Idiopathic pulmonary fibrosis is associated with common genetic variants and limited rare variants

Anna L. Peljto[1], Rachel Z. Blumhagen[1], Avram D. Walts[2], Jonathan Cardwell[1], Julia Powers[1], Tamera J. Corte[3], Joanne L Dickinson[4], Ian Glaspole[5], Yuben P. Moodley[6], Martina Koziar Vasakova[7], Elisabeth Bendstrup[8,9], Jesper R. Davidsen[10], Raphael Borie[11], Bruno Crestani[11,12], Philippe Dieude[13], Francesco Bonella[14], Ulrich Costabel[14], Gunnar Gudmundsson[15], Seamas C. Donnelly[16], Jim Egan[17], Michael T. Henry[18], Michael P. Keane[19], Marcus P. Kennedy[18], Cormac McCarthy[19], Aoife N. McElroy[16], Joshua A. Olaniyi[16], Katherine M. A. O'Reilly[20], Luca Richeldi[21,22], Paolo M. Leone[21,22], Venerino Poletti[23,24], Francesco Puppo[25], Sara Tomassetti[26], Valentina Luzzi[27], Nurdan Kokturk[28], Nesrin Mogulkoc[29], Christine A. Fiddler[30], Nikhil Hirani[31], Gisli Jenkins[32], Toby M. Maher[32,33], Philip L. Molyneaux[32], Helen Parfrey[30], Rebecca Braybrooke[34], Timothy S. Blackwell[35], Peter D. Jackson[36], Steven D. Nathan[37], Mary K. Porteous[38], Kevin K. Brown[2], Jason D. Christie[39], Harold R. Collard[40], Oliver Eickelberg[41], Elena E. Foster[42], Kevin F. Gibson[41], Marilyn Glassberg[43], Daniel Kass[41], Jonathan A. Kropski[35], David Lederer[44], Angela L. Linderholm[45], Jim Loyd[35], Susan K. Mathai[46], Sydney B. Montesi[47], Imre Noth[48], Justin M. Oldham[49], Amy J. Palmisciano[50], Cristina A. Reichner[51], Mauricio Rojas[52], Jesse Roman[53], Neil Schluger[54], Barry S. Shea[50], Jeffrey J. Swigris[2], Paul J. Wolters[40], Yingze Zhang[41], Cecilia M. A. Prele[55], Juan I. Enghelmayer[56,57], Maria Otaola[58], Christopher J. Ryerson[59], Mauricio Salinas[60], Martina Sterclova[8], Tewodros H. Gebremariam[61], Marjukka Myllärniemi[62], Roberto Carbone[25], Haruhiko Furusawa[63], Masaki Hirose[64], Yoshikazu Inoue[64], Yasunari Miyazaki[63], Ken Ohta[65], Shin Ohta[66], Tsukasa Okamoto[63], Dong Soon Kim[67], Annie Pardo[68], Moises Selman[69], Alvaro U. Aranda[70], Moo Suk Park[71], Jong Sun Park[72], Jin Woo Song[67], Maria Molina-Molina[73], Lurdes Planas-Cerezales[74], Gunilla Westergren-Thorsson[75], NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Albert V. Smith[76], Ani W. Manichaikul[77], John S. Kim[48], Stephen S. Rich[78], Elizabeth C. Oelsner[79], R. Graham Barr[79], Jerome I. Rotter[80], Josee Dupuis[81,86], George O'Connor[82], Ramachandran S. Vasan[83], Michael H. Cho[84], Edwin K. Silverman[84], Marvin I. Schwarz[1], Mark P. Steele[85], Joyce S. Lee[1], Ivana V. Yang[1], Tasha E. Fingerlin[2*], David A. Schwartz[1*]

*These authors contributed equally

[1] Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado
[2] National Jewish Health, Denver, Colorado
[3] Royal Prince Alfred Hospital and University of Sydney, Sydney, Australia
[4] Menzies Institute of Medical Research, University of Tasmania, Hobart, Australia
[5] Allergy, Asthma and Clinical Immunology, Alfred Health, Sydney, Australia
[6] University of Western Australia, Perth, Australia
[7] Department of Respiratory Medicine, University Thomayer Hospital, Prague, Czech Republic
[8] Center for Rare Lung Diseases, Department of Respiratory Diseases and Allergy, Aarhus Universitetshospital, Aarhus, Denmark

[9] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[10] South Danish Center for Interstitial Lung Diseases, Department of Respiratory Medicine, Odense University Hospital, 5000 Odense C, Denmark
[11] Service de Pneumologie A, Hopital Bichat, Paris, France
[12] Université Paris Cité, Inserm, Physiopathologie et Épidémiologie des Maladies Respiratoires, Paris, France
[13] Rheumatology Department, Bichat hospital, Paris, France
[14] Center for Interstitial and Rare Lung Diseases, Ruhrlandklinik University Hospital, University of Duisburg-Essen, Essen, Germany
[15] University of Iceland and Landspitali University Hospital, Reykjavik, Iceland
[16] Trinity College Dublin, Dublin, Ireland
[17] National Lung Transplantation Centre, Mater Misericordiae University Hospital, Dublin, Ireland
[18] Department of Respiratory Medicine, Cork University Hospital, Cork, Ireland
[19] University College Dublin, St Vincent's University Hospital, Dublin, Ireland
[20] Mater Misericordiae University Hospital, Dublin, Ireland
[21] Fondazione Policlinico A. Gemelli IRCCS, Rome, Italy
[22] Università Cattolica del Sacro Cuore, Rome, Italy
[23] Department of Diseases of the Thorax GB Morgagni Hospital, Forlí, Italy
[24] DIMES University of Bologna, Bolagna, Itali
[25] Department of Internal Medicine, University of Genoa, Genoa, Italy
[26] Department of Clinical and Experimental Medicine, Interventional Pulmonology Unit, Careggi University Hospital, Florence, Italy
[27] Interventional Pulmonology Unit, Careggi University Hospital, Florence, Italy
[28] Gazi University Faculty of Medicine, Ankara, Turkey
[29] Department of Pulmonology, Ege University Hospital, Izmir, Turkey
[30] Royal Papworth Hospital NHS Foundation Trust, Cambridge, United Kingdom
[31] University of Edinburgh, Edinburgh, United Kingdom
[32] National Heart and Lung Institute, Imperial College London, London, United Kingdom
[33] Keck Medicine of University of Southern California, Los Angeles, California
[34] Respiratory Medicine, University of Nottingham, Nottingham, United Kingdom
[35] Vanderbilt University Medical Center, Nashville, Tennessee
[36] Department of Pulmonary and Critical Care Medicine, Virginia Commonwealth University, Richmond, Virginia
[37] Inova Fairfax Hospital, Falls Church, Virginia
[38] Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania
[39] University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
[40] Department of Medicine, Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, University of California, San Francisco, California
[41] Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania
[42] Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, University of California-Davis, School of Medicine, Sacramento, California
[43] Division of Pulmonary, Critical Care and Sleep Medicine, University of Arizona College of Medicine Phoenix, Phoenix, Arizona

