Resequencing Study Confirms That Host Defense and Cell Senescence Gene Variants Contribute to the Risk of Idiopathic Pulmonary Fibrosis

Camille Moore[1,2*], Rachel Z. Blumhagen[1,2*], Ivana V. Yang[3*], Avram Walts[3*], Julie Powers[3], Tarik Walker[3], Makenna Bishop[3], Pamela Russell[1], Brian Vestal[1], Jonathan Cardwell[3], Cheryl R. Markin[4], Susan K. Mathai[3], Marvin I. Schwarz[3], Mark P. Steele[3], Joyce Lee[3], Kevin K. Brown[1], James E. Loyd[4], James D. Crapo[1,3], Edwin K. Silverman[5], Michael H. Cho[5], Judith A. James[6], Joel M. Guthridge[6], Joy D. Cogan[4], Jonathan A. Kropski[4], Jeffrey J. Swigris[1], Carol Bair[1], Dong Soon Kim[7], Wonjun Ji[7], Hocheol Kim[7], Jin Woo Song[7], Lisa A. Maier[1,2,3], Karin A. Pacheco[1,2], Nikhil Hirani[8,9], Azin S. Poon[9], Feng Li[8], R. Gisli Jenkins[10], Rebecca Braybrooke[10], Gauri Saini[10], Toby M. Maher[11‡], Philip L. Molyneaux[11], Peter Saunders[11], Yingze Zhang[12], Kevin F. Gibson[12], Daniel J. Kass[12], Mauricio Rojas[12], John Sembrat[12], Paul J. Wolters[13], Harold R. Collard[13‡], John S. Sundy[14], Thomas O'Riordan[14], Mary E. Strek[15], Imre Noth[16], Shwu-Fan Ma[16], Mary K. Porteous[17], Maryl E. Kreider[17], Namrata B. Patel[17], Yoshikazu Inoue[18], Masaki Hirose[18], Toru Arai[18], Shinobu Akagawa[19], Oliver Eickelberg[3,20‡], Isis Enlil Fernandez[20], Jürgen Behr[21], Nesrin Mogulkoc[22], Tamera J. Corte[23], Ian Glaspole[24], Sara Tomassetti[25], Claudia Ravaglia[26], Venerino Poletti[26], Bruno Crestani[27], Raphael Borie[27], Caroline Kannengiesser[27], Helen Parfrey[28], Christine Fiddler[28], Doris Rassl[28], Maria Molina-Molina[29], Carlos Machahua[29], Ana Montes Worboys[29], Gunnar Gudmundsson[30], Helgi J. Isaksson[30], David J. Lederer[31], Anna J. Podolanczuk[31], Sydney B. Montesi[32], Elisabeth Bendstrup[33], Vivi Danchel[33], Moises Selman[34], Annie Pardo[35], Michael T. Henry[36], Michael P. Keane[37,38], Peter Doran[37,38], Martina Vašáková[39], Martina Sterclova[39], Christopher J. Ryerson[40], Pearce G. Wilcox[40], Tsukasa Okamoto[3,41], Haruhiko Furusawa[3,41], Yasunari Miyazaki[41], Geoffrey Laurent[42,43†], Svetlana Baltic[42], Cecilia Prele[42,43], Yuben Moodley[42], Barry S. Shea[44], Ken Ohta[19], Maho Suzukawa[19], Osamu Narumoto[19], Steven D. Nathan[45], Drew C. Venuto[45], Merte L. Woldehanna[45], Nurdan Kokturk[46], Joao A. de Andrade[47], Tracy Luckhardt[47], Tejaswini Kulkarni[47], Francesco Bonella[48], Seamus C. Donnelly[49], Aoife McElroy[49], Michelle E. Armstong[49], Alvaro Aranda[50], Roberto G. Carbone[51], Francesco Puppo[51], Kenneth B. Beckman[52], Deborah A. Nickerson[53], Tasha E. Fingerlin[1,2,3§], and David A. Schwartz[1,3,54§]

ORCID IDs: 0000-0003-3994-874X (Y.I.); 0000-0001-7170-0360 (O.E.); 0000-0001-5258-0228 (D.J.L.).

Abstract

Rationale: Several common and rare genetic variants have been associated with idiopathic pulmonary fibrosis, a progressive fibrotic condition that is localized to the lung.

Objectives: To develop an integrated understanding of the rare and common variants located in multiple loci that have been reported to contribute to the risk of disease.

Methods: We performed deep targeted resequencing (3.69 Mb of DNA) in cases ($n = 3{,}624$) and control subjects ($n = 4{,}442$) across genes and regions previously associated with disease. We tested for associations between disease and *1*) individual common variants via logistic regression and *2*) groups of rare variants via sequence kernel association tests.

Measurements and Main Results: Statistically significant common variant association signals occurred in all 10 of the regions chosen based on genome-wide association studies. The strongest risk variant is the *MUC5B* promoter variant rs35705950, with an odds ratio of 5.45 (95% confidence interval, 4.91–6.06) for one copy of the risk allele and 18.68 (95% confidence interval, 13.34–26.17) for two copies of the risk allele ($P = 9.60 \times 10^{-295}$). In addition to identifying for the first time that rare variation in *FAM13A* is associated with disease, we confirmed the role of rare variation in the *TERT* and *RTEL1* gene regions in the risk of IPF, and found that the *FAM13A* and *TERT* regions have independent common and rare variant signals.

Conclusions: A limited number of common and rare variants contribute to the risk of idiopathic pulmonary fibrosis in each of the resequencing regions, and these genetic variants focus on biological mechanisms of host defense and cell senescence.

Keywords: targeted resequencing; idiopathic pulmonary fibrosis; genetic variants; rare variants; disease risk alleles

(Received in original form October 8, 2018; accepted in final form April 22, 2019)

Am J Respir Crit Care Med Vol 200, Iss 2, pp 199–208, Jul 15, 2019

Originally Published in Press as DOI: 10.1164/rccm.201810-1891OC on April 29, 2019
Internet address: www.atsjournals.org

