EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original Research Article

Protein biomarkers of interstitial lung abnormalities in relatives of patients with pulmonary fibrosis

Jonathan A. Rose, Mark P. Steele, Esteban J. Kosak Lopez, Gisli Thor Axelsson, Andrea G. Galecio Chao, Alan Waich, Katie Regan, Swati Gulati, Anthony H. Maeda, Sharmin Sultana, Claire Cutting, Ann-Marcia C. Tukpah, Andrew J. Synn, Mary B. Rice, Hilary J. Goldberg, Joyce S. Lee, David A. Lynch, Rachel K. Putman, Hiroto Hatabu, Benjamin A. Raby, David A. Schwartz, Ivan O. Rosas, Gary M. Hunninghake

Please cite this article as: Rose JA, Steele MP, Kosak Lopez EJ, *et al*. Protein biomarkers of interstitial lung abnormalities in relatives of patients with pulmonary fibrosis. *Eur Respir J* 2025; in press (https://doi.org/10.1183/13993003.01349-2024).

Protein Biomarkers of Interstitial Lung Abnormalities in Relatives of Patients with Pulmonary Fibrosis

Jonathan A. Rose MD, MS[1*]; Mark P. Steele MD[2*]; Esteban J. Kosak Lopez MD[1]; Gisli Thor Axelsson MD PhD[3]; Andrea G. Galecio Chao, MD[4]; Alan Waich MD[4]; Katie Regan, MS[5;] Swati Gulati MS[1]; Anthony H. Maeda MD[1]; Sharmin Sultana MD MPH[1]; Claire Cutting MD[1]; Ann-Marcia C. Tukpah MD MPH[1]; Andrew J. Synn, MD[6]; Mary B. Rice, MD, MPH[6]; Hilary J. Goldberg MD[1]; Joyce S. Lee[2]; David A. Lynch MD[7]; Rachel K. Putman, MD, MPH[1]; Hiroto Hatabu MD, PhD[8]; Benjamin A. Raby MD, MPH[1,5]; David A. Schwartz MD MPH[2†]; Ivan O. Rosas MD[4†]; Gary M. Hunninghake, MD, MPH[1†]

[1]Pulmonary and Critical Care Division, Brigham and Women's Hospital, Harvard Medical School, Boston MA, United States; [2]Department of Medicine, University of Colorado School of Medicine, Aurora, CO, United States; [3]Landspitali University Hospital, Division of Internal Medicine, Reykjavik, Iceland; [4]Pulmonary Critical Care and Sleep Medicine, Baylor College of Medicine, Houston, TX, United States; [5]Division of Pulmonary Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, United States; [6]Pulmonary, Critical Care & Sleep Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States; [7]Radiology, National Jewish Health, Denver, CO, United States; [8]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States;

*These authors contributed equally to this work.

[†]These authors contributed equally to this work.

Sources of Support: Dr. Rose is supported by National Institutes of Health (NIH) grant K08 HL173562. Dr. Cutting is supported T32 HL007633. Dr. Rice is supported by NIH grant R01 ES031252. Dr. Putman is supported by NIH grant K08 HL140087. Dr. Hatabu is supported by NIH grants R01 CA203636, U01CA209414, R01 HL111024, R01 HL135142, and R01 HL130974. Dr. Raby is supported by NIH grants R01 HL130974, R01 HL118455, P01 HL132825, U01 TR001810. Dr. Schwartz is supported through National Institutes of Health (NHLBI): P01-HL162607, R01-HL149836, R01-HL158668, and UG3/UH3 HL151865; VAMC: IO1BX005295; and DoD: W81XWH-17-1-0597. Dr. Rosas is supported by NIH grant U01 HL133232 and R01 HL130974. Dr. Hunninghake and this work were supported by NIH grants R01 HL111024, R01 HL130974, and R01 HL135142, and a generous donation from the Furman Family to our Fund for Pulmonary Fibrosis in the memory of Alfred Alves. This work was conducted with biostatistical support from UM1TR004408 award through Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health.

Conflicts of Interest: JA Rose, MP Steele, EJK Lopez, AGG Chao, A Waich, K Regan, S Gulati, A Maeda, S Sultana, C Cutting, A-MC Tukpah, AJ Synn, HJ Goldberg, DA Lynch, and RK Putman report no disclosures. GT Axelsson reports travel support for meetings from Boehringer Ingelheim. MB Rice reports lecture honoraria paid by USC, U. Utah, NYU, UNC, UVM, and Northwestern U, payment from Conservation Law Foundation for providing an

expert opinion, support for ATS registration in 2022 due to role as program committee chair, leadership of the ATS environmental health policy committee until May 2020, program committee chair 2022-2023 for the EOPH assembly of ATS, and chair-elect of the Environmental, Occupational and Population Health (EOPH) Assembly of ATS 2024. J Lee reports grants from Boehringer Ingelheim, consulting fees from Blade, Avalyn, Boehringer Ingelheim, United Therapeutics, Astra Zeneca, Elima, and Eleven P15, participation on a data safety monitoring board or advisory board for United Therapeutics (TETON trial) and Avalyn Pharma (ATLAS trial), acting as a Senior Medical Advisor for the Pulmonary Fibrosis Foundation, and receiving a research gift from Pliant Therapeutics. H Hatabu reports grants from Canon Medical Systems Inc. and Konica Minolta Inc. and consulting fees from Boehringer Ingelheim and Canon Medical Systems Inc. BA Raby reports royalties from UpToDate as an editor. DA Schwartz reports consulting fees from Vertex and being a founder and chief scientific officer of Eleven P15, Inc., a company focused on the early diagnosis and treatment of pulmonary fibrosis. IO Rosas reports grants from Boehringer Ingelheim, Genentech, and Roche and participation on the ad board for Boehringer Ingelheim, Avalyn Pharma, and Structure Therapeutics. GM Hunninghake reports consulting fees from Boehringer Ingelheim and Gerson Lehrman Group and lecture fees from Boehringer Ingelheim.

Corresponding Author:

Gary M. Hunninghake M.D., M.P.H.

Pulmonary and Critical Care Division

Department of Medicine

Brigham and Women's Hospital

75 Francis Street

Boston, MA 02115

Email: ghunninghake@bwh.harvard.edu

Phone: 617-525-9687

Fax: 617-278-6955

Author Contributions:

Study Design: J.A.R., B.A.R., D.A.S., I.O.R, G.M.H.

Acquisition, analysis, or interpretation of the data: J.A.R., G.T.A., E.J.K.L., A.G.G.C., A.R.W.C., K.R., S.G., A.H.M, S.S., C.C., A-M.C.T., A.J.S., M.B.R., H.J.G., J.S.L, D.A.L., R.K.P., H.H., M.P.S., B.A.R., D.A.S., I.O.R, G.M.H.

Critical revision of the manuscript for important intellectual content: J.A.R., G.T.A., E.J.K.L., A.G.G.C., A.R.W.C., K.R., S.G., A.H.M, S.S., C.C., A-M.C.T., A.J.S., M.B.R., H.J.G., J.S.L, D.A.L., R.K.P., H.H., M.P.S., B.A.R., D.A.S., I.O.R, G.M.H.

Statistical analysis: J.A.R., B.A.R., D.A.S., I.O.R, G.M.H.

Obtained funding: B.A.R., D.A.S., I.O.R, G.M.H.

Key Words: Interstitial lung disease, interstitial lung abnormalities, pulmonary fibrosis

Running Title: Protein Biomarkers of ILA in Relatives

Descriptor Number: 9.23 Interstitial Lung Disease

Word Counts

Title: 90 characters

Abstract: 236

Manuscript Text: 2769

Tables: 3

Figures: 3

Supplemental Tables: 7

Supplemental Figures: 3

Take Home Message:

Plasma proteins are associated with ILA in first-degree relatives of patients with pulmonary fibrosis not known to have disease, and when combined with basic demographics can accurately detect ILA. This work has implications for possible screening in this high-risk population and suggests targets warranting further investigation.

Ethical Considerations

Local IRB, and individual cohort approvals were obtained for all of the proposed analyses. Research participants gave informed consent to have data collected including peripheral blood samples and chest CT scans. The main risk of participation in this study was the social-psychological risk resulting from disclosure of medical history or genetic information. Data security and confidentiality were strictly maintained. All data was stored on secured computer servers that are kept in a locked room with key card access to authorized personnel only. Databases, and server access to these databases were password protected, and periodic checks of activity logs were run to guard against hacking. Blood samples were used by the investigators for research purposes only; the results were kept strictly confidential.

Abstract

Rationale: First-degree relatives of patients with pulmonary fibrosis (relatives) are at high risk for interstitial lung abnormalities (ILA), highlighting the need for biomarkers for risk prediction. We aimed to identify blood proteins associated with and predictive of ILA among relatives of patients with pulmonary fibrosis.

Methods: Relatives enrolled in two independent cohorts had protein levels measured using an aptamer-based proteomic platform. ILA was assessed with CT scans per Fleischner Society recommendations. Protein associations with ILA were assessed using regression, and significant proteins were used with clinical variables to detect ILA.

Results: Of 237 relatives from two independent cohorts, 26% had ILA. Seven proteins were associated with ILA in the discovery cohort after FDR-adjustment, and all remained significant after adjusting for age, gender, and smoking status. Six of seven were significant in the validation cohort including GDF15, SFTPD, and SFTPB. In a multivariable model, six proteins combined with basic demographics in the discovery cohort had AUC=0.92 (0.88 in validation cohort). LASSO modelling identified three proteins and age as predictors with an AUC=0.89. When applied to the combined cohorts, this simple model would reduce the need for CT imaging in one of every three relatives screened.

Conclusion: Peripheral blood proteins are associated with ILA in relatives of patients with pulmonary fibrosis and can be used to detect ILA. Our findings demonstrate the potential utility of blood biomarkers in this high-risk group and suggests molecular targets for future investigation.

Introduction

Interstitial lung abnormalities (ILA) are abnormal findings on chest computed tomography (CT) scans that may represent an early form of interstitial lung disease (ILD) or pulmonary fibrosis. First-degree relatives of patients with pulmonary fibrosis (relatives), whether familial or sporadic, are at high risk of developing ILA and ILD (1-3). Additionally, relatives with ILA have increased rates of 2-year imaging progression, accelerated lung function decline (4), and increased mortality (5). Based on the frequency of ILA, ILD, and adverse outcomes, biomarkers could improve early detection in this high-risk population and suggest key biological pathways for future investigation.

While protein associations of ILA have been performed in studies of the general population (6, 7), there is no prior assessment of large-scale proteomics in relatives of patients with pulmonary fibrosis in whom screening evaluations are arguably the most relevant. To achieve this objective, we performed proteomic assessments in two cohorts of relatives, using the Clinical Genetics and Screening for Pulmonary Fibrosis (CGS-PF) to discover proteins associated with ILA and a similarly ascertained cohort of relatives of patients with sporadic idiopathic pulmonary fibrosis (IPF) from Colorado to validate significant results.

Methods

Study Design and Population

CGS-PF

This study included participants enrolled in CGS-PF; protocols for this study has been previously described (2). Briefly, the CGS-PF is a longitudinal study of first-degree relatives of patients with sporadic or familial pulmonary fibrosis between 48 and 85 years of age without a known diagnosis of ILD seen at Brigham and Women's Hospital (BWH) in Boston, MA and

Baylor College of Medicine (BCM) in Houston, TX; CGS-PF enrollment is ongoing. As part of this protocol, relatives had CT scans, pulmonary function tests (PFTs), blood sample collection, and physical exam. Genotyping of the *MUC5B* promoter polymorphism (rs35705950) was measured and henceforth *MUC5B* refers to having at least one copy of the minor allele. Telomere lengths were obtained from 6-panel flow fish (Repeat Diagnostics) and henceforth low lymphocyte telomere length refers to having lymphocyte telomere length less than the 10th age-adjusted percentile. Those relatives with baseline plasma samples available as of December 2022 were included in the study, which was approved by the BWH and BCM institutional review boards.

Colorado Cohort

First-degree relatives of patients with sporadic idiopathic pulmonary fibrosis over the age of 40 years old without known cardiac or pulmonary disease were enrolled at the University of Colorado and National Jewish Health; participants had chest CT scans, blood sample collection, and completed questionnaires. This study was approved by the IRBs at the University of Colorado and National Jewish Health.

ILA Phenotyping

According to previously published protocols, participants in CGS-PF underwent prone volumetric thoracic chest CT at full inspiration. Chest CT scans were visually assessed for ILA by up to three readers (8), with ILA identified per Fleischner Society recommendations (ILA is used here to define relatives not previously suspected of having ILD) (9). For the Colorado cohort, chest CT scans were performed at local community centers using similar protocols for image acquisition and reconstruction and were interpreted by study radiologists as previously described (10).

Protein Measurements

Proteins were measured from plasma samples stored in ethylenediaminetetraacetic acid at -80°C using the SomaScan v4.1 proteomic platform with 7,596 DNA Slow-Off rate Modified Aaptamers (SOMAmers) targeting proteins from 6,411 unique genes. Protein measurements were standardized as previously described (6) using hybridization controls to mitigate within-plate variation, median signal normalization to mitigate replicate variability, and plate scaling and calibration of SOMAmers to control for inter-assay variation between analytes and batch differences between runs. Individual protein measures for each cohort were transformed separately by center-scale normalizing the base-10 logarithm transformed data..

Statistical analysis

To address sample size limitations, we adopted a discovery/validation approach, whereby suggestive protein-ILA associations were first identified in CGS-PF using univariable logistic regression. Proteins with FDR-adjusted p value <0.05 were assessed in multivariable analyses adjusting for age, gender, and smoking history to identify significant proteins in the CGS-PF cohort, and separately models were adjusted for *MUC5B* and low lymphocyte telomere length. Significant proteins were then independently validated using univariable logistic regression in the Colorado cohort. To identify clinically meaningful biomarkers that can be applied generally, ILA was treated as a binary variable with indeterminate and no ILA combined into the "without ILA" category. Separately, proteomic data from the CGS-PF and Colorado cohorts were combined, and a replication analysis was performed in the 245 of the 287 significant proteins from the prior study of ILA in AGES-Reykjavik (Age, Gene/Environment Susceptibility–Reykjavik) study (6), that were also measured in these two cohorts. In the combined cohorts, correlations between significant proteins were assessed with Pearson

correlations, and effect modification of demographics on validated ILA-protein associations in the combined cohorts was assessed using multivariable regression models with demographic-protein multiplicative interaction terms along with main effects.

Receiver operating characteristic curves (ROCs) were used to test the accuracy of clinical and multi-protein models in detecting ILA. The DeLong test was used to compare the area under the ROC curves (AUCs) for nested models; non-nested models were compared with the Venkatraman method. To simplify the model for ILA detection, LASSO (Least Absolute Shrinkage and Selection Operator) modeling was used for variable selection using the discovery cohort (CGS-PF) for training and the validation cohort (Colorado) for testing. Validated proteins, age, gender, and smoking status were included as predictors, and 200 bootstrapped samples were used with replacement. The LASSO model was fitted using the glmnet package in R with 10-fold cross-validation. The best lambda was selected based on the one-standard-error rule (11). Variables selected from the LASSO model were used in a logistic regression model to predict ILA in the discovery cohort, coefficients were applied to the validation cohort, and ROC and AUCs were calculated. Test characteristics from the validation ROC were applied to the combined cohorts to assess hypothetical screening parameters. False discovery rate (FDR) was used for multiple hypothesis testing when appropriate, and p values <0.05 were considered significant. All analyses were performed using Statistical Analysis Software version 9.4 (SAS Institute, Cary, NC) and R (Version 4.0.3).

Results

In CGS-PF, 49 of the 156 relatives (31%) had ILA at baseline and in the Colorado cohort, 13 (16%) of 81 relatives had ILA. Baseline characteristics are shown in **Table 1**. Similar to prior studies, those with ILA were of older age; in the CGS-PF cohort, there was a male

predominance, increased frequency of the *MUC5B* variant, reduced lymphocyte telomere length, and lower FVC and DLCO percent predicted.