[44] Regeneron Pharmaceuticals, Inc, Tarrytown, New York
[45] Department of Internal Medicine, University of California Davis, Davis, California
[46] Baylor University Medical Center, Dallas, Texas
[47] Massachusetts General Hospital, Boston, MA
[48] Department of Medicine, University of Virginia, Charlottesville, Virginia
[49] Department of Medicine, University of Michigan, Ann Arbor, Michigan
[50] Brown University, Providence, Rhode Island
[51] Division of Pulmonary, Critical Care and Sleep Medicine, Medstar Georgetown University Hospital, Washington, DC
[52] Division of Pulmonary, Critical Care, and Sleep Medicine, Ohio State University, Columbus, Ohio
[53] Thomas Jefferson University, Philadelphia, Pennsylvania
[54] Columbia University Medical Center, New York, New York
[55] Institute for Respiratory Health, University of Western Australia, Perth, Australia
[56] Brown University, Providence, RI
[57] Hospital de Clínicas, Universidad de Buenos Aires, Buenos Aires, Argentina
[58] Instituto De Rehabilitación Psicofísica De Buenos Aires, Buenos Aires, Argentina
[59] Department of Medicine, University of British Columbia, Vancouver, Canada
[60] Universidad de Chile, Santiago, Chile
[61] Addis Ababa University College of Health Sciences, Addis Ababa, Ethiopia
[62] Department of Pulmonary Medicine, Heart and Lung Center, Helsinki University Hospital, Helsinki, Finland
[63] Tokyo Medical and Dental University, Tokyo, Japan
[64] National Hospital Organization Kinki-Chuo Chest Medical Center, Osaka, Japan
[65] National Hospital Organization Tokyo National Hospital, Tokyo, Japan
[66] Showa University, Tokyo, Japan
[67] Asan Medical Center, University of Ulsan, Seoul, Republic of Korea
[68] Faculty of Sciences, Universidad Nacional Autónoma de México, Mexico City, Mexico
[69] Instituto Nacional de Enfermedades Respiratorias Ismael Cosío Villegas, Mexico City, Mexico
[70] CardioPulmonary Reserach Center, Alliance Pulmonary Group, Guaynabo, Puerto Rico
[71] Yonsei University College of Medicine, Severance Hospital, Seoul, Republic of Korea
[72] Seoul National University Bundang Hospital, Seoul National University, Seongnam, Republic of Korea
[73] Respiratory Department, University Hospital of Bellvitge, University of Barcelona, Barcelona, Spain
[74] Interstitial Lung Disease (ILD) Multidisciplinary Unit, Hospital Universitari Bellvitge, Hospitalet de Llobregat, Barcelona, Spain
[75] Department of Experimental Medical Science, Unit of Lung Biology, Lund University, Lund, Sweden
[76] Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan
[77] Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia
[78] Center for Public Health Genomics and Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia
[79] Departments of Medicine and Epidemiology, Columbia University Medical Center, New York, New York
[80] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute at Harbor-UCLA Medical Center, Torrance, California
[81] Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts

[82] Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts
[83] Boston University's & National Heart, Lung, and Blood Institute's Framingham Heart Study, Boston, Massachusetts
[84] Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts
[85] University of Colorado Denver School of Medicine, Pulmonary and Critical Care, Aurora, Colorado
[86] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada

Corresponding Author:
Anna L. Peljto
Department of Medicine
University of Colorado Anschutz Medical Campus
12700 E 19th Ave, RC2
Aurora, CO, 80045
(720) 281-5210
anna.peljto@cuanschutz.edu

Author Contributions: ALP, AVS, AWM, JSK, SSR, ECO, GB, JIR, JD, GO, RSV, MC, ES, TEF, IVY, and DAS contributed to the study design. ALP, RZB, JC, IVY, TEF, and DAS contributed to the analysis and interpretation of the data. ALP, TEF, and DAS contributed to drafting the manuscript. All authors contributed to the acquisition of data and provided critical reviews. All authors gave final approval for publication.

Funding:
This research was supported by the following grants:

National Institutes of Health (NHLBI): R01-HL149836 (IVY and DAS), R01-HL158668 (IVY and DAS), P01-HL0928701 (TEF and DAS), UH2/3-HL123442 (DAS), UG3 HL151865 (DAS), X01-HL134585 (TEF and DAS)

Department of Veterans Affairs Medical Center: I01BX005295 (DAS)

Department of Defense: W81XWH-17-1-0597 (DAS)

Running Title: Rare variants associations with IPF

Subject Category: 9.23 Interstitial Lung Disease

Total Word Count: 3,247

Impact: This is the first study to comprehensively assess the impact of rare variants in idiopathic pulmonary fibrosis, using an agnostic analysis strategy. These results have advanced our understanding of IPF genetics by highlighting the etiologic importance of only two well-established rare genetic variants (*TERT* and *RTEL1*), replicating common variants, and defining the heritability of IPF. In aggregate, these findings simplify the genetics of IPF.

This article has an online data supplement.

ABSTRACT

Rationale

Idiopathic pulmonary fibrosis is a rare, irreversible, and progressive disease of the lungs. Common genetic variants, in addition to non-genetic factors, have been consistently associated with IPF. Rare variants identified by candidate gene, family-based, and exome studies have also been reported to associate with IPF. However, the extent to which rare variants genome-wide may contribute to the risk of IPF remains unknown.

Objectives

We used whole-genome sequencing to investigate the role of rare variants, genome-wide, on IPF risk.

Methods

As part of the Trans-Omics for Precision Medicine Program, we sequenced 2,180 cases of IPF. Association testing focused on the aggregated effect of rare variants (minor allele frequency ≤0.01) within genes or regions. We also identified individual variants that are influential within genes and estimated the heritability of IPF based on rare and common variants.

Measurements and Main Results

Rare variants in both *TERT* and *RTEL1* were significantly associated with IPF. A single rare variant in each of the *TERT* and *RTEL1* genes was found to consistently influence the aggregated test statistics. There was no significant evidence of association with other previously reported rare variants. The SNP-heritability of IPF was estimated to be 32% (s.e. 3%).

Conclusions

Rare variants within the *TERT* and *RTEL1* genes and well-established common variants have the largest contribution to IPF risk overall. Efforts in risk profiling or development of therapies for IPF that focus on *TERT*, *RTEL1*, common variants, and environmental risk factors are likely to have the largest impact on this complex disease.

