Polygenic risk and rare variants in endotypes of idiopathic pulmonary fibrosis: a genetic analysis of population-based and case–control cohorts

Anna Duckworth, Leigh Jackson, Harry Green, Atlas Khan, Chen Wang, Andrew Condescu, Gareth Hawkes, Imre Noth, Fernando J Martinez, Ganesh Raghu, Chad A Newton, Matthew Moll, Michael Cho, Michael Gibbons, Chris J Scotton, Columbia Genomics Consortium, Krzysztof Kiryluk, Christine Kim Garcia, David Zhang*

Summary

Background Idiopathic pulmonary fibrosis (IPF) and telomere length are both strongly linked to rare and common genetic variants. Shortened telomere length might itself be causal for IPF. We aimed to evaluate whether rare and common variants compete or cooperate to confer genetic risk of IPF uniformly.

Methods In this genetic analysis, we used whole-genome sequencing (WGS) data from a discovery case–control cohort sequenced at Columbia University and validated findings using WGS data from Trans-Omics for Precision Medicine (TOPMed) and UK Biobank. In all cohorts, we identified rare damaging variants in disease-associated genes and computed control-normalised non-overlapping polygenic risk scores (PRS) for IPF and telomere length. We assessed the *MUC5B* rs35705950 single-nucleotide polymorphism (SNP), an IPF common risk variant with a large effect, independently from the polygenic scores. Telomere length in blood leukocytes was measured using a quantitative PCR assay for the discovery cohort and UK Biobank validation cohort. We conducted logistic regression (adjusting for age, sex, and principal components of ancestry) to evaluate the association between IPF risk and the *MUC5B* SNP, the IPF PRS excluding *MUC5B* (IPF-PRS-no*MUC5B*), and the PRS for telomere length in the overall cohort and analysed their effects in patient subgroups for IPF endotypes (carriers and non-carriers of rare variants stratified by telomere length cutoffs). To assess disease prediction, we calculated cross-validated area under the receiver operating receiver operating curve (AUC). We also compared the liability of IPF explained by genetic variables.

Findings The discovery cohort was recruited between April 23, 2003 and June 19, 2019 and included 777 patients with IPF and 2905 controls. We replicated the analyses in the TOPMed (1148 patients with IPF and 5202 controls) and UK Biobank (2739 patients with IPF and 395 331 controls) cohorts. 23–43% of patients with IPF had damaging rare variants or telomeres shorter than the tenth percentile. Analysis of the association of genetic variables with IPF diagnosis yielded odds ratios of 1·63 (95% CI 1·47–1·81) for telomere length PRS and 1·60 (1·44–1·77) for IPF-PRS-no*MUC5B* in the discovery cohort, with similar effect sizes for the two variables in the replication cohorts (1·47, 1·36–1·59 *vs* 1·37, 1·25–1·50 in TOPMed; 1·24, 1·19–1·29 *vs* 1·25, 1·21–1·30 in UK Biobank). The telomere length PRS had the greatest effect on disease risk in patients with IPF not harbouring rare variants and with telomere length shorter than the tenth percentile in the discovery cohort (2·02, 1·76–2·33) and UK Biobank replication cohort (1·70, 1·56–1·85). Accounting for clinical variables and all genetic variables (rare variants, *MUC5B* SNP, IPF PRS, and telomere length PRS) led to the best disease prediction in the discovery cohort (combined AUC 0·89), TOPMed cohort (0·89), and UK Biobank cohort (0·77). Rare and common variants contributed jointly to the genetic liability of IPF. The telomere length PRS accounted for 13% of the explained genetic liability of IPF in the discovery cohort and 8% and 13% in the TOPMed and UK Biobank cohorts, respectively.

Interpretation Common and rare genetic variation confer context-specific genetic risk in patients with IPF both competitively and cooperatively. In contrast to known IPF common risk variants, the telomere length PRS, which includes more than 180 genetic loci not previously associated with IPF, is associated with increased risk of disease in patients with specific IPF endotypes. Polygenic risk from telomere-associated common variants is a key feature of genetic heterogeneity in IPF.

Funding US National Institutes of Health, UK Medical Research Council, and UK National Institute for Health and Care Research.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a lethal, progressive, fibrosing disorder with no known cure. Studies have identified numerous rare and common genetic influences on the risk of developing IPF, including those that affect telomere biology.[1–5] Telomeres

Lancet Respir Med 2026; 14: 305–16

Published Online January 27, 2026 https://doi.org/10.1016/S2213-2600(25)00405-9

See Comment page 282

*Consortium members listed in the appendix (p 34)

Department of Biomedical and Clinical Sciences, University of Exeter, Exeter, UK (A Duckworth PhD, L Jackson PhD, H Green PhD, A Condescu BSc , G Hawkes PhD, Prof M Gibbons MBChB, Prof C J Scotton PhD); **NIHR Exeter Biomedical Research Centre, University of Exeter, Exeter, UK** (A Duckworth, A Condescu, Prof M Gibbons, Prof C J Scotton); **Academic Department of Respiratory Medicine, Royal Devon University Hospitals NHS Foundation Trust, Exeter, UK** (A Duckworth, Prof M Gibbons, Prof C J Scotton); **Division of Pulmonary and Critical Care, Department of Medicine, University of Virginia, Charlottesville, VA, USA** (Prof I Noth MD); **Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, University of Massachusetts, Worcester, MA, USA** (Prof F J Martinez MD); **Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, University of Washington Medical Center, Seattle, WA, USA** (Prof G Raghu MD); **Division of Pulmonary and Critical Care, Department of Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA** (C A Newton MD); **Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA** (M Moll MD, M Cho MD); **Division of Pulmonary and**

Critical Care Medicine, Department of Medicine, Mass General Brigham, Boston, MA, USA (M Moll, M Cho)**; Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, Department of Veterans Affairs, West Roxbury, MA, USA** (M Moll)**; Division of Nephrology, Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA** (A Khan PhD, C Wang PhD, Prof K Kiryluk MD)**; Precision Medicine Institute, Columbia University Irving Medical Center, New York, NY, USA** (Prof C Kim Garcia MD)**; Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA** (Prof C Kim Garcia, D Zhang MD)

Correspondence to: Dr David Zhang, Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032, USA
dz2409@cumc.columbia.edu

See Online for appendix

Research in context

Evidence before this study

We searched PubMed using the terms "common variants", "rare variants", "polygenic score", "endotype", and "IPF" for studies published between database inception and April 20, 2025, to identify integrated genetic studies of rare and common variants in idiopathic pulmonary fibrosis (IPF). We found no original articles that provided a comprehensive, combined analysis of common and rare variants associated with IPF. Nearly all studies had analysed common or rare variants in isolation; one article included analyses of both common and rare variants but reported results separately, without studying the combined effects of variants. We identified numerous reviews that discussed or conceptualised the relative contributions of rare and common genetic risk to IPF risk, but no studies provided or cited empirical data. Although IPF endotypes have been defined using transcriptomics and proteomics, we did not find any articles that used terminology based on genetic factors, although several studies distinguished IPF with rare variants or telomere length shorter than the tenth percentile.

Added value of this study

To our knowledge, this study provides the first comprehensive survey of the combined effects of common and rare risk variants in IPF. We identify subtype-specific genetic risk factors that substantially improve explanation of genetic liability and disease prediction. We uncover complex relationships between polygenic factors and rare variants, including preliminary evidence of polygenic modifiers in patients with IPF carrying rare damaging variants, which might serve as an explanation for incomplete penetrance. Genetic studies in IPF have been limited by sample size, leading to missing heritability. We show that leveraging large-scale genetic studies of causal molecular traits such as telomere length to overcome these limitations can improve our genetic understanding of IPF. These telomere-associated common variants are context-specific risk factors in some endotypes, highlighting previously unrecognised genetic heterogeneity that will be important for the future discovery of novel, reproducible genetic risk factors in IPF.

Implications of all the available evidence

A genetic basis for disease heterogeneity allows for research that focuses on relevant endotypes instead of taking an all-comer disease approach and advances precision medicine strategies for IPF. Polygenic factors that can modify the effects of a high-risk rare variant are a target for understanding disease-modifying pathways in IPF.

are hexamer repeats that serve as protective caps at the ends of chromosomes and naturally shorten with age or cellular replication. IPF risk variants in telomere maintenance genes, including *TERT, TERC,* and *RTEL1,* cause telomere shortening and have been identified in genome-wide association studies (GWAS) of common risk variants[6] and exome sequencing studies.[2,7] Large-scale GWAS of telomere length in healthy individuals have also yielded telomere-associated common variants.[8,9] Analysis of these telomere-associated single-nucleotide polymorphisms (SNPs) in our previous Mendelian randomisation study implicated telomere shortening as a causal factor for IPF.[4] About 20–30% of patients with IPF have short telomere length (ie, shorter than the tenth percentile) without harbouring a rare damaging variant in a telomere gene; polygenic risk from telomere-associated SNPs might account for unexplained short telomere length in these cases.[5] Shortened telomere length itself, with or without a rare variant, is linked to worse survival,[10–12] rapid progression,[5] and an adverse response to immunosuppression in patients with IPF.[13,14] Patients with rare damaging variants in telomere genes or with unexplained short telomeres might constitute subgroups at higher clinical risk of worse outcomes.

