prior studies, including linkage analysis, telomere length measurements, and genome-wide association studies, have disproportionately excluded African-American patients [46]. This may stem from an often-held belief that pulmonary fibrosis is more prevalent in European Americans. However, such an oversight potentially leaves an unaddressed question: Do genetic polymorphisms that contribute to PF susceptibility in non-African-American patients apply to African-American populations? Several single nucleotide polymorphisms (SNPs) associated with IPF susceptibility have been identified through genome-wide association analyses (GWAS) [14, 47, 48]. However, most of these SNPs are relatively rare and their association with outcomes remains to be clarified. Two critical gene regions, the *MUC5B* gene promoter - rs35705950, which encodes a member of the mucin family of proteins, and Toll-interacting protein (*TOLLIP*) - rs5743890, rs5743894, rs5743854, rs3750920, which regulates human Toll-like receptor signaling pathways, are areas of high conservation [49-54]. Polymorphisms in these regions are considered the most potent and replicated genetic risk factors for IPF, with established associations with disease outcomes. Individuals with shorter telomere length, have been shown to more frequently have lung high attenuation areas which is associated with the MUC5B risk allele [55]. Interestingly, it has been suggested that *MUC5B*, associated with pulmonary fibrosis onset and

mortality in European Americans, might be absent in African populations, an aspect that needs further exploration [52].

In collaboration with other research team members, I set out to elucidate this complexity. By performing targeted genotyping of the most significant pulmonary fibrosis-associated SNPs, including *TOLLIP* and *MUC5B*, we analyzed a multi-racial subset of patients with autoimmune-related PF, enrolled prospectively in the University of Chicago ILD registry. A preliminary analysis of our data uncovered striking racial disparities in these susceptibility and mortality-associated SNPs ($p<5x10-5$). Notably, the highest prevalence of the heterozygous rs35705950 G->T (*MUC5B*) and rs5743890 A->G (*TOLLIP*) risk alleles was observed in patients of European ancestry, while patients of African or Asian descent exhibited significantly lower frequencies of these SNPs among all US racial groups. Furthermore, patients of European ancestry had notably shorter peripheral blood leukocyte telomere length when compared to other US racial groups. These findings point towards a compelling link between racial ancestry and differences in these genomic factors among patients with autoimmune-related pulmonary fibrosis, suggesting potential variations in pulmonary fibrosis-associated SNPs and telomere lengths across different races.

To fill the existing knowledge gap, I spearheaded the first comprehensive natural history study of telomere length variation across racial and ethnic groups [44]. This study aimed (a) to comparatively analyze differences in leukocyte telomere length between racial/ethnic groups with pulmonary fibrosis and within control populations; (b) to investigate whether leukocyte telomere length is associated with chronological age

across racially and ethnically diverse pulmonary fibrosis cohorts and (c) to assess the relationship between leukocyte telomere length and mortality risk across ethnic groups with pulmonary fibrosis. I hypothesized that shorter telomere length is associated with increased age and greater mortality across diverse racial groups, underscoring the importance of this biomarker in all patients with pulmonary fibrosis.

Methods

Study design, age, and mortality ascertainment

This comprehensive cohort study drew upon a meticulously curated database of patients diagnosed with various subtypes of fibrotic interstitial lung disease, collectively termed as pulmonary fibrosis. These patients were enrolled in a prospective manner across four widely dispersed U.S. tertiary centers: University of Chicago, University of California San Francisco, University of California Davis, and University of Texas Southwestern, Dallas. Spanning over a significant period from September 2003 to December 2019, this study ensured a robust multidisciplinary diagnosis of pulmonary fibrosis through an exhaustive evaluation that incorporated clinical data, pulmonary function tests, high-resolution computed tomography scans, and surgical lung biopsies, adhering to the current American Thoracic Society/European Respiratory Society criteria.

A dedicated team of pulmonologists, rheumatologists, thoracic radiologists, and a thoracic pathologist collaboratively assessed subjects for the multidisciplinary diagnosis of interstitial lung disease. All participants were rigorously recruited in accordance with

consensus American Thoracic Society/European Respiratory Society criteria and were enrolled post receiving written informed consent. The study was conducted in strict compliance with guidelines set forth by relevant hospital institutional review boards (IRB: UChicago#14163A; UCSF#10-01592 & #10-00198; UCDavis#585448-7 & #875917-2; UTSW#082010-127 & #AAAS0753).

Study participants were enrolled at the diagnosis of pulmonary fibrosis. Furthermore, patients who participated in the Idiopathic Pulmonary Fibrosis Clinical Research Network clinical trials and consented to the optional genetics substudy were included, provided they had available genomic DNA. As a control group, we included healthy participants from the "Health and Retirement Study"(HRS)[NIA-U01AG009740] who had available individual-level demographic data and leukocyte telomere length measurements obtained via quantitative polymerase chain reaction assay.

For this study, race and ethnicity data were self-reported by participants and categorized following the standards set by the U.S. Census Bureau, thereby ensuring the consistency of demographic information [56]. Participants reported their chronological age at the point of study enrollment. The specified racial and ethnic categories were White (not Hispanic), Black (not Hispanic), Asian (not Hispanic), and Hispanic. Any participant fitting these predefined classifications was included in the analysis. The study focused on comparisons among the following four groups: Hispanic, non-Hispanic White, non-Hispanic Black, and non-Hispanic Asian participants, henceforth referred to as Hispanic, White, Black, and Asian participants [56]. Vital status was meticulously verified through a review of medical records and cross-referenced with the U.S. Social Security death index

for reliability. Lung transplantation status was ascertained from the Electronic Medical Records (EMR) of individual sites, ensuring a robust data collection approach. Recognizing the potential influence of age and sex on mortality in pulmonary diseases, analyses of the association of leukocyte telomere length with mortality were carefully adjusted for these variables. In all cohorts, mortality refers to all-cause mortality unless otherwise specified. Our rigorous methodology underscored the critical importance of demographic considerations in the study of pulmonary fibrosis and their potential impact on disease outcomes.

Telomere length analysis

We extracted genomic DNA from peripheral blood leukocytes of study participants. Leukocyte telomere length in peripheral blood was measured using the quantitative polymerase chain reaction (qPCR) technique, which was performed in triplicate using the RotoGene real-time PCR system by Qiagen [17, 21, 57]. LTL values were then adjusted for age, using data from healthy controls for comparison. Further, we utilized leukocytes from saliva samples taken from control subjects in the Health and Retirement Study (HRS) for comparative analysis [58]. To facilitate the comparison and analysis, we used standardized LTL values, calculated by subtracting the expected value (based on the Cronkhite 2008 study) from the observed LTL obtained through qPCR [59]. These values were depicted in the results as the telomere length (O-E, adjusted). While LTL measurements for all cohorts are represented as a T/S ratio, for the California cohort, LTL is reported in base pairs (bp). Given the potential variability in qPCR-based LTL

measurements across different study sites, we opted to standardize LTL values. We achieved this by employing z-score normalization, which adjusted for age and sex and categorized the results into quartiles. This process ensured a uniform distribution for each sample based on empirically computed standard deviations for each individual across each cohort. In our survival analyses, we classified LTL measurements below the median (TL50) and transformed TL (employing a negative log-transformation of one minus the percentile TL). I then compared the mortality hazard ratios for each percentile of TL to the highest TL percentile among subjects with pulmonary fibrosis.

Statistical analyses

Across all cohorts, I employed Fisher exact tests to investigate the associations between pairs of categorical variables. For continuous variables, I applied two-tailed t-tests or analysis of variance (ANOVA). To compare LTL between distinct diagnostic subgroups of pulmonary fibrosis, I made use of the nonparametric Mood's median test. Recognizing baseline demographic differences between Black and White participants with PF, I performed sensitivity analyses involving propensity-score matching within this selected subcohort, followed by LTL measurements analysis. I employed generalized linear models equipped with a logistic regression link to evaluate the relationship between age and LTL. I then assessed the assumption of linearity for continuous covariates using a lack-of-fit sum of squares test, which compares the linear fit to a restricted cubic spline fit with five knots, thereby ensuring the statistical model's suitability for the selected observations. Linear regression models were utilized to assess the congruity between

median standardized LTL and chronological age for all included racial/ethnic groups, stratified by sex. I analyzed rates of absolute mortality using conditional logistic regression.

The investigation leveraged Cox proportional hazards models for estimating hazard ratios and time-to-mortality analysis, with robust standard errors incorporated to account for familial correlation within the cohort. Patients were stratified by quartiles of standardized LTLs, and I used the Cochran–Armitage test to discern trends across quartiles, attributing quartile integers (1, 2, 3, and 4) as scores [17, 60]. I plotted the estimated survival functions for each quartile based on the Cox model. For the PF cohort, I considered transplant-free survival over five years, with survival time calculated from study enrollment to death, lung transplantation, loss to follow-up, or the end of the study period. My multivariable models assessing mortality were adjusted for age, sex, measures of pulmonary function, and hospital center, focusing on established confounders previously linked to both the predictor and the outcome. Survival curves were plotted using the Kaplan–Meier survival estimator, and in my Cox models, I ensured that no covariate effects violated the proportional hazards assumption. I designated a level of 0.05 as statistically significant for all two-sided P values. For data collation, I used Microsoft Excel for Mac Version 16.65, 2019, and carried out analyses using Stata (V 2019.R.16, V 2021.R17; StataCorp) and R.v.3.5.1.

Results

Our comprehensive study included a total of 7,854 individuals, comprised of 2,046 participants diagnosed with pulmonary fibrosis and 5,808 control participants from the Health and Retirement Study (**Table 2.1**).

Table 2.1 Pulmonary fibrosis (PF) and health retirement survey (HRS) control cohorts stratified by racial ancestry.

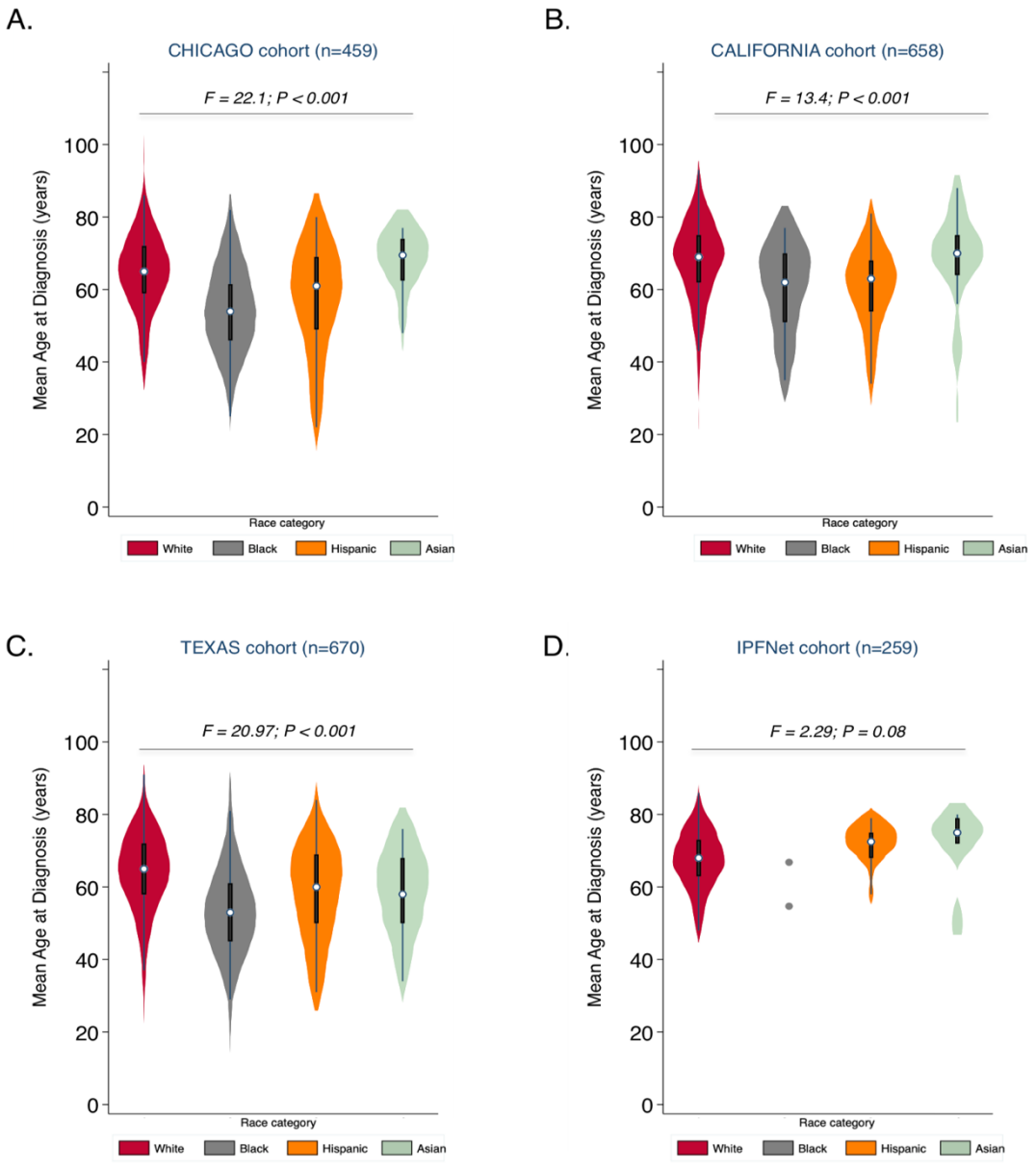
Characteristics[a]	White	Black	Hispanic	Asian	*P* value[#]
PF (*n* = 2046)	(*n* = 1613)	(*n* = 162)	(*n* = 187)	(*n* = 70)	
Age, years	66.0 (10.4)	54.9 (12.2)	59.7 (12.8)	65.7 (12.9)	<0.001
Male	918 (56.9)	44 (27.2)	90 (48.1)	39 (55.7)	<0.001
Ever Smoker	978 (60.8)	70 (43.2)	96 (51.3)	26 (37.7)	<0.001
Body Mass Index	29.2 (5.7)	29.0 (6.5)	29.8 (6.7)	25.9 (5.1)	0.001
Lung function					
FVC (% predicted)	68.4 (19.1)	59.4 (18.4)	62.0 (19.2)	66.7 (18.2)	<0.001
FEV_1 (% predicted)	74.1 (19.7)	62.9 (20.5)	65.7 (19.6)	74.4 (25.0)	<0.001
DL_{CO} (% predicted)	48.4 (19.2)	41.8 (18.9)	45.0 (17.3)	45.3 (17.1)	<0.001
PF sub-category					
IPF	653 (40.5)	13 (8.0)	49 (26.2)	20 (28.6)	<0.001
IPAF	153 (9.5)	30 (18.5)	15 (8.0)	8 (11.4)	0.006
CTD-ILD	200 (12.4)	89 (54.9)	46 (24.6)	14 (20.0)	<0.001
FHP	359 (22.3)	14 (8.6)	34 (18.2)	11 (15.7)	0.001
Unclassifiable/others	248 (15.4)	16 (9.9)	43 (23.0)	17 (24.3)	<0.001
HRS (*n* = 5808)	(*n* = 4319)	(*n* = 779)	(*n* = 614)	(*n* = 96)	
Age, years	70.2 (10.2)	67.8 (9.9)	66.0 (10.5)	65.5 (11.0)	<0.001
Male	1817 (42.1)	280 (35.9)	235 (38.3)	39 (40.6)	0.007
Ever smoker	2493 (58.0)	442 (57.1)	320 (52.7)	40 (41.7)	0.001
Body mass index	27.9 (5.7)	29.8 (6.6)	29.0 (5.6)	26.7 (5.9)	<0.001

[a]Categorical variables presented as *n* (%); continuous variables presented as means (SD).
[#]*P* value for chi-squared (categorical data) or one-way ANOVA (continuous data) comparing all four main racial groups. Patients with mixed or other racial ancestry not depicted above, *n* = 14. Exception for participants in PF cohort; smoking status, *n* = 2041; Body mass index = 1505; *FVC* forced vital capacity, *n* = 2015; *FEV1* forced expiratory volume in the first second, *n* = 1254; DL_{CO} diffusing capacity of the lungs, *n* = 1959. *ILD* interstitial lung disease; *IPF* idiopathic pulmonary fibrosis, *n* = 735, *IPAF* interstitial pneumonia with autoimmune features, *n* = 207; *CTD-ILD* connective tissue disease associated-ILD, *n* = 349; *FHP* fibrotic hypersensitivity pneumonitis, *n* = 422; unclassifiable/other ILD, *n* = 333.

The average age of PF participants was 65 years with a standard deviation (SD) of 11 years, while for the HRS control group it was slightly higher at 69 years with a SD of 10 years. It was noteworthy that Black participants consistently presented the youngest mean age across all sites, and significantly, the age at pulmonary fibrosis diagnosis was markedly lower than that of their White counterparts (Chicago: 54 vs 65 years, $P < 0.001$;

California: 60 vs 68 years, P < 0.001; Texas: 53 vs 64 years, P < 0.001; IPFNet: 61 vs 67 years, P = 0.25) (**Figure 2.1**).

Figure 2.1. Chronological age distribution across racial groups with pulmonary fibrosis.

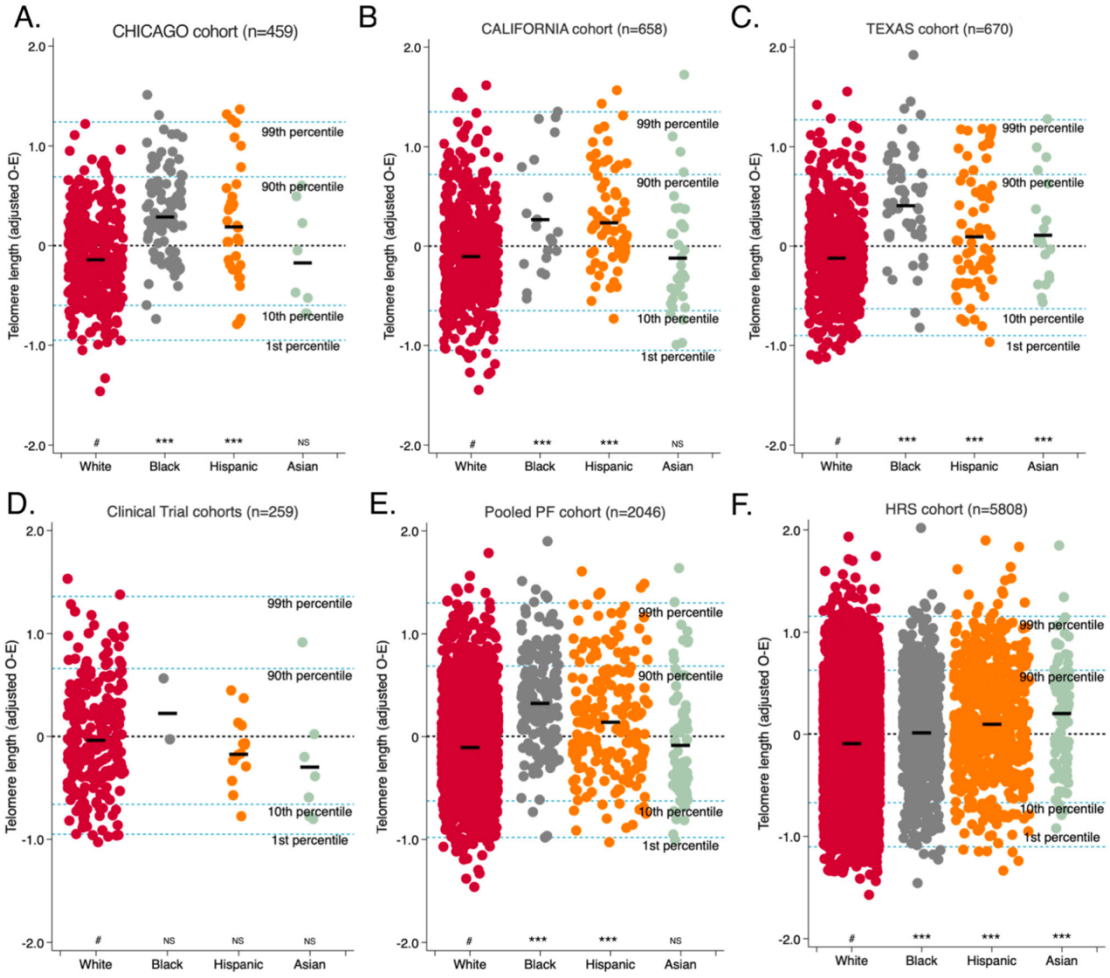


Males constituted 54% of the pulmonary fibrosis participant group and 41% of the HRS control group. Notably, the average body mass index (BMI) was relatively high in both pulmonary fibrosis participants (28 ± 6 kg/m^2) and control participants (29 ± 6 kg/m^2). The prevalence of tobacco use was strikingly similar between pulmonary fibrosis (58%) and control participants (57%). The median duration of follow-up for pulmonary fibrosis participants was 31 months, with an interquartile range (IQR) of 14-60 months. This in-depth participant profile provided a robust foundation for our subsequent analyses.

Telomere length across racial/ethnic groups

Among the pulmonary fibrosis participants, there was a distinct disparity in leukocyte telomere length across racial groups, with Black subjects exhibiting the longest LTL. The mean standardized leukocyte telomere length at pulmonary fibrosis diagnosis was significantly longer in Black participants (0.37 ± 0.49) compared to White participants (-0.06 ± 0.47) in the pooled pulmonary fibrosis cohort ($P < 0.0005$). This difference was maintained when the data was stratified by participant recruitment site. Similarly, among the HRS control participants, the mean standardized LTL was longer in Black participants (0.08 ± 0.48) as compared to White participants (-0.04 ± 0.49) ($P < 0.0005$). Notably, the discrepancy in LTL across racial groups was 3.6-fold more pronounced in the pulmonary fibrosis cohort than in the control group (**Figure 2.2**).

Figure 2.2. Leukocyte telomere length across racial groups with pulmonary fibrosis. Mean observed minus expected (O–E; age and gender-adjusted) leukocyte telomere length (TL) is longest in Black subjects with pulmonary fibrosis (PF). A. CHICAGO cohort, mean age and gender-adjusted TL(O–E) of –0·09(0·43), 0·34(0·46), 0·24(0·59), and –0·13(0·53), respectively. B. CALIFORNIA cohort, mean age and gender-adjusted TL(O–E) of –0·06(0·49), 0·32(0·56), 0·28(0·49), and –0·08(0·58), respectively. C. TEXAS cohort, mean age and gender-adjusted TL(O–E) of –0·07(0·46), 0·46(0·51), 0·14(0·57), and 0·16(0·55), respectively. D. IPFNet cohort, mean age and gender-adjusted TL(O–E) of 0·013(SD 0·50), 0·27(SD 0·41), –0·13(SD 0·33), and –0·25 (0·60), respectively. E. Pooled PF cohort, mean age and gender-adjusted TL(O–E) of –0·06(0·47), 0·37(0·49), 0·19(0·54), and –0·04(0·57), respectively. F. HRS cohort, mean age and gender-adjusted TL(O–E) of –0·04(0·49), 0·08(0·48), 0·16(0·51), and 0·19(0·53), respectively. Thick short black lines = subgroup median; black dotted line = cohort median TL. Group comparisons between white subjects (#) and other racial subgroups were conducted using the student's T-test; ***P < 0·0005; NS not significant (P ≥ 0·05).



Within the control cohort, Black, Hispanic, and Asian subjects displayed lower odds of having LTL below the median compared to White subjects (Black: OR = 0.7, 95% CI = 0.6–0.8, P < 0.001; Hispanic: OR = 0.5, 95% CI = 0.5–0.6, P < 0.001; Asian: OR = 0.6, 95% CI = 0.4–0.9, P = 0.015). A similar pattern was observed among the PF cohort, where Black and Hispanic subjects had lower odds of having LTL below the median (Black: OR = 0.2, 95% CI = 0.2–0.4, P < 0.001; Hispanic: OR = 0.5, 95% CI = 0.4–0.7, P < 0.001). However, there was no notable difference in Asian subjects compared to White subjects (OR = 0.9, 95% CI = 0.5–1.4, P = 0.55). Telomere length significantly differed between White and Black subjects (**Table 2.3**). In propensity-score matched analyses, the difference in mean LTL between Black and White subjects with pulmonary fibrosis became less pronounced (0.37 ± 0.49 vs. 0.33 ± 0.56, respectively). Finally, an examination of quartiles of standardized LTL stratified by race/ethnicity revealed a consistent trend of increasing mean LTL from the lowest quartile (Q1) to the highest quartile (Q4) across all racial groups.

Table 2.2 Mean Age and Telomere Lengths Differences Between White and Black Subjects.

Characteristic	WHITE (n=1613)	BLACK (n=162)	Absolute Mean Difference (Δ)	Delta ratio (95% CI)	*P*-value[#]
Mean TL[*]					
PF cohort	-0·06 (0·47)	0·37 (0·49)	0.43 (0.48)	3.58 (2.56 – 5.37)	< 0.001
HRS cohort	-0·04 (0·49)	0·08 (0·48)	0.12 (0.48)		
Age, years					
PF cohort	66.02 (10.40)	54.93 (12.20)	11.09 (36.74)	4.70 (3.37 – 7.17)	< 0.001
HRS cohort	70.19 (10.23)	67.83 (9.90)	2.36 (28.27)		

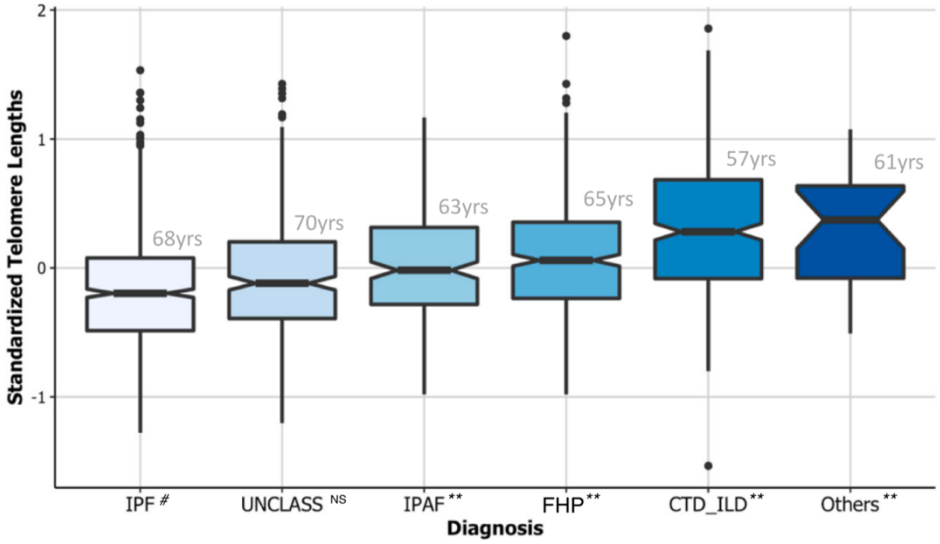
**Standardized mean observed minus expected (O–E) leukocyte telomere length (TL) across study population presented as means (SD).*
[#]P-value for the student's T-test comparing Δ between PF and HRS cohorts. PF=pulmonary fibrosis; HRS=Health Retirement Survey.

Sex, ILD subtypes, and Telomere length

The average unadjusted leukocyte telomere length consistently appeared shorter in males compared to females, a pattern that was evident across all pulmonary fibrosis cohorts as well as the Health and Retirement Study (HRS) control participants. This gender disparity manifested itself consistently across all examined racial and ethnic groups. There was a distinct variation between different subtypes of pulmonary fibrosis, both in their baseline demographics and in their standardized leukocyte telomere length measurements, post-adjustment for age and sex.

Participants diagnosed with idiopathic pulmonary fibrosis were older, with a median age of 68 years, and exhibited the shortest median leukocyte telomere length at -0.20 (interquartile range -0.49–0.08) (Fig. 2). In comparison, subjects with interstitial pneumonia accompanied by autoimmune features (median age 63 years; median leukocyte telomere length -0.02, interquartile range -0.29–0.32), fibrotic hypersensitivity pneumonitis (median age 65 years; median leukocyte telomere length 0.06, interquartile range -0.24–0.36), connective tissue disease-related interstitial lung disease (median age 57 years; median leukocyte telomere length 0.28, interquartile range -0.08–0.69), and other interstitial lung diseases (median age 61 years; median leukocyte telomere length 0.06, interquartile range -0.17–0.47) were generally younger and had longer leukocyte telomere lengths ($P < 0.0005$).

Figure 2.3. Leukocyte telomere length across clinical phenotypes of pulmonary fibrosis



In contrast, participants diagnosed with the unclassifiable subtype of pulmonary fibrosis (median age 70 years; median leukocyte telomere length -0.13, interquartile range -0.42–0.17) displayed leukocyte telomere length measurements that did not significantly differ from those with idiopathic pulmonary fibrosis (P = 0.20) (Fig. 2). Across pulmonary fibrosis subtypes, White subjects consistently exhibited the shortest leukocyte telomere lengths.

Telomere length and chronological age

The correlation between leukocyte telomere length and age revealed a decreasing pattern as age increased, in both pulmonary fibrosis patients from all participating institutions and Health and Retirement Study (HRS) control subjects (**Table 2.4**). This

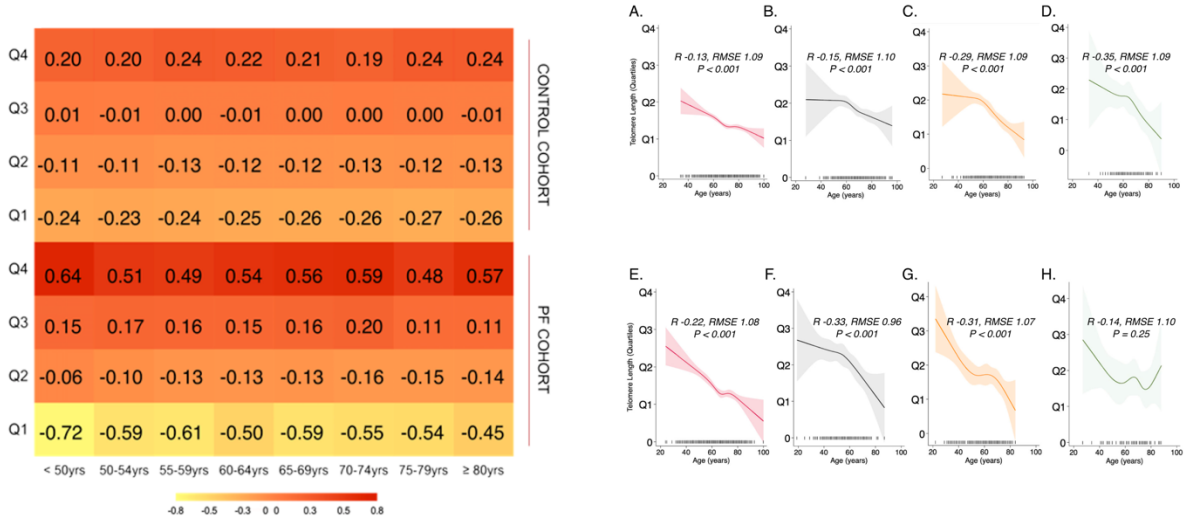
Table 2.3 Shorter telomere length occurs with increasing chronological age. Stratification by age group demonstrates shorter mean and median leukocyte telomere length with increasing chronological age among pulmonary fibrosis cohorts and control subjects.

Characteristics*	Mean Telomere Length (SD)	Median Telomere Length (IQR)
CHICAGO (n=459)		
Less than 40yrs *(n=19)*	0.59 (0.40)	0.66 (0.63)
40 – 49yrs *(n=46)*	0.30 (0.40)	0.20 (0.50)
50 – 59yrs *(n=101)*	0.11 (0.47)	0.14 (0.44)
60 – 69yrs *(n=158)*	-0.06 (0.48)	-0.02 (0.63)
70yrs or older *(n=135)*	-0.20 (0.46)	-0.22 (0.54)
CALIFORNIA (n=658)		
Less than 40yrs *(n=12)*	0.52 (0.34)	0.45 (0.66)
40 – 49yrs *(n=45)*	0.19 (0.57)	0.17 (0.53)
50 – 59yrs *(n=91)*	0.04 (0.50)	0.12 (0.57)
60 – 69yrs *(n=215)*	0.01 (0.47)	0.06 (0.60)
70yrs or older *(n=295)*	-0.07 (0.50)	-0.10 (0.65)
TEXAS (n=670)		
Less than 40yrs *(n=31)*	0.37 (0.58)	0.42 (0.69)
40 – 49yrs *(n=61)*	0.32 (0.45)	0.39 (0.49)
50 – 59yrs *(n=148)*	0.06 (0.49)	0.11 (0.62)
60 – 69yrs *(n=240)*	-0.05 (0.49)	-0.03 (0.59)
70yrs or older *(n=190)*	-0.15 (0.45)	-0.17 (0.61)
IPFNET (n=259)		
Less than 40yrs *(n=0)*	---	---
40 – 49yrs *(n=3)*	0.52 (0.72)	0.88 (1.30)
50 – 59yrs *(n=37)*	-0.05 (0.67)	-0.00 (0.68)
60 – 69yrs *(n=114)*	-0.01 (0.50)	0.03 (0.51)
70yrs or older *(n=105)*	0.02 (0.42)	0.02 (0.55)
HRS (n=5808)		
Less than 40yrs *(n=14)*	0.02 (0.19)	-0.01 (0.21)
40 – 49yrs *(n=93)*	0.05 (0.27)	0.02 (0.25)
50 – 59yrs *(n=1056)*	0.04 (0.43)	-0.02 (0.26)
60 – 69yrs *(n=1766)*	0.01 (0.46)	-0.06 (0.26)
70yrs or older *(n=2879)*	-0.02 (0.55)	-0.09 (0.25)

*Total sample size, n=7,854. *Standardized telomere length across study population presented as means (SD). Telomere lengths unadjusted for age or sex.*

negative correlation between chronological age and leukocyte telomere length persisted across racial groups in both the HRS control subjects and pulmonary fibrosis participants. When assessing leukocyte telomere length quartiles, we observed an anticipated increment in average leukocyte telomere length measurements from the first quartile (Q1) to the fourth quartile (Q4) across age categories stratified by five-year intervals. Pulmonary fibrosis participants, however, displayed a broader interquartile range for all age categories compared to the HRS control participants (**Figure 2.4**).

Figure 2.4. Telomere length interquartile range is wider in pulmonary fibrosis. Standardized leukocyte telomere length measured by qPCR demonstrates wider interquartile range variation in pulmonary fibrosis (PF) and decreases with increasing age. TL demonstrates a nonlinear negative correlation with age across diverse racial groups in healthy subjects A. Whites; B. Blacks; C. Hispanics; and D. Asians; that is altered in subjects with pulmonary fibrosis; E. Whites; F. Blacks; G. Hispanics; and H. Asians.



Upon standardization, leukocyte telomere length demonstrated a negative association with chronological age across White, Black, Hispanic, and Asian HRS control participants. This mild negative association amplified across White, Black, Hispanic, and Asian pulmonary fibrosis subjects, with the most substantial correlation observed among Black pulmonary fibrosis participants (**Table 2**).

Table 2.4 Models depicting the association of leukocyte telomere length (LTL) with age and mortality across diverse racial populations.

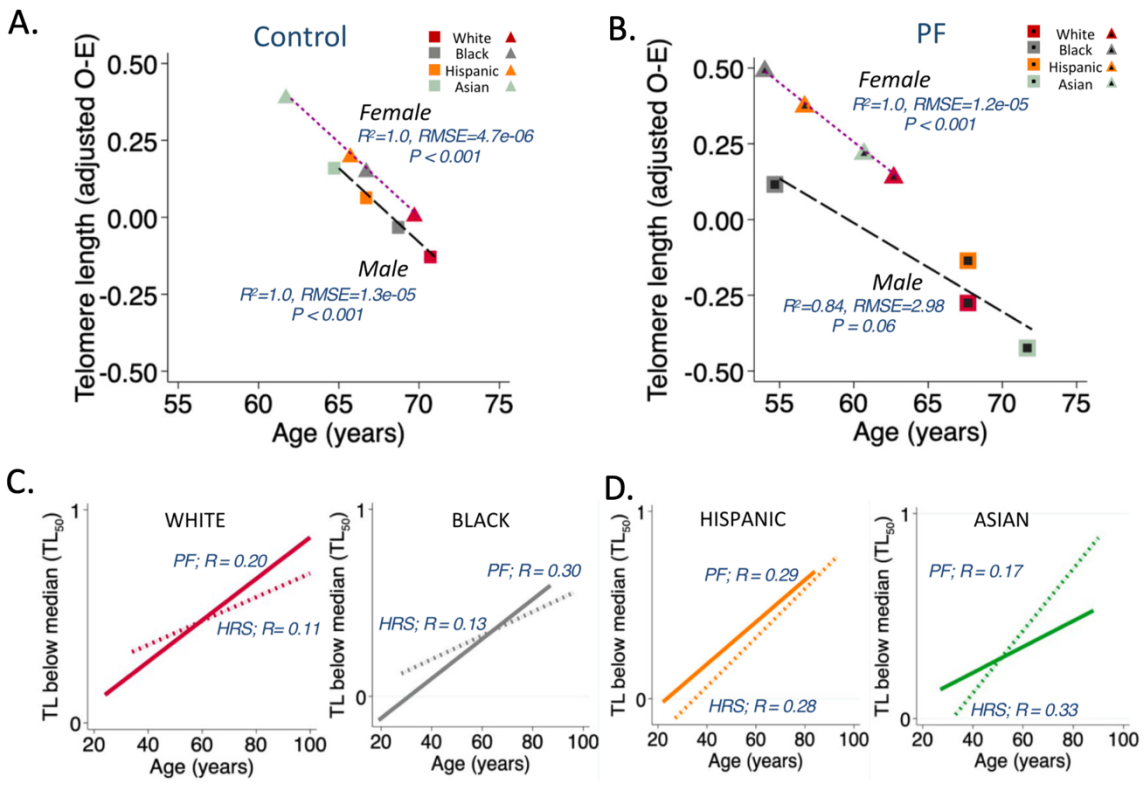
Characteristics	White	Black	Hispanic	Asian	Combined
Regression models for Age*					
Unadjusted LTL					
PF Cohort	(*n* = 1613)	(*n* = 162)	(*n* = 187)	(*n* = 70)	(*n* = 2046)
All, *R* (Root MSE)	−0.21 (0.27)	−0.29 (0.24)	−0.39 (0.28)	−0.17 (0.29)	−0.30 (0.28)
95% CI	−0.26 to −0.17	−0.43 to −0.14	−0.51 to −0.26	−0.40 to 0.07+	−0.33 to −0.26
HRS Cohort	(*n* = 4319)	(*n* = 779)	(*n* = 614)	(*n* = 96)	(*n* = 5808)
All, *R* (Root MSE)	−0.04 (0.65)	−0.02 (1.11)	−0.12 (0.59)	−0.07 (0.63)	−0.05 (0.73)
95% CI	−0.07 to −0.01	−0.10 to −0.05	−0.42 to −0.05	−0.26 to −0.14	−0.08 to −0.03
Standardized LTL**					
PF cohort	(*n* = *1613*)	(*n* = 162)	(*n* = 187)	(*n* = 70)	(*n* = 2046)
All, *R* (Root MSE)	−0.22 (1.08)	−0.33 (0.96)	−0.31 (1.07)	−0.14 (1.10)	−0.28 (1.08)
95% CI	−0.27 to −0.17	−0.46 to −0.18	−0.44 to −0.18	−0.36 to 0.10+	−0.32 to −0.24
HRS cohort	(*n* = 4319)	(*n* = 779)	(*n* = 614)	(*n* = 96)	(*n* = 5808)
All, *R* (Root MSE)	−0.13 (1.09)	−0.15 (1.10)	−0.29 (1.09)	−0.35 (1.09)	−0.17 (1.10)
95% CI	−0.16 to −0.10	−0.22 to −0.09	−0.36 to −0.21	−0.51 to −0.16	−0.19 to −0.14
PF mortality risk models	(*n* = 1613)	(*n* = 162)	(*n* = 187)	(*n* = 70)	(*n* = 2046)
#Overall crude mortality rate	8.92	5.27	9.62	6.62	8.49
(95% CI)	(8.16–9.76)	(3.91–7.10)	(7.47–12.38)	(3.84–11.40)	(7.83–9.20)
+Mortality incidence rate ratio, TL_{50}	2.10	2.09	2.79	2.67	2.31
(95% CI)	(1.71–2.59)	(1.07–4.07)	(1.66–4.69)	(0.89–8.03)	(1.93–2.78)
*Mortality hazard ratio for TL_{50}	2.21	2.22	3.40	2.11	2.47
(95% CI)	(1.79–2.72)	(1.05–4.66)	(1.88–6.14)	(0.54–8.23)	(2.05–2.97)

R Pearson's bivariate correlation coefficient. *Root MSE* root mean squared error. *PF* pulmonary fibrosis. *HRS* health retirement survey.
*Hazard Ratio for TL50 in multivariable Cox regression models adjusting for forced vital capacity, diffusing capacity of the lungs for carbon monoxide, interstitial lung disease subtype, and hospital center. For multivariable models, White *n* = 1534, Black *n* = 145, Hispanic *n* = 175, Asian *n* = 64, All (including patients with race categorized as other, *n* = 14) *n* = 1932.*$P < 0.001$ for all regression models except where denoted by +. **Standardized telomere lengths in quartiles.
#Overall crude mortality rate computed per 100 person-yrs. #*P* value for Mantel–Haenszel test statistic.
+Mortality incidence rate ratios in subjects with age- and gender-adjusted leukocyte telomere length below the median (TL_{50}) were estimated using a generalized linear model with a Poisson distribution and logistic regression link adjusting for age, gender, forced vital capacity, diffusing capacity of the lungs for carbon monoxide, and hospital center.