[1]National Jewish Health, Denver, Colorado; [2]School of Public Health; [3]Department of Medicine, and [54]Department of Immunology, University of Colorado Denver, Denver, Colorado; [4]Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee; [5]Brigham and Women's Hospital, Harvard School of Medicine, Boston, Massachusetts; [6]Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma; [7]Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; [8]MRC Centre for Inflammation Research, University of Edinburgh, Edinburgh, United Kingdom; [9]Respiratory Medicine Unit, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom; [10]Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom; [11]Royal Brompton Hospital and Imperial College, London, United Kingdom; [12]Simmons Center for Interstitial Lung Disease, University of Pittsburgh, Pittsburgh, Pennsylvania; [13]Department of Medicine, University of California, San Francisco, San Francisco, California; [14]Gilead Sciences, Foster City, California; [15]Department of Medicine, University of Chicago, Chicago, Illinois; [16]Department of Medicine, University of Virginia, Charlottesville, Virginia; [17]Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania; [18]National Hospital Organization Kinki-Chuo Chest Medical Center, Osaka, Japan; [19]National Hospital Organization Tokyo National Hospital, Tokyo, Japan; [20]Helmholtz Zentrum München, Neuherberg, Germany; [21]University Hospital Munich, Munich, Germany; [22]Department of Pulmonology, Ege University Hospital, Bornova, Izmir, Turkey; [23]Royal Prince Alfred Hospital and University of Sydney, Sydney, Australia; [24]Alfred Hospital and Monash University, Melbourne, Australia; [25]Pulmonary Medicine, GB Morgagni Hospital, Forlì, Italy; [26]Department of Diseases of the Thorax, Ospedale GB Morgagni, Forlì, Italy; [27]Université Paris Diderot and Hôpital Bichat, Paris, France; [28]Royal Papworth Hospital and Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; [29]Respiratory Department, University Hospital of Bellvitge, University of Barcelona, Barcelona, Spain; [30]National University Hospital of Iceland, University of Iceland, Reykjavik, Iceland; [31]Department of Medicine, Columbia University Irving Medical Center, New York, New York; [32]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital, Boston, Massachusetts; [33]Department of Respiratory Diseases and Allergy, Aarhus University Hospital, Aarhus, Denmark; [34]Instituto Nacional de Enfermedades Respiratorias "Ismael Cosio Villegas," México City, México; [35]Universidad Nacional Autónoma de México, México City, México; [36]Cork University Hospital and University College Cork, Cork, Ireland; [37]St. Vincent's University Hospital, Dublin, Ireland; [38]School of Medicine, University College Dublin, Dublin, Ireland; [39]Department of Respiratory Medicine, First Faculty of Medicine Charles University and Thomayer Hospital, Prague, Czech Republic; [40]University of British Columbia, Vancouver, Canada; [41]Tokyo Medical and Dental University, Tokyo, Japan; [42]Institute for Respiratory Health and; [43]Centre for Cell Therapy and Regenerative Medicine, School of Biomedical Sciences, The University of Western Australia, Perth, Australia; [44]Department of Medicine, Warren Alpert Medical School of Brown University, Providence, Rhode Island; [45]Advanced Lung Disease and Transplant Program, Inova Fairfax Hospital, Falls Church, Virginia; [46]Department of Pulmonary Medicine, Gazi University School of Medicine, Ankara, Turkey; [47]Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama; [48]Ruhrlandklinik, University Hospital, University of Duisburg-Essen, Essen, Germany; [49]Department of Medicine, Tallaght University Hospital, Trinity College Dublin, Dublin, Ireland; [50]CardioPulmonary Reserach Center, Alliance Pulmonary Group, Guaynabo, Puerto Rico; [51]Department of Internal Medicine, University of Genoa, Genoa, Italy; [52]Biomedical Genomics Center, University of Minnesota; Minneapolis, Minnesota; and [53]Northwest Genomics Center, University of Washington, Seattle, Washington

Idiopathic pulmonary fibrosis (IPF) is a progressive fibrotic lung disease that has a median survival of 3 years, is reported to affect 50,000 individuals annually in the United States, and is increasing in prevalence as our population ages (1). Although IPF is most often diagnosed in individuals with no known family history of the disease, approximately 15–25% of these individuals report a family history of idiopathic interstitial pneumonia at diagnosis. Studies of both familial and sporadic disease have identified rare mutations in telomerase (*TERT*, *TERC*, *RTEL1*, and *PARN*) and surfactant protein (*SFTPC* and *SFTPA2*) genes (2–5) and common variants in 12 genetic loci (6–9). The strongest known risk factor for both familial and sporadic IPF is a polymorphism in the distal promoter region of the *MUC5B* gene rs35705950. The

*These authors contributed equally as first authors.

†Deceased.

‡T.M.M., H.R.C., and O.E. are Associate Editors of *AJRCCM*. Their participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

§These authors contributed equally as senior authors.

This research was supported by grants from the NHLBI (R01-HL097163, P01-HL092870, and UH3-HL123442) and the Department of Defense (W81XWH-17-1-0597). The COPDGene project was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the NHLBI. The COPDGene project was 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.

Author Contributions: T.E.F. and D.A.S. designed the study and provided quality control at each step of the project. J.P., C.R.M., J. D. Crapo, E.K.S., M.H.C., J.A.J., J.M.G., J. D. Cogan, J.A.K., J.J.S., C.B., D.S.K., W.J., H.K., J.W.S., L.A.M., K.A.P., N.H., A.S.P., F.L., R.G.J., R. Braybrooke, G.S., T.M.M., P.L.M., P.S., Y.Z., K.F.G., D.J.K., M.R., J.S., P.J.W., H.R.C., J.S.S., T.O'R., M.E.S., I.N., S.-F.M., M.K.P., M.E.K., N.B.P., Y.I., M.H., T.A., S.A., O.E., I.E.F., J.B., N.M., T.J.C., I.G., S.T., C.R., V.P., B.C., R. Borie, C.K., H.P., C.F., D.R., M.M.-M., C.M., A.M.W., G.G., H.J.I., D.J.L., A.J.P., S.B.M., E.B., V.D., M.S., A.P., M.T.H., M.P.K., P.D., M.V., M.S., C.J.R., P.G.W., T.O., H.F., Y. Miyazaki, Y. Moodley, B.S.S., K.O., M.S., O.N., S.D.N., D.C.V., M.L.W., N.K., J.A.d.A., T.L., T.K., F.B., S.C.D., A.M., M.E.A., A.A., R.G.C., T.W., M.B., G.L., S.B., C.P., and F.P. provided data and samples. S.K.M., M.I.S., M.P.S, J.L., K.K.B., J.E.L., and D.A.S. coordinated the clinical evaluations and integrated the findings with clinical aspects of idiopathic pulmonary fibrosis. I.V.Y. and A.W. supervised and coordinated the laboratory work. A.W. performed the DNA preparation. D.A.N. supervised the targeted resequencing. K.B.B. supervised the AIMS genotyping. T.E.F. and D.A.S. developed the conceptual approaches to data analysis, and T.E.F. planned and supervised the analysis. C.M., R.Z.B., P.R., B.V., and J.C. performed the data cleaning and analysis. C.M., R.Z.B., T.E.F., and D.A.S. wrote the manuscript.

Correspondence and requests for reprints should be addressed to Tasha E. Fingerlin, Ph.D., National Jewish Health, 1400 Jackson Street, Denver, CO 80262. E-mail: fingerlint@njhealth.org.

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

At a Glance Commentary

Scientific Knowledge on the Subject: Rare and common genetic variants located in multiple loci have been reported to contribute to the risk of idiopathic pulmonary fibrosis.

What This Study Adds to the Field: A limited number of common and rare variants contribute substantially to the population risk of idiopathic pulmonary fibrosis, and these genetic variants focus on biological mechanisms of host defense and cell senescence.

risk variant is common (10% frequency) among individuals of European ancestry, who carry the largest burden of IPF in the United States. Although it is rare (1%) in Asia and essentially absent in Africa, the risk variant is associated with a similar risk estimate among Asian populations (10, 11).

