Originally Published in: Richard J Allen, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis.
American Journal of Respiratory and Critical Care Medicine
DOI: 10.1164/rccm.201905-1017oc
Copyright © 2019 by the American Thoracic Society
The final publication is available at https://www.atsjournals.org/doi/abs/10.1164/rccm.201905-1017OC.

Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis

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

[1]Department of Health Sciences, University of Leicester, Leicester, UK,

[2]Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain,

[3]Department of Internal Medicine, University of California Davis, Davis, USA,

[4]Division of Pulmonary & Critical Care Medicine, University of Virginia, Charlottesville, USA,

[5]Genentech, South San Francisco, USA,

[6]The University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada,

[7]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK,

[8]Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK,

[9]National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals, Nottingham, UK,

[10]Servei de Pneumologia, Laboratori de Pneumologia Experimental, IDIBELL, Spain

[11]Campus de Bellvitge, Universitat de Barcelona, Barcelona, Spain,

[12]CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Spain,

[13]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, USA,

[14]Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, USA,

[15]Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand,

[16]Department of Thoracic Medicine, University College London Hospitals, London, UK,

[17]Discovery Medicine, GlaxoSmithKline, Stevenage, UK,

[18]Respiratory Research Group, Hull York Medical School, Castle Hill Hospital, Cottingham, UK,

[19]Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, Oxford, UK,

[20]MRC Centre for Inflammation Research at the University of Edinburgh, Edinburgh, UK,

[21]UCL Respiratory Centre for Inflammation and Tissue Repair, University College London, London, UK,

[22]Academic Respiratory Unit, School of Clinical Sciences, University of Bristol, Bristol, UK,

[23]Cambridge ILD Service, Royal Papworth Hospital, Cambridge, UK,

[24]Respiratory Medicine, Nottingham University Hospitals Trust, UK,

[25]Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, USA,

[26]Simmons Center for Interstitial Lung Diseases, University of Pittsburgh, Pittsburgh, USA,

[27]Section of Pulmonary , Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, USA,

[28]Section of Pulmonary & Critical Care, Department of Medicine; The University of Chicago, Chicago, USA,

[29]Dept of Respiratory Medicine, Landspital University Hospital, Reykjavik, Iceland,

[30]Faculty of Medicine, University of Iceland, Reykjavik, Iceland,

[31]University of Iceland, Reykjavik, Iceland,

[32]Icelandic Heart Association, Kopavogur, Iceland,

[33]Department of Radiology, Brigham and Women's Hospital, Boston, USA,

[34]Center for Pulmonary Functional Imaging, Brigham and Women's Hospital, Boston, USA,

[35]Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, USA,

[36]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA,

[37]Center for Public Health Genomics, University of Virginia, Charlottesville, USA,

[38]Department of Public Health Sciences, University of Virginia, Charlottesville, USA,

[39]Division of Cardiovascular and Pulmonary Imaging, Department of Radiology, University of Iowa Carver College of Medicine, Iowa City, USA,

[40]Department of Radiology, University of Washington, Seattle, USA,

[41]Pulmonary Center, Department of Medicine, Boston University, Boston, USA,

[42]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham,USA,

[43]Center for Genomics and Personalized Medicine, Wake Forest School of Medicine, Winston-Salem, USA,

[44]Department of Biostatistics, Boston University School of Public Health, Boston, USA,

[45]Center for Genes, Environment and Health, National Jewish Health, Denver, USA,

[46]Department of Biostatistics and Informatics, University of Colorado, Denver, USA,

[4743]Institut universitaire de cardiologie et de pneumologie de Québec, Université Laval, Quebec City, Canada,

[48]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA,

[49]Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA,

[50]Merck Research Laboratories, Genetics and Pharmacogenomics, Boston, USA,

[51]Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, Canada,

[52]University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, The Netherlands,

[53]University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, The Netherlands,

[54]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK,

[55]Department of Biostatistics, University of Liverpool, Liverpool, UK,

[56]Division of Respiratory Medicine, University of Nottingham, Nottingham, UK,

[57]National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK,

[58]NIHR Respiratory Clinical Research Facility, Royal Brompton Hospital, London, UK,

[59]National Heart and Lung Institute, Imperial College, London, UK,

[60]Department of Medicine, University of Colorado Denver, Denver, USA,

[61]Department of Immunology, University of Colorado Denver, Denver, USA,

[62]Instituto Tecnológico y de Energías Renovables (ITER, S.A.), Santa Cruz de Tenerife, Spain,