The assessment of median leukocyte telomere length within each substratum of race and sex exhibited a strong correlation across racial and ethnic groups for the age-telomere length relationship in both male and female HRS control subjects ($R^2 = 1.0$; $P < 0.001$) (**Figure 2.5**). The presence of pulmonary fibrosis disrupted this uniform linear relationship across racial and ethnic groups, leading to a non-significant correlation among male pulmonary fibrosis participants ($R^2 = 0.84$; $P = 0.06$). In these patients, leukocyte telomere length was shortest in older White or Asian males. Hispanic ethnicity notably impacted the age-leukocyte telomere length relationship among controls (interaction term $P = 0.027$), whereas race did not. In contrast, in pulmonary fibrosis, Hispanic ethnicity, as well as White and Black race, significantly influenced the age-leukocyte telomere length relationship (interaction term $P < 0.001$), while Asian race did not (interaction term $P = 0.25$). The overall age/leukocyte telomere length relationship between pulmonary fibrosis and HRS control subjects across racial and ethnic categories demonstrated that pulmonary fibrosis exerted an independent effect on this relationship, as evidenced by significant interaction P values.

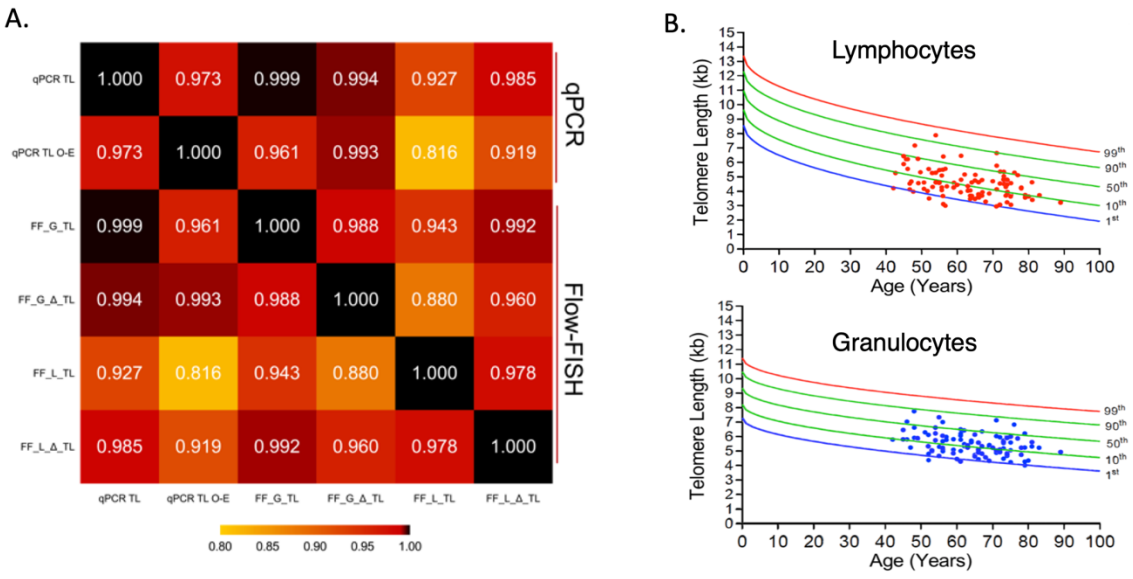
Our analysis of the association between leukocyte telomere length and chronological age, using age- and gender-adjusted leukocyte telomere length below the median (TL50), confirmed the modest association across all racial and ethnic subgroups of the HRS population ($R = 0.11–0.33$) and among pulmonary fibrosis participants ($R = 0.17–0.30$) (Table 2).

Figure 2.5. qPCR telomere length decreases with increasing age in pulmonary fibrosis. Median TL for each race (White *red*, Black *gray*, Hispanic *orange*, and Asian *green*) stratified by sex (male = squares, female = triangles) among the A. Control population (HRS); and B. PF population); statistical test: generalized linear regression of sex-stratified median TL across racial/ethnic groups. R^2 (the coefficient of determination), RMSE (root mean squared error), and P value reported for each subgroup. C. Shorter TL below the median (TL50) has greater prevalence with increasing age across all racial groups in the HRS control population (dashed lines), but this association was stronger in White and Black subjects with PF (solid lines). Subjects with PF (White n = 1613, Black, n = 162, Hispanic, n = 187, Asian, n = 70). Control subjects (White n = 4319, Black, n = 779, Hispanic, n = 614, Asian, n = 96). Other racial groups [n = 14] are not included in the graphs above. TL depicted are observed minus expected (O–E; age and gender-adjusted). Purple dotted line = fitted values for females and black dashed line = fitted values for males.



In a sub-cohort of 13 patients with ILD, we validated our qPCR measurements against telomere length (kb) measured in peripheral blood lymphocytes and granulocytes via flow-cytometry-based fluorescent in situ hybridization (FISH) by the Department of Pathology at Johns Hopkins Hospital. T/S ratios compared to granulocyte and lymphocyte telomere lengths measured using flow-FISH showed significant correlation ($p<0.01$) for both comparisons. (**Figure 2.6**)

Figure 2.6 Leukocyte subset qPCR values measurements of telomere length correlate highly with values from flow-cytometry-based fluorescent in situ hybridization (Flow-FISH).

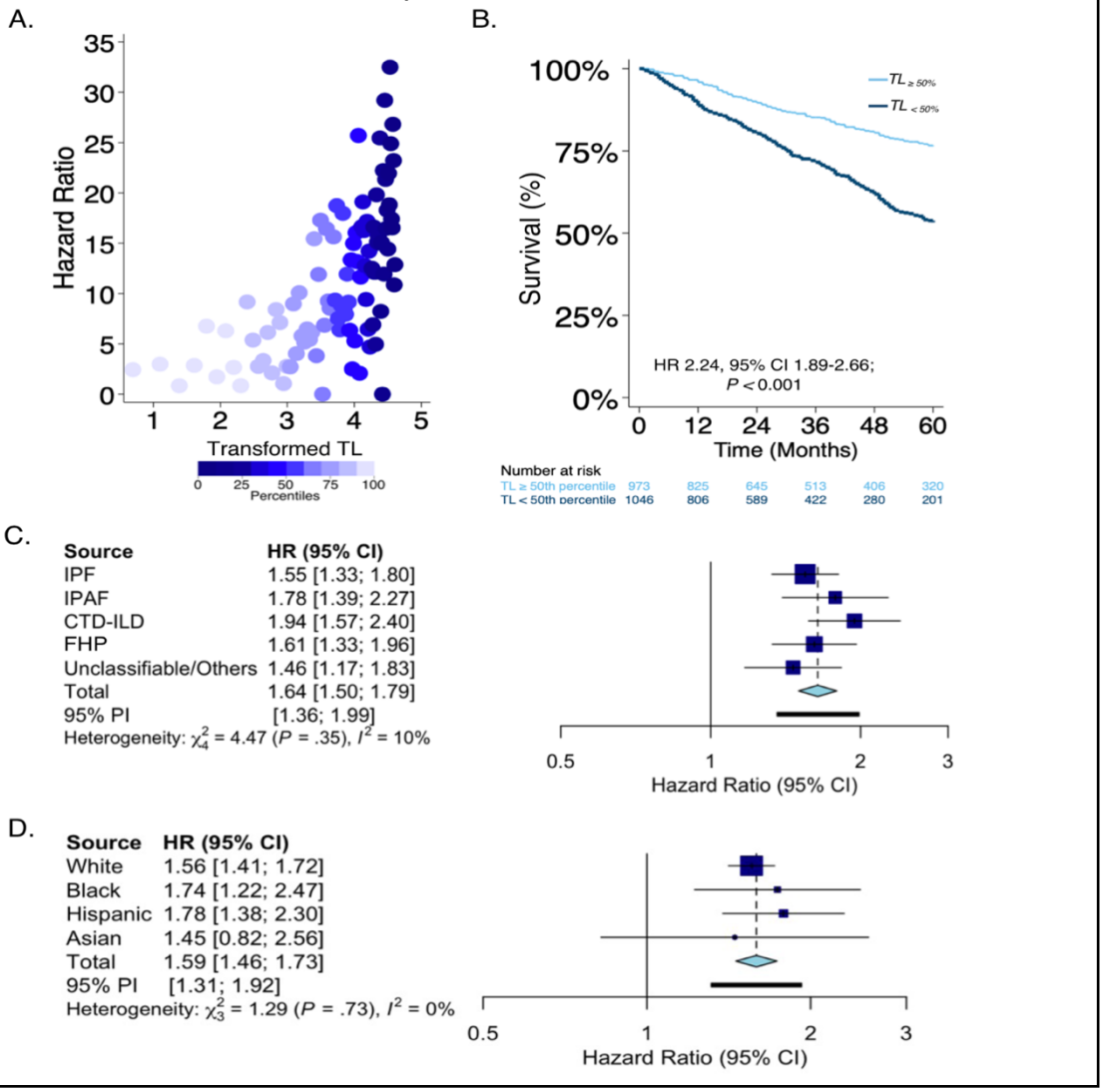


Telomere length and mortality

Pulmonary fibrosis participants displayed varying rates of absolute mortality and mortality incidence ratios, with those having age- and gender-adjusted leukocyte telomere length below the median (TL50) manifesting the lowest rates among Black participants and the highest among Hispanic participants (Table 2). As the severity of pulmonary fibrosis increased, reflected by the ILD-GAP score, leukocyte telomere length observably shortened across all racial and ethnic pulmonary fibrosis subgroups (data not shown).

In the consolidated pulmonary fibrosis population, standardized leukocyte telomere length, considered as a continuous measure, displayed an association with mortality, with shorter telomere length consistently predicting an elevated mortality risk (**Figure 2.7**). Patients with leukocyte telomere length below the median (TL_{50}) exhibited worse survival outcomes across all pulmonary fibrosis subtypes, racial or ethnic groups (**Figure 2.8**), and study locations (**Figure 2.9**) when compared with participants who had telomere length at or above the median ($TL_{\geq 50}$) (Supplementary Figs. 3–5, Adegunsoye A et al. Nature Communications 2023). This trend remained consistent, even after adjusting for age, sex, and disease severity. Each decrease in telomere length quartile was linked to an increased risk of death in pulmonary fibrosis participants, this trend held true for each subtype of pulmonary fibrosis and across the different racial and ethnic groups (**Figure 2.10**).

Figure 2.7 Shorter leukocyte telomere length (TL) consistently predicts worse survival patterns in pulmonary fibrosis (PF). *A*. Scatter plot of mortality hazard ratios (HR)* in PF by transformed TL comparing each centile of TL to the highest TL centile. The plot depicts increasing mortality hazard with shorter TL. *B*. Survival stratified by age and gender-adjusted TL below the median (TL<50%) vs. above the median (TL≥50%) in the PF cohort. Fixed effect mortality hazard estimates for quartiles of leukocyte telomere length (TLQ) adjusted for age, gender, FVC, DLCO, ILD subtype, and hospital center categorized by *C*. PF subtype, and *D*. race/ethnicity. HR depicted per quartile increase in TL. White n = 1613, Black n = 162, Hispanic n = 187, Asian n = 70, others n = 14, All patients n = 2046. The navy-blue boxes within the forest plot represent the point estimate for the Cox mortality hazard ratio for each cohort, the thin horizontal line represents its 95% confidence interval, the vertical line is the line of no effect, and the diamond represents the overall effect estimate.



Multivariable Cox proportional hazards models, adjusted for age, sex, forced vital capacity, diffusing capacity of the lungs for carbon monoxide, pulmonary fibrosis subtype, and study site, revealed that having a telomere length below the median (TL<50) was independently associated with a higher risk of death among pulmonary fibrosis subjects. This pattern maintained its consistency across different racial and ethnic groups (Hispanic, HR = 3.40, 95% CI = 1.88–6.14; Asian, HR = 2.11, 95% CI = 0.54–8.23; White, HR = 2.21 95% CI = 1.79–2.72; and Black participants, HR = 2.22, 95% CI = 1.05–4.66).

Figure 2.8 Telomere length below the 50th percentile predicts worse survival in pulmonary fibrosis across racial groups. Survival pattern by TL in (A) White subjects; (B) Black subjects; (C) Hispanic subjects; and (D) Asian subjects.

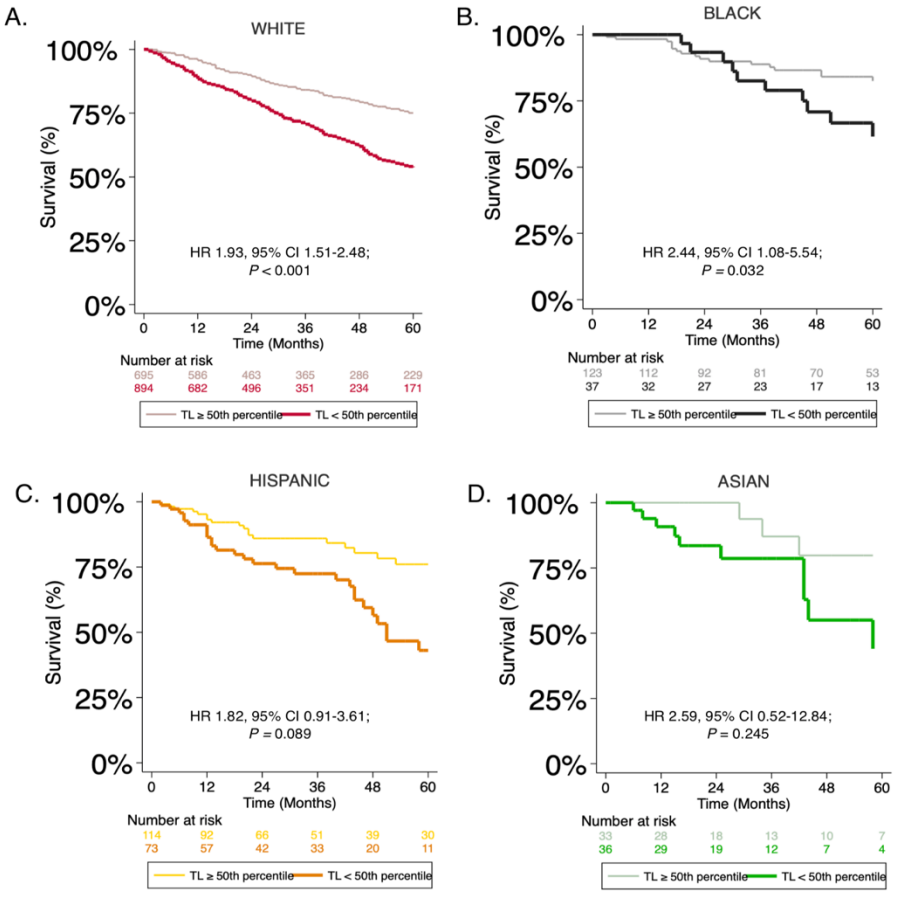


Figure 2.9 Telomere shortening below the median predicts worse pulmonary fibrosis survival across geographically disparate study locations. Survival stratified by age and gender-adjusted TL below the median (TL<50%) vs. above the median (TL≥50%) in (A) CHICAGO cohort (n=459); (B) CALIFORNIA cohort (n=658); (C) TEXAS cohort (n=670); (D) IPFNet cohort (n=259). (E) Plot of p-values for the association of mortality hazard ratios (HR) in PF with transformed TL (negative log-transformed inverse of one minus percentile TL); HR depicted for decreasing centiles of TL compared to highest centile of TL. (F) Survival stratified by age and gender-adjusted TL in quartiles for each center; TOTAL (n=2046).

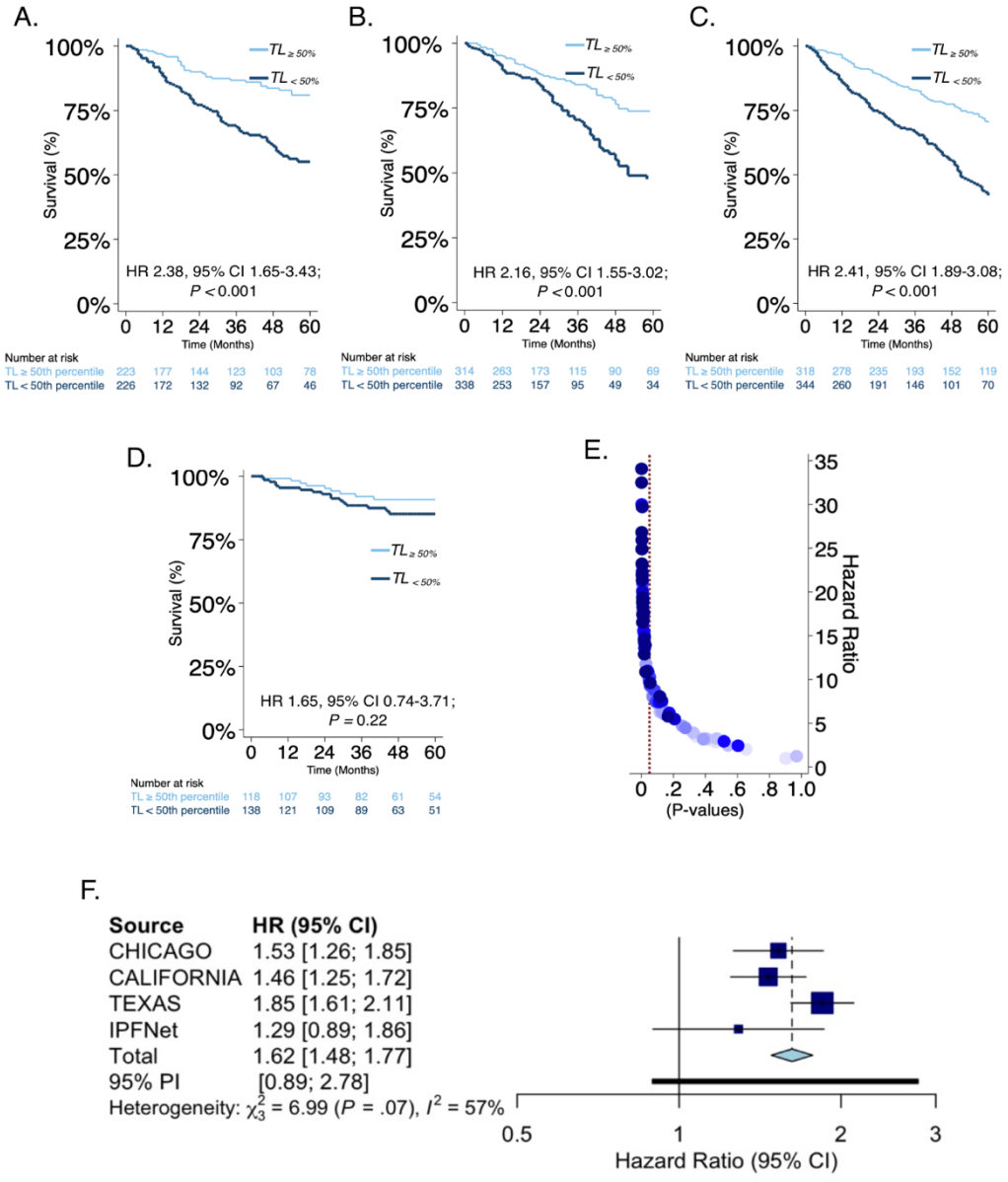
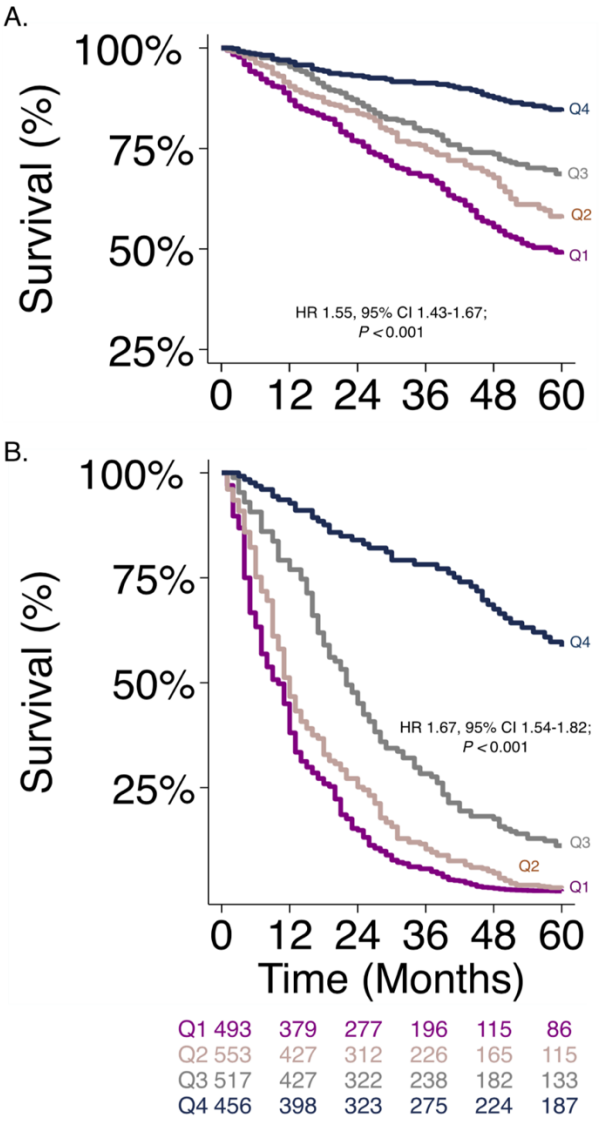


Figure 2.10 Quartiles of telomere length predict survival. A. Kaplan–Meier survival curve according to age and gender-adjusted leukocyte telomere length in quartiles (TLQ). B. Kaplan–Meier survival curves according to TLQ adjusted for age, gender, FVC, DLCO, ILD subtype, and hospital center. Cox proportional hazard ratio (HR) and 95% confidence interval of this estimate are depicted with the respective *P*-value in each plot.



Discussion

These genomic investigations into pulmonary fibrosis unveiled the crucial role leukocyte telomere length plays and its correlation with racial/ethnic disparities in patients. In surveying these genomic markers across a diverse pool of patients, I identified unique variations in telomere length and age at diagnosis across racial and ethnic groups. This exploration offers a novel perspective on pulmonary fibrosis etiopathogenesis, highlighting the necessity to factor in genetic elements when managing patients.

In patients with pulmonary fibrosis, I showed that short leukocyte telomere lengths are common and are associated with worse survival across all racial and ethnic groups. A particularly intriguing finding was the distinct racial and ethnic differences in the extent of leukocyte telomere length shortening and its correlation with chronological age. For example, Black subjects presented a 3.6-fold increase in the odds of having longer leukocyte telomere lengths and were diagnosed with pulmonary fibrosis over a decade younger than their White counterparts. This implies that the phenomenon of leukocyte telomere length shortening, associated with pulmonary fibrosis, doesn't occur uniformly across all racial and ethnic groups, potentially reflecting the chronological age at the time of pulmonary fibrosis diagnosis.

My comprehensive study demonstrated that the linear relationship between chronological age and telomere length in pulmonary fibrosis markedly diverges from healthy controls. This discrepancy might suggest that an acceleration of biological aging

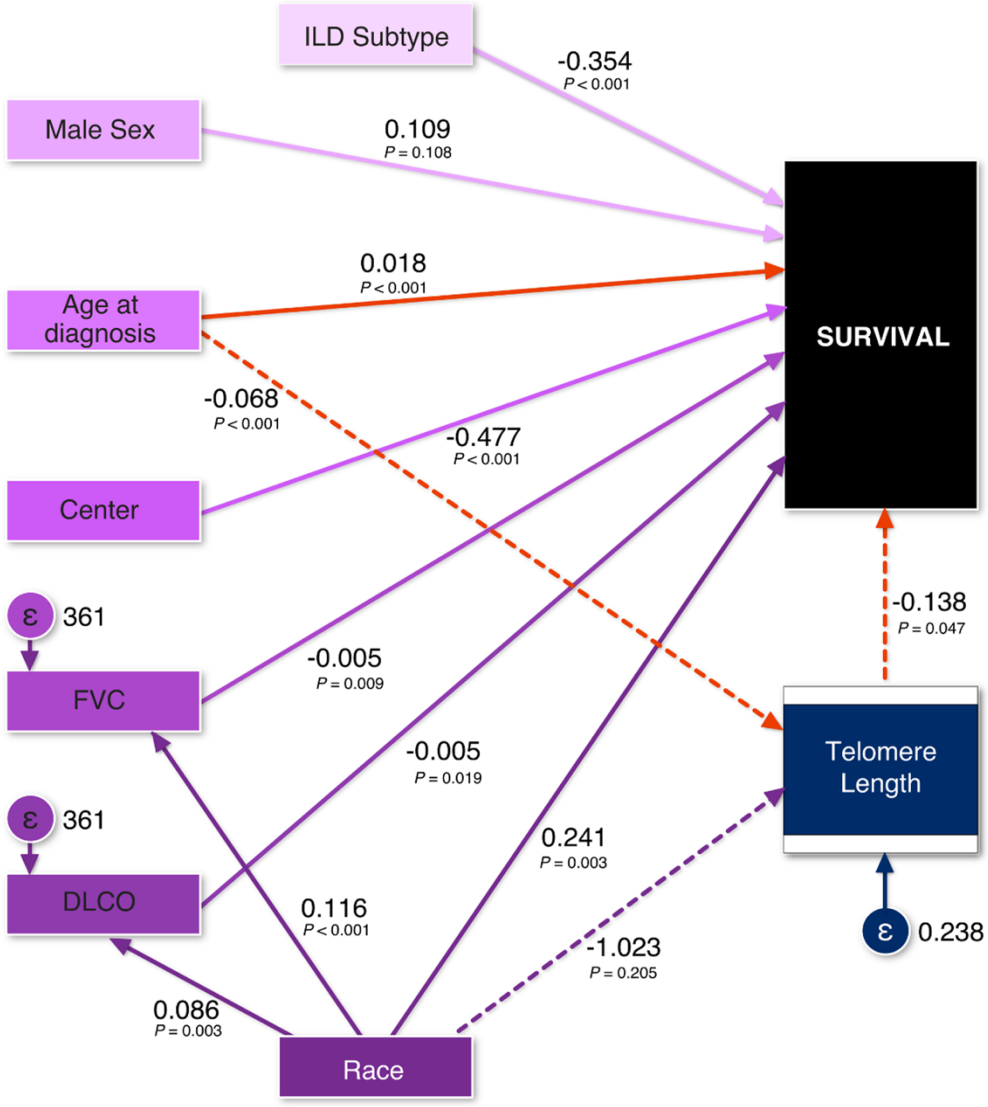
processes, early senescence, and cellular apoptosis contribute to the pathobiologic mechanisms underlying pulmonary fibrosis. Furthermore, the shortest leukocyte telomere lengths were present in older male Asian and White individuals with pulmonary fibrosis, where IPF was the most prevalent condition. This supports the theory that telomere shortening could be a causal factor for IPF, a condition that frequently manifests in families harboring pathogenic mutations in telomere-maintenance genes and short leukocyte telomere lengths [61-63]. However, when observing the Black participant group, it was noted that they exhibited less significant telomere shortening and developed lung disease at an earlier chronological age. This finding implies that aberrations in telomere homeostasis might contribute less significantly to pulmonary fibrosis development in this particular demographic. Pathological disruptions in telomere homeostasis, such as mutations in telomerase RNA component (*TERC*) or telomerase reverse transcriptase (*TERT*) genes, could result in defects within the telomerase complex. Over time, this can lead to the onset of pulmonary fibrosis characterized by extremely shortened telomeres. Conversely, it's noteworthy to mention that autoimmune conditions and most other pulmonary fibrosis causes do not typically result in such dramatically shortened telomeres. This differential effect underscores the complexity and diversity of factors contributing to pulmonary fibrosis across different populations. These findings indicate that gene-by-environment factors might have an influence on telomere length, which could explain why leukocyte telomere lengths in Black patients are not as profoundly shortened at disease onset compared to White patients.

This research also revealed a high prevalence of autoimmune-associated causes of pulmonary fibrosis among Black subjects, underscoring the varying contribution of lung stressors in the etiopathogenesis of pulmonary fibrosis across racial backgrounds. Non-hereditary factors such as inflammatory disease states, sporadic autoimmune conditions [64-66], toxic inhalation of environmental antigens [67], air pollution [5, 68], and various socioeconomic determinants of health disproportionately impact Black individuals, adding another layer of complexity to the observed racial disparities in pulmonary fibrosis.

In this study, I was able to establish a strong association of leukocyte telomere length with chronological age across racially and ethnically diverse pulmonary fibrosis cohorts. However, my findings also unveiled age-specific differences in the association of leukocyte telomere length with the risk of lung function impairment and specific pulmonary fibrosis subtypes across diverse races. This suggests that the clinical predictors of respiratory impairment might differ based on the age stratum, thereby informing targeted treatment strategies.

Despite the variation of leukocyte telomere lengths in this diverse pulmonary fibrosis population, my research revealed that shorter leukocyte telomere lengths are uniformly linked to worsened survival across all races, irrespective of the pulmonary fibrosis subtype. While genetic data predominantly stem from calculations derived from a single race, leukocyte telomere lengths hold predictive value for mortality even after adjusting for known confounding variables (**Figure 2.11**). This validation endorses the use of leukocyte telomere lengths as a valuable prognostic biomarker applicable to all racial and ethnic groups, enhancing the delivery of personalized medicine to patients.

Figure 2.11 Mediation analyses of effects of age and race on telomere length, and on pulmonary fibrosis survival.



In my investigation, I demonstrated that among patients with pulmonary fibrosis, leukocyte telomere length measurements offer considerable prognostic significance for mortality, even after adjustments are made for established confounding variables. The research demonstrated a “dose-dependent” relationship, whereby shorter leukocyte telomere lengths (categorized by quartiles) were associated with increased mortality

across different racial and ethnic groups. Moreover, a leukocyte telomere length falling below the median value proved to be a valuable prognostic biomarker applicable to all racial and ethnic demographics. This discovery holds significant clinical relevance as identifying individuals at heightened risk, those with shorter telomeres regardless of their racial or ethnic origins, could enable us to inform pharmacological strategies and guide precision disease management. By doing so, we can significantly enhance the delivery of personalized medicine to Pulmonary Fibrosis patients, thereby improving outcomes and the quality of life for patients.

This investigation, though insightful, had several constraints. Primarily, the study's cross-sectional nature inhibits longitudinal assessment of leukocyte telomere length dynamics. Moreover, the focus on race/ethnicity, a complex blend of perceived origin, culture, and genetics, can introduce variability. The method of measurement employed was qPCR, allowing analysis on archived samples but providing only average leukocyte telomere length. It is possible that shorter telomeres could be linked to mortality from pulmonary fibrosis comorbidities, as physiological age tied to telomere length may drive mortality more significantly than chronological age. Unfortunately, due to the study's retrospective design, the precise cause of death for participants remained uncertain. Although z-score standardization allowed within-group racial/ethnic differences in standardized leukocyte telomere length, the methodological heterogeneity precluded direct comparison of absolute leukocyte telomere lengths between cases and controls. Furthermore, treating all Asian participants as a single group, despite clear differences in socioeconomic position, culture, and immigration history among various Asian ethnicities,

may have introduced less precision. Lastly, residual confounding variables influencing leukocyte telomere lengths may have persisted, even after adjustments. Nevertheless, the large sample size and diverse geographic regions should mitigate these effects.

In conclusion, this innovative research offers unique insights into the etiopathogenesis of pulmonary fibrosis across racial and ethnic groups, providing a roadmap for future studies. These investigations could provide greater insight into the potential value of hormonal-based interventions in pulmonary fibrosis, looking to ameliorate telomere loss or impact clinically relevant outcomes, thus paving the way for more personalized and effective treatment strategies.

Chapter III

Prognostic and theragnostic impact of telomere length in fibrotic hypersensitivity pneumonitis

Introduction

Recent advances in clinical trials have demonstrated a significant impact of antifibrotic therapies in decelerating lung function decline in patients suffering from IPF and progressive fibrosing interstitial lung disease (PF-ILD)[69]. These promising results are replicated even in the case of scleroderma-associated ILD, despite the concurrent application of the immunosuppressive therapy mycophenolate mofetil [70].

However, the therapeutic landscape for other forms of ILD such as fibrotic hypersensitivity pneumonitis remains somewhat barren, with a distinct lack of prospective data available. My prior research suggested the potential efficacy of mycophenolate mofetil in treating fibrotic hypersensitivity pneumonitis, with findings pointing towards a decrease in adverse events, a reduction in the required prednisone dosage, and overall improvements in lung function when compared to the sole use of prednisone [71]. Nevertheless, this premise demands further exploration through prospective studies. Intriguingly, my previous research also illustrated a correlation between short leukocyte telomere length and a higher mortality rate among patients with ILD [26]. Further, another recent investigation presented evidence suggesting that leukocyte telomere length might have a notable influence on the patient's response to immunosuppressive therapy [57]. In that study, it was found that patients suffering from IPF with shorter leukocyte telomere

length experienced a heightened risk of death, the need for lung transplantation, and decline in their forced vital capacity when exposed to immunosuppressive therapy, which included mycophenolate mofetil.

Taking into consideration these intriguing findings, I sought to probe whether similar outcomes were observable in patients diagnosed with fibrotic hypersensitivity pneumonitis [60]. Consequently, I hypothesized that patients with chronic hypersensitivity pneumonitis and exhibiting short leukocyte telomere length would demonstrate a higher prevalence of both death and disease progression when compared to their counterparts with longer leukocyte telomere length. As the exploration of the relationships between telomere length, genetic factors, and treatment response continues, our hope is that our findings will contribute to the knowledge pool and eventually lead to more personalized and effective treatment strategies for individuals living with chronic hypersensitivity pneumonitis and other interstitial lung diseases.

Methods

Study design, age, and pharmacotherapeutic categorization

The study population consisted of a diverse group of participants from multiple centers. This cohort was prospectively enrolled, and comprised of consenting individuals who had a confident multidisciplinary diagnosis of chronic hypersensitivity pneumonitis. Our participants were from four institutions; the University of Chicago, University of California San Francisco, University of California Davis, and University of Texas Southwestern, Dallas. They were enrolled between September 2003 and December 2019, thus providing

us with a long-range, dynamic dataset. All procedures involved were approved by the relevant institutional review boards to ensure ethical compliance. Genomic DNA was extracted from peripheral blood leukocytes of the patients, and telomere lengths in these DNA samples was measured using quantitative PCR, performed in triplicate to ensure reliability [72]. The telomere length was adjusted for age using data from normal controls, creating a baseline for comparison across our diverse set of patients. Additionally, I took into account potential variations across different study sites by standardizing telomere lengths through normalization, ensuring that our data remained consistent and valid regardless of the source. To further delineate our data, I categorized the standardized telomere lengths into quartiles. This categorization was crucial for understanding the distribution and trends within our study cohort and allowed us to more effectively test our hypothesis regarding the role of telomere length in chronic hypersensitivity pneumonitis outcomes.

To glean the required clinical information about our study participants, I used their electronic medical records. These records were instrumental in providing us with an in-depth understanding of each patient's medical history and current health status. I also used these records to confirm the vital status of our participants, supplementing this information with the US Social Security death index for verification. In order to maintain the integrity and specificity of our study, I excluded any patients who had received azathioprine either prior to or during the study period, as this could potentially confound my results. This step led to the exclusion of 19 participants from our analysis. I applied a binary categorization to our study population based on their pharmacotherapeutic

regimen. Specifically, I differentiated between those who had been on a mycophenolate mofetil therapy of 500 mg or more per day for at least a month during the study period, and those who had not. This categorization was crucial in allowing me to understand the potential effects of this therapy on telomere length and the outcomes of patients with chronic hypersensitivity pneumonitis.

Statistical analyses

I utilized a propensity score approach, a powerful statistical tool that helped me predict the conditional probability of a patient receiving mycophenolate mofetil (MMF) treatment. This methodology ensured a balanced comparison and helped in managing potential confounding factors that could bias my results. Variables considered in the propensity score model included demographic factors such as age and sex, as well as smoking status and the use of prednisone therapy. To delve deeper into disease severity, I incorporated several physiological indices into our model. These indices included the forced vital capacity (FVC) and the diffusing capacity of the lung for carbon monoxide (DLCO), which are widely recognized as critical measures of lung function. I also considered the severity and distribution of fibrosis, the presence of usual interstitial pneumonia patterns, the degree of traction bronchiectasis, and the appearance of ground glass opacities and mosaic attenuation.

I utilized inverse probability of treatment weighting (IPTW) to estimate the average effect of treatment on time-to-event outcomes [73]. This allowed me to neutralize any bias due to confounding variables and helped create an "apples-to-apples" comparison across

different groups of patients. To visualize these results, I plotted survival functions using the Kaplan-Meier estimator and employed Cox proportional-hazards models to estimate hazard ratios. I computed the transplant-free survival time as the time from the start of immunosuppressive therapy to the occurrence of a significant event such as death, lung transplantation, loss to follow-up, or the end of the study period. In all multivariable outcome models, I adjusted for variables that were imbalanced across groups and controlled for differences between centers. In instances where missing data was encountered, I applied a statistical technique known as multiple imputation using chained equations. This allowed me to create a complete dataset for our analyses. I limited this approach to covariates with less than 20% missing data to maintain the robustness of these results.

In this investigation, I further included a comprehensive longitudinal analysis of the percentage predicted values for FVC and DLCO using linear mixed-effects models with restricted maximum likelihood modelling and an autoregressive structure. To align the time-course in all patients, I grouped pulmonary function tests into 90-day epochs. This enabled me to examine the evolution of FVC and DLCO over time and to assess their association with disease progression and outcomes. These statistical analyses were completed using Stata, a comprehensive software package that provided the necessary tools for managing, analyzing, and graphing the data.

Results

Our study focused on a cohort of 208 patients diagnosed with fibrotic hypersensitivity pneumonitis. Of these, 19 were excluded due to the receipt of azathioprine before or during the investigation period, resulting in a final group of 189 patients. This diverse cohort, comprised of individuals from various backgrounds, served as a solid foundation for the rigorous analysis that followed. The cohort consisted of a relatively balanced mix of sexes, with females representing 51% of the total. Predominantly, the cohort was white (85%), with a median age of 65 years, and a substantial percentage (47%) had a history of tobacco use. Regarding the physiological indices of disease severity, baseline measures of FVC and DLCO, both as percentages of the predicted values, stood at 65% and 53%, respectively. Moreover, around 52% of patients had undergone a surgical lung biopsy, and 75% received corticosteroids, reinforcing the severity of the disease in the cohort (**Table 3.1**).

Table 3.1 Baseline characteristics of fibrotic hypersensitivity pneumonitis (HP) study population.

Parameter	UCHICAGO (n=78)	UCDAVIS (n=49)	UCSF (n=35)	UTSW (n=27)
Age years	65±9	71±10	60±11	62±9
Male	39 (50)	23 (47)	14 (40)	16 (59)
White	66 (85)	40 (82)	30 (86)	24 (89)
Tobacco pack-years	16±24	11±18	17±13	16±27
Env. antigen	24 (31)	49 (100)	35 (100)	21 (78)
FVC %	65±20	66±20	71±15	59±19
D_{LCO} %	55±25	52±24	58±13	45±19
Surgical biopsy	43 (55)	15 (31)	25 (71)	16 (59)
Corticosteroid	54 (69)	38 (78)	26 (74)	24 (89)

Median MMF exposure time was similar between patients with TL in the first quartile (Q1) and those in the second to fourth quartiles (Q2–Q4) (10 months (IQR 5–16 months) versus 10 months (IQR 4–23 months); p=0.86). Use of corticosteroid therapy was similar between both groups (Q1 77.1% versus Q2–Q4 74.5%; p=0.72). Baseline FVC was lower in Q1 patients that received MMF when compared to those that did not receive MMF (63.2±18.2% versus 74.1±14.7% pred; p=0.029). Similarly, baseline DLCO was lower in Q2–Q4 patients who received MMF than those that did not receive MMF (48.2±20.3% versus 58.2±24.4% pred; p=0.011).