In total, common variants in 11 regions outside of the *MUC5B* region have been reproducibly associated with IPF. Several of the regions identified via genome-wide association studies (GWASs) contain multiple variants across more than one gene, each with similar associations with IPF, such that there remain questions about the most relevant variant(s) and gene(s) for risk of IPF. In addition, no large-scale studies have been undertaken to examine rare variation in either GWAS loci or genes reported to harbor rare IPF susceptibility variants. Our goal in this study was to perform deep resequencing of known regions of GWAS association in addition to genes reported to harbor rare variants associated or segregating with IPF to understand the full range of genetic variants that contribute to IPF.

Methods

Overview

To develop an integrated understanding of rare (2–5, 12) and common (6–9) variants located in multiple loci that have been reported to contribute to the risk of IPF, we performed deep targeted resequencing (3.69 Mb of DNA) in a large population of individuals with IPF ($n = 3{,}624$) and unaffected control subjects ($n = 4{,}442$). Our targeted resequencing included genes reported to have rare mutations (*TERT*, *TERC*, *RTEL1*, *PARN*, *TINF2*, *SFTPC*, *SFTPA2*, and *ABCA3*) and predominant genetic loci reported to harbor IPF genetic variants (2–7).

Here we briefly outline the methods used in this study; the online supplement provides a comprehensive description of the study methodology, including more details for each section described below.

Study Participants

We used standard criteria established by the American Thoracic Society/European Respiratory Society/Japanese Respiratory Society to select subjects with IPF (*see* Table E1 in the online supplement) (13). The control subjects were participants in the COPDGene Study or other autoimmune genetics studies and had self-reported non-Hispanic European ancestry. All of the subjects provided written informed consent as part of institutional review board (IRB)-approved protocols for recruitment at their respective institution, and the resequencing study was approved by the National Jewish Health IRB and the University of Colorado Combined IRB.

Resequencing Regions

We resequenced a region around each of 10 genetic loci of interest (3q26, 4q22, 5p15, 6p24, 7q22, 10q24, 11p15, 13q34, 15q14-15, and 19p13 (6, 7)). We additionally sequenced regions around six genes of interest based on evidence for rare variation associated with IPF (*RTEL1*, *PARN*, *TINF2*, *SFTPC*, *SFTPA2*, and *ABCA3*). The targeted resequenced regions included 3.69 Mb DNA (*see* Table E2).

Common Variant Association Analyses

Among 3,624 cases of IPF and 4,442 unaffected control subjects (Table 1), we tested for association between each common variant (minor allele frequency [MAF] $\geq 3\%$ in combined case and control group) and disease status using a logistic regression framework adjusting for sex. We performed likelihood ratio tests comparing the full models, which included sex and the variant as predictors, with a null model containing only sex. A threshold of 5×10^{-6} was used to determine statistical significance rather than a genome-wide significance level because each of the loci represented a previously replicated association signal. This corresponds to an approximate Bonferroni correction for the number of variants tested ($n = 9{,}098$; *see* below).

To determine whether independent association signals with more than one variant were present in each region, we first identified the top variant as the variant with the smallest likelihood ratio test *P* value. For all other significant common variants in the region, we then refit logistic regression models as described above, now adjusting for both the top regional variant and sex.

To investigate whether the *MUC5B* promoter variant rs35705950 modified the

Table 1. Demographics (Full Study and Subset, by Case/Control)

Demographics	Full Study		Rare Variant Subset	
	IPF Cases ($n = 3{,}624$)	Control Subjects ($n = 4{,}442$)	IPF Cases ($n = 3{,}023$)	Control Subjects ($n = 4{,}093$)
Familial status				
Familial	478 (13.2)	269 (6.1)	458 (15.2)	189 (4.6)
Sporadic	3,146 (86.8)	4,173 (93.9)	2,565 (84.8)	3,904 (95.4)
Sex				
F	1,033 (28.5)	2,610 (58.8)	884 (29.2)	2,366 (57.8)
M	2,591 (71.5)	1,832 (41.2)	2,139 (70.8)	1,727 (42.2)

Definition of abbreviation: IPF = idiopathic pulmonary fibrosis.
Data are shown as *n* (%).

Table 2. Significant Rare Variant Windows

Nearest Gene*	Locus	Start (*bp*)	Stop (*bp*)	Annotation	# Putative Functional Variants†	# Windows in Region	*P* Value‡	Adjusted Common Variant	*P* Value§
FAM13A	4q22	89661993	89663235	Intronic	0	615	—	rs2609260	0.0360
FAM13A	4q22	89710040	89711851	Intronic, exonic	1	615	8.4×10^{-6}	rs2609260	8.31×10^{-6}¶
FAM13A	4q22	89711042	89712372	Intronic, exonic	1	615	8.9×10^{-5}	rs2609260	0.0001¶
TERT	5p15	1294397	1295255	Upstream, 5′ UTR, exonic, intronic	27	345	1.7×10^{-14}	rs4449583	9.21×10^{-16}¶
TERT	5p15	1294824	1295593	Upstream, 5′ UTR, exonic, intronic	8	345	—	rs4449583	0.0083
MIR4457	5p15	1306899	1308849	Downstream	0	345	0.001	rs4449583	0.0011¶
MIR4457	5p15	1307309	1309202	Downstream	0	345	0.001	rs4449583	0.0011¶
CLPTM1L	5p15	1324810	1325822	Exonic, 3′ UTR, intronic	1	345	0.007	rs4449583	0.0236
RP13-870H17.3	11p15	1059764	1060612	Downstream	0	1,294	8.8×10^{-4}	rs35705950	0.0386
MCF2L	13q34	113724860	113726145	Intronic	0	885	0.02	rs1278769	0.0312
RNPS1	16p13	2311178	2312190	Intronic	0	165	0.02	—	0.0162
RTEL1	20q13	62324166	62324601	Exonic, intronic	20	151	0.02	—	0.0215

Definition of abbreviation: UTR = untranslated region.
*Nearest gene to the window positions using Ensembl 75.
†Annotation with SnpEff, putative functional variants defined as having a high or moderate impact.
‡Adjusted for sex with Bonferroni correction for the number of windows in the given region.
§Adjusted for sex and top common variant with Bonferroni correction for the number of windows in the given region.
¶Significant after adjustment for sex and top common variant with Bonferroni correction for the total number of windows tested ($N = 5,677$).

effect of the top common variant in each region, we fit logistic regression models that included sex, rs35705950, the top variant, and an rs35705950 × top variant interaction. Wald *t* tests for the interaction term were used to test for effect modification. To adjust for multiple comparisons, a Bonferroni correction was made to *P* values based on the number of interaction models fit ($n = 9$).

