

## ORIGINAL ARTICLE

## Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis

✉ Richard J. Allen<sup>1</sup>, Beatriz Guillen-Guio<sup>2</sup>, Justin M. Oldham<sup>3</sup>, Shwu-Fan Ma<sup>4</sup>, Amy Dressen<sup>5</sup>, Megan L. Paynton<sup>1</sup>, Luke M. Kraven<sup>1</sup>, Ma'en Obeidat<sup>6</sup>, Xuan Li<sup>6</sup>, Michael Ng<sup>7</sup>, Rebecca Braybrooke<sup>8,9</sup>, Maria Molina-Molina<sup>10,11,12</sup>, Brian D. Hobbs<sup>13,14</sup>, Rachel K. Putman<sup>14</sup>, Phuwanat Sakornsakolpat<sup>13,15</sup>, Helen L. Booth<sup>16</sup>, William A. Fahy<sup>17</sup>, Simon P. Hart<sup>18</sup>, Mike R. Hill<sup>19</sup>, Nik Hirani<sup>20</sup>, Richard B. Hubbard<sup>8,9</sup>, Robin J. McAnulty<sup>21</sup>, Ann B. Millar<sup>22</sup>, Vidyia Navaratnam<sup>8,9</sup>, Eunice Oballa<sup>17</sup>, Helen Parfrey<sup>23</sup>, Gauri Saini<sup>24</sup>, Moira K. B. Whyte<sup>20</sup>, Yingze Zhang<sup>25,26</sup>, Naftali Kaminski<sup>27</sup>, Ayodeji Adegunsoye<sup>28</sup>, Mary E. Strek<sup>28</sup>, Margaret Neighbors<sup>5</sup>, Xuting R. Sheng<sup>5</sup>, Gunnar Gudmundsson<sup>29,30</sup>, Vilmundur Gudnason<sup>30,31</sup>, Hiroto Hatabu<sup>32,33</sup>, David J. Lederer<sup>34,35</sup>, Ani Manichaikul<sup>36,37</sup>, John D. Newell, Jr.<sup>38,39</sup>, George T. O'Connor<sup>40,41</sup>, Victor E. Ortega<sup>42</sup>, Hanfei Xu<sup>43</sup>, Tasha E. Fingerlin<sup>44,45</sup>, Yohan Bossé<sup>46</sup>, Ke Hao<sup>47,48</sup>, Philippe Joubert<sup>46</sup>, David C. Nickle<sup>49</sup>, Don D. Sin<sup>6,50</sup>, Wim Timens<sup>51,52</sup>, Dominic Furniss<sup>7</sup>, Andrew P. Morris<sup>53,54,55</sup>, Krina T. Zondervan<sup>53,56</sup>, Ian P. Hall<sup>9,57</sup>, Ian Sayers<sup>9,57</sup>, Martin D. Tobin<sup>1,58</sup>, Toby M. Maher<sup>59,60\*</sup>, Michael H. Cho<sup>13,14</sup>, Gary M. Hunninghake<sup>14,33</sup>, David A. Schwartz<sup>44,61,62</sup>, Brian L. Yaspan<sup>5</sup>, Philip L. Molyneaux<sup>59,60</sup>, Carlos Flores<sup>2,12,63,64</sup>, Imre Noth<sup>4</sup>, R. Gisli Jenkins<sup>9,57‡</sup>, and Louise V. Wain<sup>1,58‡</sup>

ORCID IDs: 0000-0002-8450-3056 (R.J.A.); 0000-0002-3657-8205 (M.L.P.); 0000-0001-9564-0745 (B.D.H.); 0000-0002-4308-0974 (P.S.); 0000-0002-2914-8012 (V.N.); 0000-0002-7786-737X (G.S.); 0000-0001-5258-0228 (D.J.L.); 0000-0002-3067-3711 (Y.B.); 0000-0002-4146-6363 (W.T.); 0000-0001-9933-3216 (I.P.H.); 0000-0001-5352-069X (C.F.); 0000-0003-4951-1867 (L.V.W.).

## Abstract

**Rationale:** Idiopathic pulmonary fibrosis (IPF) is a complex lung disease characterized by scarring of the lung that is believed to result from an atypical response to injury of the epithelium. Genome-wide association studies have reported signals of association implicating multiple pathways including host defense, telomere maintenance, signaling, and cell–cell adhesion.

**Objectives:** To improve our understanding of factors that increase IPF susceptibility by identifying previously unreported genetic associations.

**Methods:** We conducted genome-wide analyses across three independent studies and meta-analyzed these results to generate the largest genome-wide association study of IPF to date (2,668 IPF cases and 8,591 controls). We performed replication in two independent studies (1,456 IPF cases and 11,874 controls) and functional analyses (including statistical fine-mapping, investigations into gene expression, and testing for enrichment of IPF susceptibility signals in

regulatory regions) to determine putatively causal genes. Polygenic risk scores were used to assess the collective effect of variants not reported as associated with IPF.

**Measurements and Main Results:** We identified and replicated three new genome-wide significant ( $P < 5 \times 10^{-8}$ ) signals of association with IPF susceptibility (associated with altered gene expression of *KIF15*, *MAD1L1*, and *DEPTOR*) and confirmed associations at 11 previously reported loci. Polygenic risk score analyses showed that the combined effect of many thousands of as yet unreported IPF susceptibility variants contribute to IPF susceptibility.

**Conclusions:** The observation that decreased *DEPTOR* expression associates with increased susceptibility to IPF supports recent studies demonstrating the importance of mTOR signaling in lung fibrosis. New signals of association implicating *KIF15* and *MAD1L1* suggest a possible role of mitotic spindle-assembly genes in IPF susceptibility.

**Keywords:** genetics; epidemiology; *KIF15*; *MAD1L1*; *DEPTOR*

(Received in original form May 17, 2019; accepted in final form November 7, 2019)

✉ This article is open access and distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

\*T.M.M. is Associate Editor of *AJRCCM*. His participation complies with American Thoracic Society requirements for recusal from review and decisions for authored works.

†These authors contributed equally to this work.

Correspondence and requests for reprints should be addressed to Louise V. Wain, Ph.D., Genetic Epidemiology Group, Department of Health Sciences, George Davies Centre, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: lww1@leicester.ac.uk.

This article has a related editorial.

This article has an online supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org).

Am J Respir Crit Care Med Vol 201, Iss 5, pp 564–574, Mar 1, 2020

Copyright © 2020 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201905-1017OC on November 11, 2019