Each quartile decrease in TL was associated with a stepwise decrease in transplant-free survival (**Figure 3.1**). Crude mortality rates were higher for Q1 patients when compared to Q2–Q4 (27.3 deaths per 100 person-years versus 8.4 deaths per 100 person-years; p=0.0002). In propensity score-adjusted analyses, Q1 patients had increased mortality overall when compared to Q2–Q4 (HR 3.29, 95% CI 1.56–6.95; p=0.002). When compared to Q1 patients who did not receive MMF, survival was improved in Q2–Q4 patients who received MMF (HR for interaction term 0.17, 95% CI 0.05–0.61; p=0.007), but not in Q2–Q4 patients who did not receive MMF (p=0.13), or in Q1 patients who received MMF (p=0.87) (figure 1c and d). Significant interaction existed between MMF and TL for Q3 (HR 0.19, 95% CI 0.05–0.70; p=0.013) and Q4 (HR 0.18, 95% CI 0.06–0.57; p=0.003), but not for Q1 (p=0.72) or Q2 (p=0.37). Seven patients were censored due to lung transplantation, all with TL above the first quartile. Those who received MMF appeared more likely to undergo lung transplantation (n=6; 10%)

compared to those who did not receive MMF (n=1, 2%; p=0.066). Importantly, irrespective of TL, annualized change in FVC and DLCO measurements did not differ with MMF use.

Figure 3.1 Quartiles of telomere length predict transplant-free survival in fibrotic hypersensitivity pneumonitis.

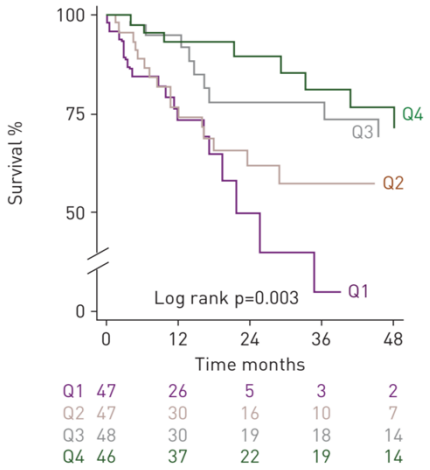


Figure 3.2 Propensity score-adjusted Cox regression theragnostic model of survival in fibrotic hypersensitivity pneumonitis.

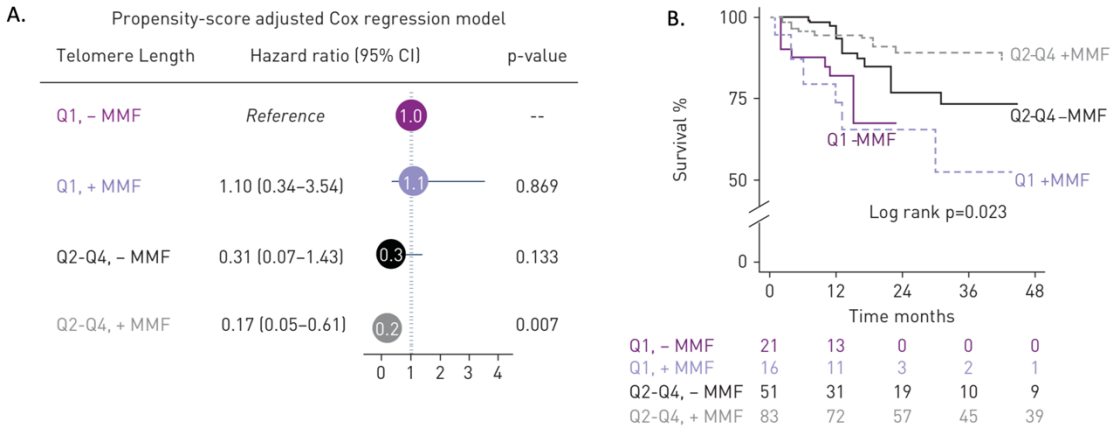


Figure 3.3 Scatterplots for lung function trajectory in fibrotic (chronic) hypersensitivity pneumonitis stratified by quartile (Q) and mycophenolate mofetil (MMF) therapy.

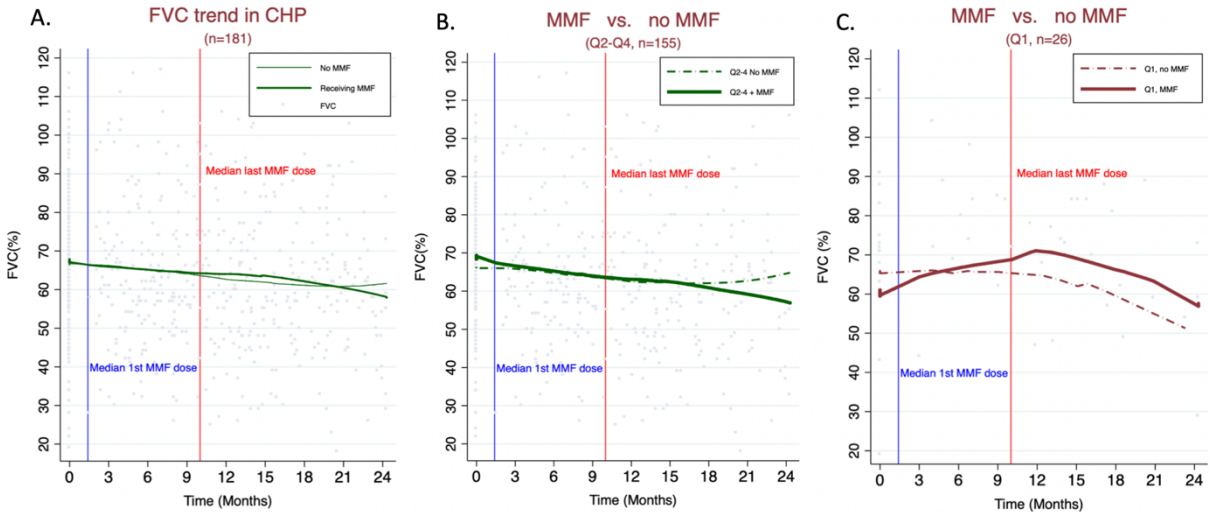
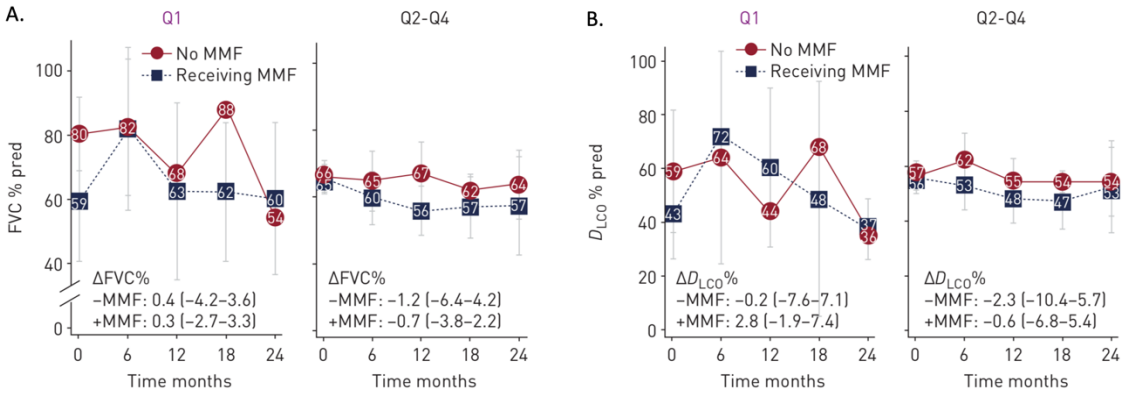


Figure 3.4 Lung function trajectory grouped by MMF therapy in fibrotic hypersensitivity pneumonitis. Forced vital capacity (FVC) % predicted trajectory from baseline for the first 24 months grouped by MMF therapy (blue squares and dotted lines) versus no MMF therapy (red circles and solid lines) for first quartile (Q1), and second to fourth quartiles (Q2–Q4) of TL. f) Diffusing capacity of the lung for carbon monoxide (DLCO) % predicted trajectory from baseline for the first 24 months grouped by MMF therapy versus no MMF therapy for patients with TL in Q1, and Q2–Q4. Δ: annualized change in pulmonary function.



Discussion

These analyses show a stepwise decrease in transplant-free survival with each quartile decrease in telomere length. Crude mortality rates were notably higher for those patients with telomere length in the first quartile when compared to the others. Through propensity score-adjusted analyses, we identified that those in the first quartile had a higher mortality rate overall. Interestingly, we found improved survival in patients who received mycophenolate mofetil and had telomere lengths in the second to fourth quartiles, compared to those in the first quartile who did not receive the treatment.

These findings were echoed in a significant interaction between the treatment and telomere length for the third and fourth quartiles. Furthermore, patients who received the treatment seemed more likely to undergo lung transplantation. Importantly, regardless of telomere length, the annualized change in FVC and DLCO measurements showed no significant difference with the use of mycophenolate mofetil. These observations underscore that mycophenolate mofetil therapy is not associated with improved survival or lung function in patients with fibrotic hypersensitivity pneumonitis and short telomeres. This parallels the association in idiopathic pulmonary fibrosis and likely reflects a shared pathway in the pathophysiological processes of advanced fibrosis underlying these two diseases.

Contrarily, in the absence of short telomeres, the improved survival associated with mycophenolate mofetil in fibrotic hypersensitivity pneumonitis may hint at fundamental differences between idiopathic pulmonary fibrosis and fibrotic

hypersensitivity pneumonitis. While idiopathic pulmonary fibrosis does not typically have an inflammatory component, fibrotic hypersensitivity pneumonitis does, and immunosuppressive therapy might alleviate exposure-related alveolar inflammation earlier in the disease course. Furthermore, while certain genetic variants are common to both diseases, the array of telomere-related mutations and susceptibility-associated gene polymorphisms appears to differ in their associations with lung function decline and survival across different types of interstitial lung diseases [26, 74]. This study revealed key associations fundamental to pharmacotherapy in patients with fibrotic hypersensitivity pneumonitis that would serve as a foundation for larger, meticulously designed studies to probe the impact of immunosuppressive therapy in patients with telomere mutation related ILDs. This will help refine treatment strategies, enhancing their precision and effectiveness in addressing this challenging condition.

Chapter IV

Telomere biology disorders associate with autoimmunity in pulmonary fibrosis.

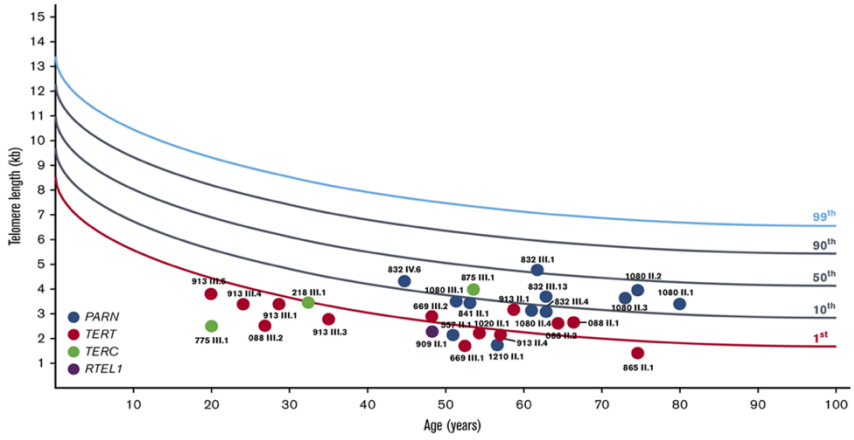
Introduction

Pulmonary fibrosis (PF) is a debilitating condition that poses significant challenges in terms of diagnosis and treatment. The past two decades have seen the emergence of telomere biology disorders (TBD) as an important factor in the pathophysiology of pulmonary fibrosis. Currently, there are 14 known genes associated with these disorders, which include *ACD, CTC1, DKC1, NAF1, NHP2, NOP10, PARN, RTEL1, TERC, TERT, TINF2, STN1, WRAP53,* and *ZCCHC8* [75-80]. Mutations in these genes result in a range of multiorgan pathologies considered part of a single disease entity, referred to as a telomeropathy, short telomere syndrome, or TBD. The phenotypes of TBDs vary depending on the age at diagnosis. Infants diagnosed with TBD typically present severe immunodeficiency, bone marrow failure, and gastrointestinal and neurologic dysfunction. Adults, on the other hand, may initially present with interstitial lung disease [20, 22, 81], emphysema [82], cryptogenic cirrhosis [83, 84], isolated macrocytosis or mild cytopenia(s), severe bone marrow failure [85, 86], or hematologic malignancies [87, 88].

Recent research efforts have been focused on understanding the prevalence of TBD in specific patient populations. Findings from these studies indicate that the prevalence of dyskeratosis congenita, a classical TBD presentation, is rare [89]. However, between 1% and 10% of adults with interstitial lung disease [20, 81], and up to 30% of those with familial pulmonary fibrosis [19, 90] harbor a pathogenic or likely pathogenic variant in a known TBD gene. In adults with hematologic disorders, the prevalence is less

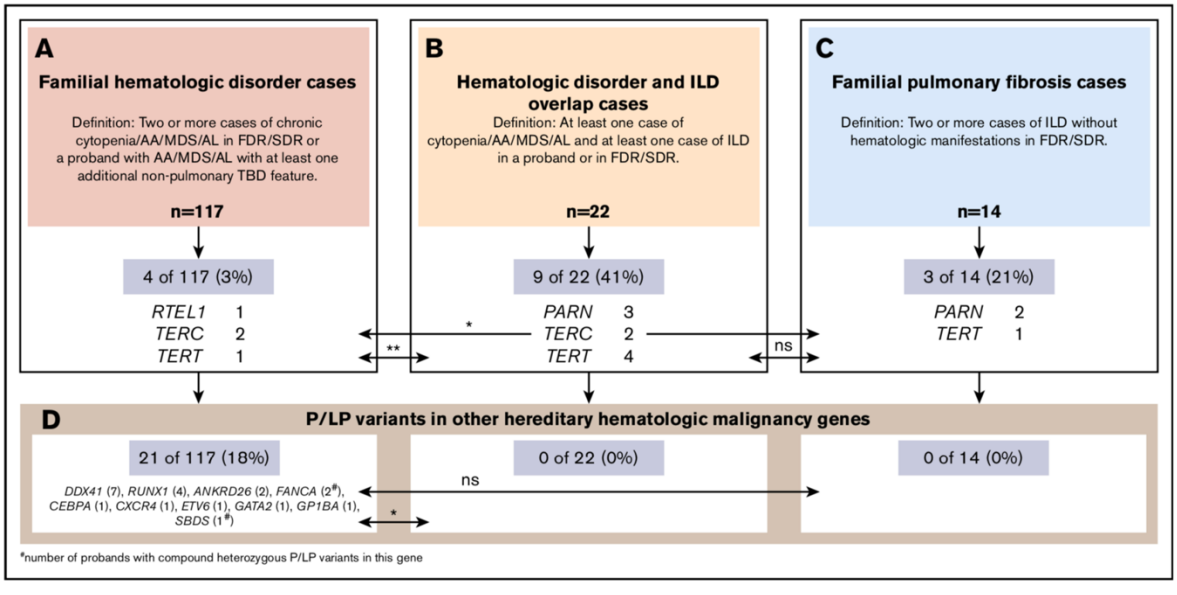
well characterized, but estimates suggest that up to 5% of patients with aplastic anemia and up to 3% of those with myelodysplastic syndrome or acute leukemia carry an inherited variant in a TBD gene [85, 91]. The importance of these genetic considerations lies in the propensity of patients with TBD to have shorter telomere lengths and experience serious complications of standard treatments for sporadic aplastic anemia or interstitial lung disease (**Figure 4.1**). Hence, it is essential to have sensitive and specific diagnostic tests to identify these patients, making the study of telomere length an area of critical importance. However, challenges in telomere length interpretation, the absence of classical mucocutaneous features in most adults with genetically diagnosed TBD, and the similarity between typical age-related and TBD-related comorbidities often make it challenging to provide a clinical TBD diagnosis without genetic testing and detailed family assessment. In this light, the investigation of the prevalence and phenotype of TBD in adults with hematologic presentations that should be enriched for TBD, those with familial clustering of chronic hematologic abnormalities, and/or aplastic anemia/myelodysplastic syndrome/acute leukemia with or without interstitial lung disease, is of utmost importance.

Figure 4.1. Telomere-related gene variants often present with shortened telomere lengths in kindreds with hematological disorders.



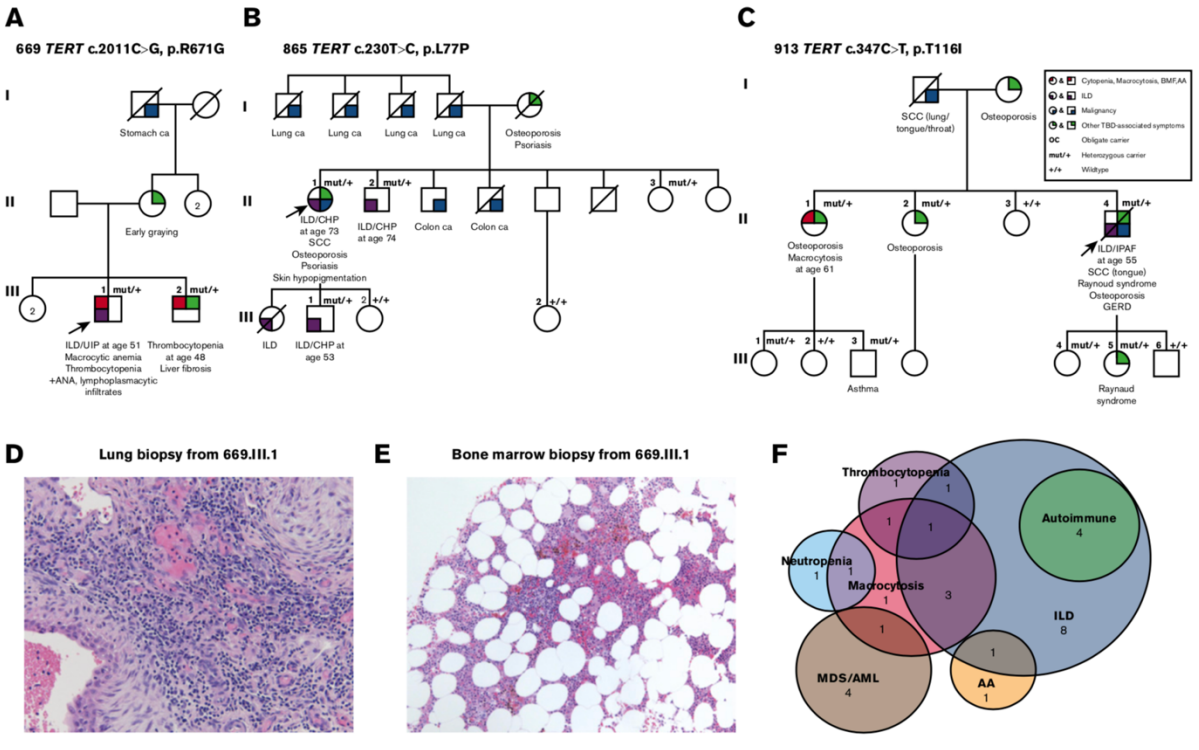
Of particular interest is prior collaborative work from our ILD group and that of Dr. Jane Churpek [92] (Feurstein S, **Adegunsoye A**, Mojsilovic D, et al. 2020) performed early in my doctoral studies that investigated the yield of genetic testing and phenotypic spectra in kindreds with TBDs. Our study of 153 adult probands found that the prevalence of inherited pathogenic or likely pathogenic variants in a TBD gene differed according to phenotype. Among families where hematologic disorders and interstitial lung disease overlap, we found a significantly greater proportion of inherited variants compared to those with familial hematologic disorder presentations without interstitial lung disease (**Figure 4.2**). These findings suggest a potentially powerful association between TBD, interstitial lung disease, and hematologic disorders.

Figure 4.2. Causative telomere biology disorders and gene variants compared in familial presentations.



In this study, a noteworthy pattern of co-occurrence of hematologic and pulmonary symptoms emerged among those carrying pathogenic or likely pathogenic variants in TBD genes [92]. If these symptoms frequently co-occur in individuals carrying variants in TBD genes beyond TERT and TERC, then personal and/or family history combination may be especially useful in selecting who to test. Interestingly, inflammatory and autoimmune-disease-related interstitial lung disease phenotypes were noted in several families. This was particularly prevalent in family 865, wherein all three members with ILD demonstrated a consistent pattern of chronic hypersensitivity pneumonitis. The presence of non-diagnostic autoimmune features was observed among four patients, marked by positive antinuclear antibodies. Notably, isolated instances of Raynaud's phenomenon and psoriasis were each identified in a single family. A microscopic examination of lung tissues revealed widespread lymphoplasmacytic infiltration in five out of six patients. Similarly, the results from bone marrow biopsies of two carriers of pathogenic or likely pathogenic variants highlighted notable lymphoplasmacytic infiltrates alongside clusters of plasma cells and eosinophils. However, comprehensive autoimmune serology panels, were not available for many carriers. The overlap and distribution of hematologic, pulmonary, and autoimmune features among our confirmed carriers of pathogenic or likely pathogenic variants is illustrated in a Venn-Euler diagram (**Figure 4.3**). The observed interplay of these conditions emphasizes the complex nature of the disease manifestation and underscores the need for further exploration into the genetic underpinnings of these disorders.

Figure 4.3. Hematologic, pulmonary, and autoimmune features overlap in genetically defined TBD kindreds. (A-C) Pedigrees of families with ILD and autoimmune features. A circle represents a woman, and a square represents a man. Roman numerals indicate generations (I, II, III). A slash through the circle or square indicates the person is deceased. The proband is indicated by the arrow. The variant carrier status is mut/1 for carriers and 1/1 for wild type. The exact disease/phenotype is detailed below the symbol. The pedigree number and the variant found in this family are designated above the pedigree. (D) Lung biopsy specimen of confirmed TBD variant carrier (669.III.1) with diffuse lymphoplasmacytic infiltrates. Depicted are fibroblastic foci with a diffuse lymphoplasmacytic infiltrate (magnification 340, hematoxylin and eosin [H&E] staining). (E) Bone marrow biopsy specimen of confirmed TBD variant carrier (669.III.1) with diffuse lymphoid aggregates. This core biopsy sample shows clusters of lymphocytic aggregates (magnification 310, H&E staining). (F) Venn-Euler diagram of the overlap of hematologic, pulmonary, and autoimmune phenotypes observed in confirmed variant carriers. The hematologic phenotype was divided into 5 different categories: MDS/AML (dark brown), macrocytosis/macrocytic anemia (red), neutropenia (teal), thrombocytopenia (purple), and AA (peach); the pulmonary phenotype observed consisted of ILD (blue). The autoimmune overlap with ILD subtypes is highlighted in green. The number within the circles represents the number of patients with this disease/combination of phenotypes.



This finding is consistent with recent reports that have identified pathogenic or likely pathogenic variants in TBD genes in populations with clinically diagnosed autoimmune disorders[24]. Similarly, individuals with known TBD pathogenic or likely pathogenic variants have been diagnosed with autoimmune diseases [24, 63, 93, 94]. Understanding the relationship between autoimmunity, lymphocyte telomere length, and TBDs is of immense importance, both clinically and mechanistically. Importantly, knowing that immunosuppressive agents may be ineffective or even harmful in the treatment of TBD-related bone marrow failure or interstitial lung disease, maintaining a high index of suspicion for TBD despite autoimmune features is critical.

Taken together, these findings underscore the need for further assessments of telomere length in larger populations of pulmonary fibrosis patients. This need is based on the rationale that autoimmune-related pulmonary fibrosis is prevalent in kindreds with telomere biology disorders. Only through this continued investigation can we improve our diagnostic and treatment strategies for this complex and challenging condition.

Methods

Study design and population

I performed a post-hoc analysis of the multicenter pulmonary fibrosis population in which patients were enrolled to investigate the relationship of leukocyte telomere length (LTL) with chronological age and mortality across racially diverse pulmonary fibrosis cohorts [44]. Patient enrollment occurred prospectively across four geographically

disparate U.S. tertiary centers: University of Chicago, University of California San Francisco, University of California Davis, and University of Texas Southwestern, Dallas over the period from September 2003 to December 2019.

In this investigation, I first extracted and assembled the cohort data from the original study, which consisted of LTL measurements among participants with pulmonary fibrosis stratified by race/ethnicity. These measurements were analyzed in connection with age and all-cause mortality and were compared with controls. The cohort was primarily divided by race and ethnicity into four categories: White, Black, Hispanic, and Asian as previously described.

Statistical analyses

Using binomial logistic regression models, odds ratios (ORs) were calculated in this analysis to determine the strength of association between telomere length and the chosen clinical predictors of respiratory impairment as the dependent variable, which included autoimmune-related interstitial lung disease, binary categorization of forced vital capacity below 50% predicted; binary categorization of diffusing capacity of the lung for carbon monoxide below 50% predicted; male sex, and IPF status. The predictive power of each clinical predictor was tested per quartile decrease in leukocyte telomere length. All data were further stratified by the age group at the time of pulmonary fibrosis diagnosis, allowing for the generation of age-specific ORs. The generated ORs and associated P-values were visualized in a comprehensive heat map. Each racial/ethnic group had its panel: White (top left, in red), Black (top right, in gray), Hispanic (bottom left, in orange),

and Asian (bottom right, in green). Each box within a panel corresponds to a specific age group (x-axis) and a particular clinical predictor of respiratory impairment (y-axis). The OR and P-value (in parenthesis) were displayed within each box, with an adjacent bar scaling the statistical significance of the odds likelihood. The color of each box conveyed the magnitude of the statistical significance, with darker colors representing higher significance. Furthermore, we examined the specific association between autoimmune-related lung fibrosis and telomere length. The data from the post-hoc analysis were depicted in box plots stratified by race and ethnicity, which illustrated the presence of longer telomere lengths associated with autoimmune-related lung fibrosis. This rigorous methodology and advanced statistical approach allowed us to delve deeper into the complex interplay of telomere biology disorders, autoimmune conditions, and pulmonary fibrosis. Importantly, given the post-hoc nature of this study, findings are exploratory and require validation in prospective cohorts. All statistical tests were two-sided, and a P-value less than 0.05 was considered statistically significant. This analysis utilized the Stata software package version 16 (StataCorp LP). All data management and statistical analyses were conducted following a pre-specified analysis plan.

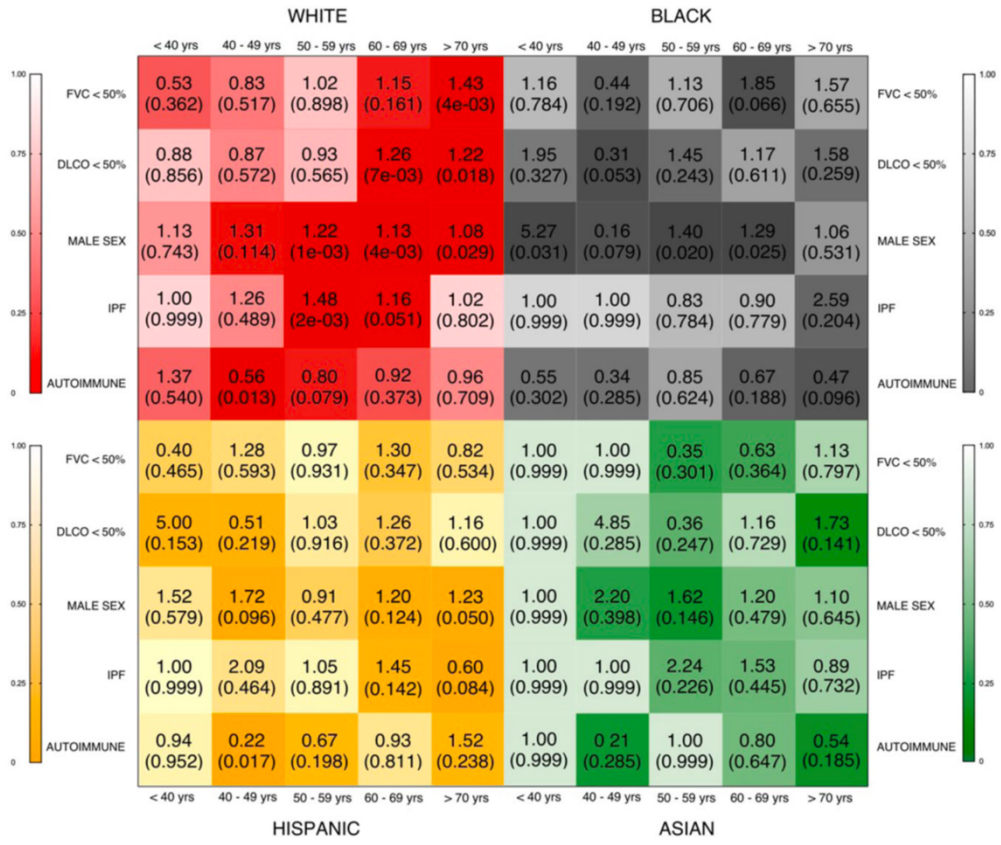
Results

With each decrease in leukocyte telomere length quartile, the age-specific odds for developing predictors of respiratory impairment, including forced vital capacity and diffusing capacity of the lungs for carbon monoxide, varied across racial groups with

pulmonary fibrosis (**Figure 4.4**). In the sixth and seventh decades of life, decreasing leukocyte telomere length was associated with increased odds for idiopathic pulmonary fibrosis in White participants, and heightened odds of being male in both Black and White participants.

In assessment of the age-stratified odds for predictors of lung fibrosis with decreasing telomere length, White (OR 0.56; P=0.013) and Hispanic (OR 0.22; P=0.017) subjects with shorter telomeres in their 5th decade of life had lower odds for autoimmune-related lung fibrosis, while the odds in Black (OR 0.34; P=0.29) and Asian (OR 0.21; P=0.29) subjects appeared lower but did not reach statistically significant thresholds [44].

Figure 4.4. Age-stratified odds for predictors of lung fibrosis with decreasing telomere length.



Among subjects with CTD-ILD or IPAF, which comprised autoimmune-related pulmonary fibrosis, standardized telomere lengths were longer for Black (0.34), Hispanic (0.25), and Asian (0.23) subjects than in White subjects (0.06; P<0.001). Similarly, among subjects without autoimmune-related pulmonary fibrosis, standardized telomere lengths were longer for Black (0.11), Hispanic (0.08), and Asian (0.01) subjects than in White subjects (-0.06; P<0.001). Overall, standardized telomere lengths were longer in subjects with autoimmune-related pulmonary fibrosis than in those without autoimmune-related pulmonary fibrosis across all racial and ethnic groups.

Figure 4.5. Autoimmune lung fibrosis is associated with longer telomere length.

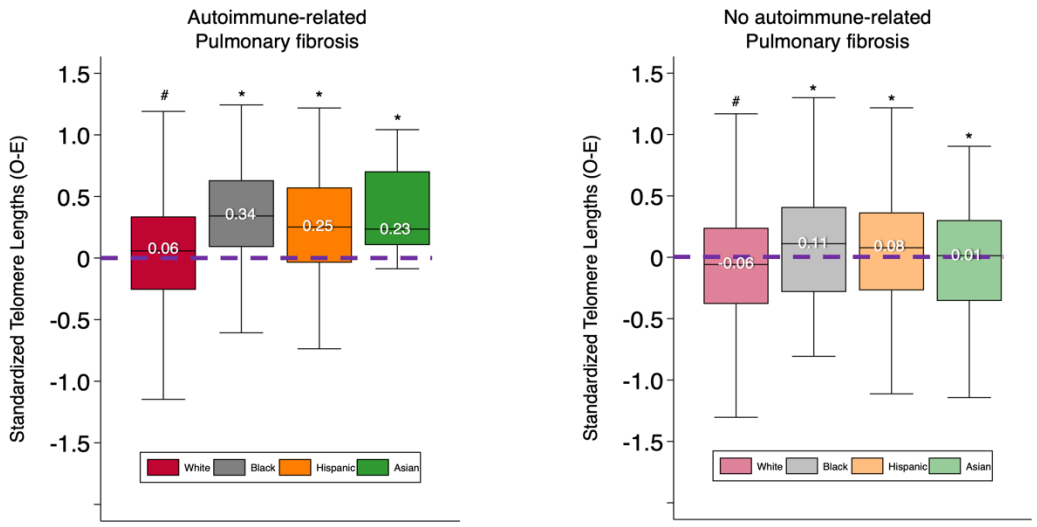


Table 4.1 Differences in Telomere Lengths Stratified by Race and autoimmune-related pulmonary fibrosis subtype.

Characteristics	WHITE	BLACK	HISPANIC	ASIAN	*P*-value[#]
Autoimmune-related PF	**(n=351)** 0.05 (0.47)	**(n=119)** 0.34 (0.38)	**(n=61)** 0.24 (0.54)	**(n=22)** 0.36 (0.35)	<0.0001
Non-Autoimmune-related PF	**(n=1242)** -0.07 (0.49)	**(n=43)** 0.14 (0.57)	**(n=126)** 0.05 (0.05)	**(n=22)** -0.01 (0.53)	0.005

*Total sample size, n=2,046. *Standardized telomere length across study population presented as means (SD). Telomere lengths unadjusted for age or sex. [#]P-value for ANOVA test comparing all four main racial groups. Patients with mixed or other racial ancestry not depicted above, n=14. PF=Pulmonary fibrosis; ILD=interstitial lung disease; IPF=Idiopathic pulmonary fibrosis, n=735; IPAF=interstitial pneumonia with autoimmune features, n=206; CTD-ILD=Connective tissue disease associated-ILD, n=349; FHP=Fibrotic hypersensitivity pneumonitis, n=418; Unclassifiable/Others, n=324.*

Discussion

This study provided additional insights into the results from our analysis of leukocyte telomere length across racially diverse pulmonary fibrosis cohorts. In this investigation, I illuminated the sophisticated interplay between autoimmunity traits and telomere length across racial groups in understanding the underpinnings of autoimmune-related pulmonary fibrosis. I also explored the implications of these findings for clinical practice, research, and public health.

My findings reveal intriguing associations between LTL and predictors of respiratory impairment. The age-specific odds for developing such predictors with decreasing LTL quartiles varied across racial groups, suggesting the existence of intricate gene-environment interactions and their implications for the onset and progression of pulmonary fibrosis. As environmental factors can modulate the expression and function of genes, it is likely that certain external influences could interact with telomere biology disorders, autoimmune genes, and HLA genes, leading to an elevated risk of autoimmune disease. Future research should therefore focus on elucidating these gene-environment interactions and their mechanistic role in autoimmune forms of pulmonary fibrosis.

Interestingly, in the sixth and seventh decades of life, decreasing LTL was associated with increased odds for idiopathic pulmonary fibrosis (IPF) in White participants and heightened odds of being male in both Black and White participants. This finding underscores the potential interplay of sex differences with disease susceptibility mechanisms for pulmonary fibrosis, further suggesting that genetics and biological sex

may play a complex role in determining individual susceptibility to pulmonary fibrosis. It might be that hormonal or genetic differences across genders modulate the impacts of telomere biology disorders and autoimmunity on the development of pulmonary fibrosis. Further investigation is warranted to substantiate this hypothesis.

Notably, I found that autoimmune-related interstitial lung disease (ILD) was three to four times more common in Black patients compared to White patients [44]. This significant racial disparity points to the importance of investigating genetic and epigenetic factors, including telomere biology and HLA genes, alongside sociocultural and environmental determinants of health that might contribute to this observed difference. Understanding the underlying mechanisms could lead to targeted interventions to reduce this disparity and improve outcomes for Black patients with pulmonary fibrosis.

In an intriguing twist, our results showed that among subjects with connective tissue disease-related ILD (CTD-ILD) or interstitial pneumonia with autoimmune features (IPAF), Black, Hispanic, and Asian subjects had longer standardized telomere lengths than White subjects. This also held true among subjects without autoimmune-related pulmonary fibrosis. Moreover, across all racial and ethnic groups, standardized telomere lengths were longer in subjects with autoimmune-related pulmonary fibrosis than in those without. This suggests that the relationship between telomere length and autoimmune-related pulmonary fibrosis is complex and multifactorial, possibly involving interplay with other genetic and environmental factors. A more nuanced understanding of these associations could have important clinical implications, potentially informing personalized treatment strategies and prognostications for patients with pulmonary fibrosis.

Overall, this study underscores the need to further investigate telomere biology disorders, autoimmunity genes, and HLA genes in larger, racially diverse cohorts of pulmonary fibrosis patients. We also highlight the importance of considering the potential influence of gene by environment interactions and sex differences on disease susceptibility. Ultimately, these avenues of research could lead to novel insights into the pathogenesis of pulmonary fibrosis and inform strategies for prevention, early detection, and personalized treatment, which in turn could greatly improve patient outcomes and quality of life.

Chapter V

Mutations in PDE4DIPP, ZNF683, SFRP5, MIR6077, RPSAP72, WASIR2, GAPDHP27 and CNTNAP3P2 drive honeycomb fibrosis in multiethnic populations

Introduction

Pulmonary fibrosis is a complex disease characterized by aberrant proliferation and differentiation of lung fibroblasts, culminating in the excessive accumulation of extracellular matrix (ECM) and, eventually, fibrotic remodeling of lung tissue [27]. This process, in its most extreme clinical phenotype is termed honeycombing fibrosis, leads to reduced lung function and, ultimately, respiratory failure. This phenotype has a devastating effect on survival and has been linked to genetics [95-99]. Several thousand genes are known to be differentially expressed in fibrotic lung tissue, implicating multiple biological pathways in the pathogenesis of pulmonary fibrosis[100-102]. These include ECM formation, degradation, signaling, growth factors, developmental pathways, immunoglobulins, complement, and chemokines. However, despite these advances, the genetic underpinnings of honeycombing fibrosis remain poorly understood, particularly in the context of multiethnic populations.

Despite significant progress in genomic analyses, the impact of ethnicity on genetic predisposition to radiologic honeycombing remains unknown. This limitation may be due to conduction of prior genetic association studies in non-Black patients as pulmonary fibrosis is thought to be more prevalent in subjects of European descent [46]. Despite

pulmonary fibrosis affecting diverse ethnicities, polymorphisms such as *MUC5B*, which has been associated with pulmonary fibrosis is notably absent in Africans [52]. I have previously conducted qualitative and quantitative honeycombing-focused studies in collaboration with experts in ILD radiology using standardized scoring methods [30, 103-106]. To assess the genetic factors associated with radiologic honeycomb fibrosis, a transcriptome-wide investigation across diverse races/ethnicities is needed. My preliminary data from the phenomics analysis of 4792 participants with pulmonary fibrosis across the US underscores the need for such an investigation [11]. This analysis demonstrated an earlier age at pulmonary fibrosis diagnosis and devastating clinical outcomes, despite similar prevalence of radiologic honeycombing among Black (37.1%) and White (37.0%) patients with pulmonary fibrosis [30].

Notably, the influence of genetic variations and environmental factors such as smoking on disease pathogenesis has been increasingly recognized. Cigarette smoke, for instance, has been shown to impact the lung epigenome and alter methylation patterns of specific promoters of genes implicated in the pathogenesis of pulmonary fibrosis[107-109]. This insight provides an impetus for further exploration of the gene-environment interaction in the pathogenesis of honeycombing fibrosis. It has been established that approximately 20% of IPF cases are familial, with causative mutations identified in genes involved in telomere function, protein folding, and secretion, impacting the function of epithelial cells [110-114]. This underscores the critical role of epithelial cell dysfunction, particularly of alveolar type 2 (AT2) cells, in IPF pathogenesis. With advanced age, AT2

cells exhibit compromised proliferative capacity, leading to defective epithelial maintenance that could potentially contribute to progressive lung fibrosis[115, 116].

Whilst gene expression profiling has provided insight into genes differentially expressed in pulmonary fibrosis, the focus of these studies has largely been on IPF as a whole rather than honeycombing fibrosis specifically. Furthermore, although diagnostic and prognostic biomarkers have been identified through peripheral blood profiling, the reliability of these markers in predicting baseline and short-term changes in forced vital capacity (FVC), a key measure of lung function, remains limited. Therapeutic approaches for IPF, including those that target inflammatory and tissue remodeling pathways, have been largely unsuccessful in the clinic [117]. This failure is attributable, in part, to the lack of comprehensive understanding of the disease and the absence of predictive biomarkers. Indeed, while studies have utilized microarray profiling and single-cell RNA sequencing to identify differentially regulated genes and pathways in IPF, our knowledge of the mechanisms that contribute to severe/end-stage IPF, characterized by severe honeycombing fibrosis, is very limited.

The objective of my study was to identify eGenes associated with radiologic honeycomb fibrosis in a multiethnic cohort of patients with diverse subtypes of pulmonary fibrosis. My central hypothesis is that genes, beyond *MUC5B*, are associated with radiologic honeycombing, differ across ethnicities and impact survival. Specifically, for this study, I hypothesize that changes in peripheral blood gene expression would be predictive of

radiologic honeycomb fibrosis. These findings can be used to derive a gene expression-based predictor of progressive pulmonary fibrosis in independent pulmonary fibrosis cohorts and will be crucial for determining pathways and cell types underpinning this gene predictors of disease progression in pulmonary fibrosis.

This dissertation chapter aims to address this critical knowledge gap by investigating the role of specific gene mutations in driving honeycombing fibrosis in multiethnic populations. Through the analysis of transcriptomic data and identification of differentially expressed genes associated with honeycombing fibrosis, I hope to provide new insights into the genetic drivers of this disease phenotype. This knowledge could contribute to the development of targeted therapeutic strategies and precision medicine approaches for this devastating disease.