[63]Instituto de Tecnologías Biomédicas (ITB), Universidad de La Laguna, Santa Cruz de Tenerife, Spain,

* Contributed equally

Corresponding author: Louise Wain (email: lvw1@leicester.ac.uk, phone: 0116 229 7252)

Author contributions

LVW, RGJ, IN, CF, RJA and JMO designed the study. RJA, BGG, AD, BLY, S-FM, MiN, MLP, LMK, MO, XL, BDH, RKP and PS analysed the data. RGJ, LVW, IN, DAS, TEF, CF, JMO, SFM, RB, MMM, HLB, WAF, SPH, MRH, NH, RBH, RJM, ABM, VN, EO, HP, GS, MKBW, YZ, NK, AA, MES, MaN, XS, IPH, IS, MDT, TMM, BLY, PLM MDH, RKP, PS, GG, VG, HH, DLJ, AM, JDN, GTO'C, VEO, HX, MHC, GMH, MO, YB, KH, PJ, DCN, DDS and WT were responsible for recruitment, screening and genotyping of cases and controls for IPF, ILA and gene expression analyses. LVW, RGJ, IN, CF, JMO and DAS supervised and coordinated the study. RJA, RGJ and LVW led the writing of the manuscript. All authors contributed to drafting and providing critical feedback on the manuscript.

Funding

R. Allen is an Action for Pulmonary Fibrosis Research Fellow. L. Wain holds a GSK/British Lung Foundation Chair in Respiratory Research. RG. Jenkins is supported by an NIHR Research Professorship (NIHR reference RP-2017-08-ST2-014). I. Noth: National Heart Lung and Blood Institute (R01HL130796). B. Guillen is funded by Agencia Canaria de Investigación, Innovación y Sociedad de la Información (TESIS2015010057) co-funded by European Social Fund. J. Oldham: National Heart Lung and Blood Institute (K23HL138190). C. Flores: Spanish Ministry of Science, Innovation and Universities (grant RTC-2017-6471-1; MINECO/AEI/FEDER, UE) co-financed by the European Regional Development Funds (ERDF) 'A way of making Europe' from the European Union, and by the agreement OA17/008 with Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology. The Spain Biobank array genotyping service was carried out 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 co-financed by ERDF. P. Molyneaux is an Action for Pulmonary Fibrosis Research Fellow. M. Obeidat is a fellow of the Parker B Francis Foundation and a Scholar of the Michael Smith Foundation for Health Research (MSFHR). B. Hobbs: NIH K08 HL136928, Parker B. Francis Research Opportunity Award. M. Cho and G Hunninghake: This work was supported by NHLBI grants R01HL113264 (M.H.C), R01HL137927 (M.H.C.), R01HL135142 (M.H.C. and G.M.H) and R01111024 (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. Maher 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. Tobin is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). The research was partially supported by the National Institute for Health Research (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. Hall 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. Sayers: MRC (G1000861) and Asthma UK (AUK-PG-2013-188). D. Furniss 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. Navaratnam is funded by an NIHR Clinical Lectureship. G. Gudmundsson is supported by project grant 141513-051 from the Icelandic Research Fund and Landspitali Scientific Fund A-2016-023, A-2017-029 and A-2018-025. A. Manichaikul and D. Lederer: MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 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, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA)

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.).

Shortened title: Genome-wide study of idiopathic pulmonary fibrosis

Subject category: 9.23 Interstitial Lung Disease

Word count:

Abstract

Rationale: Idiopathic pulmonary fibrosis (IPF) is a complex lung disease characterised 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 defence, telomere maintenance, signalling and cell-cell adhesion.

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

Methods and measurements: We conducted genome-wide analyses across three independent studies and meta-analysed 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 (1,457 IPF cases and 11,874 controls) and functional follow-up 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.

Main results: We identified and replicated three new genome-wide significant ($P<5\times10^{-8}$) signals of association with IPF susceptibility (near *KIF15*, *MAD1L1* and *DEPTOR*) and confirm 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 signalling in lung fibrosis. New signals of association implicating *KIF15* and *MAD1L1* suggest a possible role of mitotic spindle-assembly genes in IPF susceptibility.

