2. Brochard L, Slutsky A, Pesenti A. Mechanical ventilation to minimize progression of lung injury in acute respiratory failure. *Am J Respir Crit Care Med* 2017;195:438–442.
3. Venado A, Witt LJ, Kallianos K, Wolters PJ. Diaphragmatic atrophy may limit progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2020;201:e72–e73.
4. Slutsky AS. Neuromuscular blocking agents in ARDS. *N Engl J Med* 2010;363:1176–1180.
5. Moss M, Huang DT, Brower RG, Ferguson ND, Ginde AA, Gong MN, *et al*.; National Heart, Lung, and Blood Institute PETAL Clinical Trials Network. Early neuromuscular blockade in the acute respiratory distress syndrome. *N Engl J Med* 2019;380:1997–2008.
6. Puthucheary Z, Rawal J, Ratnayake G, Harridge S, Montgomery H, Hart N. Neuromuscular blockade and skeletal muscle weakness in critically ill patients: time to rethink the evidence? *Am J Respir Crit Care Med* 2012;185:911–917.
7. Sklar MC, Dres M, Fan E, Rubenfeld GD, Scales DC, Herridge MS, *et al*. Association of low baseline diaphragm muscle mass with prolonged mechanical ventilation and mortality among critically ill adults. *JAMA Netw Open* 2020;3:e1921520.
8. Yoshida T, Uchiyama A, Matsuura N, Mashimo T, Fujino Y. The comparison of spontaneous breathing and muscle paralysis in two different severities of experimental lung injury. *Crit Care Med* 2013;41:536–545.
9. Goodman SN. Toward evidence-based medical statistics: 2. The Bayes factor. *Ann Intern Med* 1999;130:1005–1013.
10. McCool FD, Conomos P, Benditt JO, Cohn D, Sherman CB, Hoppin FG Jr. Maximal inspiratory pressures and dimensions of the diaphragm. *Am J Respir Crit Care Med* 1997;155:1329–1334.





## Preclinical Pulmonary Fibrosis Circulating Protein Biomarkers

*To the Editor*:

Idiopathic pulmonary fibrosis (IPF) is characterized by progressive, irreversible scarring of the lung parenchyma that can require invasive diagnostic testing (1). Interstitial lung abnormalities (ILAs) have been described in the general population (2). Among asymptomatic first-degree relatives of patients with familial interstitial pneumonia (FIP), 14% have radiologic ILAs and 35% have interstitial abnormalities on biopsy (3). In the Framingham population, fibrotic ILAs were present in 1.8% of subjects ⩾50 years of age (4) and associated with increased risk of death (5, 6), suggesting ILAs may be a harbinger of IPF.

Because ILAs include ground glass and diffuse centrilobular nodularity can be present without fibrosis of the lung, we created the term "preclinical pulmonary fibrosis" (PrePF) (7) to identify first-degree relatives of patients with FIP (a high-risk cohort) not known to have interstitial lung disease who have features of lung fibrosis on high-resolution computed tomography.

We used proteomic analyses of plasma to identify circulating markers of IPF and then determine if IPF-associated proteins are predictive of PrePF. Some of the results of these studies have been previously reported in the form of an abstract (8).

### Methods and Findings

Subjects with IPF (American Thoracic Society and European Respiratory Society criteria) (1) and first-degree relatives of patients with FIP with no known interstitial lung disease were recruited at University of Colorado, National Jewish Health, and Vanderbilt University (COMIRB #15-1147; NJH IRB 1441a; Vanderbilt IRB #020343) (7). PrePF was defined as evidence of fibrosis (reticular abnormality or traction bronchiectasis, or honeycombing) on high-resolution computed tomography (7).

Samples were proteolyzed with the iST Kit in 96-well format (PreOmics) and analyzed by mass spectrometry (Q Exactive HF, Ultimate 3000; ThermoFisher) in a data-independent acquisition mode (9, 10). Protein identification was performed by peptide mapping (Spectronaut Pulsar) to an in-house plasma spectral library at a precursor Q value cutoff of 0.01 and using the match-between run option at a 0.1 percentile threshold. Label-free quantification was performed on the intensities of summed fragment spectra.