Methods

Study design, ILD subtypes, and categorization of honeycombing

Our study population consisted of a multiethnic group of sequentially consented participants with pulmonary fibrosis from the University of Chicago. This cohort was prospectively enrolled, and comprised of consenting individuals who had a confident multidisciplinary diagnosis of pulmonary fibrosis including IPF, fibrotic hypersensitivity pneumonitis, CTD-ILD, IPAF, and unclassifiable ILD. All subjects were evaluated and had blood samples collected between September 2019 and June 2021, thus providing us with

a long-range, dynamic dataset. All procedures involved were approved by the relevant institutional review boards to ensure ethical compliance.

Relevant demographic, clinical, pulmonary function, and laboratory data were extracted from the ILD clinic evaluations. Subjects were then classified based on the documented presence or absence of honeycombing observed on their chest Computed Tomography (CT) scans. Imaging was independently reassessed by two experienced thoracic radiologists, all the while being uninformed of any other clinical data, to verify the consistency of our findings. Both radiologists collaboratively analyzed each High-Resolution CT (HRCT) scan, reaching a consensus on whether honeycombing was present or not. Honeycombing was defined as clustered cystic airspaces, typically exhibiting comparable diameters around 3 to 10 mm, predominantly located subpleurally with well-defined walls. Ascertainment of subject categorization was performed using a previously validated and automated, deep learning-based radiomic classifier that predicted radiologic honeycombing on subject chest CT scans. This radiomic honeycombing classifier also demonstrated good test performance across a wide range of subjects and similarly discriminated survival when compared to radiologist determined honeycombing.

RNA Extraction and Processing

I performed gene expression analysis using the PAXgene Blood RNA Kit (QIAGEN GmbH) for the extraction and purification of intracellular RNA from whole blood collected in the PAXgene Blood RNA Tube. Whole blood containing peripheral blood mononuclear cells

(PBMCs) were obtained from diverse cohorts of patients diagnosed with pulmonary fibrosis. Total RNA was extracted from these cells following standard protocols, then purified and quality-checked to ensure the integrity of the RNA samples. After plating in the Gilad lab, samples were transferred to the Genomics core for expression quantification.

Bulk RNA Sequencing and Preprocessing

Bulk RNA sequencing using the NovaSeq 6000 platform (Illumina, San Diego, CA) was performed on the extracted RNA to quantify gene expression levels. The data were preprocessed following a rigorous bioinformatics workflow to remove technical noise and correct for batch effects, ensuring accurate and unbiased results (**Figure 5.1, 5.2 and 5.3**). Reads were mapped to the human (hg38) genome and counted for each gene.

Figure 5.1. Bioinformatic workflow used for the transcriptomic analysis.

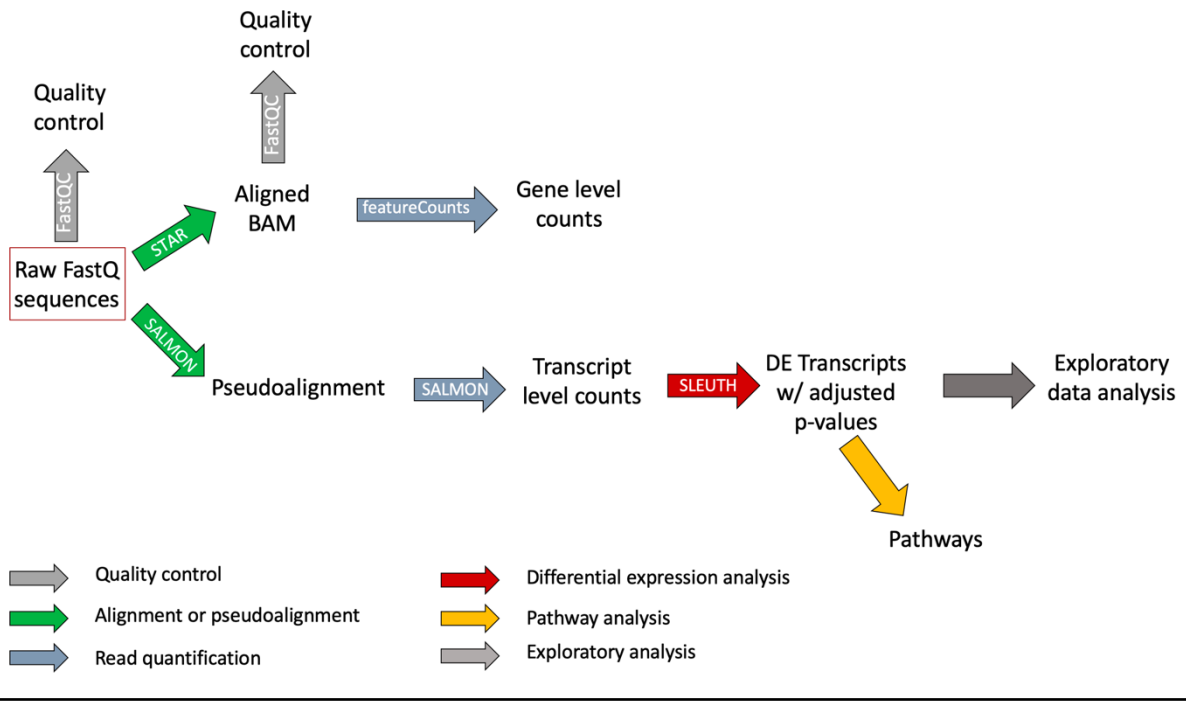


Figure 5.2 Quality control pipelines and filtering approach used for the dataset.

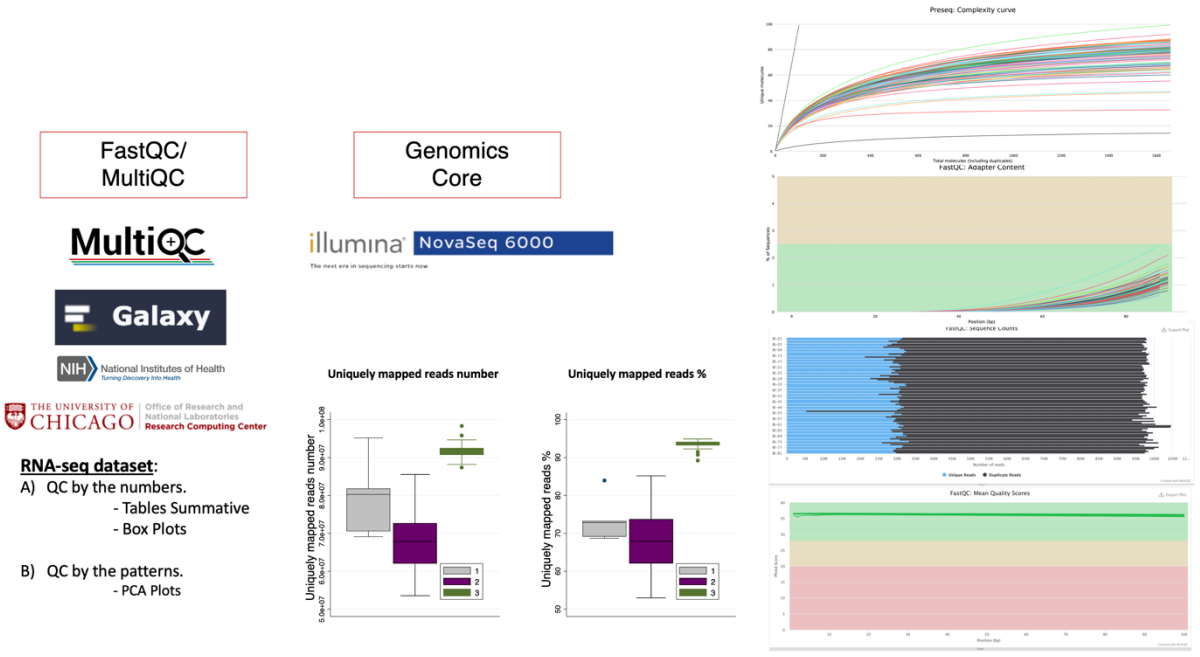


Table 5.1 Spliced Transcripts Alignment (STAR) alignment QC data

	aa (n=5)	da (n=60)	ms (n=29)	all (n=94)
Number of input reads	107579382	99646232	98054508	99577144
Average input read length	216	216	201	211
UNIQUE READS:				
Uniquely mapped reads number	79378850	67418055	91596966	75513719
Uniquely mapped reads %	73.57%	67.65%	93.41%	75.91%
Average mapped length	215	214	200	210
Number of splices: Total	50988970	43835794	56694463	48183318
Number of splices: Annotated (sjdb)	50441714	43381214	55890651	47616067
Number of splices: GT/AG	50613794	43540517	56205767	47824120
Number of splices: GC/AG	213426	168919	374587	234737
Number of splices: AT/AC	27826	20915	39210	26927
Number of splices: Non-canonical	133922	105443	74898	97535
Mismatch rate per base, %	0.20%	0.23%	0.20%	0.22%
Deletion rate per base	0.01%	0.01%	0.01%	0.01%
Deletion average length	1.72	1.67	1.75	1.70
Insertion rate per base	0.01%	0.01%	0.01%	0.01%
Insertion average length	1.51	1.48	1.54	1.50
MULTI-MAPPING READS:				
Number of reads mapped to multiple loci	26045922	26764289	4551494	19873194
% of reads mapped to multiple loci	24.44%	26.86%	4.65%	19.88%
Number of reads mapped to too many loci	50190	54711	46861	52048
% of reads mapped to too many loci	0.05%	0.05%	0.05%	0.05%
UNMAPPED READS:				
Number of reads unmapped: too many mismatches	0	0	0	0
% of reads unmapped: too many mismatches	0.00%	0.00%	0.00%	0.00%
Number of reads unmapped: too short	2069324	5364563	1825625	4097485
% of reads unmapped: too short	1.91%	5.38%	1.86%	4.11%
Number of reads unmapped: other	35096	44614	33563	40698
% of reads unmapped: other	0.03%	0.04%	0.03%	0.04%

Data Import and Transformation

The data were imported into an R environment using the read.csv function. The data matrix was transposed to orient the patient samples as rows and the gene counts as columns, following standard conventions for downstream statistical analysis. After ensuring the correct orientation and format, the data were transformed into a form suitable for the principal component analysis (PCA) and other statistical analyses.

Statistical Analysis

The statistical analysis was conducted within the R environment, utilizing packages specifically designed for the analysis of high-throughput sequencing data.

Normalization and Variance Stabilization

The raw count data obtained from the RNA-sequencing were normalized using the variance-stabilizing transformation (VST) function implemented in the DESeq2 package. This method corrects for library size differences and stabilizes the variance across the mean, permitting the use of standard statistical techniques on the normalized count data.

Principal Component Analysis

Principal Component Analysis (PCA) was performed on the VST-normalized data to visualize overall sample clustering and identify potential outliers. PCA was performed using the prcomp function in R, which computes the principal components of the data matrix. PCA is a useful technique to reduce the dimensionality of the data and visualize

patterns and structures. This unsupervised technique projects the data into a lower-dimensional space, with the axes (or principal components) chosen to maximize the variance in the data. The PCA results were saved in .csv and .txt files for further analysis.

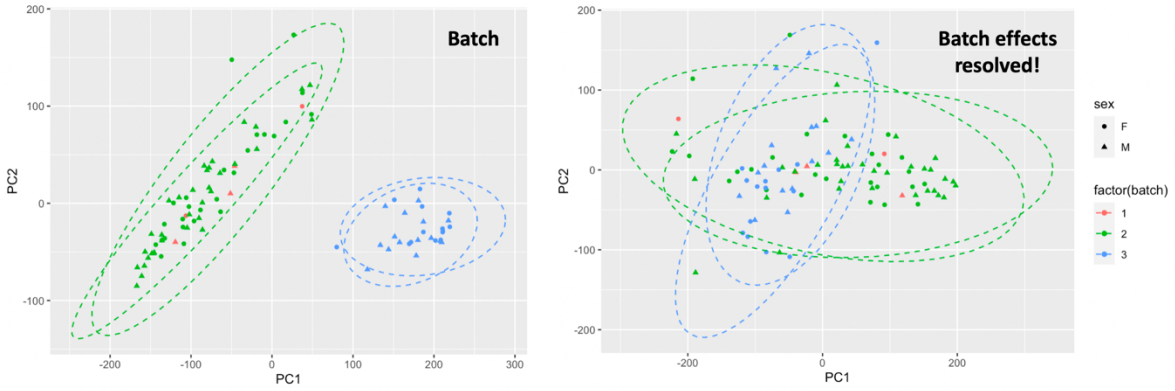
Batch Effect Adjustment

Potential batch effects were assessed visually using PCA plots and statistically using surrogate variable analysis. Where significant batch effects were detected, they were corrected using ComBat, a method that uses empirical Bayes models to adjust for known batch effects. This tool is designed to correct batch effects in high-throughput data, and this process ensures that any observed differences are due to biological variation rather than technical or batch-related factors. After batch correction, PCA was rerun on the corrected data to visualize the changes (**Figure 5.3**).

Differential Expression Analysis

Differential gene expression between the groups was assessed using DESeq2, a software in R designed specifically for differential analysis of count data. DESeq2 uses a negative binomial distribution to model the count data, then applies shrinkage estimators for both dispersion and fold change to improve stability and interpretability of estimates. This method is particularly robust for datasets with small sample sizes. DESeq2 is designed to analyze count data from high-throughput sequencing assays such as RNA-seq. This step identified genes that were differentially expressed in patients with pulmonary fibrosis, specifically those with honeycomb fibrosis.

Figure 5.3 Correction of batch effects in the transcriptomic analysis



Correction for Multiple Testing

Given the large number of simultaneous tests performed in a transcriptome-wide analysis, the Benjamini-Hochberg procedure was used to adjust the p-values for multiple testing, controlling the false discovery rate (FDR). An FDR threshold of 0.05 was used to identify differentially expressed genes.

Gene Set Enrichment Analysis

Gene ontology (GO) and pathway analysis were performed on the differentially expressed genes using the clusterProfiler package in R. This method identifies over-represented GO terms and biological pathways among the differentially expressed genes, providing insights into the biological processes affected in the disease state. GO enrichment analysis was performed to identify enriched biological processes, molecular

functions, and cellular components among the differentially expressed genes. The enrichGO function from the clusterProfiler package in R was used to conduct this analysis, using the human gene annotation database from org.Hs.eg.db. An FDR-adjusted p-value threshold of 0.05 was used to identify significantly enriched terms and pathways.

Visualization of Results

Results from the differential expression analysis were visualized using volcano plots, which plot the $\log_2$ fold changes against the negative logarithm of the adjusted p-values. This allows for the easy identification of significantly differentially expressed genes. The EnhancedVolcano package was used to generate these plots in R. A heatmap was generated to visualize the expression patterns of differentially expressed genes across the sample groups. The heatmap was produced using the ComplexHeatmap package in R, with hierarchical clustering performed on both the genes and samples to group together those with similar expression patterns. Genes were represented in rows and patient samples in columns. The color scale was set to represent Z-scores, which were calculated from the normalized gene counts.

The entire analytical process was conducted under a reproducible research framework, ensuring that all code and data manipulations can be reviewed and replicated.

Results

Demographic and clinical features of the study cohort differed by honeycombing fibrosis status at study enrollment (**Table 5.2**). Subjects with honeycombing fibrosis were older (71 +/- 10years vs 63+/-14years; P=0.002), and more likely to be male (67% vs 46%; P<0.0001). Honeycombing fibrosis was more prevalent among subjects with a clinical diagnosis of IPF (50% vs 12%), and less prevalent among subjects with a clinical diagnosis of CTD-ILD (7% vs 30%; P=0.001). DTA Honeycombing score was higher among subjects with honeycombing fibrosis compared to those without (10.8 (13.3) vs 1.4 (3.0); P<0.0001). In contrast to White subjects, Black subjects were more likely to lack honeycomb fibrosis, however these differences did not reach statistical significance (P=0.17).

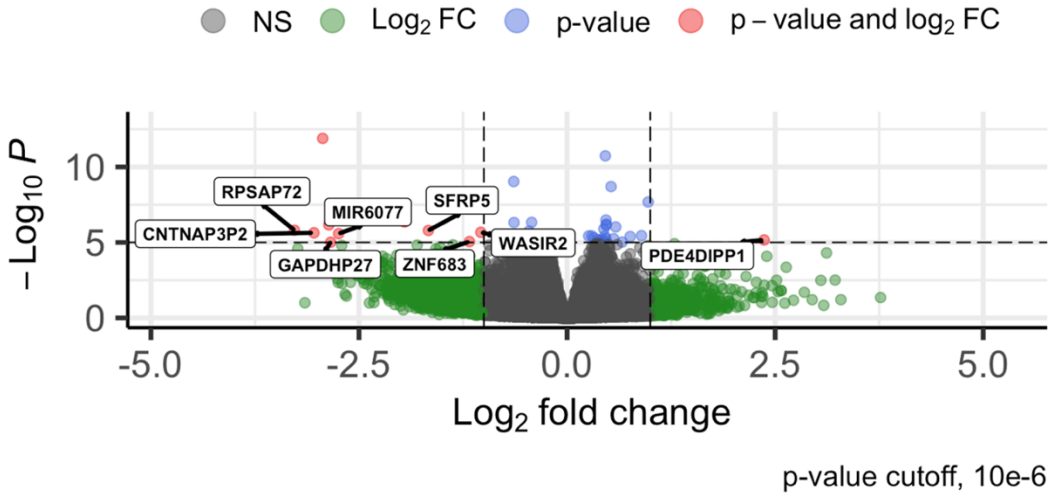
Table 5.2 Clinical Demographics in Pulmonary Fibrosis Study Cohort

Clinical Demographics	HONEYCOMBING FIBROSIS PRESENT (n=42)	HONEYCOMBING FIBROSIS ABSENT (n=52)	P-value[#]
Age, years	71 (10)	63 (14)	0.002
Male	28 (67)	24 (46)	<0.0001
Race			0.17
White	34 (81)	31 (60)	
Black	6 (14)	15 (28)	
Asian	1 (2)	3 (6)	
Hispanic	1 (2)	3 (6)	
Pulmonary fibrosis subtype			0.001
IPF	21 (50)	6 (12)	
IPAF	5 (12)	6 (12)	
CTD-ILD	3 (7)	16 (30)	
Fibrotic HP	10 (24)	18 (34)	
Unclassifiable/other ILD	3 (7)	6 (12)	
DTA Honeycombing score	10.8 (13.3)	1.4 (3.0)	<0.0001

Total sample size, n=94. [#]P-value for statistical test comparing honeycombing subgroup to participants without honeycombing. Categorical variables presented as n (%); continuous variables presented as mean (SD). White=Non-Hispanic White, Black=Non-Hispanic Black, Asian=Non-Hispanic Asian. *PF=Pulmonary fibrosis; ILD=interstitial lung disease; IPF=Idiopathic pulmonary fibrosis; IPAF=interstitial pneumonia with autoimmune features; CTD-ILD=Connective tissue disease associated-ILD; Fibrotic HP=Fibrotic hypersensitivity pneumonitis; Unclassifiable/Others.*

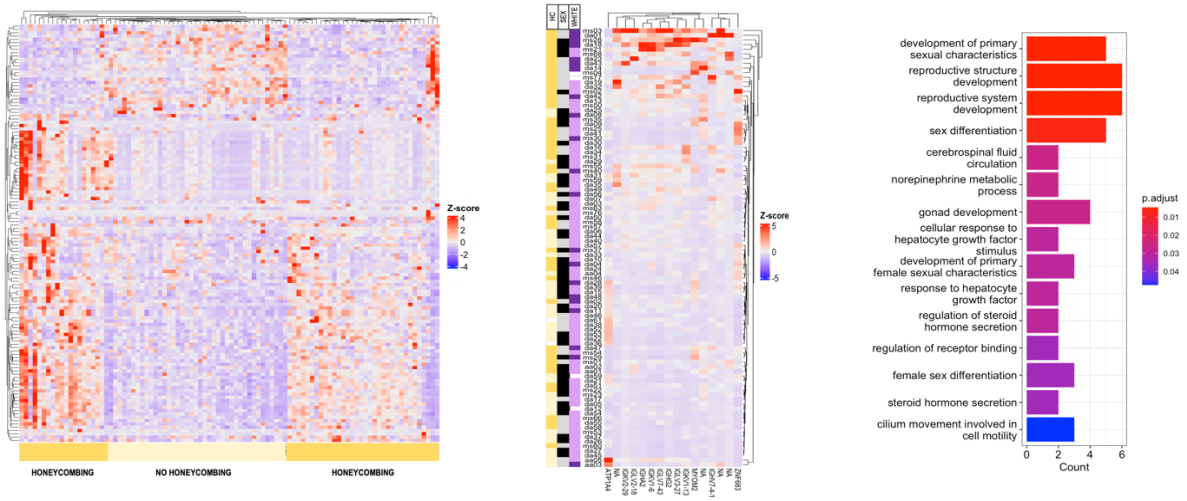
In this study, we probed the genetic landscape of pulmonary fibrosis in multiethnic populations with a particular focus on the identified mutations that drive honeycomb fibrosis. The results of our transcriptomic analyses of honeycomb fibrosis patients illuminated a myriad of differentially expressed genes (DEGs). Our approach of utilizing bulk RNA sequencing allowed us to discern significant alterations in the expression of eight particular genes —*PDE4DIPP*, *ZNF683*, *SFRP5*, *MIR6077*, *RPSAP72*, *WASIR2*, *GAPDHP27*, and *CNTNAP3P2*— all of which are now identified as crucial drivers of honeycomb fibrosis (**Figure 5.4**). Furthermore, our findings underscored *ZNF683*, a zinc finger protein coding gene located on chromosome 19, as a central player in the regulation of gene expression within this pathology. Each gene exhibited robust differential expression between honeycomb fibrosis patients and controls, underscoring their potential pathogenic roles.

Figure 5.4 Differentially expressed gene variants for honeycomb pulmonary fibrosis



A second set of genes that caught our attention included *IGHV7-4-1*, *IGKV1*, *IGLV3*, *IGHA2*, *IGLV2*, and *IGKV2*, all of which encode immunoglobulin heavy or light chain variable region proteins (**Figure 5.5**). They are primarily located on chromosomes 2, 14, and 22. *IGHV7-4-1*, an immunoglobulin heavy chain variable region protein-encoding gene located on chromosome 14, exhibited a unique expression pattern, suggesting its potential role in the immune response associated with honeycomb fibrosis. The *MYOM2* gene, which codes for a muscle-related protein and is located on chromosome 8, also displayed a distinctive expression profile indicative of its possible involvement in the myofibroblast-related, structural, and functional aspects of honeycomb fibrosis. *ATP1A4*, an ATPase family gene located on chromosome 1, emerged as another significant player in honeycomb fibrosis, underpinning the crucial role of ion transportation across cellular membranes in this disease.

Figure 5.5 Heatmap and enriched pathways for honeycombing in the pulmonary fibrosis dataset



A comprehensive Gene Ontology (GO) enrichment analysis of our RNA-seq data revealed a distinct over-representation of certain biological processes and pathways among the DEGs. Our results highlighted the involvement of multiple developmental pathways, including the development of primary sexual characteristics, reproductive structure development, sex differentiation, and gonadal development. Furthermore, other biological processes were implicated such as cerebrospinal fluid circulation and norepinephrine metabolic process. Importantly, genes involved in the cellular response to hepatocyte growth factor and the regulation of steroid hormone secretion also appeared enriched. The findings also pointed to a possible role for pathways regulating receptor binding and cellular motility, including cilia movement.

Focusing on the top 16 differentially expressed genes, we identified significant alterations in the expression of genes involved in key biological processes and pathways. Importantly, the DEGs surpassing a stringent P-value threshold of 10e-6 and a log2 fold change cut-off include *PDE4DIPP*, *ZNF683*, *SFRP5*, *MIR6077*, *RPSAP72*, *WASIR2*, *GAPDHP27*, and *CNTNAP3P2*. Each of these genes demonstrated a robust differential expression pattern in honeycomb fibrosis patients, highlighting their potential contribution to disease pathogenesis.

The gene *PDE4DIPP*, a phosphodiesterase known to regulate cyclic AMP levels, showed significantly increased expression in honeycomb fibrosis patients. Upregulation of *PDE4DIPP* suggests potential dysregulation of cyclic AMP signaling in these patients. In contrast, *ZNF683*, a transcription factor vital for natural killer cell maturation, demonstrated decreased expression, implying the involvement of the innate immune

response in honeycomb fibrosis. The gene *SFRP5*, involved in the modulation of Wnt signaling, was also observed to be downregulated. This aligns with the critical role of aberrant Wnt signaling in fibrosis [118, 119]. In tandem, *MIR6077*, a non-coding RNA, was markedly downregulated, hinting at a possible post-transcriptional regulation contributing to disease pathology.

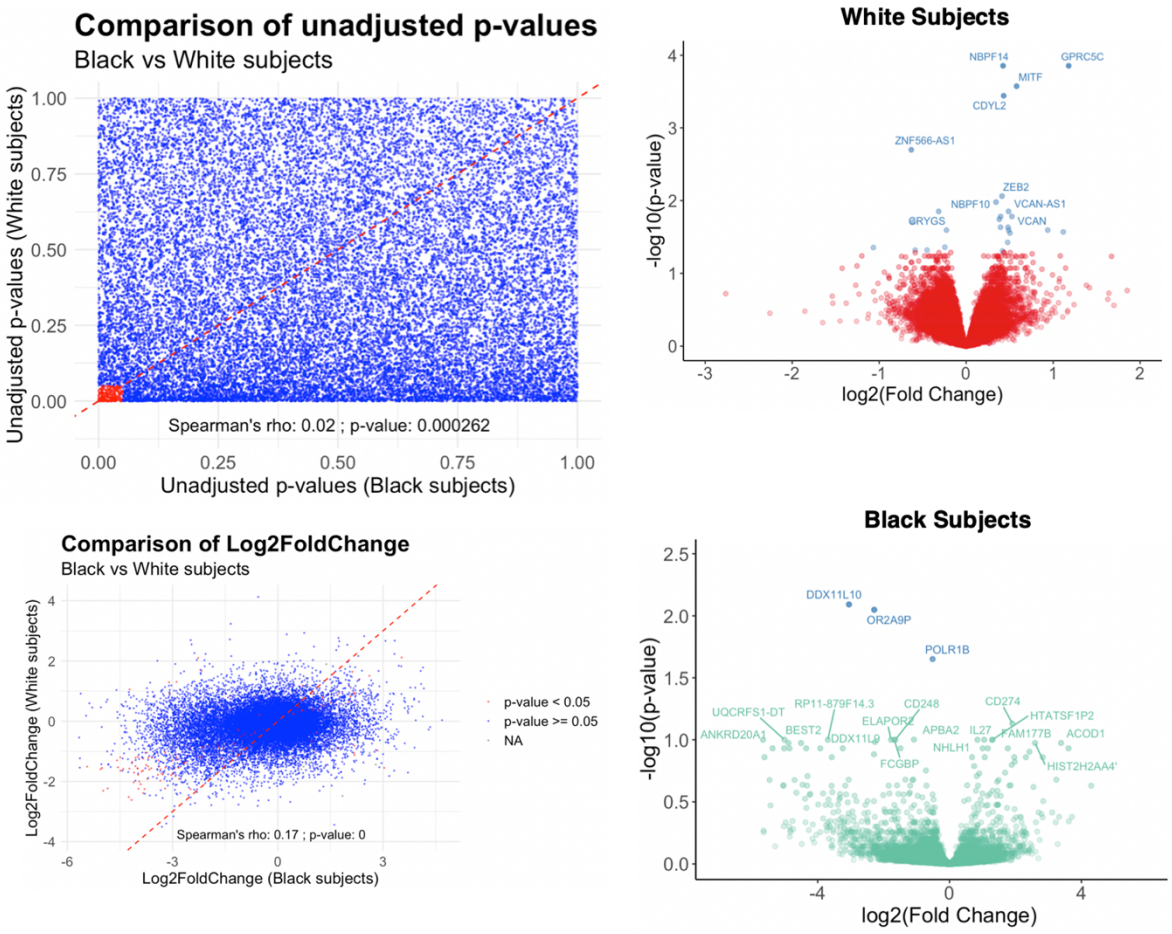
Two genes involved in ribosomal function and protein synthesis—RPSAP72 and *GAPDHP27*—were also differentially expressed. *RPSAP72* was significantly downregulated, and *GAPDHP27* was downregulated, suggesting dysregulation of protein synthesis machinery in honeycomb fibrosis. The gene *WASIR2*, associated with actin filament organization, showed decreased expression, pointing to potential cytoskeletal rearrangements. Lastly, *CNTNAP3P2*, a gene implicated in cell adhesion, was also downregulated, suggesting a disruption in cellular communication and integrity.

These results are consistent with the GO enrichment analysis which revealed significant enrichment of biological processes linked to immune responses, ECM remodeling, cell adhesion, and Wnt signaling pathway. Notably, pathways associated with innate immunity and ECM interactions were particularly enriched, further reinforcing our initial findings from the differential gene expression analysis.

Finally, we examined the expression patterns of our identified genes across different racial/ethnic groups (**Figure 5.6**). Notably, significant expression differences were found across these groups, providing compelling evidence that genetic and transcriptomic variations contribute to the observed racial disparities in the temporal incidence of honeycomb fibrosis and disparities in outcomes. Overall, our results unveil

a comprehensive transcriptional landscape of honeycomb fibrosis, highlighting the key role of identified genes, which may serve as potential targets for therapeutic interventions.

Figure 5.6 Race-stratified differentially expressed gene variants for honeycomb pulmonary fibrosis.



Discussion

Our investigation into the genetic underpinnings of honeycomb fibrosis, particularly in multiethnic populations, has revealed compelling insights into the genomic landscape of this complex disease. The novel aspect of this work lies in its focus on honeycomb fibrosis, a pathological feature that has remained understudied in comparison to broader idiopathic pulmonary fibrosis. As the first transcriptomic study of honeycombing, it provides a foundation for understanding the genetic influences driving the development and progression of this distinctive radiological hallmark of progressive fibrotic diseases.

Our results identified eight significantly differentially expressed genes—*PDE4DIPP*, *ZNF683*, *SFRP5*, *MIR6077*, *RPSAP72*, *WASIR2*, *GAPDHP27*, and *CNTNAP3P2*—emerging as critical drivers of honeycomb fibrosis. Interestingly, our results suggest a potential shift in gene expression regulation within honeycombing lesions, particularly highlighted by the differential expression of *ZNF683*. This gene, which encodes a zinc finger protein involved in gene expression regulation, has previously been associated with immune response regulation [120-122]. In contrast, our study reveals its potential role in the pathogenesis of honeycomb fibrosis, providing a unique insight into the disease's genetic landscape.

The relevance of PDE4 inhibitors in the context of IPF has gained prominence with the promising Phase 2 study results of BI 1015550 [123]. By maintaining forced vital capacity (FVC) in patients with IPF, BI 1015550 has demonstrated potential as a therapeutic agent capable of altering the course of disease progression. Its efficacy is

now being further evaluated in a larger Phase 3 trial (the FIBRONEER trial), not only in IPF but also in other phenotypes of progressive pulmonary fibrosis [123](ClinicalTrials.gov Identifier: NCT04419506). Interestingly, three PDE4 inhibitors have already received approval for the treatment of inflammatory diseases of the skin and lung. This emphasizes the utility of PDE4 inhibitors as a therapeutic strategy in conditions with an underlying inflammatory component. However, it should be noted that the efficacy of PDE4 inhibitors in fibrotic diseases has yet to be fully established. It is fascinating to consider the potential implications of our findings regarding the upregulation of PDE4DIPP in honeycomb fibrosis, in light of the recent clinical validation of PDE4 inhibitors. It's plausible that the dysregulation of cyclic AMP signaling, as suggested by the overexpression of *PDE4DIPP*, might be a therapeutic target for honeycomb fibrosis, akin to the role of PDE4 inhibitors in IPF. Nevertheless, this hypothesis is still to be thoroughly examined. Several questions remain, including whether the dysregulation of *PDE4DIPP* is directly involved in the fibrotic process and whether this pathway's modulation can have an actual therapeutic impact. Future studies will be critical in elucidating these connections and expanding our understanding of the intricate molecular mechanisms underpinning honeycomb fibrosis.

Further, our attention was drawn to a series of genes encoding immunoglobulin heavy or light chain variable region proteins. In particular, the gene *IGHV7-4-1*, an immunoglobulin heavy chain variable region protein-encoding gene, indicated potential alterations in the immune response associated with honeycomb fibrosis [124]. *MYOM2* and *ATP1A4*, genes involved in muscle-related protein synthesis and ion transportation,

respectively, further extend the potential genetic influences underpinning honeycomb fibrosis. The comprehensive Gene Ontology (GO) enrichment analysis of our RNA-seq data provides further depth to our understanding of honeycomb fibrosis at a molecular level. A plethora of biological processes and pathways, including those related to developmental pathways, cerebrospinal fluid circulation, and norepinephrine metabolic process, were highlighted, suggesting a complex interplay of various cellular functions in the disease's pathogenesis. Remarkably, genes implicated in the cellular response to hepatocyte growth factor and the regulation of steroid hormone secretion also showed significant enrichment, potentially pointing to novel pathways involved in honeycomb fibrosis.

When discussing the top differentially expressed genes, we found significant expression changes in genes known to be involved in key biological processes and pathways. It is of note that these genes demonstrated robust differential expression patterns, with each surpassing a stringent P-value threshold of 10e-6 and a log2 fold change cut-off. Interestingly, while our results reveal genes such as *PDE4DIPP*, *WASIR2*, and *CNTNAP3P2* that align with known fibrotic processes, they also unveil genes like *MIR6077*, a non-coding RNA, highlighting the potential role of post-transcriptional regulation in honeycomb fibrosis [125, 126]. These findings, together with our GO enrichment analysis results, build a consistent picture of honeycomb fibrosis as a disease with complex pathophysiology. Specifically, they emphasize the intertwining of immune responses, extracellular matrix remodeling, cell adhesion, and signaling pathways as integral to disease development and progression [27].

A closer examination of gene expression patterns across racial and ethnic groups further illuminated potential genetic and transcriptomic contributions to the disparities observed in honeycomb fibrosis incidence and outcomes. These findings underscore the critical need to consider racial and ethnic variations in genetic studies of fibrotic lung diseases and highlight the potential of personalized medicine approaches in managing these conditions [46].

Overall, our findings present a comprehensive and nuanced view of the transcriptomic landscape of honeycomb fibrosis. They underscore the potential role of the identified differentially expressed genes as novel markers of disease activity and as possible therapeutic targets. Despite these promising results, it is essential to recognize the need for further research to validate these findings and explore their mechanistic underpinnings in the pathogenesis of honeycomb fibrosis. This work marks a significant step forward in the field but also highlights the vast and complex terrain that still needs to be navigated in our pursuit of a deeper understanding of honeycomb fibrosis and the development of effective therapeutic interventions.

Chapter VI

Integrative whole genome and transcriptomic sequencing identify host defense and cell senescence gene variants contributing to pulmonary fibrosis risk.

Introduction

Pulmonary fibrosis, a progressive and often fatal fibrotic lung disease, presents a significant public health challenge across the globe, affecting thousands of individuals annually [127]. This disease presents significant heterogeneity in its clinical course, varying between rapid progression to a slow, insidious onset. A growing body of evidence underscores a substantial genetic contribution to pulmonary fibrosis, with 15-25% of patients reporting a family history of the disease [19, 128]. Importantly, this genetic underpinning exhibits variation across different racial and ethnic populations, highlighting the crucial need for more comprehensive, multiethnic studies [19, 52, 97].

Genome-wide association studies (GWAS) have been instrumental in identifying the genetic architecture of complex diseases like pulmonary fibrosis, unveiling over 20 common variant signals linked to disease susceptibility [14, 47, 48, 129-133]. Simultaneously, investigations into rare variants have implicated telomere-related genes with pulmonary fibrosis [12, 13, 23, 134, 135]. More recently, a robust study leveraging the power of whole-genome sequencing (WGS) discovered a genome-wide significant common pulmonary fibrosis susceptibility signal adjacent to the 16p telomere, alongside a putatively novel association of rare variants at ADA [136].

Common genetic variations in 17 loci have been conclusively associated with pulmonary fibrosis, with polymorphisms in the *MUC5B* gene emerging as a potent genetic risk factor [14, 49-51, 97, 137-139]. Interestingly, the distal promoter region of the *MUC5B* gene, particularly the rs35705950 polymorphism, emerges as the most potent risk factor for both sporadic and familial IPF [50, 52, 97]. While a considerable volume of genetic data has been collected over the years, the causal biological mechanisms underpinning these associations remain shrouded in mystery. A deeper understanding of these mechanisms holds the key to developing novel therapeutic strategies and advancing our knowledge of pulmonary fibrosis pathogenesis.

The functional implications of genetic variants can often be discerned through their impact on gene expression levels. This approach, known as expression quantitative trait locus (eQTL) mapping, links genetic variations with alterations in gene expression [36]. The Genotype-Tissue Expression (GTEx) project has utilized this methodology on a grand scale, revealing that the majority of genes contain regulatory genetic variants [35]. The combination of eQTL mapping with co-localization of genetic and eQTL signals offers a powerful approach to identify potentially causal genetic risk variants underlying GWAS signals. However, most eQTL analyses to date have utilized healthy tissues or generic cell lines, potentially overlooking the disease-specific transcriptional changes. Recent strides have been made to perform eQTL analyses using disease-relevant tissues and cell types [140-142], a lot of which emanated from work out of the Gilad lab. Furthermore, GTEx's initiative to perform cell type-interaction eQTL (ieQTL) analyses offers insights

into cell type-specificity of genetic regulation of gene expression across human tissues [35].

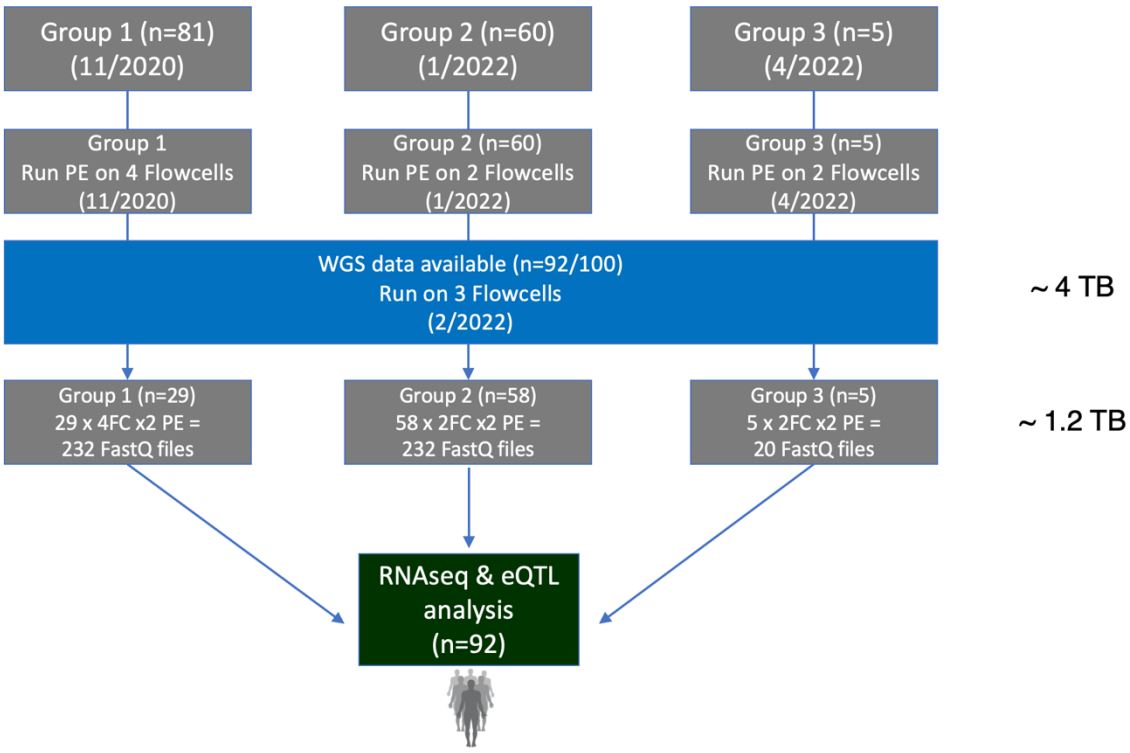
Against this background, the current study is designed to investigate the functional relevance of common genetic pulmonary fibrosis risk variants across diverse racial and ethnic groups. By integrating whole genome and transcriptome data derived from peripheral blood—a disease-relevant tissue source, especially in non-IPF phenotypes of pulmonary fibrosis where systemic involvement plays a larger role in disease pathophysiology—I aim to perform a comprehensive eQTL mapping. This will shed light on the effect of genetic variants on local (cis) and distant (trans) gene expression. Concurrently, co-localization analysis of eQTL with genetic loci will be used to prioritize potential causal risk variants and validate their functional roles.

My primary objective in this dissertation chapter, "Integrative whole genome and transcriptomic sequencing identify host defense and cell senescence gene variants contributing to pulmonary fibrosis risk," is to underscore the potential of a multi-omic approach in unraveling the complex genetic landscape of pulmonary fibrosis across diverse populations. By linking genetic variants to alterations in gene expression, I seek to unravel the molecular mechanisms contributing to the pathogenesis of pulmonary fibrosis and identify potential targets for therapeutic interventions. Ultimately, the insights garnered from our analyses will be pivotal in advancing our understanding of pulmonary fibrosis, fostering the development of personalized therapeutic strategies, and addressing health disparities in this devastating disease.

