Table of Contents

SUPPLEMENTARY MATERIAL AND METHODS .......................................................... 2

Selecting genetic determinants of circulating protein levels ..................................... 2

Colocalization analysis.............................................................................................. 3

MR analysis using multiple cis-SNPs........................................................................ 3

Fine-mapping ............................................................................................................ 4

MR analysis to test the causal relationships between possible confounders and IPF 4

Transcriptomic data in lung tissue ........................................................................... 4

Single-cell RNA sequencing data of lung tissue.................................................... 5

SUPPLEMENTARY RESULT .................................................................................... 5

Cohort characteristics ............................................................................................. 5

Transcriptomic data in lung tissue ........................................................................... 6

Single-cell RNA sequencing data of lung tissue.................................................... 6

SUPPLEMENTARY FIGURES ................................................................................... 8

Figure S1: Bivariate plots of effect sizes of the SNPs for the exposure and the outcome.................................................................................................................. 8

Figure S2. Regional Manhattan plot of the IPF GWAS at the 19p13.3 locus. ........... 9

Figure S3. Fine-mapping results of the IPF GWAS at the 19p13.3 locus. ............... 9
SUPPLEMENTARY MATERIAL AND METHODS

Selecting genetic determinants of circulating protein levels

We first identified genome-wide significant single nucleotide polymorphisms (SNPs) (p<5x10^{-8}) associated with circulating protein levels (referred to as protein quantitative trait loci (pQTL) SNPs) in the two studies[1][2] and then limited these SNPs to those that were cis-acting. The definitions of “cis-pQTL SNPs” were different between the two studies and within 1 Mb of the transcription start site of genes encoding the corresponding protein in Sun et al.[1] and within 300 kb window across the corresponding protein-coding sequence in Emilsson et al.[2] Since multiple independent pQTL SNPs were reported by conditional analyses in Sun et al., we selected multiple independent, i.e. not in linkage disequilibrium (LD) with r^2≤0.001, cis-pQTL SNPs within the 500 kb of the leading cis-pQTL SNPs[3]. The pairwise correlation of SNPs were calculated using the 503 individuals in the European subset of 1000 Genome projects[4]. Both pQTL GWASs used the SOMAscan assay which uses aptamers to measure protein levels. Each protein has its own detection reagent selected from chemically modified DNA libraries, referred to as Slow-Off rate Modified Aptamers (SOMAmers). Detailed methods of SOMAscan assay are described elsewhere[1][2]. The median variation in protein levels explained by pQTL SNPs was 5.8% (interquartile range: 2.6 – 12.4%) in Sun et al. Phenotypic variance explained by the cis-pQTL SNP was calculated by using the formula described
elsewhere[5]. From Emilsson et al., we selected the SNPs with the lowest p-values when the SNPs for the same protein with the different SOMAmers were available. Next, we assessed whether these same SNPs had been analyzed in the IPF GWAS[6], matching on rs number and position. When they were not, we identified LD proxies for these SNPs using an $r^2$ threshold of $> 0.8$ using 1000G European reference panel[4]. Both the pQTL GWASs and the IPF GWAS were built on Genome Reference Consortium Human Build 37 (GRCh37). We detected the allele flipping by inferring the forward strand alleles using allele frequency information with the “harmonise_data” function from “TwoSampleMR” R package from MR-base[7] and discarded the palindromic and ambiguous SNPs and the SNPs matched by LD proxies with minor allele frequency $> 0.42$, since we could not infer these strands correctly. This identified 558 SNPs associated with 507 plasma protein levels (540 different SOMAmers) and 731 SNPs associated with 733 plasma protein levels, respectively from the two pQTL GWAS studies. 406 proteins were overlapped and repeatedly tested by using Sun et al. and Emilsson et al. The list of proteins we analyzed is described in Table S1 and S2.