Raw intensity data were normalized via a local (retention time-dependent) method and log transformed (9). Intensities were compared in IPF versus unaffected plasma, controlling for age, sex, and family relatedness in a linear mixed-effects model. Analyses were performed in the RStudio (v.3.2.2) and the R (v.3.5.3)

**Table 1.** IPF versus No Fibrosis, Significant Proteins in Plasma

| Protein | Coefficient | *P* Value | FDR |
|---|---|---|---|
| GSN | −0.28 | $2.82 \times 10^{-12}$ | $1.04 \times 10^{-9}$ |
| C1QC | −0.33 | $1.52 \times 10^{-6}$ | 0.0003 |
| KNG1 | −0.18 | $3.33 \times 10^{-6}$ | 0.0004 |
| CLEC3B | −0.31 | $2.35 \times 10^{-5}$ | 0.0022 |
| A2M | 0.36 | $5.44 \times 10^{-5}$ | 0.0025 |
| APOA4 | −0.32 | $4.03 \times 10^{-5}$ | 0.0025 |
| FBLN1 | 0.25 | $5.94 \times 10^{-5}$ | 0.0025 |
| YTHDC2 | −0.25 | $5.00 \times 10^{-5}$ | 0.0025 |
| CRKL | −0.30 | $5.99 \times 10^{-5}$ | 0.0025 |
| SPARC | 0.59 | $7.34 \times 10^{-5}$ | 0.0027 |
| PRSS3 | 0.51 | 0.0001 | 0.0041 |
| ALB | −0.14 | 0.0002 | 0.0051 |
| LBP | 0.27 | 0.0003 | 0.0082 |
| APOA2 | −0.22 | 0.0006 | 0.015 |
| BASP1 | −0.42 | 0.0007 | 0.011 |
| APOA1 | −0.21 | 0.0010 | 0.021 |
| S100A8 | −0.83 | 0.0010 | 0.021 |
| CRISP3 | −0.50 | 0.0010 | 0.021 |
| CTBS | 0.34 | 0.0012 | 0.024 |
| C9 | 0.24 | 0.0014 | 0.024 |
| PGLYRP2 | −0.20 | 0.0014 | 0.024 |
| S100A9 | −0.65 | 0.0014 | 0.024 |
| FGG | 0.20 | 0.0015 | 0.025 |
| HP | 0.33 | 0.0023 | 0.035 |
| IGKV1D_13 | 0.76 | 0.0028 | 0.042 |

*Definition of abbreviations*: FDR = false discovery rate; IPF = idiopathic pulmonary fibrosis.
Differentially detected proteins discovered in IPF versus no lung fibrosis plasma protein analysis are shown. Analysis was controlled for age, sex, and family relatedness in a linear mixed-effects model; raw *P* values are listed as well as adjustment for multiple testing.

Supported by the U.S. Department of Defense (W81XWH-17-1-0597) (D.A.S.) and NIH R01-HL097163 (D.A.S.), UH2/3-HL123442 (D.A.S.), P01-HL092870 (D.A.S.), and K23-HL136785 (S.K.M.).

Originally Published in Press as DOI: 10.1164/rccm.202003-0724LE on August 5, 2020