Discovery and Validation of Protein Associations with ILA in Relatives

In univariate analyses in the CGS-PF (discovery) cohort, 558 proteins were associated with ILA at $p<0.05$ (**Figure 1** and **Table E1**), which is a greater number than expected by chance (320, or 5% of 6,411 proteins from unique genes). Of these proteins, seven were significant after FDR-adjustment, including growth differentiation factor 15 (GDF15), surfactant protein D (SFTPD), surfactant protein B (SFTPB), WAP four disulfide core domain protein 2 (WFDC2), CUB domain-containing protein 1 (CDCP1), GTP cyclohydrolase 1 feedback regulatory protein (GCHFR), and stratifin (SFN). These seven proteins remained significant in multivariable analyses after adjusting for age, gender, and smoking status (**Table 2**) and when additionally adjusted for *MUC5B* and low lymphocyte telomere length (**Table E2**). In the Colorado validation cohort, six of the seven proteins (GDF15, SFTPD, SFTPB, WFDC2, CDCP1, and SFN, but not GCHFR) were significant with similar effect estimates (**Table 2**). Notably, five of these proteins (GDF15, SFTPB, WFDC2, SFN, and CDCP1) were previously associated with ILA in the AGES-Rekjavik study (6). Assessment of demographics effects in the combined CGS-PF/Colorado cohorts are shown in **Tables E3** and **E4.** Correlation analyses between the six significant proteins revealed one strong correlation (GDF15 and WFDC2, $R^2=0.78$, full results in **Table E5**).

Replication of Prior Protein Associations with ILA in Combined Cohort of Relatives

The AGES-Reykjavik study previously reported proteins associated with ILA (6). As these blood biomarkers were identified with similar ILA phenotyping and protein measurements, we sought to replicate these findings in our cohorts of relatives. We assessed the 245 significant

proteins from the prior study that were also measured in our cohorts by analyzing the association with the combined CGS-PF and Colorado cohorts. We found that 42 of the 245 ILA-associated proteins in the AGES-Reykjavik study showed associations with ILA of similar magnitude and direction of association in the combined CGS-PF/Colorado relatives cohort analysis at p-value <0.05. Sixteen were significantly associated after FDR-adjustment (**Table 3**). Of the 32 proteins with an effect estimate of 50% in the AGES-Reykjavik study (i.e. odds ratio >1.5 or <0.67), all had similar effect estimates in the same direction in the combined cohorts of relatives, and 19 (59%) had p<0.05.

Protein Biomarker Assessment for the Detection of ILA

Having demonstrated significant and replicable associations of GDF15, SFTPD, SFTPB, WFDC2, CDCP1, and SFN with ILA, we hypothesized that the accuracy of multivariable models for ILA detection would be strengthened by the inclusion of these candidate biomarkers. In the CGS-PF cohort, we found that a multivariable model comprised of these six significant proteins (AUC 0.90) significantly outperformed models based solely on demographics (age, gender, smoking status; AUC 0.75; p=0.003) (ROCs shown in **Figure 2**). An integrated model that combined the six proteins and demographics had an AUC of 0.92. Additional models including *MUC5B*, low lymphocyte telomere length, and PFTs (only measured in CGS-PF) are shown in **Figure E1-E3**. When applied to the validation cohort, the multiprotein model had an AUC of 0.84, and the final model of six proteins and demographic data had an AUC of 0.88. The basic demographics model (AUC 0.80) was not statistically significantly different than the full model (p = 0.2).

Next, a machine learning computational approach was used for variable selection to simplify the proposed model. The algorithm selected three of the six proteins (GDF15, SFTPD, and

SFTPB) and age. The AUC of this simplified model in the validation cohort was 0.89 with a Youden's index corresponding to 77% sensitivity and 90% specificity (ROC curves are shown in **Figure 3;** test characteristics for local maxima in **Table E6**). As a way of demonstrating how this model could be hypothetically applied toward clinical screening, the model threshold for 100% sensitivity (also 100% negative predictive value) was applied to the combined cohorts with an ILA prevalence of 26%, which gave 107 true negatives for the total 237 participants tested. Therefore, if validated in an independent platform, implementing this model in a similar population for screening purposes would reduce the need for CT imaging in one of every three relatives screened (237/107 = 2.2; full 2x2 table displayed in **Figure E7**).

Discussion

This study is the first large-scale proteomic assessment of ILA in first-degree relatives of patients with pulmonary fibrosis, a population at high risk of developing ILA and ILD, and validates many of the findings of the one prior study of proteomics in ILA. Seven proteins were found to be significantly associated with ILA after adjusting for common risk factors and six of these were validated in an independent cohort. The consistency of these ILA-protein associations across multiple studies suggests that they may represent surrogate markers of lung injury and remodeling that occur early in the course of pulmonary fibrosis development. Importantly this study provides evidence for using a simple multiprotein model that can accurately detect ILA, which could have a role in screening at-risk individuals for pulmonary fibrosis.

Although proteins have been assessed in first-degree relatives previously using mass spectrometry (12), this is the first large-scale proteomic study to reveal multiple validated associations in two independent cohorts. Five proteins (GDF15, SFTPB, WFDC2, SFN, and

CDCP1) identified in our study were previously reported in a general population study of ILA in the AGES-Reykjavik cohort (6). GDF15, measured by immunoassays, was also shown to be associated with ILA in the Framingham Heart Study and COPDGene cohorts (7). Furthermore, GDF15, WFDC2, and CDCP1 were associated with quantitative measures of interstitial abnormalities in the COPDGene (13, 14) and CARDIA cohorts (13). This demonstrates that many protein-ILA associations are similar between relatives and ILA in the general population (presumed to be sporadic). SFTPD is the only protein associated with ILA in our study that was not significant in AGES-Reykjavik. Interestingly, SFTPD levels are associated with the presence of IPF and IPF survival in previous studies (15-17), and therefore it is unclear if the difference in SFTPD association between this study and AGES could be attributed to differences in populations or severity. However, the overall consistency with which these proteins are associated with ILA across diverse cohorts independent of methods for protein detection supports their accuracy and generalizability as candidate biomarkers.

Many of the proteins associated with ILA are enriched in lung tissue and known to be involved in the pathobiology of pulmonary fibrosis. GDF15 is a cytokine member of the TGF-β family of proteins and is elevated under stress conditions and with increasing age (18, 19). It has been implicated in various human diseases (20) and is felt to play a role in the pathogenesis of pulmonary fibrosis (21-23). SFTPD and SFTPB are essential pulmonary surfactant components (24, 25); elevated levels have been described in many respiratory diseases and are thought to represent alveolar injury and repair (26-29). SFTPB levels are increased in IPF (30, 31) and associated with disease severity in pulmonary fibrosis (32). WFDC2 is a secreted protein of the respiratory epithelium; elevated levels have been associated with IPF (33) and lung cancer (34). CDCP1 is a transmembrane protein that potentiates TGF-β activation of interstitial fibroblasts (35). Expression of SFN, a potent negative regulator of the cell cycle, is

increased in IPF airway basal cells and fibroblasts (36, 37). In total, these findings provide evidence that pathobiological processes present in established pulmonary fibrosis can be detected in subjects with ILA who are at risk of developing ILD. Although rates of progression are high in this cohort, further work is needed to understand if these proteins are associated with high-risk forms of ILA that progress and experience poor outcomes.

Our results demonstrate that peripheral blood proteins can assist in detecting ILA and may have value in screening at-risk individuals. Currently, there is no consensus on screening all relatives of patients with PF (38), but if consistent and reproducible assays can be developed with high negative predictive value, blood protein biomarkers could aid in determining who should undergo further testing (i.e. chest CT scan) and therefore could reduce some of the barriers to screening (unnecessary radiation and cost associated with CT scans). Future efforts to identify the best-fitting models for the prediction of pulmonary fibrosis are warranted across unique populations, particularly for those at high risk for poor outcomes (39) or with progressive disease (4).

There are several limitations to this study. First, although we demonstrated evidence for significant and replicable associations across populations, the small sample size may have limited our ability to detect additional significant associations. Second, pulmonary function assessments were unavailable in the validation cohort, precluding further work on these analyses. Third, significant variability in detection and accuracy across different platforms limits the interpretation across pulmonary fibrosis biomarker studies (40). While larger SomaScan platforms benefit from increased precision and a greater representation across the proteome, it seems to be at the expense of specificity (41). Although our results are consistent with prior studies and provide confidence in these associations, additional confirmation with

orthogonal assays and across the spectrum of disease severity (42) would provide further validity. Furthermore, the semiquantitative nature of Somascan makes clinical implementation challenging, specifically in identifying threshold cut-offs for individual proteins. Further biomarker development for clinical use would require the development of a quantitative, highly reproducible, low-cost method for protein detection. Fourth, some of these proteins may not be specific to pulmonary fibrosis; further studies in larger cohorts may help to validate these potential correlations independently. Fifth, it is unknown from these blood-based measurements if proteins are related to underlying disease processes, non-specific markers of injury, or are related to unknown comorbidities; more work needs to be done to understand their biologic relevance. Sixth, the lack of racial/ethnic diversity in the cohorts limits the generalizability of these findings.

In conclusion, this study identifies proteins associated with interstitial lung abnormalities in first-degree relatives of patients with pulmonary fibrosis, including GDF15, SFTPD, and SFTPB and validates many of the protein-ILA associations previously shown in the general population. When combined with basic demographic features, these proteins can improve detection of ILA. Given the high risk for ILA among relatives, this work suggests the potential for peripheral blood biomarkers to be used in clinical screening and to assist in the understanding of the pathobiologic processes that contribute to the development of pulmonary fibrosis.

Tables and Figures:

Table 1: Baseline Characteristics for ILA and Without ILA in Discovery and Validation Cohorts

Variable	CGS-PF			Colorado Cohort		
	Without ILA (N=107, 69%)	**ILA** (N=49, 31%)	**P-value**	**Without ILA** (N=68, 84%)	**ILA** (N=13, 16%)	**P-value**
Age—year, mean (SD)*	58.6 (7.65)	65.3 (8.69)	<0.001	55.2 (11.6)	68.7 (8.77)	<0.001
Male gender—N (%)	36 (33.6%)	28 (57.1%)	0.009	23 (33.8%)	8 (61.5%)	0.1
Non-hispanic white race/ethnicity—N (%)†	102 (95.3%)	49 (100%)	0.3	65 (100%)	13 (100%)	1.0
Ever smoker—N (%)	36 (33.6%)	22 (44.9%)	0.2	26 (38.2%)	3 (23.1%)	0.5
BMI—kg/m2, mean (SD)	27.5 (5.6)	27.4 (4.4)	0.9	-	-	-
Relative with FPF—N (%)	55 (51.4%)	22 (44.9%)	0.6	-	-	-
Lymphocyte telomere length <10th %ile—N (%)‡	20 (19.2%)	22 (45.8%)	<0.001	-	-	-
MUC5B promoter variant—N (%)§	38 (36.2%)	29 (60.4%)	0.008	-	-	-
Monocyte count—K/μL, mean (SD)‖	0.53 (0.18)	0.56 (0.20)	0.4	-	-	-
FVC %predicted—mean (SD)	106.2 (14.5)	99.9 (17.2)	0.02	-	-	-
DLCO %predicted—mean (SD)	90.9 (17.2)	76.6 (18.6)	<0.001	-	-	-

Definition of abbreviations: ILA = interstitial lung abnormalities; CGS-PF = Clinical Genetics and Screening for Pulmonary Fibrosis; N = number; SD = standard deviation; BMI = body mass index; FPF = familial pulmonary fibrosis; FVC = forced vital capacity; DLCO = diffusion capacity of carbon monoxide.

Comparison of categorical variables was done using Fisher exact tests and continuous variables with two-tailed Student's t test.

*Missing age data for 1 participant in the Colorado Cohort

†Missing race data for 4 participants in CGS-PF

‡Missing telomere length data for 4 participants in CGS-PF

§Missing *MUC5B* data for 3 participants in CGS-PF

‖Missing monocyte count data for 8 participants in CGS-PF

Figure 1: Discovery of single protein associations with interstitial lung abnormalities (ILA) using univariable logistic regression models. GDF15 = growth differentiation factor 15; SFTPD = surfactant protein D; SFTPB = surfactant protein B; WFDC2=WAP four-disulfide core domain protein 2.

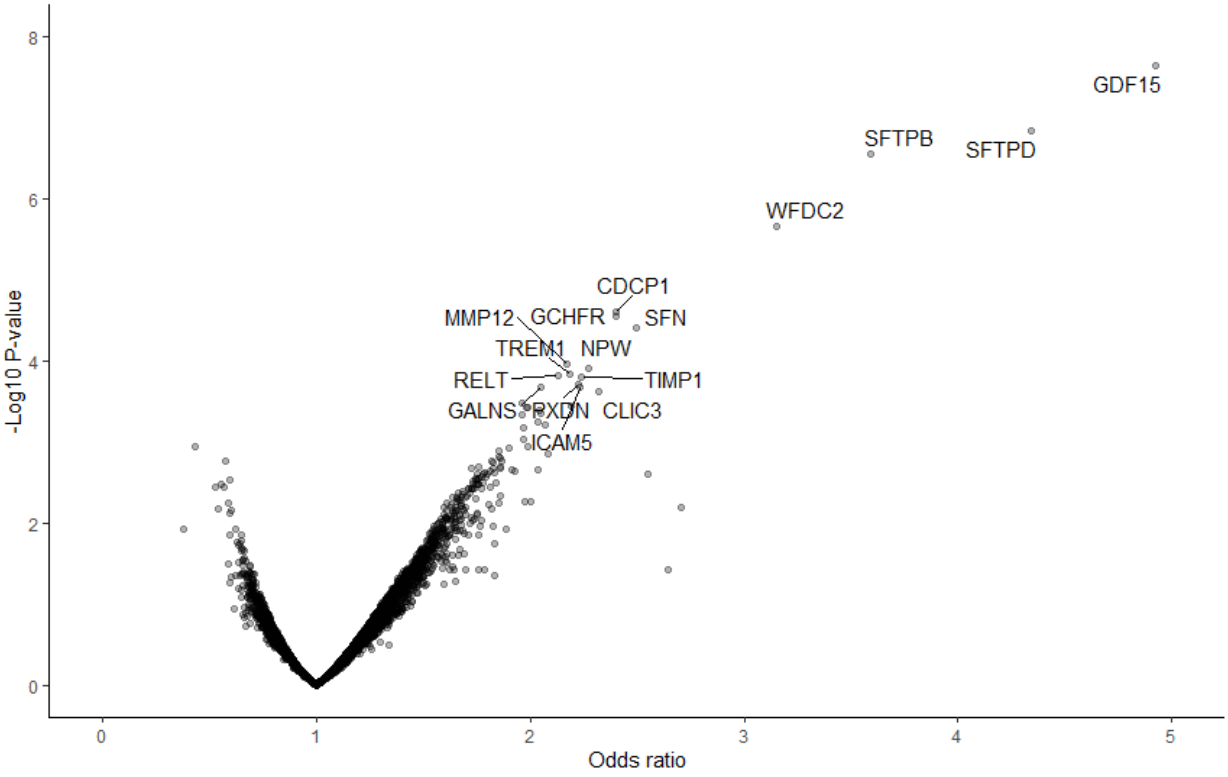


Table 2: Significant Protein Associations with ILA in Discovery and Validation Cohorts

	CGS-PF			Colorado Cohort		
Protein	**Odds Ratio**	**Confidence Interval**	**P value**	**Odds Ratio**	**Confidence Interval**	**P value**
SFTPD	4.16	2.30-7.50	2.3E-06	3.73	1.75-7.96	0.0007
GDF15	4.08	2.28-7.30	2.3E-06	3.36	1.56-7.23	0.002
SFTPB	4.00	2.29-7.00	1.1E-06	2.91	1.48-5.69	0.002
WFDC2	2.59	1.60-4.19	9.9E-05	3.50	1.72-7.14	0.0006
GCHFR	2.45	1.55-3.87	0.0001	1.32	0.72-2.40	0.4
SFN	2.26	1.42-3.58	0.0006	3.24	1.54-6.80	0.002
CDCP1	2.14	1.37-3.33	0.0008	3.03	1.50-6.14	0.002

Definition of abbreviations: CGS-PF = Clinical Genetics and Screening for Pulmonary Fibrosis; GDF15 = growth differentiation factor 15; SFTPD = surfactant protein D; SFTPB = surfactant protein B; WFDC2 = WAP four-disulfide core domain protein 2; CDCP1 = CUB domain-containing protein 1; GCHFR = GTP cyclohydrolase 1 feedback regulatory protein; SFN = stratifin.

Analyses for CGS-PF used multivariable logistic regression models with covariates age, gender, and smoking status. Analyses for Colorado Cohort used univariable logistic regression models.