Colocalization analysis

We conducted two sets of colocalization analysis using “coloc” R package[8] and eCAVIAR[9]. Coloc is a Bayesian approach that allows us to understand whether the same variants are responsible for the two GWAS signals (in this case the protein level and IPF) or they are distinct causal variants that are just in LD with each other. eCAVIAR is another probabilistic approach that accounts for the marginal statistics (i.e. Z score) obtained from the GWASs and LD structure of each locus and has been demonstrated to have higher accuracy and precision than coloc[9]. In coloc analysis, we selected the regions within 1 Mb of the lead SNPs for FUT3 (rs708686), FUT5 (rs778809) and TNFRSF6B (rs1056441) both from Sun et al.[1] and the IPF GWAS[6]. As allele frequency information was not provided in Sun et al.[1], we used 503 individuals in the European subset of 1000 Genome projects to estimate the allele frequency. We selected the exactly same regions as coloc analyses and performed eCAVIAR[9] by setting the maximum number of causal SNPs as one with otherwise the default setting. Whereas the posterior probability was estimated for each locus with coloc, the colocalization posterior probabilities (CLPP) score in eCAVIAR was assigned to each variant within the locus. CLPP is a joint-probability that the variant is causal both in the protein GWAS and the IPF GWAS. The cut-offs for colocalization we applied were 80% in coloc and 0.01 in eCAVIAR, as described previously[8][9].

MR analysis using multiple cis-SNPs

As a sensitivity analysis, we included multiple cis-SNPs to perform MR using “mr_inv” and “mr_egger” functions in “MendelianRandomization” v0.4.3[10]. Correlation matrices of SNPs were calculated using plink --r square with 503 individuals in the European subset of 1000
Genome projects. We used a fixed-effects IVW method and a random-effects MR-Egger method.

**Fine-mapping**
FINEMAP is a stochastic search algorithm to explore a set of the most important causal SNPs. To assess the posterior probability of causality of the SNPs for IPF in 19p13.3 locus, we first applied GCTA-COJO[11][12] with parameters of --cojo-win 20000, --cojo-slct, --cojo-collinear 0.9, and --cojo-p 1e-06 to define the conditionally independent SNP using the IPF GWAS summary statistics[6]. Rs708686 was defined as the conditionally independent SNP. We next calculated the posterior probability of causality of all the SNPs within 500 kb of rs708686 using FINEMAP v1.3.1[13] with parameters of --n-causal-snps 20, --corr-config 0.9, --corr-group 0.9 and --prior-std 0.21.

**MR analysis to test the causal relationships between possible confounders and IPF**
To reduce the possibility of biasing the MR estimates by horizontal pleiotropy of FUT3/5 cis-SNPs, we performed MR to test if potential confounders, namely vitamin B12, lactoperoxidase, lithostathine-1-alpha, FAM3B, CA19-9 and CEA could have an effect on IPF risk[14]. For these traits, only genetic determinants of each molecule identified in European ancestries were used. (Table S7). Since the underlying biology of these SNPs is not fully understood, we performed MR Steiger[14] using “mr_steiger” function in “TwoSampleMR” to orient the direction of causality.

**Transcriptomic data in lung tissue**
We identified several publicly available transcriptomic data performed both in IPF and control lung tissue published in peer-reviewed journals; two RNA sequencing data (SRP033095 [Ncase=8, Ncontrols=7], SRP010041[Ncase=3, Ncontrols=3]) and four microarray data (GSE21411 [Ncase=23, Ncontrol=6], GSE24206 [Ncase=11, Ncontrol=5], GSE32537[Ncase=119, Ncontrol=50], GSE35147 [Ncase=4, Ncontrol=4]). Since GSE32537 had the largest sample size and provided detailed information of phenotypes (age, sex, smoking status, pulmonary function tests), we decided to use GSE32537[15] for the analysis with “GEOquery v2.50.5” R package.

According to the original manuscript, “intensity data was log2-transformed, quantile normalized using robust multi-array average (RMA), and expression levels were summarized on a transcript level using the mean value of all probesets mapping to a transcript. Non-expressed and invariant transcripts were removed using a median variance filter, corrected by a Benjamini-Hochberg false discovery rate (FDR) of 0.10, resulting in a final dataset of 11,950 transcript measurements across 217 samples”[15]. As a quality control, we checked if FUT3 and FUT5 expression levels were associated with
age, sex, or smoking status amongst controls (N=50). Logistic regression models were fitted to assess if $FUT3$ and $FUT5$ expression levels was associated with IPF, adjusted for age, sex, and smoking status (ever vs never).