Grouped SKAT-O Rare Variant Analyses in a Subset with Verified European Ancestry

We observed 141,783 rare variants (MAF < 3%) in our subset of 7,116 subjects

Table 3. Top Common Variant Associations

SNP	Position*	Locus	Nearest Gene	Annotation	Minor Allele	MAF in Cases	OR, Aa vs. AA (95% CI)	OR, aa vs. AA (95% CI)	*P* Value†
rs2293607	169482335	3q26	*TERC*	Downstream	C	0.30	1.30 (1.18–1.43)	1.79 (1.49–2.15)	9.11×10^{-13}
rs2609260	89836819	4q22	*FAM13A*	Intronic	C	0.23	1.35 (1.22–1.50)	1.96 (1.56–2.47)	1.03×10^{-13}
rs4449583	1284135	5p15	*TERT*	Intronic	T	0.26	0.68 (0.62–0.75)	0.46 (0.39–0.55)	2.67×10^{-25}
rs2076295‡	7563232	6p24	*DSP*	Intronic	G	0.54	1.27 (1.14–1.42)	2.08 (1.83–2.37)	1.11×10^{-29}
rs6963345	99618606	7q22	*ZKSCAN1*	Intronic	A	0.44	1.35 (1.22–1.50)	1.73 (1.51–1.99)	1.89×10^{-15}
rs2488000	105671109	10q24	*OBFC1/STN1*	Intronic	T	0.08	0.70 (0.62–0.79)§	—	7.13×10^{-9}
rs35705950	1241221	11p15	*MUC5B*	Promoter	T	0.35	5.45 (4.91–6.06)	18.68 (13.34–26.17)	9.60×10^{-295}
rs1278769‡	113536627	13q34	*ATP11A*	3′ UTR	A	0.20	0.77 (0.70–0.85)	0.69 (0.56–0.86)	7.48×10^{-8}
rs35700143	40715153	15q15	*IVD*	Intronic	C	0.41	0.76 (0.68–0.84)	0.63 (0.55–0.71)	3.44×10^{-12}
rs12610495‡	4717672	19p13	*DPP9*	Intronic	G	0.34	1.22 (1.11–1.35)	1.59 (1.36–1.87)	3.11×10^{-9}

Definition of abbreviations: CI = confidence interval; MAF = minor allele frequency; OR = odds ratio; UTR = untranslated region.
The minor allele is defined as the minor allele in the combined case and control group.
*Based on NCBI build 37.
†Adjusted for sex.
‡Same SNP as identified in original genome-wide association study (6).
§OR of a versus A resulting from the dominant test.

with genetically verified European ancestry (Table 1). We created two types of variant sets for testing: one based on proximity to genes and one based on sliding windows. Association between disease status and each rare variant set was tested in SKAT-O, adjusting for sex (14). To adjust for multiple comparisons, we applied a Bonferroni correction on gene-based sets for the total number of genes tested ($N = 206$) and on window-based sets for the total number of windows tested within the region (Tables 2 and E5). All grouped analyses were performed in R v3.3.0 using the package SKAT (15). We used a 0.05

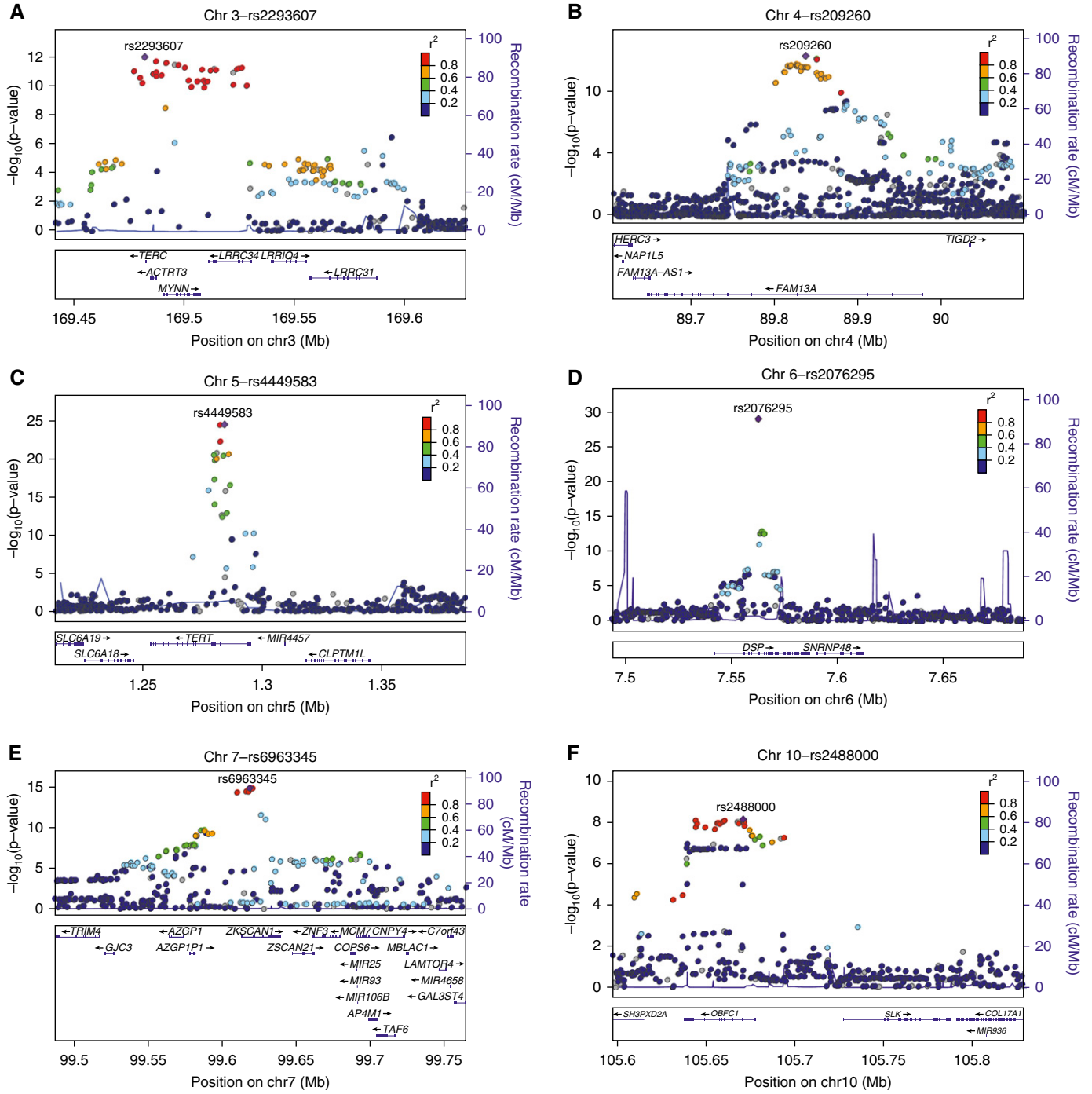

Figure 1. Regional association for common variants. Locus-specific plots corresponding to results for common variants. (*A*–*J*) For each plot, the $-\log_{10}$ *P* values (*y* axis) of the variants are shown according to their chromosomal positions (*x* axis). The significant loci are at 3q26 (*A*), 4q22 (*B*), 5p15 (*C*), 6p24 (*D*), 7q22 (*E*), 10q24 (*F*), 11p15 (*G*), 13q34 (*H*), 15q15 (*I*), and 19p13 (*J*). The estimated recombination rates from the 1000 Genomes (NCBI build 37) European population are shown as blue lines, and the genomic locations of genes within the regions of interest are shown as arrows. The SNP color represents linkage disequilibrium with the most highly associated SNP at each locus.