Table 3: Replication of Significant Protein Associations with ILA from the AGES-Reykjavik Study in the Combined Cohorts of Relatives

	AGES-Reykjavik			Combined Cohort of Relatives		
Protein	**Odds Ratio**	**Confidence Interval**	**P value**	**Odds Ratio**	**Confidence Interval**	**P value**
SFTPB	3.71	3.20-4.30	1.76×10-63	3.93	2.43-6.35	5.84×10-6
GDF15	2.06	1.80-2.36	1.43×10-21	3.09	1.96-4.89	1.54×10-4
WFDC2	2.42	2.11-2.78	1.65×10-32	2.54	1.72-3.75	2.33×10-4
MFAP5	1.70	1.51-1.91	9.57×10-15	2.32	1.52-3.54	5.95×10-3
CDCP1	1.62	1.44-1.83	6.24×10-12	2.09	1.43-3.06	6.65×10-3
ICAM5	1.62	1.44-1.84	6.38×10-11	2.03	1.40-2.95	7.85×10-3
SCGB3A1	2.43	2.13-2.77	3.29×10-36	2.02	1.39-2.94	8.81×10-3
CCL22	1.43	1.26-1.62	2.12×10-4	1.98	1.36-2.89	0.01
CCL17	1.45	1.28-1.63	6.76×10-6	1.88	1.33-2.66	0.01
CXCL11	1.48	1.31-1.67	1.30×10-6	1.87	1.32-2.65	0.01
TREM1	1.50	1.32-1.71	3.55×10-6	1.89	1.32-2.71	0.01
PLAT	1.37	1.21-1.55	1.86×10-3	1.98	1.30-3.01	0.03
CBLN4	1.35	1.19-1.53	7.04×10-3	1.81	1.25-2.65	0.04
CCL19	1.47	1.31-1.65	3.14×10-7	1.63	1.19-2.22	0.04
TNFSF13B	1.41	1.25-1.58	6.09×10-5	1.68	1.20-2.36	0.04
CCL18	1.78	1.58-2.02	1.70×10-16	1.75	1.21-2.53	0.048

Definition of abbreviations: AGES-Reykjavik = Age, Gene/Environment Susceptibility–Reykjavik; GDF15 = growth differentiation factor 15; SFTPB = surfactant protein B; WFDC2 = WAP four-disulfide core domain protein 2; CDCP1 = CUB domain-containing protein 1; TREM1 = triggering receptor expressed on myeloid cells 1; PLAT = plasminogen activator, tissue type; TIMP1 = TIMP metallopeptidase inhibitor 1; B2M = beta-2-microglobulin; CCL18 = C-C motif chemokine ligand 18; IL15RA = interleukin 15 receptor subunit alpha; COL28A1 = collagen type XXVIII alpha 1 chain.
Data for AGES-Reykjavik were published previously (6). Analyses for the combined cohorts of relatives used multivariable logistic regression models adjusted for age, gender and smoking with FDR-adjusted p-values shown.

Figure 2: Proteins and Demographics for the Detection of ILA. Receiver operating characteristic curves and area under the curves (AUC) for the detection of ILA. **A.** Logistic regression models were used in CGS-PF with clinical data and/or the six proteins significantly associated with ILA as the training set. The multiprotein model with the six validated proteins was a significant improvement over the model with demographics alone (p=0.003). The final model with demographics and the six proteins had an AUC of 0.92. **B.** The models established in CGS-PF were then tested in the independent Colorado cohort. The final model with demographics and the six proteins had an AUC of 0.88.

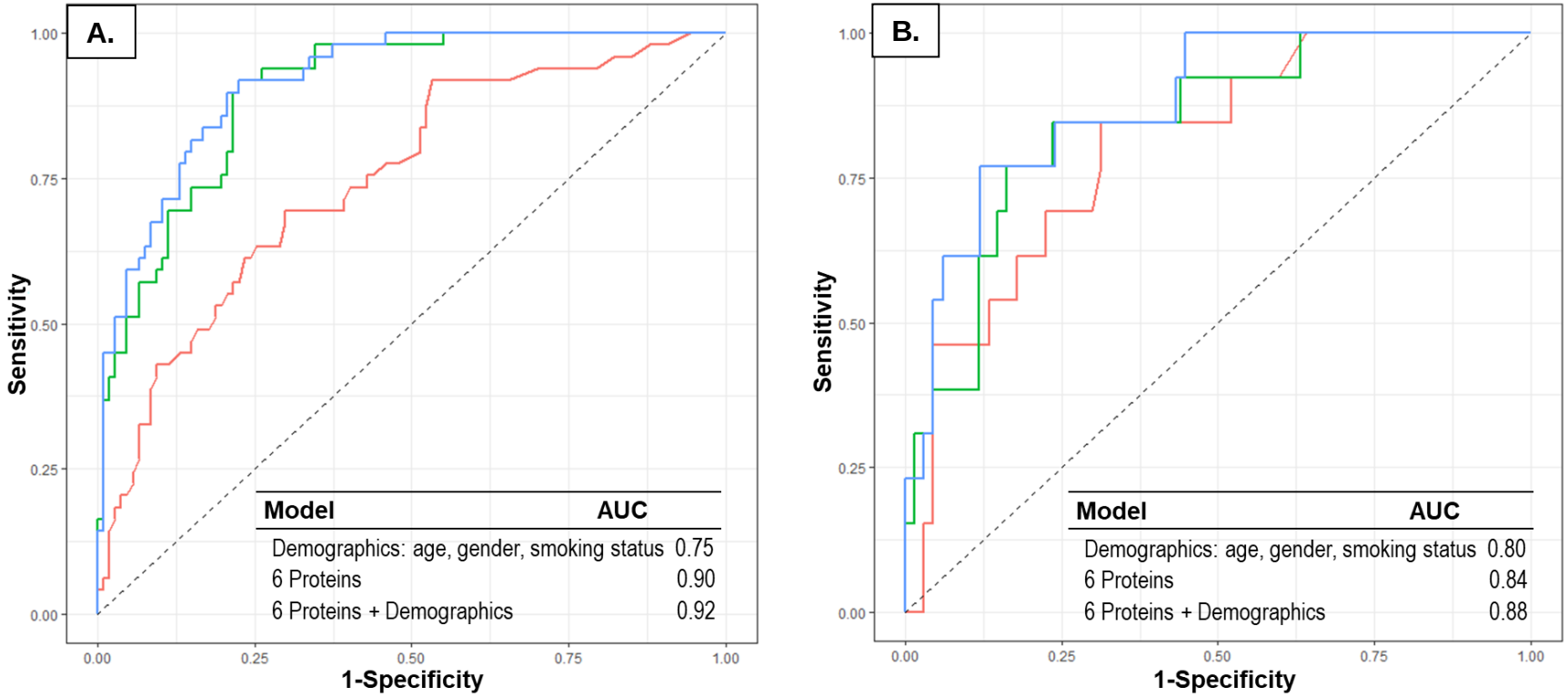
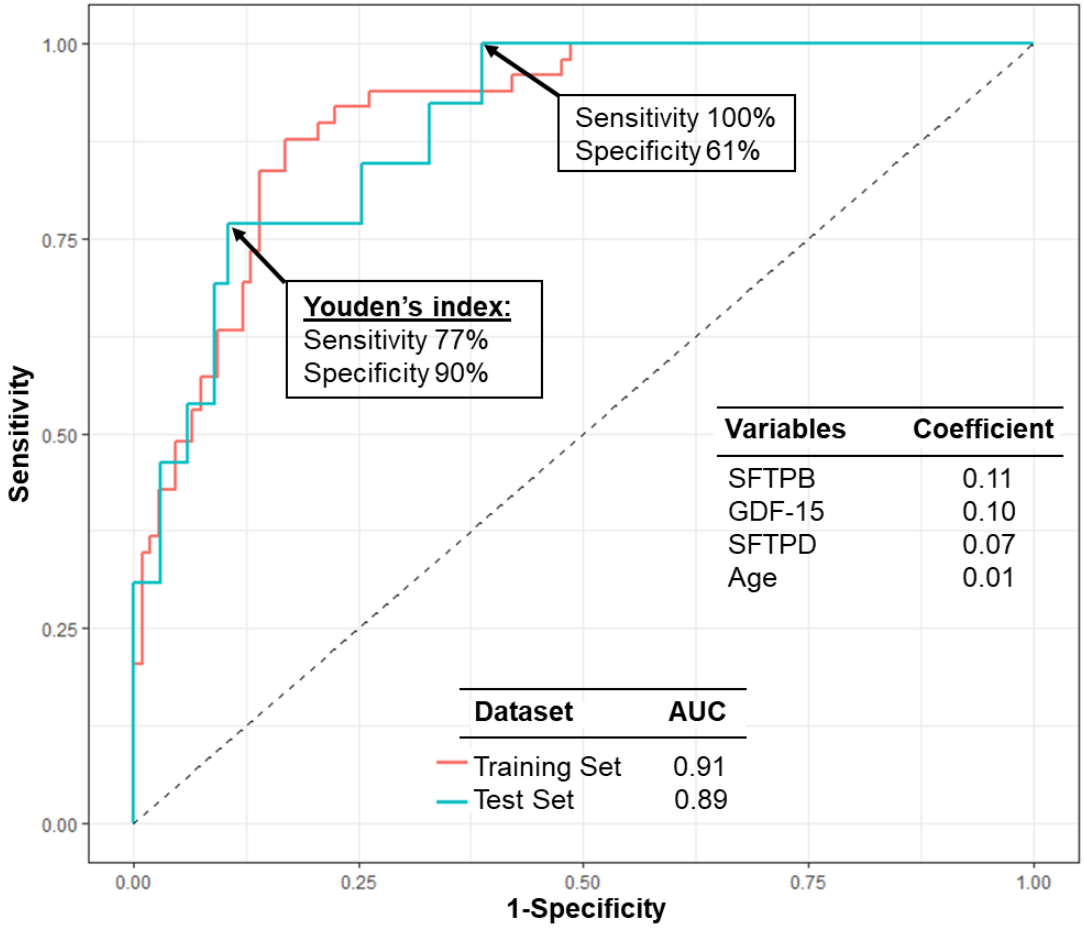


Figure 3: Simplified Model for the Detection of ILA. Receiver operating characteristic (ROC) curves and area under the curves (AUC) of the simplified model derived from machine learning for the detection of ILA. Variable selection using LASSO modelling of the discovery cohort gave four variables shown in the table inset with coefficients. These four variables were used in multivariable logistic regression for the detection of ILA in the discovery cohort (red curve). This model was then applied to the validation cohort to generate an estimate of ILA detection (blue curve) and gave an AUC of 0.89. A table of local maximas of the ROC curve from the validation cohort and corresponding test performance characteristics in the combined cohorts are shown in **Table E6**.



References:

1. Rosas IO, Ren P, Avila NA, Chow CK, Franks TJ, Travis WD, McCoy JP, Jr., May RM, Wu HP, Nguyen DM, Arcos-Burgos M, MacDonald SD, Gochuico BR. Early interstitial lung disease in familial pulmonary fibrosis. *Am J Respir Crit Care Med* 2007; 176: 698-705.
2. Hunninghake GM, Quesada-Arias LD, Carmichael NE, Martinez Manzano JM, Poli De Frias S, Baumgartner MA, DiGianni L, Gampala-Sagar SN, Leone DA, Gulati S, El-Chemaly S, Goldberg HJ, Putman RK, Hatabu H, Raby BA, Rosas IO. Interstitial Lung Disease in Relatives of Patients with Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020; 201: 1240-1248.
3. Salisbury ML, Hewlett JC, Ding G, Markin CR, Douglas K, Mason W, Guttentag A, Phillips JA, 3rd, Cogan JD, Reiss S, Mitchell DB, Wu P, Young LR, Lancaster LH, Loyd JE, Humphries SM, Lynch DA, Kropski JA, Blackwell TS. Development and Progression of Radiologic Abnormalities in Individuals at Risk for Familial Interstitial Lung Disease. *Am J Respir Crit Care Med* 2020; 201: 1230-1239.
4. Rose JA, Planchart Ferretto MA, Maeda AH, Perez Garcia MF, Carmichael NE, Gulati S, Rice MB, Goldberg HJ, Putman RK, Hatabu H, Raby BA, Rosas IO, Hunninghake GM. Progressive Interstitial Lung Disease in Relatives of Patients with Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2023; 207: 211-214.
5. Steele MP, Peljto AL, Mathai SK, Humphries S, Bang TJ, Oh A, Teague S, Cicchetti G, Sigakis C, Kropski JA, Loyd JE, Blackwell TS, Brown KK, Schwarz MI, Warren RA, Powers J, Walts AD, Markin C, Fingerlin TE, Yang IV, Lynch DA, Lee JS, Schwartz DA. Incidence and Progression of Fibrotic Lung Disease in an At-Risk Cohort. *Am J Respir Crit Care Med* 2023; 207: 587-593.
6. Axelsson GT, Gudmundsson G, Pratte KA, Aspelund T, Putman RK, Sanders JL, Gudmundsson EF, Hatabu H, Gudmundsdottir V, Gudjonsson A, Hino T, Hida T, Hobbs BD, Cho MH, Silverman EK, Bowler RP, Launer LJ, Jennings LL, Hunninghake GM, Emilsson V, Gudnason V. The Proteomic Profile of Interstitial Lung Abnormalities. *Am J Respir Crit Care Med* 2022; 206: 337-346.
7. Sanders JL, Putman RK, Dupuis J, Xu H, Murabito JM, Araki T, Nishino M, Benjamin EJ, Levy D, Ramachandran VS, Washko GR, Curtis JL, Freeman CM, Bowler RP, Hatabu H, O'Connor GT, Hunninghake GM. The Association of Aging Biomarkers, Interstitial Lung Abnormalities, and Mortality. *Am J Respir Crit Care Med* 2021; 203: 1149-1157.
8. Washko GR, Hunninghake GM, Fernandez IE, Nishino M, Okajima Y, Yamashiro T, Ross JC, Estepar RS, Lynch DA, Brehm JM, Andriole KP, Diaz AA, Khorasani R, D'Aco K, Sciurba FC, Silverman EK, Hatabu H, Rosas IO, Investigators CO. Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N Engl J Med* 2011; 364: 897-906.
9. Hatabu H, Hunninghake GM, Richeldi L, Brown KK, Wells AU, Remy-Jardin M, Verschakelen J, Nicholson AG, Beasley MB, Christiani DC, San Jose Estepar R, Seo JB, Johkoh T, Sverzellati N, Ryerson CJ, Graham Barr R, Goo JM, Austin JHM, Powell CA, Lee KS, Inoue Y, Lynch DA. Interstitial lung abnormalities detected incidentally on CT: a Position Paper from the Fleischner Society. *Lancet Respir Med* 2020; 8: 726-737.
10. Mathai SK, Humphries S, Kropski JA, Blackwell TS, Powers J, Walts AD, Markin C, Woodward J, Chung JH, Brown KK, Steele MP, Loyd JE, Schwarz MI, Fingerlin T, Yang IV, Lynch DA, Schwartz DA. MUC5B variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax* 2019; 74: 1131-1139.