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

<sup>1</sup>Department of Health Sciences, University of Leicester, Leicester, United Kingdom; <sup>2</sup>Unidad de Investigacion, Hospital Universitario Ntra. Sra. de Candelaria and <sup>34</sup>Instituto de Tecnologías Biomédicas, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; <sup>3</sup>Department of Internal Medicine, University of California Davis, Davis, California; <sup>4</sup>Division of Pulmonary and Critical Care Medicine, <sup>36</sup>Center for Public Health Genomics, and <sup>37</sup>Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia; <sup>5</sup>Genentech, South San Francisco, California; <sup>6</sup>The University of British Columbia Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada; <sup>7</sup>Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, <sup>19</sup>Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, <sup>53</sup>Wellcome Centre for Human Genetics, and <sup>56</sup>Oxford Endometriosis Care and Research Centre, Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, United Kingdom; <sup>8</sup>Division of Epidemiology and Public Health and <sup>57</sup>Division of Respiratory Medicine, University of Nottingham, Nottingham, United Kingdom; <sup>9</sup>National Institute for Health Research, Nottingham Biomedical Research Centre and <sup>24</sup>Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom; <sup>10</sup>Servei de Pneumología, Laboratori de Pneumología Experimental, Instituto de Investigación Biomédica de Bellvitge (IDIBELL), Barcelona, Spain; <sup>11</sup>Campus de Bellvitge, Universitat de Barcelona, Barcelona, Spain; <sup>12</sup>Centro de Investigación Biomédica en Red de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; <sup>13</sup>Channing Division of Network Medicine, <sup>14</sup>Division of Pulmonary and Critical Care Medicine, <sup>32</sup>Department of Radiology, and <sup>33</sup>Center for Pulmonary Functional Imaging, Brigham and Women's Hospital, Boston, Massachusetts; <sup>15</sup>Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; <sup>16</sup>Department of Thoracic Medicine, University College London Hospitals NHS Foundation Trust, London, United Kingdom; <sup>17</sup>Discovery Medicine, GlaxoSmithKline, Stevenage, United Kingdom; <sup>18</sup>Respiratory Research Group, Hull York Medical School, Castle Hill Hospital, Cottingham, United Kingdom; <sup>20</sup>Medical Research Council Centre for Inflammation Research, The University of Edinburgh, Edinburgh, United Kingdom; <sup>21</sup>UCL Respiratory Centre for Inflammation and Tissue Repair, University College London, London, United Kingdom; <sup>22</sup>Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, United Kingdom; <sup>23</sup>Cambridge Interstitial Lung Disease Service, Royal Papworth Hospital, Cambridge, United Kingdom; <sup>25</sup>Division of Pulmonary, Allergy and Critical Care Medicine and <sup>26</sup>Simmons Center for Interstitial Lung Diseases, University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>27</sup>Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut; <sup>28</sup>Section of Pulmonary and Critical Care, Department of Medicine, The University of Chicago, Chicago, Illinois; <sup>29</sup>Department of Respiratory Medicine, Landspítal University Hospital, Reykjavík, Iceland; <sup>30</sup>Faculty of Medicine University of Iceland, Reykjavík, Iceland; <sup>31</sup>Icelandic Heart Association, Kopavogur, Iceland; <sup>34</sup>Department of Medicine, College of Physicians and Surgeons and <sup>35</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York; <sup>38</sup>Division of Cardiovascular and Pulmonary Imaging, Department of Radiology, University of Iowa Carver College of Medicine, Iowa City, Iowa; <sup>39</sup>Department of Radiology, University of Washington, Seattle, Washington; <sup>40</sup>Department of Medicine, Pulmonary Center, Boston University, Boston, Massachusetts; <sup>41</sup>NHLBI's Framingham Heart Study, Framingham, Massachusetts; <sup>42</sup>Center for Precision Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina; <sup>43</sup>Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; <sup>44</sup>Center for Genes, Environment and Health, National Jewish Health, Denver, Colorado; <sup>45</sup>Department of Biostatistics and Informatics, <sup>61</sup>Department of Medicine, and <sup>62</sup>Department of Immunology, University of Colorado Denver, Denver, Colorado; <sup>46</sup>Institut universitaire de cardiologie et de pneumologie de Québec, Université Laval, Québec, Québec, Canada; <sup>47</sup>Department of Genetics and Genomic Sciences and <sup>48</sup>Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York; <sup>49</sup>Merck Research Laboratories, Genetics and Pharmacogenomics, Boston, Massachusetts; <sup>50</sup>Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; <sup>51</sup>University Medical Center Groningen, University of Groningen, Department of Pathology and Medical Biology and <sup>52</sup>Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands; <sup>54</sup>Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; <sup>55</sup>Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, United Kingdom; <sup>56</sup>National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom; <sup>59</sup>National Institute for Health Research Respiratory Clinical Research Facility, Royal Brompton Hospital, London, United Kingdom; <sup>60</sup>National Heart and Lung Institute, Imperial College, London, United Kingdom; and <sup>63</sup>Instituto Tecnológico y de Energías Renovables, S.A., Santa Cruz de Tenerife, Spain

R.J.A. is an Action for Pulmonary Fibrosis Research Fellow. L.V.W. holds a GSK/British Lung Foundation Chair in Respiratory Research. R.G.J. is supported by a National Institute for Health Research (NIHR) Research Professorship (NIHR reference RP-2017-08-ST2-014). I.N. is supported by the NHLBI (R01HL130796). B.G.-G. is funded by Agencia Canaria de Investigación, Innovación y Sociedad de la Información (TESIS2015010057) cofunded by European Social Fund. J.M.O. is supported by the NHLBI (K23HL138190). C.F. is supported by the Spanish Ministry of Science, Innovation and Universities (grant RTC-2017-6471-1; Ministerio de Ciencia e Innovación/Agenzia Estatal de Investigación/Fondo Europeo de Desarrollo Regional, Unión Europea) cofinanced by the European Regional Development Funds "A way of making Europe" from the European Union and by agreement OA17/008 with Instituto Tecnológico y de Energías Renovables to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology. The Spain Biobank array genotyping service was performed at CEGEN-PRB3-ISCIII, which is supported by PT17/0019, of the PE I+D+i 2013–2016, funded by Instituto de Salud Carlos III, and cofinanced by the European Regional Development Funds. P.L.M. is an Action for Pulmonary Fibrosis Research Fellow. M.O. is a fellow of the Parker B. Francis Foundation and a Scholar of the Michael Smith Foundation for Health Research. B.D.H. is supported by NIH K08 HL136928, Parker B. Francis Research Opportunity Award. M.H.C. and G.M.H. are supported by NHLBI grants R01HL113264 (M.H.C.), R01HL137927 (M.H.C.), R01HL135142 (M.H.C. and G.M.H.), R01111024 (G.M.H.), and R01130974 (G.M.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding body has no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. T.M.M. is supported by an NIHR Clinician Scientist Fellowship (NIHR Ref: CS-2013-13-017) and a British Lung Foundation Chair in Respiratory Research (C17-3). M.D.T. is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). The research was partially supported by the NIHR Leicester Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health. I.P.H. was partially supported by the NIHR Nottingham Biomedical Research Centre; the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. I.S. is supported by Medical Research Council (G1000861) and Asthma UK (AUK-PG-2013-188). D.F. was supported by an Intermediate Fellowship from the Wellcome Trust (097152/Z/11/Z). This work was partially supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. V.N. is funded by an NIHR Clinical Lectureship. G.G. is supported by project grant 141513-051 from the Icelandic Research Fund and Landspítali Scientific Fund A-2016-023, A-2017-029, and A-2018-025. D.J.L. and A.M. are supported by Multi-Ethnic Study of Atherosclerosis (MESA) and the MESA SNP Health Association Resource (SHARe) project are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California) and the Broad Institute of Harvard and Massachusetts Institute of Technology (Boston, Massachusetts) using the Affymetrix Genome-Wide Human SNP Array 6.0. This work was supported by NIH grants R01 HL131565 (A.M.), R01 HL103676 (D.J.L.), and R01 HL137234 (D.J.L.).

## At a Glance Commentary

### Scientific Knowledge on the Subject

**Subject:** Idiopathic pulmonary fibrosis (IPF) is a devastating disease where the lungs become scarred. It is not known what causes the scarring, but there have been 17 regions of the genome that have been reported as associated with increased susceptibility to IPF from previous genome-wide association studies. These identify host defense (particularly mucus production), cell-cell adhesion, signaling, and telomere maintenance as important processes in the development of lung fibrosis.

