AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

Resequencing Study Confirms 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], Peter Doran[37], Martina Vašáková[38], Martina Sterclova[38], Christopher J. Ryerson[39], Pearce G. Wilcox[39], Tsukasa Okamoto[3, 40], Haruhiko Furusawa[3, 40], Yasunari Miyazaki[40], Geoffrey Laurent[41,42,54], Svetlana Baltic[41], Cecilia Prele[41,42], Yuben Moodley[41], Barry S. Shea[43], Ken Ohta[19], Maho Suzukawa[19], Osamu Narumoto[19], Steven D. Nathan[44], Drew C. Venuto[44], Merte L. Woldehanna[44], Nurdan Kokturk[45], Joao A. de Andrade[46], Tracy Luckhardt[46], Tejaswini Kulkarni[46], Francesco Bonella[47], Seamus C. Donnelly[48], Aoife McElroy[48], Michelle E. Armstong[48], Alvaro Aranda[49], Roberto G. Carbone[50], Francesco Puppo[50], Kenneth B. Beckman[51], Deborah A. Nickerson[52], Tasha E. Fingerlin[1,2,3**], David A. Schwartz[1,3,53**]

* These authors contributed equally as first authors
** These authors contributed equally as senior authors

***, Associate Editor, AJRCCM (participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works).

[1] National Jewish Health, Denver, CO
[2] School of Public Health, University of Colorado Denver, Denver, CO
[3] Department of Medicine, University of Colorado Denver, Denver, CO
[4] Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN
[5] Brigham and Womens Hospital, Harvard School of Medicine, Boston, MA

[6] Oklahoma Medical Research Foundation, Oklahoma City, OK
[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, PA
[13] Department of Medicine, The University of California, San Francisco, CA
[14] Gilead Sciences, Foster City, CA
[15] Department of Medicine, University of Chicago, Chicago, IL
[16] Department of Medicine, University of Virginia, Charlottesville, VA
[17] Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA
[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, Commercial Road, 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, UK
[29] Respiratory Department. University Hospital of Bellvitge, University of Barcelona, Spain
[30] National University Hospital of Iceland, University of 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, MA
[33] Department of Respiratory Diseases and Allergy, Aarhus University Hospital, Aarhus, Denmark
[34] Instituto Nacional de Enfermedades Respiratorias "Ismael Cosio Villegas," México
[35] Universidad Nacional Autónoma de México, México City, México
[36] Cork University Hospital and University College Cork, Ireland
[37] St Vincent's University Hospital, Dublin and School of Medicine, University College Dublin, Ireland
[38] Department of Respiratory Medicine, First Faculty of Medicine Charles University and Thomayer Hospital, Prague, Czech Republic
[39] University of British Columbia, Vancouver, Canada
[40] Tokyo Medical and Dental University, Tokyo, Japan
[41] Institute for Respiratory Health, University of Western Australia, Perth, Australia
[42] Centre for Cell Therapy and Regenerative Medicine, School of Biomedical Sciences, The University of Western Australia, Perth, Australia

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

[43] Department of Medicine, Warren Alpert Medical School of Brown University, Providence, RI
[44] Advanced Lung Disease and Transplant Program, Inova Fairfax Hospital, Falls Church, VA
[45] Department of Pulmonary Medicine, Gazi University School of Medicine, Ankara, Turkey
[46] Department of Medicine, University of Alabama at Birmingham, Birmingham, AL
[47] Ruhrlandklinik, University Hospital, University of Duisburg-Essen, Essen, Germany
[48] Department of Medicine, Tallaght University Hospital & Trinity College Dublin, Ireland
[49] CardioPulmonary Reserach Center at Alliance Pulmonary Group in Guaynabo Puerto Rico
[50] Department of Internal Medicine, University of Genoa, Italy
[51] Biomedical Genomics Center, University of Minnesota; Minneapolis, MN
[52] Northwest Genomics Center, University of Washington, Seattle, WA
[53] Department of Immunology, University of Colorado, Denver, CO
[54] Deceased

Corresponding Authors:

Tasha E. Fingerlin, Ph.D.
National Jewish Health
1400 Jackson Street
Denver, CO 80262
Phone: 303-270-2487
fingerlint@njhealth.org

David A. Schwartz, M.D.
University of Colorado
12631 East 17th Avenue, B178
Aurora, CO 80045
Phone: 303-724-1783
david.schwartz@ucdenver.edu

AUTHOR CONTRIBUTIONS

TEF and DAS designed the study and provided quality control at each step of the project; JP, CRM, JDC, EKS, MHC, JAJ, JMG, JDC, JAK, JJS, CB, DSK, WJ, HK, JWS, LAM, KAP, NH, ASP, FL, RGJ, RB, GS, TMM, PLM, PS, YZ, KFG, DJK, MR, JS, PJW, HRC, JS, TOR, MES, IN S-FM, MKP, MRK, NBP, YI, MH, TA, SA, OE, IEF, JB, NM, TJC, IG, ST, CR, VP, BC, RB, CK, HP, CF, DR, MMM, CM, AMW, GG, HJI, DJL, AJP, SBM, EB, VD, MS, AP, MT, MPK, PD, MV, MS, CJR, PGW, TO, HP, YM, BSS, KO, MS, ON, SDN, DCV, MLW, NK, JAD, TL, TK, FB, SCD, AM, MEA, AA, RGC, FP, provided data and samples; SKM, MIS, MPS, JL, KKB, JEL, DAS coordinated the clinical evaluations and integrated the findings with clinical aspects of IPF; IVY and AW supervised and coordinated the laboratory work; AW performed the DNA preparation; DAN supervised the targeted resequencing; KBB supervised the AIMS genotyping; TEF and DAS developed the conceptual approaches to data analysis and TEF planned and

supervised the analysis; CM, RZB, PAR, JC and BPV performed the data cleaning and analysis; CM, RZB, TEF and DAS wrote the manuscript.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

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.

Objective: 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 controls (N=4,442) across genes and regions previously associated with disease. We tested for association between disease and a) individual common variants via logistic regression and b) groups of rare variants via a sequence kernel association test.

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 OR of 5.45 (95% CI: 4.91-6.06) for one copy of the risk allele and 18.68 (95% CI: 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 risk of IPF in the *TERT* and *RTEL1* gene regions, 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; these genetic variants focus on biological mechanisms of host defense and cell senescence.

Abstract word count: 250.
MeSH terms: DNA Sequence Analysis,

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a progressive fibrotic lung disease with a median survival of 3 years, is reported to affect 50,000 individuals annually in the U.S., and is increasing in prevalence as our population ages (1). While IPF is most often diagnosed in an individual with no prior known family history, approximately 15%-25% 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 *MUC5B* gene, rs35705950. The risk variant is common (10% frequency) among individuals of European ancestry, which carry the largest burden of IPF in the U.S. While 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 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 in order to understand the full range of genetic variants contributing to IPF.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

METHODS

Overview

To develop this integrated understanding of rare (2-5, 12) and common variants (6-9) located in multiple loci that have been reported to contribute to the risk of IPF, we have performed deep targeted resequencing (3.69 Mb of DNA) in a large population of patients 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 for the study; the online supplement methods provide a comprehensive description of the study methodology, including more detail on each section described below.

Study Participants

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

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 6 genes of interest based on evidence for rare variation associated with IPF (*RTEL1*, *PARN*, *TINF2*, *SFTPC*, *SFTPA2*, and *ABCA3*). The targeted resequenced regions include 3.69 Mb DNA (see Table E2 in the online data supplement).

Common Variant Association Analyses

Among 3,624 cases of IPF and 4,442 unaffected controls (Table 1), we tested for association between each common variant (Minor Allele Frequency (MAF) ≥ 3% in combined case and control group) and disease status using a logistic regression framework adjusting for sex; we performed likelihood ratio tests (LRT) comparing the full models, which included sex and the variant as predictors, to 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 since each of the loci represented a previously replicated association signal. This corresponds to an approximate Bonferroni correction for the number of variants tested (# variants tested = 9,098; see below).

To determine if 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 LRT p-value. For all other significant common variants in the region, we then re-fit

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

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 effect of the top common variant in each region, we fit logistic regression models that included sex, rs35705950, the top variant, and an rs35705950 x 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 (# models = 9).

Grouped SKAT-O Rare Variant Analyses in subset with verified European Ancestry

141,783 rare variants (MAF < 3%) were observed in our subset of 7,116 subjects with genetically-verified European ancestry (Table 1). We created two types of variant sets for testing, one based on proximity to genes and another 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 (Table 3; Supplemental Table E5). All grouped analyses were performed in R v3.3.0 using the package SKAT (15). We used a 0.05 significance threshold for Bonferroni adjusted p-values.

To identify rare variant sets independently associated with disease status after adjusting for the common signals found in single variant association tests, we re-tested 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.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

RESULTS

Common Variant Analyses

We tested for association between IPF and common variants (allele frequency ≥3%) using 8,066 subjects, including 3,624 cases of IPF (3,146 sporadic cases and 478 independent cases from families with ≥2 cases of idiopathic interstitial pneumonia) and 4,442 unaffected controls (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 2 and Figure 1). We did not identify significant common variants in the 6 resequencing regions designed around genes chosen based on 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 OR of 5.45 (95% CI: 4.91-6.06) for one copy of the risk allele and 18.68 (95% CI: 13.34-26.17) for two copies of the risk allele ($p=9.60 \times 10^{-295}$). Testing only significant common variants within each GWAS resequencing region, there were 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 2 are the only independent signals detected among common variants (Minor Allele Frequency (MAF) ≥ 3%). In addition, these variants are associated with similar risk in familial and sporadic IPF (see Table E3 in the online data supplement). The potential classification and prediction utility of these variants are explored via the 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 to other common variants, rs35705950 has the best potential for risk prediction (Figure 2). Since several of these variants have been identified previously in independent populations, the ROC curves and positive predictive values are not as susceptible to estimation bias that is usually present when estimating such values from the same sample as that used for association testing. In particular, 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 Methods). 141,783 rare variants were observed in our subset of 7,116 subjects and these variants were assembled into testing sets for grouped variant association analyses. We created two types of variant sets, one based on proximity to genes and another based on sliding windows. For gene-based sets, all rare variants within 20kb 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 overlap of 25 rare variants between adjacent windows. Thirty-four gene-based rare variant sets were

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

significantly associated with disease status after adjusting for sex and multiple comparisons (see Table E4 in the online data supplement); the genes were annotated to 7 of our resequencing regions. Twenty of the 34 significant gene-based sets were in the *MUC5B* region on 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 adjusting for sex and multiple comparisons within each region (see Table E5 in the online data supplement). Significant windows of variants were identified in 8 of the resequencing regions. Fifty-five of the significant windows were found on chr11, and 17 significant windows were on chr5. Significant windows of variants were also identified on chr4, chr6, chr10, chr13, chr16, and chr20.

Functional Variant Subset Analysis

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

Joint Common and Grouped Rare Variants

Four of the 34 significant gene-based sets were located on chromosome 20 (*ARFRP1, TNFRSF6B, ZGPAT, AL121845.3),* 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 chromosome 5 remained significant after adjustment for the top common variant located in the intronic region of the gene (p=0.01; see Table E7 in the online data supplement).

In total, we identified 12 rare-variant window-based sets with evidence of independent association with IPF (Table 3). Ten window-based sets in regions with a common signal were significant after adjusting for the top common variant in the region (Table 3). There were three significant windows within the gene body of *FAM13A* on chr4 after adjusting for the *FAM13A* intronic variant, rs2609260 (Figure 3); 2 of these windows 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' UTR region, exon 1, and intronic regions of *TERT*. Two other windows on chr5 are downstream of micro RNA *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 chromosome 11 was significant after

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

adjustment for the *MUC5B* promoter variant, rs35705950 (see Figure E1 in the online data supplement); 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 chromosome 13 was significant after adjusting for rs1278769 (3' UTR of *ATP11A*), and is located in an intron of *MCF2L* (Figure E1). Two other window-based sets on chromosomes 16 and 20 did not occur in regions with a significant common variant. The significant window on chromosome 16 was located within the intron of *RNPS1* (Figure E1), and the significant window on chromosome 20 overlaps *RTEL1*. The *RTEL1* significant window is also near *ARFRP1, TNFRSF6B, ZGPAT,* and *AL121845.3* such that it is 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 of the sets of rare variants included in the significant windows after adjustment for the common variant, we characterized each variant according to putative function and report the number of variants with a high or moderate (17) functional annotation (Table 3, sixth column). The most significant window in the *TERT* gene contained 27 variants with high or moderate functional impact (17) and the most significant window in the *RTEL1* gene contained 20 high or moderate impact variants.

eQTLs 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 in the online data supplement), as we have previously reported (6, 19). On chr15, rs35700143 is an eQTL for *BAHD1 (*see Figure E3 in the online data supplement), a nuclear protein that promotes heterochromatic gene silencing (20) and that when repressed, contributes to the induction of interferon (IFN)-stimulated genes (21). And on chr19, rs12601495 is an eQTL for *DPP9* (see Figure E4 in the online data supplement).