Abstract Word Count: 245

Key Words: Whole Genome Sequencing, Interstitial Lung Disease, TOPMed, Genetic Association Studies, Telomerase

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a rare, irreversible, and progressive disease of the lungs that affects an estimated 5 million individuals worldwide and is associated with a median survival of 3-5 years.(1-3) IPF is associated with older age, cigarette smoking, and other environmental exposures.(4, 5) In addition, both rare mutations and common genetic variants are reported to contribute to the etiology of IPF, although the heritability of the disease remains unknown. Within 11p15, we discovered a gain-of-function(6) promoter variant in *MUC5B* (rs35705950) that is the dominant genetic risk factor for IPF, present in >50% of affected patients.(7, 8) Genome-wide association studies have identified and validated several other genetic loci with more moderate associations with IPF risk, including genes involved in telomerase maintenance, host defense, and cell-cell adhesion.(7-15) Candidate gene studies, family-based studies, and exome sequencing analyses have also identified rare mutations that associate with IPF.(16-26) However, critical unresolved questions concerning the genetics of IPF remain, including: 1) the extent and types of rare variants genome-wide that contribute to risk, 2) relative contribution of rare vs. common variants to risk, and 3) the genetic heritability of IPF. To address these questions, we have conducted a whole-genome sequencing study of IPF with genome-wide analysis of rare variation.

METHODS

Study Populations and Sequencing

To comprehensively investigate the role of rare variants in the development of IPF, we collected DNA from patients that were diagnosed with IPF according to criteria established by the

American Thoracic Society/European Respiratory Society from institutions across the United States, Europe, and Australia (Table E1). Whole-genome sequencing was performed by the National Institutes of Health (NIH) Trans-Omics for Precision Medicine (TOPMed) Program.(27) Subjects with IPF were compared to out-of-study controls selected from other TOPMed study populations, and identified as unaffected (without evidence of interstitial lung disease). Within the TOPMed program, sequencing of our case and control populations was performed at multiple centers (Table 1). TOPMed sequencing data from the IPF samples were made available in separate "data freezes", which we used to define our discovery and validation case populations. Samples that were included in TOPMed Freeze 8 were used for the discovery phase of the analysis, in which 1,264 IPF cases were compared to 1,257 unaffected controls selected from the COPDGene study (Table 1). The validation cohort comprised 916 IPF cases and 1,200 unaffected controls that were selected from the FHS and MESA studies, and included in TOPMed Freeze 10. Since the vast majority of our IPF cases were non-Hispanic white, we filtered our case and control samples to those with European ancestry using ancestry informative principal components, in order to minimize population stratification (see Supplemental Methods).

Statistical Analysis

We conducted a preliminary analysis of common genetic variants with a minor allele frequency (MAF) >0.01, genome-wide, using the combined discovery and validation cohorts. For our primary analyses of rare variants, we used SKAT-O(28) to conduct association testing of the aggregated effect of rare variants within genes or regions, defined as those with a MAF $\leq$0.01. All analyses were adjusted for sex as a covariate in the models, as well as principal components

of genetic ancestry to control for any residual fine-scale population stratification. We used a Bonferroni correction for the effective number of tests (K_{eff}) in each analysis, which is based on the estimated minimum achievable p-value for each test.(28) Any gene or window-based variant sets with a p-value<0.05/K_{eff} were considered genome-wide significant and included in validation testing ($1.76x10^{-5}$, $3.22x10^{-6}$, $3.24x1^{-8}$ for the LOF, LOF/missense, and window-based analyses respectively). Our primary analysis strategy included loss-of-function (LOF) variants aggregated within gene-based sets. For this primary analysis, gene sets with a p-value<$5x10^{-4}$ were also included in validation testing based on moderate association. Our pre-specified secondary analyses included LOF and missense variants aggregated within gene-based sets, and comprehensive testing of all rare variants aggregated into non-overlapping windows across the genome based on spatial clustering.(29) Variant sets that had been previously reported in the literature and were moderately associated with IPF in our secondary analyses (p-value<$5x10^{-5}$ for missense variant analysis or p-value<$5x10^{-7}$ for window-based analysis), were also included in validation testing. We used a Bonferroni correction to assess significance in the validation cohort, adjusting for a total of 9 tests (p-value<$5.5x10^{-3}$). We used the Rare Variant Influential Filtering Tool (RIFT)(30) to identify individual variants that had a strong influence on the aggregated test statistic for variant sets that were significantly associated with IPF in the validation cohort. For each analysis strategy, we performed a meta-analysis combining statistics from the discovery and validation cohorts. Finally, we used a genome-based restricted maximum likelihood method (GREML)(31) to estimate SNP-heritability in the combined dataset of discovery and validation samples, using all measured variants. Additional details of the

sample selection, variant filtering, and statistical methods can be found in the Supplemental Methods.

RESULTS

In our preliminary analysis of common variants, using the discovery and validation cohorts combined, we observed genome-wide significant (p-value<$5x10^{-8}$) associations with loci previously identified in genetic studies of IPF, including variants in *MUC5B, TERT, TERC, DSP,* and others (Table E2). In addition, we identified two novel associations between IPF and variants in the third intron of *MCL1* (OR=0.77; 95% CI 0.71-0.84; p-value=$6.41x10^{-09}$) and the first intron of RNA gene ENSG00000260803 (OR=1.72; 95% CI 1.42-2.08; p-value=$3.12x10^{-08}$). The findings from this common variant analysis are consistent with previous genome-wide association studies, and validate the utility of our case and control populations.

In our primary rare variant analysis, which only included loss-of-function (LOF) variants aggregated into gene-based sets, none of the genes met criteria for genome-wide significance, but there were five genes that met our criteria for moderate association in the discovery cohort (Table 2; Figure 1). *ALOX15B* and *RTEL1-TNFRSF6B* (a read-through transcription between *RTEL1* and *TNFRSF6B*) were most strongly associated with IPF (p-value=$2.81x10^{-5}$ and p-value=$3.49x10^{-5}$, respectively). Rare variants in the *RTEL1* (p-value=$1.11x10^{-4}$), *UNC93A* (p-value=$3.44x10^{-4}$), and *NFX1* (p-value=$4.67x10^{-4}$) genes were also moderately associated with IPF. These five genes were tested in our independent validation cohort of 916 IPF cases and 1,200 unaffected controls of European ancestry using a Bonferroni p-value threshold for significance that was corrected for a total of nine tests (p<$5.5x10^{-3}$). In the validation analysis, only the *RTEL1* gene

was statistically significant after adjustment for multiple testing (p-value=2.53×10^{-3}). None of the other associations from our LOF analysis strategy replicated within the validation cohort; however, *NFX1* was nominally significant (p-value=0.03).

In a prespecified secondary analysis that included missense variants in addition to the LOF variants, aggregated into gene-base sets, *TERT* (p-value=3.25×10^{-16}) and *RTEL1* (p-value=7.49×10^{-9}) were both strongly associated with IPF, exceeding the criteria for genome-wide significance in our discovery cohort (Table 2; Figure 2). The third strongest association signal in this analysis was the *SPDL1* gene (p=2.73×10^{-5}). Since a rare missense mutation within the *SPDL1* gene has been previously reported to be associated with IPF,(32) we included this gene in our validation testing of rare LOF and missense variants along with *TERT* and *RTEL1*. In the validation cohort, the association with *TERT* was replicated with p-value=9.39×10^{-8}. The associations with *RTEL1* and *SPDL1* did not reach our Bonferroni-corrected significance threshold for the validation cohort, but were nominally associated with IPF (p-value=2.07×10^{-2} and p-value=1.63×10^{-2}, respectively).