Previous studies have begun to uncover genetic heterogeneity in IPF related to the common *MUC5B* promoter polymorphism, other common risk variants, and rare mutations.[5,10,15,16] Despite these advances, the genetic risk still unaccounted for in IPF is sizeable. Existing studies show that 75% of patients with familial IPF do not have a genetic diagnosis[5] and that known common risk variants explain only about 10% of disease liability in the general population.[17] However, these studies use all-comer IPF cohorts and focus on rare or common variants in isolation, obfuscating the potential role of gene interaction and disease heterogeneity in genetic risk.

Identification of endotype-specific genetic risk in IPF can enable precision diagnostics and guide targeted therapy. Clarifying how rare high-risk mutations interact with polygenic risk might reveal genetic modifiers and clues to resiliency. These human genetic insights could power rational, precise drug design targeting the origins of IPF. In this study, we conducted, to our knowledge, the first integrated assessment of rare damaging variants and common risk variants in patients with IPF stratified by genetic endotypes. We leveraged existing large-scale genetic studies of telomere length heritability to improve the genetic understanding of IPF risk.

Methods

Patient cohorts

In this study, we conducted genetic analyses of three cohorts of patients with IPF and controls. The Columbia discovery cohort of patients with IPF and controls has been described previously.[5,7] The institutional review board at Columbia University Medical Center (AAAS0753, AAAS7495, and AAAP0052) approved this

study. All study participants provided written consent for research use of clinical and genetic data. Cohorts from the Trans-Omics for Precision Medicine (TOPMed) programme[18] and UK Biobank[19] were used for replication of the analyses; participants in these cohorts consented to research per cohort-specific protocol. For the TOPMed cohort, individuals with IPF (Database of Genotypes and Phenotypes accession number phs001607) were classified as cases, and participants of the Multi-Ethnic Study of Atherosclerosis (MESA; phs001416) and Framingham Heart study (phs000974) were classified as controls. Analysis of data from UK Biobank was carried out under UK Biobank application number 103356. UK Biobank protocols were approved by the National Research Ethics Service Committee. For the UK Biobank cohort, ICD-10 code J84.1 was used to define IPF cases, and exclusion of all J84 codes defined controls. Further details are provided in the appendix (pp 2–3).

Genome sequencing, rare variant definitions, and genomic analysis

Whole-genome sequencing (WGS) for the Columbia cohort was done on Illumina's NovaSeq 6000 platform according to standard protocols, with raw sequencing reads aligned to the hg19 reference genome. WGS data from the discovery cohort and the TOPMed and UK Biobank cohorts were similarly aligned, called, and genotyped on their respective pipelines (appendix pp 3–4). Rare qualifying variants in disease-associated genes predicted to be deleterious (*TERT, TERC, RTEL1, PARN, DKC1, TINF2, NAF1, ZCCHC8, SFTPC, SFTPA1/2*, and *KIF15*) were identified using in-silico tools as previously described.[5,7]

Polygenic risk scores

To compute polygenic risk scores (PRS), we took a pruning and thresholding approach to include genome-wide significant autosomal risk variants identified from GWAS summary statistics. We used independent GWAS[6,9] data to generate weighted scores, identifying 16 SNPs related to IPF and 190 SNPs related to telomere length (appendix pp 25–29). Published β coefficients for each SNP were used to weight minor allele counts to generate weighted scores.[5] Three genome-wide significant loci were associated with both IPF and telomere length (*TERT* rs7705526, *TERC* rs2293607, and *RTEL1* rs115610405; appendix pp 5, 30). Given its large effect, we studied the *MUC5B* rs35705950 promoter polymorphism separately from the polygenic scores, as previously described.[20]

For the primary analysis, we defined a PRS for IPF using 12 SNPs that excluded the three overlapping telomere-associated SNPs and the *MUC5B* SNP, hereafter termed the IPF-PRS-no*MUC5B*. We defined a telomere length PRS using the 190 SNPs identified as being related to telomere length. For each cohort, common risk variants and PRS were normalised to controls and represented as Z scores. For concordance with the IPF-PRS-no*MUC5B*, we negatively inverted the β values of the SNPs constituting the telomere length PRS so that higher polygenic score predicted lower telomere length. For sensitivity analyses, we used lasso regregression to examine multiple alternative polygenic scores that redistributed the three overlapping telomere-associated SNPs, as well as a score derived from an independent GWAS study[8] and an alternative IPF polygenic score[20] (appendix p 30). We also conducted sensitivity analyses on individuals of solely European ancestry and those of non-European ancestry. We conducted sensitivity analyses by stratifying patients with IPF into those with and without a family history of disease, or by specific genes affected by rare variants.

Telomere length measurement

For the Columbia and UK Biobank cohorts, peripheral leukocyte telomere length was measured via quantitative PCR as previously described.[21,22] In both cohorts, genomic DNA was isolated from blood leukocytes at the time of study enrolment and quality control measures were done. Age-adjusted percentiles were computed using a multi-ancestry panel of control individuals for the Columbia cohort. For the UK Biobank, internal age-adjusted percentiles were computed using measured ancestry-specific telomere length values across 5-year age-bands due to differences by ancestry groups.[22]

Genetic liability

To estimate the proportion of disease risk explained by genetic variables, we used the liability threshold model, which conceptually transforms the observed probability of a binary trait into a continuous liability scale whereby exceeding a threshold results in development of that trait. We estimated the genetic liability explained across a range of disease prevalences for each genetic predictor and IPF endotype. We estimated the proportion of variance explained by calculating the correlation attributed to the predictor in a linear model and comparing full and reduced models adjusted for other genetic predictors, as well as age, sex, and principal components of ancestry. We converted the proportion of variance explained to a liability scale accounting for case–control ascertainment across a range of observed prevalences of IPF in the general population.

Statistical analysis

We conducted logistic regression to assess the association of genetic variables with IPF diagnosis in each cohort, adjusting for age, sex, and principal components of ancestry. We selected five principal components on the basis of visualisation of flattening of variance explained on a scree plot (appendix p 6). We conducted analyses using all individuals with IPF versus controls as well as those with specific endotypes of IPF versus controls. We assessed gene interaction by

assessing the significance of pairwise interaction terms between genetic variables. To assess disease prediction, we calculated area under the receiver operating characteristic curve (AUC) with ten-fold cross-validation using R packages pROC and caret. Comparison of AUCs was done using DeLong's test. Comparisons of non-parametric continuous data were done using the Wilcoxon rank sum test for two groups or Kruskal–Wallis test for multiple groups. For meta-analyses, we used a random-effects model with the R package metafor. To estimate genetic correlation between traits, we conducted cross-trait linkage disequilibrium score regression[23] using published GWAS summary statistics. To minimise bias from population stratification, we focused on studies of individuals of European ancestry. We used linkage disequilibrium scores estimated from the 1000 Genomes Project[24] European reference panel for regression weights. All p values less than 0·05 were considered significant, with Bonferroni correction applied for multiple comparisons. Statistical analyses were done with R statistical analysis software, version 4.4.0.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The Columbia discovery cohort[7] included 777 unrelated individuals with IPF and 2905 non-IPF control individuals recruited between April 23, 2003, and