11. Greenwood CJ, Youssef GJ, Letcher P, Macdonald JA, Hagg LJ, Sanson A, McIntosh J, Hutchinson DM, Toumbourou JW, Fuller-Tyszkiewicz M, Olsson CA. A comparison of penalised regression methods for informing the selection of predictive markers. *PLoS One* 2020; 15: e0242730.
12. Mathai SK, Cardwell J, Metzger F, Powers J, Walts AD, Kropski JA, Eickelberg O, Hauck SM, Yang IV, Schwartz DA. Preclinical Pulmonary Fibrosis Circulating Protein Biomarkers. *Am J Respir Crit Care Med* 2020; 202: 1720-1724.
13. Choi B, Liu GY, Sheng Q, Amancherla K, Perry A, Huang X, San Jose Estepar R, Ash SY, Guan W, Jacobs DR, Jr., Martinez FJ, Rosas IO, Bowler RP, Kropski JA, Banovich NE, Khan SS, San Jose Estepar R, Shah R, Thyagarajan B, Kalhan R, Washko GR. Proteomic Biomarkers of Quantitative Interstitial Abnormalities in COPDGene and CARDIA Lung Study. *Am J Respir Crit Care Med* 2024.
14. Ash S, Doyle TJ, Choi B, San Jose Estepar R, Castro V, Enzer N, Kalhan R, Liu G, Bowler R, Wilson DO, San Jose Estepar R, Rosas IO, Washko GR. Utility of peripheral protein biomarkers for the prediction of incident interstitial features: a multicentre retrospective cohort study. *BMJ Open Respir Res* 2024; 11.
15. Todd JL, Neely ML, Overton R, Durham K, Gulati M, Huang H, Roman J, Newby LK, Flaherty KR, Vinisko R, Liu Y, Roy J, Schmid R, Strobel B, Hesslinger C, Leonard TB, Noth I, Belperio JA, Palmer SM, investigators I-PR. Peripheral blood proteomic profiling of idiopathic pulmonary fibrosis biomarkers in the multicentre IPF-PRO Registry. *Respir Res* 2019; 20: 227.
16. Greene KE, King TE, Jr., Kuroki Y, Bucher-Bartelson B, Hunninghake GW, Newman LS, Nagae H, Mason RJ. Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J* 2002; 19: 439-446.
17. Barlo NP, van Moorsel CH, Ruven HJ, Zanen P, van den Bosch JM, Grutters JC. Surfactant protein-D predicts survival in patients with idiopathic pulmonary fibrosis. *Sarcoidosis Vasc Diffuse Lung Dis* 2009; 26: 155-161.
18. Wang D, Day EA, Townsend LK, Djordjevic D, Jorgensen SB, Steinberg GR. GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease. *Nat Rev Endocrinol* 2021; 17: 592-607.
19. Patel S, Alvarez-Guaita A, Melvin A, Rimmington D, Dattilo A, Miedzybrodzka EL, Cimino I, Maurin AC, Roberts GP, Meek CL, Virtue S, Sparks LM, Parsons SA, Redman LM, Bray GA, Liou AP, Woods RM, Parry SA, Jeppesen PB, Kolnes AJ, Harding HP, Ron D, Vidal-Puig A, Reimann F, Gribble FM, Hulston CJ, Farooqi IS, Fafournoux P, Smith SR, Jensen J, Breen D, Wu Z, Zhang BB, Coll AP, Savage DB, O'Rahilly S. GDF15 Provides an Endocrine Signal of Nutritional Stress in Mice and Humans. *Cell Metab* 2019; 29: 707-718 e708.
20. Wischhusen J, Melero I, Fridman WH. Growth/Differentiation Factor-15 (GDF-15): From Biomarker to Novel Targetable Immune Checkpoint. *Front Immunol* 2020; 11: 951.
21. Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol* 2014; 9: 157-179.
22. Radwanska A, Cottage CT, Piras A, Overed-Sayer C, Sihlbom C, Budida R, Wrench C, Connor J, Monkley S, Hazon P, Schluter H, Thomas MJ, Hogaboam CM, Murray LA. Increased expression and accumulation of GDF15 in IPF extracellular matrix contribute to fibrosis. *JCI Insight* 2022; 7.

23. Zhang Y, Jiang M, Nouraie M, Roth MG, Tabib T, Winters S, Chen X, Sembrat J, Chu Y, Cardenes N, Tuder RM, Herzog EL, Ryu C, Rojas M, Lafyatis R, Gibson KF, McDyer JF, Kass DJ, Alder JK. GDF15 is an epithelial-derived biomarker of idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2019; 317: L510-L521.
24. Sorensen GL. Surfactant Protein D in Respiratory and Non-Respiratory Diseases. *Front Med (Lausanne)* 2018; 5: 18.
25. Martinez-Calle M, Olmeda B, Dietl P, Frick M, Perez-Gil J. Pulmonary surfactant protein SP-B promotes exocytosis of lamellar bodies in alveolar type II cells. *FASEB J* 2018; 32: 4600-4611.
26. Lomas DA, Silverman EK, Edwards LD, Locantore NW, Miller BE, Horstman DH, Tal-Singer R, Evaluation of CLtIPSEsi. Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD. *Eur Respir J* 2009; 34: 95-102.
27. Sin DD, Tammemagi CM, Lam S, Barnett MJ, Duan X, Tam A, Auman H, Feng Z, Goodman GE, Hanash S, Taguchi A. Pro-surfactant protein B as a biomarker for lung cancer prediction. *J Clin Oncol* 2013; 31: 4536-4543.
28. Leung JM, Mayo J, Tan W, Tammemagi CM, Liu G, Peacock S, Shepherd FA, Goffin J, Goss G, Nicholas G, Tremblay A, Johnston M, Martel S, Laberge F, Bhatia R, Roberts H, Burrowes P, Manos D, Stewart L, Seely JM, Gingras M, Pasian S, Tsao MS, Lam S, Sin DD, Pan-Canadian Early Lung Cancer Study G. Plasma pro-surfactant protein B and lung function decline in smokers. *Eur Respir J* 2015; 45: 1037-1045.
29. Park J, Pabon M, Choi AMK, Siempos, II, Fredenburgh LE, Baron RM, Jeon K, Chung CR, Yang JH, Park CM, Suh GY. Plasma surfactant protein-D as a diagnostic biomarker for acute respiratory distress syndrome: validation in US and Korean cohorts. *BMC Pulm Med* 2017; 17: 204.
30. Selman M, Lin HM, Montano M, Jenkins AL, Estrada A, Lin Z, Wang G, DiAngelo SL, Guo X, Umstead TM, Lang CM, Pardo A, Phelps DS, Floros J. Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. *Hum Genet* 2003; 113: 542-550.
31. Kahn N, Rossler AK, Hornemann K, Muley T, Grunig E, Schmidt W, Herth FJF, Kreuter M. C-proSP-B: A Possible Biomarker for Pulmonary Diseases? *Respiration* 2018; 96: 117-126.
32. Papaioannou AI, Kostikas K, Manali ED, Papadaki G, Roussou A, Spathis A, Mazioti A, Tomos I, Papanikolaou I, Loukides S, Chainis K, Karakitsos P, Griese M, Papiris S. Serum Levels of Surfactant Proteins in Patients with Combined Pulmonary Fibrosis and Emphysema (CPFE). *PLoS One* 2016; 11: e0157789.
33. Raghu G, Richeldi L, Jagerschmidt A, Martin V, Subramaniam A, Ozoux ML, Esperet CA, Soubrane C. Idiopathic Pulmonary Fibrosis: Prospective, Case-Controlled Study of Natural History and Circulating Biomarkers. *Chest* 2018; 154: 1359-1370.
34. Li J, Li J, Hao H, Lu F, Wang J, Ma M, Jia B, Zhuo M, Wang J, Chi Y, Zhai X, Wang Y, Wu M, An T, Zhao J, Yang F, Wang Z. Secreted proteins MDK, WFDC2, and CXCL14 as candidate biomarkers for early diagnosis of lung adenocarcinoma. *BMC Cancer* 2023; 23: 110.
35. Noskovicova N, Heinzelmann K, Burgstaller G, Behr J, Eickelberg O. Cub domain-containing protein 1 negatively regulates TGF-beta signaling and myofibroblast differentiation. *Am J Physiol Lung Cell Mol Physiol* 2018; 314: L695-L707.

36. Jaeger B, Schupp JC, Plappert L, Terwolbeck O, Artysh N, Kayser G, Engelhard P, Adams TS, Zweigerdt R, Kempf H, Lienenklaus S, Garrels W, Nazarenko I, Jonigk D, Wygrecka M, Klatt D, Schambach A, Kaminski N, Prasse A. Airway basal cells show a dedifferentiated KRT17(high)Phenotype and promote fibrosis in idiopathic pulmonary fibrosis. *Nat Commun* 2022; 13: 5637.
37. Munguia-Reyes A, Balderas-Martinez YI, Becerril C, Checa M, Ramirez R, Ortiz B, Melendez-Zajgla J, Pardo A, Selman M. R-Spondin-2 Is Upregulated in Idiopathic Pulmonary Fibrosis and Affects Fibroblast Behavior. *Am J Respir Cell Mol Biol* 2018; 59: 65-76.
38. Borie R, Kannengiesser C, Antoniou K, Bonella F, Crestani B, Fabre A, Froidure A, Galvin L, Griese M, Grutters JC, Molina-Molina M, Poletti V, Prasse A, Renzoni E, van der Smagt J, van Moorsel CHM. European Respiratory Society statement on familial pulmonary fibrosis. *Eur Respir J* 2023; 61.
39. Rose JA, Menon AA, Hino T, Hata A, Nishino M, Lynch DA, Rosas IO, El-Chemaly S, Raby BA, Ash SY, Choi B, Washko GR, Silverman EK, Cho MH, Hatabu H, Putman RK, Hunninghake GM. Suspected Interstitial Lung Disease in COPDGene Study. *Am J Respir Crit Care Med* 2023; 207: 60-68.
40. Raffield LM, Dang H, Pratte KA, Jacobson S, Gillenwater LA, Ampleford E, Barjaktarevic I, Basta P, Clish CB, Comellas AP, Cornell E, Curtis JL, Doerschuk C, Durda P, Emson C, Freeman CM, Guo X, Hastie AT, Hawkins GA, Herrera J, Johnson WC, Labaki WW, Liu Y, Masters B, Miller M, Ortega VE, Papanicolaou G, Peters S, Taylor KD, Rich SS, Rotter JI, Auer P, Reiner AP, Tracy RP, Ngo D, Gerszten RE, O'Neal WK, Bowler RP, Consortium NT-OfPM. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics* 2020; 20: e1900278.
41. Katz DH, Robbins JM, Deng S, Tahir UA, Bick AG, Pampana A, Yu Z, Ngo D, Benson MD, Chen ZZ, Cruz DE, Shen D, Gao Y, Bouchard C, Sarzynski MA, Correa A, Natarajan P, Wilson JG, Gerszten RE. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci Adv* 2022; 8: eabm5164.
42. Bowman WS, Newton CA, Linderholm AL, Neely ML, Pugashetti JV, Kaul B, Vo V, Echt GA, Leon W, Shah RJ, Huang Y, Garcia CK, Wolters PJ, Oldham JM. Proteomic biomarkers of progressive fibrosing interstitial lung disease: a multicentre cohort analysis. *Lancet Respir Med* 2022; 10: 593-602.

Supplemental Figures and Tables

Table E1: Suggestive Protein Associations (p-value <0.05) with ILA in CGS-PF Cohort

Rank	SOMAID	Protein	Odds Ratio	95% CI	p-value	FDR-adjusted p-value
1	seq.4374.45	GDF15	4.93	2.82 - 8.62	2.3E-08	0.0002
2	seq.19590.46	SFTPD	4.34	2.51 - 7.50	1.5E-07	0.0005
3	seq.10672.75	SFTPB	3.59	2.21 - 5.86	2.8E-07	0.0006
4	seq.11388.75	WFDC2	3.15	1.96 - 5.07	2.2E-06	0.0035
5	seq.16818.200	CDCP1	2.4	1.60 - 3.61	2.4E-05	0.0303
6	seq.21118.48	GCHFR	2.4	1.59 - 3.62	2.8E-05	0.0303
7	seq.4829.43	SFN	2.5	1.61 - 3.86	3.9E-05	0.0354
8	seq.4496.60	MMP12	2.17	1.47 - 3.22	0.0001	0.0800
9	seq.9986.14	NPW	2.27	1.49 - 3.46	0.0001	0.0800
10	seq.9266.1	TREM1	2.19	1.46 - 3.27	0.0001	0.0800
11	seq.5115.31	RELT	2.13	1.44 - 3.14	0.0001	0.0800
12	seq.2211.9	TIMP1	2.24	1.47 - 3.40	0.0002	0.0800
13	seq.13463.1	PXDN	2.22	1.46 - 3.38	0.0002	0.0900
14	seq.5124.62	ICAM5	2.23	1.46 - 3.41	0.0002	0.0900
15	seq.21724.22	GALNS	2.05	1.40 - 3.00	0.0002	0.0900
16	seq.24693.5	CLIC3	2.32	1.48 - 3.64	0.0002	0.1000
17	seq.3438.10	FSTL3	1.96	1.36 - 2.84	0.0003	0.1200
18	seq.2212.69	PLAT	2.19	1.42 - 3.37	0.0004	0.1200
19	seq.11200.52	CD93	1.98	1.36 - 2.89	0.0004	0.1200
20	seq.3519.3	CCL17	1.99	1.36 - 2.90	0.0004	0.1200
21	seq.23640.10	EHD2	2.04	1.37 - 3.04	0.0004	0.1300
22	seq.16614.27	RSPO1	2.05	1.37 - 3.05	0.0004	0.1300
23	seq.7211.2	RNASE1	1.96	1.34 - 2.86	0.0005	0.1300
24	seq.14684.17	CAPN2	2.03	1.36 - 3.05	0.0006	0.1500
25	seq.6252.62	SCGB3A1	2.07	1.36 - 3.14	0.0006	0.1600
26	seq.6645.53	POSTN	1.97	1.33 - 2.91	0.0007	0.1600
27	seq.10702.1	COL28A1	1.97	1.32 - 2.94	0.0009	0.2200
28	seq.23622.128	INIP	0.44	0.26 - 0.72	0.0011	0.2500
29	seq.13431.74	SHISAL2A	1.99	1.31 - 3.00	0.0012	0.2500
30	seq.15533.97	MSR1	1.9	1.29 - 2.80	0.0012	0.2500
31	seq.14088.38	IGFBP6	1.85	1.27 - 2.69	0.0013	0.2600
32	seq.2597.8	VEGFA	2.08	1.33 - 3.26	0.0014	0.2800
33	seq.13738.8	INHBA	1.85	1.26 - 2.71	0.0015	0.2900
34	seq.19601.15	ASB9	1.86	1.27 - 2.72	0.0016	0.2900
35	seq.3326.58	CADM1	1.82	1.25 - 2.65	0.0017	0.3000
36	seq.21495.134	GPR37	1.87	1.26 - 2.76	0.0017	0.3000
37	seq.6580.29	PZP	0.58	0.41 - 0.81	0.0017	0.3000
38	seq.5301.7	CCL11	1.82	1.25 - 2.66	0.0018	0.3000
39	seq.4234.8	IL1RL1	1.76	1.23 - 2.52	0.0020	0.3000
40	seq.22124.94	DYTN	1.86	1.25 - 2.76	0.0021	0.3000

41	seq.9316.67	WFDC1	1.86	1.25 - 2.76	0.0021	0.3000
42	seq.18253.8	SSBP1	1.73	1.22 - 2.45	0.0021	0.3000
43	seq.21705.33	METRNL	1.83	1.24 - 2.68	0.0021	0.3000
44	seq.17366.6	DCUN1D1	1.92	1.26 - 2.91	0.0022	0.3000
45	seq.24320.3	TRIM72	2.04	1.29 - 3.21	0.0022	0.3000
46	seq.3073.51	IL18BP	1.81	1.24 - 2.66	0.0023	0.3000
47	seq.17692.2	BTN3A3	1.93	1.26 - 2.94	0.0023	0.3000
48	seq.11102.22	REG4	1.83	1.24 - 2.71	0.0024	0.3000
49	seq.2654.19	TNFRSF1A	1.79	1.23 - 2.60	0.0024	0.3000
50	seq.5630.48	CD300A	1.8	1.23 - 2.63	0.0024	0.3000
51	seq.5496.49	SPON1	1.78	1.23 - 2.58	0.0025	0.3000
52	seq.4834.61	EPHA2	1.75	1.22 - 2.52	0.0025	0.3000
53	seq.20516.11	SCN3B	2.55	1.39 - 4.67	0.0025	0.3000
54	seq.7258.5	PRG2	1.77	1.22 - 2.57	0.0027	0.3100
55	seq.3044.3	CCL18	1.79	1.22 - 2.62	0.0027	0.3100
56	seq.17678.28	VSIG4	1.74	1.21 - 2.51	0.0028	0.3100
57	seq.22047.46	COL5A1	1.76	1.21 - 2.54	0.0029	0.3100
58	seq.17836.17	S100A16	1.76	1.21 - 2.55	0.0029	0.3100
59	seq.3007.7	SIGLEC9	0.59	0.42 - 0.84	0.0030	0.3100
60	seq.23038.63	UNC5B	1.74	1.21 - 2.52	0.0030	0.3100
61	seq.6223.5	GUCA2B	1.75	1.21 - 2.54	0.0030	0.3100
62	seq.6546.41	DKK2	1.73	1.20 - 2.48	0.0030	0.3100
63	seq.23380.16	L3HYPDH	1.78	1.21 - 2.60	0.0032	0.3100
64	seq.21667.57	PDE1B	1.84	1.23 - 2.76	0.0032	0.3100
65	seq.18893.26	ADGRG1	1.73	1.20 - 2.50	0.0033	0.3100
66	seq.11193.27	HNF1A	1.76	1.21 - 2.57	0.0033	0.3100
67	seq.17685.9	APOA4	1.76	1.21 - 2.57	0.0033	0.3100
68	seq.3364.76	CTSV	0.56	0.38 - 0.82	0.0033	0.3100
69	seq.9638.2	TIGIT	1.74	1.20 - 2.53	0.0034	0.3100
70	seq.21382.70	ELAC1	1.75	1.20 - 2.55	0.0034	0.3100
71	seq.6321.65	PCDHGA10	1.81	1.21 - 2.70	0.0036	0.3100
72	seq.9313.27	CBLN1	0.57	0.39 - 0.83	0.0036	0.3100
73	seq.11647.6	FZD10	0.53	0.34 - 0.81	0.0036	0.3100
74	seq.8348.4	EPHB2	1.73	1.19 - 2.50	0.0037	0.3100
75	seq.7085.81	CDSN	1.71	1.19 - 2.46	0.0037	0.3100
76	seq.21391.17	CGB3	1.73	1.19 - 2.50	0.0037	0.3100
77	seq.11196.31	COL6A3	1.77	1.20 - 2.60	0.0038	0.3100
78	seq.4971.1	CTSZ	1.74	1.19 - 2.52	0.0038	0.3100
79	seq.2687.2	MIA	1.71	1.19 - 2.47	0.0038	0.3100
80	seq.18901.26	HSPA1B	1.72	1.19 - 2.49	0.0038	0.3100
81	seq.20988.63	RP9	1.68	1.18 - 2.40	0.0041	0.3300
82	seq.2609.59	CST3	1.66	1.17 - 2.35	0.0042	0.3300
83	seq.9931.20	KRT1	1.71	1.18 - 2.48	0.0043	0.3300
84	seq.7861.9	ROR2	1.69	1.18 - 2.42	0.0045	0.3400
85	seq.21545.51	MUSTN1	1.71	1.18 - 2.47	0.0046	0.3400
86	seq.10738.11	FBLN5	1.86	1.21 - 2.85	0.0046	0.3400