**Table 2.** PrePF versus No Fibrosis, Plasma Protein Analysis

| Protein | Protein Name | Coefficient | 95% CI | *P* Value | FDR |
|---|---|---|---|---|---|
| *GSN* | *Gelsolin* | *−0.14* | *−0.22 to −0.07* | *0.0002* | *0.003* |
| *S100A9* | *Protein S100-A9* | *−0.73* | *−1.11 to −0.35* | *0.0002* | *0.003* |
| *CRKL* | *Crk-like protein* | *−0.23* | *−0.37 to −0.10* | *0.0006* | *0.005* |
| *LBP* | *LPS-binding protein* | *0.21* | *0.08 to 0.35* | *0.0013* | *0.006* |
| *C1QC* | *Complement C1q subcomponent subunit C* | *−0.22* | *−0.35 to −0.09* | *0.0011* | *0.006* |
| *S100A8* | *Protein S100-A8* | *−0.67* | *−1.13 to −0.25* | *0.0021* | *0.009* |
| *BASP1* | *Brain acid soluble protein 1* | *−0.32* | *−0.55 to −0.10* | *0.0042* | *0.015* |
| *SPARC* | *SPARC* | *0.35* | *0.09 to 0.61* | *0.0075* | *0.024* |
| *APOA4* | *Apolipoprotein A-IV* | *−0.18* | *−0.32 to −0.05* | *0.0093* | *0.026* |
| *C9* | *Complement component C9* | *0.18* | *0.04 to 0.31* | *0.011* | *0.027* |
| *ALB* | *Serum albumin* | *−0.08* | *−0.15 to −0.02* | *0.014* | *0.031* |
| *CRISP3* | *Cysteine-rich secretory protein 3* | *−0.32* | *−0.61 to −0.04* | *0.023* | *0.049* |
| APOA1 | Apolipoprotein A-I | −0.12 | −0.24 to −0.01 | 0.026 | 0.050 |
| PRSS3 | Trypsin-3 | 0.27 | 0.03 to 0.51 | 0.029 | 0.051 |
| YTHDC2 | Probable ATP-dependent RNA helicase YTHDC2 | −0.12 | −0.24 to −0.01 | 0.034 | 0.058 |
| PGLYRP2 | *N*-acetylmuramoyl-L-alanine amidase | −0.13 | −0.25 to −0.01 | 0.038 | 0.057 |
| CLEC3B | Tetranectin | −0.14 | −0.27 to −0.01 | 0.044 | 0.062 |
| APOA2 | Apolipoprotein A-II | −0.12 | −0.23 to −0.002 | 0.047 | 0.062 |
| A2M | Alpha-2-macroglobulin | 0.16 | 0.0 to 0.32 | 0.047 | 0.062 |
| CTBS | Di-*N*-acetylchitobiase | 0.13 | −0.05 to 0.31 | 0.147 | 0.184 |
| HP | Haptoglobin | 0.14 | −0.06 to 0.34 | 0.180 | 0.214 |
| FGG | Fibrinogen gamma chain | 0.06 | −0.06 to 0.18 | 0.327 | 0.371 |
| FBLN1 | Fibulin-1 | 0.05 | −0.06 to 0.17 | 0.351 | 0.381 |
| IGKV1D-13 | Ig kappa variable 1D-13 | 0.11 | −0.30 to 0.52 | 0.603 | 0.628 |
| KNG1 | Kininogen-1 | −0.006 | −0.08 to 0.07 | 0.874 | 0.873 |

*Definition of abbreviations*: CI = confidence interval; FDR = false discovery rate; PrePF = preclinical pulmonary fibrosis.
Proteins found to be significant in the analysis of subjects with idiopathic pulmonary fibrosis versus those without pulmonary fibrosis were examined in the plasma of subjects with PrePF versus those without pulmonary fibrosis. Analysis was controlled for age, sex, and family relatedness in a linear mixed-effects model; raw *P* values are listed as well as adjustment for multiple testing. Proteins with FDR < 0.05 are italicized.

environment using the *lme4* package. Proteins differentially detected (false discovery rate [FDR] < 0.05) in the IPF versus unaffected analysis were then tested in PrePF versus unaffected plasma using the same model.

Plasma samples were filtered to include the oldest unaffected member per family while maximizing the number of PrePF subjects. Top differentially detected, uncorrelated proteins were used to generate a predictive model. The *caret* R package was used to train models and receiver operating characteristic curves. Models were developed with only age and sex, and uncorrelated proteins were iteratively added. The model with the highest area under the curve (AUC) was selected.