DISCUSSION

In the largest study of IPF to date, we have refined the signal of association across each of the regions we examined that were previously identified by 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 GWAS studies, showing the power of such studies among those with European Ancestry. However, in other regions such as those on Chromosome 7 and 15, the resequencing study provided substantial localization over the GWAS studies. Across all of the regions, we did not find evidence for 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 risk of IPF in the *TERT* and *RTEL1* gene regions, and found two regions with independent common and rare variant signals (*FAM13A* and *TERT* regions). The identification of rare variation in *FAM13A* associated with IPF is

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

particularly interesting in light of the opposite associations previously reported with emphysema and 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* previously reported as 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 are non-differential with respect to case-control status since cases and controls were randomized on plates and calls were made using the entire study sample jointly. While each of the cases and controls 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 Supplemental Methods). Since we do not have the same data on 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 can exclude neither the possibility of modest rare variant effects on IPF risk in these regions nor 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, ER 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), 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

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

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 non-genetic 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 lung eventually coalesce and present as IPF. Our aggregate genetic findings across multiple studies suggest that reduced host defense plays a dominant role in this dynamic relationship with the environment, however, cell senescence also appears to be a critical mechanism 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.

REFERENCES

1. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, Swigris JJ, Taniguchi H, Wells AU. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017; 3: 17074.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi PG, Scuirba J, Richards T, Juan-Guardela B, Vij R, Han MK, Martinez FJ, Kossen K, Seiwert SD, Christie JD, Nicolae DL, Kaminski N, Garcia JG. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *The Lancet Respiratory Medicine* 2013; 1: 309-317.
8. 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 genetics* 2016; 17: 74.
9. 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,

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

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. *The Lancet Respiratory medicine* 2017; 5: 869-880.

10. Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L, Fan Y, Zhang D, Wang Y. 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, Kohno N, Guzman J, Costabel U. 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, 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. *The Lancet Respiratory medicine* 2018.
13. 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.
14. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, 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.
15. SL S, L M, M W. SKAT: SNP-Set (Sequence) Kernel Association Test. 2017. 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, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. Ensembl 2018. *Nucleic Acids Res* 2018; 46: D754-d761.
17. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 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* 2012; 6: 80-92.
18. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, Biospecimen Collection Source Site N, Biospecimen Collection

Source Site R, Biospecimen Core Resource V, Brain Bank Repository-University of Miami Brain Endowment B, Leidos Biomedical-Project M, Study E, Genome Browser Data I, Visualization EBI, Genome Browser Data I, Visualization-Ucsc Genomics Institute UoCSC, Lead a, Laboratory DA, Coordinating C, management NIHp, Biospecimen c, Pathology, e QTLmwg, Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* 2017; 550: 204-213.

19. Mathai SK, Pedersen BS, Smith K, Russell P, Schwarz MI, Brown KK, Steele MP, Loyd JE, Crapo JD, Silverman EK, Nickerson D, Fingerlin TE, Yang IV, Schwartz DA. 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, Kerneis-Golsteyn S, Regnault B, Seeler JS, Muchardt C, Feunteun J, Cossart P. Human BAHD1 promotes heterochromatic gene silencing. *Proc Natl Acad Sci U S A* 2009; 106: 13826-13831.
21. Lebreton A, Lakisic G, Job V, Fritsch L, Tham TN, Camejo A, Mattei PJ, Regnault B, Nahori MA, Cabanes D, Gautreau A, Ait-Si-Ali S, Dessen A, Cossart P, Bierne H. 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, Bosse Y, Shrine N, Artigas MS, Wain LV, Hall IP, Jackson VE, Wyss AB, London SJ, North KE, Franceschini N, Strachan DP, Beaty TH, Hokanson JE, Crapo JD, Castaldi PJ, Chase RP, Bartz TM, Heckbert SR, Psaty BM, Gharib SA, Zanen P, Lammers JW, Oudkerk M, Groen HJ, Locantore N, Tal-Singer R, Rennard SI, Vestbo J, Timens W, Pare PD, Latourelle JC, Dupuis J, O'Connor GT, Wilk JB, Kim WJ, Lee MK, Oh YM, Vonk JM, de Koning HJ, Leng S, Belinsky SA, Tesfaigzi Y, Manichaikul A, Wang XQ, Rich SS, Barr RG, Sparrow D, Litonjua AA, Bakke P, Gulsvik A, Lahousse L, Brusselle GG, Stricker BH, Uitterlinden AG, Ampleford EJ, Bleecker ER, Woodruff PG, Meyers DA, Qiao D, Lomas DA, Yim JJ, Kim DK, Hawrylkiewicz I, Sliwinski P, Hardin M, Fingerlin TE, Schwartz DA, Postma DS, MacNee W, Tobin MD, Silverman EK, Boezen HM, Cho MH. 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, 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.
24. 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.
25. 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.
26. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L, Yang IV, Schwartz DA. Idiopathic Pulmonary Fibrosis: A Genetic Disease That Involves Mucociliary Dysfunction of the Peripheral Airways. *Physiological reviews* 2016; 96: 1567-1591.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

27. Hancock LA, Hennessy C, Solomon GM, Dobrinskikh E, Estrella A, Hara N, Hill DB, Kissner WJ, Markovetz MR, Villalon DE, Voss ME, G.J. T, Carroll KS, Shi Y, Schwarz MI, Thelin WR, Rowe SM, Yang IV, Evans CM, Schwartz DA. Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. *Under Review at Nature Communications* 2018.
28. Nakano Y, Yang IV, Walts AD, Watson AM, Helling BA, Fletcher AA, Lara AR, Schwarz MI, Evans CM, Schwartz DA. 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, Boucher RC, Rubinstein M. A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* 2012; 337: 937-941.

TABLE 1. DEMOGRAPHICS (FULL AND SUBSET, BY CASE/CONTROL)

Demographics	Full Study		Rare Variant Subset	
	IPF Cases N = 3,624	Controls N = 4,442	IPF Cases N = 3,023	Controls 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				
Female	1033 (28.5)	2610 (58.8)	884 (29.2)	2366 (57.8)
Male	2591 (71.5)	1832 (41.2)	2139 (70.8)	1727 (42.2)

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

TABLE 2. Top Common Variant Associations

SNP	Position[a]	Locus	Nearest Gene	Annotation	Minor Allele	MAF in cases	OR Aa vs AA (95% CI)	OR aa vs AA (95% CI)	*P*[b]
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[c]	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)[d]	-	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[c]	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[c]	4717672	19p13	*DPP9*	Intronic	G	0.34	1.22 (1.11-1.35)	1.59 (1.36- 1.87)	3.11×10^{-9}

OR, odds ratio. The minor allele is defined as the minor allele in the combined case and control group.
[a]Based on NCBI Build 37. [b]Adjusted for sex. [c] Same SNP as identified in original GWAS (Fingerlin et al. 2013) [d]OR of a vs A resulting from dominant test.

TABLE 3. Significant Rare Variant Windows

Nearest Gene[a]	Locus	Start (bp)	Stop (bp)	Annotation	# Putative Functional variants[b]	# Windows in region	p-value[c]	Adjusted Common Variant	p-value[d]
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	1294	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

[a]Nearest gene to the window positions using Ensembl 75
[b]Annotation withSnpEff, putative functional variants defined as high or moderate impact
[c]Adjusted for sex with Bonferroni correction for the number of windows in the given region
[d]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

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE 1. Regional Association for Common Variants

FIGURE 2. Receiver Operator Characteristic (ROC) Curve and Positive Predictive Value

FIGURE 3. Rare variant Associations at FAM13A (top), TERT (middle) and RTEL (bottom) regions

FIGURE LEGENDS

Figure 1. 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. SNP color represents LD with the most highly associated SNP at each locus.

Figure 2. Receiver Operator Characteristic (ROC) Curve and Positive Predictive Value 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.

Figure 3. Rare variant signals at *FAM13A* (top), *TERT* (middle), *RTEL1* (bottom) regions. For both plots, the -$\log_{10}$ adjusted P values (y axis) from SKAT-O of the gene (green) and window (blue) sets are shown according to their chromosomal positions (x axis). The gene sets are labeled with their gene name where the dark green region represents the gene body and the lighter region represents the 20kb region upstream/downstream region. Rare variants located within the gene ± 20kb were included in the testing for that gene. For the *FAM13A* and *TERT* regions, results correspond to adjusting for sex and the common variant (rs2609260 and rs4449583, respectively, for *FAM13A* and *TERT*) where the *RTEL1* region results are adjusted only for sex. The horizontal red line corresponds to significance level of 0.05.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

ACKNOWLEDGEMENTS

We gratefully acknowledge the individuals who participated in the studies that contributed to this work and all the clinical and research staff at each site for their efforts.

FUNDING

This research was supported by the National Heart, Lung and Blood Institute (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 National Heart, Lung, and Blood Institute. 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.

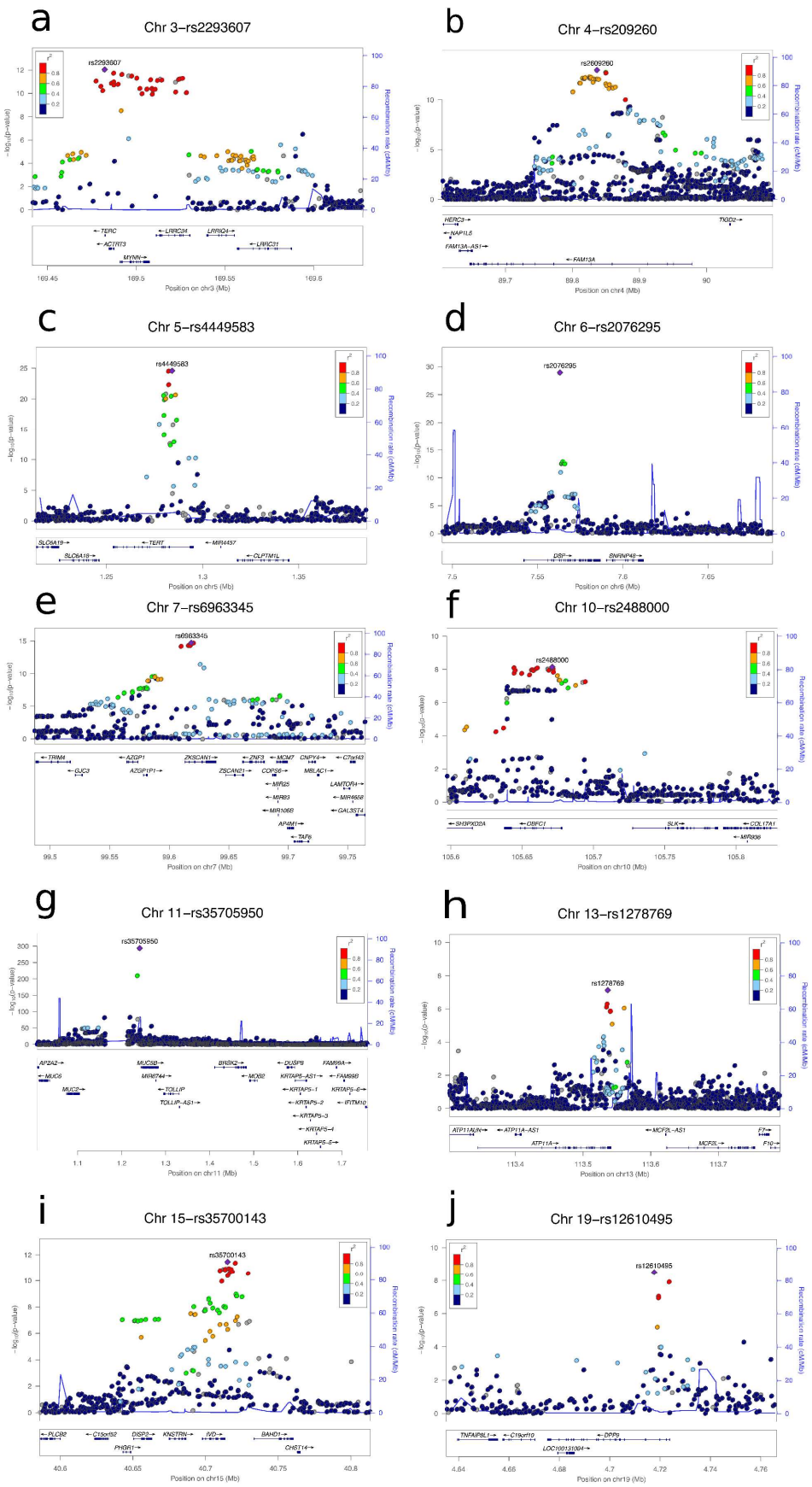
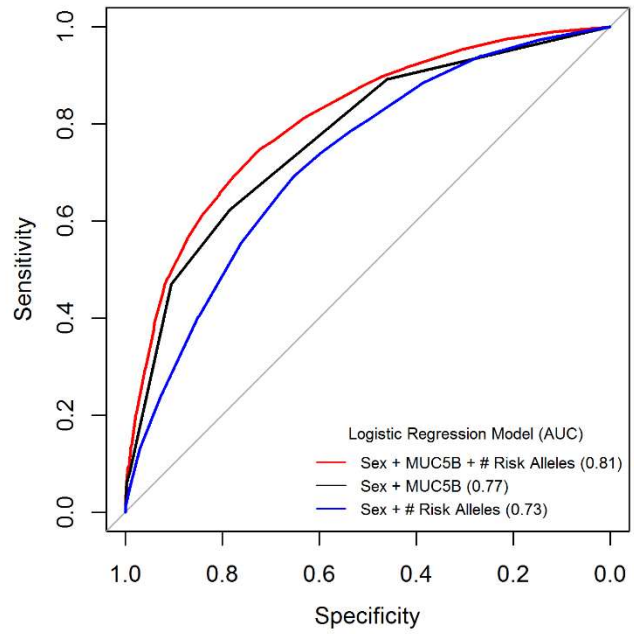