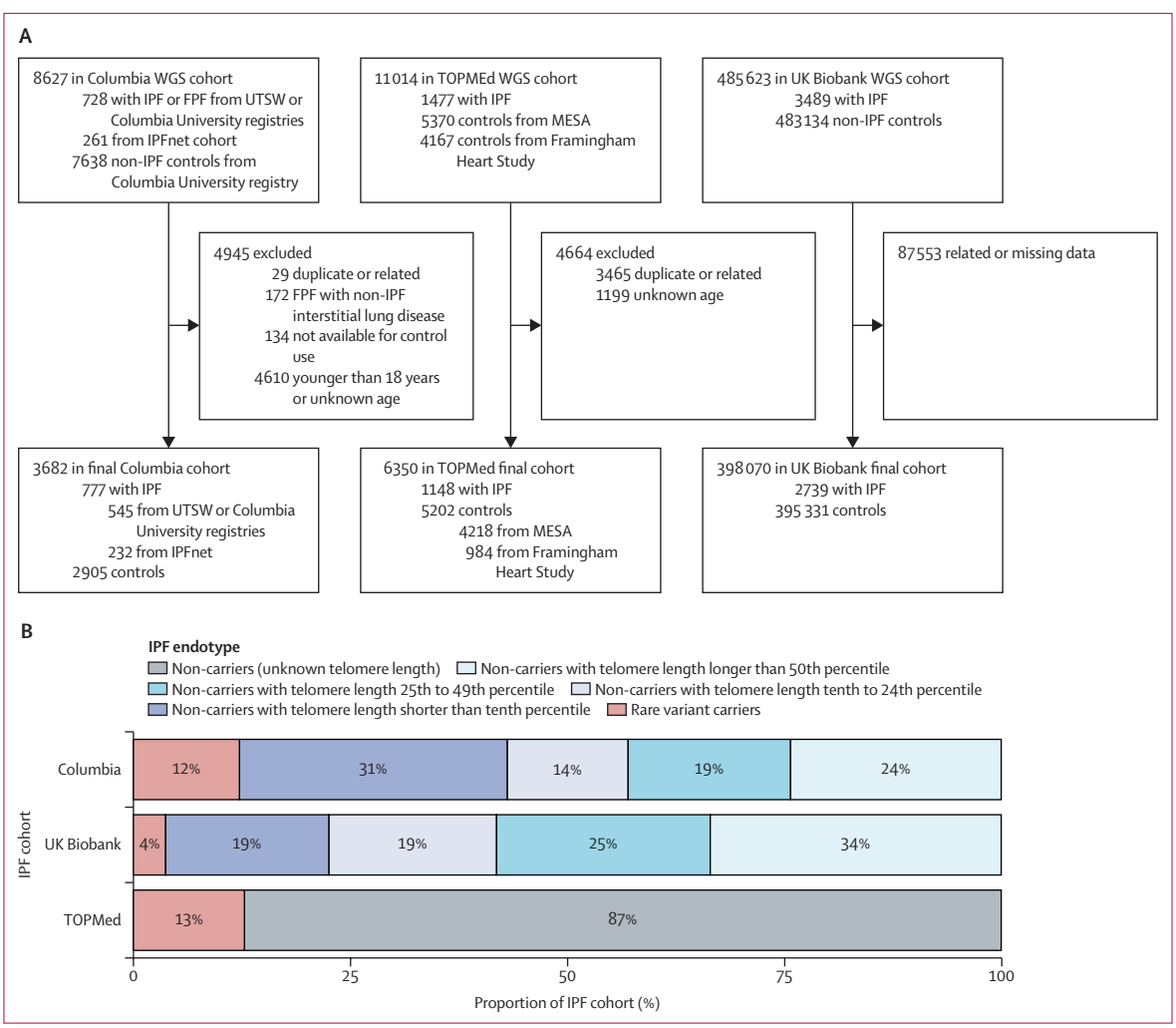


Figure 1: **Study cohorts and proportion of IPF endotypes**

(A) Flow diagrams for the three study cohorts. (B) Relative proportion of endotypes in each IPF cohort. Peripheral blood telomere length is presented as age-adjusted percentiles. Individuals without rare variants were subdivided on the basis of telomere length percentiles (shorter than tenth percentile, tenth to 24th percentile, 25th to 49th percentile, and longer than the 50th percentile). Telomere length percentiles were not available for the TOPMed cohort. FPF=familial pulmonary fibrosis. IPF=idiopathic pulmonary fibrosis. IPFnet=IPF Clinical Research Network. MESA=Multi-Ethnic Study of Atherosclerosis. UTSW=University of Texas Southwestern. WGS=whole-genome sequencing.

June 19, 2019. The TOPMed cohort included 1148 unrelated individuals with IPF and 5202 unrelated, population-based controls from the MESA (recruited between 2004 and 2007) and Framingham Heart studies.[18] The UK Biobank cohort[19] included 2739 unrelated individuals with IPF and 395 331 unrelated controls recruited between 2006 and 2010 (figure 1A). The three IPF cohorts had a median age of 64–67 years; 61–72% of participants were male, 28–39% were female, and 84–96% were of European ancestry (appendix p 30).

In the Columbia cohort, 94 (12%) of 777 patients with IPF were carriers of rare damaging variants in IPF-associated genes. In the replication cohorts, 143 (12%) of 1148 with IPF in the TOPMed cohort and 108 (4%) of 2739 patients with IPF in the UK Biobank cohort were rare variant carriers (appendix p 31). Telomere length measurement was only available for the Columbia and UK Biobank cohorts. For these cohorts, we subdivided patients with IPF into disease endotypes, namely carriers of rare variants and non-carriers stratified by telomere length (shorter than tenth percentile, tenth to 24th percentile, 25th to 49th percentile, and longer than 50th percentile; figure 1B; appendix pp 31–32). In most non-carriers, telomere length was shorter than the median for age; rare variant carriers and non-carriers with telomere length below the tenth percentile constituted 23–43% of those with IPF.

We found a significant difference in the distribution of all PRS between individuals with IPF and controls (all $p<0·0001$) in the three cohorts (appendix p 7). In the Columbia cohort there was heterogeneity in the distribution of telomere length PRS among individuals with IPF stratified by telomere length ($p=5·9\times10^{-5}$; appendix p 10). In multivariable analysis (adjusting for age, sex, genetic variables, and principal components of ancestry), we found significant independent associations between IPF risk and rare variants (OR 12·01, 95% CI 7·54–19·52; $p=1·0\times10^{-24}$), *MUC5B* SNP (2·00, 1·85–2·15; $p=4·0\times10^{-72}$), IPF-PRS-no*MUC5B* (1·60, 1·44–1·77; $p=9·4\times10^{-19}$), and telomere length PRS (1·63, 1·47–1·81; $p=5·5\times10^{-21}$; table). We did not find consistent evidence of a significant pairwise interaction term between genetic variables (appendix p 11). In the TOPMed and UKBiobank cohorts, we similarly identified independent associations between IPF risk and rare variants (18·2, 12·0–28·2 for TOPMed and 3·28, 2·66–4·01 for UK Biobank), *MUC5B* SNP (1·86, 1·77–1·97 for TOPMed and 1·45, 1·41–1·49 for UK Biobank), IPF-PRS-no*MUC5*B (1·37, 1·25–1·50 for TOPMed and 1·25, 1·21–1·30 for UK Biobank), and telomere length PRS (1·47, 1·36–1·59 for TOPMed and 1·24, 1·19–1·29 for UK Biobank). AUC analysis showed that adding genetic variables to a clinical model of age and sex improved disease prediction (figure 2). Analysis