**Single-cell RNA sequencing data of lung tissue**

We further investigated the expression profiles of $FUT3$ and $FUT5$ in lungs at single-cell resolution. We used two publicly available datasets, GSE136831 (Ncase=12, Ncontrol=10) [16] and GSE135893 (Ncase=32, Ncontrol=28)[17], both of which were sequenced with 10x Genomics Chromium platform. For GSE136831 data, we created Seurat object by applying “Read10x” function to GSE136831_AllCells.GeneIDs.txt.gz, GSE136831_AllCells.cellBarcodes.txt.gz and GSE136831_RawCounts_Sparse.mtx.gz, followed by “CreateSeuratObject” function in “Seurat v3.2.3” package. Meta data was obtained from GSE136831_AllCells.Samples.CellType_MetadataTable.txt.gz. This dataset had already been pre-processed and cells were kept if >12% of transcriptome was from intron-spanning reads, <20% were mitochondrial origin, and with at least 1,000 unique genes captured. For GSE135893 data, GSE135893_ILD_annotated_fullsize.rds was used for the downstream analysis. This dataset had already been pre-processed and cells containing less than 1,000 nFeature_RNA and more than 25% percentage of mitochondrial genes were filtered out.

$FUT3$ and $FUT5$ were not well detected in either GSE135893 or GSE136831 ($FUT3$: 2.9% out of total cells had non-zero counts in GSE135893 and 0.66% in GSE136831, $FUT5$: 0% in GSE135893 and 0.13% in GSE136831). Although clearly such data should be interpreted with caution, we decided to analyze $FUT3$, which were relatively more expressed.

We applied three sets of statistical analyses to compare $FUT3$ expression levels between IPF and controls, stratified by cell types annotated in the original manuscripts. First, we compared $FUT3$ expression level treating each individual cell as an independent sample by applying Wilcoxon rank sum test using “wilcox.test” R package, whose results were described in the main text. Second, we averaged the $FUT3$ expression for each subject to create a single “sample” representative for each cell type. Last, we applied linear mixed model to account for the dependency of subjects using “lme4” R package with the following formula.

\[
glmer(\text{expression} \sim \text{diagnosis} + (1 \mid \text{subject}), \text{family} = \text{gaussian})
\]

where $\text{expression}$ denotes UMI counts transformed by using “SCTransform” function.

**SUPPLEMENTARY RESULT**

**Cohort characteristics**

The IPF GWAS was a meta-analysis of three distinct cohorts, which in total consisted of 2,668 cases and 8,591 controls[6]; a Chicago-based study with 541 IPF cases and 542 controls[18], a
Colorado-based study with 1,515 fibrotic idiopathic interstitial pneumonia cases and 4,683 controls[19][20], and a UK-based study with 612 IPF cases and 3,366 controls[21]. The mean age was 67.3 years for cases and 64.7 years for controls, respectively, 69.3% of cases were males and 57.1% of controls were males. 72.5% of cases were ever smokers and 66.1% of controls were ever smokers (Table 1).

In Chicago study[18], IPF cases were selected from the University of Chicago and University of Pittsburgh via the Lung Tissue Research Consortium (LTRC), and the Correlating Outcomes with biomedical Markers to Estimate Time-progression in IPF (COMET) study and the controls were selected from the database of genotypes and phenotypes (dbGaP) and healthy individuals recruited from the University of Pittsburgh. All individuals were unrelated, of European-American ancestry.

In the Colorado Study[19][20], 1,515 fibrotic idiopathic interstitial pneumonia cases were recruited from the National Jewish Health IIP population, InterMune IPF trials, UCSF, Vanderbilt University IIP population and the National Heart, Lung and Blood Institute Lung Tissue Research Consortium. 4,683 controls were generated at Centre d’Etude du Polymorphisme Humain and approved for use as controls in other studies. Controls were selected such that they were genetically similar to the cases based on IBS (identical by state) estimates. All individuals were self-reported as non-Hispanic white.