Methods

The methodology I employed for this investigation encompasses a multi-tiered approach involving comprehensive analyses of genomic, transcriptomic, and clinical phenotype data to identify potential gene variants that contribute to pulmonary fibrosis risk (**Figure 6.1**). The study also highlights the power of multi-omic approaches in decoding the intricate genetic landscape of pulmonary fibrosis across diverse populations.

Figure 6.1 Schematic for transcriptomic and eQTL studies of the pulmonary fibrosis population



Subjects and Tissue Samples

All human tissue was collected after appropriate IRB and ethical review for the protection of human subjects through the University of Chicago. All subjects met American Thoracic Society/European Respiratory Society criteria for diagnosis of pulmonary fibrosis. Whole genome sequencing and Gene Expression Profiling DNA and RNA were isolated using standard methods.

Gene Expression Studies and Data Analysis

Gene expression profiling was performed using the PAXgene Blood RNA Kit (QIAGEN GmbH) for the extraction and purification of intracellular RNA from whole blood collected in the PAXgene Blood RNA Tube. Total RNA was extracted from these cells following standard protocols, then purified and quality-checked to ensure the integrity of the RNA samples. After plating in the Gilad lab, samples were transferred to the Genomics core for expression quantification where bulk RNA sequencing using the NovaSeq 6000 platform (Illumina, San Diego, CA) was performed on the extracted RNA to quantify gene expression levels. RNA sequencing data were available in 92 subjects with pulmonary fibrosis and WGS data available; 40 with honeycombing fibrosis and 52 with non-honeycombing pulmonary fibrosis. Expression array data were normalized using cyclic loess. The R package, limma, was used to fit linear models for expression data and P values were based on the moderated t statistic followed by false discovery rate (FDR) correction to obtain q values.

Whole Genome Sequencing, Telomere Lengths, and Ancestral analyses

I performed whole-genome sequencing (WGS) for all included subjects. The acquired WGS data was then subjected to rigorous quality control (QC) checks to ensure its robustness and reliability (**Figures 6.2 and 6.3**). Subsequent to QC validation, we matched the sequence to the genotype, a vital step in discerning gene variations associated with honeycombing fibrosis.

Simultaneously, we determined the racial ancestry of study participants utilizing the "Admixture" program to analyze PLINK (.bed) files [143]. This software facilitates the estimation of ancestry in a model-based manner from large autosomal SNP genotype datasets, where the individuals are unrelated.

Figure 6.2 Bioinformatic workflow used for the eQTL analysis

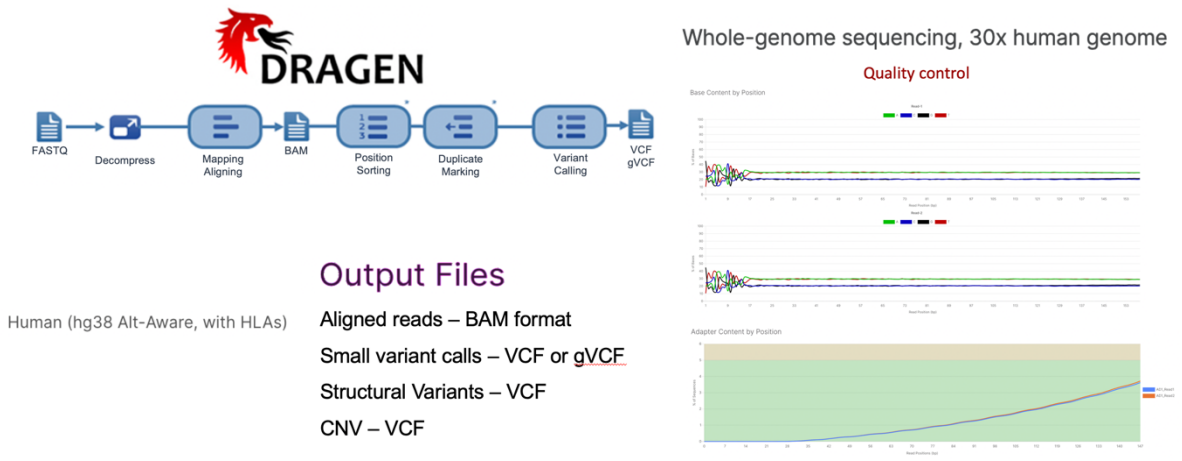
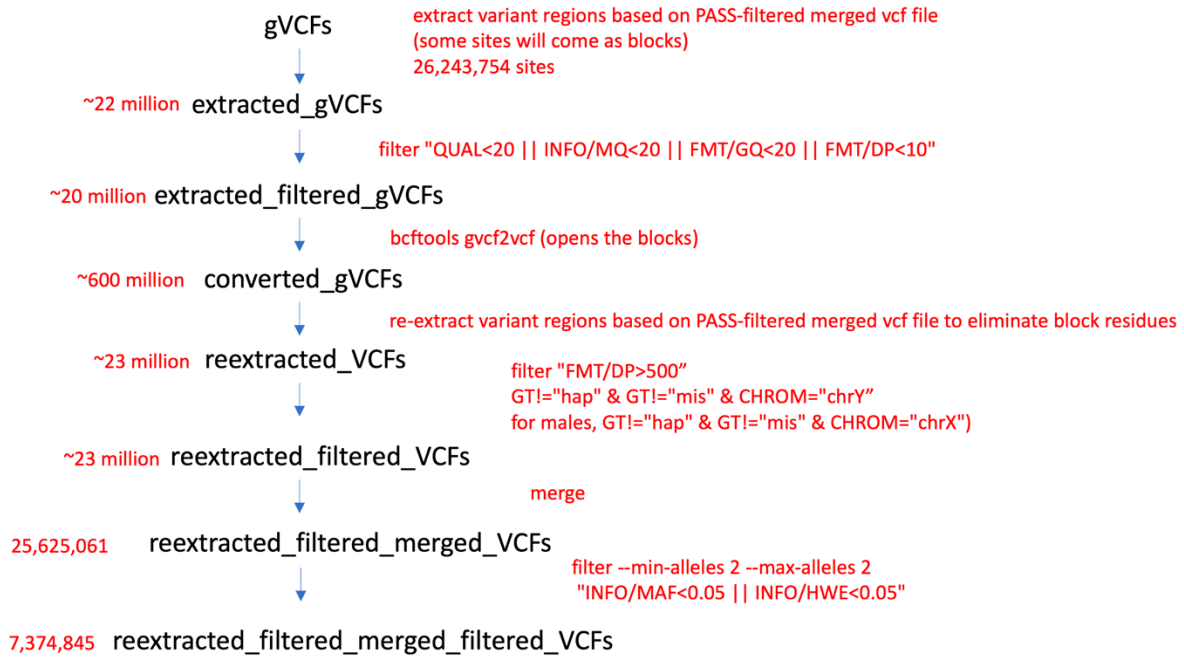


Figure 6.3 Pre-processing and filtering of whole genome sequencing genome variant call files.



Additionally, telomere lengths were estimated from the created WGS bam files using the Telseq tool in bash [144, 145]. Telseq measurements offered the estimated telomere length for each of 10 possible telomere repeats and provided a summary of statistics for the input BAM file, such as the number of mapped and unmapped reads, and the number of duplicate reads. These measurements were comparatively analyzed in tandem with qPCR and FlowFISH telomere length measurements from the same individuals with pulmonary fibrosis.

eQTL discovery

Next, I prepared the input files from WGS, RNAseq, and clinical phenotype data for expression quantitative trait loci (eQTL) analyses. eQTL analysis is a powerful approach used to link genetic variants to alterations in gene expression. In the context of our study, it can provide valuable insights into the molecular mechanisms contributing to the pathogenesis of pulmonary fibrosis. Our quantitative trait loci (QTL) models were run in QTLtools [146], a toolset for molecular QTL discovery and analysis, where we fit expression data to WGS genotyping data and adjusted for age, sex, genotype and phenotype principal components, and probabilistic estimation of expression residual factors.

We then performed gene expression quantification and generated Principal Component Analysis (PCA) plots to visualize and understand the variance in our dataset. Our eQTL discovery comprised several stages: nominal, permutation, conditional, and full pass. All these steps are necessary for a robust and comprehensive cis-QTL analysis. The nominal pass identifies potential SNP-gene associations, the permutation pass corrects for multiple testing issues, and the conditional pass uncovers multiple independent cis-QTLs for the same gene. This integrative approach maximizes the likelihood of identifying true cis-QTLs while minimizing the risk of false positives. Subsequently, we determined the annotation density at QTLs, analyzed QTL enrichment in annotations, and investigated the overlap of QTLs with Genome-Wide Association Study (GWAS) hits.

Colocalization analysis and SNP association tests

In our study, we employed Bayesian statistical colocalization analysis to dissect the intricate genetic associations underlying pulmonary fibrosis [147]. This powerful statistical approach allowed us to probabilistically assign shared genetic variants to multiple traits, thus revealing the interconnected genetic architecture of disease risk. Utilizing this robust method, we were able to integrate multiple datasets and decipher complex genotype-phenotype relationships, a critical step in illuminating the genetic underpinnings of pulmonary fibrosis. Moreover, we also performed single nucleotide polymorphism (SNP) association tests considering the genotype of each SNP under an additive model (0, 1, or 2 effect alleles) and tested for association with pulmonary fibrosis using logistic regression. In the association testing we included the covariates age, sex, the first 4 phenotype principal components, and the first genotype principal component which accounted for estimates of ancestry principal components. To complement our genetic analyses, we also investigated gene expression. For this, RNA sequencing data were available, which we subjected to an additive effects linear model to test for eQTL effect sizes (β) between the four core-region SNPs and each of the expressed genes at the extended locus. Gene expression counts were normalized, and we used linear regression considering additive genotype effects on gene expression.

Overall, the integrated approach adopted in this study, encompassing whole-genome sequencing, eQTL analysis, SNP association tests, and gene expression

studies, provides a comprehensive methodology to identify host defense and cell senescence gene variants contributing to the risk of pulmonary fibrosis.

Results

Demographic Characteristics

The current study cohort comprised 100 participants with an average age of 66 years, exhibiting a standard deviation of 12 years. Predominantly, males formed the majority of the study population (55%), and about one-third of the participants self-identified as non-White subjects. Among the non-White subjects, autoimmune-related subtypes of pulmonary fibrosis were notably prevalent. In contrast, white subjects were predominantly afflicted with fibrotic Hypersensitivity Pneumonitis (HP) and unclassifiable subtypes. However, these differences did not reach statistical significance.

Table 6.1 Clinical Demographics in the Pulmonary Fibrosis Study Cohort

Clinical Demographics	WHITE SUBJECTS (n=68)	NON-WHITE SUBJECTS (n=32)	P-value[#]
Age, years	67 (12)	65 (13)	0.46
Male	40 (59)	15 (47)	0.26
Pulmonary fibrosis subtype			0.18
IPF	20 (29)	7 (22)	
IPAF	6 (9)	6 (19)	
CTD-ILD	11 (16)	10 (31)	
Fibrotic HP	24 (35)	7 (22)	
Unclassifiable/other ILD	7 (10)	2 (6)	
Radiologic honeycombing	35 (51)	10 (31)	0.06
DTA Honeycombing score	6.2 (10.5)	4.5 (9.3)	0.46
UIP pattern on chest CT	35 (54)	9 (29)	0.023
Mortality patterns			
Survival time, months	41 (37)	33 (34)	0.28
Number of deaths[*]	7 (11)	2 (6)	0.46

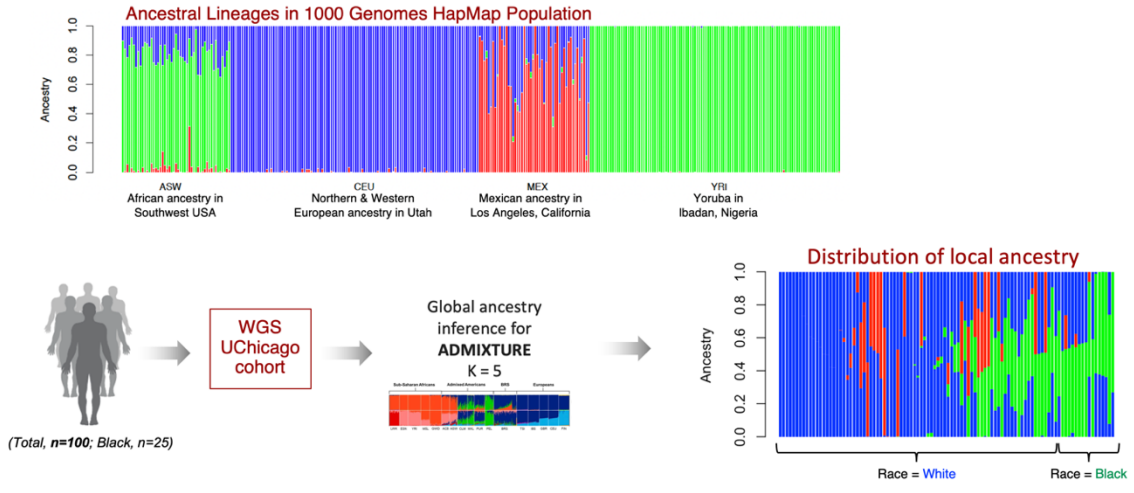
Total sample size, n=100. [#]P-value for statistical test comparing White subgroup to Non-White subgroup. Categorical variables presented as n (%); continuous variables presented as mean (SD). Non-White subjects include Non-Hispanic Black subjects (n=23), Non-Hispanic Asian subjects (n=5), and Hispanic subjects (n=4). UIP pattern = Usual Interstitial Pneumonia pattern. *[*]Lung transplantation counted as events. PF=Pulmonary fibrosis; ILD=interstitial lung disease; IPF=Idiopathic pulmonary fibrosis; IPAF=interstitial pneumonia with autoimmune features; CTD-ILD=Connective tissue disease associated-ILD; Fibrotic HP=Fibrotic hypersensitivity pneumonitis; Unclassifiable/Others.*

A crucial finding was the presence of radiologic Usual Interstitial Pneumonia (UIP) pattern being more prominent among the White cohort. This cohort exhibited increased high-resolution computed tomography (HRCT) evidence of honeycombing, a critical component of the UIP pattern. Despite the slight disparity, the mortality patterns remained similar between White and non-White subjects, but with significant variability in survival time within both groups.

Genetic Ancestry

Subsequent to the demographic analysis, a global ancestry inference analysis was performed at a K=5 threshold. The results demonstrated considerable discordance between self-identified race and nominal ancestral markers across all analyzed racial and ethnic populations (**Figure 6.4**). A substantial proportion of individuals who self-identified as Black had a high fraction of northern and western European ancestral markers within their genome, highlighting that self-identified race is not equivalent to ancestry.

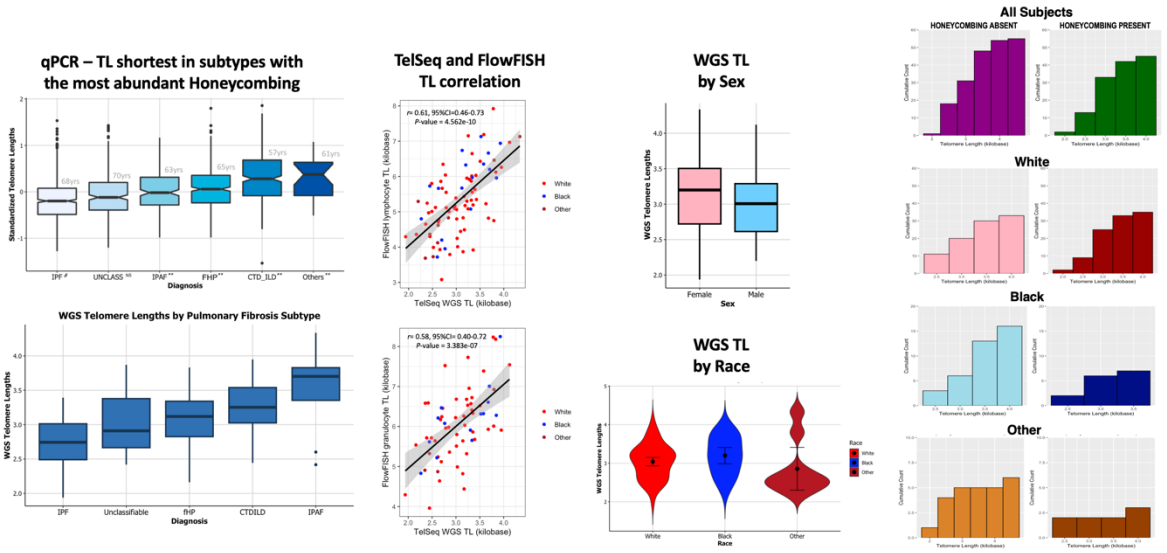
Figure 6.4 Global ancestry inference demonstrates that self-identified race is not equivalent to ancestry.



WGS-derived Telomere Length Differences

Moving onto genomic analysis, telomere length measurements derived from whole-genome sequencing (WGS) echoed similar patterns as observed with qPCR derived telomere length (**Figure 6.5**). There was modest to high correlation in the telomere length measurements between Telseq, and FlowFISH derived telomere indices measured independently on the granulocyte (r=0.58, 95% CI=0.40-0.72; P< 0.0001) and lymphocyte (r=0.61, 95% CI=0.46-0.73; P<0.0001) subsets of peripheral leukocytes. Shorter telomere lengths were observed in the pulmonary fibrosis subtypes with the most abundant radiologic honeycombing patterns present. Subjects with IPF, and unclassifiable ILD had the shortest telomere length measurements. While those with autoimmune-related pulmonary fibrosis, and background connected tissue disease had higher telomere lengths. Corroboratively, WGS-derived telomere lengths were notably longer in females and in Black subjects.

Figure 6.5 Characteristics of qPCR, WGS, and Flow-FISH telomere lengths in the pulmonary fibrosis dataset.



Moreover, telomere length was shorter among subjects exhibiting honeycomb fibrosis, with this difference being more pronounced in Black and other non-white populations.

Widespread Identification of Gene Variants in Blood from Pulmonary Fibrosis Patients

We observed modest correlation between Principal Component Analyses (PCAs) for WGS and RNA-seq, which persisted after filtering out uninformative XY-linked genes (**Figure 6.6-6.7**). The quantitative trait loci (QTL) discovery process yielded 2,291 loci identified after false discovery rate (FDR) correction at P<0.05 (**Table 6.2**). The majority of identified QTL loci were clustered around the gene transcription start site, indicating their potential functional relevance.

Figure 6.6 Heatmap of sex-linked genes and Spearman correlation of phenotype vs. genotype principal components.

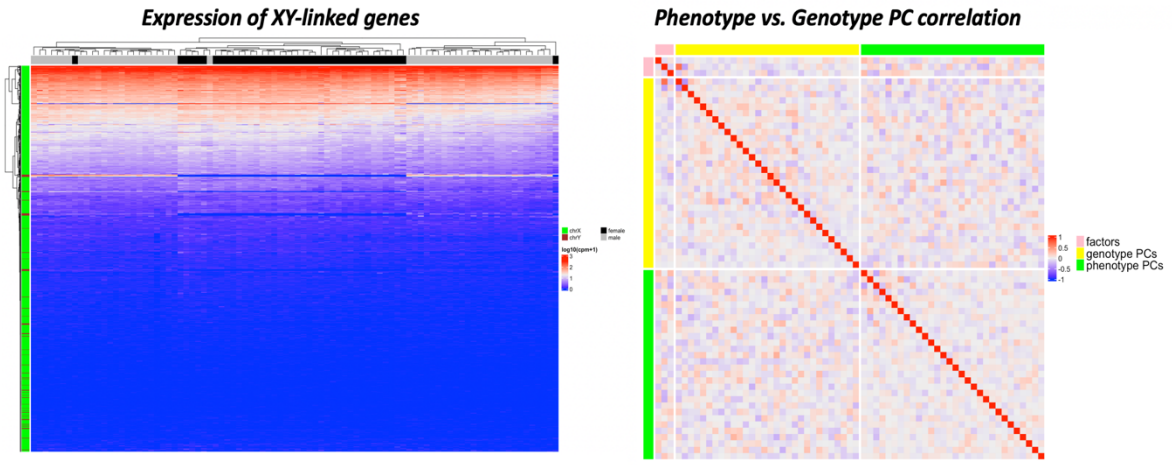


Figure 6.7 Heatmap of phenotype-genotype correlation from regenerated variant call files

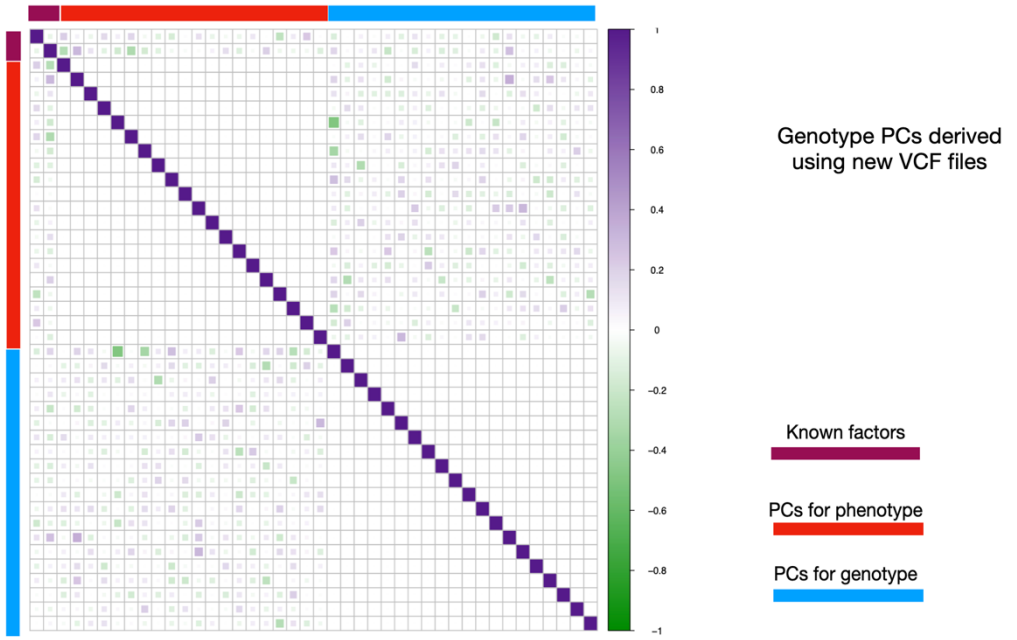


Table 6.2 cis eQTL results for genes identified in the NHGRI Catalog of Published Genome-Wide Association Studies.

pid	top_variant_id	top_variant_chr	nom_P	adj_P	qval	gene_symbol
ENSG00000096696.15	rs35881062	chr6	7.98371E-11	2.32433E-07	9.28873271151691E-06	DSP
ENSG00000142002.19	rs8108044	chr19	5.26702E-07	0.00173178	0.0197569857261667	DPP9
ENSG00000128928.10	rs77531164	chr15	2.14145E-06	0.00377632	0.0367609279161425	IVD
ENSG00000138640.15	rs9991237	chr4	2.00631E-05	0.0271119	0.159529324514877	FAM13A
ENSG00000164362.21	rs424761	chr5	1.54287E-05	0.0528856	0.240935594801318	TERT
ENSG00000140694.18	rs537233264;rs71389717	chr16	9.13932E-05	0.106247	0.354117176300258	PARN
ENSG00000196126.12	rs9269130	chr6	4.08678E-05	0.114216	0.36920951015535	HLA-DRB1
ENSG00000124762.14	rs7739790	chr6	0.000103616	0.131692	0.395810569553979	CDKN1A
ENSG00000068650.19	rs12877526	chr13	7.75703E-05	0.142609	0.410795816515081	ATP11A
ENSG00000164342.14	rs6840864	chr4	7.1067E-05	0.169477	0.44131283623377	TLR3
ENSG00000186868.18	rs67340844	chr17	0.000200027	0.205817	0.47932114180929	MAPT
ENSG00000078902.16	rs4881750	chr11	0.00010344	0.209544	0.483218878342341	TOLLIP
ENSG00000105329.11	rs546672523	chr19	0.000188358	0.209901	0.483450206461168	TGFB1
ENSG00000167972.14	rs11860424	chr16	0.000157491	0.220204	0.492141346996675	ABCA3
ENSG00000258366.12	rs6090241	chr20	0.000177203	0.273442	0.534450121232253	RTEL1
ENSG00000092330.19	rs12433068	chr14	0.00026732	0.295315	0.549848692073068	TINF2
ENSG00000171757.17	rs9849779	chr3	0.000546992	0.400111	0.610455790499408	LRRC34
ENSG00000130826.19	rs1402882622;rs58675912	chrX	0.00866376	0.696331	0.723173270210483	DKC1
ENSG00000117983.17	rs140845051	chr11	0.000888353	0.818055	0.749972125792683	MUC5B
ENSG00000136689.20	rs10593176;rs745343381;rs796778806	chr2	0.00182386	0.832396	0.753558778970648	IL1RN
ENSG00000141510.18	rs35353536	chr17	0.00123759	0.845592	0.756250116515678	TP53
ENSG00000170776.22	rs16977847	chr15	0.00184922	0.903823	0.770277245541903	AKAP13

Integration of informative genetic loci and gene expression in blood tissue

The QQ-Plot of QTLs showed optimal distribution of identified loci and the annotation density at QTLs and assessment of distance to transcription start site for identified QTLs demonstrated that the majority of variants discovered were clustered around the gene transcription start site (**Figures 6.8-6.10**). The eQTL permutation analyses identified several genes that represent potential functional loci in the pathogenesis of pulmonary fibrosis (**Figure 6.11**). These top hit genes include loci within the *HLA, ZNF, TMEM, LINC, MAP3K,* and *GATD* gene families. Overlaying the dataset with the NHGRI-EBI catalog of human GWAS studies led to the discovery of novel variants in three genes previously associated with Idiopathic Pulmonary Fibrosis (IPF): *IVD, DPP9,* and *DSP* (**Table 6.2**).

Figure 6.8 eQTL permutation pass after FDR adjustment.

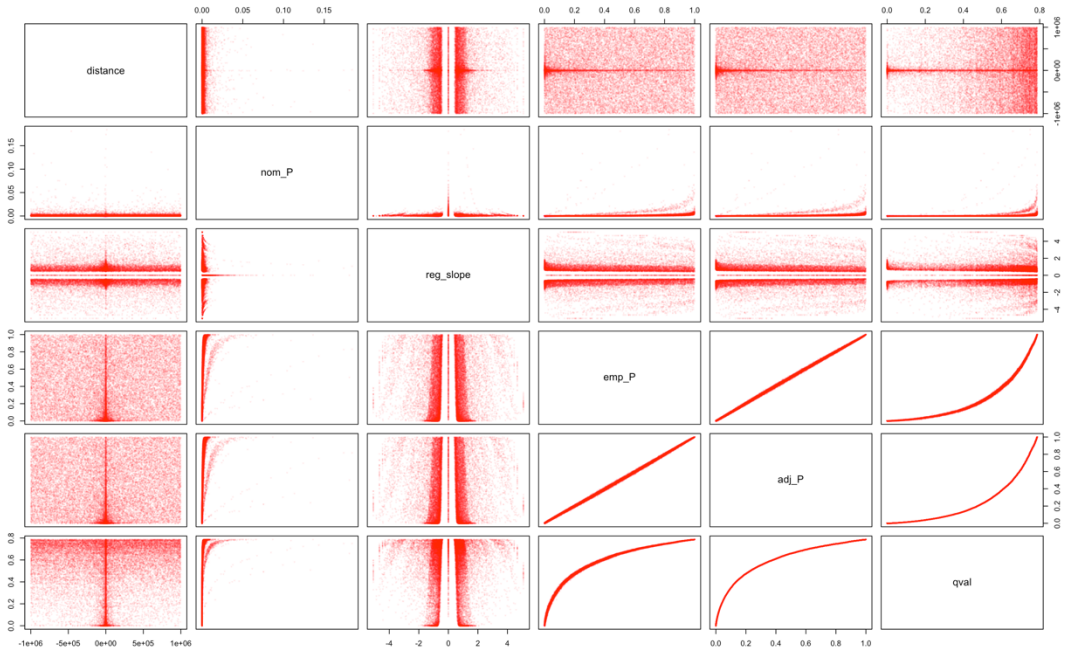


Figure 6.9 QQ plots depicting QTL distribution.

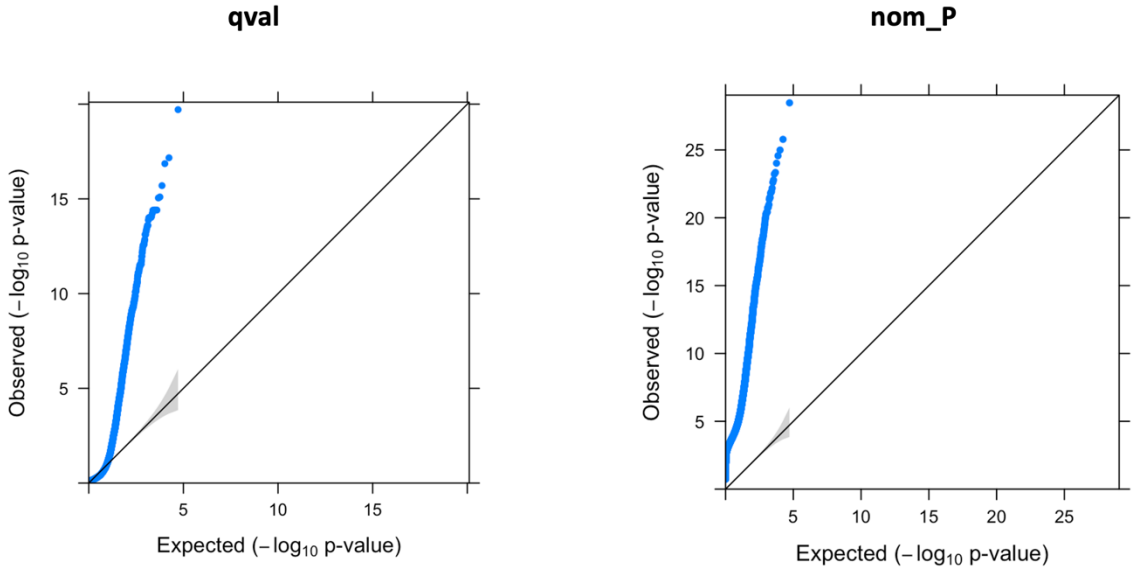


Figure 6.10 QTL annotation density and distance to transcription start site.

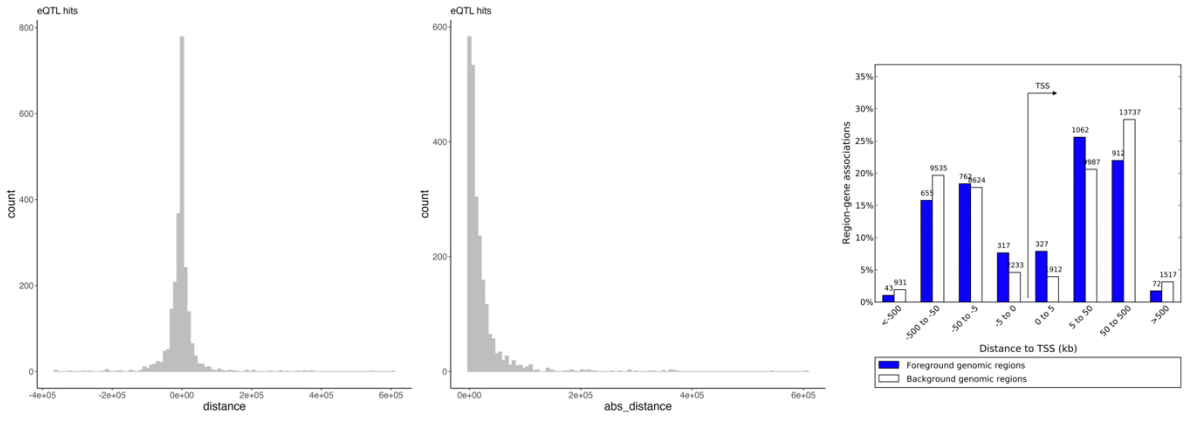
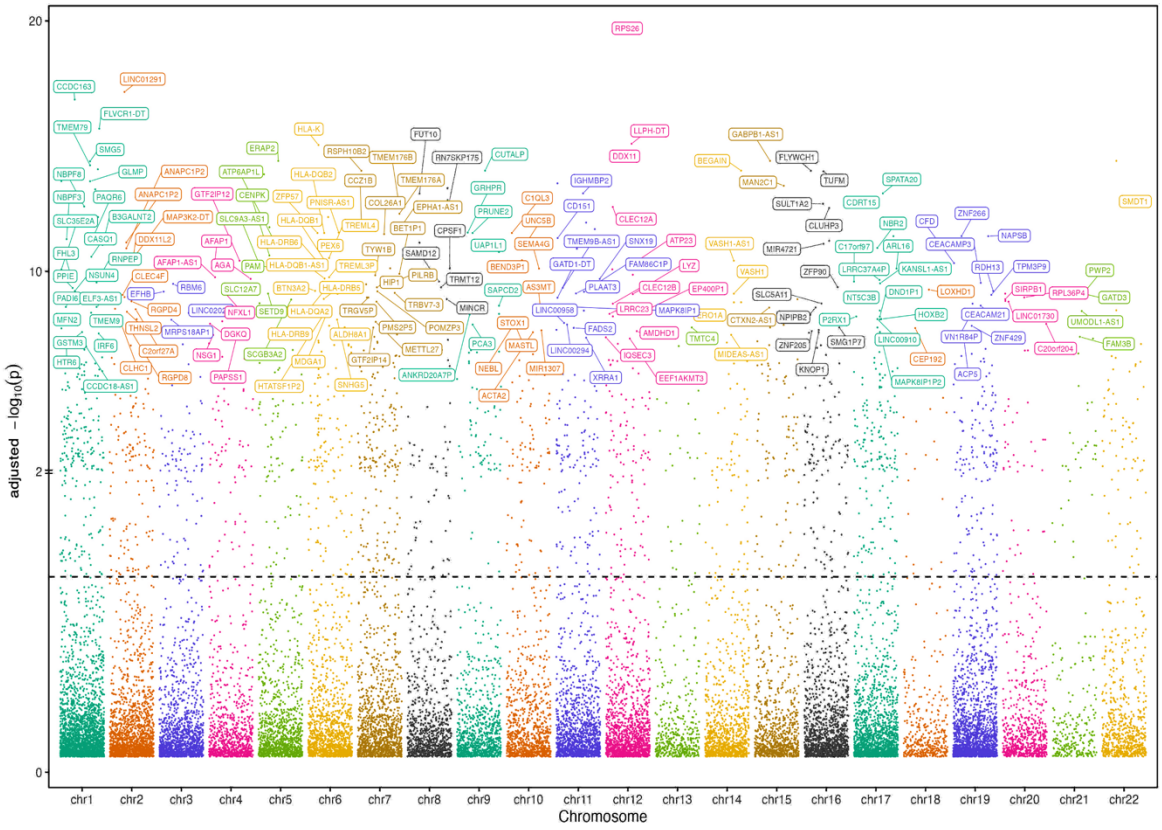


Figure 6.11 QTL discovery with permutation hits after FDR correction



Top gene pathways identified through gene ontology KEGG analyses included autoimmune thyroid disease, viral inflammatory organ involvement and Type 1 diabetes mellitus, all of which have been associated with the pathogenesis of pulmonary fibrosis [148-154]. These analyses also identified the involvement of immunologic T-cell driven mechanisms, cellular structure, and MHC protein complex genes linked to HLA pathways **(Figure 6.12-6.14**).

Figure 6.12 Gene ontology (GO) KEGG Enrichment Analysis

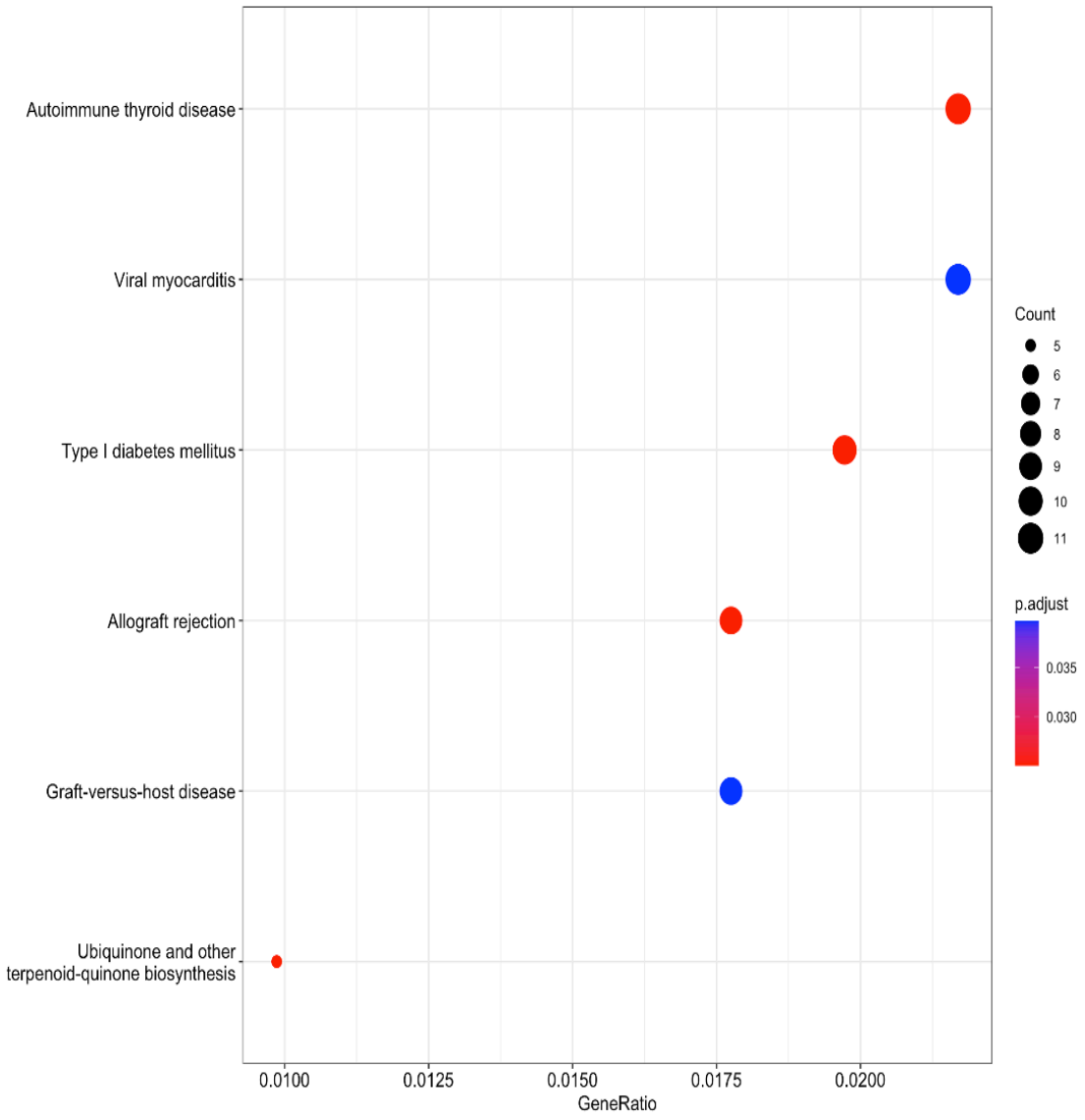


Figure 6.13 Overrepresented eQTL identified genes.