	Columbia cohort			TOPMed cohort			UK Biobank cohort		
	Odds ratio (95% CI)	p value	AUC	Odds ratio (95% CI)	p value	AUC	Odds ratio (95% CI)	p value	AUC
Clinical model	..	..	0·74	..	..	0·64	..	..	0·73
Age	1·06 (1·05–1·07)	$6·1\times10^{-71}$	..	1·02 (1·01–1·03)	$5·9\times10^{-10}$	..	1·12 (1·11–1·12)	$1·3\times10^{-274}$	..
Male sex	1·85 (1·54–2·22)	$9·9\times10^{-12}$	..	2·43 (2·12–2·79)	$7·0\times10^{-37}$	..	1·75 (1·62–1·89)	$1·9\times10^{-45}$	..
MUC5B rs35705950	2·08 (1·95–2·23)	$5·5\times10^{-104}$	0·77	1·87 (1·78–1·97)	$6·1\times10^{-123}$	0·85	1·44 (1·40–1·48)	$4·1\times10^{-136}$	0·61
IPF-PRS-no*MUC5B*	1·61 (1·49–1·76)	$2·4\times10^{-29}$	0·71	1·45 (1·35–1·57)	$2·4\times10^{-21}$	0·80	1·25 (1·20–1·29)	$1·8\times10^{-33}$	0·57
Telomere length PRS	1·55 (1·43–1·68)	$4·0\times10^{-26}$	0·71	1·52 (1·42–1·63)	$1·4\times10^{-33}$	0·81	1·24 (1·20–1·29)	$9·2\times10^{-30}$	0·57
Rare variants	8·77 (6·01–13·01)	$2·6\times10^{-28}$	0·69	13·82 (9·49–20·57)	$1·6\times10^{-40}$	0·79	3·11 (2·52–3·78)	$7·8\times10^{-28}$	0·54
MUC5B rs35705950 + clinical	2·02 (1·88–2·17)	$1·3\times10^{-85}$	0·85	1·86 (1·77–1·97)	$2·3\times10^{-116}$	0·87	1·45 (1·41–1·49)	$4·2\times10^{-140}$	0·75
IPF-PRS-no*MUC5B* + clinical	1·63 (1·49–1·79)	$3·0\times10^{-26}$	0·82	1·43 (1·33–1·55)	$3·6\times10^{-19}$	0·82	1·25 (1·21–1·30)	$1·2\times10^{-33}$	0·74
Telomere length PRS + clinical	1·58 (1·45–1·73)	$2·4\times10^{-24}$	0·81	1·52 (1·42–1·63)	$1·8\times10^{-32}$	0·83	1·24 (1·19–1·28)	$3·6\times10^{-29}$	0·74
Rare variants + clinical	12·05 (7·88–18·8)	$2·8\times10^{-29}$	0·83	15·53 (10·6–23·4)	$6·0\times10^{-42}$	0·83	3·23 (2·62–3·95)	$2·6\times10^{-29}$	0·74
Combined polygenic + clinical	..	..	0·87	..	..	0·88	..	..	0·76
MUC5B rs35705950	2·01 (1·87–2·16)	$6·2\times10^{-77}$	..	1·84 (1·74–1·94)	$3·3\times10^{-107}$	..	1·45 (1·41–1·49)	$4·7\times10^{-140}$	..
IPF-PRS-no*MUC5B*	1·56 (1·41–1·72)	$4·3\times10^{-18}$	..	1·36 (1·24–1·48)	$5·8\times10^{-12}$	..	1·25 (1·20–1·29)	$2·2\times10^{-33}$	..
Telomere length PRS	1·64 (1·49–1·82)	$1·6\times10^{-22}$	..	1·46 (1·35–1·57)	$4·0\times10^{-22}$	..	1·24 (1·19–1·29)	$4·1\times10^{-29}$	..
Combined polygenic + rare + clinical	..	..	0·89	..	..	0·89	..	..	0·77
MUC5B rs35705950	2·00 (1·85–2·15)	$4·0\times10^{-72}$	..	1·86 (1·77–1·97)	$7·8\times10^{-104}$	..	1·45 (1·41–1·49)	$4·4\times10^{-140}$	..
IPF-PRS-no*MUC5B*	1·60 (1·44–1·77)	$9·4\times10^{-19}$	..	1·37 (1·25–1·50)	$7·6\times10^{-12}$	..	1·25 (1·21–1·30)	$1·0\times10^{-33}$	..
Telomere length PRS	1·63 (1·47–1·81)	$5·5\times10^{-21}$	..	1·47 (1·36–1·59)	$1·8\times10^{-21}$	..	1·24 (1·19–1·29)	$5·0\times10^{-29}$	..
Rare variants	12·01 (7·54–19·52)	$1·0\times10^{-24}$	..	18·21 (11·98–28·22)	$2·5\times10^{-40}$	..	3·28 (2·66–4·01)	$9·7\times10^{-30}$	..

The clinical model included genetic predictors, age, sex, and five principal components of ancestry. All associations with genetic predictors were adjusted for five principal components of ancestry. Common variant genetic variables are coded as control-normalised Z-transformed scores; each one unit increase in odds ratio represents one SD increase in genetic predictor. Rare variant variable is coded as presence or absence; each one unit increase in odds ratio represents presence of rare variant. AUC=area under the receiver operating characteristic curve. IPF-PRS-no*MUC5B*=idiopathic pulmonary fibrosis polygenic risk score excluding the *MUC5B* rs35705950 polymorphism and overlapping telomere-associated loci (appendix pp 2–4). PRS=polygenic risk score.

Table: **Associations between polygenic scores and IPF in multiple cohorts**

in all three cohorts showed that inclusion of clinical variables with all rare and common genetic variables, including telomere length PRS, led to the best prediction of IPF, with an AUC value of 0·89 in the Columbia cohort, 0·89 in the TOPMed cohort, and 0·77 in the UK Biobank cohort (table, figure 2). Sensitivity analysis with redistribution of the overlapping telomere-associated loci between polygenic scores showed similar findings (appendix p 33).

As polygenic scores had independent effects, we explored their effects on risk for different IPF endotypes. In the Columbia cohort (figure 3A), the *MUC5B* polymorphism was similarly associated with IPF across all endotypes (overall OR 2·02, 95% CI 1·88–2·17), despite having decreased prevalence in rare variant carriers (appendix p 32). IPF-PRS-no*MUC5B* was associated with IPF risk in analysis of non-carriers of rare variants versus controls

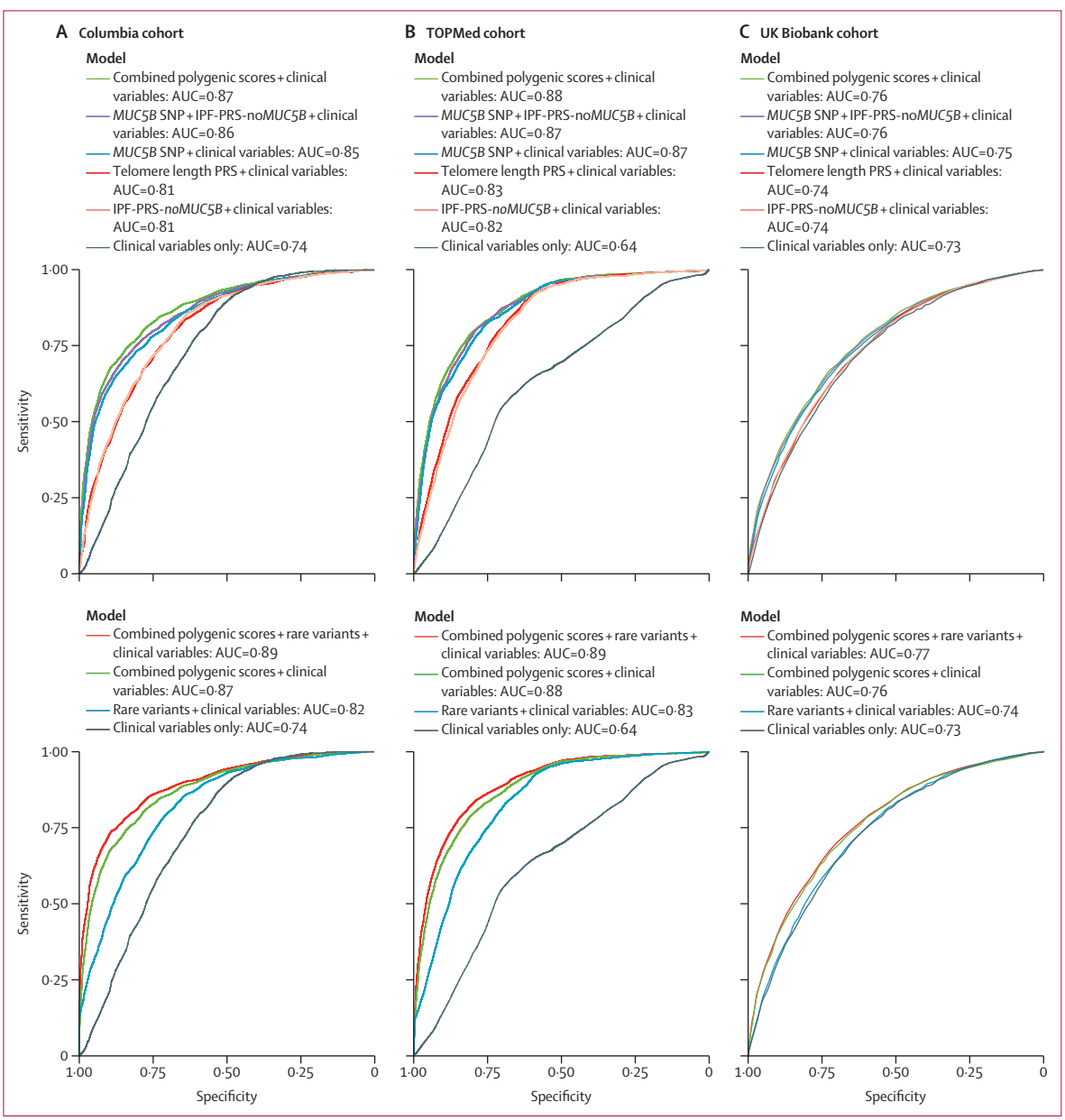


Figure 2: **Receiver operator curves of clinical and genetic predictors of IPF for the Columbia (A), TOPMed (B), and UK Biobank (C) cohorts**
The clinical model includes age and sex. In the upper plots, receiver operating characteristic curves combine the *MUC5B* SNP, polygenic scores, and clinical predictors, including age and sex; in all cohorts, prediction improved with the addition of individual genetic predictors to clinical predictors. In the lower plots, a model combining rare variants, common variants including the *MUC5B* SNP and polygenic scores, and clinical predictors resulted in the best disease prediction. IPF-PRS-no*MUC5B*=idiopathic pulmonary fibrosis polygenic risk score excluding *MUC5B* SNP and overlapping telomere-associated loci. AUC=area under the receiver operating characteristic curve. PRS=polygenic risk score. SNP=single-nucleotide polymorphism.