In our final analysis of all rare variants, spatially aggregated within non-overlapping windows, none reached our genome-wide significance threshold. At a more moderate level of significance, a ~3200bp window (chr11:1284193-1287389) within the *TOLLIP* gene at 11p15 was associated with IPF (p-value=1.45×10^{-7}; Table 2). Since previous studies have reported an association between IPF and variants within the *TOLLIP* gene,(10) we tested this window for association within the validation cohort, where the strength of the association was similar (p-value=2.89×10^{-6}). Given the proximity of this region to the influential gain-of-function polymorphism within the promoter of the *MUC5B* gene, rs35705950, we repeated the test of

association, adjusting for the *MUC5B* variant. After adjustment for the *MUC5B* promoter polymorphism, the window within *TOLLIP* was no longer associated with IPF (p-value=0.79).

The results of a meta-analyses of the discovery and validation cohorts largely reflected the findings from the individual cohorts (Table 3; Figure 3). In the meta-analysis of LOF variants, *RTEL1* had the strongest association signal, with a p-value just below the threshold for genome-wide significance (p-value=4.25×10^{-6}), followed by *RTEL1-TNFRSF6B* (p-value=9.12×10^{-5}), *SPSB2* (p-value=1.02×10^{-4}), and *PARN* (p =1.43×10^{-4}). In the LOF/missense variant meta-analysis, both *TERT* and *RTEL1* reached genome-wide significance levels. We tested the association with *TERT*, adjusting for the previously identified common IPF risk variant rs4449583 within *TERT.* (13) The aggregate test statistic remained significant (p-value=3.47×10^{-21}), indicating that rare variants within *TERT* influence IPF risk independent of the effect of this common *TERT* variant. We also tested the association with *RTEL1* after adjusting for the recently identified IPF risk variant rs41308092 (33). The aggregate test statistic for *RTEL1* remained significant (p-value=6.11×10^{-11}). The evidence for association with *SPDL1* was just below the level of genome-wide significance for the meta-analysis. In the window-based analysis, the window within *TOLLIP* was significantly associated with IPF, but not independent of the *MUC5B* variant, as demonstrated by the model adjusting for the rs35705950 genotype (p-value=1.0). None of the other p-values for rare variant associations changed substantially after adjustment for previously identified common IPF risk variants that reside on the same chromosome or the *MUC5B* promoter variant (Table 3).

We did not adjust for age in our analyses, since age was missing for >10% of our cases. However, we performed a sensitivity analysis where association testing was repeated in the

discovery cohort with age as a covariate. Interestingly, we found that the p-values for *TERT*, *RTEL1*, and *SPDL1* decreased after adjustment for age. The other p-values increased slightly, to a degree expected based on the reduced sample size (Table E4).

Based on a previous study that found IPF patients without the *MUC5B* risk (T) allele at rs35705950 had a higher burden of rare missense or LOF variants in *TERT* than those without the risk allele,(34) we examined the frequency of rare variants in *TERT* and *RTEL1* among IPF cases within strata defined by carriage of the *MUC5B* risk allele (GG vs. GT/TT). We did not find a significant difference in the burden of rare LOF and missense alleles in TERT. However, the burden of rare LOF and missense alleles in *RTEL1* was greater in cases without the risk allele than cases carrying one or more copies of the risk allele (0.003 vs. 0.001, p-value=0.02).

We applied a recently developed statistical method, the Rare Variant Influential Filtering Tool (RIFT),(30) to identify variants within the *RTEL1*, *TERT*, and *SPDL1* variant sets that had a strong influence on the aggregate test statistic. A single variant in the *RTEL1* LOF variant set, rs373740199, was classified as influential in both the LOF and LOF/missense variant sets, and in both cohorts (Figure E1). This variant is within the 30th exon of *RTEL1*, and was previously identified in an exome sequencing study.(26) The minor allele was present at a frequency of 0.17% among IPF cases and absent among controls. A previously reported IPF risk variant in *TERT*,(24) rs199422297, was influential in the *TERT* LOF/missense variant set across cohorts (Figure E2). This is a stop-gain variant within the 5th exon of *TERT*, and the minor allele was present at a frequency of 0.25% among IPF cases, and absent among controls. These influential *TERT* and *RTEL1* variants are reported in dbSNP to be rare among Europeans (MAF<0.01%) and absent in other populations. A single variant, rs116483731, first identified by exome-wide

association,(32) was also classified as influential in the *SPDL1* LOF/missense variant set across cohorts (Figure E3). This variant is in the second exon of *SPDL1*, and the minor allele frequency was 2.2% among cases, and 0.8% among controls (OR=2.86; 95% CI 1.96-4.17). In dbSNP, the MAF is reported as 0.7% among Europeans, 0.07% among Africans, and is absent among other populations. While other variants may have contributed to the aggregate association test statistics, and to the overall risk of IPF, our analyses suggest that a single rare variant is largely responsible for observed associations in each of the *RTEL1*, *TERT*, and *SPDL1* genes.

We compared the minor allele counts of the three identified rare, influential variants in *TERT*, *RTEL1*, and *SPDL1* in cases with and without a family history of disease. Among IPF cases with non-missing family history data, there were 1,065 sporadic cases and 837 cases with a family history of disease included in our analyses. There was no difference in the proportion of familial and sporadic IPF cases carrying the identified influential minor alleles in *RTEL1* or *SPDL1*. While the minor allele of the influential rare variant in *TERT,* rs199422297, was observed among both sporadic and familial cases, the cases with a family history of disease were more likely to carry the TERT minor allele (2/1,065 sporadic vs. 8/837 familial; p-value=0.03).

Finally, we used a genome-based restricted maximum likelihood method, GREML-LMDS,(31) to estimate the heritability of IPF. Using whole-genome sequence data from the combined discovery and validation cohorts, we estimated the SNV-heritability of IPF to be 32% (s.e. 3%).

DISCUSSION

Our findings indicate that rare variants in *RTEL1*, *TERT*, and likely *SPDL1* contribute to the risk of IPF. While these genes have been reported by others to contain rare variants associated with IPF, (13, 17, 20, 23-26, 32) we have found that a single rare variant in each of the implicated genes (*RTEL1*, *TERT*, and *SPDL1*) could be largely responsible for the observed associations. Our whole-genome sequence analysis also suggests that rare variants identified in more focused studies of familial pulmonary fibrosis, including *TERC*,(17) the surfactant protein genes,(16, 18, 21, 22, 35, 36) *TINF2,*(37) and *ABCA3*(38) do not appear to substantially contribute to the overall risk of IPF, at least in a sample of this size (Table E3). In order to further assess any potential effect of individual, previously reported rare variants within our combined dataset, we tabulated the number of minor alleles observed among cases and controls (Table E5). Although additional exceedingly rare variants may prove to be risk factors in unique families or relevant to specific IPF subtypes, given their frequency, these rare variants will only influence risk for a very small proportion of the IPF population. Moreover, our common variant analysis highlights the importance of telomerase maintenance, host defense, and cell-cell adhesion genes in the development of IPF, and our overall analysis estimated IPF heritability to be 32%. In aggregate, our results have narrowed the focus of IPF genetics to a few well-established rare variants and replicated common variants.