87	seq.5349.69	DLL1	1.7	1.18 - 2.45	0.0048	0.3500
88	seq.13381.49	B4GALT1	1.66	1.17 - 2.36	0.0048	0.3500
89	seq.5763.67	DEFB104A	1.64	1.16 - 2.30	0.0049	0.3500
90	seq.4534.10	PRSS22	1.68	1.17 - 2.40	0.0049	0.3500
91	seq.11696.7	CRABP2	1.75	1.18 - 2.58	0.0051	0.3500
92	seq.13101.60	SOST	1.66	1.16 - 2.36	0.0051	0.3500
93	seq.7244.16	IGFLR1	1.67	1.16 - 2.40	0.0053	0.3600
94	seq.7921.65	FJX1	1.67	1.16 - 2.39	0.0053	0.3600
95	seq.8235.48	CHGB	1.66	1.16 - 2.37	0.0055	0.3700
96	seq.20066.19	FZD5	2	1.23 - 3.27	0.0055	0.3700
97	seq.7928.183	TPST1	1.97	1.22 - 3.19	0.0055	0.3700
98	seq.4831.4	SELL	0.59	0.40 - 0.86	0.0056	0.3700
99	seq.2888.49	C7	1.85	1.20 - 2.87	0.0057	0.3700
100	seq.9561.21	VCAN	1.61	1.15 - 2.26	0.0058	0.3700
101	seq.21883.17	BMP5	1.68	1.16 - 2.43	0.0058	0.3700
102	seq.9880.33	TDO2	1.81	1.19 - 2.75	0.0059	0.3700
103	seq.5801.72	NGF	1.63	1.15 - 2.31	0.0060	0.3700
104	seq.6576.1	ART4	1.68	1.16 - 2.44	0.0063	0.3800
105	seq.2480.58	TIMP3	1.68	1.16 - 2.43	0.0063	0.3800
106	seq.3396.54	REN	1.65	1.15 - 2.36	0.0064	0.3800
107	seq.10713.151	CD300LB	1.6	1.14 - 2.24	0.0064	0.3800
108	seq.8476.11	CHGA	2.71	1.32 - 5.54	0.0064	0.3800
109	seq.24215.8	PRRX1	1.63	1.15 - 2.32	0.0064	0.3800
110	seq.3508.78	CCL22	1.7	1.16 - 2.49	0.0064	0.3800
111	seq.9547.29	EMID1	1.82	1.18 - 2.79	0.0066	0.3800
112	seq.3378.49	KLK7	0.54	0.35 - 0.84	0.0066	0.3800
113	seq.10460.1	CHIT1	1.68	1.16 - 2.45	0.0067	0.3800
114	seq.20448.7	POLR2F	1.69	1.15 - 2.47	0.0069	0.3800
115	seq.20577.5	FZD8	1.65	1.15 - 2.37	0.0069	0.3800
116	seq.21232.39	KLK3	1.61	1.14 - 2.28	0.0070	0.3800
117	seq.25112.6	RPS6KB2	1.66	1.15 - 2.39	0.0070	0.3800
118	seq.7210.25	APLP1	0.6	0.41 - 0.87	0.0071	0.3800
119	seq.3362.61	CHRDL1	1.65	1.15 - 2.38	0.0071	0.3800
120	seq.6713.4	LRP11	1.64	1.14 - 2.35	0.0072	0.3800
121	seq.6918.183	CCK	1.64	1.14 - 2.37	0.0074	0.4000
122	seq.21548.20	GGT5	0.6	0.41 - 0.87	0.0075	0.4000
123	seq.18819.21	PPIC	1.75	1.16 - 2.64	0.0077	0.4000
124	seq.8243.55	SPINK1	1.75	1.16 - 2.65	0.0078	0.4000
125	seq.7127.3	APOA2	1.67	1.14 - 2.44	0.0078	0.4000
126	seq.3403.1	TPSB2	1.7	1.15 - 2.51	0.0078	0.4000
127	seq.6603.18	ANOS1	1.6	1.13 - 2.25	0.0080	0.4000
128	seq.20576.71	FZD7	1.64	1.14 - 2.37	0.0081	0.4000
129	seq.22076.34	ATOH1	1.67	1.14 - 2.44	0.0081	0.4000
130	seq.6373.54	DLK1	1.63	1.13 - 2.34	0.0083	0.4100
131	seq.21526.88	IZUMO1	1.6	1.13 - 2.28	0.0084	0.4100
132	seq.3336.50	TFPI	1.72	1.15 - 2.59	0.0085	0.4100

133	seq.3339.33	THBS2	1.61	1.13 - 2.30	0.0087	0.4100
134	seq.13112.179	FSTL1	1.6	1.13 - 2.28	0.0088	0.4100
135	seq.24891.54	EPS8L2	1.63	1.13 - 2.36	0.0088	0.4100
136	seq.12766.33	GPR142	1.61	1.13 - 2.30	0.0088	0.4100
137	seq.13943.38	DPY30	1.65	1.13 - 2.39	0.0088	0.4100
138	seq.20966.7	SENP8	1.61	1.13 - 2.30	0.0089	0.4100
139	seq.15535.3	PRSS27	1.58	1.12 - 2.23	0.0090	0.4100
140	seq.15468.14	CFHR1	1.74	1.15 - 2.64	0.0090	0.4100
141	seq.10620.21	MSMB	1.68	1.14 - 2.48	0.0091	0.4100
142	seq.9211.19	SERPINF1	1.73	1.15 - 2.62	0.0092	0.4100
143	seq.9756.6	TAGLN	1.61	1.12 - 2.30	0.0092	0.4100
144	seq.18814.21	INHBA|INHBC	1.63	1.13 - 2.36	0.0092	0.4100
145	seq.7841.84	ESAM	1.67	1.13 - 2.45	0.0093	0.4100
146	seq.18925.24	PSMA5	1.78	1.15 - 2.74	0.0094	0.4100
147	seq.3234.23	CCDC80	1.6	1.12 - 2.28	0.0094	0.4100
148	seq.21433.3	BTN3A2	1.65	1.13 - 2.41	0.0095	0.4100
149	seq.8807.13	SNX1	1.58	1.12 - 2.23	0.0097	0.4100
150	seq.9468.8	LMAN2	1.65	1.13 - 2.41	0.0097	0.4100
151	seq.8005.1	MXRA7	1.59	1.12 - 2.27	0.0098	0.4200
152	seq.6408.2	INHBC	1.72	1.14 - 2.59	0.0101	0.4200
153	seq.8903.1	COX6C	1.64	1.13 - 2.40	0.0102	0.4300
154	seq.9294.45	MFAP2	1.57	1.11 - 2.21	0.0102	0.4300
155	seq.13553.4	DCUN1D3	1.56	1.11 - 2.20	0.0105	0.4300
156	seq.9013.60	NTN1	1.59	1.11 - 2.27	0.0105	0.4300
157	seq.25233.2	MPI	1.68	1.13 - 2.50	0.0107	0.4400
158	seq.19252.67	PSMB2	1.83	1.15 - 2.91	0.0108	0.4400
159	seq.9748.31	GSTM3	1.55	1.11 - 2.17	0.0110	0.4400
160	seq.3214.3	NRP1	1.77	1.14 - 2.74	0.0111	0.4400
161	seq.9359.9	DLK2	1.59	1.11 - 2.29	0.0113	0.4400
162	seq.3521.16	CGA|TSHB	1.59	1.11 - 2.28	0.0115	0.4400
163	seq.8032.23	MCCD1	1.63	1.12 - 2.38	0.0115	0.4400
164	seq.9580.5	LAMC2	1.64	1.12 - 2.41	0.0115	0.4400
165	seq.11178.21	SVEP1	1.58	1.11 - 2.26	0.0116	0.4400
166	seq.7006.4	COL25A1	1.65	1.12 - 2.43	0.0117	0.4400
167	seq.6990.44	SDF2L1	1.59	1.11 - 2.28	0.0117	0.4400
168	seq.9049.2	SLC16A3	1.55	1.10 - 2.17	0.0118	0.4400
169	seq.8908.14	KCNE5	0.63	0.43 - 0.90	0.0118	0.4400
170	seq.23379.60	C16orf72	1.89	1.15 - 3.10	0.0118	0.4400
171	seq.3059.50	TNFSF13B	1.57	1.10 - 2.22	0.0119	0.4400
172	seq.2590.69	ROR1	1.6	1.11 - 2.30	0.0119	0.4400
173	seq.4157.2	F2	0.38	0.18 - 0.81	0.0120	0.4400
174	seq.22578.17	ROBO2	1.63	1.11 - 2.38	0.0120	0.4400
175	seq.22527.4	PDLIM3	1.59	1.11 - 2.29	0.0121	0.4400
176	seq.6929.10	ST8SIA3	1.68	1.12 - 2.53	0.0122	0.4400
177	seq.6440.31	MFAP5	1.67	1.12 - 2.49	0.0123	0.4400
178	seq.8983.7	GOLM1	1.56	1.10 - 2.22	0.0124	0.4400

179	seq.16597.11	GLRX5	1.54	1.10 - 2.16	0.0125	0.4400
180	seq.13133.73	LTBP4	1.59	1.11 - 2.29	0.0125	0.4400
181	seq.21340.38	HLF	1.59	1.11 - 2.30	0.0125	0.4400
182	seq.22094.15	C2orf73	1.65	1.11 - 2.44	0.0126	0.4400
183	seq.6375.75	XXYLT1	1.56	1.10 - 2.20	0.0126	0.4400
184	seq.9076.25	PENK	1.66	1.11 - 2.46	0.0127	0.4400
185	seq.13094.75	RSPO3	1.57	1.10 - 2.25	0.0128	0.4400
186	seq.19768.13	CSTB	1.56	1.10 - 2.21	0.0128	0.4400
187	seq.2944.66	NBL1	1.56	1.10 - 2.22	0.0133	0.4500
188	seq.6387.61	DEFA5	1.56	1.10 - 2.22	0.0133	0.4500
189	seq.6359.50	POMGNT2	1.59	1.10 - 2.31	0.0133	0.4500
190	seq.18397.5	AKR1C4	1.53	1.09 - 2.15	0.0135	0.4500
191	seq.10435.2	AGA	1.69	1.11 - 2.56	0.0135	0.4500
192	seq.22530.111	PHYHD1	1.6	1.10 - 2.33	0.0136	0.4500
193	seq.3485.28	B2M	1.54	1.09 - 2.17	0.0136	0.4500
194	seq.3340.53	THBS4	1.59	1.10 - 2.29	0.0137	0.4500
195	seq.22960.8	FCGR2B	1.54	1.09 - 2.18	0.0138	0.4500
196	seq.7251.64	C1QTNF3	1.63	1.10 - 2.40	0.0139	0.4500
197	seq.6605.17	IGFALS	0.65	0.46 - 0.92	0.0139	0.4500
198	seq.3046.31	RETN	1.56	1.09 - 2.23	0.0140	0.4500
199	seq.16610.13	LRP10	1.63	1.10 - 2.40	0.0140	0.4500
200	seq.15530.33	EPHB4	1.56	1.09 - 2.22	0.0141	0.4500
201	seq.2249.25	HCE004333	0.6	0.40 - 0.90	0.0142	0.4500
202	seq.9332.6	DEFB136	1.76	1.12 - 2.77	0.0142	0.4500
203	seq.9518.95	PLA1A	1.71	1.11 - 2.64	0.0143	0.4500
204	seq.20561.15	CNTN6	1.57	1.09 - 2.24	0.0144	0.4500
205	seq.23176.17	DAG1	1.62	1.10 - 2.38	0.0144	0.4500
206	seq.13661.193	CST5	1.61	1.10 - 2.37	0.0147	0.4600
207	seq.19568.17	IL15	1.54	1.09 - 2.18	0.0148	0.4600
208	seq.4148.49	PAPPA	1.6	1.10 - 2.34	0.0149	0.4600
209	seq.4155.3	TNC	1.54	1.09 - 2.18	0.0152	0.4700
210	seq.3038.9	CXCL11	1.53	1.09 - 2.17	0.0154	0.4700
211	seq.19127.1	HSPB6	1.56	1.09 - 2.24	0.0154	0.4700
212	seq.5688.65	CBLN4	1.59	1.09 - 2.32	0.0155	0.4700
213	seq.19241.31	RBP5	1.54	1.09 - 2.19	0.0156	0.4700
214	seq.11159.14	MUL1	1.6	1.09 - 2.34	0.0156	0.4700
215	seq.13123.3	FLRT3	1.57	1.09 - 2.26	0.0156	0.4700
216	seq.5139.32	UNC5C	1.55	1.09 - 2.21	0.0157	0.4700
217	seq.19141.22	DAP	1.52	1.08 - 2.13	0.0158	0.4700
218	seq.2992.59	IL17RA	1.55	1.09 - 2.21	0.0158	0.4700
219	seq.23250.3	TCEAL4	1.65	1.10 - 2.49	0.0160	0.4700
220	seq.17770.42	EMC8	1.54	1.08 - 2.20	0.0160	0.4700
221	seq.10445.20	APOM	0.65	0.45 - 0.92	0.0163	0.4700
222	seq.19748.3	MVD	1.54	1.08 - 2.18	0.0163	0.4700
223	seq.5631.83	MLN	1.55	1.08 - 2.21	0.0164	0.4700
224	seq.8093.13	MYORG	1.53	1.08 - 2.18	0.0165	0.4700