Fig 1

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE 2. Receiver Operator Characteristic (ROC) Curve and Positive Predictive Value



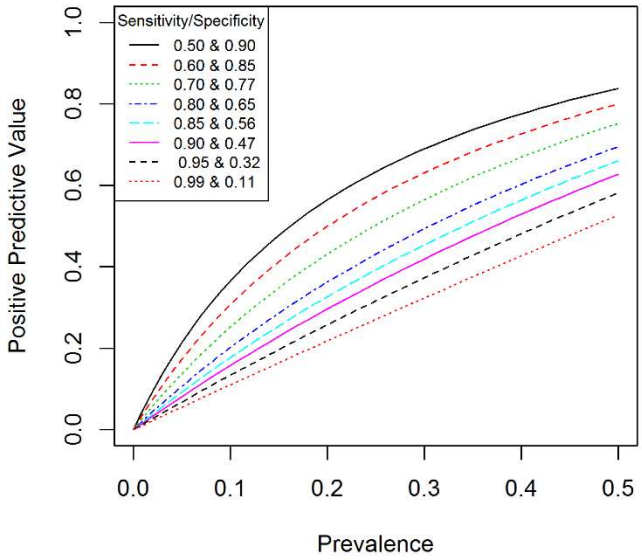
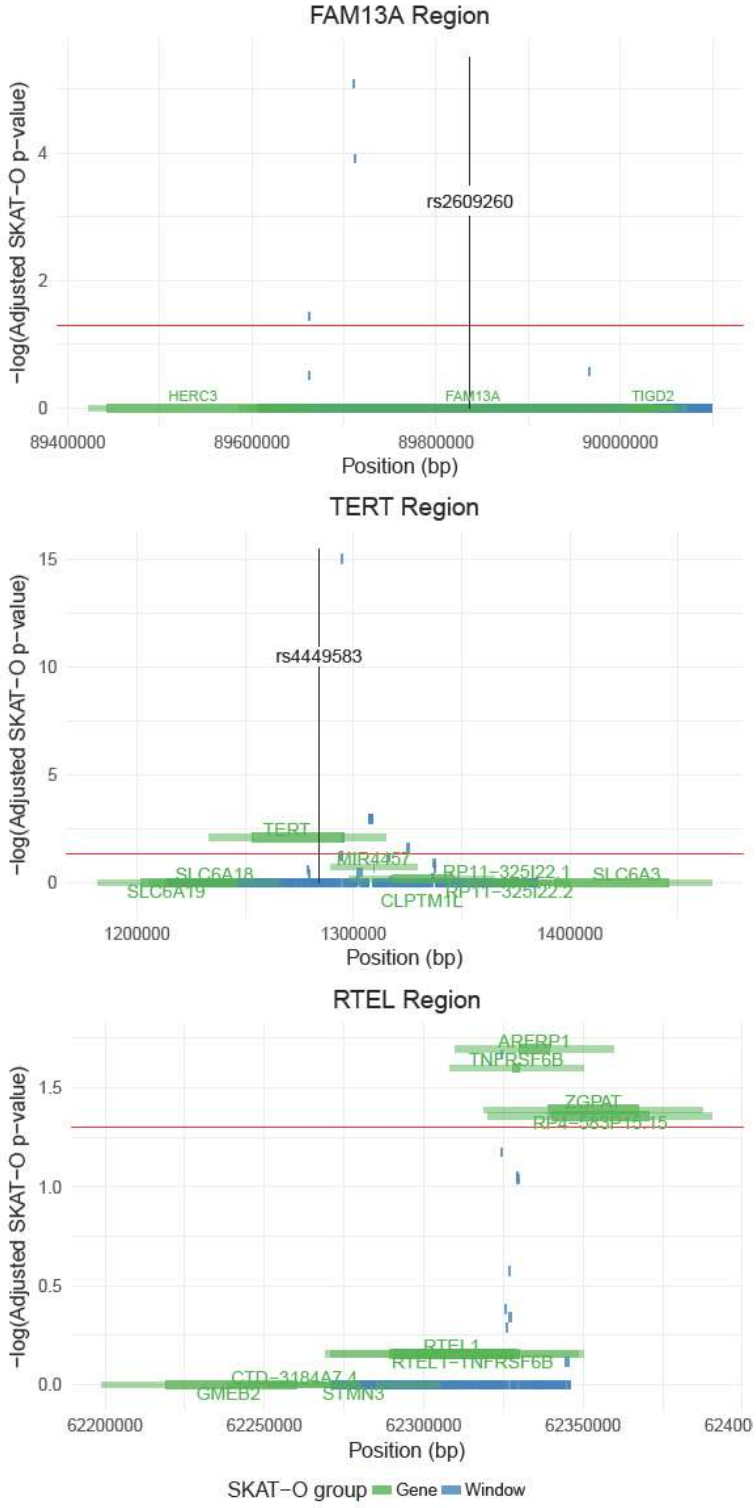


FIGURE 3. Rare variant signals at FAM13A (top), TERT (middle) and RTEL (bottom) regions



AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

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

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

ONLINE DATA SUPPLEMENT

ONLINE METHODS

STUDY PARTICIPANTS

We used standard criteria established by the American Thoracic Society/European Respiratory Society/Japanese Respiratory Society to determine the diagnostic classification and select IPF cases (Table E1) (1). We excluded cases with known explanations for the development of fibrotic IIP, including infections, systemic disorders or relevant exposures (e.g. asbestos). Eligible control subjects from the COPDGene Study were ages 45 to 80, with at least 10 pack-years of smoking. For this analysis, only non-Hispanic White subjects with normal post-bronchodilator spirometry were included. The other controls were collected as part of separate studies of auto-immune genetics studies. These controls were selected to have no known auto-immune disease and had self-reported non-Hispanic European ancestry. To maximize power and minimize potential confounding by ancestry, we included only self-reported non-Hispanic, white participants. All subjects provided written informed consent as part of institutional review board (IRB)-approved protocols for their recruitment at their respective institution, and the resequencing study was approved by the National Jewish Health IRB and the University of Colorado Combined Institutional Review Board (COMIRB).

DEFINITION OF RESEQUENCING REGIONS

To define a target region around each of 10 genetic loci of interest (3q26, 4q22, 5p15, 6p24, 7q22, 10q24, 11p15, 13q34, 15q14-15, and 19p13; (2, 3)), we followed a multi-step process. We defined target regions by extending from either side of each of the 10

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

top GWAS SNPs that reached genome-wide significance in a meta-analysis (2). We extended the region sequentially if there was another SNP with $P<1 \times 10^{-04}$ within 500 Kb of the current boundary, then we extended these boundaries by an additional 20 Kb on each side. If either end overlapped a gene using the Ensemble v83 hg19 gene annotation, the bounds were extended to the end of that gene. Next, each region was extended by 20 Kb on each side, and then each region was manually extended to the end of the LD block on either side defined visually by the hg19 linkage data from 1,000 Genomes CEU using the D' statistic. Finally, all regions were extended an additional 20 Kb on either end. We additionally targeted regions around 6 genes of interest based on evidence for rare variation associated with IPF (*RTEL1*, *PARN*, *TINF2*, *SFTPC*, *SFTPA2*, and *ABCA3*), defining region boundaries manually by centering a region around each gene with at least an additional 18 Kb flank on each end of the gene. The targeted regions that were resequenced include 3.69 Mb DNA, and are presented in Table E2.

DNA Preparation

Genomic DNA was isolated from whole blood stored in either K_2 EDTA tubes or PAX DNA tubes (Cat #761115, Qiagen). For blood collected in K_2 EDTA tubes, genomic DNA was isolated with the QIAamp DNA blood midi kit (Cat# 51183, Qiagen, San Diego, CA). For blood collected in PAX DNA tubes, genomic DNA was isolated with the PAXgene Blood DNA Kit (Cat #761133, Qiagen). DNA concentrations were measured with the Qubit™ dsDNA BR Assay Kit (Cat #Q32853, ThermoFisher Waltham, MA). A Freedom EVO Liquid Handling System (TECAN, Baldwin Park, CA) was use to aliquot

500ng of DNA (50uL at 100ng/uL) to 96 well plates. To minimize any batch effects, an equal number of case and control samples were distributed in a random order on each plate when possible. In addition, one case and one control sample was plated twice on each plate to confirm the plate's orientation and screen for any errors in the sample manifest.

Northwest Genomics Center (NWGC)

All targeted capture sequencing was performed at the University of Washington Northwest Genomics Center (NWGC). Samples were assigned unique barcode tracking numbers and had a detailed sample manifest (i.e., identification number/code, sex, DNA concentration, barcode, extraction method) linked to each sample within our laboratory information management system. Initial QC entailed DNA quantification and a gender validation assay. Samples were failed if: (1) the total amount, concentration, or integrity of DNA was too low; and/or (2) sex-typing is inconsistent with the sample manifest.

Library Production and Targeted Capture

Library construction and targeted capture were conducted via an automated (Perkin-Elmer Janus II) process in 96-well plate format. 500ng of genomic DNA was subjected to a series of shotgun library construction steps, including fragmentation through acoustic sonication (Covaris), end-polishing and A-tailing ligation of sequencing adaptors and PCR amplification with 8bp unique dual barcodes for multiplexing. Libraries undergo targeted capture using a Roche/Nimblegen Seqcap EZ custom designed probe (5.5 mb). Briefly, 500ng of shotgun library was hybridized to biotinylated

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

capture probes for 72 hours. Enriched fragments were recovered via streptavidin beads and PCR amplified. To facilitate optimal flow cell loading, the library concentration was determined by triplicate qPCR and molecular weight distributions verified on the Agilent Bioanalyzer (consistently 150 ± 15bp).