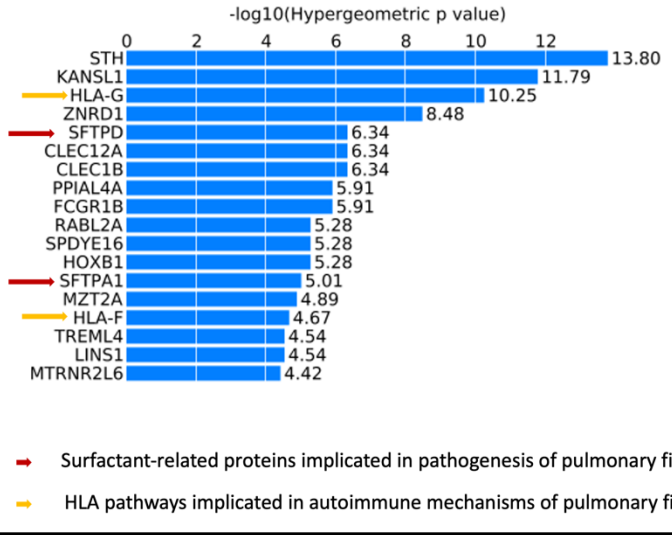
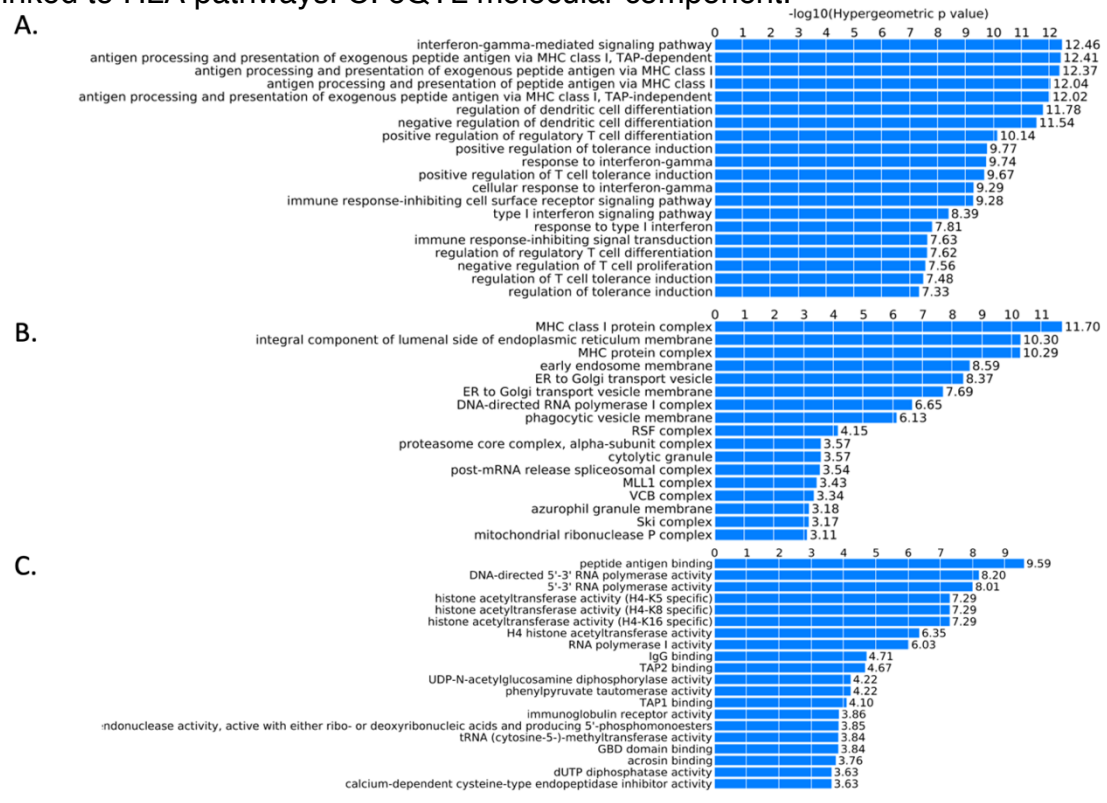


Figure. 6.14 Overrepresented eQTL gene pathways. A. Overrepresentation of immunologic gene ontology pathways, primarily T-cell driven mechanisms. B. eQTL cellular component. Several MHC protein complex genes noted which are linked to HLA pathways. C. eQTL molecular component.



IVD (Isobutyryl-CoA Dehydrogenase)

The gene *IVD* encodes an enzyme involved in the catabolism of the amino acid valine [155]. This metabolic process is critical in the provision of energy for cells and tissues, suggesting that disruptions to this process could potentially affect cellular homeostasis and contribute to disease pathology. Mutations in the *IVD* gene have been known to cause isobutyryl-CoA dehydrogenase deficiency, a rare metabolic disorder [155-157]. Although a direct link between this gene and pulmonary fibrosis remains elusive, the implication of a metabolic gene accentuates the role of altered cellular energetics in the pathogenesis of pulmonary fibrosis.

DPP9 (Dipeptidyl-Peptidase 9)

The *DPP9* gene encodes a protein that belongs to the S9B family in clan SC of the serine proteases [158, 159]. These proteins have dipeptidyl-peptidase activity and are thought to play a role in cellular processes, potentially influencing the balance between cell proliferation and cell death[160, 161]. Dysregulation of such processes could potentially contribute to fibrotic remodeling in pulmonary fibrosis, although specific roles and mechanisms involving *DPP9* in the pathogenesis of pulmonary fibrosis warrant further investigation.

DSP (Desmoplakin)

DSP encodes a protein that is a critical component of desmosomes, cell-cell junctions between epithelial, myocardial, and certain other cell types [162, 163].

Desmosomes maintain the structural integrity of these tissues, and mutations in *DSP* have been associated with various dermatological and cardiac diseases [162, 163]. Considering the crucial role of epithelial injury and dysregulation in the pathogenesis of pulmonary fibrosis, the implication of a gene critical for epithelial cell integrity suggests a potential mechanistic link that warrants further investigation.

Notably, none of the previously identified variants within these three genes were in LD with our eQTL significant hits.

The overlay of eQTL identified loci with three recent IPF GWAS studies, specifically those by Allen et al.[14], Peljto et al.[164], and a meta-GWAS [48] from the International IPF Genetics consortium, identified significant variant hits in four genes: *LRRC37A4P, KANSL1-AS1, LINC02210,* and *MIR4315-1* (**Figure 6.15**) (**Table 6.3-6.4**).

Table 6.3 cis eQTL top permutation hits and accompanying multi-GWAS p-values.

pid	top_variant_id	top_variant_chr	nom_P	adj_P	qval	gene_symbol	meta_gwas_pval	allen_gwas_pval	peljto_gwas_pval
ENSG00000197728.11	rs1131017	chr12	3.3616E-29	9.5278E-25	1.953298765764345E-20	RPS26	0.154652603144041	0.1854769155563983	0.6366
ENSG00000204792.2	rs12366	chr2	1.65166E-26	6.64191E-22	6.808305487789399E-18	LINC01291	0.0208574726684615	0.0433637309045467	0.3266
ENSG00000280670.3	rs138478267;rs58962251	chr1	2.72236E-25	2.00539E-21	1.370419828247999E-17	CCDC163	NA	NA	NA
ENSG00000198468.11	rs11120049	chr1	1.03476E-25	3.88543E-20	1.991384586016796E-16	FLVCR1-DT	1	0.952104508785226	0.05369
ENSG00000239335.4	rs1370126	chr12	9.56988E-25	1.94841E-19	7.988889036716086E-16	LLPH-DT	0.595261266236448	0.116746411956424	0.6807
ENSG00000230795.3	rs13201160	chr6	4.53714E-24	2.62944E-19	8.984378881884344E-16	HLA-K	0.502876095538885	0.491834152995857	NA
ENSG00000273295.1	rs5760119	chr22	3.29914E-22	1.88341E-18	3.861187712196466E-15	ENSG00000273295	0.468446908425735	0.604298733820454	0.5584
ENSG00000164308.17	rs2927608	chr5	1.64208E-23	1.56626E-18	3.861187712196466E-15	ERAP2	0.527012805129425	0.435926772091405	0.5511
ENSG00000013573.17	rs10771802	chr12	6.06892E-24	1.70657E-18	3.861187712196466E-15	DDX11	NA	NA	NA
ENSG00000244879.10	rs504781	chr15	6.6713E-23	1.5916E-18	3.861187712196466E-15	GABPB1-AS1	0.182132044085486	0.642455383745169	0.4637
ENSG00000198952.8	rs10542877;rs398053427;rs755624161	chr1	1.53796E-22	2.14677E-18	4.001002989699055E-15	SMG5	NA	NA	0.3762
ENSG00000163472.19	rs10542877;rs398053427;rs755624161	chr1	3.2538E-22	3.39169E-18	5.794433037053195E-15	TMEM79	NA	NA	0.3762
ENSG00000286619.1	rs9438393	chr1	1.28423E-22	5.31645E-18	8.384060912635515E-15	ENSG00000286619	0.222674967806327	0.911923644923777	0.8608
ENSG00000226752.10	rs29001669;rs72298225	chr9	3.80221E-21	6.9176E-18	9.454536228751846E-15	CUTALP	NA	NA	0.4318
ENSG00000183092.18	rs11628965	chr14	4.4304E-22	6.50647E-18	9.454536228751846E-15	BEGAIN	0.419002194909154	0.710588463335882	0.2981
ENSG00000059122.17	rs781299714	chr16	1.17061E-22	7.43327E-18	9.52436364939086E-15	FLYWCH1	NA	NA	NA
ENSG00000169402.16	rs4476910	chr7	2.43076E-23	7.90775E-18	9.536303248182669E-15	RSPH10B2	NA	NA	0.03383
ENSG00000178952.11	rs7187776	chr16	1.33945E-21	8.85057E-18	1.008033069153376E-14	TUFM	0.988529488632586	0.596547226230265	0.8871
ENSG00000250155.3	rs1025288	chr5	5.55706E-21	1.22691E-17	1.3238389315603E-14	ENSG00000250155	0.838879533741699	0.249179803474062	0.4321
ENSG00000272004.1	rs79113395	chr1	3.80739E-21	2.42463E-17	2.485372691719526E-14	ENSG00000272004	NA	NA	NA
ENSG00000198715.13	rs10542877;rs398053427;rs755624161	chr1	1.91776E-21	2.5816E-17	2.522062057045566E-14	GLMP	NA	NA	0.3762
ENSG00000279672.2	rs10678686	chr11	1.05895E-21	3.27848E-17	3.055103313738444E-14	ENSG00000279672	NA	NA	0.842
ENSG00000140400.18	rs4393535	chr15	5.19696E-21	4.261E-17	3.798041813073366E-14	MAN2C1	NA	NA	0.9647
ENSG00000287576.1	rs1563152	chr8	1.83052E-21	5.56035E-17	4.749708218849411E-14	ENSG00000287576	0.0647636667876307	0.233578661049401	0.5334
ENSG00000286966.1	rs1174656	chr1	8.03939E-21	6.4379E-17	5.279347645462144E-14	ENSG00000286966	0.0235045205409007	0.0344353190750692	0.5776
ENSG00000172728.16	rs2732317	chr8	1.11792E-20	1.01954E-16	7.658239362895552E-14	FUT10	0.1883542067207	0.205004496804203	0.2246
ENSG00000006282.21	rs9890200	chr17	1.04129E-20	1.04595E-16	7.658239362895552E-14	SPATA20	0.259933729629069	0.0421819023836256	0.8359
ENSG00000132740.10	rs200447892	chr11	1.64364E-21	9.96384E-17	7.658239362895552E-14	IGHMBP2	NA	NA	NA
ENSG00000254859.1	rs6558372	chr8	4.74139E-21	1.97145E-16	1.393682389958881E-13	ENSG00000254859	NA	NA	NA
ENSG00000183172.9	rs61527837	chr22	4.0103E-21	2.36113E-16	1.613521244780911E-13	SMDT1	0.133345840915684	0.0477566736716064	0.3849
ENSG00000197165.11	rs28772958	chr16	3.408166E-20	3.0274E-16	2.002092580513896E-13	SULT1A2	0.939376477121766	0.382603731573791	0.6298
ENSG00000270231.5	rs61807936	chr1	3.25928E-19	3.97541E-16	2.5468771143403E-13	NBPF8	NA	NA	0.05044
ENSG00000172322.14	rs12231872	chr12	3.97596E-21	4.31525E-16	2.6808225304836E-13	CLEC12A	0.572211891056593	0.730877674125994	0.9545
ENSG00000223510.7	rs6502327	chr17	2.53213E-19	4.56401E-16	2.751970117146699E-13	CDRT15	0.68502252815966	0.514205335450361	0.7145
ENSG00000106565.18	rs201144751	chr7	3.12674E-20	4.71403E-16	2.761215744834488E-13	TMEM176B	NA	NA	NA
ENSG00000131797.14	rs74579683	chr16	2.09198E-19	5.13501E-16	2.924252274392077E-13	CLUHP3	NA	NA	0.3942
ENSG00000275558.1	rs1563152	chr8	1.62623E-20	6.6651E-16	3.693014298326443E-13	RN7SKP175	0.0647636667876307	0.233578661049401	0.5334
ENSG00000002933.9	rs4725943	chr7	1.11779E-19	9.19649E-16	4.961517754238288E-13	TMEM176A	NA	NA	0.09317
ENSG00000279942.1	rs1757171	chr6	7.07525E-20	1.30497E-15	6.796507135182311E-13	ENSG00000279942	0.0600793153994227	0.240386753344089	0.08981
ENSG00000261000.1	rs6668883	chr1	4.03212E-19	1.32608E-15	6.796507135182311E-13	ENSG00000261000	0.225897337624653	0.254390663778108	0.1763
ENSG00000146574.16	rs4476910	chr7	6.4409E-20	2.07482E-15	1.037462988968669E-12	CCZ1B	NA	NA	0.03383
ENSG00000286943.1	rs12577614	chr11	3.09838E-19	2.30426E-15	1.124755788082736E-12	ENSG00000286943	0.32445049908056	0.523346874335811	0.5443
ENSG00000250081.1	rs468690	chr5	2.14148E-18	3.60958E-15	1.720934176689435E-12	ENSG00000250081	0.0595900021616888	0.0135901736221241	0.8354
ENSG00000254851.1	rs1242127	chr11	4.16111E-19	4.40847E-15	2.054051161745515E-12	ENSG00000254851	NA	NA	0.132
ENSG00000228506.4	rs3087961	chr6	1.28488E-18	6.13939E-15	2.796976079248911E-12	PNISR-AS1	0.348312142022991	0.438477625802843	0.9142
ENSG00000137106.18	rs59967810	chr9	1.87559E-19	6.30359E-15	2.809352076093777E-12	GRHPR	0.613855886143683	0.647795237196819	0.1566
ENSG00000204644.10	rs2535238	chr6	5.10394E-20	7.09951E-15	2.929314224110296E-12	ZFP57	0.375424014643159	0.116389122896836	0.7227
ENSG00000232629.9	rs9273328	chr6	8.38197E-20	6.76475E-15	2.929314224110296E-12	HLA-DQB2	NA	NA	0.854
ENSG00000106772.19	rs620737	chr9	3.99073E-19	7.12111E-15	2.929314224110296E-12	PRUNE2	0.285761364241358	0.325021853672818	0.1127

Table 6.4. cis eQTL–GWAS significant variant hits and variants in linkage disequilibrium

GWAS significant variants among identified eQTL top variants

pid	top_variant_id	top_variant_chr	nom_P	adj_P	qval	gene_symbol	meta_gwas_pval	allen_gwas_pval	peljto_gwas_pval
ENSG00000214425.11	rs62054424	chr17	3.87739E-17	1.07996E-12	2.18553221544453E-10	LRRC37A4P	2.00861899567476E-19	5.80969030819687E-15	1.729E-07
ENSG00000214401.6	rs199447	chr17	4.03774E-15	1.58309E-12	3.00509283855166E-10	KANSL1-AS1	1.14457574359788E-17	2.12916903108744E-14	2.214E-08
ENSG00000204650.14	rs2942166	chr17	7.88015E-13	2.39101E-09	1.65602056185231E-07	LINC02210	3.77830228640565E-19	1.02677727407205E-14	1.855E-07
ENSG00000284055.1	rs62054424	chr17	9.22354E-07	0.00192314	0.0215444719799554	MIR4315-1	2.00861899567476E-19	5.80969030819687E-15	1.729E-07

Identified eQTL variants in LD with GWAS significant variants

pid	top_variant_id	top_variant_chr	adj_P	qval	gene_symbol	meta_gwas_pval	allen_gwas_pval	peljto_gwas_pval
ENSG00000177697.19	rs1130276	chr11	1.27E-14	4.65E-12	CD151	5.22E-05	0.009154723	0.04329
ENSG00000214425.11	rs62054424	chr17	1.08E-12	2.19E-10	LRRC37A4P	2.01E-19	5.81E-15	1.73E-07
ENSG00000214401.6	rs199447	chr17	1.58E-12	3.01E-10	KANSL1-AS1	1.14E-17	2.13E-14	2.21E-08
ENSG00000204650.14	rs2942166	chr17	2.39E-09	1.66E-07	LINC02210	3.78E-19	1.03E-14	1.86E-07
ENSG00000176681.15	rs2532240	chr17	0.00108983	0.01337083	LRRC37A	NA	NA	0.01864

LRRC37A4P and KANSL1-AS1

The *LRRC37A4P* gene, a leucine-rich repeat-containing member, has been implicated in immune-mediated genetic pathways that result in pulmonary function impairment and increased lung cancer susceptibility [165, 166]. Although classified as a pseudogene, these types of genes can exhibit regulatory roles influencing the expression of other genes.

Similarly, the *KANSL1-AS1* gene is known to play a role in lung health. As a cell type-specific and disease-associated eQTL, it appears to have a critical role in the function of ciliated epithelial cells [165, 166]. These cells, present in the lining of the respiratory tract, are key to maintaining lung health, suggesting that dysregulation of this gene might contribute to the pathogenesis of pulmonary fibrosis.

LINC02210 and MIR4315-1

The long intergenic non-protein-coding RNA 2210 (*LINC02210*) and *MIR4315-1* are both non-coding RNAs, which have been shown to be potent regulators of gene expression[166-168]. Their roles in pulmonary fibrosis, however, remain to be elucidated.

In addition to these genes, analysis of variants in LD with significant cis eQTL–GWAS loci identified variants in two additional genes, *CD151* and *LRRC37A*. The *CD151* molecule (Raph Blood Group) is a member of the transmembrane 4 superfamily, also known as the tetraspanin family [169, 170]. *CD151* is involved in facilitating signal transduction events that play a role in the regulation of cell development, activation, growth, and motility, implying its potential involvement in the cellular processes underpinning pulmonary fibrosis. *LRRC37A* (Leucine Rich Repeat Containing 37A) [166], is located on the human Y chromosome and belongs to a larger gene family characterized by leucine-rich repeat (LRR) domains critical for protein-protein interactions and mediate various cellular functions. It's suggested role in immune function and signal transduction could impact progression of pulmonary fibrosis.

After conditioning upon the top SNPs obtained from the permutation analyses, the obtained significant SNP-gene associations remained consistent suggesting that their effects were accounted for in the model even after repetition of this association test for all other SNPs in the region (**Table 6.4**).

Table 6.5 cis eQTL conditional pass identified loci.

pid	variant_id	variant_chr	rank_of_assoc	back_nomP	gene_symbol
ENSG00000171055.15	rs4670149	chr2	1	1.03029E-06	FEZ2
ENSG00000160226.16	rs2236441	chr21	1	1.2487E-06	CFAP410
ENSG00000160307.10	rs2839349	chr21	1	4.05383E-06	S100B
ENSG00000160310.19	rs9979822	chr21	1	4.71581E-07	PRMT2
ENSG00000215790.9	rs11486028	chr1	1	9.53617E-08	SLC35E2A
ENSG00000187010.21	rs3093625	chr1	1	9.29007E-07	RHD
ENSG00000220785.10	rs380464	chr1	1	1.40878E-07	MTMR9LP
ENSG00000189269.12	rs16994633	chr22	1	4.62876E-09	DRICH1
ENSG00000288153.1	rs73881424	chr22	1	1.03946E-06	ENSG00000288153
ENSG00000213672.8	rs4955427	chr3	1	4.79264E-06	NCKIPSD
ENSG00000145736.15	rs79998530	chr5	1	2.40558E-06	GTF2H2
ENSG00000153113.24	rs78749432	chr5	1	2.48429E-06	CAST
ENSG00000164307.13	rs27033	chr5	1	1.23914E-06	ERAP1
ENSG00000145949.12	rs2038760	chr6	1	2.00445E-07	MYLK4
ENSG00000281831.1	rs9258899	chr6	1	1.48286E-07	HCP5B
ENSG00000204622.12	rs75029704	chr6	1	1.04965E-06	HLA-J
ENSG00000204520.14	rs2596529	chr6	1	7.81828E-07	MICA
ENSG00000198502.6	rs202190613;rs9281791	chr6	1	2.47139E-08	HLA-DRB5
ENSG00000237541.4	rs115879259	chr6	1	1.74768E-07	HLA-DQA2
ENSG00000178665.16	rs143665817	chr7	1	8.12646E-07	ZNF713
ENSG00000287985.1	rs202034997	chr7	1	3.40964E-06	ENSG00000287985
ENSG00000129422.15	rs13252921	chr8	1	3.23185E-07	MTUS1
ENSG00000120156.22	rs1322048	chr9	1	6.93658E-07	TEK
ENSG00000214401.6	rs1230063	chr17	1	1.84284E-06	KANSL1-AS1
ENSG00000228696.11	rs369332489	chr17	1	6.47939E-07	ARL17B
ENSG00000185829.19	rs9891103	chr17	1	6.6936E-07	ARL17A
ENSG00000238083.10	rs74521229	chr17	1	2.5713E-06	LRRC37A2
ENSG00000170412.18	rs899596	chr17	1	2.24314E-06	GPRC5C
ENSG00000271553.1	rs6467197	chr7	1	7.34547E-07	ENSG00000271553
ENSG00000112031.16	rs3215595;rs397776045;rs397811977	chr6	1	2.53339E-07	MTRF1L
ENSG00000131238.18	rs67280643;rs869034434	chr1	1	7.50503E-07	PPT1
ENSG00000162598.13	rs1709072	chr1	1	2.02124E-06	C1orf87
ENSG00000286619.1	rs879065773	chr1	1	2.57909E-06	ENSG00000286619
ENSG00000284713.2	rs72932523	chr11	1	1.65265E-06	SMIM38
ENSG00000109861.17	rs570279	chr11	1	8.10183E-07	CTSC
ENSG00000254851.1	rs543312	chr11	1	2.22607E-06	ENSG00000254851
ENSG00000251364.9	rs11041415	chr11	1	6.176E-07	SYT9-AS1
ENSG00000236935.3	rs11608180	chr11	1	1.27121E-06	ENSG00000236935
ENSG00000224769.1	rs13326315	chr3	1	5.06054E-07	MUC20P1
ENSG00000242086.8	rs370033802	chr3	1	1.38943E-06	MUC20-OT1
ENSG00000176945.18	rs879686846	chr3	1	2.78493E-07	MUC20
ENSG00000251301.8	rs11177250	chr12	1	1.45449E-06	LINC02384
ENSG00000175164.16	rs1291122587	chr9	1	7.19086E-11	ABO
ENSG00000255050.1	rs2290416	chr8	1	8.78192E-08	ENSG00000255050
ENSG00000147813.16	rs2290416	chr8	1	4.08404E-08	NAPRT
ENSG00000010278.15	rs59522682	chr12	1	1.93979E-07	CD9
ENSG00000225231.2	rs74065410	chr12	1	1.59491E-06	LINC02470
ENSG00000197728.11	rs111914426	chr12	1	4.66875E-07	RPS26
ENSG00000233608.4	rs3791529	chr2	1	4.27206E-07	TWIST2
ENSG00000204792.2	rs72903286	chr2	1	4.0009E-07	LINC01291
ENSG00000236209.1	rs13412501	chr2	1	6.44921E-07	ENSG00000236209
ENSG00000168874.13	rs62163128	chr2	1	4.73001E-07	ATOH8
ENSG00000144115.18	rs12992162	chr2	1	1.35622E-07	THNSL2
ENSG00000269001.5	rs6509734	chr19	1	1.18051E-06	ZNF818P
ENSG00000241015.2	rs4803123	chr19	1	6.65919E-07	TPM3P9
ENSG00000160113.6	rs1043963	chr19	1	2.61831E-07	NR2F6
ENSG00000197013.11	rs7251393	chr19	1	1.42108E-07	ZNF429
ENSG00000259363.6	rs2603218	chr15	1	1.00834E-07	ENSG00000259363
ENSG00000259363.6	rs2603218	chr15	1	1.00834E-07	ENSG00000259363
ENSG00000205853.12	rs2281034	chr22	1	2.4476E-06	RFPL3S
ENSG00000249310.2	rs12628403	chr22	1	2.93568E-07	APOBEC3B-AS1
ENSG00000273243.1	rs28685010	chr22	1	3.05035E-08	ENSG00000273243
ENSG00000092036.19	rs59423232	chr14	1	6.4646E-07	HAUS4
ENSG00000259132.1	rs59423232	chr14	1	8.7344E-07	ENSG00000259132
ENSG00000126790.12	rs113870438	chr14	1	1.81183E-07	L3HYPDH
ENSG00000197253.13	rs75105689	chr16	1	7.32839E-07	TPSB2
ENSG00000241525.4	rs9899842	chr17	1	1.12334E-06	ENSG00000241525
ENSG00000167711.14	rs11654472	chr17	1	1.55783E-06	SERPINF2

Colocalization Analysis

Our research undertook Bayesian statistical colocalization analysis to probe into the intricate genetic associations involved in pulmonary fibrosis, focusing on honeycombing fibrosis and UIP. With specific emphasis on the loci near the *LRRC37A4P*, *KANSL1-AS1*, *LINC02210*, *MIR4315-1, LRRC37A*, and *CD151* genes, the Bayesian hypotheses testing aided in the discernment of likely causal variants for the phenotypes under investigation. The Bayesian probabilities (H_0-H_4) obtained from the analysis are indicative of the varying hypotheses regarding the potential shared causal variant(s). They represent: no association with either trait (H_0), association with trait 1 only (H_1), association with trait 2 only (H_2), two distinct causal variants (H_3), and a shared causal variant (H_4) (**Table 6.6**).

Table 6.6. Colocalization analysis for cis eQTL and GWAS significant variant hits including variants in linkage disequilibrium.

GENE	No. of SNPs tested	H_0	H_1	H_2	H_3	H_4
LRRC37A4P	3.07e+03	6.71e-20	1.69e-16	2.68e-04	6.74e-01	3.26e-01
KANSL1-AS1	2.80e+03	7.09e-19	1.72e-16	2.83e-03	6.85e-01	3.12e-01
LINC02210	2.93e+03	1.19e-18	1.67e-16	4.74e-03	6.68e-01	3.27e-01
LRRC37A	2.71e+03	8.30e-17	1.54e-16	3.32e-01	6.13e-01	5.50e-02
CD151	1.29e+03	3.73e-104	2.19e-100	1.70e-04	9.99e-01	1.76e-05

Figure 6.15. Region plots of six novel pulmonary fibrosis susceptibility loci from discovery eQTL analysis and GWAS. Each point represents a variant with chromosomal position on the x axis and the 2log(P-value) on the y axis. Variants are colored in by linkage disequilibrium with the sentinel variant. Blue lines show the recombination rate, and gene locations are shown at the bottom of the plot. Region plots are shown for the five validated novel pulmonary fibrosis susceptibility loci, i.e., (A) the susceptibility signal on chromosome 17 near *LRRC37A4P*, (B) the susceptibility signal on chromosome 17 near *KANSL1-AS1*, (C) the susceptibility signal on chromosome 17 near *LINC02210*, (D) the susceptibility signal on chromosome 17 near *MIR4315-1*, (E) the susceptibility signal on chromosome 11 near *CD151*, and (F) the susceptibility signal on chromosome 17 near *LRRC37A*.

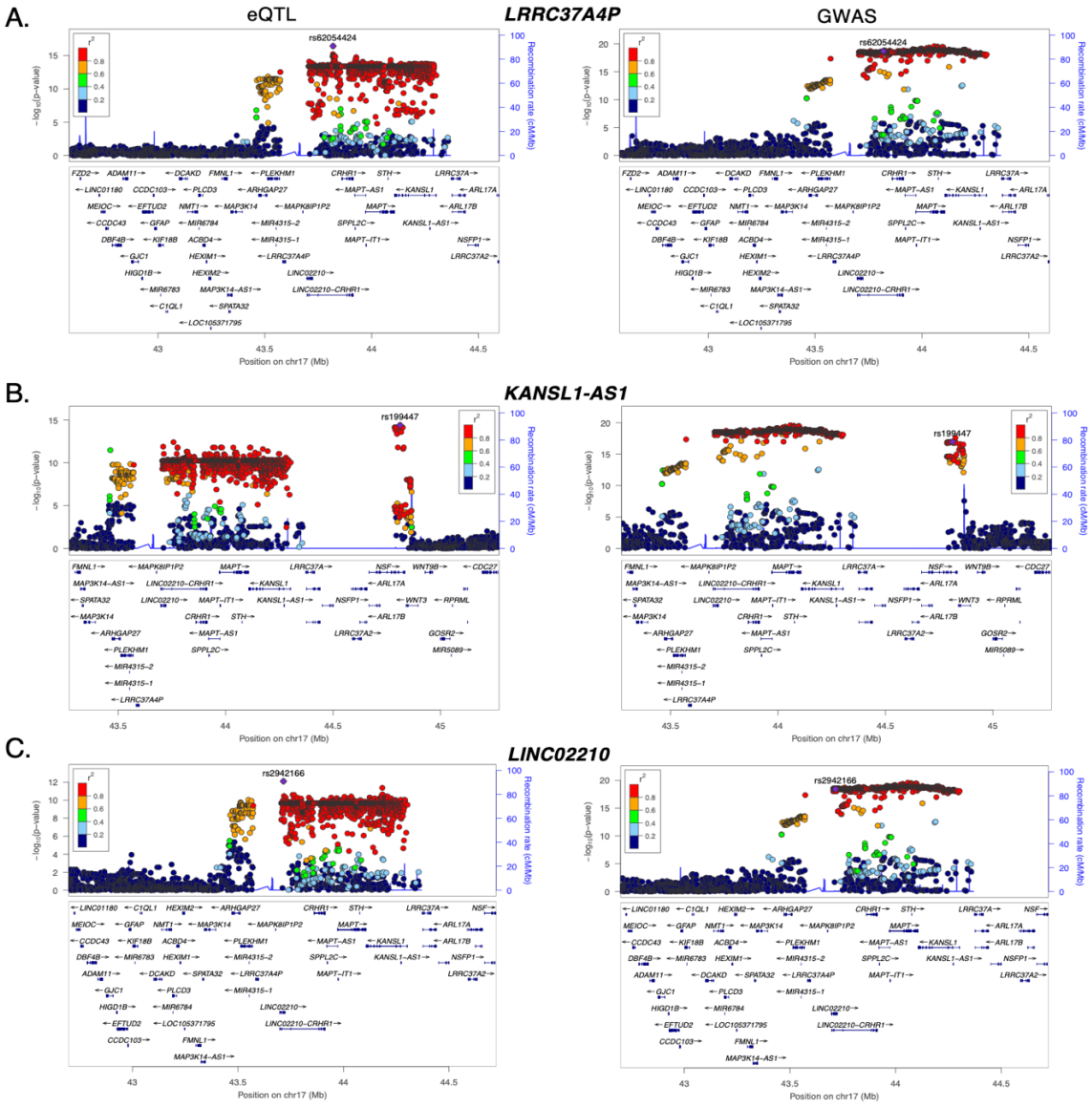
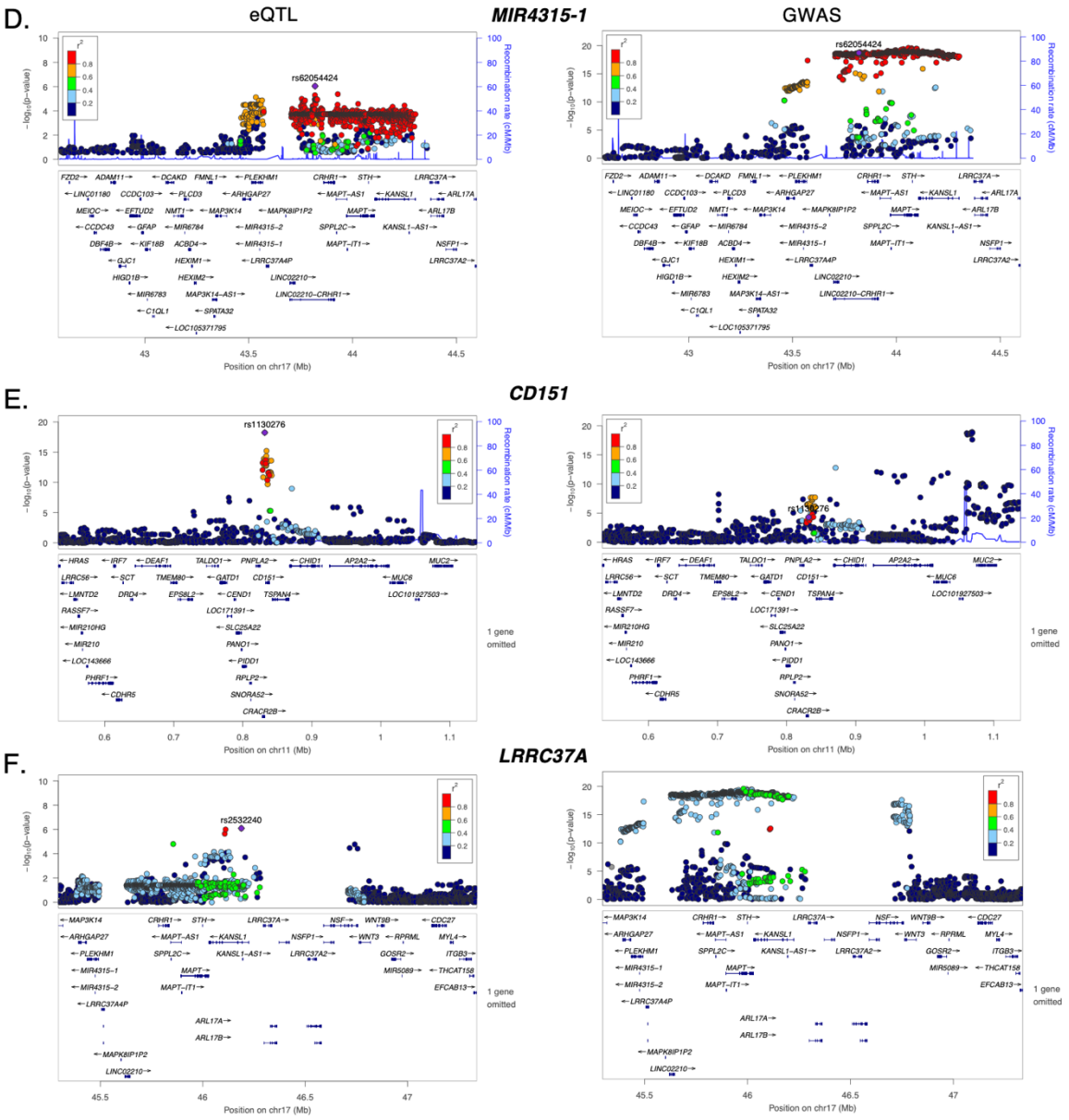


Figure 6.15, continued.



Starting with *LRRC37A4P*, the Bayesian analysis proposed a likely scenario of two distinct causal variants, each influencing a separate trait (H_3=6.74e-01), although the possibility of a shared causal variant is non-negligible (H_4=3.26e-01). The locus near *KANSL1-AS1* presented a similar scenario (H_3=6.85e-01, H_4=3.12e-01), implying the

existence of distinct causal variants for each trait but also indicating the potential for shared causal factors. Analyzing the leading QTL variant near the *LINC02210* gene suggested an even stronger possibility of two distinct causal variants (H_3=6.68e-01) and a significant probability of a shared causal variant (H_4=3.27e-01). This result further highlighted the complex genetic architecture involved in pulmonary fibrosis, necessitating the consideration of both shared and distinct genetic influences on related phenotypes.

While the lead SNP for *MIR4315-1* was significant in both QTL and GWAS analyses, there were no other variants in LD with this SNP in either QTL and GWAS studies.

Assessing the locus near the *LRRC37A* gene presented a nuanced case where both distinct (H_3=6.13e-01) and shared (H_4=5.50e-02) causal variants might be at play, with a significant possibility of association with trait 2 only (H_2=3.32e-01). In the case of the *CD151* gene, the Bayesian probabilities analysis strongly suggested (H_3=9.99e-01) the presence of two distinct causal variants impacting pulmonary fibrosis, each influencing a separate trait. This suggests a complex genetic landscape around *CD151*, underscoring the polygenic nature of pulmonary fibrosis (**Figure 6.15**).

Collectively, these analyses underscore the intricate and multifaceted nature of the genetic influences contributing to pulmonary fibrosis. This research lays critical groundwork for untangling the complex genetic landscape of pulmonary fibrosis. It underlines the importance of a nuanced view of the genetic architecture, taking into account both distinct and shared genetic influences across phenotypes. As we continue to unravel these findings, our understanding of the genetic contributions to honeycombing

fibrosis and UIP within the broader context of pulmonary fibrosis will be further refined, leading to more targeted therapeutic approaches and predictive models.

Finally, a notable finding was the discordance in regression coefficient (R^2) between subjects of European ancestry and those of African Ancestry for most significant variants in *CD151*, *KANSL1-AS1*, and *LRRC37A4P*. Trans eQTL top variant analyses for each gene was unrevealing for significant variants in GWAS studies. These observations underscore the importance of studying pulmonary fibrosis in diverse racial groups to capture the breadth of genetic variations influencing disease risk and progression.

In conclusion, the identified genes and variants represent promising starting points for further research into the complex genetic underpinnings of pulmonary fibrosis. Their potential functional implications highlight key areas such as cellular energetics, cell junction integrity, immune regulation, and non-coding RNA-mediated gene regulation as integral to disease pathogenesis. However, the findings also emphasize the need for deeper investigation, given the observed racial differences in variant associations and the multifaceted nature of pulmonary fibrosis pathogenesis.

To sum up, this study's findings reveal a complex genetic landscape associated with pulmonary fibrosis risk across diverse racial groups. The identified genetic variants, coupled with eQTL analyses, shed light on potential molecular mechanisms underlying the development of honeycombing and UIP patterns in pulmonary fibrosis.

Discussion

The findings obtained in this study underscore the intricate interplay of genetic factors contributing to the risk and progression of pulmonary fibrosis, with a particular focus on honeycombing fibrosis and UIP patterns. This investigation holds significant novelty as it represents the first eQTL study of honeycombing fibrosis conducted within multiethnic populations.

The demographic analysis emphasized significant differences in pulmonary fibrosis subtypes based on self-identified race, with autoimmune-related subtypes being prevalent among non-White subjects and fibrotic HP and unclassifiable subtypes being predominant among White subjects. An intriguing finding was the presence of radiologic UIP patterns more common in the White cohort, with an increased incidence of honeycombing, which might be an epiphenomenon of the increased propensity for an IPF diagnosis in older White individuals [171, 172]. However, mortality patterns did not differ significantly based on race, suggesting that genetic factors influencing disease progression may not directly translate to survival outcomes.

A key revelation of this investigation is the discordance between self-identified race and genetic ancestry, thus spotlighting the inadequacy of race as a surrogate for genetic ancestry in biomedical research. The consideration of genetic ancestry over self-identified race could contribute to a more nuanced understanding of disease risk, progression, and treatment outcomes, particularly in diseases like pulmonary fibrosis where genetic underpinnings are increasingly recognized.

Differences in telomere length, a marker of cellular aging, across the different pulmonary fibrosis subtypes and racial groups were another notable finding. Specifically, shorter telomere lengths were associated with the most abundant radiologic honeycombing patterns. This finding implies a potential link between cellular senescence, represented by telomere shortening, and honeycombing fibrosis. It also resonates with prior research demonstrating the role of cellular senescence in the pathogenesis of IPF, although the relationship with UIP needs further exploration [173, 174].

In terms of gene identification, the study discovered novel variants in three genes previously linked with IPF: *IVD*, *DPP9*, and *DSP*. While these genes are associated with various cellular functions such as amino acid catabolism (IVD), cellular processes including cell proliferation and death (*DPP9*), and cell-cell junction integrity (DSP), their precise roles in the pathogenesis of pulmonary fibrosis, specifically honeycombing fibrosis and UIP, warrant further investigation.

This study further discovered significant variant hits in four genes—*LRRC37A4P*, *KANSL1-AS1, LINC02210,* and *MIR4315-1*—when overlaying eQTL identified loci with recent IPF GWAS studies. Of these, *LRRC37A4P* and *KANSL1-AS1* are particularly compelling due to their implicated roles in immune-mediated genetic pathways and lung health, respectively. However, a deeper understanding of how these genes contribute to the development and progression of pulmonary fibrosis and specifically, their role in the formation of honeycombing and UIP patterns, is required.

Finally, the investigation revealed variants in LD with cis eQTL–GWAS significant variant hits in two additional genes, *CD151* and *LRRC37A*. The *CD151* molecule has a

key role in regulating cell development, activation, growth, and motility, suggesting its potential involvement in the cellular processes underpinning pulmonary fibrosis [169, 170].

In conclusion, the current investigation highlights a complex landscape of genetic variations associated with the risk and progression of pulmonary fibrosis across diverse racial groups. By unearthing these novel genetic insights, this study significantly advances our understanding of the genetic underpinnings of honeycombing fibrosis and UIP patterns. However, it also highlights the necessity for deeper investigation and the integration of these findings with pathological and clinical features of the disease.

These results ultimately underscore the complexity and heterogeneity of pulmonary fibrosis, revealing that it is not only a disease marked by diverse pathological and clinical manifestations but also by complex genetic underpinnings. As the first eQTL study of honeycombing fibrosis in multiethnic populations, it sets the stage for future investigations aimed at refining these findings and their implications for improving the diagnosis, risk stratification, and treatment of this devastating disease.

Chapter VII

Discussion, Future Directions, and Research Impact

Discussion

The goal of my doctoral training in Genetics Genomics and Systems Biology was to unravel the complex intersection between race, telomeres, and genomics in the context of pulmonary fibrosis, which has emerged as a central theme throughout this doctoral dissertation. As I parsed out the genetic underpinnings of pulmonary fibrosis my research aimed to comprehensively understand the interconnectedness of these elements in a way that profoundly shapes the discourse on the etiopathogenesis, diagnosis, and management of this enigmatic and devastating disease.