### What This Study Adds to the Field:

By combining all previous IPF genome-wide association studies, we have identified three novel regions of the genome identified with IPF risk and confirmed 11 of the 17 previously reported regions. The three novel regions implicate the genes DEPTOR, KIF15, and MAD1L1. These findings support recent research that shows mTOR signaling promotes lung fibrogenesis and also implicate spindle-assembly genes in the development of IPF.

Idiopathic pulmonary fibrosis (IPF) is a devastating lung disease characterized by the buildup of scar tissue. It is believed that damage to the alveolar epithelium is followed by an aberrant wound-healing response leading to the deposition of dense fibrotic tissue, reducing the lungs' flexibility and inhibiting gas transfer (1). Treatment options are limited, and half of individuals diagnosed with IPF die within 3 to 5 years (1, 2). Two drugs (pirfenidone and nintedanib) have been approved for the treatment of IPF, but neither offer a cure, and they only slow disease progression.

IPF is associated with a number of environmental and genetic factors.

Identifying regions of the genome contributing to disease risk improves our understanding of the biological processes underlying IPF and helps in the development of new treatments (3). To date, genome-wide association studies (4–8) (GWAS) have reported 17 common variant (minor allele frequency [MAF] >5%) signals associated with IPF, stressing the importance of host defense, telomere maintenance, cell-cell adhesion, and signaling with respect to disease susceptibility. The sentinel (most strongly associated) variant, rs35705950, in one of these signals that maps to the promoter region of the *MUC5B* gene has a much larger effect on disease susceptibility than other reported risk variants with each copy of the risk allele associated with a fivefold increase in odds of disease (9). Despite this, the variant rs35705950 has a risk allele frequency of only 35% in cases (compared with 11% in the general population) and so does not explain all IPF risk. Rare variants (MAF <1%) in telomere-related and surfactant genes have also been implicated in familial pulmonary fibrosis and sporadic IPF (10, 11).

In this study, we aimed to identify previously unreported genetic associations with IPF to improve our understanding of disease susceptibility and generate new hypotheses about disease pathogenesis. We conducted a large GWAS of IPF susceptibility by utilizing all European cases and controls recruited to any previously reported IPF GWAS (5–8) and meta-analyzing the results. This was followed by replication in individuals not previously included in IPF GWAS and bioinformatic analysis of gene expression data to identify the genes underlying the identified association signals. As specific IPF-associated variants have also been shown to overlap with other related respiratory traits including lung function in the general population, chronic obstructive pulmonary disease (COPD) (with genetic effects in opposite directions between COPD and IPF) (12–14), and interstitial lung abnormalities (ILAs) (which might be a

precursor lesion for IPF) (15), we tested for association of the IPF susceptibility variants with these respiratory phenotypes in independent datasets. Finally, using polygenic risk scores, we tested whether there was still a substantial contribution to IPF risk from genetic variants with as yet unconfirmed associations with IPF susceptibility.

Some of the results of these studies have been previously reported in the form of two abstracts and a preprint (16–18).

## Methods

### Study Cohorts

We analyzed genome-wide data from three previously described independent IPF case-control collections (named here as the Chicago [5], Colorado [6], and UK [8] studies; please refer to the online supplement for summaries of these collections). Two more independent case-control collections (named here as the UUS [United States, United Kingdom, and Spain] and Genentech studies) were included as replication datasets. The new UUS study recruited cases from the United States, United Kingdom, and Spain and selected controls from UK Biobank (19) (full details on the recruitment, genotyping, and quality control of UUS cases and controls can be found in the online supplement). The previously described (20) Genentech study consisted of cases from three IPF clinical trials and controls from four non-IPF clinical trials (see the online supplement). All studies were restricted to unrelated individuals of European ancestry, and we applied stringent quality control measures (full details of the quality control measures of each study can be found in the online supplement and Figure E1 in the online supplement). All studies diagnosed cases using American Thoracic Society and European Respiratory Society guidelines (21–23) and had appropriate institutional review board or ethics approval.

Genotype data for the Colorado, Chicago, UK, and UUS studies were imputed separately using the Haplotype Reference

Author Contributions: R.J.A., J.M.O., C.F., I.N., R.G.J., and L.V.W. designed the study. R.J.A., B.G.-G., S.-F.M., A.D., M.L.P., L.M.K., M.O., X.L., M. Ng, B.D.H., R.K.P., P.S., D.F., A.P.M., K.T.Z., and B.L.Y. analyzed the data. J.M.O., S.-F.M., M.O., R.B., M.M.-M., R.K.P., P.S., H.L.B., W.A.F., S.P.H., M.R.H., N.H., R.B.H., R.J.M., A.B.M., V.N., E.O., H.P., G.S., M.K.B.W., Y.Z., N.K., A.A., M.E.S., M. Neighbors, X.R.S., G.G., V.G., H.H., D.J.L., A.M., J.D.N., G.T.O'C., V.E.O., H.X., T.E.F., Y.B., K.H., P.J., D.C.N., D.D.S., W.T., I.P.H., I.S., M.D.T., T.M.M., M.H.C., G.M.H., D.A.S., B.L.Y., P.L.M., C.F., I.N., R.G.J., and L.V.W. were responsible for recruitment, screening and genotyping of cases and controls for idiopathic pulmonary fibrosis, interstitial lung abnormalities, and gene expression analyses. J.M.O., D.A.S., C.F., I.N., R.G.J., and L.V.W. supervised and coordinated the study. R.J.A., R.G.J., and L.V.W. led the writing of the manuscript. All authors contributed to drafting and providing critical feedback on the manuscript.

Data availability statement: Full summary statistics for the genome-wide meta-analysis can be accessed from <https://github.com/genomicsITER/PFgenetics>.

Consortium r1.1 panel (24) (see the online supplement). For individuals in the Genentech study, genotypes were derived from whole-genome sequencing data. Duplicated individuals between studies were removed (see the online supplement).

### Identification of IPF Susceptibility Signals

In each of the Chicago, Colorado, and UK studies separately, a genome-wide analysis of IPF susceptibility, using SNPTEST (25) v2.5.2, was conducted adjusting for the first 10 principal components to account for fine-scale population structure. Only biallelic autosomal variants that had a minor allele count  $\geq 10$  were in the Hardy-Weinberg Equilibrium ( $P > 1 \times 10^{-6}$ ) and were well-imputed (imputation quality  $R^2 > 0.5$ ) in at least two studies were included. A genome-wide meta-analysis of the association summary statistics was performed across the Chicago, Colorado, and UK studies using R v3.5.1 (discovery stage). Conditional analyses were performed to identify independent association signals in each locus (see the online supplement).

Sentinel variants (defined as the variant in an association signal where no other variants within 1 Mb showed a stronger association) of the novel signals reaching genome-wide significance in the meta-analysis ( $P < 5 \times 10^{-8}$ ), and nominally significant ( $P < 0.05$ ) with consistent direction of effect in each study, were further tested in the replication samples. We considered novel signals to be associated with IPF susceptibility if they reached a Bonferroni-corrected threshold ( $P < 0.05/\text{number of signals followed up}$ ) in a meta-analysis of the UUS and Genentech studies (replication stage; see the online supplement). Previously reported signals with  $P < 5 \times 10^{-8}$  in the discovery meta-analysis were deemed a confirmed association.