Clustering/Sequencing

Barcoded libraries were pooled using liquid handling robotics prior to clustering (Illumina cBot) and loading. Massively parallel sequencing-by-synthesis with fluorescently labeled, reversibly terminating nucleotides was carried out on the HiSeq 4000 sequencer, generating 100x100bp paired end reads. The mean number of reads per sample was 8,246,968 (4,123,484 pairs). Figure E8 shows the distribution of read counts by sample. We reverted the bam files provided by the sequencing center to fastq format using SamToFastq in Picard Tools version 1.114 (https://software.broadinstitute.org/gatk/documentation/tooldocs/4.0.2.0/picard_sam_SamToFastq.php). We aligned the reads to hg19 with BWA-MEM version 0.7.12 (4). We called variants with the Genome Analysis Toolkit (GATK) version 3.5 (5), following the best practices workflow (6). Our variant calling pipeline identified 39,505 existing dbSNP (build 137) variants and 274,524 novel variants in the targeted regions. Overall, 89% of identified variants are single nucleotide polymorphisms. Our callset has a Ti/Tv ratio of 2.42 for dbSNP variants, 2.26 for novel variants, and 2.28 overall. The concordance rate with dbSNP build 137 was 89.32%. We converted the final VCF file to PLINK format for downstream analysis using PLINK version 1.9 (7). Our complete variant calling pipeline with command lines, including non-GATK steps, is available at

https://github.com/FingerlinGroup/variant-calling-pipeline-IPF-resequencing/releases/tag/1.0.

ANCESTRY-INFORMATIVE MARKERS

We obtained genotypes at 116 ancestry-informative markers (AIMs), 91 from the panel for determining continental ancestry (8) and 25 for fine mapping non-Hispanic White ancestry (9). Prior to genotyping, all samples were quality controlled by real-time Q-PCR quantitation ("QC1") and uniplex genotyping using Taqman ("QC2"). Samples that failed QC1 or QC2, although carried forward through genotyping, were later removed from analysis.

AIMs genotyping was accomplished with Sequenom iPLEXassays. 116 SNPs were efficiently placed into a set of 4 multiplex wells with 37, 35, 31 and 13 SNPs in each. Sequenom iPLEX genotyping is based on multiplexed locus-specific PCR amplification, multiplexed single-based extension (SBE) from locus-specific amplicons, and multiplexed resolution of SBE products base calling using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOD) mass spectrometry.

Primers for the Sequenom assay were purchased from IDT (Coralville, Iowa), and all steps of the iPLEX procedure were carried out using reagents and methods from Sequenom (San Diego, CA) according to the manufacturer's instructions. Reactions were carried out in 384-well plates and analyzed using the Sequenom MassARRAY Analyzer 4 system with iPLEX Gold reagents and SpectroCHIP arrays. Results were

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

analyzed using a combination of commercial software (Typer 4, Sequenom) and custom tools for data management.

STATISTICAL ANALYSIS

Exclusion of Variants

206,767 variants were observed in our final analysis cohort. We limited our analysis to bi-allellic variants, including insertions and deletions (<50 bp; N= 180,370). In addition, we observed 1,090 multi-allelic variants where two alleles were observed at frequencies greater than 3% and the other alleles were only observed once. For the purposes of our analysis, we treated these variants as bi-allellic by excluding the subject(s) with the single copy allele(s) from the statistical testing for that variant.

Variants missing in more than 5% of subjects were excluded from analysis. In addition, we excluded variants exhibiting differential missingness between cases and controls as follows: for each variant, we compared rates of missingness between cases and controls using a $\chi 2$ test. For variants with missing rates between 2% and 5%, those with a $\chi 2$ test p-value less than 0.05 were excluded. For variants with less than 2% missing, those with a p-value less than 0.001 were excluded. We also evaluated departures from HWE separately in cases and controls via a 1-degree-of-freedom $\chi 2$ goodness-of-fit test and excluded variants with p-values less than 0.0001 in controls.

Variants that met the above criteria with MAF greater than 3% were classified as “common” and were tested in single variant association analyses, while those with MAF

less than 3% were considered “uncommon/rare” and were tested in grouped analyses (see below). MAF was calculated on cases and controls combined.

Variant QC for AIM’s

Genotype data on continental AIM’s were excluded if they failed QC1, QC2, or had < 90% call rate. For subjects with duplicate genotype data, we selected the sample with the highest call rate. This resulted in 5,271 subjects with continental AIMs data.

Exclusion of Individuals

Resequencing data was available for 8,883 samples from 8,367 subjects. Twenty-three samples from 21 subjects had low capture efficiency (< 40% on target reads) and were removed from the analysis (8,860 samples from 8,346 subjects). We calculated kinship using KING v2.1 software between duplicates (subjects repeated either within a plate or across plates) to confirm sample identity and identified a further 34 samples that had uncertain subject identity (8,826 samples from 8,325 subjects; (10)). For intentional duplicates confirmed via kinship, we selected the subject’s sample having the highest percent on-target reads (N = 8,325). We estimated pairwise kinship coefficients on the resulting 8,325 samples using the same procedure as for duplicates and flagged groups of samples with a kinship > 0.35. For groups of samples with matching phenotype and sex, we kept one sample and removed the remaining; otherwise, all samples were removed. This resulted in removal of 159 samples (N = 8,166). When available, we used AIMs data to confirm European ancestry, as not all subjects had prior GWAS data. AIMs from 1000Genomes Phase 3 data on 2,504 individuals were used as

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

representatives of global populations to infer ancestry-informative principal components (PCs). We projected the first two PCs onto the subset of our study population with AIMs data (11). Putative non-European samples having PC1 or PC2 > 4 s.d. of the European reference groups were flagged as outliers and removed from all subsequent analyses (N=100). This resulted in a final analysis cohort of 8,066 subjects, including 3,624 cases of IPF (3,146 sporadic cases and 478 cases from families with ≥2 cases of idiopathic interstitial pneumonia) and 4,442 unaffected controls (Figure E5). The QC steps above are outlined in Figure E5 in the online data supplement.

Additional Exclusions for Rare Variant Testing

We could not verify European ancestry through genetic data for 950 subjects who either (i) did not have AIMs or (ii) were not included in previous GWAS studies. Since the common variant analyses are a follow-up of replicated associations that had been adjusted for fine-scale European ancestry and the total number of individuals mis-classified is likely less than 10% based on our other studies, we did not remove these individuals from the common variant analyses. Because rare variant analyses are more susceptible to false positive results due to even very low levels of population stratification, these subjects were excluded from the rare variant analyses. All grouped variant analyses were performed on the subset of 7,116 subjects with verified European ancestry.

SINGLE VARIANT ASSOCIATION ANALYSES

We tested for association between each common variant and disease status using a logistic regression framework. We assumed a dominant genetic model for variants with MAF < 10%. For variants with MAF ≥ 10%, we used a genotypic (two degree-of-freedom) model, which allows a more general consideration of genetic effects. All models were adjusted for sex. To test whether a variant was associated with disease status controlling for sex, we performed likelihood ratio tests (LRT) comparing the full models, which included sex and the variant as predictors, to 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 since each of the loci represented a previously replicated association signal.

To determine if independent association signals with more than one variant were present in each region, variants with significant LRT p-values were carried forward into region-specific conditional models. Within each region, we first identified the top variant as the variant with the smallest LRT p-value. For all significant common variants in the region, we then re-fit logistic regression models as described above, now adjusting for both the top regional variant and sex. Likelihood ratio tests were used to test whether a variant was associated with disease status after adjusting for the top variant and sex.

To investigate whether the *MUC5B* promoter variant, rs35705950, modified the effect of the top common variant in each region, we fit logistic regression models that included sex, rs35705950, the top variant, and an rs35705950 x top variant interaction. For variants with MAF ≥ 10%, we assumed an additive genetic model, and we assumed a

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

dominant genetic model for variants with MAF < 10%. 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. Logistic regression models were fit in R v3.3.0 using the stats package and likelihood ratio tests were performed with the lmtest package (12).

GROUPED (SKATO) VARIANT ASSOCIATION ANALYSES ON AIMS SUBSET

141,783 rare variants were observed in our subset of 7,116 subjects with genetically-verified European ancestry. These variants were assembled into testing sets for grouped variant association analyses. We created two types of variant sets, one based on proximity to genes and another based on sliding windows. For gene-based sets, all rare variants within 20kb upstream (or downstream) of the start (or stop) position of a gene were grouped together based on Ensembl 75 annotation (13). For window-based sets, we constructed sliding windows of 50 consecutive rare variants with overlap of 25 rare variants between adjacent windows. Association between disease status and each rare variant set was tested in SKAT-O, adjusting for sex (14). SKAT-O combines a burden test and a non-burden sequence kernel association test (SKAT) to maximize power to detect association between rare variants and phenotypes. In total, a rare-variant signal was tested in 206 gene-based sets and 5,677 windows. To adjust for multiple comparisons, we applied a Bonferroni correction. P-values for gene-based sets were adjusted for 206 total tests across all regions, while the window-based tests were adjusted for the total number of windows within a region. All grouped analyses were

performed in R v3.3.0 using the package SKAT (15). We used a 0.05 significance threshold for the Bonferroni adjusted p-values.

Exon-Only Grouped Variant Association Analyses

In our primary grouped variant analyses, we considered a broad definition of a gene in order to form gene-based analysis sets. Grouping variants into these relatively large regions could potentially result in a loss of power as they presumably carry a large number of non-causal variants. As a sensitivity analysis, we considered a more restrictive definition and created gene-based sets that only included functional rare variants that occurred in the exon of each gene. This resulted in a total of 100 functional gene-based sets. For all variants, effect prediction was determined using SnpEff v4.3t (16). Variants with an impact score of high (exp. stop gained, frameshift variant) or moderate (exp. missense variant, in-frame deletion) were considered functional.

JOINT SINGLE AND GROUPED VARIANT ASSOCIATION ANALYSES

To identify rare variant sets independently associated with disease status, after adjusting for the common signals found in single variant association tests, we re-tested 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.

POWER ANALYSES JUSTIFYING ANALYTIC APPROACH

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

To develop our statistical analysis plan, we calculated power to detect dominant, additive, and recessive odds ratios of 1.5 across a range of minor allele frequencies using a likelihood ratio test. We performed these calculations for three different logistic regression models, each employing different coding of the genetic effect: a genotypic or 2 degree of freedom model, an additive model, and a dominant model. We also calculated detectable odds ratios at 80% power for each model. All calculations assumed an alpha level of 5×10^{-6} and a total sample size of 8,066 subjects (3,624 cases and 4,442 controls). When the true underlying genetic effect is recessive, the two degree of freedom model has higher power and can detect smaller odds ratios than the additive and dominant models. However, when the true genetic effect is dominant or additive, all three logistic regression models (2 degree of freedom, additive and dominant), have similar power for minor allele frequencies greater than 10%. Therefore, we decided to use the two degree of freedom test for variants with minor allele frequencies greater than 10%. For minor allele frequencies less than 10%, we chose to use the dominant logistic regression model since: 1. dominant and additive models have similar power to detect additive and dominant effects and 2. the number of subjects homozygous for the minor allele is limited at low minor allele frequencies.

REFERENCES

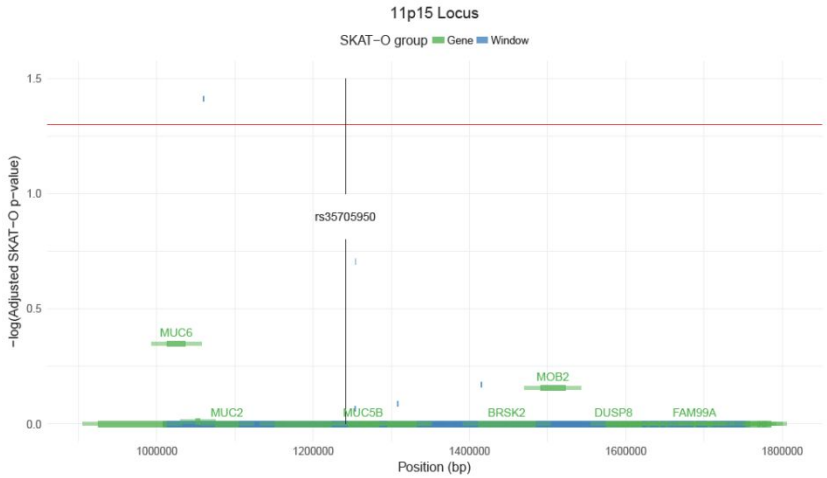
1. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, Flaherty KR, Wells A, Martinez FJ, Azuma A, Bice TJ, Bouros D, Brown KK, Collard HR, Duggal A, Galvin L, Inoue Y, Jenkins RG, Johkoh T, Kazerooni EA, Kitaichi M, Knight SL, Mansour G, Nicholson 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. 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.
3. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi PG, Scuirba J, Richards T, Juan-Guardela B, Vij R, Han MK, Martinez FJ, Kossen K, Seiwert SD, Christie JD, Nicolae DL, Kaminski N, Garcia JG. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *The Lancet Respiratory Medicine* 2013; 1: 309-317.
4. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754-1760.
5. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; 20: 1297-1303.
6. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43: 491-498.
7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559-575.
8. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, Seldin MF. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human mutation* 2009; 30: 69-78.
9. Huckins LM, Boraska V, Franklin CS, Floyd JA, Southam L, Gcan, Wtccc, Sullivan PF, Bulik CM, Collier DA, Tyler-Smith C, Zeggini E, Tachmazidou I, Gcan, Wtccc. Using ancestry-informative markers to identify fine structure across 15 populations of European origin. *Eur J Hum Genet* 2014; 22: 1190-1200.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