A total of 328 samples were analyzed. Six were excluded because of hemolysis and six because of internal quality control failures, leaving 316 samples in the analysis. Of these, 34 had IPF, and 282 were first-degree relatives of patients with FIP (240 without radiologic lung fibrosis, 42 had PrePF). Those with PrePF or IPF were older and more likely to be male and have the IPF-associated *MUC5B* promoter variant rs35705950 (minor allele frequency 0.29 and 0.32, respectively, vs. 0.21 in unaffected subjects). Unaffected subjects were from families with FIP, so they were enriched for the *MUC5B* promoter variant compared with other studies (11).

Comparison of IPF ($n = 34$) to first-degree relatives without lung fibrosis ($n = 240$) revealed 25 plasma proteins differentially detected (FDR < 0.05) (Table 1). These 25 proteins were examined in the first-degree relatives with PrePF ($n = 42$) versus those without lung fibrosis ($n = 240$), revealing that 12 of the 25 plasma proteins remained differentially detected (GSN [gelsolin], S100-A9, CRKL [Crk-like protein], LBP [LPS-binding protein], C1QC [C1q subcomponent subunit C], S100A8, BASP1 [brain acid soluble protein 1], SPARC or osteonectin [secreted protein acidic and rich in cysteine], APOA4 [apolipoprotein A-IV], C9, ALB [albumin], and CRISP3 [cysteine-rich secretory protein 3]) (Table 2). The directionality of the plasma protein differences remained constant in terms of affected (IPF or PrePF) versus unaffected subjects.

Using the *cor* function in R and using a cutoff of 0.5, we found two correlated proteins (GSN and S100A8) and removed them from predictive modeling. Plasma samples were reviewed to create a data set with only one member per family while maximizing cases of PrePF, leaving 31 first-degree relatives with PrePF and 99 without evidence of lung fibrosis. The 12 proteins significant among subjects with PrePF were included in predictive modeling. When compared with a model using age and sex alone, including the top four proteins (S100A9, LBP, CRISP3, and CRKL) improved the model performance based on AUC. The AUC for the model including age, sex, and the four proteins was 0.86 (95% confidence interval [CI], 0.82–0.89) versus 0.77 (95% CI, 0.72–0.82) for the model using only age and sex; the lack of overlap in 95% CIs for the AUCs indicates improved predictive utility for the model including the four proteins (Figure 1). Adding *MUC5B* genotype to the models did not improve predictive ability (AUC, 0.79; 95% CI, 0.74–0.83). Adding *MUC5B* genotype to the aforementioned four proteins with age and sex did not improve the AUC (0.82; 95% CI, 0.78–0.86).