We found that the estimated SNP-heritability for IPF (based on all measured rare and common variants) was 32% (s.e. 3%). This estimate is similar to our previous estimates of 28% (s.e. 2%) to 31% (s.e. 3%), which were based only on common variants, excluding the *MUC5B* variant.(8) Based on these results, we hypothesize that the majority of IPF heritability can be explained by common genetic variation. However, larger sequencing datasets are needed to

explicitly estimate the contribution of rare variation to the overall heritability of IPF. Additionally, given the relatively high heritability of IPF, common variants could be used to identify early interstitial lung disease, especially among unaffected family members.(39, 40) Although early interstitial lung disease is known to have a poor prognosis,(40-42) screening guidelines for early interstitial lung disease have not been established and therapeutic intervention for early interstitial lung disease has not been studied.

While identifying common IPF risk variants was not a primary aim of this study, preliminary analyses that included common variants identified two previously unreported loci that were significantly associated with IPF. Within these loci, the variants with the strongest associations include an indel in *MCL1* (an apoptosis regulator in the BCL2 family at 1q21.2) and an indel in a lncRNA gene at 16p13.3. These indel variants may not have been well represented by the markers included in previous GWA studies, and will require validation in an independent cohort.

This is the first whole-genome sequencing study of IPF, with comprehensive assessment of rare variant associations outside of the exome. However, this study also has some limitations. While this study included one of the largest collections of IPF patients to date, the identification of rare variants is highly dependent on sample size, and lower frequency IPF risk alleles could possibly be identified by larger studies. Consequently, extremely rare variants in genes, such as *TERC*, *TINF2*, *ABCA3*, and the surfactant protein genes, previously identified through targeted candidate gene studies may play a role in the heritability of IPF, however, given their frequency, were not identified in our study population and will only influence risk for a very small proportion of the IPF population. In addition, the minor allele frequency

threshold used to define rare variants (MAF≤1%) is somewhat arbitrary, and the power to identify aggregated variant sets that are associated with IPF will depend on the distribution of allele frequencies among risk variants included in a set. This study was also limited to subjects of European ancestry, and there are likely different rare variants that influence IPF risk in populations of other ancestries. Finally, the effect of rare variants may depend on age, sex, family history of disease, or other common variant genotypes. Additional analyses will be required to understand how interactions among genetic and non-genetic risk factors contribute to the etiology of IPF.

ACKNOWLEDGEMENTS

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI).

WGS for "NHLBI TOPMed: Whole Genome Sequencing in Familial and Sporadic Idiopathic Pulmonary Fibrosis" (phs001607) was performed at the Broad Institute Genomics Platform and the McDonnell Genome Institute (HHSN268201600034I, HHSN268201600037I).

WGS for "NHLBI TOPMed: Genetic Epidemiology of COPD" (phs000951) was performed at the Broad Institute Genomics Platform (HHSN268201500014C). The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: http://www.copdgene.org/directory.

WGS for "NHLBI TOPMed: Whole Genome Sequencing in the Framingham Heart Study" (phs000974) was performed at the Broad Institute (3R01HL092577-06S1, 3U54HG003067-12S2). The Framingham Heart Study (FHS) acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible. Dr.

Vasan is supported in part by the Evans Medical Foundation and the Jay and Louis Coffman Endowment from the Department of Medicine, Boston University School of Medicine.

WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/HSN26800004). The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A fill list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org.

Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN2682018000002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN2682018000001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

REFERENCES

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

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

3. Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. *Orphanet J Rare Dis* 2008; 3: 8.

4. Baumgartner KB, Samet JM, Coultas DB, Stidley CA, Hunt WC, Colby TV, Waldron JA. Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case-control study. Collaborating Centers. *Am J Epidemiol* 2000; 152: 307-315.

5. Pinheiro GA, Antao VC, Wood JM, Wassell JT. Occupational risks for idiopathic pulmonary fibrosis mortality in the United States. *Int J Occup Environ Health* 2008; 14: 117-123.

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

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

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

9. Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, Taniguchi H, Kubo M, Kamatani N, Nakamura Y, Pirfenidone Clinical Study G. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J Med Genet* 2008; 45: 654-656.

10. 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 MK, 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* 2013; 1: 309-317.

11. Fingerlin TE, Zhang W, Yang IV, 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, du Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isaksson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Maher TM, Molyneaux PL, Wells AU, Moffatt MF, Selman M, Pardo A, Kim DS, Crapo JD, Make BJ, Regan EA, Walek DS, Daniel JJ, Kamatani Y, Zelenika D, Murphy E, Smith K, McKean D, Pedersen BS, Talbert J, Powers J, Markin CR, Beckman KB, Lathrop M, Freed B, Langefeld CD, Schwartz DA. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016; 17: 74.

12. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, Guillen-Guio B, Ma SF, Okamoto T, John AE, Obeidat M, Yang IV, Henry A, Hubbard RB, Navaratnam V, Saini G, Thompson N, Booth HL, Hart SP, Hill MR, Hirani N, Maher TM, McAnulty RJ, Millar AB, Molyneaux PL, Parfrey H, Rassl DM, Whyte MKB, Fahy WA, Marshall RP, Oballa E, Bosse Y, Nickle DC, Sin DD, Timens W, Shrine N, Sayers I, Hall IP, Noth I, Schwartz DA, Tobin MD, Wain LV, Jenkins RG. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017; 5: 869-880.