225	seq.16594.44	FAIM	1.53	1.08 - 2.18	0.0165	0.4700
226	seq.15580.2	EPHA7	1.55	1.08 - 2.23	0.0165	0.4700
227	seq.19376.74	NNMT	1.58	1.09 - 2.31	0.0166	0.4700
228	seq.20564.53	THY1	1.57	1.09 - 2.28	0.0167	0.4700
229	seq.5637.81	NTNG1	1.51	1.08 - 2.13	0.0168	0.4700
230	seq.8480.29	EFEMP1	1.53	1.08 - 2.17	0.0168	0.4700
231	seq.10924.258	CYB5D2	1.56	1.08 - 2.24	0.0170	0.4700
232	seq.14054.17	IL15RA	1.54	1.08 - 2.19	0.0171	0.4700
233	seq.14048.7	IL1RAP	0.63	0.43 - 0.92	0.0171	0.4700
234	seq.9590.10	COX7A2L	1.54	1.08 - 2.21	0.0176	0.4800
235	seq.8467.9	INHBA|INHBB	1.56	1.08 - 2.24	0.0177	0.4800
236	seq.24934.96	OGFR	1.55	1.08 - 2.22	0.0179	0.4900
237	seq.19631.13	KNG1	0.64	0.45 - 0.93	0.0181	0.4900
238	seq.10424.31	NPDC1	1.61	1.08 - 2.39	0.0181	0.4900
239	seq.20984.142	DUSP19	1.84	1.11 - 3.04	0.0183	0.4900
240	seq.7959.34	CDH7	1.57	1.08 - 2.29	0.0186	0.4900
241	seq.16780.6	HSPA1A	1.55	1.08 - 2.23	0.0186	0.4900
242	seq.8982.65	THBS3	1.55	1.08 - 2.22	0.0187	0.4900
243	seq.18896.23	HS6ST3	0.63	0.43 - 0.93	0.0187	0.4900
244	seq.23566.6	APOL2	1.52	1.07 - 2.15	0.0188	0.4900
245	seq.18830.1	ITLN1	1.56	1.08 - 2.27	0.0188	0.4900
246	seq.20175.17	COL9A3	1.57	1.08 - 2.29	0.0191	0.5000
247	seq.19492.5	KAAG1	1.52	1.07 - 2.16	0.0191	0.5000
248	seq.14108.15	TGFB1	1.52	1.07 - 2.16	0.0195	0.5000
249	seq.24926.9	KHSRP	1.53	1.07 - 2.18	0.0196	0.5000
250	seq.18918.86	PDE4A	1.52	1.07 - 2.16	0.0196	0.5000
251	seq.16858.384	RCN3	1.6	1.08 - 2.38	0.0197	0.5000
252	seq.11116.16	C11orf87	1.59	1.08 - 2.36	0.0200	0.5100
253	seq.13660.76	ANGPT2	1.54	1.07 - 2.22	0.0200	0.5100
254	seq.19108.50	MECP2	1.5	1.07 - 2.10	0.0200	0.5100
255	seq.7156.2	FUT10	1.55	1.07 - 2.25	0.0202	0.5100
256	seq.21942.14	WNT16	1.57	1.07 - 2.31	0.0203	0.5100
257	seq.6388.21	CCDC126	0.65	0.45 - 0.94	0.0205	0.5100
258	seq.21282.21	TNIP1	1.54	1.07 - 2.23	0.0206	0.5100
259	seq.13499.30	F8	1.54	1.07 - 2.22	0.0206	0.5100
260	seq.4968.50	CAPG	1.51	1.06 - 2.13	0.0206	0.5100
261	seq.9627.15	VIT	1.54	1.07 - 2.23	0.0208	0.5100
262	seq.13114.50	LUM	1.53	1.07 - 2.19	0.0209	0.5100
263	seq.10521.10	MXRA8	0.66	0.46 - 0.94	0.0209	0.5100
264	seq.6506.54	TMED10	1.48	1.06 - 2.08	0.0209	0.5100
265	seq.15453.3	AMBP	1.55	1.07 - 2.24	0.0210	0.5100
266	seq.11715.1	POU2F1	1.67	1.08 - 2.59	0.0211	0.5100
267	seq.12386.11	RNPEP	1.52	1.07 - 2.18	0.0211	0.5100
268	seq.11494.4	ELL2	1.52	1.06 - 2.17	0.0212	0.5100
269	seq.15441.6	GM2A	1.49	1.06 - 2.10	0.0213	0.5100
270	seq.17515.6	HSPA13	1.52	1.06 - 2.18	0.0213	0.5100

271	seq.8374.5	NGFR	1.54	1.07 - 2.21	0.0214	0.5100
272	seq.4413.3	SLPI	1.54	1.07 - 2.23	0.0215	0.5100
273	seq.6923.1	PLOD2	1.51	1.06 - 2.15	0.0215	0.5100
274	seq.8450.36	ADCYAP1	1.51	1.06 - 2.16	0.0217	0.5100
275	seq.7083.74	MATN4	1.5	1.06 - 2.11	0.0218	0.5100
276	seq.7192.37	IFNLR1	1.49	1.06 - 2.09	0.0218	0.5100
277	seq.14747.9	CRLF1	1.49	1.06 - 2.09	0.0219	0.5100
278	seq.23029.3	SH3GL3	0.65	0.45 - 0.94	0.0222	0.5100
279	seq.10743.13	SLITRK1	0.67	0.47 - 0.94	0.0225	0.5200
280	seq.7835.2	BNIP3L	1.55	1.06 - 2.25	0.0227	0.5200
281	seq.22490.16	LCNL1	1.48	1.06 - 2.08	0.0228	0.5200
282	seq.19238.12	GLUL	1.48	1.06 - 2.09	0.0229	0.5200
283	seq.15426.5	NEU1	1.52	1.06 - 2.19	0.0230	0.5200
284	seq.17750.8	RRM2	1.49	1.06 - 2.11	0.0230	0.5200
285	seq.20237.38	OAT	1.49	1.06 - 2.10	0.0231	0.5200
286	seq.9050.170	RSRP1	1.55	1.06 - 2.26	0.0232	0.5200
287	seq.3040.59	CCL3	1.51	1.06 - 2.17	0.0233	0.5200
288	seq.12548.75	RRAGC	1.5	1.06 - 2.12	0.0236	0.5300
289	seq.10710.23	ZPBP2	1.69	1.07 - 2.66	0.0237	0.5300
290	seq.17699.43	TIMELESS	1.61	1.06 - 2.44	0.0241	0.5300
291	seq.18926.7	PSMA3	1.58	1.06 - 2.36	0.0242	0.5300
292	seq.23200.25	C1QTNF4	1.5	1.05 - 2.13	0.0242	0.5300
293	seq.22115.2	CREBL2	1.54	1.06 - 2.25	0.0243	0.5300
294	seq.13552.7	SWAP70	1.5	1.05 - 2.12	0.0244	0.5300
295	seq.15395.15	GSTM1	1.5	1.05 - 2.14	0.0245	0.5300
296	seq.22104.39	CDX1	1.64	1.07 - 2.51	0.0245	0.5300
297	seq.10505.12	RNF43	1.59	1.06 - 2.38	0.0246	0.5300
298	seq.16318.12	ACVRL1	1.66	1.07 - 2.59	0.0246	0.5300
299	seq.9191.8	TFF2	1.51	1.05 - 2.17	0.0248	0.5300
300	seq.7999.23	ENTPD1	1.51	1.05 - 2.17	0.0248	0.5300
301	seq.4924.32	MMP1	1.51	1.05 - 2.16	0.0251	0.5300
302	seq.7122.31	LGR5	1.47	1.05 - 2.06	0.0251	0.5300
303	seq.12574.36	EDN2	1.52	1.05 - 2.19	0.0253	0.5400
304	seq.5609.92	TAFA5	1.56	1.06 - 2.31	0.0257	0.5400
305	seq.8991.115	FCRL4	1.51	1.05 - 2.17	0.0260	0.5400
306	seq.12578.13	ARF3	1.57	1.06 - 2.34	0.0260	0.5400
307	seq.23522.1	GIMD1	1.51	1.05 - 2.17	0.0261	0.5400
308	seq.5134.52	HAVCR2	1.5	1.05 - 2.15	0.0261	0.5400
309	seq.17454.15	EGFL6	1.49	1.05 - 2.12	0.0264	0.5400
310	seq.10956.82	GCSH	1.48	1.05 - 2.10	0.0264	0.5400
311	seq.11516.7	FABP1	1.54	1.05 - 2.24	0.0264	0.5400
312	seq.8069.85	CD3E	1.54	1.05 - 2.27	0.0265	0.5400
313	seq.12872.35	CNGA2	1.6	1.06 - 2.43	0.0266	0.5400
314	seq.8288.27	APOH	1.49	1.05 - 2.12	0.0266	0.5400
315	seq.8006.12	DNAJB12	1.51	1.05 - 2.16	0.0266	0.5400
316	seq.12475.48	CLIC5	1.51	1.05 - 2.17	0.0267	0.5400

317	seq.9058.1	TIMM21	1.49	1.05 - 2.12	0.0269	0.5400
318	seq.5643.2	LCN8	1.47	1.04 - 2.07	0.0271	0.5400
319	seq.24912.40	WDR26	1.47	1.04 - 2.07	0.0271	0.5400
320	seq.3050.7	VWF	1.51	1.05 - 2.17	0.0271	0.5400
321	seq.6392.7	CCN5	1.5	1.05 - 2.15	0.0271	0.5400
322	seq.5128.53	SLAMF6	1.52	1.05 - 2.21	0.0272	0.5400
323	seq.17796.15	IMPA1	1.49	1.05 - 2.12	0.0272	0.5400
324	seq.22496.21	LYRM1	1.5	1.05 - 2.15	0.0272	0.5400
325	seq.2640.3	MST1R	0.66	0.46 - 0.95	0.0272	0.5400
326	seq.12361.102	RRAS2	1.57	1.05 - 2.34	0.0273	0.5400
327	seq.5661.15	IL18	1.48	1.04 - 2.09	0.0274	0.5400
328	seq.8107.12	TMEM234	1.5	1.05 - 2.14	0.0278	0.5400
329	seq.19553.14	STX1A	1.48	1.04 - 2.11	0.0278	0.5400
330	seq.6379.62	ADAMTSL2	1.47	1.04 - 2.07	0.0279	0.5400
331	seq.6213.10	CGB2	1.5	1.04 - 2.15	0.0280	0.5400
332	seq.24690.1	CABP2	1.46	1.04 - 2.05	0.0281	0.5400
333	seq.21356.6	B9D2	1.47	1.04 - 2.06	0.0281	0.5400
334	seq.5404.53	TNFRSF21	1.48	1.04 - 2.10	0.0282	0.5400
335	seq.12663.1	TST	1.51	1.05 - 2.19	0.0282	0.5400
336	seq.24422.36	SPATA24	0.66	0.45 - 0.96	0.0284	0.5400
337	seq.11242.33	TGM1	1.52	1.05 - 2.21	0.0284	0.5400
338	seq.17837.5	CLIC2	1.48	1.04 - 2.10	0.0287	0.5400
339	seq.20987.21	SIX6	1.48	1.04 - 2.09	0.0288	0.5400
340	seq.10666.7	GNPTG	1.5	1.04 - 2.16	0.0290	0.5400
341	seq.4549.78	FUT5	1.5	1.04 - 2.15	0.0291	0.5400
342	seq.9250.87	DEFA1	1.46	1.04 - 2.06	0.0291	0.5400
343	seq.19143.38	CYB5R2	1.58	1.05 - 2.39	0.0291	0.5400
344	seq.5646.20	RNASE6	1.54	1.04 - 2.27	0.0292	0.5400
345	seq.14249.68	MID2	1.61	1.05 - 2.46	0.0294	0.5400
346	seq.19289.29	UROD	1.51	1.04 - 2.19	0.0294	0.5400
347	seq.23258.56	COA7	1.54	1.04 - 2.27	0.0294	0.5400
348	seq.20134.27	MARCO	1.51	1.04 - 2.19	0.0296	0.5400
349	seq.24407.31	KIF3C	0.66	0.45 - 0.96	0.0297	0.5400
350	seq.24711.21	SHOX	1.46	1.04 - 2.07	0.0298	0.5400
351	seq.20591.48	GLDN	1.48	1.04 - 2.10	0.0299	0.5400
352	seq.5701.81	CLEC3B	0.67	0.46 - 0.96	0.0300	0.5400
353	seq.2523.31	CCL5	1.61	1.05 - 2.48	0.0302	0.5400
354	seq.11646.4	CHST9	1.46	1.04 - 2.06	0.0302	0.5400
355	seq.5751.14	LETMD1	1.45	1.04 - 2.04	0.0303	0.5400
356	seq.10531.18	NRAS	1.5	1.04 - 2.16	0.0303	0.5400
357	seq.7809.22	SPATA9	1.49	1.04 - 2.14	0.0306	0.5400
358	seq.9177.6	FAM3B	1.5	1.04 - 2.18	0.0306	0.5400
359	seq.4718.5	PPIB	0.64	0.43 - 0.96	0.0307	0.5400
360	seq.3122.6	DIABLO	1.49	1.04 - 2.13	0.0308	0.5400
361	seq.10030.8	MUTYH	1.47	1.04 - 2.09	0.0308	0.5400
362	seq.14227.21	MYL6B	1.47	1.04 - 2.09	0.0308	0.5400

363	seq.7243.8	IFNB1	1.47	1.04 - 2.09	0.0308	0.5400
364	seq.9265.10	GLIPR1	1.45	1.03 - 2.03	0.0309	0.5400
365	seq.10514.5	PTGDS	1.47	1.04 - 2.07	0.0309	0.5400
366	seq.5121.3	SEMA6B	1.47	1.03 - 2.10	0.0318	0.5500
367	seq.12557.18	NOVA1	1.49	1.04 - 2.13	0.0318	0.5500
368	seq.10575.31	PUF60	1.54	1.04 - 2.28	0.0318	0.5500
369	seq.16315.105	FLT1	0.59	0.37 - 0.96	0.0322	0.5600
370	seq.24892.8	SHCBP1L	1.49	1.03 - 2.16	0.0322	0.5600
371	seq.10490.3	RPN1	1.5	1.03 - 2.16	0.0322	0.5600
372	seq.13492.44	HSPA9	1.46	1.03 - 2.06	0.0324	0.5600
373	seq.15594.47	HTRA1	1.49	1.03 - 2.15	0.0326	0.5600
374	seq.5508.62	CTSD	1.46	1.03 - 2.07	0.0327	0.5600
375	seq.16322.10	MZB1	1.5	1.03 - 2.16	0.0328	0.5600
376	seq.2809.25	ADAMTS4	1.45	1.03 - 2.05	0.0328	0.5600
377	seq.2436.49	CXCL16	1.55	1.04 - 2.31	0.0329	0.5600
378	seq.13744.37	IL3RA	1.47	1.03 - 2.08	0.0329	0.5600
379	seq.20928.39	TCEAL7	1.51	1.03 - 2.21	0.0331	0.5600
380	seq.6217.23	QSOX1	1.49	1.03 - 2.14	0.0333	0.5600
381	seq.9728.4	CCNB1IP1	1.52	1.03 - 2.24	0.0334	0.5600
382	seq.5463.22	GAS1	1.46	1.03 - 2.07	0.0334	0.5600
383	seq.16585.16	B3GNT4	1.46	1.03 - 2.08	0.0334	0.5600
384	seq.4964.67	ERAP1	0.69	0.49 - 0.97	0.0335	0.5600
385	seq.3208.2	MATN3	0.66	0.45 - 0.97	0.0336	0.5600
386	seq.16609.106	KIRREL2	1.45	1.03 - 2.05	0.0339	0.5600
387	seq.23683.79	TRIM55	1.46	1.03 - 2.06	0.0339	0.5600
388	seq.4314.12	DCTPP1	1.5	1.03 - 2.18	0.0341	0.5600
389	seq.11104.13	CHI3L1	1.45	1.03 - 2.06	0.0344	0.5600
390	seq.13651.54	ZFP91	1.48	1.03 - 2.14	0.0345	0.5600
391	seq.5708.1	LEAP2	1.64	1.04 - 2.59	0.0345	0.5600
392	seq.21647.9	SRM	1.47	1.03 - 2.11	0.0346	0.5600
393	seq.5069.9	CD55	1.51	1.03 - 2.21	0.0347	0.5600
394	seq.8406.17	IGF1	1.45	1.03 - 2.05	0.0349	0.5600
395	seq.7097.8	CDY1	1.45	1.03 - 2.06	0.0349	0.5600
396	seq.4464.10	SIGLEC1	1.46	1.03 - 2.08	0.0350	0.5600
397	seq.3175.51	ADAMTS13	0.69	0.49 - 0.97	0.0351	0.5600
398	seq.18871.24	AIF1L	1.45	1.03 - 2.04	0.0354	0.5600
399	seq.13118.5	SMOC1	1.45	1.03 - 2.04	0.0355	0.5600
400	seq.2652.15	PLAUR	1.52	1.03 - 2.23	0.0356	0.5600
401	seq.15513.108	PRSS8	1.54	1.03 - 2.30	0.0358	0.5600
402	seq.11566.48	KRT72	1.45	1.03 - 2.06	0.0358	0.5600
403	seq.4179.57	YWHAG	1.59	1.03 - 2.46	0.0363	0.5600
404	seq.3292.75	CD48	1.45	1.02 - 2.05	0.0366	0.5600
405	seq.7926.13	SPINT3	1.43	1.02 - 1.99	0.0367	0.5600
406	seq.10048.7	CBFB	1.47	1.02 - 2.10	0.0368	0.5600
407	seq.19383.131	CINP	1.46	1.02 - 2.07	0.0368	0.5600
408	seq.5934.1	FTH1|FTL	1.49	1.02 - 2.17	0.0369	0.5600