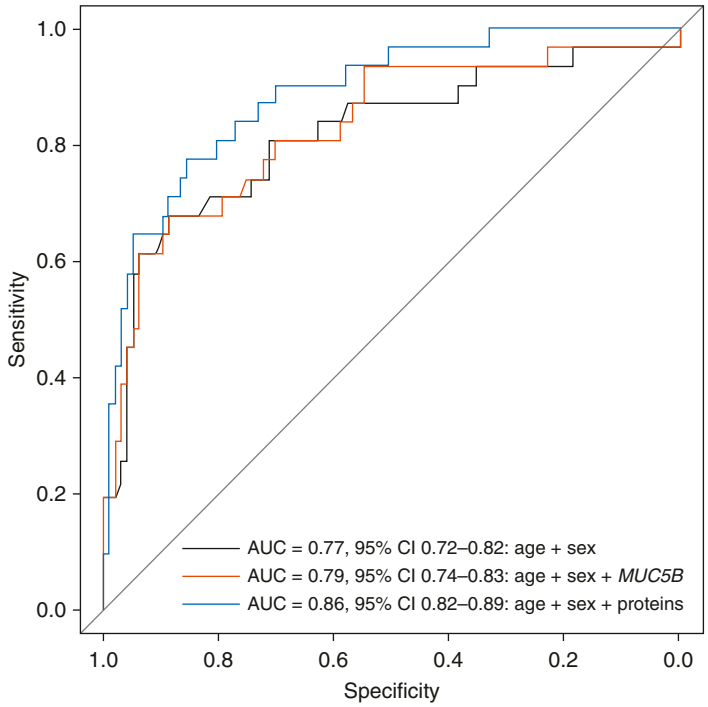


**Figure 1.** Predictive model for preclinical pulmonary fibrosis using top plasma proteins and patient characteristics. When compared with a model utilizing age and sex alone, including the top four proteins (S100A9, LBP, CRISP3, and CRKL) in addition to age and sex in a predictive model for preclinical pulmonary fibrosis improved the receiver operating characteristic curve performance based on comparing areas under the curve (AUCs). The AUC for the model including age, sex, and the four proteins was 0.86 (95% confidence interval [CI], 0.82–0.89; sensitivity, 0.77; specificity, 0.90; blue line) versus the AUC of 0.77 (95% CI, 0.72–0.82; sensitivity, 0.68; specificity, 0.89; black line) for the model using only age and sex. The negative predictive value and the positive predictive values of the age and sex model were 0.90 and 0.66 versus 0.93 and 0.70 for the model including age, sex, and protein levels. Adding *MUC5B* genotype to age and sex did not significantly improve the predictive ability of the model (red line) compared with including only age and sex (black line) or including the top four proteins (blue line).

To interrogate the consistency of the findings in a different blood sample, serum samples from first-degree relatives with PrePF ($n = 26$) and subjects without fibrosis ($n = 129$) were analyzed in a similar fashion to plasma proteins. Ten of the previously discovered 12 proteins were able to be detected in serum samples; S100A9 and S100A8 could not be measured in serum and so could not be compared. Nine of these 10 serum proteins showed consistent changes in directionality. Seven of those nine approached statistical significance but did not meet an FDR $< 0.05$ (ALB, GSN, C9, LBP, CRISP3, CRKL, SPARC); one did reach significance (C1QC, FDR = 0.02) (Table 3).

## Discussion

Circulating proteins have been associated with IPF but are not in clinical use (12). We focused on a high-risk cohort, first-degree

**Table 3.** Serum Protein Analyses, PrePF versus No Fibrosis

| Protein | Coefficient | *P* Value | FDR | Same Direction as Plasma? |
|---|---|---|---|---|
| ALB | −0.07 | 0.04 | 0.07 | Yes |
| APOA4* | 0.06 | 0.34 | 0.40 | No |
| GSN | −0.09 | 0.04 | 0.08 | Yes |
| C9 | 0.18 | 0.06 | 0.09 | Yes |
| LBP | 0.20 | 0.03 | 0.07 | Yes |
| C1QC | −0.14 | 0.002 | 0.02 | Yes |
| CRISP3 | −0.32 | 0.04 | 0.07 | Yes |
| BASP1 | −0.04 | 0.56 | 0.58 | Yes |
| CRKL | −0.13 | 0.08 | 0.12 | Yes |
| SPARC | 0.27 | 0.01 | 0.05 | Yes |

*Definition of abbreviations*: FDR = false discovery rate; PrePF = preclinical pulmonary fibrosis.
Of the 12 significant proteins identified in plasma analysis, 10 were able to be detected in serum samples to allow for comparison between groups.
Analysis was controlled for family relatedness.
*Indicates different directionality than in the plasma samples.

relatives of patients with FIP, and found that in addition to age and sex, circulating proteins (S100A9, LBP, CRISP3, and CRKL) may be useful in identifying subjects with PrePF.

The identification of PrePF may play an important role in the development of clinical care of pulmonary fibrosis because other investigators have illustrated that ILAs in first-degree relatives of both patients with FIP and subjects with IPF progress (6, 13, 14). Clinically, how to address PrePF and/or ILAs is an important question because approved medical therapies (nintedanib and pirfenidone) slow down disease progression but do not reverse existing fibrosis. Therefore, there is rationale to study the role of early treatment in this disease before patients develop irreversible lung fibrosis.