In UK study[21], 612 IPF cases recruited from nine different centres in the UK. All diagnoses were made in accordance with accepted ATS/ERS criteria[22][23]. 3,366 controls selected from UK Biobank such that they had no history of any interstitial lung disease (defined by hospital episode statistics and cause of death) and followed a similar age, sex and smoking distribution to the cases.

Transcriptomic data in lung tissue
Using microarray-based transcriptomic data in whole lungs (GSE32537), both FUT3 and FUT5 expression levels were not associated with age, sex, and smoking status (Figure S4, Table S10) in control lung tissue (N=50). Next, we confirmed that low FUT3 expression level was associated with increased risk of IPF (OR: 0.50 per 1 SD increase, 95%CI: 0.31-0.80, p=3.4x10^{-3}), but FUT5 was not significantly associated with IPF (OR: 0.72 per 1 SD increase, 95%CI: 0.46-1.1, p=0.14, Figure 4, Table S8).

Single-cell RNA sequencing data of lung tissue
FUT3 were mainly expressed in epithelial cells in two datasets according to the annotation of the original manuscripts (Figure S5). Both in GSE136831 and GSE135893, there were distinct patterns of subgroups in epithelial cells between IPF and control lung tissues; alveolar type 2 cells (AT2) were decreased and ciliated cells and basal cells were increased in IPF lungs, which is in line with previous studies[24, 25] (Figure S6).
*FUT3* expression in AT2 cells tended to be lower in IPF lungs than normal lungs (p=1.9x10^{-48} in GSE135893 and p=0.16 in GSE136831), which is concordant with our MR evidence ([Figure S7, Table S11](#)).

On the other hand, *MUC5B* positive cells defined in GSE135893 and ciliated cells defined in GSE136831 had modestly higher *FUT3* expression in IPF than in controls. ([Figure S7,8](#)), although further validation is required.
SUPPLEMENTARY FIGURES

Figure S1: Bivariate plots of effect sizes of the SNPs for the exposure and the outcome.

(A) FUT3 multiple cis-SNPs

(B) FUT5 multiple cis-SNPs

Each point represents the per allele effect size estimate of a SNP (lines from each point are 95% CI for the effect size) Both (A) FUT3 and (B) FUT5 demonstrated consistent estimates of the slope by IVW and MR-Egger methods accounting for correlated variants.
Figure S2. Regional Manhattan plot of the IPF GWAS at the 19p13.3 locus.

Each point represents a variant with chromosomal position on the x axis and the -log10(P value) on the y axis. Variants are colored in by linkage disequilibrium with rs708686. Blue=rs12610495 (top hit on chr19, which was near DPP9 gene.) Red = rs708686 (cis-pQTL SNP for FUT3) Green = rs778806 (cis-QTL SNPs for FUT5.)

Figure S3. Fine-mapping results of the IPF GWAS at the 19p13.3 locus.

For 500 kbp region around the lead SNP; rs708686 on 19p13.3 locus, we applied statistical fine mapping to calculate log10 Bayes factors (BF) for each SNP as a measure of their posterior probability for causality. Conditional independence testing was implemented using GCTA-COJO and log10BF were estimated using FINEMAP.
Figure S4: Scatter plots of standardized log-transformed $FUT3$ and $FUT5$ expression amongst control lung tissue (N=50).

- **$FUT3$ expression**
  - slope = -0.012
  - p = 0.17

- **$FUT5$ expression**
  - slope = 0.014
  - p = 0.12
Figure S5: *FUT3* expression (UMI counts) per cell stratified by annotated cellular type.

Y axis is UMI read counts per each cell. The read counts were normalized using “SCTransform” function in “Seurat” package. Black dots represent the mean value per each cell type.
Figure S6: Cell type proportions of epithelial cells in two scRNA-seq datasets in IPF and control lungs.

The cell type annotations were defined by clustering analyses in the original manuscripts.
Figure S7: FUT3 expression comparison between IPF and control lung epithelial cells.

Y axis is read counts per each cell. The read counts were normalized using “SCTransform” function in “Seurat” package. Black dots represent the mean value per each cell type. *: p-value<0.05, **: p-value<0.005, ***: p-value<0.0005, ****: p-value<0.00005. P-values were calculated by Mann-Whitney’s U test using “wilcox.test” R function, treating each individual cell as an independent sample. For other sensitivity analyses, please refer to Table S11.
Figure S8: Comparison of the fraction of FUT3 positive cells between IPF and control lungs.