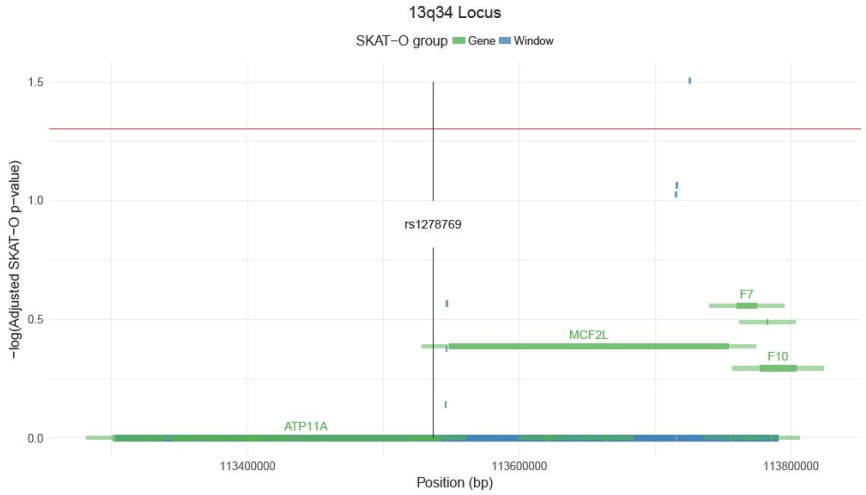
10. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010; 26: 2867-2873.
11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904-909.
12. A. Z, T. H. Diagnostic Checking in Regression Relationships. 2002.
13. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. Ensembl 2018. *Nucleic Acids Res* 2018; 46: D754-d761.
14. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, 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.
15. SL S, L M, M W. SKAT: SNP-Set (Sequence) Kernel Association Test. 2017. Available from: https://cran.r-project.org/web/packages/SKAT/index.html.
16. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 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* 2012; 6: 80-92.

FIGURE E1.

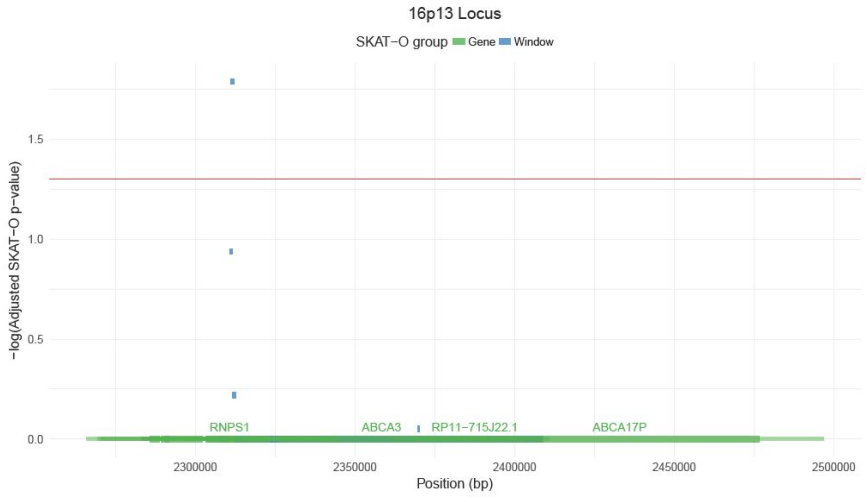
A.



B.



C.



AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE E2.

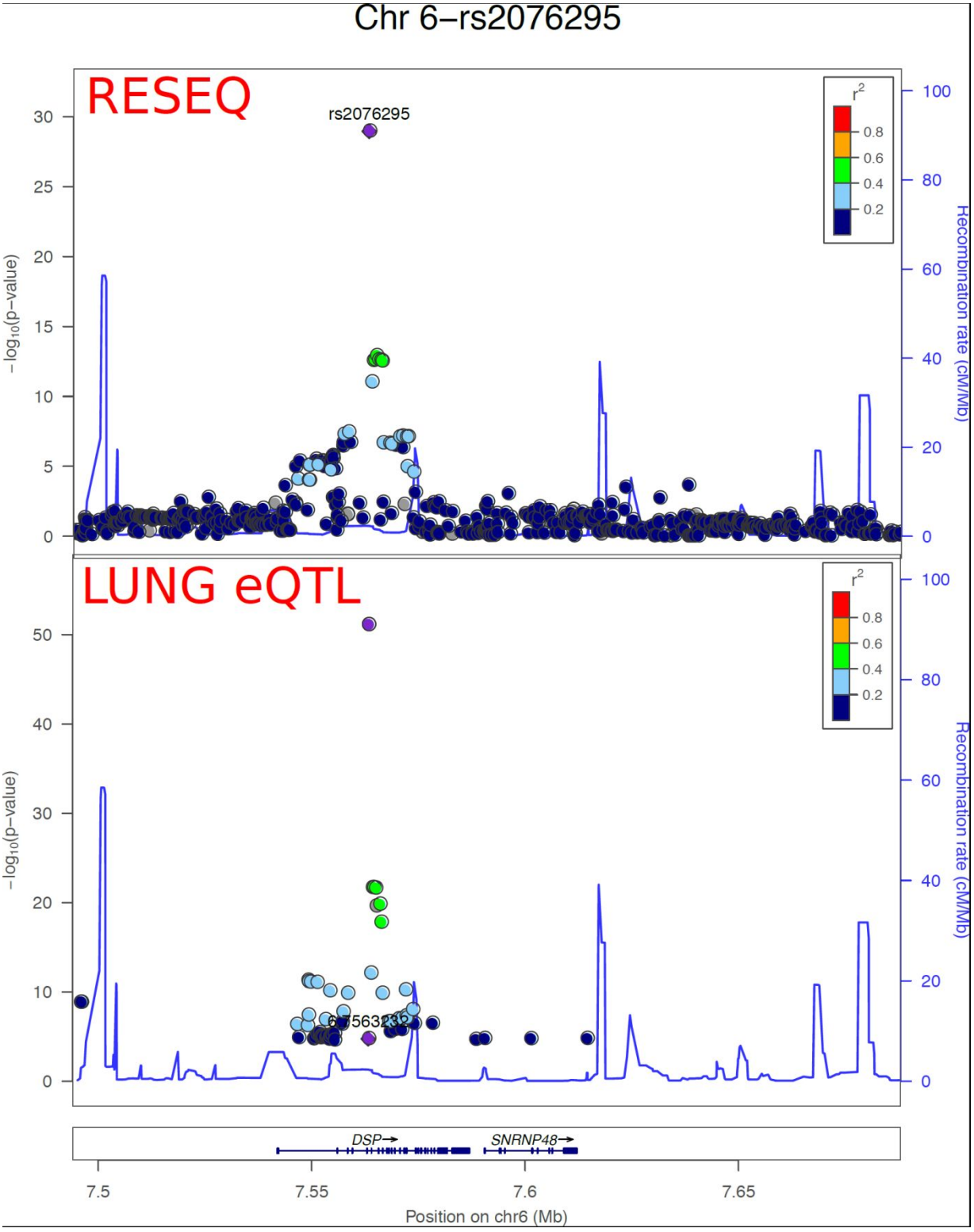
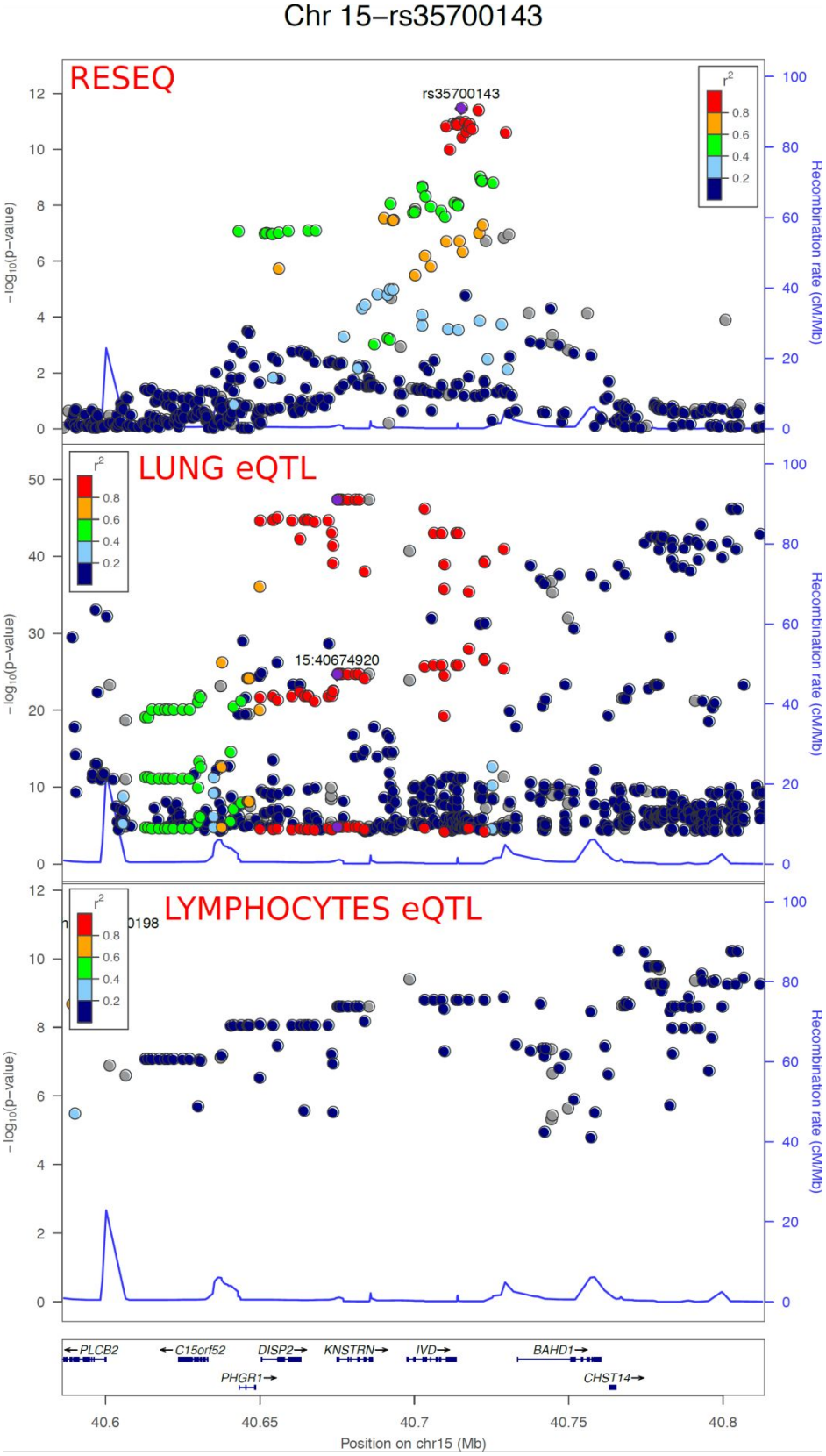


FIGURE E3.



AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE E4.