One limitation of this study is that the subjects included in these analyses were not true "control subjects"—those without disease were first-degree relatives from families with FIP. As numerous studies have now illustrated (3, 7), first-degree relatives from families with FIP are at high risk for developing abnormal lung parenchyma. However, the "No Fibrosis" family members included in this investigation were those that did not have radiologic evidence of lung fibrosis. Though this may be considered a limitation of study design, we believe that this would bias our study toward the null hypothesis and would not lead to false-positive findings.

This study is also limited by the lack of a validation cohort, and validation in independent cohorts are required before these findings can be generalized. Further validation is particularly important because serum data showed consistent trends for most but not all of the plasma protein findings.

In conclusion, circulating plasma proteins are differentially detected in IPF, and some are common to subjects with IPF and PrePF. Further study and validation of these findings in independent cohorts is necessary. ■

**Author disclosures** are available with the text of this letter at www.atsjournals.org.

**Acknowledgment:** The authors thank Marvin I. Schwarz, Kevin K. Brown, Mark P. Steele, Joyce S. Lee, and James E. Loyd for clinical phenotyping and recruitment of cohort subjects; David A. Lynch for radiology reviews of the cohort; and Tasha E. Fingerlin and Weiming Zhang for guidance on statistical analyses.

Susan K. Mathai, M.D.*
*University of Colorado School of Medicine*
*Aurora, Colorado*
and
*Baylor University Medical Center at Dallas*
*Dallas, Texas*

Jonathan Cardwell, M.S.
*University of Colorado School of Medicine*
*Aurora, Colorado*

Fabian Metzger, Ph.D.
*Helmholtz Zentrum München*
*Munich, Germany*

Julia Powers, M.H.S.
Avram D. Walts, M.S.
*University of Colorado School of Medicine*
*Aurora, Colorado*

Jonathan A. Kropski, M.D.
*Vanderbilt University School of Medicine*
*Nashville, Tennessee*

Oliver Eickelberg, M.D.‡
*University of Colorado School of Medicine*
*Aurora, Colorado*

Stefanie M. Hauck, Ph.D.
*Helmholtz Zentrum München*
*Munich, Germany*

Ivana V. Yang, Ph.D.§
David A. Schwartz, M.D., M.P.H.§
*University of Colorado School of Medicine*
*Aurora, Colorado*

ORCID IDs: 0000-0002-5899-8083 (S.K.M.); 0000-0002-8923-1344 (J.A.K.); 0000-0001-7170-0360 (O.E.).

*Corresponding author (e-mail: susan.mathai@bswhealth.org).
‡O.E. is Associate Editor of *AJRCCM*. His participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.
§These authors contributed equally to this work.

## References

1. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, *et al*.; American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Society. Diagnosis of idiopathic pulmonary fibrosis: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018;198:e44–e68.
2. Steele MP, Speer MC, Loyd JE, Brown KK, Herron A, Slifer SH, *et al*. Clinical and pathologic features of familial interstitial pneumonia. *Am J Respir Crit Care Med* 2005;172:1146–1152.
3. Kropski JA, Pritchett JM, Zoz DF, Crossno PF, Markin C, Garnett ET, *et al*. Extensive phenotyping of individuals at risk for familial interstitial pneumonia reveals clues to the pathogenesis of interstitial lung disease. *Am J Respir Crit Care Med* 2015;191:417–426.
4. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, *et al*. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013;368:2192–2200.
5. Araki T, Nishino M, Zazueta OE, Gao W, Dupuis J, Okajima Y, *et al*. Paraseptal emphysema: prevalence and distribution on CT and association with interstitial lung abnormalities. *Eur J Radiol* 2015;84:1413–1418.
6. Putman RK, Hatabu H, Araki T, Gudmundsson G, Gao W, Nishino M, *et al*.; Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators; COPDGene Investigators. Association between interstitial lung abnormalities and all-cause mortality. *JAMA* 2016;315:672–681.
7. Mathai SK, Humphries S, Kropski JA, Blackwell TS, Powers J, Walts AD, *et al*. MUC5B variant is associated with visually and quantitatively detected preclinical pulmonary fibrosis. *Thorax* 2019;74:1131–1139.
8. Mathai SK, Metzger F, Cardwell J, Kropski J, Powers J, Walts AD, *et al*. Circulating plasma proteins differentially detected in idiopathic pulmonary fibrosis and in subjects with pre-clinical pulmonary fibrosis [abstract]. *Am J Respir Crit Care Med* 2019;199:A1237.
9. Lepper MF, Ohmayer U, von Toerne C, Maison N, Ziegler AG, Hauck SM. Proteomic landscape of patient-derived CD4+ T cells in recent-onset type 1 diabetes. *J Proteome Res* 2018;17:618–634.
10. Niersmann C, Hauck SM, Kannenberg JM, Röhrig K, von Toerne C, Roden M, *et al*. Omentin-regulated proteins combine a pro-inflammatory phenotype with an anti-inflammatory counterregulation in human adipocytes: a proteomics analysis. *Diabetes Metab Res Rev* 2019;35:e3074.
11. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, *et al*. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–1512.
12. O'Dwyer DN, Norman KC, Xia M, Huang Y, Gurczynski SJ, Ashley SL, *et al*. The peripheral blood proteome signature of idiopathic