GSE135893

GSE136831

Manuscript_Identity
- ATI
- ATII
- Aberrant_Basaloid
- Basal
- Ciliated
- Club
- Goblet
- Ionocyte
- Mesothelial
- PNEC
REFERENCES

1. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, Burgess S, Jiang T, Paige E, Surendran P, Oliver-Williams C, Kamat MA, Prins BP, Wilcox SK, Zimmerman ES, Chi A, Bansal N, Spain SL, Wood AM, Morrell NW, Bradley JR, Janjic N, Roberts DJ, Ouweland WH, Todd JA, Soranzo N, Suhre K, Paul DS, Fox CS, Plenge RM, et al. Genomic atlas of the human plasma proteome. Nature [Internet] Springer US; 2018; 558: 73–79Available from: http://dx.doi.org/10.1038/s41586-018-0175-2.

2. Emilsson V, Ilkov M, Lamb JR, Finkel N, Gudmundsson EF, Pitts R, Hoover H, Gudmundsdottir V, Horman SR, Aspelund T, Shu L, Trifonov V, Sigurdsson S, Manolescu A, Zhu J, Olafsson Ö, Jakobsdottir J, Lesley SA, To J, Zhang J, Harris TB, Launer LJ, Zhang B, Eiriksdottir G, Yang X, Orth AP, Jennings LL, Gudnason V. Co-regulatory networks of human serum proteins link genetics to disease. Science (80-. ). 2018; 361: 769–773.

3. Zheng J, Haberland V, Baird D, Walker V, Haycock P, Gutteridge A, Richardson TG, Staley J, Elsworth B, Burgess S, Sun BB, Danesh J, Runz H, Maranville JC, Martin HM, Yarmolinsky J, Laurin C, Holmes M V., Liu J, Estrada K, McCarthy L, Hurle M, Waterworth D, Nelson MR, Butterworth AS, Smith GD, Hemani G, Scott RA, Gaunt TR. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. bioRxiv [Internet] 2019; : 627398Available from: https://www.biorxiv.org/content/10.1101/627398v1.

4. Consortium T 1000 GP, Auton A, Abecasis GR, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs (Principal Investigator) RA, et al. A global reference for human genetic variation. Nature [Internet] The Author(s); 2015; 526: 68–74Available from: https://doi.org/10.1038/nature15393.

5. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM, Stephens M. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One 2015; 10: 1–20.

6. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, Kraven LM, Obeidat M, Li X, Ng M, Braybrooke R, Molina-Molina M, Hobbs BD, Putman RK, Sakornsakolpat P, Booth HL, Fahy WA, Hart SP, Hill MR, Hirani N, Hubbard RB, McAnulty RJ, Millar AB, Navaratnam V, Oballa E, Parfrey H, Saini G, Whyte MKB, Zhang Y, Kaminski N, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. Am. J. Respir. Crit. Care Med. 2020; 201: 564–574.

7. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess
S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihaib HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. Loos R, editor. *Elife* [Internet] eLife Sciences Publications, Ltd; 2018; 7: e34408Available from: https://doi.org/10.7554/eLife.34408.

8. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* 2014; 10.

9. Hormozdiari F, van de Bunt M, Segrè A V., Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* 2016; 99: 1245–1260.

10. Yavorska OO, Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* 2017; 46: 1734–1739.

11. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; 88: 76–82.

12. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, Martin NG, Montgomery GW, Weedon MN, Loos RJ, Frayling TM, McCarthy MI, Hirschhorn JN, Goddard ME, Visscher PM. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* [Internet] Nature Publishing Group; 2012; 44: 369–375Available from: http://dx.doi.org/10.1038/ng.2213.

13. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; 32: 1493–1501.

14. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. Li J, editor. *PLOS Genet.* [Internet] Public Library of Science; 2017 [cited 2020 Jul 31]; 13: e1007081Available from: https://dx.plos.org/10.1371/journal.pgen.1007081.