(1·69, 1·54–1·86) but not in carriers versus controls (1·23, 1·00–1·51). Telomere length PRS was associated with IPF risk in analysis of rare variant carriers versus controls (1·63, 1·33–1·99) and non-carriers versus controls (1·58, 1·44–1·73), with the highest OR in the group with telomere length shorter than the tenth percentile (2·02, 1·76–2·33) and lowest OR in the group with telomere length longer than the 50th percentile (1·23, 1·06–1·44). In sensitivity analyses, the use of alternative polygenic scores, including a PRS for IPF that retained the overlapping telomere-associated loci, yielded similar endotype-specific associations (appendix p 12). We also observed similar associations for both in familial IPF with a positive family history and for sporadic IPF without a family history (appendix p 13).

Similar to findings in the Columbia cohort, the effect of IPF-PRS-no*MUC5B* on disease risk was attenuated in in analysis of rare variant carriers versus non-carriers compared with controls in the TOPMed cohort (OR 1·30, 95% CI 1·08–1·57 for carriers *vs* 1·46, 1·34–1·59 for

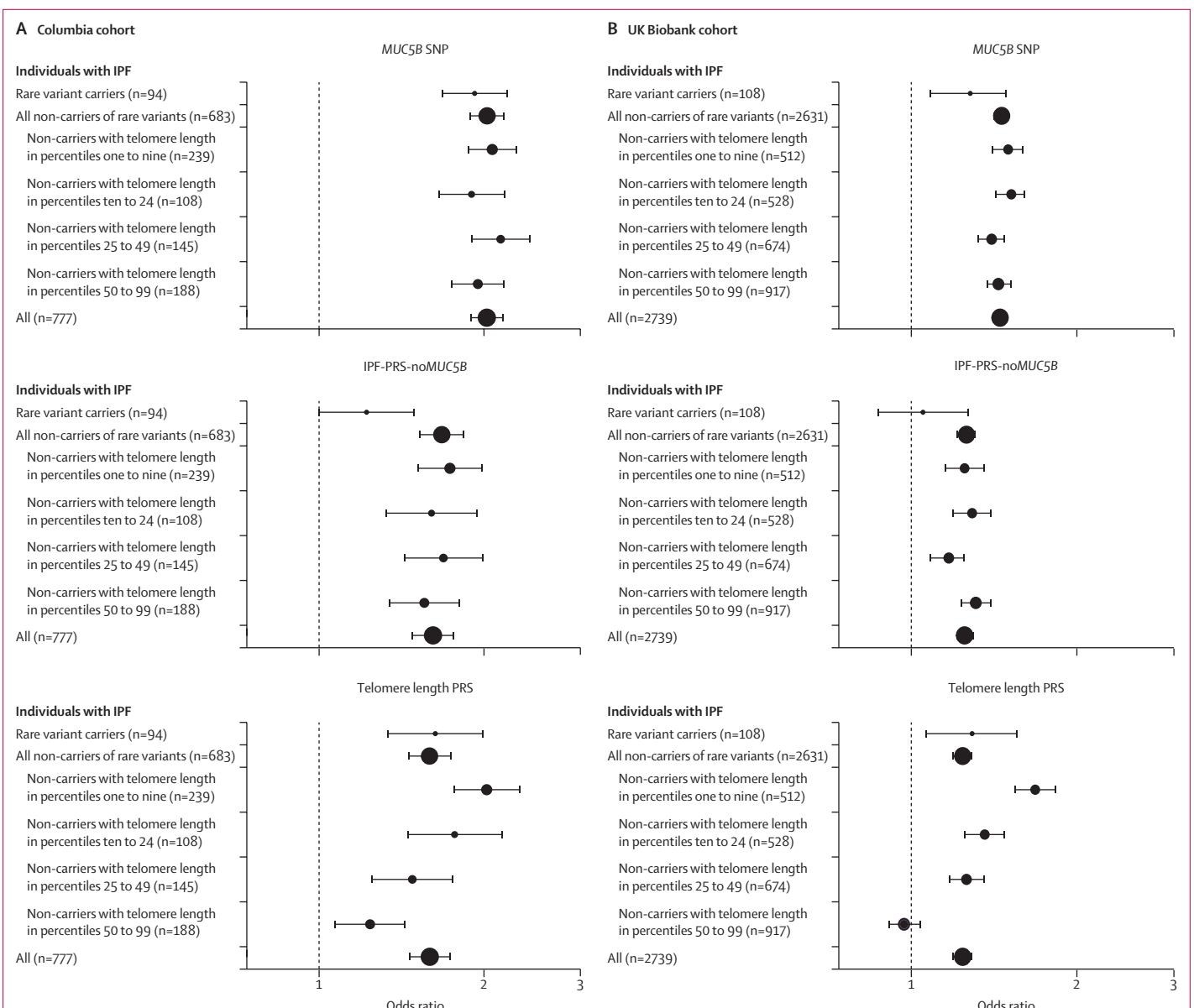
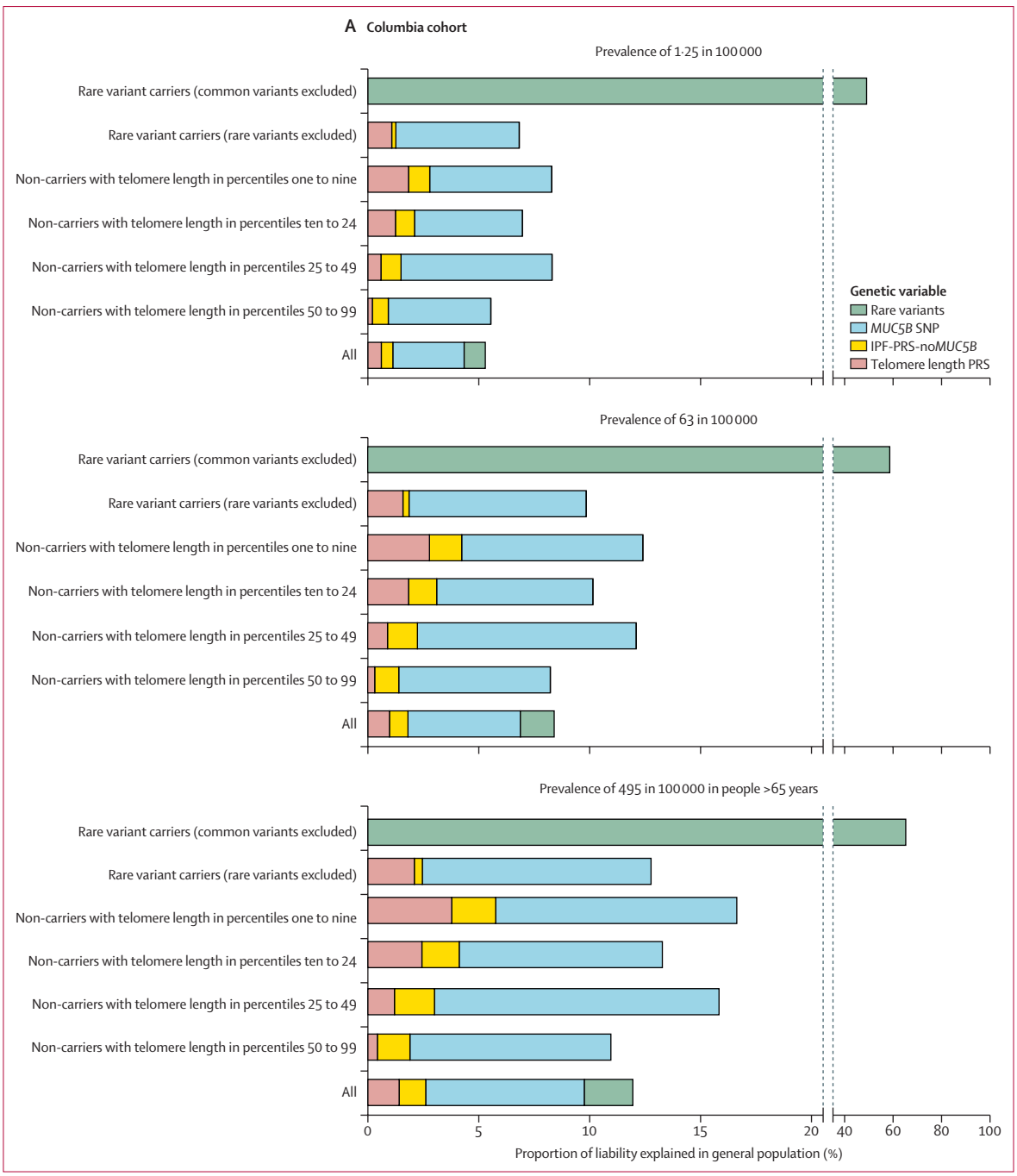


Figure 3: **Associations of genetic predictors and polygenic scores with IPF risk across endotypes of disease**
Odds ratios with 95% CIs (whiskers) calculated for the Columbia cohort (A) and UK Biobank cohort (B) for IPF overall and by disease endotype. All associations were adjusted for age, sex, and five principal components of ancestry. Genetic predictors are displayed as control-normalised Z-transformed scores; each one unit increase in OR represents one SD increase in genetic predictor. Dot size indicates the relative size of each IPF subset. IPF=idiopathic pulmonary fibrosis. IPF-PRS-no*MUC5B*=idiopathic pulmonary fibrosis polygenic risk score excluding *MUC5B* SNP and overlapping telomere-associated loci. PRS=polygenic risk score. SNP=single-nucleotide polymorphism.



