Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis


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 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 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 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 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).

Measurements and Main Results: We identified and replicated three new genome-wide significant (P < 5 × 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.

Keywords: genetics; epidemiology; KIF15; MAD1L1; DEPTOR
At a Glance Commentary

Scientific Knowledge on the 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.

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


Data availability statement: Full summary statistics for the genome-wide meta-analysis can be accessed from https://github.com/genomicsITER/IPgenetics.
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 ≥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 ($P < 0.05$/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 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 particular 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 FEV1, FVC, the ratio FEV1/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 $r^2 = 0.1$. 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 (r² > 0.5) variants with minor allele count ≥ 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 × 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 RETL1 (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

<table>
<thead>
<tr>
<th></th>
<th>Chicago</th>
<th>Colorado</th>
<th>UK</th>
<th>UUS</th>
<th>Genentech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td>Cases</td>
<td>Controls</td>
<td>Cases</td>
</tr>
<tr>
<td>n Genotyping</td>
<td>541</td>
<td>542</td>
<td>1,515</td>
<td>4,683</td>
<td>612</td>
</tr>
<tr>
<td>array/sequencing</td>
<td>Affymetrix 6.0</td>
<td>SNP array</td>
<td>Illumina Human 660W Quad BeadChip</td>
<td>Affymetrix UK BiLEVE array</td>
<td>Affymetrix UK BiLEVE and UK Biobank arrays</td>
</tr>
<tr>
<td>Imputation panel</td>
<td>HRC</td>
<td>—</td>
<td>HRC</td>
<td>—</td>
<td>HRC</td>
</tr>
<tr>
<td>Age, yr. mean</td>
<td>68</td>
<td>63*</td>
<td>66</td>
<td>—</td>
<td>70*</td>
</tr>
<tr>
<td>1* Age only available for 103 Chicago controls.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex, M, %</td>
<td>71†</td>
<td>47§</td>
<td>68</td>
<td>49</td>
<td>72*</td>
</tr>
<tr>
<td>Ever smokers, %</td>
<td>72</td>
<td>42</td>
<td>—</td>
<td>—</td>
<td>72.0*</td>
</tr>
</tbody>
</table>

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

†Age only available for 481 of the Genentech controls.
‡Age available for 602 UK cases.
§Sex only available for 500 Chicago cases.
¶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.
signals were not associated with a risk score for inclusion of variants in the score.

DEPTOR

3

ATP11A

IPF risk allele with decreased FVC were significant 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 to 25.

Figure 1. Manhattan plot of discovery analysis results. The x axis shows chromosomal position, and the y axis shows the −log(P value) for each variant in the discovery genome-wide analysis. The red line shows genome-wide significance (P < 5 × 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.

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 (Table 3 and E6). After adjustments for multiple testing (P < 5.2 × 10−7), the previously reported variants at FAM13A, DSP, and IVD were associated with increased FVC, and variants at FAM13A, DSP, 7q22.1 (ZKSCAN1), and ATP11A were associated with increased FEV1/FVC. Similarly, for the three novel susceptibility variants, all showed at least a nominal association with decreased FVC and increased FEV1/FVC. We observed a nominally significant association of the MUC5B IPF risk allele with decreased FVC and increased FEV1/FVC. The IPF risk alleles at MAPT were significantly associated with both increased FEV1 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 MLL, 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, MADIL1, 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 P1 for inclusion of variants in the score was increased, the risk score became more significant reaching a plateau at around P1 = 0.2 with risk score P < 3.08 × 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 P1 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). TGFβ–induced DEPTOR suppression can stimulate collagen synthesis (42), and the importance of mTORC1 signaling via
4E-BP1 for TGFβ-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. RTELI, a gene involved in telomere elongation regulation, has not previously been identified in an IPF GWAS; however, the collective effect of rare variants in RTELI 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/FVC ratio. In an independent dataset of COPD patients,

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.
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

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

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

<table>
<thead>
<tr>
<th>Chr</th>
<th>rsid of Sentinel Variant</th>
<th>Annotation</th>
<th>eQTL</th>
<th>FEV₁</th>
<th>FVC</th>
<th>FEV₁/FVC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lung Tissue</td>
<td>Nonlung Tissue</td>
<td>β [95% CI]</td>
<td>P Value</td>
</tr>
<tr>
<td>3</td>
<td>rs78238620 Intron (KIF15)</td>
<td>—</td>
<td>↓ KIF15</td>
<td>—</td>
<td>0.011 [-0.022 to 0.000]</td>
<td>0.069</td>
</tr>
<tr>
<td>7</td>
<td>rs12699415 Intron (MAD1L1)</td>
<td>—</td>
<td>↓ MAD1L1</td>
<td>TMEM42</td>
<td>-0.007 [-0.012 to -0.002]</td>
<td>0.011</td>
</tr>
<tr>
<td>8</td>
<td>rs28513081 Intron (DEPTOR)</td>
<td>↓ DEPTOR</td>
<td>↓ RP11-760H22.2</td>
<td>↓ RP11-760H22.2</td>
<td>0.001 [-0.004 to 0.000]</td>
<td>0.822</td>
</tr>
</tbody>
</table>

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 β 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 FEV1. 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.

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


26. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al.; 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/NCl; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NRDR; Biospecimen Collection Source Site—RPIC; 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.


31. Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH National Human Genome Research Institute Project Management; ELSI Study; 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.