15. Yang I V., Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, Rosen R, Neidermyer AJ, McKean DF, Groshong SD, Cool C, Cosgrove GP, Lynch DA, Brown KK, Schwartz MI, Fingerlin TE, Schwartz DA. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* [Internet] Thorax; 2013 [cited 2021 Mar 10]; 68: 1114–1121Available from: https://pubmed.ncbi.nlm.nih.gov/23783374/.

16. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, Chu SG, Raby BA, Deluliii G, Januszyk M, Duan Q, Arnett HA, Siddiqui A, Washko GR, Homer R, Yan X, Rosas IO, Kaminski N. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* [Internet] American
17. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, Peter L, Chung MI, Taylor CJ, Jetter C, Raju L, Roberson J, Ding G, Wood L, Sucre JMS, Richmond BW, Serezani AP, McDonnell WJ, Mallal SB, Bacchetta MJ, Loyd JE, Shaver CM, Ware LB, Bremner R, Walia R, Blackwell TS, Banovich NE, Kropski JA. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. Sci. Adv. [Internet] American Association for the Advancement of Science; 2020 [cited 2021 Mar 12]; 6: eaba1983Available from: www.ipfcellatlas.com.

18. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi P, Scuirba J, Richards TJ, Juan-Guardela BM, Vij R, Han MLK, Martinez FJ, Kossen K, Seiwert SD, Christie JD, Nicolae D, Kaminski N, Garcia JGN. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: A genome-wide association study. Lancet Respir. Med. [Internet] Elsevier Ltd; 2013; 1: 309–317Available from: http://dx.doi.org/10.1016/S2213-2600(13)70045-6.

19. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, Collard HR, Wolters PJ, Bradford WZ, Kossen K, Seiwert SD, Du Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isaksson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Cogan JD, Mason WR, Maher TM, Molyneaux PL, Wells AU, Moffatt MF, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat. Genet. 2013; 45: 613–620.

20. Fingerlin TE, Zhang W, Yang I V., Ainsworth HC, Russell PH, Blumhagen RZ, Schwarz MI, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch DA, Groshong S, Collard HR, Wolters PJ, Bradford WZ, Kossen K, Seiwert SD, Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isakssson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Maher TM, Molyneaux PL, Wells AU, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for autoimmunity in fibrotic idiopathic interstitial pneumonia. BMC Genet. [Internet] BMC Genetics; 2016; 17: 1–12Available from: http://dx.doi.org/10.1186/s12863-016-0377-2.

21. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, Guillen-Guio B, Ma SF, Okamoto T, John AE, Obeidat M, Yang I V., Henry A, Hubbard RB, Navaratnam V, Saini G, Thompson N, Booth HL, Hart SP, Hill MR, Hirani N, Maher TM, McAnulty RJ, Millar AB, Molyneaux PL, Parfrey H, Rassl DM, Whyte MKB, Fahy WA, Marshall RP, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. Lancet Respir. Med. 2017; 5: 869–880.

22. Travis WD, Costabel U, Hansell DM, King TE, Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, Behr J, Bouros D, Brown KK, Colby T V., Collard HR,
Cordeiro CR, Cottin V, Crestani B, Drent M, Dudden RF, Egan J, Flaherty K, Hogaboam C, Inoue Y, Johkoh T, Kim DS, Kitaichi M, Loyd J, Martinez FJ, Myers J, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am. J. Respir. Crit. Care Med.* 2013; 188: 733–748.

23. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby T V., Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE, Kondoh Y, Myers J, Müller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, et al. An Official ATS/ERS/JRS/ALAT Statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *Am. J. Respir. Crit. Care Med.* 2011; 183: 788–824.

24. Parimon T, Yao C, Stripp BR, Noble PW, Chen P. Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis. *Int. J. Mol. Sci.* [Internet] MDPI AG; 2020 [cited 2021 Mar 17]; 21: 2269Available from: https://www.mdpi.com/1422-0067/21/7/2269.

25. Plantier L, Crestani B, Wert SE, Dehoux M, Zweytick B, Guenther A, Whitsett JA. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. [cited 2021 Mar 17]; Available from: http://thorax.bmj.com/.