(Figure 4 continues on next page)

non-carriers; appendix p 14) and UK Biobank cohort (1·05, 0·87–1·27 for carriers *vs* 1·26, 1·21–1·31; figure 3B). Telomere length PRS had a heterogeneous effect on disease risk in analysis of non-carriers versus controls in the UK Biobank cohort, with a significant association in the group with telomere length shorter than the tenth percentile (1·70, 1·56–1·85) but not in the group with telomere length longer than the 50th percentile (0·95, 0·93–1·06). Analysis of rare variant carriers separated by specific gene or by telomere-related and telomere-unrelated genes showed that the telomere length PRS and *MUC5B* SNP had largely similar effects (appendix pp 15–16).

In ancestry-stratified analysis, all associations remained consistent in individuals of European ancestry (appendix p 17). IPF-PRS-no*MUC5B* and telomere length PRS did not have a consistent pattern of association in non-European individuals (appendix p 18).

To conceptualise the relative contributions from each genetic variable, we compared the proportion of IPF risk conferred by using liability estimates across a range of IPF prevalences reported in the general population as well as in individuals older than 65 years[25,26] (figure 4A). For patients with IPF carrying rare variants, we reported the liability explained by rare variants and polygenic factors separately due to potential collinearity between genetic factors in this group. In the Columbia cohort, rare variants alone explained most of the liability (40–60%) for IPF with rare variants. Comparing polygenic risk factors, the PRS for telomere length explained more disease liability than the IPF-PRS-no*MUC5B* (1·1–2·0% *vs* 0·2–0·4%) in IPF with rare variants. The *MUC5*B SNP explained a major fraction of liability (60–61%) in patients with IPF without rare variants. Comparing polygenic scores in non-carriers, the telomere length PRS explained more liability than the IPF-PRS-no*MUC5B*, especially for non-carriers with telomere length shorter than the tenth percentile (1·8–3·6% *vs* 1·0–2·0%) or between the tenth and 24th percentiles (1·2–2·3% *vs* 0·9–1·8%). The endotype-specific genetic liability explained by the telomere length PRS persisted in sensitivity analyses that included only individuals of non-European ancestry (appendix p 19) or that used alternative polygenic scores (appendix pp 20–21). Across all patients with IPF, *MUC5B* polymorphism explained 3·1–6·9% of overall liability, rare variants explained 0·9–2·1%, the IPF-PRS-no*MUC5B* explained 0·5–1·2%, and telomere length PRS explained 0·6–1·3%. The telomere length PRS accounted for up to 13% of the genetic liability of IPF. We observed similar proportions of explained genetic liability overall and for specific endotypes in the TOPMed and UK Biobank replication cohorts (figure 4B–D). Across multiple cohorts, combining rare variants and telomere length PRS accounted for about 30% of the genetic liability.

To assess the genetic overlap between IPF and telomere length in the context of other epidemiologically associated traits, we estimated pairwise genetic correlation between IPF,[6] telomere length,[9] smoking,[27] gastro-oesophageal reflux disease,[28] monocyte count,[29] rheumatoid arthritis,[30] and forced vital capacity[31] using summary statistics from large-scale GWAS studies for all traits (appendix pp 22–23). Telomere length was the only trait that had significant genetic correlation with IPF after correcting for multiple comparison testing (r=–0·35; p=3·6×10^{-6}).

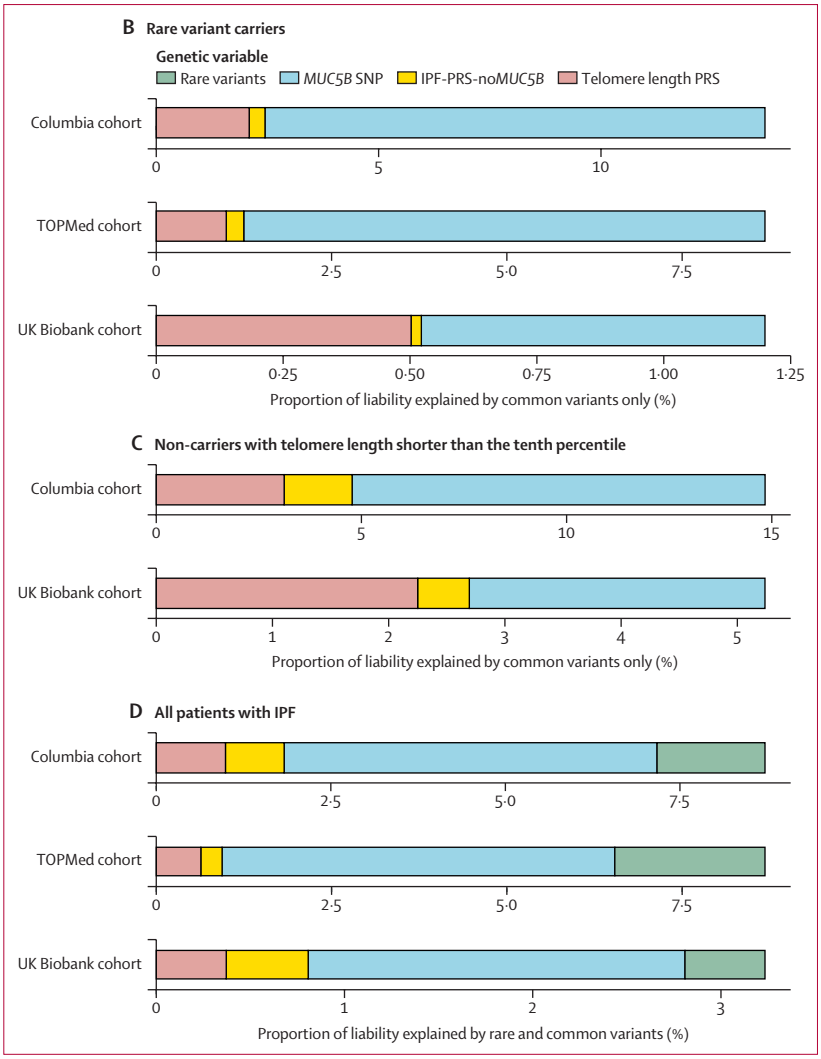


Figure 4: **Liability of developing IPF in the general population explained by known rare and common genetic risk factors**
(A) Proportion of liability in the Columbia cohort estimated by each genetic predictor, assuming a range of observed prevalences of IPF in the general population and in individuals older than 65 years. (B) Liability explained by common variants in patients with IPF carrying rare variants. (C) Liability explained by common variants in patients with IPF not carrying rare variants and with telomere length shorter than the tenth percentile. (D) Liability explained by all genetic variables in all patients with IPF. The assumed disease prevalence in (B–D) was 63 in 100 000. IPF-PRS-no*MUC5B*=idiopathic pulmonary fibrosis polygenic risk score excluding *MUC5B* SNP and overlapping telomere-associated loci. PRS=polygenic risk score. SNP=single-nucleotide polymorphism.

Discussion

This study used three independent cohorts comprising more than 4500 individuals with IPF and 400 000 controls to systematically characterise the interplay of rare and common genetic risk factors in genetic endotypes of IPF (summarised in the appendix, p 24). We observed large effects from polygenic risk of telomere shortening, especially in the 23–43% of individuals with IPF who were carriers of a damaging rare variant or had very short telomere length. Our study highlights the utility of understanding genetic heterogeneity for uncovering novel genetic risk factors for IPF.