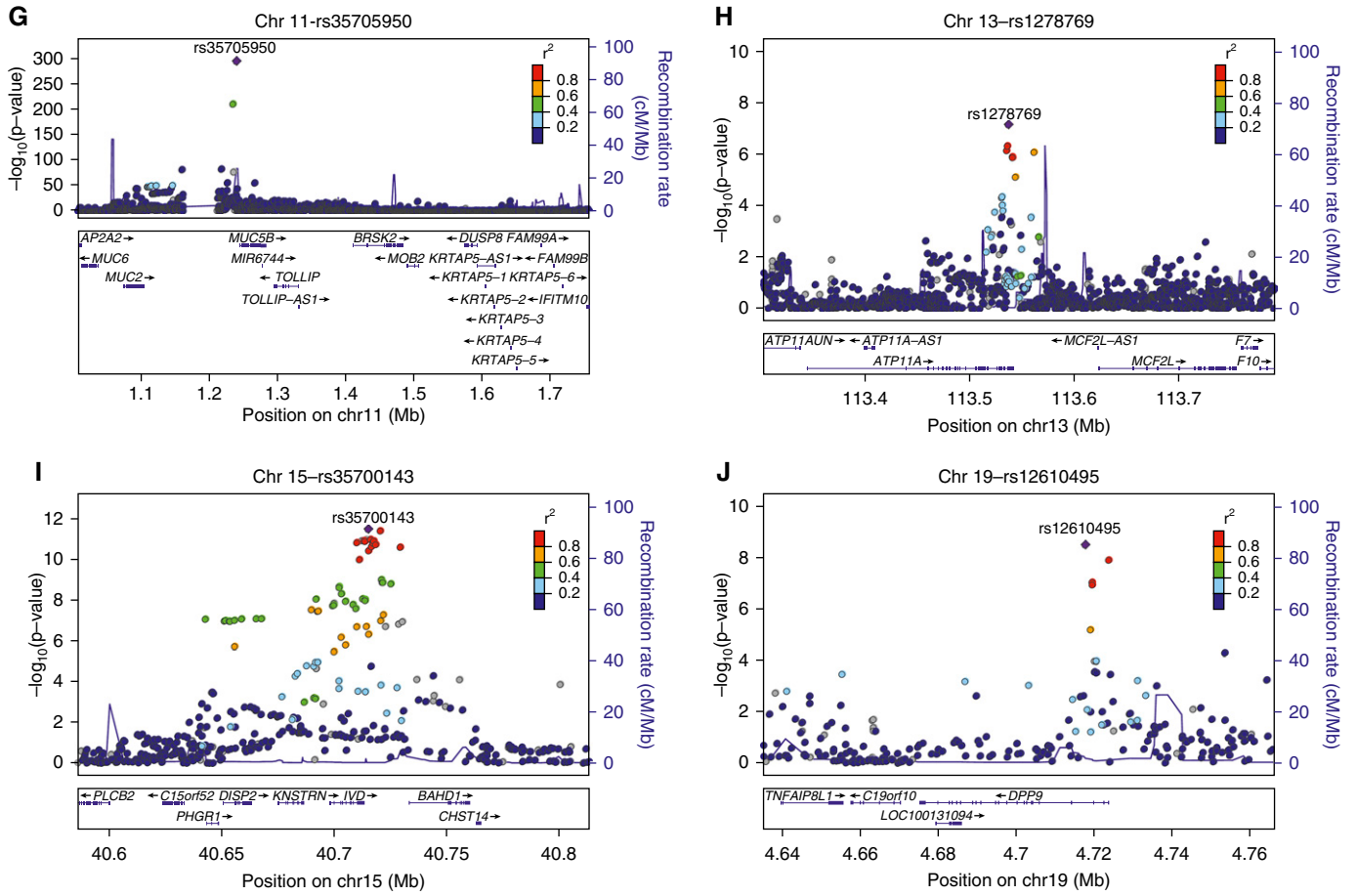

Figure 1. (Continued).

significance threshold for Bonferroni-adjusted *P* values.

To identify rare variant sets that were independently associated with disease status after adjusting for the common signals found in single variant association tests, we retested gene- and window-based sets adjusting for both sex and the top common variant in the region using SKAT-O. Bonferroni corrections were made to *P* values as described above.

Results

Common Variant Analyses

We tested for association between IPF and common variants (allele frequency ⩾ 3%) in 8,066 subjects, including 3,624 cases of IPF (3,146 sporadic cases and 478 independent cases from families with two or more cases of idiopathic interstitial pneumonia) and 4,442 unaffected control subjects (Table 1). Among the 9,098 common variants included in the single variant analyses, 992 met our statistical significance threshold of 5×10^{-6}. These signals occurred in all 10 of the GWAS loci resequencing regions (Table 3 and Figure 1). We did not identify significant common variants in the six resequencing regions designed around genes chosen based on a prior association between rare variation and IPF (*TERT*, *TERC*, *RTEL1*, *PARN*, *TINF2*, *SFTPC*, *SFTPA2*, and *ABCA3*).

The strongest common IPF risk variant is the previously identified *MUC5B* promoter variant rs35705950, with an odds ratio of 5.45 (95% confidence interval, 4.91–6.06) for one copy of the risk allele and 18.68 (95% confidence interval, 13.34–26.17) for two copies of the risk allele ($P = 9.60 \times 10^{-295}$). When we tested only significant common variants within each GWAS resequencing region, we found no additional common variants that were statistically significantly associated with IPF after adjusting for the top variant in that region. Therefore, the IPF risk variants reported in Table 3 are the only independent signals detected among common variants (MAF ⩾ 3%). In addition, these variants are associated with similar risks in familial and sporadic IPF (*see* Table E3). We explored the potential classification and prediction utility of these variants via receiver operating characteristic curves and representative positive predictive values for using sex, the *MUC5B* promoter variant (rs35705950), and the number of other risk variants. In comparison with other common variants, rs35705950 has the best potential for risk prediction (Figure 2). Because several of these variants have been identified previously in independent populations, the receiver operator characteristic curves and positive predictive values are not as susceptible to the estimation bias that is usually present when such values are estimated from the same sample that was used for association testing. In particular,