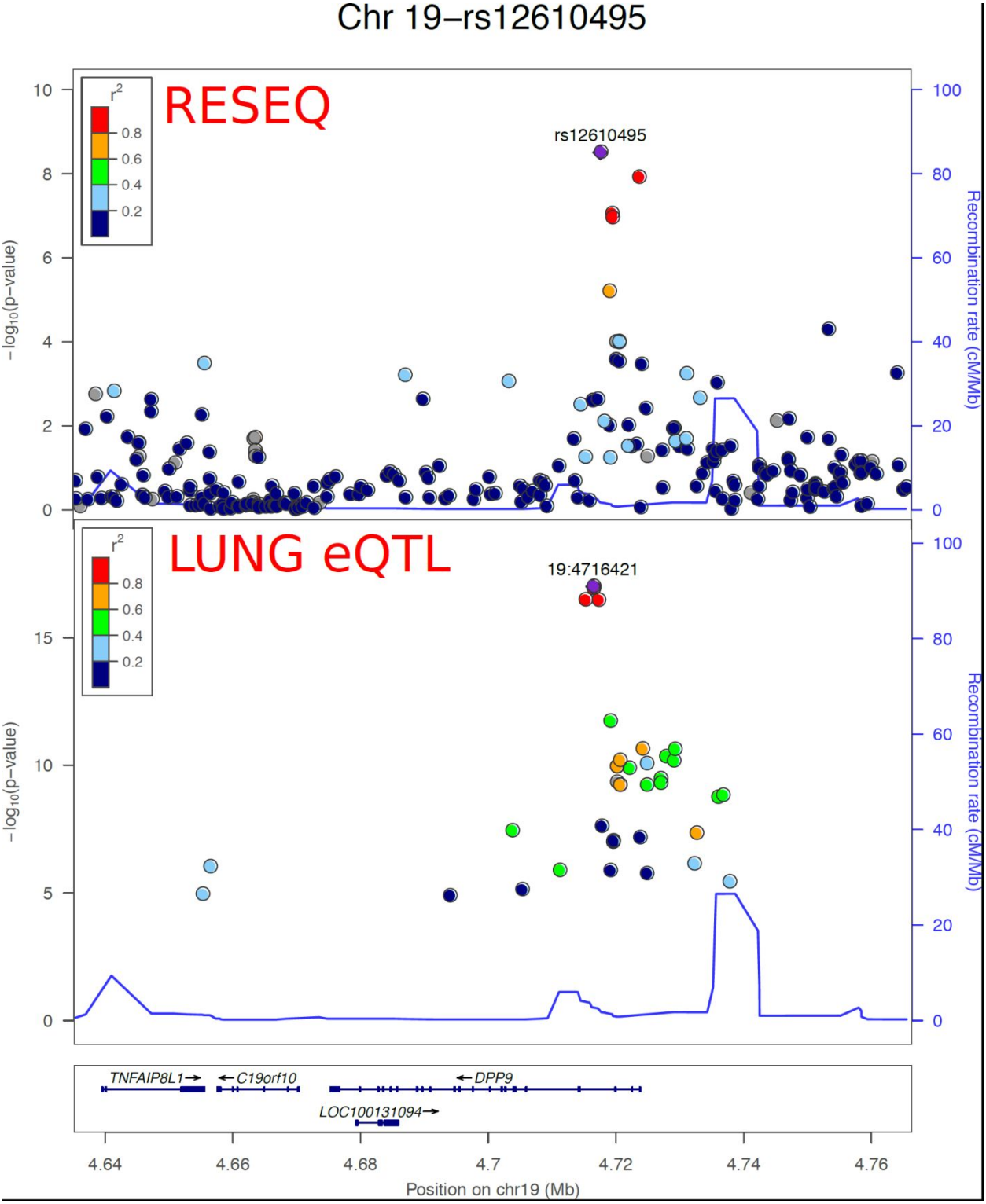
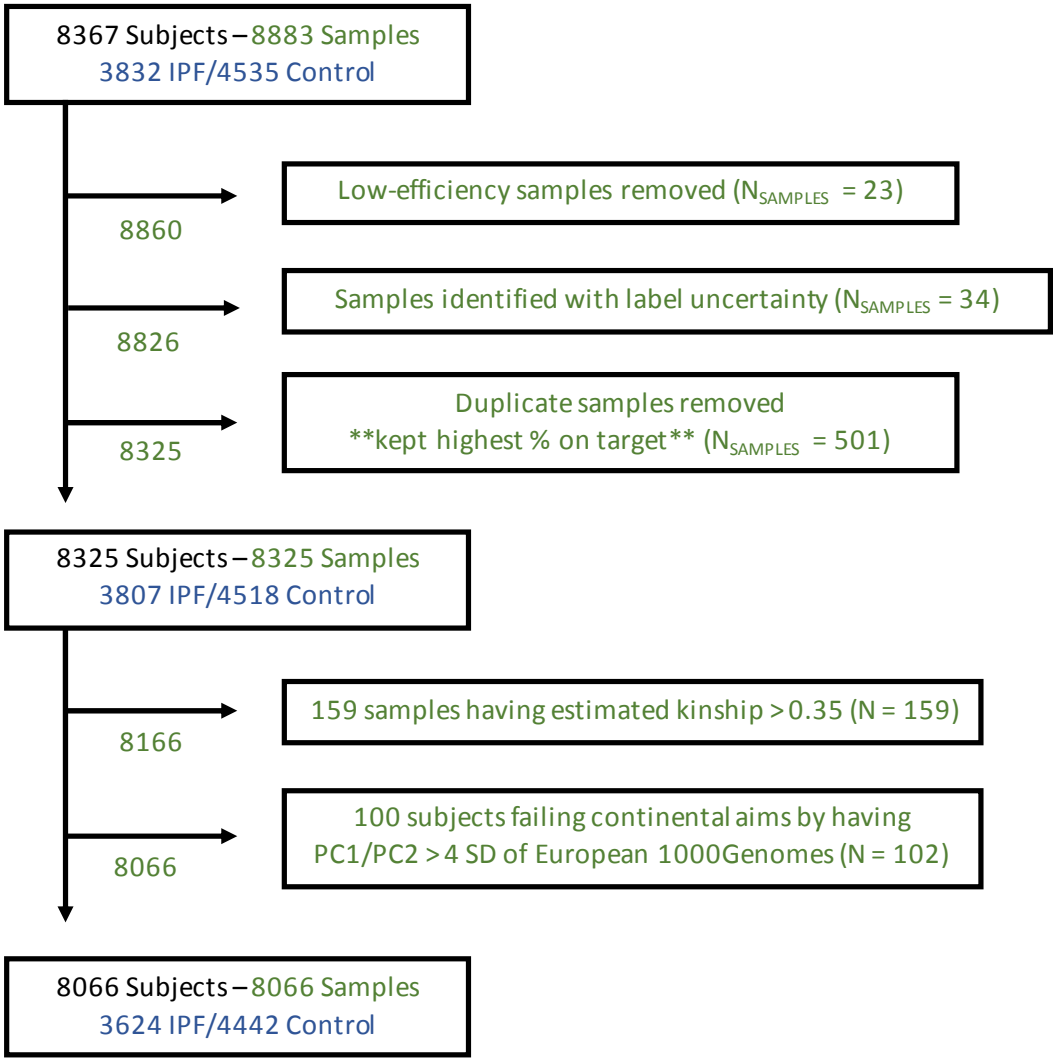


FIGURE E5. CONSORT DIAGRAM



AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE E6.

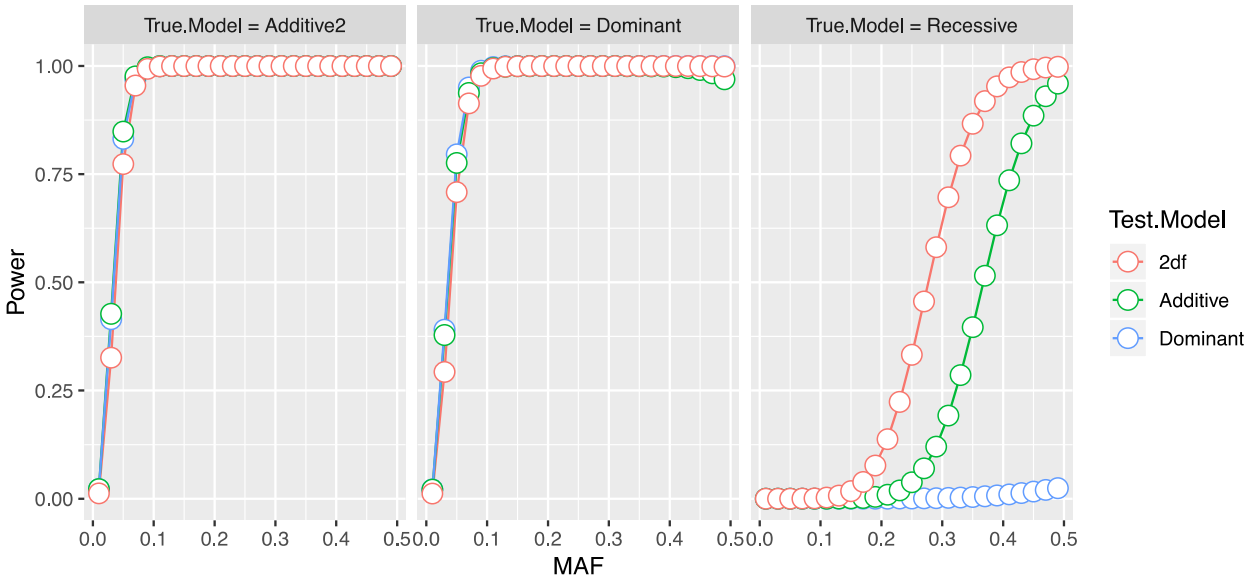
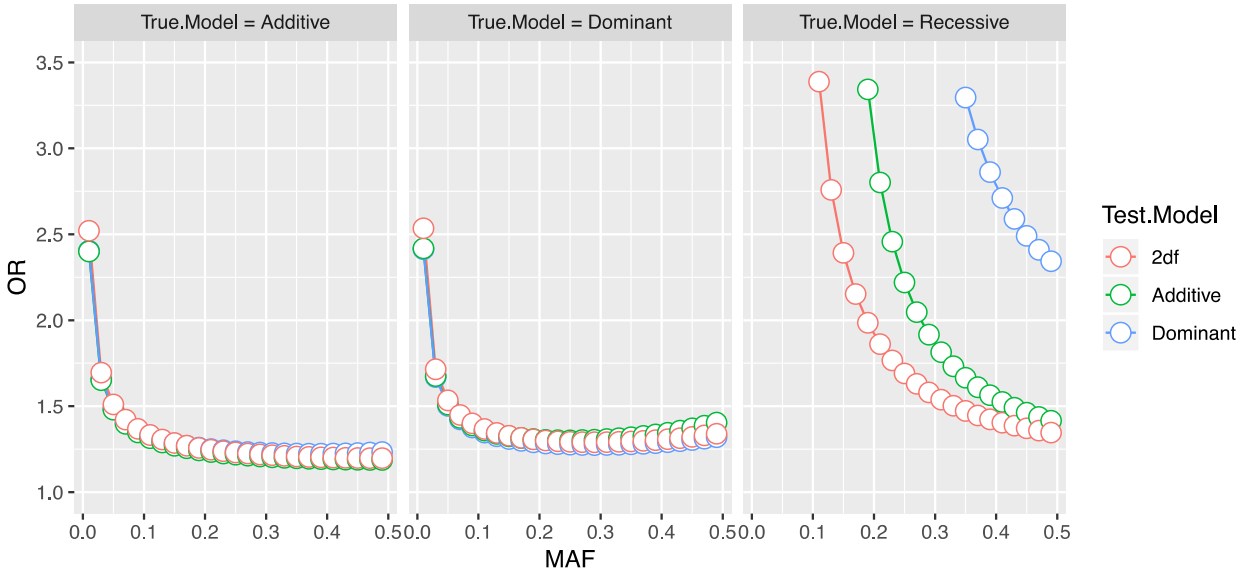


FIGURE E7.



AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

FIGURE E8. The distribution of number of sequenced reads per sample (N = 9,280) is shown. Individual reads are counted such that a read pair equals two reads.