Studying the genetic basis of associated molecular phenotypes can substantially improve our mechanistic understanding of diseases and lead to therapies that focus on treating causal traits. Successful examples of this approach include genetic studies of plasma concentrations of LDL cholesterol for coronary artery disease,[32] of fetal haemoglobin for haemoglobinopathies,[33] and of serum IgA for IgA nephropathy.[34] Similarly, our

approach leveraged GWAS studies of telomere length involving hundreds of thousands of participants to identify novel IPF genetic associations, capitalising on robust polygenic scores from well-powered analyses. Genetic studies of molecular traits such as telomere length often have larger sample sizes and yield more informative genetic findings than studies of a rare diagnosis like IPF.

We highlight a complex relationship between polygenic risk factors and rare variants in IPF. Non-*MUC5B* common risk variants had a diminished effect in carriers of rare damaging variants, suggesting that these genetic variables compete to confer risk of IPF. In contrast, the *MUC5B* SNP and telomere-associated common variants cooperated with rare risk variants in terms of risk. Most of these rare variants are also found in telomere maintenance genes, suggesting that rare and common variants that govern telomere length confer IPF risk additively. Rare telomere gene variants confer upwards of 40-fold increased odds of IPF and are linked to monogenic forms of disease[7] but have variable penetrance. Studies of familial pulmonary fibrosis kindreds[35] often identify asymptomatic rare variant carriers and have found that the *MUC5B* SNP might associate with early fibrotic lung disease. Polygenic modification has also been proposed to explain incomplete penetrance in other monogenic disorders, including cardiovascular disease, hereditary cancer, and chronic kidney disease.[36,37] Our findings pave the way to study polygenic modifiers as an explanation for variable penetrance in the families of patients with IPF who are rare variant carriers.

Historically, the genetic link between telomere shortening and IPF has been attributed to rare variants.[1–3] We found that telomere-associated common variants can also explain a substantial portion of genetic risk of IPF. In patients with IPF who have unexplained short telomere length, the polygenic risks of IPF and telomere shortening converged, highlighting a parallel genetic pathway for future studies of disease mechanism for this specific endotype. Furthermore, testing of telomere length and genetic testing for IPF are both available clinically,[38] allowing for identification of this endotype in clinical practice. Telomere-associated common variants remained enriched in up to 65–75% of patients with IPF with telomere lengths shorter than average for age (<50th percentile), suggesting that a spectrum of telomere-associated polygenic risk extends to most patients with IPF. Given the abundant evidence linking rare variants or short telomere lengths to reduced survival in IPF,[11,12] these findings suggest that rare and common telomere-associated variants not only underlie genetic heterogeneity but also drive the heterogeneity of disease outcomes.

We found that the overall genetic architecture of IPF correlated significantly with genetic determinants of telomere length but not with other epidemiologically associated traits. This parallel analysis complements our previous study implicating telomere shortening as causal for IPF.[4] Although genome-wide significant loci from GWAS studies of IPF and telomere length overlap to a minimal degree, the extent of genetic correlation between these two traits suggests that many other variants below the significance threshold are also overlapping. Our findings indicate that the extent of this overlap might be related to the genetic heterogeneity of the IPF GWAS cohort. As some common variants have attenuated effects in rare variant carriers with familial pulmonary fibrosis, a sporadic IPF cohort without rare variant carriers could be enriched for polygenic discovery. Similarly, a hypothetical GWAS analysis of an IPF cohort without short telomeres would likely yield different results compared with analysis of an IPF cohort in which all patients have short telomere lengths. Notably, the genetic locus in *TERT* was identified in a previous GWAS of IPF[39] but not in a similarly powered contemporaneous GWAS.[40] Imbalance of genetic endotypes might have led to discrepant results in these studies, highlighting the importance of understanding genetic heterogeneity in all-comer IPF cohorts.

Our focus on telomere length indicated that there were multiple endotypes with shared telomere shortening but distinct genetic influences. Unlike those carrying rare variants in telomere-related genes, patients with IPF who have unexplained short telomere length are also enriched in other IPF-related polygenic risk factors. Furthermore, although rare variants in IPF are largely found in telomere-related genes that govern telomerase activity, genes implicated by telomere-associated common variants are involved in additional pathways such as nucleotide metabolism.[8] Cell-based studies[41] and clinical trials (NCT06817590) have further evaluated thymidine replacement for increasing telomere length. The findings highlight the value of distinguishing these endotypes for exploring distinct causal pathways and identifying druggable targets for therapeutics.

Our study had several limitations. Due to the variability in clinical phenotyping of cohorts, some relevant risk factors were not uniformly available, including smoking, environmental exposures, and other comorbid conditions. Telomere length correlates genetically with some of these risk factors. We cannot definitively exclude confounding or effect modification from non-genetic factors, and the effect of gene–environment interactions will require further investigation. Since all major GWAS studies on IPF have focused on individuals of European ancestry, polygenic scores and associations might not be portable across non-European ancestries. Our sensitivity analyses using non-European individuals were limited by sample size and suggested a consistent direction of effect for polygenic scores. Future studies with diverse IPF cohorts will be needed to identify pan-ancestry genetic effects. Analyses of ancestrally diverse cohorts are crucial for understanding the full spectrum of pathogenic genetic

risk and identifying new disease mechanisms or treatment targets. Patients with IPF in the UK Biobank cohort were depleted of rare variants compared with clinical cohorts. This phenomenon might be related to enrolment bias in the UK Biobank, which preferentially recruited healthy older individuals, potentially excluding some individuals with pathogenic rare variants, who tend to be younger and sicker. Diagnostic certainty of IPF can be challenging in biobank cohorts, and clinical misclassification might also lead to discrepancy in prevalence for rare variants. Clinical IPF cohorts might also be biased towards enrolling those with a genetic aetiology; a recent analysis of the USA-wide Pulmonary Fibrosis Foundation Registry showed a 7% prevalence of rare variants in patients with IPF using the same definition, similar to this study.[15] We were unable to draw definitive conclusions for rare variant carriers regarding specific genes, although we observed consistent associations in the most well-represented genes. We focused on genetically classified IPF endotypes, but other complementary or orthogonal endotypes might exist. Lastly, our work forms the basis for incorporating genetics into precision diagnosis of IPF, but additional work is needed to translate these findings into daily clinical practice—including harmonisation of telomere length measurements, improved accessibility to genetic counselling and testing, and broader adoption of polygenic assessments in clinical diagnosis.

In conclusion, polygenic background and rare variation contribute to the genetic risk of IPF both independently and in tandem. Although attenuated, polygenic modifiers of disease risk exist even in patients with IPF who carry rare damaging variants. We found that genetically predicted short telomeres intersected with IPF risk not only in the purview of damaging rare variants but also in the context of polygenic risk in particular endotypes. Accounting for these novel genetic risk factors both improves disease prediction and explains additional genetic liability. Taken together, our findings underscore the importance of understanding IPF heterogeneity from rare and common genetic influences that might be relevant for future studies.

Contributors

DZ conceived the project and designed the study. DZ, AD, LJ, HG, AC, and GH conducted the analysis. AK, CW, CKG, and KK generated genetic data for the discovery cohort. CKG, IN, FJM, GR, CAN, and the Columbia Genomics Consortium enrolled and contributed samples for the discovery cohort. MM and MC provided analytic support for alternative polygenic scores. MG and CJS provided supervisory support to AD and facilitated access to UK Biobank. DZ and AD verified the data. DZ, AD, and CKG interpreted the results and drafted the manuscript. All authors had direct access to the raw data. As the corresponding author, DZ was responsible for the decision to submit the manuscript and DZ verifies that authors have seen and approve of the final text.

Declaration of interests

GH reports grant support from the EU Innovative Health Initiative Joint Undertaking. CAN reports editorial board participation for CHEST, committee participation for the American Thoracic Society, working group participation for the Pulmonary Fibrosis Foundation, and consulting fees from Boehringer Ingelheim and Medpace, unrelated to the current work. IN reports grant support from Veracyte and consulting fees from Boehringer Ingelheim and Sanofi, unrelated to the current work. FJM reports support from Boehringer Ingelheim, Biogen, Bristol-Myers Squibb, DevPro, GlaxoSmithKline, Nitto, Promedior/Roche, Vicore, and Chiesi; consulting fees from AstraZeneca, Boehringer Ingelheim, Bristol-Myers Squibb, Chiesi, Endeavor, Excalibur, GlaxoSmithKline, Lung Therapeutics/Aileron, Novartis, RS Biotherapeutics, Two XR, and Hoffman Laroche; and participation in data safety monitoring boards for Boehringer Ingelheim, Endeavor Biomedicine, and Pliant, unrelated to the current work. MM reports consulting fees from 2ndMD, TheaHealth, Axon Advisors, Dialectica, Sanofi, and Verona Pharma, unrelated to the current work. MC reports grant support from Bayer and consulting fees from Apogee and BMS, unrelated to the current work. KK reports grant support from the IgA Nephropathy Foundation and consulting fees from HIBIO, unrelated to the current work. CKG reports grant support from the US Department of Defense, research support from AstraZeneca, and equity or stocks from Rejuvenation Technologies. DZ reports grant support from the Francis Family Foundation, committee participation for the American Thoracic Society, and consulting fees from Boehringer Ingelheim. All other authors declare no competing interests.