13. Moore C, Blumhagen RZ, Yang IV, Walts A, Powers J, Walker T, Bishop M, Russell P, Vestal B, Cardwell J, Markin CR, Mathai SK, Schwarz MI, Steele MP, Lee J, Brown KK, Loyd JE, Crapo JD, Silverman EK, Cho MH, James JA, Guthridge JM, Cogan JD, Kropski JA, Swigris JJ, Bair C, Kim DS, Ji W, Kim H, Song JW, Maier LA, Pacheco KA, Hirani N, Poon AS, Li F, Jenkins RG, Braybrooke R, Saini G, Maher TM, Molyneaux PL, Saunders P, Zhang Y, Gibson KF, Kass DJ, Rojas M, Sembrat J, Wolters PJ, Collard HR, Sundy JS, O'Riordan T, Strek ME, Noth I, Ma SF, Porteous MK, Kreider ME, Patel NB, Inoue Y, Hirose M, Arai T, Akagawa S, Eickelberg O, Fernandez IE, Behr J, Mogulkoc N, Corte TJ, Glaspole I, Tomassetti S, Ravaglia C, Poletti V, Crestani B, Borie R, Kannengiesser C, Parfrey H, Fiddler C, Rassl D, Molina-Molina M, Machahua C, Worboys AM, Gudmundsson G, Isaksson HJ, Lederer DJ, Podolanczuk AJ, Montesi SB, Bendstrup E, Danchel V, Selman M, Pardo A, Henry MT, Keane MP, Doran P, Vasakova M, Sterclova M, Ryerson CJ, Wilcox PG, Okamoto T, Furusawa H, Miyazaki Y, Laurent G, Baltic S, Prele C, Moodley Y, Shea BS, Ohta K, Suzukawa M, Narumoto O, Nathan SD, Venuto DC, Woldehanna ML, Kokturk N, de Andrade JA, Luckhardt T, Kulkarni T, Bonella F, Donnelly SC, McElroy A, Armstong ME, Aranda A, Carbone RG, Puppo F, Beckman KB, Nickerson DA, Fingerlin TE, Schwartz DA. Resequencing Study Confirms That Host Defense and Cell Senescence Gene Variants Contribute to the Risk of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 200: 199-208.

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

Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020; 201: 564-574.

15. Zhang D, Povysil G, Kobeissy PH, Li Q, Wang B, Amelotte M, Jaouadi H, Newton CA, Maher TM, Molyneaux PL, Noth I, Martinez FJ, Raghu G, Todd JL, Palmer SM, Haefliger C, Platt A, Petrovski S, Garcia JA, Goldstein DB, Garcia CK. Rare and Common Variants in KIF15 Contribute to Genetic Risk of Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2022.

16. Nogee LM, Dunbar AE, 3rd, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 2001; 344: 573-579.

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

18. Lawson WE, Grant SW, Ambrosini V, Womble KE, Dawson EP, Lane KB, Markin C, Renzoni E, Lympany P, Thomas AQ, Roldan J, Scott TA, Blackwell TS, Phillips JA, 3rd, Loyd JE, du Bois RM. Genetic mutations in surfactant protein C are a rare cause of sporadic cases of IPF. *Thorax* 2004; 59: 977-980.

19. Thomas AQ, Lane K, Phillips J, 3rd, Prince M, Markin C, Speer M, Schwartz DA, Gaddipati R, Marney A, Johnson J, Roberts R, Haines J, Stahlman M, Loyd JE. Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med* 2002; 165: 1322-1328.

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

21. van Moorsel CH, van Oosterhout MF, Barlo NP, de Jong PA, van der Vis JJ, Ruven HJ, van Es HW, van den Bosch JM, Grutters JC. Surfactant protein C mutations are the basis of a significant portion of adult familial pulmonary fibrosis in a dutch cohort. *Am J Respir Crit Care Med* 2010; 182: 1419-1425.

22. Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, DiMaio JM, Kinch LN, Grishin NV, Garcia CK. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am J Hum Genet* 2009; 84: 52-59.

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

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

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

26. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL, Frankel C, Mebane CM, Ren Z, Bridgers J, Urban TJ, Malone CD, Finlen Copeland A, Brinkley C, Allen AS, O'Riordan T, McHutchison JG, Palmer SM, Goldstein DB. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2017; 196: 82-93.

27. Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee SB, Tian X, Browning BL, Das S, Emde AK, Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM, Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA, Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen YI, Cho MH, Choi SH, Chung MK, Clish CB, Correa A, Curran JE, Custer B, Darbar D, Daya M, de Andrade M, DeMeo DL, Dutcher SK, Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C, Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR, Irvin MR, Johnsen JM, Johnson AD, Kaplan R, Kardia SLR, Kelly T, Kelly S, Kenny EE, Kiel DP, Klemmer R, Konkle BA, Kooperberg C, Kottgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin KH, Liu C, Loos RJF, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK, Mathias RA, McManus DD, McGarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD, Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS, Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S, Schwartz DA, Seo JS, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith JA, Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ, Vasan RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng LC, Willer CJ, Zhang Y, Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS, Silverman EK, Qasba P, Gan W, Consortium NT-OfPM, Papanicolaou GJ, Nickerson DA, Browning SR, Zody MC, Zollner S, Wilson JG, Cupples LA, Laurie CC, Jaquish CE, Hernandez RD, O'Connor TD, Abecasis GR. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; 590: 290-299.

28. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Team NGESP-ELP, Christiani DC, Wurfel MM, Lin X. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 2012; 91: 224-237.

29. Loehlein Fier H, Prokopenko D, Hecker J, Cho MH, Silverman EK, Weiss ST, Tanzi RE, Lange C. On the association analysis of genome-sequencing data: A spatial clustering approach for partitioning the entire genome into nonoverlapping windows. *Genet Epidemiol* 2017; 41: 332-340.

30. Blumhagen RZ, Schwartz DA, Langefeld CD, Fingerlin TE. Identification of Influential Variants in Significant Aggregate Rare Variant Tests. *Hum Hered* 2021: 1-13.

31. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, Robinson MR, Perry JR, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, LifeLines Cohort S, Esko T, Milani L, Magi R, Metspalu A, Hamsten A, Magnusson PK, Pedersen NL, Ingelsson E, Soranzo N, Keller MC, Wray NR, Goddard ME, Visscher PM. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* 2015; 47: 1114-1120.

32. Dhindsa RS, Mattsson J, Nag A, Wang Q, Wain LV, Allen R, Wigmore EM, Ibanez K, Vitsios D, Deevi SVV, Wasilewski S, Karlsson M, Lassi G, Olsson H, Muthas D, Monkley S, Mackay A, Murray L, Young S, Haefliger C, FinnGen C, Maher TM, Belvisi MG, Jenkins G, Molyneaux PL, Platt A, Petrovski S. Identification of a missense variant in SPDL1 associated with idiopathic pulmonary fibrosis. *Commun Biol* 2021; 4: 392.

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

34. Dressen A, Abbas AR, Cabanski C, Reeder J, Ramalingam TR, Neighbors M, Bhangale TR, Brauer MJ, Hunkapiller J, Reeder J, Mukhyala K, Cuenco K, Tom J, Cowgill A, Vogel J, Forrest WF, Collard HR, Wolters PJ, Kropski JA, Lancaster LH, Blackwell TS, Arron JR, Yaspan BL. Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. *Lancet Respir Med* 2018; 6: 603-614.

35. Guillot L, Epaud R, Thouvenin G, Jonard L, Mohsni A, Couderc R, Counil F, de Blic J, Taam RA, Le Bourgeois M, Reix P, Flamein F, Clement A, Feldmann D. New surfactant protein C gene mutations associated with diffuse lung disease. *J Med Genet* 2009; 46: 490-494.