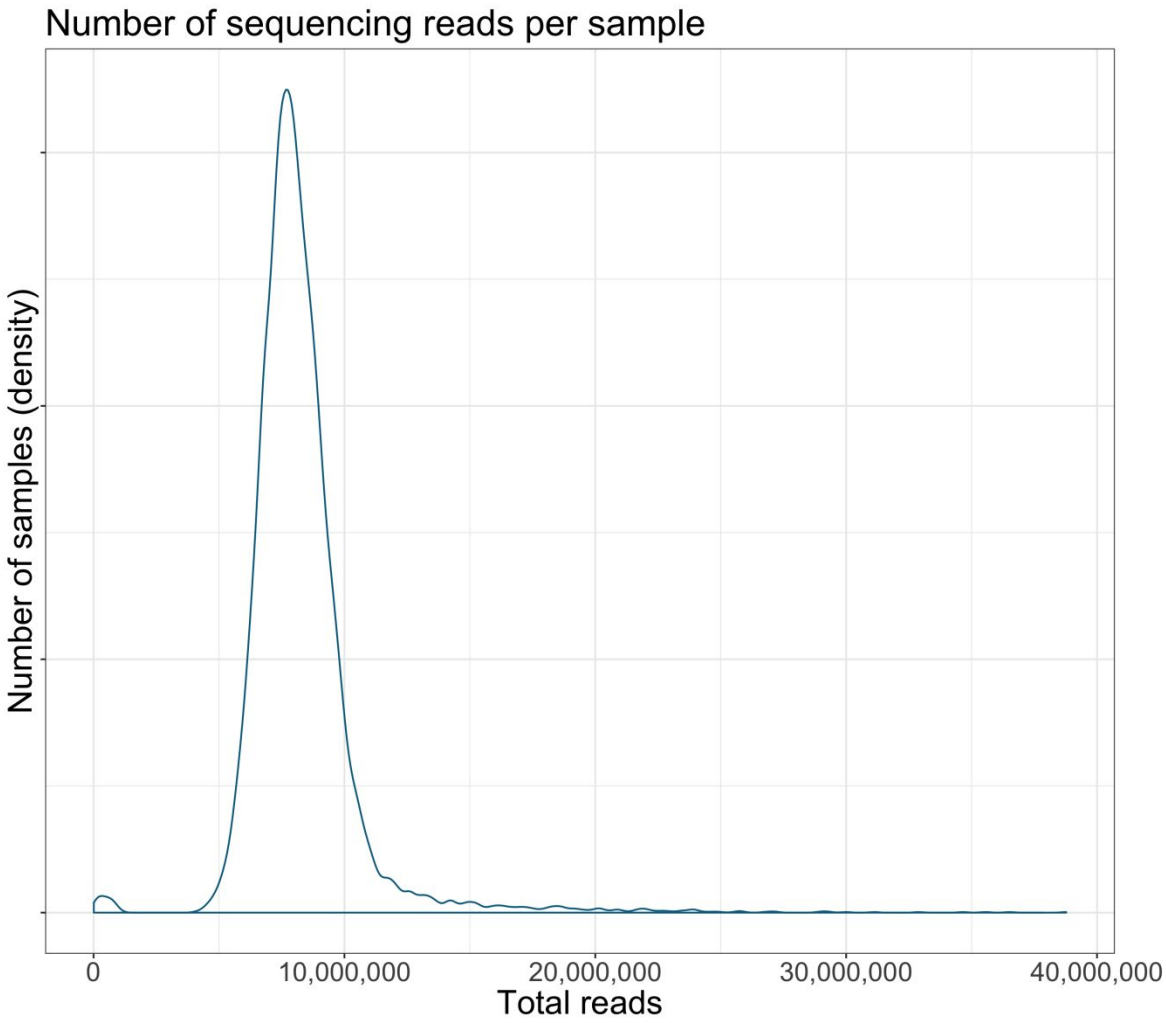


TABLE E1. Cohort Summary

Institution	IPF Cases	Controls
Lung Foundation, Australia	121	0
Columbia University	68	0
COPDGene Study	0	1017
Cork University Hospital	34	0
University of Colorado Denver	432	262
Thomayer Hospital	29	0
Aarhaus University Hospital	61	1
University College Dublin	29	0
Royal Brompton Hospital, London, and Imperial College London	62	0
University of Edinburgh	187	0
Hospital Bichat	57	0
Helmholtz Zentrum München	17	0
Gilead Sciences Inc	305	0
National University Hospital of Iceland (Landspitali)	68	0
GB Morgagni Hospital	95	0
Lung Tissue Research Consortium	157	0
Bellvitge Biomedical Research Institute	94	11
National Jewish Health	100	66
University of Chicago	226	0
University of Nottingham	435	0
Oklahoma Medical Research Foundation	0	1978
University of Pennsylvania	52	7
Lung Institute of Western Australia	17	0
University of Pittsburgh Medical Center	289	182
Sarcoidosis GWAS Controls	0	851
Brown University	10	0
Ege University Hospital	136	57
Gazi University School of Medicine	4	0
University of California San Francisco	153	0
Vanderbilt University	386	10

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

TABLE E2. Targeted Resequencing Regions

Loci	Chr	Start	Stop	# genes	Size (mb)	Gene Targeted/GWAS
3q26	3	169441500	169627900	14	0.19	GWAS
4q22	4	89606000	90098800	9	0.49	GWAS
5p15	5	1213261	1385214	8	0.17	GWAS
6p24	6	7494000	7688000	7	0.19	GWAS
7q22	7	99487014	99764762	32	0.28	GWAS
8p21	8	22000933	22040241	6	0.04	*SFTPC*
10q22	10	81297357	81338464	5	0.04	*SFTPA2*
10q24	10	105597131	105828991	7	0.23	GWAS
11p15	11	1008000	1758985	28	0.75	GWAS
13q34	13	113303000	113791175	15	0.49	GWAS
14q12	14	24685358	24730482	12	0.05	*TINF2*
15q15	15	40585957	40813514	25	0.23	GWAS
16p13	16	2307628	2408996	12	0.10	*ABCA3*
16p13	16	14511306	14742381	7	0.23	*PARN*
19p13	19	4635100	4766391	10	0.13	GWAS
20q13	20	62270912	62345855	9	0.08	*RTEL1*

TABLE E3. Top Common Signals by Familial and Sporadic IPF Status

	Familial Cases (478) vs All Controls (4442)		Sporadic Cases (3146) vs All Controls (4442)	
SNP	**OR Aa vs AA (95% CI)**	**OR aa vs AA (95% CI)**	**OR Aa vs AA (95% CI)**	**OR aa vs AA (95% CI)**
rs2293607	1.41 (1.15-1.72)	1.64 (1.14- 2.37)	1.27 (1.15-1.41)	1.82 (1.50- 2.20)
rs2609260	1.20 (0.97-1.47)	1.37 (0.84- 2.22)	1.38 (1.25-1.54)	2.09 (1.65- 2.64)
rs4449583	0.81 (0.67-0.99)	0.44 (0.29- 0.65)	0.66 (0.59-0.73)	0.47 (0.39- 0.56)
rs2076295	1.34 (1.04-1.71)	2.46 (1.89- 3.20)	1.27 (1.13-1.43)	2.03 (1.78- 2.33)
rs6963345	1.44 (1.15-1.80)	2.07 (1.58- 2.73)	1.35 (1.21-1.50)	1.69 (1.47- 1.96)
rs2488000	0.69 (0.53-0.90)[c]	-	0.70 (0.62-0.80)[c]	-
rs35705950	6.13 (4.98-7.55)	20.01 (12.53-31.94)	5.31 (4.76-5.92)	19.01 (13.49-26.78)
rs1278769	0.66 (0.54-0.82)	0.65 (0.40- 1.04)	0.79 (0.71-0.88)	0.71 (0.57- 0.90)
rs35700143	0.75 (0.61-0.93)	0.53 (0.40- 0.71)	0.76 (0.68-0.85)	0.65 (0.56- 0.74)
rs12610495	1.05 (0.86-1.29)	1.30 (0.94- 1.80)	1.25 (1.13-1.38)	1.64 (1.39- 1.93)

OR, odds ratio. The minor allele is defined as the minor allele in the corresponding case group and full control group.
[a]Based on NCBI Build 37.
[b]Adjusted for sex.
[c]OR of a vs A resulting from dominant test.

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

TABLE E4. Significant Rare Variant Gene-Based Results Prior to Adjustment for Common Variant

Ensembl Gene ID[a]	Gene Name	Locus	start	# variants	p-value[b]
ENSG00000206777	*RNU6-637P*	3q26	169437974	394	0.04
ENSG00000138640	*FAM13A*	4q22	89647105	13121	0.003
ENSG00000164362	*TERT*	5p15	1253261	3826	2.0×10^{-5}
ENSG00000263670	*MIR4457*	5p15	1309424	1626	0.004
ENSG00000049656	*CLPTM1L*	5p15	1317858	3041	0.009
ENSG00000207547	*MIR25*	7q22	99691182	1605	0.04
ENSG00000207757	*MIR93*	7q22	99691390	1606	0.04
ENSG00000208036	*MIR106B*	7q22	99691615	1615	0.05
ENSG00000221838	*AP4M1*	7q22	99699171	1812	0.05
ENSG00000183020	*AP2A2*	11p15	924893	1439	0.006
ENSG00000184956	*MUC6*	11p15	1012820	2696	0.03
ENSG00000254872	*RP13-870H17.3*	11p15	1049879	2161	0.001
ENSG00000198788	*MUC2*	11p15	1074874	3275	1.9×10^{-15}
ENSG00000215182	*MUC5AC*	11p15	1151579	2796	0.005
ENSG00000078902	*TOLLIP*	11p15	1295600	3312	0.02
ENSG00000267950	*AC136297.1*	11p15	1295808	1913	0.01
ENSG00000255153	*TOLLIP-AS1*	11p15	1330998	1702	0.008
ENSG00000174672	*BRSK2*	11p15	1411128	5834	0.006
ENSG00000182208	*MOB2*	11p15	1490686	3738	2.0×10^{-5}
ENSG00000184545	*DUSP8*	11p15	1575273	2304	1.7×10^{-4}
ENSG00000233930	*KRTAP5-AS1*	11p15	1592582	2696	2.0×10^{-5}
ENSG00000205869	*KRTAP5-1*	11p15	1605571	1518	2.8×10^{-4}
ENSG00000205867	*KRTAP5-2*	11p15	1618408	1502	9.3×10^{-5}
ENSG00000196224	*KRTAP5-3*	11p15	1628794	1571	0.002
ENSG00000185940	*KRTAP5-5*	11p15	1651032	1603	0.004
ENSG00000231487	*AP006285.7*	11p15	1683813	1746	0.005
ENSG00000205865	*FAM99B*	11p15	1704498	1734	2.3×10^{-4}
ENSG00000227306	*AP006285.6*	11p15	1709526	1702	2.3×10^{-4}
ENSG00000205864	*KRTAP5-6*	11p15	1718424	1642	4.3×10^{-4}
ENSG00000128944	*KNSTRN*	15q15	40674921	1745	0.04
ENSG00000243509	*TNFRSF6B*	20q13	62328020	2245	0.03
ENSG00000101246	*ARFRP1*	20q13	62329995	2129	0.02

ENSG00000197114	*ZGPAT*	20q13	62338816	1660	0.04
ENSG00000273154	*RP4-583P15.15*	20q13	62340215	1577	0.04

[a]Ensembl version 75
[b]Adjusted for sex with Bonferroni correction for the number of gene-based tests (n = 206)

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

TABLE E5. Significant Rare Variant Window-Based Results Prior to Adjustment for Common Variant

Nearest Gene[a]	Locus	Start (bp)	Stop (bp)	# windows in region	p-value[b]
FAM13A	4q22	89710040	89711851	615	8.4×10^{-6}
FAM13A	4q22	89711042	89712372	615	8.9×10^{-5}
TERT	5p15	1276192	1277167	345	2.5×10^{-5}
TERT	5p15	1276697	1277530	345	0.004
TERT	5p15	1286032	1286647	345	0.018
TERT	5p15	1294397	1295255	345	1.7×10^{-14}
TERT	5p15	1295605	1297059	345	0.007
TERT	5p15	1297716	1298887	345	0.02
MIR4457	5p15	1306899	1308849	345	0.001
MIR4457	5p15	1307309	1309202	345	0.001
CLPTM1L	5p15	1316145	1317074	345	1.5×10^{-5}
CLPTM1L	5p15	1316685	1317699	345	0.03
CLPTM1L	5p15	1320524	1321398	345	0.04
CLPTM1L	5p15	1320915	1322039	345	0.04
CLPTM1L	5p15	1324810	1325822	345	0.007
CLPTM1L	5p15	1336672	1337695	345	0.007
CLPTM1L	5p15	1337144	1337976	345	0.003
CLPTM1L	5p15	1340386	1341392	345	0.02
CLPTM1L	5p15	1351873	1353039	345	0.03
RP3-336K20__B.2	6p24	7668129	7669865	269	0.01
SH3PXD2A	10q24	105603576	105605359	321	0.005
MUC6	11p15	1031017	1031909	1294	2.3×10^{-9}
MUC6	11p15	1031561	1032362	1294	2.3×10^{-9}
RP13-870H17.3	11p15	1059764	1060612	1294	8.8×10^{-4}
RP13-870H17.3	11p15	1060346	1061077	1294	4.6×10^{-4}
RP13-870H17.3	11p15	1060619	1061499	1294	0.001
MUC2	11p15	1092369	1093069	1294	0.008
MUC2	11p15	1098701	1099918	1294	4.2×10^{-8}
MUC2	11p15	1099269	1100753	1294	6.0×10^{-8}
MUC2	11p15	1105360	1106823	1294	0.05
MUC2	11p15	1106830	1108135	1294	4.3×10^{-4}
MUC2	11p15	1107538	1108754	1294	5.3×10^{-4}