### Characterization of Signals and Functional Effects

To further refine our association signals to include only variants with the highest probabilities of being causal, Bayesian fine-mapping was undertaken. This approach takes all variants within the associated locus and, using the GWAS association results, calculates the probability of each variant being the true causal variant (under the assumptions that

there is one causal variant and that the causal variant has been measured). The probabilities are then combined across variants to define the smallest set of variants that is 95% likely to contain the causal variant (i.e., the 95% credible set) for each IPF susceptibility signal (see the online supplement).

To identify which genes might be implicated by the IPF susceptibility signals, we identified whether any variants in the credible sets were genic coding variants and defined as deleterious (using Variant Effect Predictor [VEP] [26]). In addition, we tested to see if any of the credible set variants were associated with gene expression using three expression quantitative trait loci (eQTL) resources (the Lung eQTL study [ $n = 1,111$ ] [27-29], the NESDA-NTR [Netherlands Study of Depression and Anxiety-Netherlands Twin Register] blood eQTL database [ $n = 4,896$ ] [30], and 48 tissues in GTEx [31] [ $n$  between 80 and 491]; see the online supplement). Where IPF susceptibility variants were found to be associated with expression levels of a gene, we tested whether the same variant was likely to be causal both for differences in gene expression and IPF susceptibility. We only report associations with gene expression where the probability of the same variant driving both the IPF susceptibility signal and gene expression signal exceeded 80% (see the online supplement).

To investigate whether the IPF susceptibility variants that were in noncoding regions of the genome might be in regions with regulatory functions (for example, in regions of open chromatin), we investigated the likely functional impact of those variants using DeepSEA (deep learning-based sequence analyzer) (32). Taking all of the IPF susceptibility variants together, we tested for overall enrichment in regulatory regions specific to particular cell and tissue types using FORGE (functional element overlap analysis of the results of GWAS experiments) (33) and GARFIELD (GWAS analysis of regulatory or functional information enrichment with LD correction) (34). Finally, we investigated whether the genes that were near to the IPF susceptibility variants were more likely to be differentially expressed between IPF cases and controls in four lung epithelial cell types, using SNPsea (35). More details are provided in the online supplement.

### Shared Genetic Susceptibility with Other Respiratory Traits

As previous studies have reported shared genetic susceptibility for IPF and other lung traits (12, 13, 15), we investigated whether the new and previously reported IPF susceptibility signals were associated with quantitative lung function measures in a GWAS of 400,102 individuals (36) or with ILAs in a GWAS comparing 1,699 individuals with an ILA and 10,247 controls (37). Lung function measures investigated were FEV<sub>1</sub>, FVC, the ratio FEV<sub>1</sub>/FVC (used in the diagnosis of COPD), and peak expiratory flow. We applied a Bonferroni corrected  $P$  value threshold to define variants also associated with ILAs or lung function.

### Polygenic Risk Scores

The contribution of as yet unreported variants to IPF susceptibility was assessed using polygenic risk scores. For each individual in the UUS study, the weighted score was calculated as the number of risk alleles, multiplied by the effect size of the variant (as a weighting), summed across all variants included in the score. Effect sizes were taken from the discovery GWAS and independent variants selected using a linkage disequilibrium  $r^2 \leq 0.1$ . As we wanted to explore the contribution from as yet unreported variants, we excluded variants within 1 Mb of each IPF susceptibility locus from the risk score calculation (see the online supplement).

The score was tested to identify whether it was associated with IPF susceptibility, adjusting for 10 principal components to account for fine-scale population structure, using PRSice v1.25 (38). We altered the number of variants included in the risk score calculation using a sliding  $P$  threshold ( $P_T$ ) such that the variant had to have a  $P$  value  $< P_T$  in the genome-wide meta-analysis to be included in the score. This allows us to explore whether variants that do not reach statistical significance in GWAS of current size contribute to disease susceptibility. We used the recommended significance threshold of  $P < 0.001$  for determining significantly associated risk scores (38).

## Results

Following quality control, 541 cases and 542 controls from the Chicago study, 1,515 cases

and 4,683 controls from the Colorado study, and 612 cases and 3,366 controls from the UK study were available (Table 1 and Figure E1) to contribute to the discovery stage of the genome-wide susceptibility analysis (Figure 1). For the replication stage of the GWAS, after quality control, there were 792 cases and 10,000 controls available in the UUS study and 664 cases and 1,874 controls available in the Genentech study (see the online supplement).

To identify new signals of association, we meta-analyzed the genome-wide association results for IPF susceptibility for the Chicago, Colorado, and UK discovery studies. This gave a maximum sample size of up to 2,668 cases and 8,591 controls for 10,790,934 well-imputed ( $R^2 > 0.5$ ) variants with minor allele count  $\geq 10$  in each study and which were available in two or more of the studies (Figure E2).

Three novel signals (in 3p21.31 [near *KIF15*, Figure 2A], 7p22.3 [near *MAD1L1*, Figure 2B], and 8q24.12 [near *DEPTOR*, Figure 2C]) showed a genome-wide significant ( $P < 5 \times 10^{-8}$ ) association with IPF susceptibility in the discovery meta-analysis and were also significant after adjusting for multiple testing ( $P < 0.01$ ) in the replication stage comprising 1,467 IPF cases and 11,874 controls (Tables 2 and E1). Two additional loci were genome-wide significant in the genome-wide discovery analysis but did not reach significance in the replication studies. The sentinel variants

of these two signals were a low-frequency intronic variant in *RTEL1* (MAF = 2.1%, replication  $P = 0.012$ ) and a rare intronic variant in *HECTD2* (MAF = 0.3%, replication  $P = 0.155$ ). Conditional analyses did not identify any additional independent association signals at the new or previously reported IPF susceptibility loci (Figure E5).

To identify the likely causal genes for each new signal, we investigated whether any of the variants were also associated with changes in gene expression (Table 3). The sentinel variant (rs78238620) of the novel signal on chromosome 3 was a low-frequency variant (MAF = 5%) in an intron of *KIF15* with the minor allele being associated with increased susceptibility to IPF and decreased expression of *KIF15* in brain tissue and the nearby gene *TMEM42* in thyroid (31) (Figure E7 and Tables E2 and E3i). The IPF risk allele for the novel chromosome 7 signal (rs12699415, MAF = 42%) was associated with decreased expression of *MAD1L1* in heart tissue (31) (Figure E8 and Tables E2 and E3ii). For the signal on chromosome 8, the sentinel variant (rs28513081) was located in an intron of *DEPTOR*, and the IPF risk allele was associated with decreased expression of *DEPTOR* (in colon, lung, and skin [27–29, 31]) and RP11-760H22.2 (in colon and lung [31]). The risk allele was also associated with increased expression of *DEPTOR* (in whole blood [30]), *TAF2* (in colon [31]), RP11-760H22.2 (in adipose

[31]), and KB-1471A8.1 (in adipose and skin [31], Figure E9 and Tables E2 and E3iii). There were no variants predicted to be highly deleterious within the fine-mapped signals for any of the loci.

We confirmed genome-wide significant associations with IPF susceptibility for 11 of the 17 previously reported signals (in or near *TERC*, *TERT*, *DSP*, 7q22.1, *MUC5B*, *ATP11A*, *IVD*, *AKAP13*, *KANSL1*, *FAM13A*, and *DPP9*; Table E1 and Figure E4). The signal at *FAM13A*, while genome-wide significant in the discovery meta-analysis, was not significant in the Chicago study. This was the only signal reaching genome-wide significance in the discovery genome-wide meta-analysis that did not reach at least nominal significance in each study in the discovery analysis. Three further previously reported signals at 11p15.5 (near *MUC5B*) were no longer genome-wide significant after conditioning on the *MUC5B* promoter variant (Table E1), consistent with previous reports (6, 39).