409	seq.14603.51	KIAA0040	1.47	1.02 - 2.11	0.0372	0.5600
410	seq.25117.17	PGM5	1.5	1.02 - 2.18	0.0373	0.5600
411	seq.9889.42	AFAP1L1	1.45	1.02 - 2.05	0.0373	0.5600
412	seq.21658.15	PHB2	1.45	1.02 - 2.07	0.0374	0.5600
413	seq.6366.38	TXNDC15	1.5	1.02 - 2.20	0.0375	0.5600
414	seq.10746.24	DKK3	1.45	1.02 - 2.06	0.0376	0.5600
415	seq.21106.206	CETN1	1.42	1.02 - 1.99	0.0377	0.5600
416	seq.3151.6	IL2RA	1.42	1.02 - 1.98	0.0377	0.5600
417	seq.11347.9	TALDO1	1.46	1.02 - 2.10	0.0377	0.5600
418	seq.13449.25	SF3B4	1.76	1.03 - 3.00	0.0378	0.5600
419	seq.21141.9	LRATD2	1.44	1.02 - 2.02	0.0378	0.5600
420	seq.22049.24	YWHAH	1.62	1.03 - 2.56	0.0379	0.5600
421	seq.20568.3	FREM1	1.47	1.02 - 2.10	0.0379	0.5600
422	seq.15324.58	FTL	1.49	1.02 - 2.16	0.0380	0.5600
423	seq.10818.36	SMPD1	1.45	1.02 - 2.05	0.0380	0.5600
424	seq.15667.39	BMP4	1.45	1.02 - 2.07	0.0381	0.5600
425	seq.17696.1	TICAM2	1.64	1.03 - 2.63	0.0381	0.5600
426	seq.14636.25	RIDA	1.44	1.02 - 2.04	0.0381	0.5600
427	seq.9053.16	CRIP2	1.5	1.02 - 2.20	0.0381	0.5600
428	seq.11838.130	PIEZO1	1.7	1.03 - 2.80	0.0381	0.5600
429	seq.3389.7	SERPINA5	2.64	1.05 - 6.63	0.0383	0.5600
430	seq.14069.61	CA4	1.78	1.03 - 3.09	0.0384	0.5600
431	seq.23365.7	FNDC8	1.51	1.02 - 2.24	0.0384	0.5600
432	seq.6077.63	ADA2	1.48	1.02 - 2.13	0.0384	0.5600
433	seq.22589.3	FZD2	1.46	1.02 - 2.09	0.0385	0.5600
434	seq.9365.91	EYS	1.43	1.02 - 2.01	0.0385	0.5600
435	seq.4874.3	ANG	1.47	1.02 - 2.11	0.0389	0.5600
436	seq.12787.47	ZNF134	1.53	1.02 - 2.31	0.0391	0.5600
437	seq.13998.26	ADSS1	1.42	1.02 - 2.00	0.0393	0.5600
438	seq.5452.71	ASGR1	1.44	1.02 - 2.05	0.0393	0.5600
439	seq.18242.8	SPA17	1.47	1.02 - 2.13	0.0395	0.5600
440	seq.5635.66	TREM2	1.44	1.02 - 2.03	0.0395	0.5600
441	seq.2201.17	COL18A1	1.47	1.02 - 2.11	0.0395	0.5600
442	seq.19561.216	PLXND1	1.45	1.02 - 2.07	0.0396	0.5600
443	seq.6590.54	NRP2	1.47	1.02 - 2.12	0.0397	0.5600
444	seq.5085.18	IL20RA	1.42	1.02 - 2.00	0.0397	0.5600
445	seq.6550.4	ICAM4	0.68	0.47 - 0.98	0.0397	0.5600
446	seq.9243.10	LYG1	1.41	1.02 - 1.95	0.0398	0.5600
447	seq.21528.12	BMF	1.48	1.02 - 2.15	0.0398	0.5600
448	seq.5002.76	MMP14	1.42	1.02 - 1.99	0.0399	0.5600
449	seq.8099.42	SPON2	1.45	1.02 - 2.07	0.0399	0.5600
450	seq.14124.6	EFNA2	1.44	1.02 - 2.04	0.0403	0.5600
451	seq.9248.36	MYDGF	1.43	1.02 - 2.01	0.0403	0.5600
452	seq.15363.32	APOA5	0.69	0.48 - 0.98	0.0404	0.5600
453	seq.6973.111	IGF2	0.66	0.44 - 0.98	0.0404	0.5600
454	seq.7139.14	SLITRK4	1.43	1.02 - 2.01	0.0405	0.5600

455	seq.7241.12	CD247	1.53	1.02 - 2.29	0.0405	0.5600
456	seq.18198.51	HBQ1	1.44	1.02 - 2.03	0.0405	0.5600
457	seq.9300.13	NECTIN1	1.44	1.02 - 2.03	0.0405	0.5600
458	seq.6520.87	MGP	1.47	1.02 - 2.13	0.0406	0.5600
459	seq.13122.19	FLRT2	1.44	1.02 - 2.04	0.0407	0.5600
460	seq.17682.1	CD46	1.43	1.02 - 2.01	0.0408	0.5600
461	seq.16588.10	HSPA5	1.45	1.02 - 2.07	0.0410	0.5600
462	seq.8407.84	NOTCH2	0.69	0.48 - 0.98	0.0410	0.5600
463	seq.21537.33	SCLY	1.41	1.01 - 1.97	0.0410	0.5600
464	seq.22154.37	GATAD1	1.52	1.02 - 2.27	0.0412	0.5600
465	seq.9039.47	TOR1AIP1	1.49	1.02 - 2.18	0.0415	0.5600
466	seq.4342.10	ICAM1	1.45	1.01 - 2.07	0.0416	0.5600
467	seq.12895.28	DGKB	1.44	1.01 - 2.05	0.0416	0.5600
468	seq.18241.18	CPOX	1.42	1.01 - 1.98	0.0420	0.5600
469	seq.8368.102	TNFRSF1B	1.48	1.01 - 2.16	0.0422	0.5600
470	seq.9369.174	LRRC4C	1.42	1.01 - 2.00	0.0422	0.5600
471	seq.14273.19	PREP	1.47	1.01 - 2.13	0.0423	0.5600
472	seq.23569.53	STARD10	1.44	1.01 - 2.04	0.0424	0.5600
473	seq.8028.22	SPINK5	1.46	1.01 - 2.09	0.0424	0.5600
474	seq.20549.1	IZUMO4	0.71	0.51 - 0.99	0.0424	0.5600
475	seq.23593.14	MB21D2	0.67	0.46 - 0.99	0.0426	0.5600
476	seq.6521.35	NPTX2	1.45	1.01 - 2.08	0.0426	0.5600
477	seq.17832.12	IDI2	1.47	1.01 - 2.14	0.0427	0.5600
478	seq.9327.3	LYPD1	1.42	1.01 - 1.99	0.0427	0.5600
479	seq.21236.12	TMED9	1.44	1.01 - 2.06	0.0428	0.5600
480	seq.5526.53	TNFRSF18	1.45	1.01 - 2.09	0.0428	0.5600
481	seq.15542.19	CKMT1A	1.45	1.01 - 2.08	0.0428	0.5600
482	seq.13476.16	KIN	1.4	1.01 - 1.95	0.0428	0.5600
483	seq.2948.58	GHR	0.69	0.49 - 0.99	0.0428	0.5600
484	seq.2358.19	FLT4	0.7	0.50 - 0.99	0.0429	0.5600
485	seq.5613.75	LAT2	1.46	1.01 - 2.09	0.0429	0.5600
486	seq.12731.12	PLEKHA7	1.42	1.01 - 1.98	0.0429	0.5600
487	seq.23282.19	RAB37	1.47	1.01 - 2.13	0.0429	0.5600
488	seq.13470.43	PTH1R	0.65	0.43 - 0.99	0.0430	0.5600
489	seq.11391.69	MVK	0.66	0.44 - 0.99	0.0431	0.5600
490	seq.18342.2	PSAT1	1.43	1.01 - 2.04	0.0431	0.5600
491	seq.22810.41	TIPIN	0.66	0.44 - 0.99	0.0432	0.5600
492	seq.9361.7	PCDHAC2	1.41	1.01 - 1.97	0.0432	0.5600
493	seq.3082.9	ULBP2	1.44	1.01 - 2.06	0.0435	0.5600
494	seq.15635.4	SMOC2	1.43	1.01 - 2.03	0.0436	0.5600
495	seq.9573.108	TXNDC11	1.47	1.01 - 2.12	0.0436	0.5600
496	seq.12612.37	PSMB1	1.47	1.01 - 2.14	0.0437	0.5600
497	seq.23627.3	SDHAF4	1.48	1.01 - 2.16	0.0437	0.5600
498	seq.18909.11	EXOSC8	1.44	1.01 - 2.06	0.0437	0.5600
499	seq.3009.3	TGFBR3	1.45	1.01 - 2.08	0.0439	0.5600
500	seq.25465.42	MLIP	1.44	1.01 - 2.05	0.0439	0.5600

501	seq.19572.10	CEBPG	0.65	0.42 - 0.99	0.0440	0.5600
502	seq.20511.3	F11R	1.44	1.01 - 2.05	0.0440	0.5600
503	seq.14186.13	RFFL	1.45	1.01 - 2.07	0.0441	0.5600
504	seq.4922.13	CCL19	1.43	1.01 - 2.02	0.0442	0.5600
505	seq.9015.1	PRG3	1.42	1.01 - 2.00	0.0442	0.5600
506	seq.9409.11	TPSAB1	1.44	1.01 - 2.04	0.0444	0.5600
507	seq.22985.160	IGFBP2	1.43	1.01 - 2.03	0.0444	0.5600
508	seq.12817.1	GEM	0.62	0.39 - 0.99	0.0445	0.5600
509	seq.18303.39	NRGN	1.83	1.02 - 3.31	0.0445	0.5600
510	seq.4906.35	F5	1.44	1.01 - 2.06	0.0446	0.5600
511	seq.21491.7	AOC3	1.46	1.01 - 2.12	0.0450	0.5600
512	seq.5255.22	PDE4D	0.67	0.46 - 0.99	0.0450	0.5600
513	seq.14025.18	TNFRSF9	1.43	1.01 - 2.02	0.0451	0.5600
514	seq.2773.50	IL10	1.42	1.01 - 2.00	0.0451	0.5600
515	seq.10419.1	SCARA5	1.43	1.01 - 2.02	0.0453	0.5600
516	seq.19257.11	CLTA	1.46	1.01 - 2.10	0.0454	0.5600
517	seq.3470.1	SELE	1.47	1.01 - 2.15	0.0456	0.5600
518	seq.6383.90	TLL1	1.5	1.01 - 2.24	0.0456	0.5600
519	seq.5903.91	HSPA8	1.41	1.01 - 1.96	0.0457	0.5600
520	seq.8929.7	BDP1	0.67	0.45 - 0.99	0.0457	0.5600
521	seq.23618.23	ZNF593	1.48	1.01 - 2.18	0.0457	0.5600
522	seq.15381.45	DDR2	1.46	1.01 - 2.12	0.0458	0.5600
523	seq.3045.72	PTN	1.43	1.01 - 2.03	0.0459	0.5600
524	seq.9974.8	DLL3	1.44	1.01 - 2.06	0.0459	0.5600
525	seq.8660.5	OLFML3	1.45	1.01 - 2.08	0.0459	0.5600
526	seq.12695.62	KLHL12	1.41	1.01 - 1.98	0.0459	0.5600
527	seq.7131.8	CETP	1.41	1.01 - 1.97	0.0459	0.5600
528	seq.12976.49	SORBS3	1.5	1.01 - 2.24	0.0462	0.5600
529	seq.22841.16	ZNF696	0.68	0.46 - 0.99	0.0462	0.5600
530	seq.10990.21	LRRK2	0.6	0.37 - 0.99	0.0464	0.5600
531	seq.20535.68	IL17RE	1.46	1.01 - 2.11	0.0466	0.5600
532	seq.11313.100	PCBD1	1.47	1.01 - 2.15	0.0466	0.5600
533	seq.10339.48	ENO2	1.49	1.01 - 2.21	0.0467	0.5600
534	seq.13575.40	DVL2	1.49	1.01 - 2.22	0.0469	0.5600
535	seq.12738.43	NISCH	0.68	0.47 - 0.99	0.0469	0.5600
536	seq.6911.103	CLEC6A	1.39	1.00 - 1.93	0.0470	0.5600
537	seq.8520.8	ADAM30	1.41	1.00 - 1.99	0.0471	0.5600
538	seq.15304.1	REG3A	1.41	1.00 - 1.98	0.0473	0.5700
539	seq.6361.49	PTPRR	1.39	1.00 - 1.93	0.0474	0.5700
540	seq.7880.9	PSMA1	0.69	0.48 - 1.00	0.0479	0.5700
541	seq.10396.6	MCL1	1.49	1.00 - 2.21	0.0481	0.5700
542	seq.21572.91	XYLT2	1.47	1.00 - 2.15	0.0485	0.5700
543	seq.9188.119	CXCL9	1.38	1.00 - 1.90	0.0488	0.5700
544	seq.12684.5	SCIN	1.47	1.00 - 2.16	0.0488	0.5700
545	seq.9759.13	IDO1	1.48	1.00 - 2.20	0.0488	0.5700
546	seq.13668.44	TYRO3	1.43	1.00 - 2.04	0.0488	0.5700

547	seq.25051.104	FTO	1.41	1.00 - 1.99	0.0491	0.5700
548	seq.3042.7	MB	1.42	1.00 - 2.03	0.0492	0.5700
549	seq.21107.5	NDUFAF1	1.4	1.00 - 1.97	0.0492	0.5700
550	seq.8986.2	PHGDH	1.41	1.00 - 1.98	0.0492	0.5700
551	seq.21604.2	CUTA	1.43	1.00 - 2.05	0.0493	0.5700
552	seq.19282.3	RAB13	1.43	1.00 - 2.04	0.0494	0.5700
553	seq.24903.7	KLHL14	1.41	1.00 - 1.98	0.0495	0.5700
554	seq.5879.51	DCTN2	1.42	1.00 - 2.00	0.0496	0.5700
555	seq.15589.1	GC	0.71	0.50 - 1.00	0.0497	0.5700
556	seq.3025.50	FGF2	1.46	1.00 - 2.12	0.0497	0.5700
557	seq.8009.121	SURF1	1.42	1.00 - 2.00	0.0498	0.5700
558	seq.15482.12	A2ML1	1.39	1.00 - 1.92	0.0498	0.5700

Table E2: Significant Protein Associations with ILA in Discovery Cohort (CGS-PF) with Adjustment for *MUC5B* and Low Lymphocyte Telomere Length

	MUC5B*-Adjusted Association with ILA			**Telomere Length-Adjusted Association with ILA†**		
Protein	**Odds Ratio**	**Confidence Interval**	**P-value**	**Odds Ratio**	**Confidence Interval**	**P-value**
SFTPD	5.49	2.72-11.1	2.0 E-06	3.70	2.02-6.79	2.3E-05
SFTPB	3.59	2.05-6.28	7.4 E-06	3.80	2.12-6.81	7.1E-06
GDF15	3.56	1.98-6.42	2.3 E-05	3.50	1.92-6.37	4.0E-05
WFDC2	2.14	1.33-3.45	0.002	2.42	1.48-3.95	0.0004
GCHFR	2.10	1.31-3.36	0.002	2.09	1.30-3.34	0.002
SFN	1.90	1.18-3.06	0.008	2.01	1.26-3.21	0.003
CDCP1	1.89	1.19-2.98	0.007	1.93	1.20-3.11	0.006

Definition of abbreviations: CGS-PF = Clinical Genetics and Screening for Pulmonary Fibrosis; GDF15 = growth differentiation factor 15; SFTPD = surfactant protein D; SFTPB = surfactant protein B; WFDC2 = WAP four-disulfide core domain protein 2; CDCP1 = CUB domain-containing protein 1; GCHFR = GTP cyclohydrolase 1 feedback regulatory protein; SFN = stratifin.