pulmonary fibrosis is distinct from normal and is associated with novel immunological processes. *Sci Rep* 2017;7:46560. [Published erratum appears in *Sci Rep* 7:46860.]
13. Araki T, Putman RK, Hatabu H, Gao W, Dupuis J, Latourelle JC, *et al*. Development and progression of interstitial lung abnormalities in the Framingham heart study. *Am J Respir Crit Care Med* 2016;194: 1514–1522.
14. Salisbury ML, Hewlett JC, Ding G, Markin CR, Douglas K, Mason W, *et al*. 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.





## Choosing the Better Global Lung Initiative 2012 Equation in South African Population Groups

*To the Editor*:

Spirometry is an effective and widely available technique to measure lung function. Correct interpretation of spirometry is imperative when used to diagnose and manage lung pathology. The Global Lung Initiative 2012 ($GLI_{2012}$) provides robust and representative reference equations for lung function in four ethnic groups; however, the $GLI_{2012}$ is limited in data from African populations, and "Black" equations in $GLI_{2012}$ were solely derived using data from African Americans. For populations lacking reference range equations and for individuals of mixed ethnic origin, the $GLI_{2012}$ taskforce provided a composite "Other" equation (1).

The Pan African Thoracic Society is reluctant to endorse the use of the $GLI_{2012}$ "Other" or "Black" equations in Africa without evidence of their applicability in African populations (2). In this study, we aimed to collect spirometry data in healthy South Africans to determine if the "Black" or "Other" $GLI_{2012}$ reference equations were a good fit, or whether new reference equations are required. We hypothesized that the $GLI_{2012}$ "Black" reference equations will not fit black South African adults and children. Some of the results of this study have been previously reported in the form of abstracts (3–5).

In this cross-sectional population-based study, healthy children and adults between the age of 5 and 95 years were recruited from two provinces in South Africa: KwaZulu-Natal and the Western Cape. South Africa has a population of over 57 million people who belong to four major ethnic groups: Black African (80.9%), Mixed Ethnicity (8.8%), Caucasian (7.8%), and Indian/Asian (2.5%) (6). In line with $GLI_{2012}$ recommendations, we recruited a representative sample of at least 300 participants for each ethnic population (7). Participants were recruited between August 1, 2017, and July 31, 2018.

Anthropometric measurements were obtained and spirometry was performed as per international recommendations. Spirometry data were converted to *z*-scores using the $GLI_{2012}$ Desktop Software for Large Datasets (version 1.3.4 Build 3, April 7, 2013) and summarized by ethnic group. A good fit was determined if the average *z*-score was not statistically or physiologically different from an average *z*-score of zero (SD of 1). A difference of more than 0.5 *z*-scores from zero was considered to be clinically significant, as it represents a difference greater than sampling variability (7).