Of the 14 IPF susceptibility signals (i.e., the 11 previously reported signals we confirmed and three novel signals), the only variant predicted to have a potential functional effect on gene regulation through disruption of chromatin structure or transcription factor binding motifs (using DeepSEA) was rs2013701 (in an intron of *FAM13A*), which was associated with a change in DNase I hypersensitivity in

**Table 1.** Demographics of Study Cohorts

	Chicago		Colorado		UK		UUS		Genentech	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
<i>n</i>	541	542	1,515	4,683	612	3,366	792	10,000	664	1,874
Genotyping array/sequencing	Affymetrix 6.0 SNP array		Illumina Human 660W Quad BeadChip		Affymetrix BiLEVE array	Affymetrix UK BiLEVE and UK Biobank arrays	Affymetrix UK Biobank and Spain Biobank arrays	Affymetrix UK BiLEVE and UK Biobank arrays	Affymetrix UK Biobank and Spain Biobank arrays	HiSeq X Ten platform (Illumina)
Imputation panel	HRC		HRC		HRC		HRC		HRC	
Age, yr, mean	68	63*	66	—	70†	65	69	58	68	—
Sex, M, %	71‡	47§	68	49	70.8	70.0	75.2	72.1	73.5	27.1
Ever smokers, %	72	42	—	—	72.9	70.0	68.7††	68.0	67.3	18.1**

Definition of abbreviations: HRC = Haplotype Reference Consortium; UUS = United States, United Kingdom, and Spain.

\*Age only available for 103 Chicago controls.

†Age available for 602 UK cases.

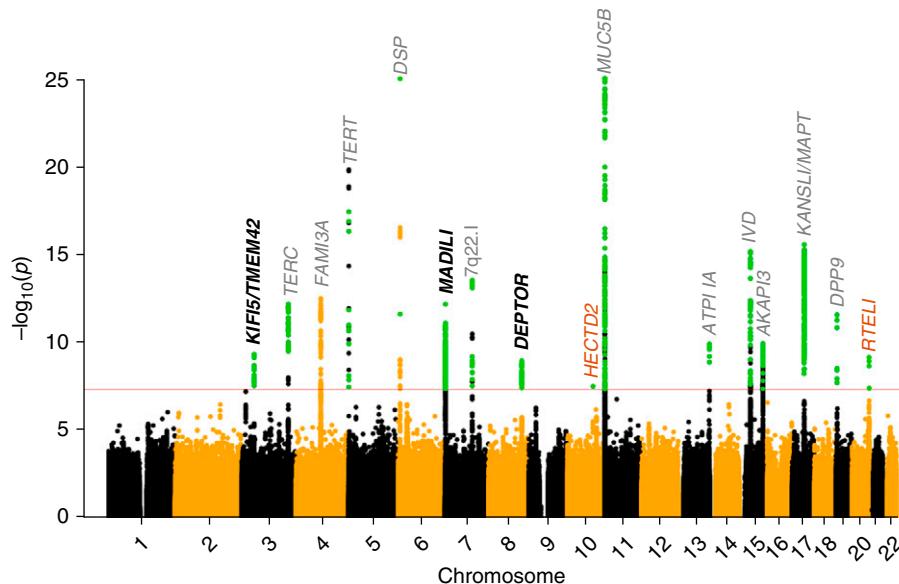
‡Sex only available for 500 Chicago cases.

§Sex only available for 510 Chicago controls.

||Smoking status only recorded for 236 UK cases.

††Smoking status only recorded for 753 idiopathic pulmonary fibrosis cases in UUS.

\*\*Smoking status only recorded for 481 of the Genentech controls.



**Figure 1.** Manhattan plot of discovery analysis results. The x axis shows chromosomal position, and the y axis shows  $-\log(P)$  for each variant in the discovery genome-wide analysis. The red line shows genome-wide significance ( $P < 5 \times 10^{-8}$ ), and variants in green met the criteria for further study in the replication analysis (i.e., reached genome-wide significance in the discovery meta-analysis and had  $P < 0.05$  and consistent direction of effects in each study). Genes in gray are previously reported signals that reach significance in the discovery genome-wide meta-analysis. Genes in black are the novel signals identified in the discovery analysis that reach genome-wide significance when meta-analyzing discovery and replication samples. The signals that did not replicate are shown in red. For ease of visualization the y axis has been truncated at 25.

18 cell types and FOXA1 in the T-47D cell line (a breast cancer cell line derived from a pleural effusion, Table E4). The 14 IPF susceptibility signals were found to be enriched in DNase I hypersensitivity site regions in multiple tissues including fetal lung tissue (Figures E10 and E11). No enrichment in differential expression in airway epithelial cells between IPF cases and healthy controls was observed for the 14 IPF susceptibility signals when using SNPsea (Table E5).

Previous studies have reported an overlap of genetic association loci between lung function and IPF (12). We undertook a lookup of the 14 IPF susceptibility loci in the largest GWAS of lung function in the general population published to date (36). The sentinel variants of 12 of the 14 IPF susceptibility loci were at least nominally associated ( $P < 0.05$ ) with one or more lung function trait in general population studies (Tables 3 and E6). After adjustments for multiple testing ( $P < 5.2 \times 10^{-4}$ ), the previously reported variants at *FAM13A*, *DSP*, and *IVD* were associated with decreased FVC, and variants at *FAM13A*, *DSP*, 7q22.1 (*ZKSCAN1*), and *ATP1A*

were associated with increased  $FEV_1/FVC$ . Similarly, for the three novel susceptibility variants, all showed at least a nominal association with decreased FVC and increased  $FEV_1/FVC$ . We observed a nominally significant association of the *MUC5B* IPF risk allele with decreased FVC and increased  $FEV_1/FVC$ . The IPF risk alleles at *MAPT* were significantly associated with both increased  $FEV_1$  and FVC. To determine how the variants identified for IPF susceptibility are related to differences in lung function between cases and controls, we investigated whether variants known to be associated with lung function show an association in our IPF GWAS. Of the 279 variants reported (36) as associated with lung function (Table E7), 8 showed an association with lung function after corrections for multiple testing (located in or near *MCL1*, *DSP*, *ZKSCAN1*, *OBFC1*, *IVD*, *MAPT*, and two signals in *FAM13A*).