MUC2	11p15	1112699	1113428	1294	0.03
MUC2	11p15	1114407	1115269	1294	0.02
MUC2	11p15	1114934	1115763	1294	1.5×10^{-5}
MUC2	11p15	1116155	1117403	1294	3.7×10^{-13}
MUC2	11p15	1116708	1118039	1294	2.8×10^{-7}
MUC2	11p15	1121320	1122366	1294	4.9×10^{-16}
MUC2	11p15	1121909	1122699	1294	9.9×10^{-17}
MUC5AC	11p15	1134851	1136245	1294	0.01
MUC5AC	11p15	1135640	1137157	1294	1.1×10^{-5}
MUC5AC	11p15	1145210	1146300	1294	2.9×10^{-7}
MUC5AC	11p15	1145797	1146779	1294	3.3×10^{-7}
MUC5AC	11p15	1153966	1155014	1294	0.005
MUC5AC	11p15	1217999	1218829	1294	0.02
MUC5AC	11p15	1218460	1219160	1294	1.4×10^{-5}
MUC5AC	11p15	1229607	1230494	1294	1.0×10^{-8}
MUC5B	11p15	1243337	1244304	1294	0.04
MUC5B	11p15	1247637	1248452	1294	0.01
MUC5B	11p15	1250910	1251800	1294	0.007
MUC5B	11p15	1251345	1252116	1294	0.03
MUC5B ,RP11-532E4.2	11p15	1264595	1265214	1294	0.03
TOLLIP	11p15	1288864	1290709	1294	1.7×10^{-4}
TOLLIP	11p15	1289704	1291367	1294	0.003
TOLLIP	11p15	1297176	1298070	1294	0.04
TOLLIP	11p15	1307347	1308466	1294	5.8×10^{-26}
TOLLIP	11p15	1307887	1309361	1294	2.6×10^{-27}
TOLLIP	11p15	1326506	1327169	1294	0.005
TOLLIP	11p15	1326880	1327728	1294	1.9×10^{-5}
TOLLIP-AS1	11p15	1332783	1334043	1294	0.01
TOLLIP-AS1	11p15	1333502	1335934	1294	0.02
TOLLIP-AS1	11p15	1339172	1340355	1294	0.03
TOLLIP-AS1	11p15	1366062	1367234	1294	0.03
TOLLIP-AS1	11p15	1366687	1367843	1294	0.03
BRSK2	11p15	1414275	1416131	1294	0.01
DUSP8	11p15	1566693	1568076	1294	0.005
DUSP8	11p15	1567430	1568684	1294	0.005

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

KRTAP5-AS1	11p15	1597159	1599085	1294	4.6×10^{-4}
KRTAP5-AS1	11p15	1597628	1600004	1294	0.04
KRTAP5-AS1	11p15	1614984	1616627	1294	0.05
KRTAP5-AS1	11p15	1615811	1617311	1294	0.02
KRTAP5-5	11p15	1663772	1664515	1294	4.7×10^{-12}
KRTAP5-5	11p15	1664106	1664979	1294	6.0×10^{-13}
AP006285.7	11p15	1684447	1685698	1294	0.05
AP006285.6	11p15	1709360	1709956	1294	5.7×10^{-14}
AP006285.6	11p15	1709653	1710298	1294	5.7×10^{-5}
MCF2L	13q34	113724860	113726145	885	0.02
RNPS1, AC009065.1	16p13	2311178	2312190	165	0.02
RTEL1, RTEL1-TNFRSF6B	20q13	62324166	62324601	151	0.02

[a]Nearest gene to the window positions using Ensembl 75
[b]Adjusted for sex with Bonferroni correction for the number of windows in the given region

TABLE E6. Significant Rare Variant Functional Results

Ensembl Gene ID[a]	Gene Name	Locus	# total variants	# high impact	# moderate impact	p-value unadjusted[b]	p-value adjusted[c]
ENSG00000164362	TERT	5p15	172	13	159	8.2×10^{-15}	2.07×10^{-13}
ENSG00000196224	*KRTAP5-3*	11p15	20	1	19	0.014	1
ENSG00000215182	MUC5AC	11p15	68	2	66	0.020	1
ENSG00000258366	*RTEL1*	20q13	244	34	210	0.00023	-
ENSG00000026036	*RTEL1-TNFRSF6B*	20q13	272	42	230	0.00068	-

[a]Ensembl version 75

[b]Adjusted for sex with Bonferroni correction for the number of total tests (n = 100)

[c]Adjusted for sex and the top common variant in the region (rs4449583 for locus 5p15 and rs35705950 for locus 11p15) with Bonferroni correction for the number of total unconditional tests (n = 100)

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

TABLE E7. Significant Rare Variant Gene-Based Results After Adjusting for Top Common Variant (if Present)

Gene Name	Ensembl Gene ID[a]	Locus	# variants	p-value[b]	Adjusted Common Variant
TERT	ENSG00000164362	5p15	3826	0.01	rs4449583
ARFRP1	ENSG00000101246	20q13	2129	0.02	-
TNFRSF6B	ENSG00000243509	20q13	2245	0.03	-
ZGPAT	ENSG00000197114	20q13	1660	0.04	-
AL121845.3	ENSG00000273154	20q13	1577	0.04	-

[a]Ensembl version 75
[b]Adjusted for sex and top common variant (if present) with Bonferroni correction for the number of gene-based tests (n = 206)

TABLE E8. Functional Variants Located within Signficant Rare Window-Based Tests (Corresponding to Results in Table 3)

Variant	Ref/Alt Allele	Case MAC/ Control MAC	Impact	Significant Window (chr:start)
chr4:89711758	C/T	0/1	moderate	chr4:89710040;chr4:89711042
rs121918661	C/T	2/1	moderate	chr5:1294397
chr5:1294409	C/G	1/0	moderate	chr5:1294397
chr5:1294475	A/G	1/0	moderate	chr5:1294397
chr5:1294493	C/T	3/1	moderate	chr5:1294397
chr5:1294553	C/T	1/0	moderate	chr5:1294397
chr5:1294555	A/T	1/0	moderate	chr5:1294397
chr5:1294559	C/A	1/0	moderate	chr5:1294397
chr5:1294570	A/G	1/0	moderate	chr5:1294397
rs199422291	C/T	1/0	moderate	chr5:1294397
chr5:1294585	A/C	2/0	moderate	chr5:1294397
chr5:1294612	G/A	2/0	moderate	chr5:1294397
chr5:1294624	G/T	1/0	moderate	chr5:1294397
chr5:1294654	G/C	1/0	moderate	chr5:1294397
chr5:1294655	T/C	1/0	moderate	chr5:1294397
chr5:1294678	C/T	0/1	moderate	chr5:1294397
chr5:1294693	A/C	1/0	moderate	chr5:1294397
chr5:1294694	G/C	1/0	moderate	chr5:1294397
chr5:1294698	G/T	1/0	moderate	chr5:1294397
chr5:1294777	G/A	1/0	moderate	chr5:1294397
chr5:1294890	C/A	0/1	moderate	chr5:1294397;chr5:1294824
chr5:1294933	C/G	1/0	moderate	chr5:1294397;chr5:1294824
chr5:1294947	TG/T	1/0	high	chr5:1294397;chr5:1294824
chr5:1294981	C/T	1/0	moderate	chr5:1294397;chr5:1294824
chr5:1295005	G/A	1/0	high	chr5:1294397;chr5:1294824
chr5:1295022	A/C	2/0	moderate	chr5:1294397;chr5:1294824
chr5:1295023	C/G	1/0	moderate	chr5:1294397;chr5:1294824
chr5:1295056	G/A	1/0	moderate	chr5:1294397;chr5:1294824
chr5:1324915	C/T	1/0	moderate	chr5:1324810
rs200505378	G/A	0/1	moderate	chr20:62324166
rs141717966	C/T	0/4	moderate	chr20:62324166
chr20:62324197	C/T	0/2	moderate	chr20:62324166
chr20:62324238	C/A	0/1	moderate	chr20:62324166
chr20:62324266	G/C	6/1	moderate	chr20:62324166
chr20:62324281	G/C	2/0	moderate	chr20:62324166

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

chr20:62324323	G/A	1/0	moderate	chr20:62324166
rs142969505	C/G	1/0	moderate	chr20:62324166
chr20:62324336	A/G	0/1	moderate	chr20:62324166
chr20:62324339	A/G	0/1	moderate	chr20:62324166
chr20:62324514	G/A	4/0	moderate	chr20:62324166
chr20:62324521	C/A	1/0	moderate	chr20:62324166
chr20:62324531	C/T	0/1	high	chr20:62324166
rs115464632	G/C	2/2	moderate	chr20:62324166
chr20:62324549	A/G	0/1	moderate	chr20:62324166
chr20:62324564	C/T	11/3	high	chr20:62324166
chr20:62324577	A/G	1/0	moderate	chr20:62324166
rs144034326	C/T	6/0	moderate	chr20:62324166
chr20:62324600	C/T	7/0	high	chr20:62324166
rs146221660	G/A	1/1	moderate	chr20:62324166

SUPPLEMENTAL FIGURE LEGENDS

Figure E1. Rare variant signals at (A) 11p15, (B) 13q34 and (C) 16p13 loci. The -$\log_{10}$ adjusted P values (y axis) from SKAT-O of the gene (green) and window (blue) sets are shown according to their chromosomal positions (x axis). The gene sets are labeled with their gene name where the dark green region represents the gene body and the lighter region represents the 20kb region upstream/downstream region. Rare variants located within the gene ± 20kb were included in the testing for that gene. A total of 9 genes were tested at the 11p15 locus, 15 genes tested at the 13q34 locus and 12 genes tested at the 16p13 locus. For clarity, only a few gene names for each locus are shown. In regions which included a significant common variant, the position is noted with a vertical black line.

Figure E2. Common variant results (top) with GTEx v7 Lung eQTL results (bottom) for the 6p24 locus. The estimated recombination rates from 1000 Genomes (NCBI Build 37) are shown as blue lines, and the genomic locations of genes within the regions of interest are shown as arrows. SNP color in GTEx plot represents LD with the most highly associated SNP in our study (rs2076295).

Figure E3. Common variant results (top) with GTEx v7 Lung eQTL results (bottom) for the 15q15 locus. The estimated recombination rates from 1000 Genomes (NCBI Build 37) are shown as blue lines, and the genomic locations of genes within the regions of interest are shown as arrows. SNP color in GTEx plot represents LD with the most highly associated SNP in our study (rs35700143).

AJRCCM Articles in Press. Published on 29-April-2019 as 10.1164/rccm.201810-1891OC

Figure E4. Common variant results (top) with GTEx v7 Lung eQTL results (bottom) for the 19p13locus. The estimated recombination rates from 1000 Genomes (NCBI Build 37) are shown as blue lines, and the genomic locations of genes within the regions of interest are shown as arrows. SNP color in GTEx plot represents LD with the most highly associated SNP in our study (rs12610495).

Figure E5. QC steps are outlined that resulted in a final analysis cohort of 8,066 subjects, including 3,624 cases of IPF (3,146 sporadic cases and 478 cases from families with ≥2 cases of idiopathic interstitial pneumonia) and 4,442 unaffected controls.

Figure E6: Power to detect additive, dominant and recessive odds ratios of 1.5, using a likelihood ratio test, assuming alpha = 5×10^{-6} and a total sample size of 8,066 subjects (3,624 cases and 4,442 controls). Test models include 3 different logistic regression models, each employing different coding of the genetic effect. We consider genotypic or 2 degree of freedom coding, additive, and dominant coding of the genetic effect.

Figure E7: Detectable additive, dominant and recessive odds ratios, using a likelihood ratio test, assuming 80% power, alpha = 5×10^{-6} and a total sample size of 8,066 subjects (3,624 cases and 4,442 controls). Test models include 3 different logistic regression models, each employing different coding of the genetic effect. We consider genotypic or 2 degree of freedom coding, additive, and dominant coding of the genetic effect.