A seminal theme from my investigation was the pivotal role of telomeres, the nucleoprotein structures that safeguard chromosome ends, especially in relation to racial and ethnic disparities in pulmonary fibrosis. In chapter II, I meticulously investigated leukocyte telomere length across a diverse patient population and demonstrated the broad prevalence of shortened telomeres, which consistently correlated with worse survival outcomes. Notably, the dynamic interplay between leukocyte telomere length and chronological age varied significantly across racial and ethnic groups, suggesting a potential divergence from healthy controls. This discovery extends the work of Alder et

al., who first reported an association between shortened telomeres and pulmonary fibrosis in 2008 [175], and later confirmed by multiple studies, including a 2017 study by Newton et al. [26, 176, 177], which established that short telomeres are a risk factor for progression of the disease. My findings also echoed prior research associating telomere dysfunctions with both familial and sporadic cases of pulmonary fibrosis and importantly, further illuminated the impact of telomere shortening on the age of disease onset and prognosis. As telomere biology and its associated genetic determinants intersect with both physiological aging and cellular senescence processes, my results corroborated the involvement of an accelerated biological aging mechanism in pulmonary fibrosis, a view supported by other recent studies.

More interestingly, I noticed the dynamic interplay between leukocyte telomere length and chronological age varied significantly across racial and ethnic groups. This observation implies a potential divergence from the normal aging process, as reflected in healthy controls, a view supported by previous work from several investigators [178-180] including Selman et al., 2014 [181], and Alder et al., 2011 [182], both of whom indicated accelerated biological aging mechanisms in pulmonary fibrosis. This unraveling of the connection between cellular senescence, represented by telomere shortening, and pulmonary fibrosis gives us a deeper appreciation of the complex pathogenesis of the disease.

An intriguing aspect of my investigation pertains to the differential response to mycophenolate mofetil therapy based on telomere length among patients with fibrotic hypersensitivity pneumonitis. I demonstrated in chapter III, that patients with longer

telomeres who received this therapy showed improved survival outcomes compared to those with shorter telomeres who did not receive the treatment. This concept is corroborated by similar observations in IPF pharmacognostics where Newton et al showed in a post hoc analysis of the PANTHER-IPF trial that IPF patients with short telomeres who received prednisone, azathioprine and N-acetylcysteine had higher mortality profiles [57]. This observation expands upon the recognized concept of personalized medicine, suggesting the potential value of considering telomere length in treatment decisions. My findings also resonate with earlier studies by Le Saux et al. [183] and a 2020 study by Piñeiro-Hermida et al., [184] which emphasized the potential therapeutic benefits of using telomerase activators for pulmonary fibrosis. As such, considering telomere length in treatment decisions further solidifies the principle of personalized medicine, an approach increasingly encouraged by the wider medical community.

Furthermore, my work described in chapter IV established intriguing associations between leukocyte telomere length and predictors of respiratory impairment. The odds for developing such predictors with decreasing leukocyte telomere length quartiles differed across racial groups, suggesting intricate gene-environment interactions and their implications for the onset and progression of pulmonary fibrosis. The modulation of gene expression and function by environmental factors could interact with telomere biology disorders, autoimmune genes, and HLA genes, escalating the risk of autoimmune disease. This finding aligns with previous research by Newton et al., 2019 [26], which showed telomere length as a predictor of disease progression and survival. It also aligns

with work by Armanios et al., 2007 [20], which reported the relationship between telomere length and familial pulmonary fibrosis. Therefore, exploring these gene-environment interactions and their mechanistic role in autoimmune forms of pulmonary fibrosis will further enhance our understanding of this complex disease.

In chapter V, insights gleaned from my studies on honeycomb fibrosis, a hallmark of progressive fibrotic diseases, were particularly novel. I discovered significantly differentially expressed genes in peripheral blood of patients with honeycombing, such as *PDE4DIPP, ZNF683, SFRP5, MIR6077, RPSAP72, WASIR2, GAPDHP27,* and *CNTNAP3P2,* further corroborating the multi-dimensional genetic landscape of this disease. The genetic architecture of honeycombing fibrosis was observed to be highly heterogeneous, consistent with previous studies on pulmonary fibrosis but focused on this distinctive radiological feature for the first time. The differential expression of these genes in the peripheral blood of patients with honeycombing could indicate a shift in the molecular regulation of honeycombing fibrosis, highlighting potential new therapeutic targets. My findings echo the work of Huang et al., 2023 [185], who studied the idiopathic subtype of pulmonary fibrosis and found significantly differentially expressed genes in areas of parenchymal fibrosis. They found that the central-associated genes signature score in myofibroblasts significantly correlated with previously published activated fibroblasts signature confirming that active myofibroblast features contribute to disease progression in lung fibrosis, thus further reinforcing the concept of the multi-dimensional genetic landscape in pulmonary fibrosis. As one of the hallmarks of advanced fibrotic diseases, honeycombing fibrosis has remained understudied in comparison to broader

idiopathic pulmonary fibrosis. By examining this feature, our work opens avenues for novel therapeutic targets, expanding the horizon for pulmonary fibrosis treatment.

My study also revealed significant differences in pulmonary fibrosis subtypes based on self-identified race, an observation clinically similar to that of a 2012 diagnosis-focused study by Swigris et al. [9], which highlighted the racial disparities in ILD. Pulmonary fibrosis, marked by its remarkable genetic heterogeneity, displays differential phenotypic manifestations based on race and ethnicity, with the autoimmune-associated subtype more common among non-White subjects and the fibrotic hypersensitivity pneumonitis and unclassifiable subtypes being predominant among White subjects [11]. In chapter VI, I delved deeper into the genetic underpinnings of these disease manifestations. The exploration of these genes, intricately woven into the fabric of pulmonary fibrosis, offers a profound perspective on the disease's pathogenesis. Through the lens of genomics, I identified genetic variants in *IVD*, *DPP9*, and *DSP*, genes previously associated with IPF, as well as novel variant hits in *LRRC37A4P, KANSL1-AS1, LINC02210, MIR4315-1, CD151* and *LRRC37A*. My observations mirrored the findings of other researchers such as Fingerlin et al., 2013 [186], and Stuart et al., 2019 [22], who also identified novel genetic variants associated with IPF, further underscoring the potential contribution of genomics to disease pathogenesis.

Overall, this investigation illuminated the critical interplay of race, telomeres, and genomics in pulmonary fibrosis. By drawing upon the collective insights of these diverse yet interconnected realms, we offered a novel perspective on the etiopathogenesis of this disease. My findings underscore the necessity for a nuanced understanding of these

interrelationships, which is essential for the development of personalized therapeutic strategies. While our investigation advances the field of pulmonary fibrosis, it also highlights the need for further research to validate these findings and their implications for the diagnosis, risk stratification, and treatment of this complex disease.

Future Directions

The intricate and dynamic interplay between race, telomere biology, and genomics within the context of pulmonary fibrosis revealed by this doctoral dissertation provides a roadmap for future studies. While the nuances of telomere biology, race-dependent variations in disease phenotype, fibrosis-associated genetic polymorphisms, and gene-by environment explorations have indeed charted a detailed course providing valuable insights that serve to enhance our collective understanding of pulmonary fibrosis, much remains yet to be done. Of course, the nature of scientific enquiry is such that each revelation often uncovers a myriad of new questions, inviting further exploration and study.

A critical revelation from my research has been the marked difference in pulmonary fibrosis susceptibility, presentation, and progression across racially diverse populations. This finding underscores the role of genetic diversity and complex gene-environment interactions in shaping the landscape of pulmonary fibrosis. Consequently, my future work will focus on disentangling these intricate relationships, investigating the gene-by-environment interactions, and their role in racial disparities observed in the incidence and

outcomes of pulmonary fibrosis. How much of an impact does ambient particulate matter such as PM2.5, smoking and other inhaled air pollutants have on propagating these gene-by-environment interactions [187, 188]? Are they causal, promoting factors, or an epiphenomenon. My research has shown promise in using iPSCs as a powerful tool for investigating disease mechanisms in pulmonary fibrosis. I have also generated a robust pipeline of PBMC derived human iPSCs from subjects with pulmonary fibrosis that will enable me to study more precisely the impact of gene variants on disease pathogenesis and progression in a manner that is unencumbered by these environmental factors including diet, stress, medications and other potential confounders. This platform will also enable me to study gene regulation throughout fibroblast development using recently developed genomic techniques. Moving forward, I will also integrate the use of lung organoids with iPSC technology in better defining the impact of genomics in susceptibility to pulmonary fibrosis. The Gilad lab has developed specific expertise in use of organoids as a more precise genomic platform for assessing the impact of genetic variation in human disease. This will enable me to model disease processes more accurately in vitro and further elucidate the molecular mechanisms contributing to pulmonary fibrosis. This approach could enhance our understanding of the role of identified genetic variants and telomere biology in disease pathogenesis, improving our ability to develop effective therapeutic strategies.

One particularly promising area of research is exploring the mechanisms underlying the genetic associations with pulmonary fibrosis, with a particular focus on telomere biology. The association between shortened telomeres and transplant-free

survival underscores the critical role of telomere biology in pulmonary fibrosis. To concretize this association, my future work will focus on functional studies investigating the biological processes affected by telomere shortening in lung cells. These studies could also leverage iPSCs and lung organoids [34], enabling us to examine cellular changes associated with telomere shortening in a controlled in vitro setting. The quartile decreases in transplant-free survival associated with telomere length presents a compelling area for further exploration. Specifically, my planned investigations into the causal mechanisms of this association could elucidate the processes underlying mortality rates in pulmonary fibrosis patients. These studies could employ newer genetic techniques to refine our understanding of the implicated telomere-related genetic loci, alongside comprehensive DNA methylation studies. My findings also supported the potential role of PDE4 inhibitors as therapeutic agents in fibrotic diseases [123]. Others and I will explore the validation of these compounds in clinical settings. Importantly, parallel investigations into the potential therapeutic role of other differentially expressed genes identified in my study could present novel treatment strategies for pulmonary fibrosis.

Transcriptomic profiling studies performed in our research have shed light on the differential gene expressions across diverse forms of pulmonary fibrosis, providing a glimpse into the genetic underpinnings of the disease. However, bridging the gap between transcriptomic variations and functional significance remains a challenge. As such, I plan to undertake more robust functional genomic studies, such as Chromatin Immunoprecipitation sequencing (ChIP-Seq) and Transcriptome-Wide Association

Studies (TWAS), which could help to connect these disparate pieces of the puzzle. Applying newer TWAS approaches to examine the association between gene expression variation across racially diverse populations with pulmonary fibrosis would be tremendously insightful. These studies could help to identify genetic variants that impact gene expression and contribute to disease susceptibility and progression. This approach would also be critical for investigating the impact of genetic ancestry on pulmonary fibrosis and could harness the benefits of using nominal ancestral markers to determine the influence of ancestry on disease susceptibility, course and outcomes. Given the apparent discordance between self-identified race and genetic ancestry, my future work would prioritize investigating the influence of genetic ancestry on pulmonary fibrosis. This could involve analyzing genetic markers indicative of ancestry in patient cohorts and correlating these with disease susceptibility, progression, and outcomes.

My findings suggest the existence of racial disparities in pulmonary fibrosis susceptibility, progression, and outcome. Therefore, an immediate and feasible next step would involve the initiation of large-scale, multi-center cohort studies to explore these disparities. The aim would be to uncover specific gene-environment interactions contributing to these differences, which could involve the collection of comprehensive environmental, lifestyle, and clinical data alongside extensive genomic profiling. I have already begun acquiring data on such a cohort and have currently collated a robust multinational cohort of 11,000 deeply phenotyped subjects with diverse forms of pulmonary fibrosis. This comprises on of the largest longitudinal lung function repositories in pulmonary fibrosis with over 75,000 lung function measurements obtained with

concurrent serial chest HRCT scans and biorepository genetic data acquired in over 2,000 subjects. This will facilitate the assessment of predictors of progressive pulmonary fibrosis across diverse racial groups. Building on the unique findings from the eQTL analyses and the observed higher prevalence of autoimmunity in Black patients with pulmonary fibrosis, I plan to embark on a series of studies focused on HLA-related immunologic and autoimmune mechanisms. Firstly, leveraging the recently advanced techniques in HLA genotyping [189, 190], I will extend the profiling of HLA gene variants across larger, racially diverse cohorts of patients with pulmonary fibrosis. This will help to identify HLA gene variant patterns that might be more common in particular racial groups or associated with increased disease severity. Given the role of HLA molecules in presenting antigens to T cells, I'm particularly interested in exploring whether these novel HLA gene variants could alter antigen presentation, leading to altered immune responses and potentially autoimmune mechanisms. This would involve in vitro functional studies, where iPSC-derived antigen-presenting cells with specific HLA gene variants are co-cultured with T cells to observe any changes in T cell activation. Secondly, considering the higher prevalence of autoimmunity in Black patients, I plan to investigate whether these HLA gene variants correlate with autoantibody production. This would entail serological studies to identify any autoantibodies unique to patients with these HLA gene variants and in vitro assays to examine whether these autoantibodies could contribute to lung injury. Lastly, given the emerging field of HLA peptidomics, it would be fascinating to probe whether these HLA gene variants result in an altered peptide repertoire presented to T cells [191, 192]. Such studies could involve mass spectrometry analyses to

characterize the peptide sequences presented by the HLA molecules associated with these variants. These detailed, actionable research projects not only have the potential to elucidate the functional implications of the HLA gene variants identified in our study, but they could also illuminate the role of HLA genes and autoimmunity in the racial disparities observed in pulmonary fibrosis, laying the groundwork for potential therapeutic interventions.

Lastly, future work should strive to augment the translational potential of these findings. As the ultimate goal of research is to translate laboratory discoveries into clinical interventions that improve patient outcomes, I strongly advocate for further research that emphasizes the application of these findings to clinical practice. This could involve developing prognostic models based on identified genetic variations, race, and telomere lengths, which can inform personalized treatment strategies. In conclusion, the exploration of the interplay between race, telomeres, and genomics within the context of pulmonary fibrosis has paved the way for a plethora of intriguing future research directions. As we chart the course forward, it is clear that our journey in elucidating the intricacies of pulmonary fibrosis is only beginning. The vast expanse of knowledge yet to be explored is indeed challenging but is nonetheless a testament to the complexity and beauty of human biology. With continued research efforts, we can enhance our understanding of this devastating disease, thereby paving the way for novel, more effective therapeutic interventions and improved patient outcomes.

Research Impact

The research encapsulated in this doctoral thesis marks a significant leap forward in understanding the labyrinthine genetic landscape and telomere biology within the field of pulmonary fibrosis, with an added layer of complexity introduced by racial diversity. The implications of these findings are profound, not only enriching our grasp of the molecular underpinnings of this intricate disease but also setting the stage for advancing the development of tailored therapeutic strategies and improving patient management.

The investigation of racial disparities in pulmonary fibrosis through the lens of genomic distinctions provides a transformative perspective on disease heterogeneity. By establishing genetic ancestry as a significant determinant of disease phenotype and progression, this research challenges conventional biomedical paradigms that employ self-identified race as a proxy, thus catalyzing a shift towards more genetically-informed research methodologies.

Simultaneously, this work unravels the nexus between telomere length and pulmonary fibrosis, highlighting the role of telomeres as critical biomarkers of disease progression and survival outcomes. The novel link established between telomere length, fibrotic hypersensitivity pneumonitis, and idiopathic pulmonary fibrosis could signal the presence of shared pathophysiological pathways across diverse fibrotic diseases. This crucial finding could influence future research trajectories aimed at deciphering shared molecular mechanisms and potential therapeutic targets across ILDs.

Moreover, the exploration of honeycombing fibrosis, an understudied yet critical facet of pulmonary fibrosis, has identified a set of eight genes that could serve as key genetic drivers of disease progression. This revelation could significantly impact the broader understanding of the disease's genetic landscape, potentially triggering the development of novel therapeutic strategies targeting these genes.

Conducting eQTL studies as part of this dissertation has been instrumental in unearthing novel gene variants and their potential influence on pulmonary fibrosis risk and progression. The discovery of novel variants in genes, including those in the HLA family, could herald a shift towards a more personalized approach to managing and treating this disease.

Collectively, these research findings hold transformative potential in informing the future direction of pulmonary fibrosis research, emphasizing the need for studies that are attuned to the genetic ancestry, telomere biology, and unique genetic signatures of the disease. These insights may also inform clinical practice by providing robust evidence to support treatment decisions based on genetic and telomeric characteristics.

In essence, this dissertation represents a critical milestone in the realm of pulmonary fibrosis and interstitial lung diseases. The significant impacts of this research extend beyond augmenting our understanding of the disease's molecular mechanisms, to fundamentally influencing patient care and management. This work underscores the importance of a multi-faceted, integrative, and nuanced approach to complex diseases like pulmonary fibrosis, underscoring the transformative power of meticulous scientific inquiry in enhancing patient outcomes and advancing the field.

References

1. Podolanczuk AJ, Richeldi L, Martinez FJ. The Future of Clinical Trials in Idiopathic Pulmonary Fibrosis. *JAMA* 2023; 329: 1554-1555.

2. Manichaikul A, Wang XQ, Sun L, Dupuis J, Borczuk AC, Nguyen JN, Raghu G, Hoffman EA, Onengut-Gumuscu S, Farber EA, Kaufman JD, Rabinowitz D, Stukovsky KDH, Kawut SM, Hunninghake GM, Washko GR, O'Connor GT, Rich SS, Barr RG, Lederer DJ. Genome-wide association study of subclinical interstitial lung disease in MESA. *Respir Res* 2017; 18: 97.

3. Adegunsoye A, Oldham JM, Bellam SK, Chung JH, Chung PA, Biblowitz KM, Montner S, Lee C, Hsu S, Husain AN, Vij R, Mutlu G, Noth I, Churpek MM, Strek ME. African-American race and mortality in interstitial lung disease: a multicentre propensity-matched analysis. *Eur Respir J* 2018; 51.

4. Pugashetti JV, Adegunsoye A, Wu Z, Lee CT, Srikrishnan A, Ghodrati S, Vo V, Renzoni EA, Wells AU, Garcia CK, Chua F, Newton CA, Molyneaux PL, Oldham JM. Validation of Proposed Criteria for Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2023; 207: 69-76.

5. Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Shirude S, Naghavi M, Mokdad AH, Murray CJL. Trends and Patterns of Differences in Chronic Respiratory Disease Mortality Among US Counties, 1980-2014. *JAMA* 2017; 318: 1136-1149.

6. Vaz Fragoso CA, McAvay G, Gill TM, Concato J, Quanjer PH, Van Ness PH. Ethnic differences in respiratory impairment. *Thorax* 2014; 69: 55-62.

7. Cooke CR, Erickson SE, Eisner MD, Martin GS. Trends in the incidence of noncardiogenic acute respiratory failure: the role of race. *Crit Care Med* 2012; 40: 1532-1538.

8. Heron M, Hoyert DL, Murphy SL, Xu J, Kochanek KD, Tejada-Vera B. Deaths: final data for 2006. *Natl Vital Stat Rep* 2009; 57: 1-134.

9. Swigris JJ, Olson AL, Huie TJ, Fernandez-Perez ER, Solomon J, Sprunger D, Brown KK. Ethnic and racial differences in the presence of idiopathic pulmonary fibrosis at death. *Respir Med* 2012; 106: 588-593.

10. Adegunsoye A, Vela M, Saunders M. Racial Disparities in Pulmonary Fibrosis and the Impact on the Black Population. *Arch Bronconeumol* 2022; 58: 590-592.

11. Adegunsoye A, Freiheit E, White EN, Kaul B, Newton CA, Oldham JM, Lee CT, Chung J, Garcia N, Ghodrati S, Vij R, Jablonski R, Flaherty KR, Wolters PJ, Garcia CK,

Strek ME. Evaluation of Pulmonary Fibrosis Outcomes by Race and Ethnicity in US Adults. *JAMA Netw Open* 2023; 6: e232427.

12. Salisbury ML, Markin CR, Wu P, Cogan JD, Mitchell DB, Liu Q, Loyd JE, Lancaster LH, Kropski JA, Blackwell TS. Peripheral Blood Telomere Attrition in Persons at Risk for Familial Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2023; 207: 208-211.

13. Kropski JA, Pritchett JM, Zoz DF, Crossno PF, Markin C, Garnett ET, Degryse AL, Mitchell DB, Polosukhin VV, Rickman OB, Choi L, Cheng DS, McConaha ME, Jones BR, Gleaves LA, McMahon FB, Worrell JA, Solus JF, Ware LB, Lee JW, Massion PP, Zaynagetdinov R, White ES, Kurtis JD, Johnson JE, Groshong SD, Lancaster LH, Young LR, Steele MP, Phillips Iii JA, Cogan JD, Loyd JE, Lawson WE, Blackwell TS. Extensive phenotyping of individuals at risk for familial interstitial pneumonia reveals clues to the pathogenesis of interstitial lung disease. *Am J Respir Crit Care Med* 2015; 191: 417-426.

14. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, Kraven LM, Obeidat M, Li X, Ng M, Braybrooke R, Molina-Molina M, Hobbs BD, Putman RK, Sakornsakolpat P, Booth HL, Fahy WA, Hart SP, Hill MR, Hirani N, Hubbard RB, McAnulty RJ, Millar AB, Navaratnam V, Oballa E, Parfrey H, Saini G, Whyte MKB, Zhang Y, Kaminski N, Adegunsoye A, Strek ME, Neighbors M, Sheng XR, Gudmundsson G, Gudnason V, Hatabu H, Lederer DJ, Manichaikul A, Newell JD, Jr., O'Connor GT, Ortega VE, Xu H, Fingerlin TE, Bosse Y, Hao K, Joubert P, Nickle DC, Sin DD, Timens W, Furniss D, Morris AP, Zondervan KT, Hall IP, Sayers I, Tobin MD, Maher TM, Cho MH, Hunninghake GM, Schwartz DA, Yaspan BL, Molyneaux PL, Flores C, Noth I, Jenkins RG, Wain LV. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020; 201: 564-574.

15. Kam MLW, Nguyen TTT, Ngeow JYY. Telomere biology disorders. *NPJ Genom Med* 2021; 6: 36.

16. Hoa S, Hudson M, Troyanov Y, Proudman S, Walker J, Stevens W, Nikpour M, Assassi S, Mayes MD, Wang M, Baron M, Fritzler MJ, Canadian Scleroderma Research G, Australian Scleroderma Interest G, Genetics versus Environment in Scleroderma Outcome S. Single-specificity anti-Ku antibodies in an international cohort of 2140 systemic sclerosis subjects: clinical associations. *Medicine (Baltimore)* 2016; 95: e4713.

17. Stuart BD, Lee JS, Kozlitina J, Noth I, Devine MS, Glazer CS, Torres F, Kaza V, Girod CE, Jones KD, Elicker BM, Ma SF, Vij R, Collard HR, Wolters PJ, Garcia CK. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. *Lancet Respir Med* 2014; 2: 557-565.

18. Ley B, Newton CA, Arnould I, Elicker BM, Henry TS, Vittinghoff E, Golden JA, Jones KD, Batra K, Torrealba J, Garcia CK, Wolters PJ. The MUC5B promoter polymorphism and telomere length in patients with chronic hypersensitivity pneumonitis: an observational cohort-control study. *Lancet Respir Med* 2017; 5: 639-647.

19. Adegunsoye A, Vij R, Noth I. Integrating Genomics Into Management of Fibrotic Interstitial Lung Disease. *Chest* 2019; 155: 1026-1040.

20. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, Lawson WE, Xie M, Vulto I, Phillips JA, 3rd, Lansdorp PM, Greider CW, Loyd JE. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 2007; 356: 1317-1326.

21. Diaz de Leon A, Cronkhite JT, Katzenstein AL, Godwin JD, Raghu G, Glazer CS, Rosenblatt RL, Girod CE, Garrity ER, Xing C, Garcia CK. Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. *PLoS One* 2010; 5: e10680.

22. Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, Choi M, Dharwadkar P, Torres F, Girod CE, Weissler J, Fitzgerald J, Kershaw C, Klesney-Tait J, Mageto Y, Shay JW, Ji W, Bilguvar K, Mane S, Lifton RP, Garcia CK. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet* 2015; 47: 512-517.

23. Cogan JD, Kropski JA, Zhao M, Mitchell DB, Rives L, Markin C, Garnett ET, Montgomery KH, Mason WR, McKean DF, Powers J, Murphy E, Olson LM, Choi L, Cheng DS, Blue EM, Young LR, Lancaster LH, Steele MP, Brown KK, Schwarz MI, Fingerlin TE, Schwartz DA, Lawson WE, Loyd JE, Zhao Z, Phillips JA, 3rd, Blackwell TS. Rare variants in RTEL1 are associated with familial interstitial pneumonia. *Am J Respir Crit Care Med* 2015; 191: 646-655.

24. Juge PA, Borie R, Kannengiesser C, Gazal S, Revy P, Wemeau-Stervinou L, Debray MP, Ottaviani S, Marchand-Adam S, Nathan N, Thabut G, Richez C, Nunes H, Callebaut I, Justet A, Leulliot N, Bonnefond A, Salgado D, Richette P, Desvignes JP, Liote H, Froguel P, Allanore Y, Sand O, Dromer C, Flipo RM, Clement A, Beroud C, Sibilia J, Coustet B, Cottin V, Boissier MC, Wallaert B, Schaeverbeke T, Dastot le Moal F, Frazier A, Menard C, Soubrier M, Saidenberg N, Valeyre D, Amselem S, consortium F, Boileau C, Crestani B, Dieude P. Shared genetic predisposition in rheumatoid arthritis-interstitial lung disease and familial pulmonary fibrosis. *Eur Respir J* 2017; 49.

25. Kropski JA, Young LR, Cogan JD, Mitchell DB, Lancaster LH, Worrell JA, Markin C, Liu N, Mason WR, Fingerlin TE, Schwartz DA, Lawson WE, Blackwell TS, Phillips

JA, 3rd, Loyd JE. Genetic Evaluation and Testing of Patients and Families with Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2017; 195: 1423-1428.

26. Newton CA, Oldham JM, Ley B, Anand V, Adegunsoye A, Liu G, Batra K, Torrealba J, Kozlitina J, Glazer C, Strek ME, Wolters PJ, Noth I, Garcia CK. Telomere length and genetic variant associations with interstitial lung disease progression and survival. *Eur Respir J* 2019; 53.

27. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N Engl J Med* 2018; 379: 797-798.

28. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, Flaherty KR, Wells A, Martinez FJ, Azuma A, Bice TJ, Bouros D, Brown KK, Collard HR, Duggal A, Galvin L, Inoue Y, Jenkins RG, Johkoh T, Kazerooni EA, Kitaichi M, Knight SL, Mansour G, Nicholson AG, Pipavath SNJ, Buendia-Roldan I, Selman M, Travis WD, Walsh S, Wilson KC, American Thoracic Society ERSJRS, Latin American Thoracic S. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2018; 198: e44-e68.

29. Morell F, Villar A, Montero MA, Munoz X, Colby TV, Pipvath S, Cruz MJ, Raghu G. Chronic hypersensitivity pneumonitis in patients diagnosed with idiopathic pulmonary fibrosis: a prospective case-cohort study. *Lancet Respir Med* 2013; 1: 685-694.

30. Adegunsoye A, Oldham JM, Bellam SK, Montner S, Churpek MM, Noth I, Vij R, Strek ME, Chung JH. Computed Tomography Honeycombing Identifies a Progressive Fibrotic Phenotype with Increased Mortality across Diverse Interstitial Lung Diseases. *Ann Am Thorac Soc* 2019; 16: 580-588.

31. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019; 47: D1005-D1012.

32. Alysandratos KD, Russo SJ, Petcherski A, Taddeo EP, Acin-Perez R, Villacorta-Martin C, Jean JC, Mulugeta S, Rodriguez LR, Blum BC, Hekman RM, Hix OT, Minakin K, Vedaie M, Kook S, Tilston-Lunel AM, Varelas X, Wambach JA, Cole FS, Hamvas A, Young LR, Liesa M, Emili A, Guttentag SH, Shirihai OS, Beers MF, Kotton DN. Patient-specific iPSCs carrying an SFTPC mutation reveal the intrinsic alveolar epithelial dysfunction at the inception of interstitial lung disease. *Cell Rep* 2021; 36: 109636.

33. Castro-Vinuelas R, Sanjurjo-Rodriguez C, Pineiro-Ramil M, Hermida-Gomez T, Rodriguez-Fernandez S, Oreiro N, de Toro J, Fuentes I, Blanco FJ, Diaz-Prado S. Generation and characterization of human induced pluripotent stem cells (iPSCs) from hand osteoarthritis patient-derived fibroblasts. *Sci Rep* 2020; 10: 4272.

34. Adegunsoye A. GN, Gilad Y. Induced Pluripotent Stem Cells in Disease Biology and the Evidence for their in vitro Utility. *Annual Review Genetics* 2023; 2023: 57.

35. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020; 369: 1318-1330.

36. Umans BD, Battle A, Gilad Y. Where Are the Disease-Associated eQTLs? *Trends Genet* 2021; 37: 109-124.

37. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; 131: 861-872.

38. Strikoudis A, Cieslak A, Loffredo L, Chen YW, Patel N, Saqi A, Lederer DJ, Snoeck HW. Modeling of Fibrotic Lung Disease Using 3D Organoids Derived from Human Pluripotent Stem Cells. *Cell Rep* 2019; 27: 3709-3723 e3705.

39. Simsek S, Zhou T, Robinson CL, Tsai SY, Crespo M, Amin S, Lin X, Hon J, Evans T, Chen S. Modeling Cystic Fibrosis Using Pluripotent Stem Cell-Derived Human Pancreatic Ductal Epithelial Cells. *Stem Cells Transl Med* 2016; 5: 572-579.

40. Steele-Stallard HB, Pinton L, Sarcar S, Ozdemir T, Maffioletti SM, Zammit PS, Tedesco FS. Modeling Skeletal Muscle Laminopathies Using Human Induced Pluripotent Stem Cells Carrying Pathogenic LMNA Mutations. *Front Physiol* 2018; 9: 1332.

41. Lee S, Park C, Han JW, Kim JY, Cho K, Kim EJ, Kim S, Lee SJ, Oh SY, Tanaka Y, Park IH, An HJ, Shin CM, Sharma S, Yoon YS. Direct Reprogramming of Human Dermal Fibroblasts Into Endothelial Cells Using ER71/ETV2. *Circ Res* 2017; 120: 848-861.

42. Ruaro B, Salton F, Braga L, Wade B, Confalonieri P, Volpe MC, Baratella E, Maiocchi S, Confalonieri M. The History and Mystery of Alveolar Epithelial Type II Cells: Focus on Their Physiologic and Pathologic Role in Lung. *Int J Mol Sci* 2021; 22.

43. Pardo A, Selman M. Lung Fibroblasts, Aging, and Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc* 2016; 13 Suppl 5: S417-S421.

44. Adegunsoye A, Newton CA, Oldham JM, Ley B, Lee CT, Linderholm AL, Chung JH, Garcia N, Zhang D, Vij R, Guzy R, Jablonski R, Bag R, Voogt RS, Ma SF, Sperling

AI, Raghu G, Martinez FJ, Strek ME, Wolters PJ, Garcia CK, Pierce BL, Noth I. Telomere length associates with chronological age and mortality across racially diverse pulmonary fibrosis cohorts. *Nat Commun* 2023; 14: 1489.

45. Armanios M. Syndromes of telomere shortening. *Annu Rev Genomics Hum Genet* 2009; 10: 45-61.

46. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; 51: 584-591.

47. Allen RJ, Oldham JM, Jenkins DA, Leavy OC, Guillen-Guio B, Melbourne CA, Ma SF, Jou J, Kim JS, Clean UPIPFIotPTC, Fahy WA, Oballa E, Hubbard RB, Navaratnam V, Braybrooke R, Saini G, Roach KM, Tobin MD, Hirani N, Whyte MKB, Kaminski N, Zhang Y, Martinez FJ, Linderholm AL, Adegunsoye A, Strek ME, Maher TM, Molyneaux PL, Flores C, Noth I, Gisli Jenkins R, Wain LV. Longitudinal lung function and gas transfer in individuals with idiopathic pulmonary fibrosis: a genome-wide association study. *Lancet Respir Med* 2023; 11: 65-73.

48. Allen RJ, Stockwell A, Oldham JM, Guillen-Guio B, Schwartz DA, Maher TM, Flores C, Noth I, Yaspan BL, Jenkins RG, Wain LV, International IPFGC. Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. *Thorax* 2022; 77: 829-833.

49. Juge PA, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, Kannengiesser C, Ottaviani S, Oka S, Tohma S, Tsuchiya N, Rojas-Serrano J, Gonzalez-Perez MI, Mejia M, Buendia-Roldan I, Falfan-Valencia R, Ambrocio-Ortiz E, Manali E, Papiris SA, Karageorgas T, Boumpas D, Antoniou K, van Moorsel CHM, van der Vis J, de Man YA, Grutters JC, Wang Y, Borie R, Wemeau-Stervinou L, Wallaert B, Flipo RM, Nunes H, Valeyre D, Saidenberg-Kermanac'h N, Boissier MC, Marchand-Adam S, Frazier A, Richette P, Allanore Y, Sibilia J, Dromer C, Richez C, Schaeverbeke T, Liote H, Thabut G, Nathan N, Amselem S, Soubrier M, Cottin V, Clement A, Deane K, Walts AD, Fingerlin T, Fischer A, Ryu JH, Matteson EL, Niewold TB, Assayag D, Gross A, Wolters P, Schwarz MI, Holers M, Solomon JJ, Doyle T, Rosas IO, Blauwendraat C, Nalls MA, Debray MP, Boileau C, Crestani B, Schwartz DA, Dieude P. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. *N Engl J Med* 2018; 379: 2209-2219.

50. Helling BA, Gerber AN, Kadiyala V, Sasse SK, Pedersen BS, Sparks L, Nakano Y, Okamoto T, Evans CM, Yang IV, Schwartz DA. Regulation of MUC5B Expression in Idiopathic Pulmonary Fibrosis. *Am J Respir Cell Mol Biol* 2017; 57: 91-99.

51. Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, Lara AR, Schwarz MI, Evans CM, Schwartz DA. MUC5B Promoter Variant rs35705950

Affects MUC5B Expression in the Distal Airways in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2016; 193: 464-466.

52. Yang IV, Fingerlin TE, Evans CM, Schwarz MI, Schwartz DA. MUC5B and Idiopathic Pulmonary Fibrosis. *Ann Am Thorac Soc* 2015; 12 Suppl 2: S193-199.

53. Oldham JM, Ma SF, Martinez FJ, Anstrom KJ, Raghu G, Schwartz DA, Valenzi E, Witt L, Lee C, Vij R, Huang Y, Strek ME, Noth I, Investigators IP. TOLLIP, MUC5B, and the Response to N-Acetylcysteine among Individuals with Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2015; 192: 1475-1482.

54. Borie R, Crestani B, Dieude P, Nunes H, Allanore Y, Kannengiesser C, Airo P, Matucci-Cerinic M, Wallaert B, Israel-Biet D, Cadranel J, Cottin V, Gazal S, Peljto AL, Varga J, Schwartz DA, Valeyre D, Grandchamp B. The MUC5B variant is associated with idiopathic pulmonary fibrosis but not with systemic sclerosis interstitial lung disease in the European Caucasian population. *PLoS One* 2013; 8: e70621.

55. Kim JS, Manichaikul AW, Hoffman EA, Balte P, Anderson MR, Bernstein EJ, Madahar P, Oelsner EC, Kawut SM, Wysoczanski A, Laine AF, Adegunsoye A, Ma JZ, Taub MA, Mathias RA, Rich SS, Rotter JI, Noth I, Garcia CK, Barr RG, Podolanczuk AJ. MUC5B, telomere length and longitudinal quantitative interstitial lung changes: the MESA Lung Study. *Thorax* 2023; 78: 566-573.

56. Arias E, Heron M, Xu J. United States Life Tables, 2014. *Natl Vital Stat Rep* 2017; 66: 1-64.

57. Newton CA, Zhang D, Oldham JM, Kozlitina J, Ma SF, Martinez FJ, Raghu G, Noth I, Garcia CK. Telomere Length and Use of Immunosuppressive Medications in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 200: 336-347.

58. Tang J, Sheng C, Wu YY, Yan LL, Wu C. Association of Joint Genetic and Social Environmental Risks With Incident Myocardial Infarction: Results From the Health and Retirement Study. *J Am Heart Assoc* 2023; 12: e028200.

59. Cronkhite JT, Xing C, Raghu G, Chin KM, Torres F, Rosenblatt RL, Garcia CK. Telomere shortening in familial and sporadic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008; 178: 729-737.

60. Adegunsoye A, Morisset J, Newton CA, Oldham JM, Vittinghoff E, Linderholm AL, Strek ME, Noth I, Garcia CK, Wolters PJ, Ley B. Leukocyte telomere length and mycophenolate therapy in chronic hypersensitivity pneumonitis. *Eur Respir J* 2021; 57.

61. Duckworth A, Gibbons MA, Allen RJ, Almond H, Beaumont RN, Wood AR, Lunnon K, Lindsay MA, Wain LV, Tyrrell J, Scotton CJ. Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a mendelian randomisation study. *Lancet Respir Med* 2021; 9: 285-294.

62. Borie R, Tabeze L, Thabut G, Nunes H, Cottin V, Marchand-Adam S, Prevot G, Tazi A, Cadranel J, Mal H, Wemeau-Stervinou L, Bergeron Lafaurie A, Israel-Biet D, Picard C, Reynaud Gaubert M, Jouneau S, Naccache JM, Mankikian J, Menard C, Cordier JF, Valeyre D, Reocreux M, Grandchamp B, Revy P, Kannengiesser C, Crestani B. Prevalence and characteristics of TERT and TERC mutations in suspected genetic pulmonary fibrosis. *Eur Respir J* 2016; 48: 1721-1731.

63. Newton CA, Batra K, Torrealba J, Kozlitina J, Glazer CS, Aravena C, Meyer K, Raghu G, Collard HR, Garcia CK. Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. *Eur Respir J* 2016; 48: 1710-1720.

64. Ramos PS, Shedlock AM, Langefeld CD. Genetics of autoimmune diseases: insights from population genetics. *J Hum Genet* 2015; 60: 657-664.

65. McNearney TA, Reveille JD, Fischbach M, Friedman AW, Lisse JR, Goel N, Tan FK, Zhou X, Ahn C, Feghali-Bostwick CA, Fritzler M, Arnett FC, Mayes MD. Pulmonary involvement in systemic sclerosis: associations with genetic, serologic, sociodemographic, and behavioral factors. *Arthritis Rheum* 2007; 57: 318-326.

66. Mittoo S, Gelber AC, Christopher-Stine L, Horton MR, Lechtzin N, Danoff SK. Ascertainment of collagen vascular disease in patients presenting with interstitial lung disease. *Respir Med* 2009; 103: 1152-1158.

67. Ash M, Boyce JK. Racial disparities in pollution exposure and employment at US industrial facilities. *Proc Natl Acad Sci U S A* 2018; 115: 10636-10641.

68. Tessum CW, Apte JS, Goodkind AL, Muller NZ, Mullins KA, Paolella DA, Polasky S, Springer NP, Thakrar SK, Marshall JD, Hill JD. Inequity in consumption of goods and services adds to racial-ethnic disparities in air pollution exposure. *Proc Natl Acad Sci U S A* 2019; 116: 6001-6006.

69. Adegunsoye A. Interstitial lung disease: a decade of progress and hope. *Lancet Respir Med* 2023; 11: 510-512.

70. Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, Raghu G, Sauter W, Girard M, Alves M, Clerisme-Beaty E, Stowasser S, Tetzlaff K, Kuwana M, Maher TM, Investigators ST. Nintedanib for Systemic Sclerosis-Associated Interstitial Lung Disease. *N Engl J Med* 2019; 380: 2518-2528.

71. Adegunsoye A, Oldham JM, Fernandez Perez ER, Hamblin M, Patel N, Tener M, Bhanot D, Robinson L, Bullick S, Chen L, Hsu S, Churpek M, Hedeker D, Montner S, Chung JH, Husain AN, Noth I, Strek ME, Vij R. Outcomes of immunosuppressive therapy in chronic hypersensitivity pneumonitis. *ERJ Open Res* 2017; 3.

72. Faust HE, Golden JA, Rajalingam R, Wang AS, Green G, Hays SR, Kukreja J, Singer JP, Wolters PJ, Greenland JR. Short lung transplant donor telomere length is associated with decreased CLAD-free survival. *Thorax* 2017; 72: 1052-1054.

73. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015; 34: 3661-3679.

74. Ley B, Torgerson DG, Oldham JM, Adegunsoye A, Liu S, Li J, Elicker BM, Henry TS, Golden JA, Jones KD, Dressen A, Yaspan BL, Arron JR, Noth I, Hoffmann TJ, Wolters PJ. Rare Protein-Altering Telomere-related Gene Variants in Patients with Chronic Hypersensitivity Pneumonitis. *Am J Respir Crit Care Med* 2019; 200: 1154-1163.