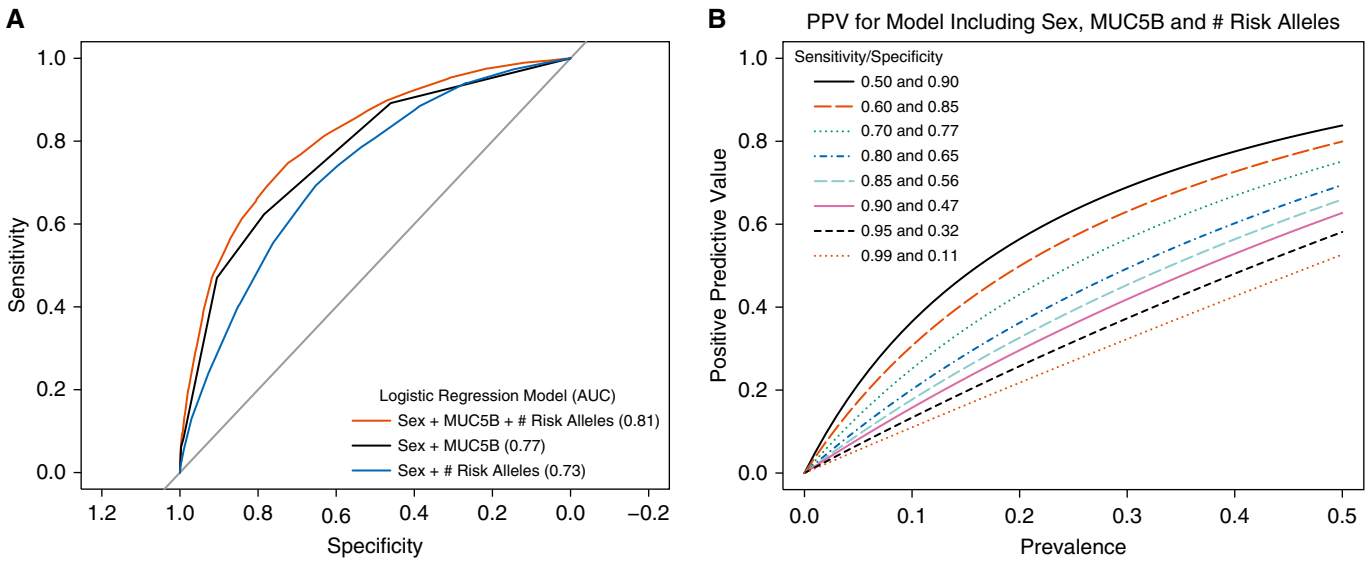

Figure 2. Receiver operator characteristic curves and positive predictive values (PPVs) using the same subjects as for all analyses. The top common variant from each region is included in a logistic regression model. The MUC5B promoter variant rs35705950 is included as a single variable with an additive (on the log-odds scale) effect of the variant. The other variants are included as a summary burden variable, where the variable is the number of putative risk variants carried by the individual. AUC = area under the curve.

only 226 cases overlap between our original GWAS and this study.

After Bonferroni correction, we did not identify significant interactions between the *MUC5B* promoter variant and the top common variants in each region (all adjusted P values > 0.10).

Grouped Rare Variant Analyses

We tested for association between IPF and groups of uncommon and rare variants among the subset of 7,116 subjects with verified European ancestry (*see* online supplement). We observed 141,783 rare variants in our subset of 7,116 subjects and assembled these variants into testing sets for grouped variant association analyses. We created two types of variant sets: one based on proximity to genes and one based on sliding windows. For gene-based sets, all rare variants within 20 kb upstream (or downstream) of the start (or stop) position of a gene were grouped together based on Ensembl 75 annotation (16). For window-based sets, we constructed sliding windows of 50 consecutive rare variants with an overlap of 25 rare variants between adjacent windows. Thirty-four gene-based rare variant sets were significantly associated with disease status after adjustment for sex and multiple comparisons (*see* Table E4); the genes were annotated to seven of our resequencing regions. Twenty of the 34 significant gene-based sets were in the *MUC5B* region on chromosome 11 (chr11), with the most significant P value corresponding to *MUC2* ($P = 1.9 \times 10^{-15}$). Seventy-nine window-based sets were significantly associated with disease status after adjustment for sex and multiple comparisons within each region (*see* Table E5). Significant windows of variants were identified in eight of the resequencing regions. Fifty-five of the significant windows were found on chr11, and 17 significant windows were found on chr5. Significant windows of variants were also identified on chr4, chr6, chr10, chr13, chr16, and chr20.

Functional Variant Subset Analysis

The results from our functional variant analysis were qualitatively the same (*see* Table E6): we identified three genes that were associated with IPF when we tested only variants in exons that were predicted to have a high or moderate impact (15). The most significant gene was *TERT*, which included 172 functional variants, 13 of which were annotated as high impact ($P = 2.07 \times 10^{-13}$; Table E6). The remaining two significant genes, *RTEL* and *RTEL1-TNFRSF6B*, contained 244 and 272 functional variants ($P = 0.00023$ and 0.00068), respectively.

Joint Common and Grouped Rare Variants

Four of the 34 significant gene-based sets were located on chr20 (*ARFRP1*, *TNFRSF6B*, *ZGPAT*, and *AL121845.3*), which was chosen as a resequencing region based on previous studies of *RTEL1*; these associations are likely driven by a significant window of 50 variants in *RTEL1* (*see* below). The remaining 30 significant gene-based rare variant sets and 77 of the 79 significant window-based sets occurred in loci that also had a significant common risk variant. To identify rare variant signals that are independent of the common variants associated with IPF, we tested the gene- and window-based rare variant sets adjusting for sex and the top common variant in the region. Of the 30 gene-based sets in regions with a common signal, only the *TERT* gene on chr5 remained significant after adjustment for the top common variant located in the intronic region of the gene ($P = 0.01$; *see* Table E7).

In total, we identified 12 rare-variant window-based sets with evidence of an independent association with IPF (Table 2). Ten window-based sets in regions with a common signal were significant after adjustment for the top common variant in the region (Table 2). There were three significant windows within the gene body of *FAM13A* on chr4 after adjustment for the *FAM13A* intronic variant, rs2609260 (Figure 3), and two of these windows