Abstract word count:

Key words: Genetics, Epidemiology, KIF15, MAD1L1, DEPTOR

Introduction

Idiopathic pulmonary fibrosis (IPF) is a devastating lung disease characterised by the build-up 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 3-5 years[1,2]. Two drugs (pirfenidone and nintedanib) have been approved for the treatment of IPF, but neither offer a cure and 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 defence, telomere maintenance, cell-cell adhesion and signalling 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 five-fold 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 utilising all European cases and controls recruited to any previously reported IPF GWAS[5-8] and meta-analysing 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 a still 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 an abstract and preprint[16-18].

Methods

Study cohorts

We analysed 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 **Appendix** for summaries of these collections). Two more independent case-control collections (named here as the UUS and Genentech studies) were included as replication

datasets. The new UUS study recruited cases from the USA, UK 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 **Appendix**). The previously described[20] Genentech study consisted of cases from three IPF clinical trials and controls from four non-IPF clinical trials (**Appendix**). 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 **Appendix** and **Figure E1**). 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 Consortium (HRC) r1.1 panel[24] (**Appendix**). For individuals in the Genentech study, genotypes were derived from whole-genome sequencing data. Duplicated individuals between studies were removed (**Appendix**).

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 bi-allelic autosomal variants that had a minor allele count ≥10, were in Hardy-Weinberg Equilibrium (P>1×10^{-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 (**Appendix**).

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 ($P<5\times10^{-8}$), 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, **Appendix**). Previously reported signals with $P<5\times10^{-8}$ in the discovery meta-analysis were deemed as a confirmed association.

Characterisation 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 (**Appendix**).

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 VEP[26]). In addition, we tested to see if any of the credible set variants were associated with gene expression using three eQTL resources (the Lung eQTL study

[n=1,111][27-29], the NESDA-NTR blood eQTL database [n=4,896][30] and 48 tissues in GTEx[31] [n between 80 and 491], **Appendix**). 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% (**Appendix**).

To investigate whether the IPF susceptibility variants that were in non-coding 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[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[33] and GARFIELD[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 **Appendix**.

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 ILA in a GWAS comparing 1,699 individuals with an ILA and 10,247 controls[37]. Lung function measures investigated were, FEV_1 (volume of air an individual can forcibly exhale in the first second), FVC (total volume of air that can be forcibly exhaled), the ratio FEV_1/FVC (used in the diagnosis of COPD) and PEF (the peak expiratory flow). We

applied a Bonferroni corrected *P* value threshold to define variants also associated with ILA 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 an LD $r^2 \leq 0.1$. As we wanted to explore the contribution from as-yet unreported variants, we excluded variants within 1Mb of each IPF susceptibility locus from the risk score calculation (**Appendix**).

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].

Data availability statement

Full summary statistics for the genome-wide meta-analysis can be downloaded at https://github.com/genomicsITER-developers/PFgenetics/blob/master/README.md

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, 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 793 cases and 10,000 controls available in the UUS study and 664 cases and 1,874 controls available in the Genentech study (**Appendix**).

To identify new signals of association, we meta-analysed 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 ≥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 2i**], 7p22.3 [near *MAD1L1*, **Figure 2ii**] and 8q24.12 [near *DEPTOR*, **Figure 2iii**]) showed a genome-wide significant ($P<5x10^{-8}$) 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. 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, Tables E2 and E3i**). The IPF risk allele for the novel chromosome 7 signal (rs12699415, MAF=42.0%) was associated with decreased expression of *MAD1L1* in heart tissue[31] (**Figure E8, 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, 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, Figure E4**). The signal at *FAM13A,* whilst 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 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 foetal lung tissue (**Figure 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[40]. We undertook a look-up 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 E6**). After adjustments for multiple testing ($P<5.2\times10^{-4}$), the previously reported variants at *FAM13A*, *DSP* and *IVD* were associated with decreased FVC and variants at *FAM13A*, *DSP*, 7q22.1 (*ZKSCAN1*) and *ATP11A* 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 whether the variants identified for IPF susceptibility are driven by 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**), eight 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 aetiology between IPF and ILA[37,41,42], we investigated whether our three new signals, and the 11 previously reported signals, were associated with ILA in the largest ILA GWAS reported to date[37]. Eight of the IPF susceptibility loci were at least nominally significantly associated with either ILA or subpleural ILA 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 ILA (although the rare risk allele at *HECTD2* that did not replicate in our study showed some association with an increased risk of subpleural ILA [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 1Mb). The polygenic risk score was significantly associated with increased IPF susceptibility despite exclusion of the known genetic association signals (including *MUC5B*).