Data sharing

Genotype data for the discovery cohort allowable under consent are available in the Database of Genotypes and Phenotypes (dbGAP; https://dbgap.ncbi.nlm.nih.gov/home/) under the following projects: Pulmonary Fibrosis and Telomerase Dysfunction (phs002692); Genomics of Glomerular Disorders (phs002480); and Genomic Translation for ALS Care (phs02973). Genetic and phenotypic data for the TOPMed cohort are available under the following dbGAP projects: Multi-Ethnic Study of Atherosclerosis (MESA; phs0001416); Framingham Heart Study (phs000974); and Idiopathic Pulmonary Fibrosis (IPF; phs001607). Genetic and phenotypic data from the UK Biobank are available through applications to the UK Biobank. Additional study results are available on reasonable request made to the corresponding author.

Acknowledgments

The authors thank the research participants for their participation and enrolment in study cohorts. This study was funded by the National Institute for Health and Care Research (NIHR) Exeter Biomedical Research Centre (grants to AD, LJ, HG, and AC), the UK Medical Research Council (grants to AD, LJ, and GH), and US National Institutes of Health (to CAN, IN, FJM, MM, MC, KK, CKG, and DZ). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. We thank the investigators and institutions who supported the National Heart, Lung, and Blood Institute cohorts (IPF, MESA, and the Framingham Heart Study). We also acknowledge the Collaborative Group of Genetic Studies of IPF for access to summary statistics of previous GWAS of idiopathic pulmonary fibrosis. The Columbia Genomics Consortium includes genomic data from a number of projects and contact investigators: the Columbia University Biobank (Muredach Reilly), Glomerulonephropathy (CureGN consortium members and non-CureGN members; Ali Gharavi, Krzysztof Kiryluk, and Simone Sanna-Cherchi), Genomic Translation for ALS Care (GTAC; Matthew Harms), Pulmonary Fibrosis (Christine Garcia), Atopic disease (Joshua Milner), and Prenatal disease (Ronald Wapner). Only the contact investigator or investigators for each project are listed.

References

1 Tsakiri KD, Cronkhite JT, Kuan PJ, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci USA* 2007; **104:** 7552–57.

2 Stuart BD, Choi J, Zaidi S, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet* 2015; **47:** 512–17.

3 Armanios MY, Chen JJ, Cogan JD, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 2007; **356:** 1317–26.

4 Duckworth A, Gibbons MA, Allen RJ, et al. Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a mendelian randomisation study. *Lancet Respir Med* 2021; **9:** 285–94.

5 Zhang D, Newton CA, Wang B, et al. Utility of whole genome sequencing in assessing risk and clinically relevant outcomes for pulmonary fibrosis. *Eur Respir J* 2022; **60:** 2200577.

6 Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2020; **201:** 564–74.

7 Zhang D, Povysil G, Kobeissy PH, et al. Rare and common variants in *KIF15* contribute to genetic risk of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2022; **206:** 56–69.

8 Li C, Stoma S, Lotta LA, et al. Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *Am J Hum Genet* 2020; **106:** 389–404.

9 Codd V, Wang Q, Allara E, et al. Polygenic basis and biomedical consequences of telomere length variation. *Nat Genet* 2021; **53:** 1425–33.

10 Dressen A, Abbas AR, Cabanski C, et al. Analysis of protein-altering variants in telomerase genes and their association with *MUC5B* common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. *Lancet Respir Med* 2018; **6:** 603–14.

11 Stuart BD, Lee JS, Kozlitina J, et al. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. *Lancet Respir Med* 2014; **2:** 557–65.

12 Adegunsoye A, Newton CA, Oldham JM, et al. Telomere length associates with chronological age and mortality across racially diverse pulmonary fibrosis cohorts. *Nat Commun* 2023; **14:** 1489.

13 Newton CA, Zhang D, Oldham JM, et al. Telomere length and use of immunosuppressive medications in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; **200:** 336–47.

14 Zhang D, Adegunsoye A, Oldham JM, et al. Telomere length and immunosuppression in non-idiopathic pulmonary fibrosis interstitial lung disease. *Eur Respir J* 2023; **62:** 2300441.

15 Alonso-Gonzalez A, Jaspez D, Lorenzo-Salazar JM, et al. Rare variants and survival of patients with idiopathic pulmonary fibrosis. *Lancet Respir Med* 2025; **13:** 495–504.

16 Peljto AL, Blumhagen RZ, Walts AD, et al. Idiopathic pulmonary fibrosis is associated with common genetic variants and limited rare variants. *Am J Respir Crit Care Med* 2023; **207:** 1194–202.

17 Leavy OC, Ma SF, Molyneaux PL, et al. Proportion of idiopathic pulmonary fibrosis risk explained by known common genetic loci in European populations. *Am J Respir Crit Care Med* 2021; **203:** 775–78.

18 Taliun D, Harris DN, Kessler MD, et al, and the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; **590:** 290–99.

19 Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562:** 203–09.

20 Moll M, Peljto AL, Kim JS, et al. A polygenic risk score for idiopathic pulmonary fibrosis and interstitial lung abnormalities. *Am J Respir Crit Care Med* 2023; **208:** 791–801.

21 Cronkhite JT, Xing C, Raghu G, et al. Telomere shortening in familial and sporadic pulmonary fibrosis. *Am J Respir Crit Care Med* 2008; **178:** 729–37.

22 Codd V, Denniff M, Swinfield C, et al. Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nat Aging* 2022; **2:** 170–79.

23 Bulik-Sullivan B, Finucane HK, Anttila V, et al, and the ReproGen Consortium, and the Psychiatric Genomics Consortium, and the Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47:** 1236–41.

24 Auton A, Brooks LD, Durbin RM, et al, and the 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526:** 68–74.

25 Raghu G, Chen SY, Hou Q, Yeh WS, Collard HR. Incidence and prevalence of idiopathic pulmonary fibrosis in US adults 18–64 years old. *Eur Respir J* 2016; **48**: 179–86.

26 Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D. Incidence and prevalence of idiopathic pulmonary fibrosis: review of the literature. *Eur Respir Rev* 2012; **21:** 355–61.

27 Saunders GRB, Wang X, Chen F, et al, and the 23andMe Research Team, and the Biobank Japan Project. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* 2022; **612:** 720–24.

28 An J, Gharahkhani P, Law MH, et al, and the BEACON, and the 23andMe Research Team. Gastroesophageal reflux GWAS identifies risk loci that also associate with subsequent severe esophageal diseases. *Nat Commun* 2019; **10:** 4219.

29 Barton AR, Sherman MA, Mukamel RE, Loh PR. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nat Genet* 2021; **53:** 1260–69.

30 Ishigaki K, Sakaue S, Terao C, et al, and the BioBank Japan Project. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat Genet* 2022; **54:** 1640–51.

31 Shrine N, Izquierdo AG, Chen J, et al, and the China Kadoorie Biobank Collaborative Group, and the Qatar Genome Program Research (QGPR) Consortium. Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nat Genet* 2023; **55:** 410–22.

32 Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466:** 707–13.

33 Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows *BCL11A* associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci USA* 2008; **105:** 1620–25.

34 Liu L, Khan A, Sanchez-Rodriguez E, et al. Genetic regulation of serum IgA levels and susceptibility to common immune, infectious, kidney, and cardio-metabolic traits. *Nat Commun* 2022; **13:** 6859.

35 McGroder CF, Zhang D, Choudhury M, et al. Radiographic lung abnormalities in first-degree relatives of patients with different subtypes of pulmonary fibrosis. *Chest* 2023; **163:** 1471–75.

36 Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun* 2020; **11:** 3635.

37 Khan A, Shang N, Nestor JG, et al. Polygenic risk alters the penetrance of monogenic kidney disease. *Nat Commun* 2023; **14:** 8318.

38 Zhang D, Eckhardt CM, McGroder C, et al. Clinical impact of telomere length testing for interstitial lung disease. *Chest* 2024; **166:** 1071–81.

39 Noth I, Zhang Y, Ma SF, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013; **1:** 309–17.

40 Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; **45:** 613–20.

41 Mannherz W, Agarwal S. Thymidine nucleotide metabolism controls human telomere length. *Nat Genet* 2023; **55:** 568–80.