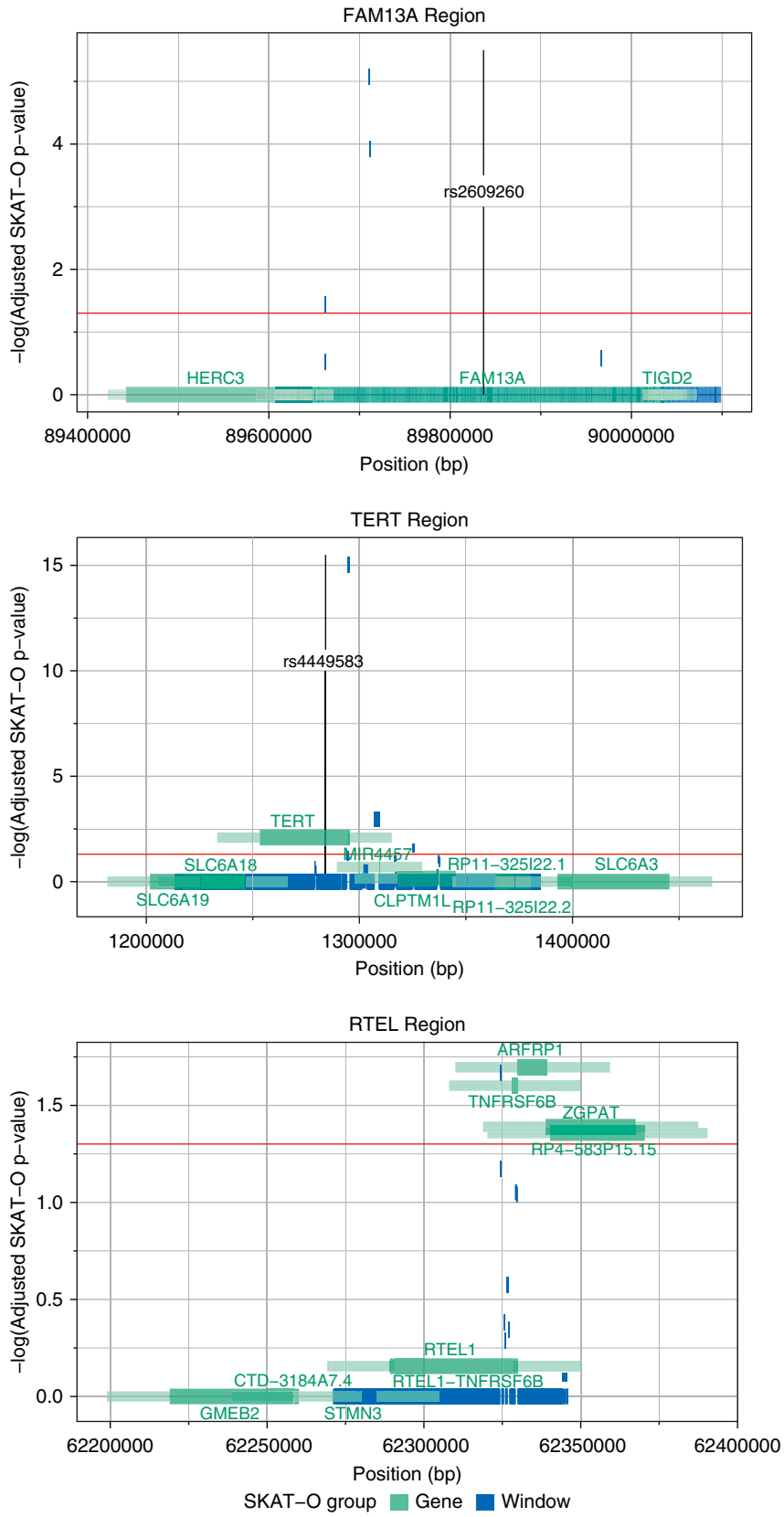

Figure 3. Rare variant signals at *FAM13A* (top), *TERT* (middle), and *RTEL1* (bottom) regions. For all plots, the $-\log_{10}$ adjusted *P* values (*y* axis) from SKAT-O of the gene (green) and window (blue) sets

overlap various exons, depending on the specific transcript. There were five significant windows on chr5 after adjustment for the intronic *TERT* SNP, rs4449583 (Figure 3). Two overlapping rare variant windows, including the most significant window on chr5 ($P = 9.21 \times 10^{-16}$), span the 5′ untranslated region (UTR), exon 1, and intronic regions of *TERT*. Two other windows on chr5 are downstream of microRNA *MIR4457*, and one spans the last exon, intron, and 3′ UTR of *CLPTM1L*, a membrane protein involved in apoptosis. One rare variant window in the *MUC5B* region on chr11 was significant after adjustment for the *MUC5B* promoter variant rs35705950 (*see* Figure E1). This window is downstream of and closest to *RP13-870H17.3* and is upstream of the *MUC2* gene, a mucin that is minimally expressed in the lung. One rare variant window on chr13 was significant after adjustment for rs1278769 (3′ UTR of *ATP11A*) and is located in an intron of *MCF2L* (Figure E1). Two other window-based sets on chr16 and chr20 did not occur in regions with a significant common variant. The significant window on chr16 was located within the intron of *RNPS1* (Figure E1) and the significant window on chr20 overlaps *RTEL1*. The *RTEL1* significant window is also near *ARFRP1*, *TNFRSF6B*, *ZGPAT*, and *AL121845.3*, and thus was included in the gene-based tests for each of those genes. Given the lack of other significant windows in these genes, we presume that the gene-based signals for these genes are driven by the *RTEL1* significant window.

Within each set of rare variants included in the significant windows after adjustment for the common variant, we characterized each variant according to its putative function and reported the number of variants with a high or moderate (17) functional annotation (Table 2, sixth column). The most significant window in the *TERT* gene contained 27 variants with a high or moderate functional impact (17), and the most significant window in the *RTEL1* gene contained 20 high- or moderate-impact variants.

Expression Quantitative Trait Loci in Significant Regions of Association

In addition to rs35705950 on chr11, three of the significant common variants in other regions are expression quantitative trait loci

(eQTLs) for expression of genes in lung tissue from the GTEx consortium (18). On chr6, rs2076295 is an eQTL for *DSP* (*see* Figure E2), as we have previously reported (6, 19). On chr15, rs35700143 is an eQTL for *BAHD1* (*see* Figure E3), a nuclear protein that promotes heterochromatic gene silencing (20) and when repressed contributes to the induction of IFN-stimulated genes (21). On chr19, rs12601495 is an eQTL for *DPP9* (*see* Figure E4).

Discussion

In the largest study of IPF to date, we refined the signal of association across each of the regions we examined that were previously identified by a GWAS. In some of the regions, including those of *MUC5B* and *DSP*, the most significantly associated variant remained the same as that identified in our GWASs, showing the power of such studies among subjects of European ancestry. However, in other regions, such as those on chr7 and chr15, the resequencing study provided substantial localization over the GWASs. Across all of the regions, we found no evidence of multiple independent associations.

In addition to identifying for the first time that rare/uncommon variation in *FAM13A*, *MIR4457*, *CLPTM1L*, *RP13-870H17.3*, *MCF2L*, and *RNPS1* is associated with IPF, we confirmed the role of rare variation in the *TERT* and *RTEL1* gene regions in the risk of IPF, and found two regions with independent common and rare variant signals (the *FAM13A* and *TERT* regions). The identification of rare variation in *FAM13A* associated with IPF is particularly interesting in light of the opposite associations previously observed in subjects with emphysema versus IPF with the C allele at rs2609260 (22) (the C allele is a risk allele for IPF but a protective allele for emphysema).

The rare and uncommon variant association signals will require further study. The aggregate tests of association do not implicate individual variants, so we do not know which of the variants in either a gene or window are responsible for the associations we observed. We note, however, that the functional variants in *TERT* that were previously reported to be relevant for pulmonary fibrosis (3, 23) are in the most significant window we identified in *TERT*. We did not use capillary sequencing to validate the rare variants included in the windows or genes that were significantly associated with disease, so we cannot exclude the possibility of some genotype call errors. However, we used strict thresholds for genotype quality scores for each variant included in any of our analyses, and expect that any errors would be nondifferential with respect to case/control status, as cases and control subjects were randomized on plates, and calls were made using the entire study sample jointly. Although each case and control subject used in these aggregate tests of association for rare and uncommon variants had verified European ancestry, that verification was based on disparate sets of data: some individuals had genome-wide SNP data from a variety of chip types, and others had ancestry-informative markers (as described in the online supplement). Because we did not have the same data for each person outside of our resequencing regions, we could not create a unified variable representing ancestry that would be appropriate in the statistical models.