As the *P*-threshold (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 $P<3.08\times10^{-23}$ 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 DEP Domain containing MTOR interacting protein. *DEPTOR* inhibits mTOR 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[43] and the importance of mTORC1 signalling via 4E-BP1 for TGFβ induced collagen synthesis has recently been demonstrated in fibrogenesis[44]. *MAD1L1*, implicated by a new signal on chromosome 7 and eQTL analyses of non-lung tissue, is a mitotic checkpoint gene, mutations in which have been associated with multiple cancers including lung cancer[45,46]. 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)[46,47]. This could suggest that *MAD1L1* may increase IPF susceptibility through reduced telomerase activity. Another

spindle-assembly related gene[48], *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* have been reported as associated with IPF susceptibility[52-55]. The ubiquitin E3 ligase encoded by *HECTD2* has been shown to have a pro-inflammatory role in the lung and other *HECTD2* variants may be protective against acute respiratory distress syndrome[56]. 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_1/FVC ratio. In an independent

dataset of 400,102 individuals, eight of the IPF signals were associated with decreased FVC and with a comparatively weaker effect on FEV_1. 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 whilst 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 data set. 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 data sets. 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 non-lung tissues but not in either of the two (non-fibrotic) 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 signalling in pulmonary fibrosis as well as the possible implication of mitotic spindle-assembly genes.

Acknowledgements

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. Genome-wide summary statistics are available on request via the corresponding author.

Table 1: Demographics of study cohorts

	Chicago		Colorado		UK		UUS		Genentech	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
n	541	542	1,515	4,683	612	3,366	793	10,000	664	1,874
Genotyping array /sequencing	Affymetrix 6.0 SNP array		Illumina Human 660W Quad BeadChip		Affymetrix UK BiLEVE array	Affymetrix UK BiLEVE and UK Biobank arrays	Affymetrix UK Biobank and Spain Biobank arrays	Affymetrix UK BiLEVE and UK Biobank arrays	HiSeq X Ten platform (Illumina)	
Imputation panel	HRC		HRC		HRC		HRC		-	
Age (mean)	68	63 [a]	66	-	70 [b]	65	69	58	68	-
Sex (% males)	71% [c]	47% [d]	68%	49%	71%	70%	75.2%	72.1%	73.5%	27.1%
% ever smokers	72%	42%	-	-	72.9% [e]	70%	68.7% [f]	68%	67.3%	18.1% [g]

[a] Age only available for 103 Chicago controls
[b] Age available for 602 UK cases
[c] Sex only available for 500 Chicago cases
[d] Sex only available for 510 Chicago controls
[e] Smoking status only recorded for 236 UK cases
[f] Smoking status only recorded for 753 IPF cases in UUS
[g] Smoking status only recorded for 481 of the Genentech controls

Table 2 - Discovery and replication association analysis results for the five signals reaching significance in the discovery GWAS that have not previously reported as associated with IPF

The minor allele is the effect allele and the minor allele frequency (MAF) is taken from across the studies used in the discovery meta-analysis.

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*	OR [95% CI]	*P*	OR [95% CI]	*P*
3	44902386	rs78238620	*KIF15*	T	A	5.3%	1.58 [1.37, 1.83]	5.12×10^{-10}	1.48 [1.24, 1.77]	1.43×10^{-5}	1.54 [1.38, 1.73]	4.05×10^{-14}
7	1909479	rs12699415	*MAD1L1*	G	A	42.0%	1.28 [1.19, 1.37]	7.15×10^{-13}	1.29 [1.18, 1.41]	2.27×10^{-8}	1.28 [1.21, 1.35]	9.38×10^{-20}
8	120934126	rs28513081	*DEPTOR*	A	G	42.8%	0.82 [0.76, 0.87]	1.20×10^{-9}	0.87 [0.80, 0.95]	0.002	0.83 [0.79, 0.88]	1.93×10^{-11}
10	93271016	rs537322302	*HECTD2*	C	G	0.3%	7.82 [3.77, 16.2]	3.43×10^{-8}	1.75 [0.81, 3.78]	0.155	3.85 [2.27, 6.54]	6.25×10^{-7}
20	62324391	rs41308092	*RTEL1*	G	A	2.1%	2.12 [1.67, 2.69]	7.65×10^{-10}	1.45 [1.08, 1.94]	0.012	1.82 [1.51, 2.19]	2.24×10^{-10}