36. Legendre M, Butt A, Borie R, Debray MP, Bouvry D, Filhol-Blin E, Desroziers T, Nau V, Copin B, Dastot-Le Moal F, Hery M, Duquesnoy P, Allou N, Bergeron A, Bermudez J, Cazes A, Chene AL, Cottin V, Crestani B, Dalphin JC, Dombret C, Doray B, Dupin C, Giraud V, Gondouin A, Gouya L, Israel-Biet D, Kannengiesser C, Le Borgne A, Leroy S, Longchampt E, Lorillon G, Nunes H, Picard C, Reynaud-Gaubert M, Traclet J, de Vuyst P, Coulomb L'Hermine A, Clement A, Amselem S, Nathan N. Functional assessment and phenotypic heterogeneity of SFTPA1 and SFTPA2 mutations in interstitial lung diseases and lung cancer. *Eur Respir J* 2020; 56.

37. Alder JK, Stanley SE, Wagner CL, Hamilton M, Hanumanthu VS, Armanios M. Exome sequencing identifies mutant TINF2 in a family with pulmonary fibrosis. *Chest* 2015; 147: 1361-1368.

38. Campo I, Zorzetto M, Mariani F, Kadija Z, Morbini P, Dore R, Kaltenborn E, Frixel S, Zarbock R, Liebisch G, Hegermann J, Wrede C, Griese M, Luisetti M. A large kindred of pulmonary fibrosis associated with a novel ABCA3 gene variant. *Respir Res* 2014; 15: 43.

39. 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.

40. 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.

41. Putman RK, Gudmundsson G, Axelsson GT, Hida T, Honda O, Araki T, Yanagawa M, Nishino M, Miller ER, Eiriksdottir G, Gudmundsson EF, Tomiyama N, Honda H, Rosas IO, Washko GR, Cho MH, Schwartz DA, Gudnason V, Hatabu H, Hunninghake GM. Imaging Patterns Are Associated with Interstitial Lung Abnormality Progression and Mortality. *Am J Respir Crit Care Med* 2019; 200: 175-183.

42. 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* 2022.

TABLES AND FIGURES

Table 1: Summary of discovery and validation cohorts by case-control status

	Discovery		**Validation**	
	Cases	**Controls**	**Cases**	**Controls**
N	1264	1257	916	1200
Sequencing Center	Washington University	Broad Institute	Broad Institute	Broad Institute
Age, Mean (SD)	65.5 (9.4)	59.4 (6.3)	67.3 (9.1)	70.0 (9.4)
Male, N (%)	895 (70.8%)	676 (53.8%)	609 (66.5%)	792 (66.0%)
Ever smoker, N (%)	783 (68.1%)	1198 (95.3%)	563 (67.5%)	742 (61.8%)

Table 2: Significant rare variant sets identified in the discovery cohort

Variant Filtering, Aggregation Unit	**Gene/Window**	**P-value, discovery cohort**	**P-value, validation cohort§**
Loss-of-function, by gene*	*UNC93A*	3.44×10^{-4}	0.346
	NFX1	4.67×10^{-4}	0.029
	ALOX15B	2.81×10^{-5}	0.705
	RTEL1	1.11×10^{-4}	2.53×10^{-3}
	RTEL1-TNFRSF6B	3.49×10^{-5}	0.106
Loss-of-function or missense, by gene†	*TERT*	3.25×10^{-16}	9.39×10^{-8}
	RTEL1	7.49×10^{-9}	0.021
	SPDL1	2.73×10^{-5}	0.016
All rare, by window‡	Chr11:1284193-1287310 (*TOLLIP*)	1.45×10^{-7}	2.89×10^{-6}

***Genes included in validation analysis based on $p<5 \times 10^{-4}$**
†Genes included in validation analysis based on genome-wide significance ($p<3.2 \times 10^{-6}$) or $p<5 \times 10^{-5}$ for previously reported genes
‡Genes included in validation analysis based on genome-wide significance ($p<3.2 \times 10^{-8}$) or $p<5 \times 10^{-7}$ for previously reported genes
§Significance in validation cohort assessed at $p<5.5 \times 10^{-3}$

Table 3: Meta-analysis p-values for rare variant sets included in validation analysis, with and without adjustment for common variants

Variant Filtering, Aggregation Unit	**Chr**	**Gene/Window**	**Variant Set P-value**	**Common Variant Covariate**	**Variant Set P-value, adjusted for common variant**
Loss-of-function, by gene	6	*UNC93A (ENSG00000112494)*	2.22×10^{-3}	rs2076295	2.18×10^{-3}
				rs35705950	9.70×10^{-4}
	9	*NFX1 (ENSG00000086102)*	2.84×10^{-4}	rs35705950	9.70×10^{-3}
	17	*ALOX15B (ENSG00000179593)*	1.28×10^{-2}	rs1981997	1.56×10^{-2}
				rs35705950	1.73×10^{-2}
	20	*RTEL1 (ENSG00000258366)*	4.25×10^{-6}	rs35705950	1.05×10^{-5}
				rs41308092	4.22×10^{-6}
	20	*RTEL1-TNFRSF6B (ENSG00000026036)*	9.12×10^{-5}	rs35705950	6.71×10^{-5}
Loss-of-function or missense, by gene	5	*TERT (ENSG00000164362)*	2.74×10^{-21}	rs4449583	3.47×10^{-21}
				rs35705950	2.39×10^{-19}
	5	*SPDL1 (ENSG00000040275)*	4.81×10^{-6}	rs4449583	5.14×10^{-6}
				rs35705950	1.43×10^{-6}
	20	*RTEL1 (ENSG00000258366)*	1.00×10^{-10}	rs35705950	3.83×10^{-11}
				rs41308092	6.11×10^{-11}
All rare, by window	11	11:1284193-1287310 (*TOLLIP*)	1.04×10^{-11}	rs35705950	1.00

Figure 1: Manhattan plot for rare loss-of-function variants (A) and loss-of-function/missense variants (B), aggregated by gene, in the discovery cohort. Horizontal lines represent the genome-wide significance thresholds adjusted for the effective number of tests (solid), and the total number of tests (dashed).