As interstitial lung abnormalities may be a precursor to IPF in a subset of patients, and there have been previous reports of shared genetic etiology between IPF and ILAs (37, 40, 41), we investigated whether our three new signals and the 11 previously

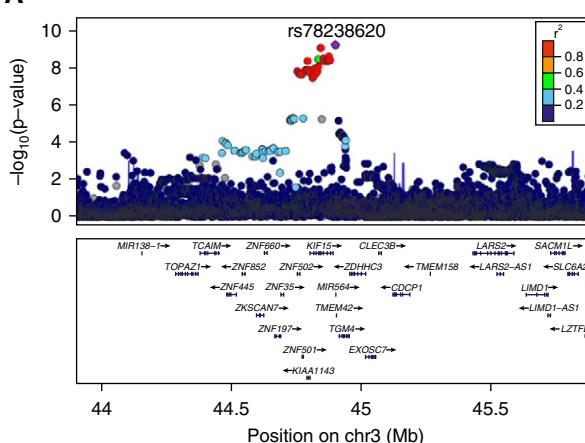
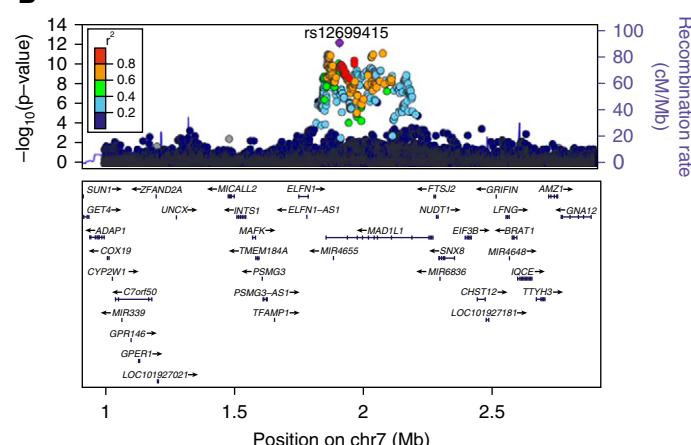
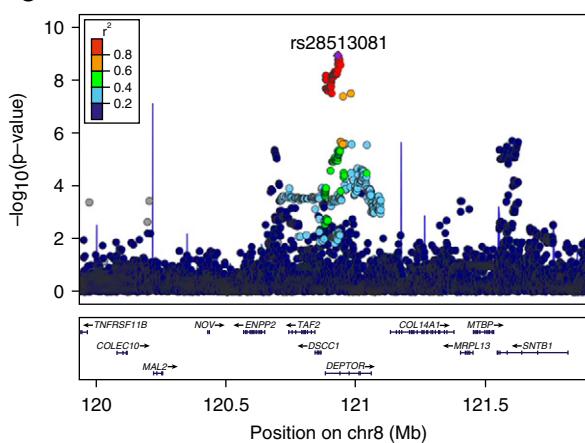
reported signals were associated with ILAs in the largest ILA GWAS reported to date (37). Eight of the IPF susceptibility loci were at least nominally significantly associated with either ILAs or subpleural ILAs with consistent direction of effects (i.e., the allele associated with increased IPF risk was also associated with increased ILA risk). The new *KIF15*, *MAD1L1*, and *DEPTOR* signals were not associated with ILAs (although the rare risk allele at *HECTD2* that did not replicate in our study showed some association with an increased risk of subpleural ILAs [ $P = 0.003$ ] with a large effect size similar to that observed in the IPF discovery meta-analysis).

To quantify the impact of as yet unreported variants on IPF susceptibility, polygenic risk scores were calculated excluding the 14 IPF susceptibility variants (as well as all variants within 1 Mb). The polygenic risk score was significantly associated with increased IPF susceptibility despite exclusion of the known genetic association signals (including *MUC5B*). As the  $P_T$  for inclusion of variants in the score was increased, the risk score became more significant reaching a plateau at around  $P_T = 0.2$  with risk score  $P < 3.08 \times 10^{-23}$  and explaining around 2% of the phenotypic variation (Figure E12), suggesting that there is a modest but statistically significant contribution of additional as yet undetected variants to IPF susceptibility. Further increasing  $P_T$  beyond 0.2 did not improve the predictive accuracy of the risk score.

## Discussion

We undertook the largest GWAS of IPF susceptibility to date and identified three novel signals of association that implicated genes not previously known to be important in IPF.

The strongest evidence for the new signal on chromosome 8 implicates *DEPTOR*, which encodes the dishevelled, Egl-10 and Pleckstrin domain-containing mTOR-interacting protein. *DEPTOR* inhibits mTOR (mammalian target of rapamycin) kinase activity as part of both the mTORC1 and mTORC2 protein complexes. The IPF risk allele at this locus was associated with decreased gene expression of *DEPTOR* in lung tissue (Table E2). TGF $\beta$ -induced *DEPTOR* suppression can stimulate collagen synthesis (42), and the importance of mTORC1 signaling via

**A****B****C**

**Figure 2.** Region plots of three novel idiopathic pulmonary fibrosis susceptibility loci from discovery genome-wide meta-analysis. Each point represents a variant with chromosomal position on the x axis and the  $-\log(P$  value) on the y axis. Variants are colored in by linkage disequilibrium with the sentinel variant. Blue lines show the recombination rate, and gene locations are shown at the bottom of the plot. Region plots are shown for the three replicated novel idiopathic pulmonary fibrosis susceptibility loci, i.e., (A) the susceptibility signal on chromosome 3 near *KIF15*, (B) the susceptibility signal on chromosome 7 near *MAD1L1*, and (C) the susceptibility signal on chromosome 8 near *DEPTOR*.

4E-BP1 for TGF $\beta$ -induced collagen synthesis has recently been demonstrated in fibrogenesis (43). *MAD1L1*, implicated by a new signal on chromosome 7 and eQTL analyses of nonlung tissue, is a mitotic checkpoint gene, mutations in which have been associated with multiple cancers including lung cancer (44, 45). Studies have shown that *MAD1*, a homolog of *MAD1L1*, can inhibit *TERT* activity (or possibly enforce expression of *TERT* when the promoter E-box is mutated) (45, 46). This could suggest that *MAD1L1* may increase IPF susceptibility through reduced telomerase activity. Another spindle-assembly-related gene (47), *KIF15*, was implicated by the new signal on chromosome 3 (along with *TMEM42*).

The genome-wide study also identified two signals that were not replicated after

multiple testing adjustments. *RTEL1*, a gene involved in telomere elongation regulation, has not previously been identified in an IPF GWAS; however, the collective effect of rare variants in *RTEL1* has been reported as associated with IPF susceptibility (48–54). The ubiquitin E3 ligase encoded by *HECTD2* has been shown to have a proinflammatory role in the lung, and other *HECTD2* variants may be protective against acute respiratory distress syndrome (55). However, the lack of replication for these signals in our data suggests that further exploration of their relationship to interstitial lung diseases is warranted.

By combining the largest available GWAS datasets for IPF, we were able to confirm 11 of 17 previously reported signals. Conditional analysis at the 11p15.5 region indicated that previously reported signals at

*MUC2* and *TOLLIP* were not independent of the association with the *MUC5B* promoter variant. Previously reported signals at *EHMT2*, *OBFC1*, and *MDGA2* were only found to be associated in one of the discovery studies and showed no evidence of an association with IPF susceptibility in the other two discovery studies. Only the 11 signals that we confirmed in our data were included in subsequent analyses.