Table 3 – Gene expression and spirometric results for the three novel IPF susceptibility loci
Annotation of the variant was taken from VEP. A list of all variants included in the credible sets with their annotations and eQTL results can be found in **Table E3**. For colocalisation, only genes where there was a greater than 80% probability of colocalisation between the IPF risk signal and gene expression of that gene are reported in this table. In the colocalisation 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 colocalisation analyses can be found in **Table E2**. The spirometric results for the three novel IPF risk loci are taken from Shrine et al 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**.

Chr	rsid of sentinel variant	Annotation	eQTL		FEV_1		FVC		FEV_1 / FVC	
			Lung tissue	Non-lung tissue	β [95% CI]	*P*	β [95% CI]	*P*	β [95% CI]	*P*
3	rs78238620	Intron (*KIF15*)	-	↓ *KIF15* ↓ *TMEM42*	−0.011 [−0.022, 0.000]	0.069	−0.022 [-0.033, 0.011]	2.92×10^{-4}	0.017 [0.006, 0.028]	0.005
7	rs12699415	Intron (*MAD1L1*)	-	↓ *MAD1L1*	−0.007 [−0.012, −0.002]	0.011	−0.011 [−0.016, −0.007]	1.41×10^{-5}	0.008 [0.003, 0.012]	0.005
8	rs28513081	Intron (*DEPTOR*)	↓ *DEPTOR* ↓ RP11-760H22.2	↕ *DEPTOR* ↕ RP11-760H22.2 ↑ KB-1471A8.1 ↑ *TAF2*	0.001 [−0.004, 0.006]	0.822	−0.005 [−0.010, −0.001]	0.045	0.011 [0.006, 0.016]	4.22×10^{-5}

Figure 1 - Manhattan plot of discovery analysis results

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 ($P<5\times10^{-8}$) 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 labelled in grey are previously reported signals that reach significance in the discovery genome-wide meta-analysis. Genes labelled in black are the novel signals identified in the discovery analysis that reach genome-wide significance when meta-analysing discovery and replication samples. The signals which did not replicate are shown by red labels. For ease of visualisation the y axis has been truncated at 25.