In this population, *TOLLIP*, *SFTPC*, *SFTPA2*, *PARN*, and *TINF2* do not appear to contribute to the risk of IPF, although we cannot exclude the possibility of modest rare variant effects on IPF risk in these regions or the existence of highly penetrant rare/private familial mutations in familial pulmonary fibrosis. In our *a priori* power analyses, we had >90% power to detect rare variants (1% frequency) with odds ratios of 12, but less than 50% power to detect rare variants with odds ratios of 6.

Given the relative importance of the gain-of-function *MUC5B* promoter variant in the development of IPF, we speculate that IPF is caused by recurrent injury/repair/regeneration at the bronchoalveolar junction secondary to overexpression of *MUC5B*, mucociliary dysfunction, retention of particles, endoplasmic reticulum stress, and disruption of normal reparative and regenerative mechanisms in the distal lung (24–27). As stem cells attempt to regenerate injured bronchiolar and alveolar epithelium, excess expression of *MUC5B* may disrupt normal developmental pathways and hijack the normal reparative mechanisms in the distal lung, leading to chronic fibroproliferation and a regenerative process that results in honeycomb cyst formation. Based on the relationship between the *MUC5B* promoter variant and excess expression of *MUC5B* specifically at the bronchoalveolar junction (28), too much MUC5B may impair mucociliary function (29) and cause excess retention of inhaled substances (air pollutants, cigarette smoke, microorganisms, etc.), and over time, the foci of lung injury may lead to scar tissue and persistent fibroproliferation that expands and displaces normal lung tissue. However, the importance of genetic variants in telomerase genes, along with the age-related aspect of IPF, suggests that both host defense and cell senescence contribute to the pathogenesis of this progressive fibrotic lung disease.

In the absence of clear evidence of epistasis, our findings suggest that a number of genetic and nongenetic risk factors may independently (and additively) contribute to the risk and pathogenic heterogeneity of IPF. IPF appears to occur in genetically susceptible individuals who are exposed to one of a number of environmental agents that repeatedly and microscopically injure the terminal airspace. Given the multifocal, spatial heterogeneity of IPF (1), one could speculate that genetically susceptible bronchoalveolar lung units may be disproportionately injured by different environmental exposures at different points in time, and in aggregate these isolated regions of the lung eventually coalesce and present as IPF. Our aggregate genetic findings across multiple studies suggest that a reduced host defense plays a dominant role in this dynamic relationship with the environment; however, cell senescence also appears to be a critical mechanism

Figure 3. (Continued). are shown according to their chromosomal positions (*x* axis). The gene sets are labeled with their gene name; the dark green region represents the gene body and the lighter region represents the 20 kb region upstream/downstream region. Rare variants located within the gene ±20 kb were included in the testing for that gene. For the *FAM13A* and *TERT* regions, results correspond to adjustments for sex and the common variant (rs2609260 and rs4449583 for *FAM13A* and *TERT*, respectively), whereas the *RTEL1* region results are adjusted only for sex. The horizontal red line corresponds to a significance level of 0.05.

in resisting environmental stress and repopulating injured bronchoalveolar epithelia.

In summary, our findings demonstrate that the *MUC5B* promoter variant rs35705950 is the strongest known risk factor for IPF, genetic or otherwise, and that other loci contain both rare and common variation that independently contribute to IPF. Given their frequency, rare variants in *TERT*, *FAM13A*, *RTEL1*, and a few other genes have proportionately small effects on the genetic risk of IPF at the population level. ■

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

Acknowledgment: The authors gratefully acknowledge the individuals who participated in the studies that contributed to this work, and the clinical and research staff at each site for their efforts.

References

1. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, *et al*. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017;3:17074.
2. Nogee LM, Dunbar AE III, 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.
3. Armanios MY, Chen JJ, Cogan JD, Alder JK, Ingersoll RG, Markin C, *et al*. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med* 2007;356:1317–1326.
4. Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, *et al*. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am J Hum Genet* 2009;84:52–59.
5. Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, *et al*. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet* 2015;47: 512–517.
6. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, *et al*. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45:613–620.
7. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, *et al*. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1:309–317.
8. Fingerlin TE, Zhang W, Yang IV, Ainsworth HC, Russell PH, Blumhagen RZ, *et al*. 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.
9. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, *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.
10. Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L, *et al*. Mucin 5B promoter polymorphism is associated with susceptibility to interstitial lung diseases in Chinese males. *PLoS One* 2014;9: e104919.
11. Horimasu Y, Ohshimo S, Bonella F, Tanaka S, Ishikawa N, Hattori N, *et al*. MUC5B promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. *Respirology* 2015;20:439–444.
12. Dressen A, Abbas AR, Cabanski C, Reeder J, Ramalingam TR, Neighbors M, *et al*. Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. *Lancet Respir Med* 2018;6:603–614.
13. 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.
14. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, *et al*.; NHLBI GO Exome Sequencing Project—ESP Lung Project Team. 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.
15. Comprehensive R Archive Network. SKAT: SNP-Set (Sequence) Kernel Association Test; 2017 [accessed 2018 Sept 1]. Available from: https://cran.r-project.org/web/packages/SKAT/index.html.
16. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, *et al*. Ensembl 2018. *Nucleic Acids Res* 2018;46:D754–D761.
17. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, *et al*. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.
18. Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods Groups—Analysis Working Group; Enhancing GTEx (eGTEx) Groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH Program Management; Biospecimen Collection; Pathology; eQTL Manuscript Working Group. Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–213. [Published erratum appears in *Nature* 553:530.]
19. Mathai SK, Pedersen BS, Smith K, Russell P, Schwarz MI, Brown KK, *et al*. Desmoplakin variants are associated with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;193:1151–1160.
20. Bierne H, Tham TN, Batsche E, Dumay A, Leguillou M, Kernéis-Golsteyn S, *et al*. Human BAHD1 promotes heterochromatic gene silencing. *Proc Natl Acad Sci USA* 2009;106:13826–13831.
21. Lebreton A, Lakisic G, Job V, Fritsch L, Tham TN, Camejo A, *et al*. A bacterial protein targets the BAHD1 chromatin complex to stimulate type III interferon response. *Science* 2011;331:1319–1321.
22. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, *et al*.; COPDGene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators; International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
23. Tsakiri KD, Cronkhite JT, Kuan PJ, Xing C, Raghu G, Weissler JC, *et al*. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci USA* 2007;104:7552–7557.
24. 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.
25. Helling BA, Gerber AN, Kadiyala V, Sasse SK, Pedersen BS, Sparks L, *et al*. Regulation of MUC5B expression in idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2017;57:91–99.
26. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L, *et al*. Idiopathic pulmonary fibrosis: a genetic disease that involves mucociliary dysfunction of the peripheral airways. *Physiol Rev* 2016; 96:1567–1591.
27. Hancock LA, Hennessy C, Solomon GM, Dobrinskikh E, Estrella A, Hara N, *et al*. Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Nat Commun* 2018;9:5363.
28. Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, *et al*. MUC5B promoter variant rs35705950 affects MUC5B expression in the distal airways in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016;193:464–466.
29. Button B, Cai LH, Ehre C, Kesimer M, Hill DB, Sheehan JK, *et al*. A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* 2012;337:937–941.