The IPF susceptibility signals at *DSP*, *FAM13A*, 7q22.1 (*ZKSCAN1*), and 17q21.31 (*MAPT*) have also been reported as associated with COPD, although with opposite effects (i.e., the allele associated with increased risk of IPF being associated with decreased risk of COPD). Spirometric diagnosis of COPD was based on a reduced FEV<sub>1</sub>/FVC ratio. In an independent dataset

**Table 2.** Discovery and Replication Association Analysis Results for the Five Signals Reaching Significance in the Discovery Genome-Wide Association Studies that Have Not Previously Been Reported as Associated with Idiopathic Pulmonary Fibrosis

Chr	Pos	rsid	Locus	Major Allele	Minor Allele	MAF (%)	Discovery Meta-Analysis		Replication Meta-Analysis		Meta-Analysis of Discovery and Replication	
							OR [95% CI]	P Value	OR [95% CI]	P Value	OR [95% CI]	P Value
3	44902386	rs78238620	<i>KIF15</i>	T	A	5.3	1.58 [1.37–1.83]	$5.12 \times 10^{-10}$	1.48 [1.24–1.77]	$1.43 \times 10^{-5}$	1.54 [1.38–1.73]	$4.05 \times 10^{-14}$
7	1909479	rs12699415	<i>MAD1L1</i>	G	A	42.0	1.28 [1.19–1.37]	$7.15 \times 10^{-13}$	1.29 [1.18–1.41]	$2.27 \times 10^{-8}$	1.28 [1.21–1.35]	$9.38 \times 10^{-20}$
8	120934126	rs28513081	<i>DEPTOR</i>	A	G	42.8	0.82 [0.76–0.87]	$1.20 \times 10^{-9}$	0.87 [0.80–0.95]	0.002	0.83 [0.79–0.88]	$1.93 \times 10^{-11}$
10	93271016	rs537322302	<i>HECTD2</i>	C	G	0.3	7.82 [3.77–16.2]	$3.43 \times 10^{-8}$	1.75 [0.81–3.78]	0.155	3.85 [2.27–6.54]	$6.25 \times 10^{-7}$
20	62324391	rs41308092	<i>RTEL1</i>	G	A	2.1	2.12 [1.67–2.69]	$7.65 \times 10^{-10}$	1.45 [1.08–1.94]	0.012	1.82 [1.51–2.19]	$2.24 \times 10^{-10}$

Definition of abbreviations: Chr = chromosome; CI = confidence interval; MAF = minor allele frequency; OR = odds ratio; Pos = position; rsid = reference SNP cluster ID. The minor allele is the effect allele, and the MAF is taken from across the studies used in the discovery meta-analysis.

**Table 3.** Gene Expression and Spirometric Results for the Three Novel IPF Susceptibility Loci

Chr	rsid of Sentinel Variant	Annotation	eQTL		FEV <sub>1</sub>		FVC		FEV <sub>1</sub> /FVC	
			Lung Tissue	Nonlung Tissue	$\beta$ [95% CI]	P Value	$\beta$ [95% CI]	P Value	$\beta$ [95% CI]	P Value
3	rs78238620	Intron ( <i>KIF15</i> )	—	↓ <i>KIF15</i> ↓ <i>TMEM42</i>	−0.011 [−0.022 to 0.000]	0.069	−0.022 [−0.033 to 0.011]	$2.92 \times 10^{-4}$	0.017 [0.006 to 0.028]	0.005
7	rs12699415	Intron ( <i>MAD1L1</i> )	—	↓ <i>MAD1L1</i>	−0.007 [−0.012 to −0.002]	0.011	−0.011 [−0.016 to −0.007]	$1.41 \times 10^{-5}$	0.008 [0.003 to 0.012]	0.005
8	rs28513081	Intron ( <i>DEPTOR</i> )	↓ <i>DEPTOR</i> ↓ RP11-760H22.2	↑ <i>DEPTOR</i> ↑ RP11-760H22.2 ↑ KB-1471A8.1 ↑ <i>TAF2</i>	0.001 [−0.004 to 0.006]	0.822	−0.005 [−0.010 to −0.001]	0.045	0.011 [0.006 to 0.016]	$4.22 \times 10^{-5}$

Definition of abbreviations: Chr = chromosome; CI = confidence interval; eQTL = expression quantitative trait loci; IPF = idiopathic pulmonary fibrosis; rsid = reference SNP cluster ID. Annotation of the variant was taken from Variant Effect Predictor (VEP). A list of all variants included in the credible sets with their annotations and eQTL results can be found in Table E3. For colocalization, only genes where there was a greater than 80% probability of colocalization between the IPF risk signal and gene expression of that gene are reported in this table. In the colocalization column, ↑ denotes that the allele that increases IPF risk was associated with increased expression of the gene, ↓ denotes that the IPF risk allele was associated with decreased expression of the gene, and ‡ denotes that the IPF risk allele was associated with increased expression in some tissues and decreased expression in others. Full results from the eQTL and colocalization analyses can be found in Table E2. The spirometric results for the three novel IPF risk loci are taken from Shrine and colleagues (36) using the allele associated with increased IPF risk as the effect allele, with  $\beta$  being the change in z-score units. Results for all IPF risk variants can be found in Table E6.

of 400,102 individuals, eight of the IPF signals were associated with decreased FVC and with a comparatively weaker effect on FEV<sub>1</sub>. This is consistent with the lung function abnormalities associated with IPF, as well as the decreased risk of COPD. Of note, only around 3% of previously reported lung function signals (36) also showed association with IPF susceptibility in our study. This suggests that while some IPF susceptibility variants might represent genes and pathways that are important in general lung health, others are likely to represent more disease-specific processes.

Using polygenic risk scores, we demonstrated that, despite the relatively large proportion of disease susceptibility explained by the known genetic signals of association reported here, IPF is highly polygenic with potentially hundreds (or thousands) of as yet unidentified variants associated with disease susceptibility.

A strength of our study was the large sample size compared with previous GWAS and the availability of an independent replication dataset. A limitation of our study was that the controls used were generally

younger in all studies included, and there were differences in sex and smoking distributions in some of the studies. As age, sex, and smoking status were not available for all individuals in four of our datasets, we were unable to adjust for these variables without substantially reducing our sample size. However, cases and controls in the UUS and UK datasets were matched for age, sex, and smoking. The three novel signals replicated in all of the discovery and replication datasets, providing reassurance that the signals we report are robust despite differences between the datasets. As we had limited information beyond IPF diagnosis status for a large proportion of the individuals included in the studies, we cannot rule out some association with other age-related conditions that are comorbid with IPF. However, other age-related conditions were not excluded from either the cases or controls. For the signals near *KIF15* and *MAD1L1*, there was substantial evidence for an association with gene expression in nonlung tissues but not in either of the two (nonfibrotic) lung tissue eQTL datasets. This could reflect cell

type-specific effects that are missed when studying whole tissue or effects that are disease-dependent. Finally, our study was not designed to identify rare functional variant associations. As both common and rare variants are known to be important in IPF susceptibility (39), this is a limitation of our study.

In summary, we report new biological insights into IPF susceptibility and demonstrate that further studies to identify the genetic determinants of IPF susceptibility are needed. Our new signals of association with IPF susceptibility provide increased support for the importance of mTOR signaling in pulmonary fibrosis as well as the possible implication of mitotic spindle-assembly genes. ■

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

**Acknowledgment:** This research has been conducted using the UK Biobank Resource under application 8389. This research used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester.