A total of 4,223 participants were recruited; of these, 546 (13%) were excluded. Exclusions included those who were acutely unwell or had a previous diagnosis of respiratory, cardiac, or neuromuscular disease. Past and current smokers were also excluded as per GLI methodology and the American Thoracic Society recommendations (8). Tests with missing data, failing quality control, or with *z*-scores greater than $\pm 5$ were excluded. Demographic characteristics of the final cohort (3,676 participants) are included in Table 1. Observed *z*-scores from Black African participants ($n = 2{,}116$) showed that the $GLI_{2012}$ "Other" had the best fit for this group (Figure 1; mean *z*-score $\pm$ SD of $0.13 \pm 1.28$ for $FEV_1$, $0.13 \pm 1.32$ for FVC, and $-0.01 \pm 0.87$ for $FEV_1$/FVC).

The "Other" equations were also the best fit for the Mixed Ethnicity group ($n = 693$) (Figure 1; mean *z*-scores were $0.22 \pm 1.44$ for $FEV_1$, $0.24 \pm 1.56$ for FVC, and $-0.02 \pm 0.85$ for $FEV_1$/FVC). The "Northeast Asian" equations had a similar average *z*-score but had much wider variability. The Caucasian participants ($n = 343$) demonstrated a good fit with the $GLI_{2012}$ "Caucasian" equation (Figure 1; mean *z*-scores were $0.21 \pm 1.22$ for $FEV_1$, $0.19 \pm 1.24$ for FVC, and $0.02 \pm 0.91$ for $FEV_1$/FVC). Participants of Asian ancestry ($n = 524$) demonstrated a good fit to the "Southeast Asian" and "Black" equation. (Figure 1; Southeast Asian mean *z*-scores were $-0.18 \pm 1.03$ for $FEV_1$, $-0.13 \pm 1.09$ for FVC, and $-0.1 \pm 0.93$ for $FEV_1$/FVC; Black equation mean *z*-scores were $0.15 \pm 1.03$ for $FEV_1$, $0.04 \pm 1.07$ for FVC, and $0.23 \pm 0.87$ for $FEV_1$/FVC). Across all ethnic groups, the $FEV_1$/FVC ratio *z*-scores were close to zero (Figure 1).

In this large, representative sample of the South African population, we found that the $GLI_{2012}$ "Caucasian" fit the Caucasian population well. For the Indian population, both the Black and the Southeast Asian equations demonstrated a good fit. As the Southeast Asian data reflects the ethnic background of the Indian population best, we determined that Southeast Asian showed the best fit but that a larger data set would be useful to confirm this. The $GLI_{2012}$ "Other" equations fit the Black African and Mixed Ethnicity populations well. In South Africa, the black population largely represents a mixture of Bantu and Khoi-San ancestry. As these genetic groups predominate in wider Southern Africa, it may be appropriate to extrapolate our conclusions to the Southern African region.

However, previous studies investigating the use of $GLI_{2012}$ equations in Africa are relatively scarce and have provided conflicting results from cohorts in different regions of Africa

Supported by a grant from the Medical Research Council of South Africa and Astrazeneca Investigator Self-Initiated Research Grant. S.-J.S. was supported by a Wellcome Trust Institutional Strategic Support Fund from Imperial College, London, UK.

Author Contributions: S.-J.S.: data handling, analysis, and interpretation and drafting and revising the manuscript. D.M.G.: study conception and design, data analysis and interpretation, and drafting and revising the manuscript. R.P.M.: data analysis and interpretation and revising the manuscript. G.L.H.: study conception and design, data analysis and interpretation, and drafting and revising the manuscript. S.S.: study conception and design, data analysis and interpretation, and drafting and revising the manuscript. R. Mphahlele: data interpretation and drafting and revising the manuscript. R. Masekela: study conception and design, data analysis and interpretation, and drafting and revising the manuscript.

Originally Published in Press as DOI: 10.1164/rccm.202005-2085LE on August 6, 2020