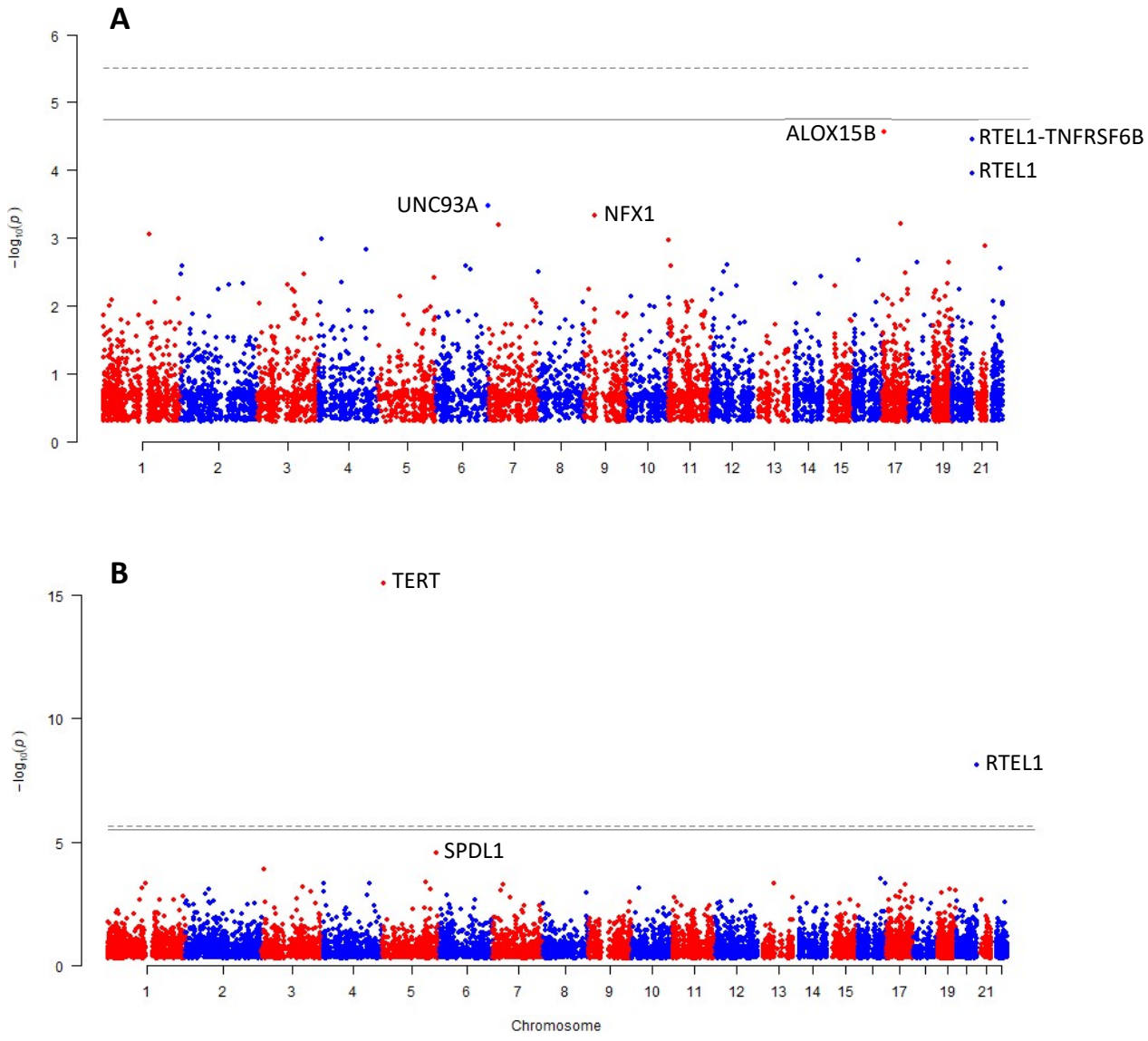
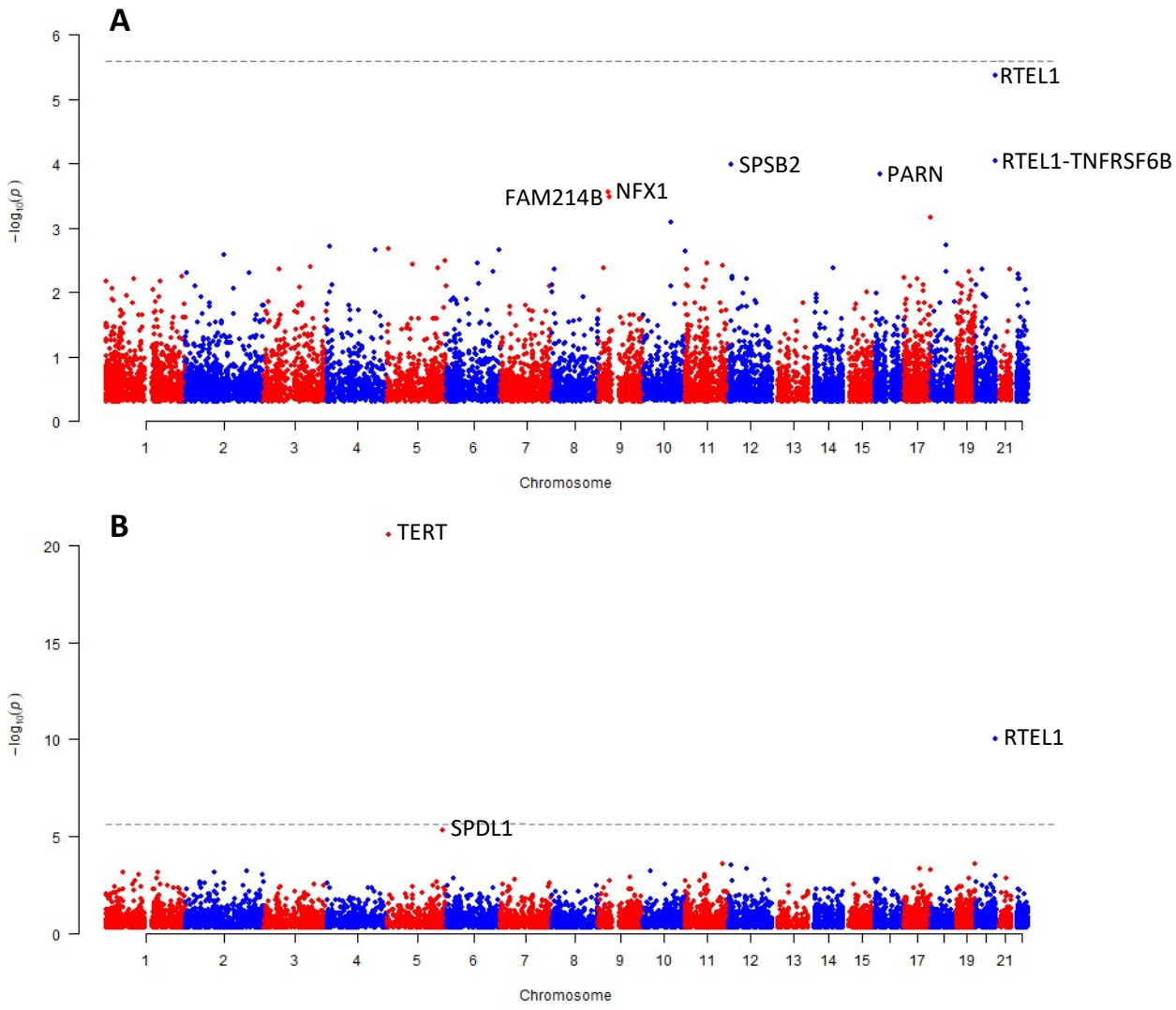
Figure 2: Manhattan plots for meta-analysis with rare loss-of-function variants (A) and loss-of-function/missense variants (B). Horizontal lines represent the genome-wide significance thresholds adjusted for the total number of tests