## References

1. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med* 2018;378:1811–1823.
2. Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183:431–440.
3. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856–860.
4. Musheroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, et al.; Pirfenidone Clinical Study Group. A genome-wide association study identifies an association of a common variant in *TERT* with susceptibility to idiopathic pulmonary fibrosis. *J Med Genet* 2008;45:654–656.
5. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1:309–317.
6. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45:613–620.
7. Fingerlin TE, Zhang W, Yang IV, Ainsworth HC, Russell PH, Blumhagen RZ, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016;17:74.
8. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017;5:869–880.
9. Zhu QQ, Zhang XL, Zhang SM, Tang SW, Min HY, Yi L, et al. Association between the MUC5B promoter polymorphism rs35705950 and idiopathic pulmonary fibrosis: a meta-analysis and trial sequential analysis in Caucasian and Asian populations. *Medicine (Baltimore)* 2015;94:e1901.
10. Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q, et al. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. *BMJ Open Respir Res* 2014; 1:e000057.
11. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL, et al. An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *Am J Respir Crit Care Med* 2017;196:82–93.
12. Hobbs BD, de Jong K, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al.; COPDGene Investigators; ECLIPSE Investigators; LifeLines Investigators; SPIROMICS Research Group; International COPD Genetics Network Investigators; UK BiLEVE Investigators; International COPD Genetics Consortium. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet* 2017;49:426–432.
13. Sakornnakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, et al.; SpiroMeta Consortium; International COPD Genetics Consortium. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet* 2019;51:494–505.
14. Chilosi M, Poletti V, Rossi A. The pathogenesis of COPD and IPF: distinct horns of the same devil? *Respir Res* 2012;13:3.
15. Putman RK, Gudmundsson G, Araki T, Nishino M, Sigurdsson S, Gudmundsson EF, et al. The MUC5B promoter polymorphism is associated with specific interstitial lung abnormality subtypes. *Eur Respir J* 2017;50:1700537.
16. Allen RJ, Oldham JM, Fingerlin TE, et al. Polygenicity of idiopathic pulmonary fibrosis [abstract]. *Genet Epidemiol* 2018;42:684.
17. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Fingerlin TE, Ng M, et al. Genome-wide meta-analysis of susceptibility to idiopathic pulmonary fibrosis [abstract]. Presented at the American Society of Human Genetics 68th Annual Meeting. October 16–20, 2018, San Diego, CA. Abstract 3313T, p. 493.
18. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis [preprint]. bioRxiv; 2019 [accessed 2019]

May 14]. Available from: <https://www.biorxiv.org/content/10.1101/636761v1>.

19. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.

20. Dressen A, Abbas AR, Cabanski C, Reeder J, Ramalingam TR, Neighbors M, et al. Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. *Lancet Respir Med* 2018;6:603–614.

21. American Thoracic Society; European Respiratory Society. American Thoracic Society/European Respiratory Society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias: this joint statement of the American Thoracic Society (ATS) and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002;165:277–304. [Published erratum appears in *Am J Respir Crit Care Med* 166:426.]

22. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al.; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis. Evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183:788–824.

23. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al.; American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Society. Diagnosis of idiopathic pulmonary fibrosis: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018;198:e44–e68.

24. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al.; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279–1283.

25. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906–913.

26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol* 2016;17:122.

27. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012;8:e1003029.

28. Lamontagne M, Couture C, Postma DS, Timens W, Sin DD, Paré PD, et al. Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs. *PLoS One* 2013;8:e70220.

29. Obeidat M, Miller S, Probert K, Billington CK, Henry AP, Hodge E, et al. GSTCD and INTS12 regulation and expression in the human lung. *PLoS One* 2013;8:e74630.

30. Jansen R, Hottenga JJ, Nivard MG, Abdellaoui A, Laport B, de Geus EJ, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet* 2017;26:1444–1451.

31. Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; eQTL manuscript working group. Genetic effects on gene expression across human tissues. *Nature* 2017;550:204–213.

32. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 2015;12:931–934.

33. Dunham I, Kulesha E, Iotchkova V, Morganella S, Birney E. FORGE: a tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions [preprint]. F1000 Research; 2015 [accessed 2020 Jan 15]. Available from: <https://f1000research.com/articles/4-18>.

34. Iotchkova V, Ritchie GR, Geihs M, et al. GARFIELD: GWAS analysis of regulatory or functional information enrichment with LD correction [preprint]. bioRxiv; 2016 [accessed 2019 Feb 11]. Available from: <https://www.biorxiv.org/content/10.1101/085738v1>.

35. Slowikowski K, Hu X, Raychaudhuri S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* 2014;30:2496–2497.

36. Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019;51:481–493.

37. Hobbs BD, Putman RK, Araki T, Nishino M, Gudmundsson G, Gudnason V, et al. Overlap of genetic risk between interstitial lung abnormalities and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;200:1402–1413.

38. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. *Bioinformatics* 2015;31:1466–1468.

39. Moore C, Blumhagen RZ, Yang IV, Walts A, Powers J, Walker T, et al. Resequencing study confirms that host defense and cell senescence gene variants contribute to the risk of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2019;200:199–208.

40. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013;368:2192–2200.

41. Putman RK, Hunninghake GM, Dieffenbach PB, Barragan-Bradford D, Serhan K, Adams U, et al. Interstitial lung abnormalities are associated with acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2017;195:138–141.

42. Das F, Bera A, Ghosh-Choudhury N, Abboud HE, Kasinath BS, Choudhury GG. TGF $\beta$ -induced deopter suppression recruits mTORC1 and not mTORC2 to enhance collagen I ( $\alpha$ 2) gene expression. *PLoS One* 2014;9:e109608.

43. Woodcock HV, Eley JD, Guillotin D, Platé M, Nanthakumar CB, Martufi M, et al. The mTORC1/4E-BP1 axis represents a critical signaling node during fibrogenesis. *Nat Commun* 2019;10:6.

44. Thompson MJ, Rubbi L, Dawson DW, Donahue TR, Pellegrini M. Pancreatic cancer patient survival correlates with DNA methylation of pancreas development genes. *PLoS One* 2015;10:e0128814.

45. Coe BP, Lee EH, Chi B, Girard L, Minna JD, Gazdar AF, et al. Gain of a region on 7p22.3, containing MAD1L1, is the most frequent event in small-cell lung cancer cell lines. *Genes Chromosomes Cancer* 2006;45:11–19.

46. Kang JU, Koo SH, Kwon KC, Park JH, Kim JM. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet Cytogenet* 2008;182:1–11.

47. Tanenbaum ME, Macúrek L, Janssen A, Geers EF, Alvarez-Fernández M, Medema RH. Kif15 cooperates with eg5 to promote bipolar spindle assembly. *Curr Biol* 2009;19:1703–1711.

48. Alder JK, Chen JJ, Lancaster L, Danoff S, Su SC, Cogan JD, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci USA* 2008;105:13051–13056.

49. Stuart BD, Lee JS, Kozlitina J, Noth I, Devine MS, Glazer CS, et al. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. *Lancet Respir Med* 2014;2:557–565.

50. McDonough JE, Martens DS, Tanabe N, Ahangari F, Verleden SE, Maes K, et al. A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. *Respir Res* 2018;19:132.

51. Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet* 2015;47:512–517.

52. Kropski JA, Loyd JE. Telomeres revisited: RTEL1 variants in pulmonary fibrosis. *Eur Respir J* 2015;46:312–314.

53. Kannengiesser C, Borie R, Ménard C, Réocreux M, Nitschke P, Gazal S, et al. Heterozygous RTEL1 mutations are associated with familial pulmonary fibrosis. *Eur Respir J* 2015;46:474–485.

54. Deng Y, Li Z, Liu J, Wang Z, Cao Y, Mou Y, et al. Targeted resequencing reveals genetic risks in patients with sporadic idiopathic pulmonary fibrosis. *Hum Mutat* 2018;39:1238–1245.

55. Coon TA, McKelvey AC, Lear T, Rajbhandari S, Dunn SR, Connelly W, et al. The proinflammatory role of HECTD2 in innate immunity and experimental lung injury. *Sci Transl Med* 2015;7:295ra109.