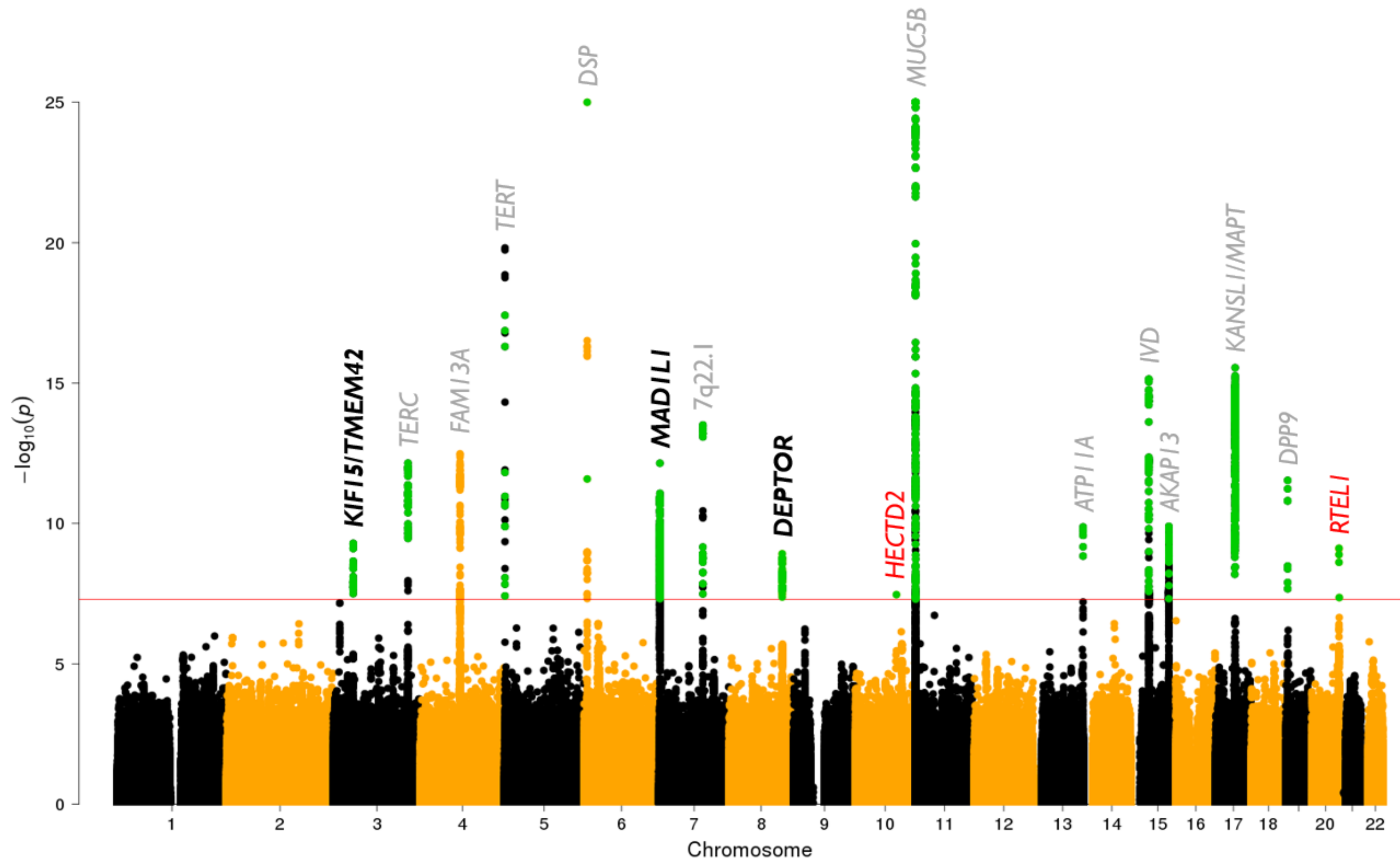
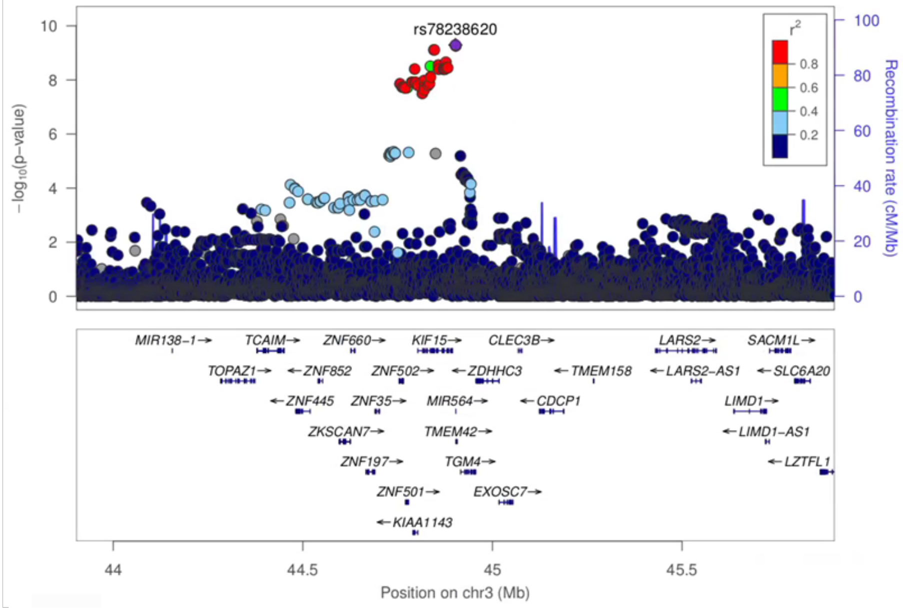