*Analyses used multivariable logistic regression models with covariates age, gender, smoking status, and at least one copy of *MUC5B* promotor polymorphism.

†Analyses used multivariable logistic regression models with covariates age, gender, smoking status, and low lymphocyte telomere length.

Table E3: Assessment of Effect Modification of Demographics on Validated ILA-Protein Associations in the Combined Cohorts (P-Values for Interaction Term)

Protein	Age	Gender	Smoking History
SFTPD	0.6	0.3	0.8
GDF15	0.2	0.1	0.6
SFTPB	0.6	0.6	0.6
WFDC2	0.3	0.2	0.5
CDCP1	0.04	0.9	0.5
SFN	0.7	0.4	0.05

*Multivariable regression models were used with variable-protein multiplicative interaction terms along with age, gender, smoking status, and protein. P-values are for the interaction term are shown.

Table E4: Stratified Analyses for ILA-Protein Associations with Effect Modification

		CDCP1		SFN	
Stratum	**N**	**OR (95% CI)**	**P value**	**OR (95% CI)**	**P value**
Age <60	120	2.9 (1.6-5.8)	0.001	-	-
Age ≥60	116	1.9 (1.2-3.1)	0.01	-	-
Never smoker	149	-	-	4.2 (2.2-9.3)	<0.001
Ever smoker	87	-	-	1.5 (0.9-2.7)	0.1

Multivariable regression models were used with the protein, age, gender, and smoking history except when stratified on the variable.

Table E5: Correlation of Proteins Significantly Associated with ILA in Combined Cohorts

	GDF15	SFTPD	SFTPB	WFDC2	CDCP1	SFN
GDF15	1.00	0.43	0.40	0.78	0.63	0.56
SFTPD	0.43	1.00	0.48	0.46	0.36	0.31
SFTPB	0.40	0.48	1.00	0.49	0.43	0.34
WFDC2	0.78	0.46	0.49	1.00	0.57	0.55
CDCP1	0.63	0.36	0.43	0.57	1.00	0.42
SFN	0.56	0.31	0.34	0.55	0.42	1.00

Figure E1: Detection of ILA Using *MUC5B* in CGS-PF. Receiver operating characteristic curves and area under the curves (AUC) for the detection of ILA in CGS-PF using multivariable logistic regression with different combinations of demographics (age, gender, smoking status), the *MUC5B* genotype, and the six validated proteins significantly associated with ILA (GDF15, SFTPB, SFTPD, WFDC2, CDCP1, SFN). There is only a borderline statistical difference between the model with demographics and the *MUC5B* genotype and the model with demographics alone (p=0.05). The model with demographics and the top 6 validated proteins is significantly better than demographics and the *MUC5B* genotype (p=0.0008), and there is no difference when adding *MUC5B* genotype to demographics and the top 6 proteins (p=0.1).

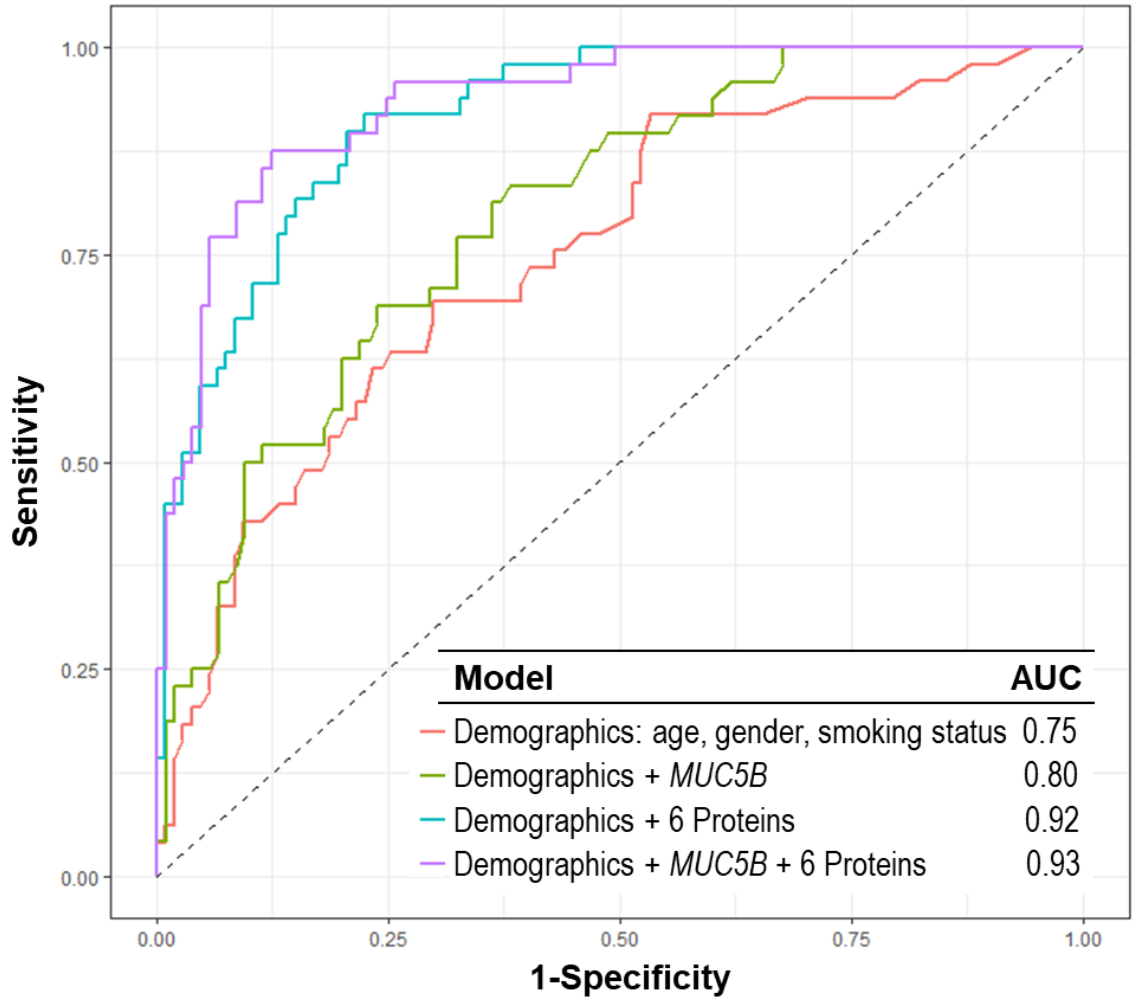


Figure E2: Detection of ILA Using Lymphocyte Telomere Length in CGS-PF. Receiver operating characteristic curves and area under the curves (AUC) for the detection of ILA in CGS-PF using multivariable logistic regression with different combinations of demographics (age, gender, smoking status), reduced lymphocyte telomere length (TL), and the six validated proteins significantly associated with ILA (GDF15, SFTPB, SFTPD, WFDC2, CDCP1, SFN). There is a borderline difference between the model with demographics and TL and the model with demographics alone (p=0.07). The model with demographics and the top 6 validated proteins is significantly better than demographics and TL (p=0.0005), and there is no difference when adding TL to demographics and the top 6 proteins (p=0.8).

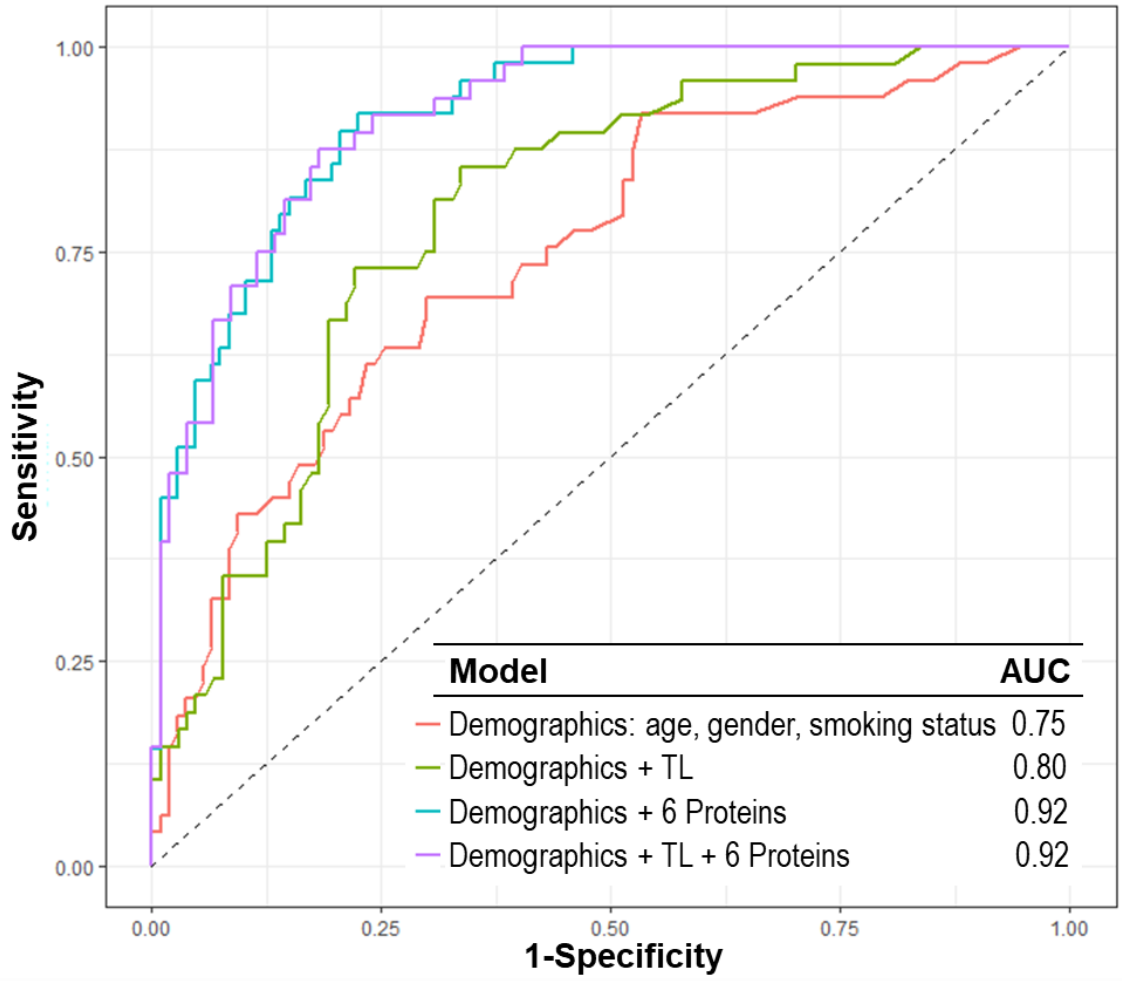


Figure E3: Detection of ILA Using PFTs in CGS-PF. Receiver operating characteristic curves and area under the curves (AUC) for the detection of ILA in CGS-PF using multivariable logistic regression with different combinations of demographics (age, gender, smoking status), PFTs (FVC and DLCO percent predicted), and the six validated proteins significantly associated with ILA (GDF15, SFTPB, SFTPD, WFDC2, CDCP1, SFN). The model with demographics and PFTs is significantly better than the model with demographics alone (p=0.04), the model with demographics and the top 6 validated proteins is significantly better than demographics and PFTs (p=0.002), and there is no difference when adding PFTs to demographics and the top 6 proteins (p=0.5).

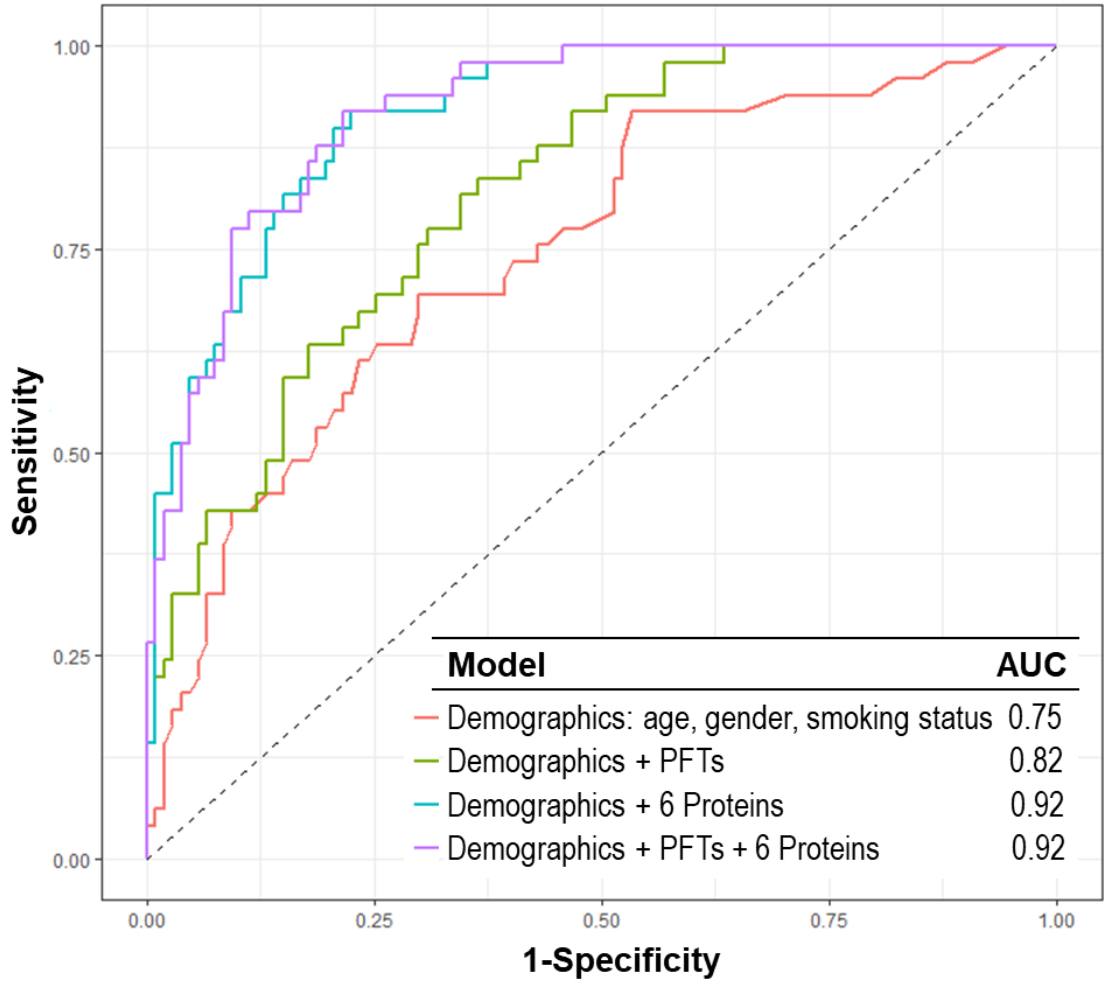


Table E6: Test Characteristics of Various Thresholds of the Simplified Model Applied to Combined Cohorts. The table below lists all of the local maximas of the ROC curve of the AUC curve from the test set and corresponding test performance characteristics. Positive predictive value (PPV) and negative predictive value (NPV) were calculated based on the combined cohorts of 237 participants (62 with ILA and 175 without). The test characteristics corresponding to the Youden's index are highlighted in the table in bold.

Sensitivity	Specificity	PPV	NPV
100%	61%	48%	100%
92%	67%	50%	96%
85%	75%	54%	93%
77%	**90%**	**72%**	**92%**
69%	91%	73%	89%
54%	94%	76%	85%
46%	97%	85%	84%
31%	100%	100%	80%

Figure E7: 2x2 Table of the Simplified Model Threshold with 100% Sensitivity Applied to Combined Cohorts. 2x2 table of the final model applied to the combined cohorts at a threshold with 100% sensitivity, and thus 100% negative predictive value (NPV). The corresponding specificity and positive predictive value (PPV) are shown. This gives 107 true negatives among the 237 participants tested. Therefore, for every 2.2 people tested, 1 CT scan would be prevented.

		Actual (CT)			
		Positive	Negative		
Test (3 Protein + Age Model)	Positive	62	68	130	PPV 0.48
	Negative	0	107	107	NPV 1.0
		62	175	237	
		Sensitivity 1.0	Specificity 0.61		