75. Bertuch AA. The molecular genetics of the telomere biology disorders. *RNA Biol* 2016; 13: 696-706.

76. Savage SA, Bertuch AA. The genetics and clinical manifestations of telomere biology disorders. *Genet Med* 2010; 12: 753-764.

77. Gable DL, Gaysinskaya V, Atik CC, Talbot CC, Jr., Kang B, Stanley SE, Pugh EW, Amat-Codina N, Schenk KM, Arcasoy MO, Brayton C, Florea L, Armanios M. ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. *Genes Dev* 2019; 33: 1381-1396.

78. Shay JW, Wright WE. Telomeres and telomerase: three decades of progress. *Nat Rev Genet* 2019; 20: 299-309.

79. Stanley SE, Armanios M. The short and long telomere syndromes: paired paradigms for molecular medicine. *Curr Opin Genet Dev* 2015; 33: 1-9.

80. Tummala H, Walne A, Collopy L, Cardoso S, de la Fuente J, Lawson S, Powell J, Cooper N, Foster A, Mohammed S, Plagnol V, Vulliamy T, Dokal I. Poly(A)-specific ribonuclease deficiency impacts telomere biology and causes dyskeratosis congenita. *J Clin Invest* 2015; 125: 2151-2160.

81. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, Rosenblatt RL, Shay JW, Garcia CK. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A* 2007; 104: 7552-7557.

82. Stanley SE, Chen JJ, Podlevsky JD, Alder JK, Hansel NN, Mathias RA, Qi X, Rafaels NM, Wise RA, Silverman EK, Barnes KC, Armanios M. Telomerase mutations in smokers with severe emphysema. *J Clin Invest* 2015; 125: 563-570.

83. Calado RT, Brudno J, Mehta P, Kovacs JJ, Wu C, Zago MA, Chanock SJ, Boyer TD, Young NS. Constitutional telomerase mutations are genetic risk factors for cirrhosis. *Hepatology* 2011; 53: 1600-1607.

84. Hartmann D, Srivastava U, Thaler M, Kleinhans KN, N'Kontchou G, Scheffold A, Bauer K, Kratzer RF, Kloos N, Katz SF, Song Z, Begus-Nahrmann Y, Kleger A, von Figura G, Strnad P, Lechel A, Gunes C, Potthoff A, Deterding K, Wedemeyer H, Ju Z, Song G, Xiao F, Gillen S, Schrezenmeier H, Mertens T, Ziol M, Friess H, Jarek M, Manns MP, Beaugrand M, Rudolph KL. Telomerase gene mutations are associated with cirrhosis formation. *Hepatology* 2011; 53: 1608-1617.

85. Keel SB, Scott A, Sanchez-Bonilla M, Ho PA, Gulsuner S, Pritchard CC, Abkowitz JL, King MC, Walsh T, Shimamura A. Genetic features of myelodysplastic syndrome and aplastic anemia in pediatric and young adult patients. *Haematologica* 2016; 101: 1343-1350.

86. Yamaguchi H, Calado RT, Ly H, Kajigaya S, Baerlocher GM, Chanock SJ, Lansdorp PM, Young NS. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N Engl J Med* 2005; 352: 1413-1424.

87. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica* 2018; 103: 30-39.

88. Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, Lansdorp PM, Hogge D, Chanock SJ, Estey EH, Falcao RP, Young NS. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. *Proc Natl Acad Sci U S A* 2009; 106: 1187-1192.

89. Alder JK, Hanumanthu VS, Strong MA, DeZern AE, Stanley SE, Takemoto CM, Danilova L, Applegate CD, Bolton SG, Mohr DW, Brodsky RA, Casella JF, Greider CW, Jackson JB, Armanios M. Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* 2018; 115: E2358-E2365.

90. Stanley SE, Gable DL, Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, Khalil SE, DeZern AE, Rojas-Duran MF, Applegate CD, Alder JK, Parry EM, Gilbert WV, Armanios M. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. *Sci Transl Med* 2016; 8: 351ra107.

91. Yamaguchi H, Baerlocher GM, Lansdorp PM, Chanock SJ, Nunez O, Sloand E, Young NS. Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome. *Blood* 2003; 102: 916-918.

92. Feurstein S, Adegunsoye A, Mojsilovic D, Vij R, West DePersia AH, Rajagopal PS, Osman A, Collins RH, Kim RH, Gore SD, Greenberg P, Godley LA, Li Z, Del Gaudio D, Subramanian HP, Das S, Walsh T, Gulsuner S, Segal JP, Husain AN, Gurbuxani S, King MC, Strek ME, Churpek JE. Telomere biology disorder prevalence and phenotypes in adults with familial hematologic and/or pulmonary presentations. *Blood Adv* 2020; 4: 4873-4886.

93. Borie R, Bouvry D, Cottin V, Gauvain C, Cazes A, Debray MP, Cadranel J, Dieude P, Degot T, Dominique S, Gamez AS, Jaillet M, Juge PA, Londono-Vallejo A, Mailleux A, Mal H, Boileau C, Menard C, Nunes H, Prevot G, Quetant S, Revy P, Traclet J, Wemeau-Stervinou L, Wislez M, Kannengiesser C, Crestani B. Regulator of telomere length 1 (RTEL1) mutations are associated with heterogeneous pulmonary and extra-pulmonary phenotypes. *Eur Respir J* 2019; 53.

94. George G, Rosas IO, Cui Y, McKane C, Hunninghake GM, Camp PC, Raby BA, Goldberg HJ, El-Chemaly S. Short telomeres, telomeropathy, and subclinical extrapulmonary organ damage in patients with interstitial lung disease. *Chest* 2015; 147: 1549-1557.

95. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, Nishino M, Araki T, Zazueta OE, Kurugol S, Ross JC, San Jose Estepar R, Murphy E, Steele MP, Loyd JE, Schwarz MI, Fingerlin TE, Rosas IO, Washko GR, O'Connor GT, Schwartz DA. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013; 368: 2192-2200.

96. Chung JH, Peljto AL, Chawla A, Talbert JL, McKean DF, Rho BH, Fingerlin TE, Schwarz MI, Schwartz DA, Lynch DA. CT Imaging Phenotypes of Pulmonary Fibrosis in the MUC5B Promoter Site Polymorphism. *Chest* 2016; 149: 1215-1222.

97. Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, Lee JS, Ji W, Schwarz MI, Yang IV, Schwartz DA, Fingerlin TE. The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries. *Chest* 2015; 147: 460-464.

98. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, Silveira LJ, Lindell KO, Steele MP, Loyd JE, Gibson KF, Seibold MA, Brown KK, Talbert JL, Markin C, Kossen K, Seiwert SD, Murphy E, Noth I, Schwarz MI, Kaminski N, Schwartz DA. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013; 309: 2232-2239.

99. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, Fingerlin TE, Zhang W, Gudmundsson G, Groshong SD, Evans CM, Garantziotis S, Adler KB, Dickey BF, du Bois RM, Yang IV, Herron A, Kervitsky D, Talbert JL, Markin C, Park J, Crews AL, Slifer SH, Auerbach S, Roy MG, Lin J, Hennessy CE, Schwarz MI, Schwartz DA. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011; 364: 1503-1512.

100. Herazo-Maya JD, Sun J, Molyneaux PL, Li Q, Villalba JA, Tzouvelekis A, Lynn H, Juan-Guardela BM, Risquez C, Osorio JC, Yan X, Michel G, Aurelien N, Lindell KO, Klesen MJ, Moffatt MF, Cookson WO, Zhang Y, Garcia JGN, Noth I, Prasse A, Bar-Joseph Z, Gibson KF, Zhao H, Herzog EL, Rosas IO, Maher TM, Kaminski N. Validation of a 52-gene risk profile for outcome prediction in patients with idiopathic pulmonary fibrosis: an international, multicentre, cohort study. *Lancet Respir Med* 2017; 5: 857-868.

101. Kim SK, Jung SM, Park KS, Kim KJ. Integrative analysis of lung molecular signatures reveals key drivers of idiopathic pulmonary fibrosis. *BMC Pulm Med* 2021; 21: 404.

102. Konigsberg IR, Borie R, Walts AD, Cardwell J, Rojas M, Metzger F, Hauck SM, Fingerlin TE, Yang IV, Schwartz DA. Molecular Signatures of Idiopathic Pulmonary Fibrosis. *Am J Respir Cell Mol Biol* 2021; 65: 430-441.

103. Chung JH, Oldham JM, Montner SM, Vij R, Adegunsoye A, Husain AN, Noth I, Lynch DA, Strek ME. CT-Pathologic Correlation of Major Types of Pulmonary Fibrosis: Insights for Revisions to Current Guidelines. *AJR Am J Roentgenol* 2018; 210: 1034-1041.

104. Chung JH, Cox CW, Montner SM, Adegunsoye A, Oldham JM, Husain AN, Vij R, Noth I, Lynch DA, Strek ME. CT Features of the Usual Interstitial Pneumonia Pattern: Differentiating Connective Tissue Disease-Associated Interstitial Lung Disease From Idiopathic Pulmonary Fibrosis. *AJR Am J Roentgenol* 2018; 210: 307-313.

105. Chung JH, Montner SM, Adegunsoye A, Lee C, Oldham JM, Husain AN, MacMahon H, Noth I, Vij R, Strek ME. CT Findings, Radiologic-Pathologic Correlation, and Imaging Predictors of Survival for Patients With Interstitial Pneumonia With Autoimmune Features. *AJR Am J Roentgenol* 2017; 208: 1229-1236.

106. Adegunsoye A, Oldham JM, Demchuk C, Montner S, Vij R, Strek ME. Predictors of survival in coexistent hypersensitivity pneumonitis with autoimmune features. *Respir Med* 2016; 114: 53-60.

107. Zong D, Liu X, Li J, Ouyang R, Chen P. The role of cigarette smoke-induced epigenetic alterations in inflammation. *Epigenetics Chromatin* 2019; 12: 65.

108. Yang IV, Pedersen BS, Rabinovich E, Hennessy CE, Davidson EJ, Murphy E, Guardela BJ, Tedrow JR, Zhang Y, Singh MK, Correll M, Schwarz MI, Geraci M, Sciurba FC, Quackenbush J, Spira A, Kaminski N, Schwartz DA. Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 1263-1272.

109. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, Guan W, Xu T, Elks CE, Aslibekyan S, Moreno-Macias H, Smith JA, Brody JA, Dhingra R, Yousefi P, Pankow JS, Kunze S, Shah SH, McRae AF, Lohman K, Sha J, Absher DM, Ferrucci L, Zhao W, Demerath EW, Bressler J, Grove ML, Huan T, Liu C, Mendelson MM, Yao C, Kiel DP, Peters A, Wang-Sattler R, Visscher PM, Wray NR, Starr JM, Ding J, Rodriguez CJ, Wareham NJ, Irvin MR, Zhi D, Barrdahl M, Vineis P, Ambatipudi S, Uitterlinden AG, Hofman A, Schwartz J, Colicino E, Hou L, Vokonas PS, Hernandez DG, Singleton AB, Bandinelli S, Turner ST, Ware EB, Smith AK, Klengel T, Binder EB, Psaty BM, Taylor KD, Gharib SA, Swenson BR, Liang L, DeMeo DL, O'Connor GT, Herceg Z, Ressler KJ, Conneely KN, Sotoodehnia N, Kardia SL, Melzer D, Baccarelli AA, van Meurs JB, Romieu I, Arnett DK, Ong KK, Liu Y, Waldenberger M, Deary IJ, Fornage M, Levy D, London SJ. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* 2016; 9: 436-447.

110. Zhang K, Xu L, Cong YS. Telomere Dysfunction in Idiopathic Pulmonary Fibrosis. *Front Med (Lausanne)* 2021; 8: 739810.

111. Kropski JA, Blackwell TS. Progress in Understanding and Treating Idiopathic Pulmonary Fibrosis. *Annu Rev Med* 2019; 70: 211-224.

112. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J* 2015; 45: 1717-1727.

113. Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q, Wegner DJ, Cole FS, Hamvas A. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. *BMJ Open Respir Res* 2014; 1: e000057.

114. Fernandez BA, Fox G, Bhatia R, Sala E, Noble B, Denic N, Fernandez D, Duguid N, Dohey A, Kamel F, Edwards L, Mahoney K, Stuckless S, Parfrey PS, Woods MO. A Newfoundland cohort of familial and sporadic idiopathic pulmonary fibrosis patients: clinical and genetic features. *Respir Res* 2012; 13: 64.

115. Parimon T, Yao C, Stripp BR, Noble PW, Chen P. Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis. *Int J Mol Sci* 2020; 21.

116. Yao C, Guan X, Carraro G, Parimon T, Liu X, Huang G, Mulay A, Soukiasian HJ, David G, Weigt SS, Belperio JA, Chen P, Jiang D, Noble PW, Stripp BR. Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2021; 203: 707-717.

117. Johannson KA, Chaudhuri N, Adegunsoye A, Wolters PJ. Treatment of fibrotic interstitial lung disease: current approaches and future directions. *Lancet* 2021; 398: 1450-1460.

118. Tong S, Ji Q, Du Y, Zhu X, Zhu C, Zhou Y. Sfrp5/Wnt Pathway: A Protective Regulatory System in Atherosclerotic Cardiovascular Disease. *J Interferon Cytokine Res* 2019; 39: 472-482.

119. Liu J, Xiao Q, Xiao J, Niu C, Li Y, Zhang X, Zhou Z, Shu G, Yin G. Wnt/beta-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduct Target Ther* 2022; 7: 3.

120. Li X, Chen M, Wan Y, Zhong L, Han X, Chen X, Xiao F, Liu J, Zhang Y, Zhu D, Xiang J, Liu J, Huang H, Hou J. Single-cell transcriptome profiling reveals the key role of ZNF683 in natural killer cell exhaustion in multiple myeloma. *Clin Transl Med* 2022; 12: e1065.

121. Kitakaze M, Uemura M, Hara T, Chijimatsu R, Motooka D, Hirai T, Konno M, Okuzaki D, Sekido Y, Hata T, Ogino T, Takahashi H, Miyoshi N, Ofusa K, Mizushima T, Eguchi H, Doki Y, Ishii H. Cancer-specific tissue-resident memory T-cells express ZNF683 in colorectal cancer. *Br J Cancer* 2023; 128: 1828-1837.

122. Greenough TC, Straubhaar JR, Kamga L, Weiss ER, Brody RM, McManus MM, Lambrecht LK, Somasundaran M, Luzuriaga KF. A Gene Expression Signature That Correlates with CD8+ T Cell Expansion in Acute EBV Infection. *J Immunol* 2015; 195: 4185-4197.

123. Richeldi L, Azuma A, Cottin V, Hesslinger C, Stowasser S, Valenzuela C, Wijsenbeek MS, Zoz DF, Voss F, Maher TM, Trial I. Trial of a Preferential Phosphodiesterase 4B Inhibitor for Idiopathic Pulmonary Fibrosis. *N Engl J Med* 2022; 386: 2178-2187.

124. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform* 2011; 12: 449-462.

125. Soni DK, Biswas R. Role of Non-Coding RNAs in Post-Transcriptional Regulation of Lung Diseases. *Front Genet* 2021; 12: 767348.

126. Bi G, Liang J, Zhao M, Zhang H, Jin X, Lu T, Zheng Y, Bian Y, Chen Z, Huang Y, Besskaya V, Zhan C, Wang Q, Tan L. miR-6077 promotes cisplatin/pemetrexed resistance in lung adenocarcinoma via CDKN1A/cell cycle arrest and KEAP1/ferroptosis pathways. *Mol Ther Nucleic Acids* 2022; 28: 366-386.

127. Wijsenbeek M, Cottin V. Spectrum of Fibrotic Lung Diseases. *N Engl J Med* 2020; 383: 958-968.

128. Garcia-Sancho C, Buendia-Roldan I, Fernandez-Plata MR, Navarro C, Perez-Padilla R, Vargas MH, Loyd JE, Selman M. Familial pulmonary fibrosis is the strongest risk factor for idiopathic pulmonary fibrosis. *Respir Med* 2011; 105: 1902-1907.

129. Allen RJ. Using Families to Study Pulmonary Fibrosis Genetics. *Am J Respir Crit Care Med* 2023; 207: 1264-1266.

130. Oldham JM, Allen RJ, Lorenzo-Salazar JM, Molyneaux PL, Ma SF, Joseph C, Kim JS, Guillen-Guio B, Hernandez-Beeftink T, Kropski JA, Huang Y, Lee CT, Adegunsoye A, Pugashetti JV, Linderholm AL, Vo V, Strek ME, Jou J, Munoz-Barrera A, Rubio-Rodriguez LA, Hubbard R, Hirani N, Whyte MKB, Hart S, Nicholson AG, Lancaster L, Parfrey H, Rassl D, Wallace W, Valenzi E, Zhang Y, Mychaleckyj J, Stockwell A, Kaminski N, Wolters PJ, Molina-Molina M, Banovich NE, Fahy WA, Martinez FJ, Hall IP, Tobin MD, Maher TM, Blackwell TS, Yaspan BL, Jenkins RG, Flores C, Wain LV, Noth I. PCSK6 and Survival in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2023; 207: 1515-1524.

131. Partanen JJ, Happola P, Zhou W, Lehisto AA, Ainola M, Sutinen E, Allen RJ, Stockwell AD, Leavy OC, Oldham JM, Guillen-Guio B, Cox NJ, Hirbo JB, Schwartz DA, Fingerlin TE, Flores C, Noth I, Yaspan BL, Jenkins RG, Wain LV, Ripatti S, Pirinen M, International IPFGC, Global Biobank Meta-Analysis I, Laitinen T, Kaarteenaho R, Myllarniemi M, Daly MJ, Koskela JT. Leveraging global multi-ancestry meta-analysis in the study of idiopathic pulmonary fibrosis genetics. *Cell Genom* 2022; 2: 100181.

132. Leavy OC, Ma SF, Molyneaux PL, Maher TM, Oldham JM, Flores C, Noth I, Jenkins RG, Dudbridge F, Wain LV, Allen RJ. Proportion of Idiopathic Pulmonary Fibrosis Risk Explained by Known Common Genetic Loci in European Populations. *Am J Respir Crit Care Med* 2021; 203: 775-778.

133. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, Guillen-Guio B, Ma SF, Okamoto T, John AE, Obeidat M, Yang IV, Henry A, Hubbard RB, Navaratnam V, Saini G, Thompson N, Booth HL, Hart SP, Hill MR, Hirani N, Maher TM, McAnulty RJ, Millar AB, Molyneaux PL, Parfrey H, Rassl DM, Whyte MKB, Fahy WA, Marshall RP, Oballa E, Bosse Y, Nickle DC, Sin DD, Timens W, Shrine N, Sayers I, Hall IP, Noth I, Schwartz DA, Tobin MD, Wain LV, Jenkins RG. Genetic

variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017; 5: 869-880.

134. Liu Q, Zhou Y, Cogan JD, Mitchell DB, Sheng Q, Zhao S, Bai Y, Ciombor KK, Sabusap CM, Malabanan MM, Markin CR, Douglas K, Ding G, Banovich NE, Nickerson DA, Blue EE, Bamshad MJ, Brown KK, Schwartz DA, Phillips JA, 3rd, Martinez-Barricarte R, Salisbury ML, Shyr Y, Loyd JE, Kropski JA, Blackwell TS. The Genetic Landscape of Familial Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2023; 207: 1345-1357.

135. Sivley RM, Sheehan JH, Kropski JA, Cogan J, Blackwell TS, Phillips JA, Bush WS, Meiler J, Capra JA. Three-dimensional spatial analysis of missense variants in RTEL1 identifies pathogenic variants in patients with Familial Interstitial Pneumonia. *BMC Bioinformatics* 2018; 19: 18.

136. Donoghue LJ, Stockwell AD, Neighbors M, Sheng RX, Prabhakaran R, Wolters PJ, Lancaster LH, Kropski JA, Blackwell TS, McCarthy MI, Yaspan BL. Identification of a Genetic Susceptibility Locus for Idiopathic Pulmonary Fibrosis in the 16p Subtelomere Using Whole-Genome Sequencing. *Am J Respir Crit Care Med* 2023; 207: 941-944.

137. Okamoto T, Dobrinskikh E, Hennessy CE, Liu N, Schwarz MI, Evans CM, Fontenot AP, Yang IV, Schwartz DA. Muc5b plays a role in the development of inflammation and fibrosis in hypersensitivity pneumonitis induced by Saccharopolyspora rectivirgula. *Am J Physiol Lung Cell Mol Physiol* 2022; 323: L329-L337.

138. Borie R, Cardwell J, Konigsberg IR, Moore CM, Zhang W, Sasse SK, Gally F, Dobrinskikh E, Walts A, Powers J, Brancato J, Rojas M, Wolters PJ, Brown KK, Blackwell TS, Nakanishi T, Richards JB, Gerber AN, Fingerlin TE, Sachs N, Pulit SL, Zappala Z, Schwartz DA, Yang IV. Colocalization of Gene Expression and DNA Methylation with Genetic Risk Variants Supports Functional Roles of MUC5B and DSP in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2022; 206: 1259-1270.

139. Putman RK, Gudmundsson G, Araki T, Nishino M, Sigurdsson S, Gudmundsson EF, Eiriksdottir G, Aspelund T, Ross JC, San Jose Estepar R, Miller ER, Yamada Y, Yanagawa M, Tomiyama N, Launer LJ, Harris TB, El-Chemaly S, Raby BA, Cho MH, Rosas IO, Washko GR, Schwartz DA, Silverman EK, Gudnason V, Hatabu H, Hunninghake GM. The MUC5B promoter polymorphism is associated with specific interstitial lung abnormality subtypes. *Eur Respir J* 2017; 50.

140. Knowles DA, Burrows CK, Blischak JD, Patterson KM, Serie DJ, Norton N, Ober C, Pritchard JK, Gilad Y. Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *Elife* 2018; 7.

141. Banovich NE, Li YI, Raj A, Ward MC, Greenside P, Calderon D, Tung PY, Burnett JE, Myrthil M, Thomas SM, Burrows CK, Romero IG, Pavlovic BJ, Kundaje A, Pritchard JK, Gilad Y. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res* 2018; 28: 122-131.

142. Sarkar AK, Tung PY, Blischak JD, Burnett JE, Li YI, Stephens M, Gilad Y. Discovery and characterization of variance QTLs in human induced pluripotent stem cells. *PLoS Genet* 2019; 15: e1008045.

143. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009; 19: 1655-1664.

144. Taub MA, Conomos MP, Keener R, Iyer KR, Weinstock JS, Yanek LR, Lane J, Miller-Fleming TW, Brody JA, Raffield LM, McHugh CP, Jain D, Gogarten SM, Laurie CA, Keramati A, Arvanitis M, Smith AV, Heavner B, Barwick L, Becker LC, Bis JC, Blangero J, Bleecker ER, Burchard EG, Celedon JC, Chang YPC, Custer B, Darbar D, de Las Fuentes L, DeMeo DL, Freedman BI, Garrett ME, Gladwin MT, Heckbert SR, Hidalgo BA, Irvin MR, Islam T, Johnson WC, Kaab S, Launer L, Lee J, Liu S, Moscati A, North KE, Peyser PA, Rafaels N, Seidman C, Weeks DE, Wen F, Wheeler MM, Williams LK, Yang IV, Zhao W, Aslibekyan S, Auer PL, Bowden DW, Cade BE, Chen Z, Cho MH, Cupples LA, Curran JE, Daya M, Deka R, Eng C, Fingerlin TE, Guo X, Hou L, Hwang SJ, Johnsen JM, Kenny EE, Levin AM, Liu C, Minster RL, Naseri T, Nouraie M, Reupena MS, Sabino EC, Smith JA, Smith NL, Su JL, Taylor JG, Telen MJ, Tiwari HK, Tracy RP, White MJ, Zhang Y, Wiggins KL, Weiss ST, Vasan RS, Taylor KD, Sinner MF, Silverman EK, Shoemaker MB, Sheu WH, Sciurba F, Schwartz DA, Rotter JI, Roden D, Redline S, Raby BA, Psaty BM, Peralta JM, Palmer ND, Nekhai S, Montgomery CG, Mitchell BD, Meyers DA, McGarvey ST, Network NC, Mak AC, Loos RJ, Kumar R, Kooperberg C, Konkle BA, Kelly S, Kardia SL, Kaplan R, He J, Gui H, Gilliland FD, Gelb BD, Fornage M, Ellinor PT, de Andrade M, Correa A, Chen YI, Boerwinkle E, Barnes KC, Ashley-Koch AE, Arnett DK, Consortium NT-OfPM, Hematology TO, Hemostasis Working G, Group TOSVW, Laurie CC, Abecasis G, Nickerson DA, Wilson JG, Rich SS, Levy D, Ruczinski I, Aviv A, Blackwell TW, Thornton T, O'Connell J, Cox NJ, Perry JA, Armanios M, Battle A, Pankratz N, Reiner AP, Mathias RA. Genetic determinants of telomere length from 109,122 ancestrally diverse whole-genome sequences in TOPMed. *Cell Genom* 2022; 2.

145. Alder JK, Sutton RM, Iasella CJ, Nouraie M, Koshy R, Hannan SJ, Chan EG, Chen X, Zhang Y, Brown M, Popescu I, Veatch M, Saul M, Berndt A, Methe BA, Morris A, Pilewski JM, Sanchez PG, Morrell MR, Shapiro SD, Lindell KO, Gibson KF, Kass

DJ, McDyer JF. Lung transplantation for idiopathic pulmonary fibrosis enriches for individuals with telomere-mediated disease. *J Heart Lung Transplant* 2022; 41: 654-663.

146. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. *Nat Commun* 2017; 8: 15452.

147. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; 10: e1004383.

148. Oldham JM, Kumar D, Lee C, Patel SB, Takahashi-Manns S, Demchuk C, Strek ME, Noth I. Thyroid Disease Is Prevalent and Predicts Survival in Patients With Idiopathic Pulmonary Fibrosis. *Chest* 2015; 148: 692-700.

149. Adegunsoye A, Oldham JM, Husain AN, Chen L, Hsu S, Montner S, Chung JH, Vij R, Noth I, Strek ME. Autoimmune Hypothyroidism As a Predictor of Mortality in Chronic Hypersensitivity Pneumonitis. *Front Med (Lausanne)* 2017; 4: 170.

150. Enomoto T, Usuki J, Azuma A, Nakagawa T, Kudoh S. Diabetes mellitus may increase risk for idiopathic pulmonary fibrosis. *Chest* 2003; 123: 2007-2011.

151. Gribbin J, Hubbard R, Smith C. Role of diabetes mellitus and gastro-oesophageal reflux in the aetiology of idiopathic pulmonary fibrosis. *Respir Med* 2009; 103: 927-931.

152. Garcia-Sancho Figueroa MC, Carrillo G, Perez-Padilla R, Fernandez-Plata MR, Buendia-Roldan I, Vargas MH, Selman M. Risk factors for idiopathic pulmonary fibrosis in a Mexican population. A case-control study. *Respir Med* 2010; 104: 305-309.

153. Rangarajan S, Bone NB, Zmijewska AA, Jiang S, Park DW, Bernard K, Locy ML, Ravi S, Deshane J, Mannon RB, Abraham E, Darley-Usmar V, Thannickal VJ, Zmijewski JW. Metformin reverses established lung fibrosis in a bleomycin model. *Nat Med* 2018; 24: 1121-1127.

154. Moore BB, Moore TA. Viruses in Idiopathic Pulmonary Fibrosis. Etiology and Exacerbation. *Ann Am Thorac Soc* 2015; 12 Suppl 2: S186-192.

155. Vockley J, Ensenauer R. Isovaleric acidemia: new aspects of genetic and phenotypic heterogeneity. *Am J Med Genet C Semin Med Genet* 2006; 142C: 95-103.

156. Couce ML, Aldamiz-Echevarria L, Bueno MA, Barros P, Belanger-Quintana A, Blasco J, Garcia-Silva MT, Marquez-Armenteros AM, Vitoria I, Vives I, Navarrete R, Fernandez-Marmiesse A, Perez B, Perez-Cerda C. Genotype and phenotype

characterization in a Spanish cohort with isovaleric acidemia. *J Hum Genet* 2017; 62: 355-360.

157. Vockley J, Rogan PK, Anderson BD, Willard J, Seelan RS, Smith DI, Liu W. Exon skipping in IVD RNA processing in isovaleric acidemia caused by point mutations in the coding region of the IVD gene. *Am J Hum Genet* 2000; 66: 356-367.

158. Saso K, Miyoshi N, Fujino S, Sasaki M, Yasui M, Ohue M, Ogino T, Takahashi H, Uemura M, Matsuda C, Mizushima T, Doki Y, Eguchi H. Dipeptidyl Peptidase 9 Increases Chemoresistance and is an Indicator of Poor Prognosis in Colorectal Cancer. *Ann Surg Oncol* 2020; 27: 4337-4347.

159. Gabrilovac J, Cupic B, Zapletal E, Kraus O, Jakic-Razumovic J. Dipeptidyl peptidase 9 (DPP9) in human skin cells. *Immunobiology* 2017; 222: 327-342.

160. Wilson CH, Indarto D, Doucet A, Pogson LD, Pitman MR, McNicholas K, Menz RI, Overall CM, Abbott CA. Identifying natural substrates for dipeptidyl peptidases 8 and 9 using terminal amine isotopic labeling of substrates (TAILS) reveals in vivo roles in cellular homeostasis and energy metabolism. *J Biol Chem* 2013; 288: 13936-13949.

161. Huang M, Zhang X, Toh GA, Gong Q, Wang J, Han Z, Wu B, Zhong F, Chai J. Structural and biochemical mechanisms of NLRP1 inhibition by DPP9. *Nature* 2021; 592: 773-777.

162. Hatzfeld M, Keil R, Magin TM. Desmosomes and Intermediate Filaments: Their Consequences for Tissue Mechanics. *Cold Spring Harb Perspect Biol* 2017; 9.

163. Al-Jassar C, Bikker H, Overduin M, Chidgey M. Mechanistic basis of desmosome-targeted diseases. *J Mol Biol* 2013; 425: 4006-4022.

164. Peljto AL, Blumhagen RZ, Walts AD, Cardwell J, Powers J, Corte TJ, Dickinson JL, Glaspole I, Moodley YP, Vasakova MK, Bendstrup E, Davidsen JR, Borie R, Crestani B, Dieude P, Bonella F, Costabel U, Gudmundsson G, Donnelly SC, Egan J, Henry MT, Keane MP, Kennedy MP, McCarthy C, McElroy AN, Olaniyi JA, O'Reilly KMA, Richeldi L, Leone PM, Poletti V, Puppo F, Tomassetti S, Luzzi V, Kokturk N, Mogulkoc N, Fiddler CA, Hirani N, Jenkins RG, Maher TM, Molyneaux PL, Parfrey H, Braybrooke R, Blackwell TS, Jackson PD, Nathan SD, Porteous MK, Brown KK, Christie JD, Collard HR, Eickelberg O, Foster EE, Gibson KF, Glassberg M, Kass DJ, Kropski JA, Lederer D, Linderholm AL, Loyd J, Mathai SK, Montesi SB, Noth I, Oldham JM, Palmisciano AJ, Reichner CA, Rojas M, Roman J, Schluger N, Shea BS, Swigris JJ, Wolters PJ, Zhang Y, Prele CMA, Enghelmayer JI, Otaola M, Ryerson CJ, Salinas M, Sterclova M, Gebremariam TH, Myllarniemi M, Carbone RG, Furusawa H, Hirose M, Inoue Y, Miyazaki Y, Ohta K, Ohta S, Okamoto T, Kim DS,

Pardo A, Selman M, Aranda AU, Park MS, Park JS, Song JW, Molina-Molina M, Planas-Cerezales L, Westergren-Thorsson G, Smith AV, Manichaikul AW, Kim JS, Rich SS, Oelsner EC, Barr RG, Rotter JI, Dupuis J, O'Connor G, Vasan RS, Cho MH, Silverman EK, Schwarz MI, Steele MP, Lee JS, Yang IV, Fingerlin TE, Schwartz DA. Idiopathic Pulmonary Fibrosis Is Associated with Common Genetic Variants and Limited Rare Variants. *Am J Respir Crit Care Med* 2023; 207: 1194-1202.

165. Kachuri L, Johansson M, Rashkin SR, Graff RE, Bosse Y, Manem V, Caporaso NE, Landi MT, Christiani DC, Vineis P, Liu G, Scelo G, Zaridze D, Shete SS, Albanes D, Aldrich MC, Tardon A, Rennert G, Chen C, Goodman GE, Doherty JA, Bickeboller H, Field JK, Davies MP, Dawn Teare M, Kiemeney LA, Bojesen SE, Haugen A, Zienolddiny S, Lam S, Le Marchand L, Cheng I, Schabath MB, Duell EJ, Andrew AS, Manjer J, Lazarus P, Arnold S, McKay JD, Emami NC, Warkentin MT, Brhane Y, Obeidat M, Martin RM, Relton C, Davey Smith G, Haycock PC, Amos CI, Brennan P, Witte JS, Hung RJ. Immune-mediated genetic pathways resulting in pulmonary function impairment increase lung cancer susceptibility. *Nat Commun* 2020; 11: 27.

166. Campoy E, Puig M, Yakymenko I, Lerga-Jaso J, Caceres M. Genomic architecture and functional effects of potential human inversion supergenes. *Philos Trans R Soc Lond B Biol Sci* 2022; 377: 20210209.

167. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006; 34: D140-144.

168. Goff LA, Davila J, Swerdel MR, Moore JC, Cohen RI, Wu H, Sun YE, Hart RP. Ago2 immunoprecipitation identifies predicted microRNAs in human embryonic stem cells and neural precursors. *PLoS One* 2009; 4: e7192.

169. Kang M, Ryu J, Lee D, Lee MS, Kim HJ, Nam SH, Song HE, Choi J, Lee GH, Kim TY, Lee H, Kim SJ, Ye SK, Kim S, Lee JW. Correlations between transmembrane 4 L6 family member 5 (TM4SF5), CD151, and CD63 in liver fibrotic phenotypes and hepatic migration and invasive capacities. *PLoS One* 2014; 9: e102817.

170. Perez MD, Seu L, Lowman KE, Moylan DC, Tidwell C, Samuel S, Duverger A, Wagner FH, Carlin E, Sharma V, Pope B, Raman C, Erdmann N, Locke J, Hu H, Sabbaj S, Kutsch O. The tetraspanin CD151 marks a unique population of activated human T cells. *Sci Rep* 2020; 10: 15748.

171. Kaul B, Lee JS, Zhang N, Vittinghoff E, Sarmiento K, Collard HR, Whooley MA. Epidemiology of Idiopathic Pulmonary Fibrosis among U.S. Veterans, 2010-2019. *Ann Am Thorac Soc* 2022; 19: 196-203.

172. Dove EP, Olson AL, Glassberg MK. Trends in Idiopathic Pulmonary Fibrosis-related Mortality in the United States: 2000-2017. *Am J Respir Crit Care Med* 2019; 200: 929-931.

173. Schafer MJ, White TA, Iijima K, Haak AJ, Ligresti G, Atkinson EJ, Oberg AL, Birch J, Salmonowicz H, Zhu Y, Mazula DL, Brooks RW, Fuhrmann-Stroissnigg H, Pirtskhalava T, Prakash YS, Tchkonia T, Robbins PD, Aubry MC, Passos JF, Kirkland JL, Tschumperlin DJ, Kita H, LeBrasseur NK. Cellular senescence mediates fibrotic pulmonary disease. *Nat Commun* 2017; 8: 14532.

174. Kellogg DL, Kellogg DL, Jr., Musi N, Nambiar AM. Cellular Senescence in Idiopathic Pulmonary Fibrosis. *Curr Mol Biol Rep* 2021; 7: 31-40.

175. Alder JK, Chen JJ, Lancaster L, Danoff S, Su SC, Cogan JD, Vulto I, Xie M, Qi X, Tuder RM, Phillips JA, 3rd, Lansdorp PM, Loyd JE, Armanios MY. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci U S A* 2008; 105: 13051-13056.

176. Newton CA, Kozlitina J, Lines JR, Kaza V, Torres F, Garcia CK. Telomere length in patients with pulmonary fibrosis associated with chronic lung allograft dysfunction and post-lung transplantation survival. *J Heart Lung Transplant* 2017; 36: 845-853.

177. Courtwright AM, El-Chemaly S. Telomeres in Interstitial Lung Disease: The Short and the Long of It. *Ann Am Thorac Soc* 2019; 16: 175-181.

178. Schneider JL, Rowe JH, Garcia-de-Alba C, Kim CF, Sharpe AH, Haigis MC. The aging lung: Physiology, disease, and immunity. *Cell* 2021; 184: 1990-2019.

179. Budinger GRS, Kohanski RA, Gan W, Kobor MS, Amaral LA, Armanios M, Kelsey KT, Pardo A, Tuder R, Macian F, Chandel N, Vaughan D, Rojas M, Mora AL, Kovacs E, Duncan SR, Finkel T, Choi A, Eickelberg O, Chen D, Agusti A, Selman M, Balch WE, Busse P, Lin A, Morimoto R, Sznajder JI, Thannickal VJ. The Intersection of Aging Biology and the Pathobiology of Lung Diseases: A Joint NHLBI/NIA Workshop. *J Gerontol A Biol Sci Med Sci* 2017; 72: 1492-1500.

180. Armanios M. Telomeres and age-related disease: how telomere biology informs clinical paradigms. *J Clin Invest* 2013; 123: 996-1002.

181. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am J Respir Crit Care Med* 2014; 189: 1161-1172.

182. Alder JK, Barkauskas CE, Limjunyawong N, Stanley SE, Kembou F, Tuder RM, Hogan BL, Mitzner W, Armanios M. Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A* 2015; 112: 5099-5104.

183. Le Saux CJ, Davy P, Brampton C, Ahuja SS, Fauce S, Shivshankar P, Nguyen H, Ramaseshan M, Tressler R, Pirot Z, Harley CB, Allsopp R. A novel telomerase activator suppresses lung damage in a murine model of idiopathic pulmonary fibrosis. *PLoS One* 2013; 8: e58423.

184. Pineiro-Hermida S, Autilio C, Martinez P, Bosch F, Perez-Gil J, Blasco MA. Telomerase treatment prevents lung profibrotic pathologies associated with physiological aging. *J Cell Biol* 2020; 219.

185. Huang Y, Guzy R, Ma SF, Bonham CA, Jou J, Schulte JJ, Kim JS, Barros AJ, Espindola MS, Husain AN, Hogaboam CM, Sperling AI, Noth I. Central lung gene expression associates with myofibroblast features in idiopathic pulmonary fibrosis. *BMJ Open Respir Res* 2023; 10.

186. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, Collard HR, Wolters PJ, Bradford WZ, Kossen K, Seiwert SD, du Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isaksson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Cogan JD, Mason WR, Maher TM, Molyneaux PL, Wells AU, Moffatt MF, Selman M, Pardo A, Kim DS, Crapo JD, Make BJ, Regan EA, Walek DS, Daniel JJ, Kamatani Y, Zelenika D, Smith K, McKean D, Pedersen BS, Talbert J, Kidd RN, Markin CR, Beckman KB, Lathrop M, Schwarz MI, Schwartz DA. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613-620.

187. Goobie GC, Nouraie M, Zhang Y, Kass DJ, Ryerson CJ, Carlsten C, Johannson KA. Air Pollution and Interstitial Lung Diseases: Defining Epigenomic Effects. *Am J Respir Crit Care Med* 2020; 202: 1217-1224.

188. Goobie GC, Carlsten C, Johannson KA, Khalil N, Marcoux V, Assayag D, Manganas H, Fisher JH, Kolb MRJ, Lindell KO, Fabisiak JP, Chen X, Gibson KF, Zhang Y, Kass DJ, Ryerson CJ, Nouraie SM. Association of Particulate Matter Exposure With Lung Function and Mortality Among Patients With Fibrotic Interstitial Lung Disease. *JAMA Intern Med* 2022; 182: 1248-1259.

189. Truong L, Matern B, D'Orsogna L, Martinez P, Tilanus MGJ, De Santis D. A novel multiplexed 11 locus HLA full gene amplification assay using next generation sequencing. *HLA* 2020; 95: 104-116.

190. Anzar I, Sverchkova A, Samarakoon P, Ellingsen EB, Gaudernack G, Stratford R, Clancy T. Personalized HLA typing leads to the discovery of novel HLA alleles and tumor-specific HLA variants. *HLA* 2022; 99: 313-327.

191. Nagler A, Kalaora S, Barbolin C, Gangaev A, Ketelaars SLC, Alon M, Pai J, Benedek G, Yahalom-Ronen Y, Erez N, Greenberg P, Yagel G, Peri A, Levin Y, Satpathy AT,

Bar-Haim E, Paran N, Kvistborg P, Samuels Y. Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics. *Cell Rep* 2021; 35: 109305.

192. Karnaukhov V, Paes W, Woodhouse IB, Partridge T, Nicastri A, Brackenridge S, Shcherbinin D, Chudakov DM, Zvyagin IV, Ternette N, Koohy H, Borrow P, Shugay M. HLA variants have different preferences to present proteins with specific molecular functions which are complemented in frequent haplotypes. *Front Immunol* 2022; 13: 1067463.