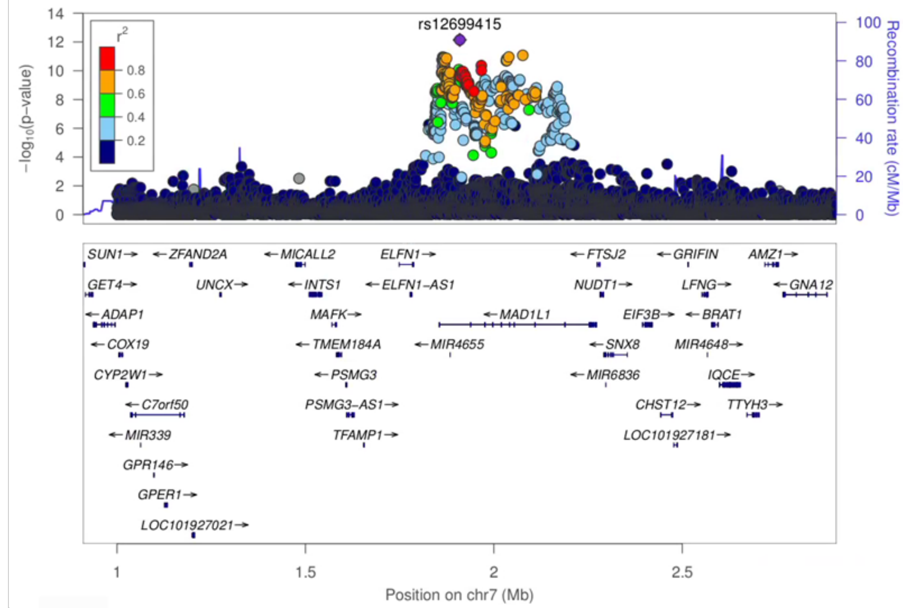
Figure 2 - Region plots of three novel IPF susceptibility loci from discovery genome-wide meta-analysis

Each point represents a variant with chromosomal position on the x axis and the −log(*P* value) on the y axis. Variants are coloured in by LD 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 IPF susceptibility loci, i.e. i) the susceptibility signal on chromosome 3 near *KIF15*, ii) the susceptibility signal on chromosome 7 near *MAD1L1* and iii) the susceptibility signal on chromosome 8 near *DEPTOR*.

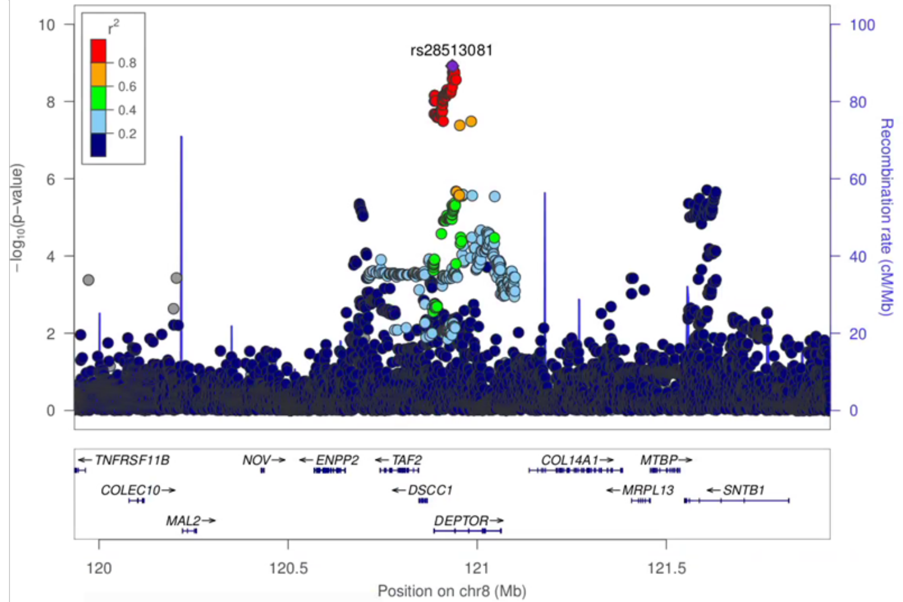
i)



ii)



iii)



References

1. Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med*. 2018;378(19):1811-1823.

2. Ley B, Collard HR, King Jr TE. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2011;183(4):431-440.

3. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47(8):856-860.

4. Mushiroda T, Wattanapokayakit S, Takahashi A, et al. 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(10):654-656.

5. Noth I, Zhang Y, Ma S, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: A genome-wide association study. *The Lancet Respiratory Medicine*. 2013;1(4):309-317.

6. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013;45(6):613-620.

7. Fingerlin TE, Zhang W, Yang IV, 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 genetics*. 2016;17(1):1.

8. Allen RJ, Porte J, Braybrooke R, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of european ancestry: A genome-wide association study. *The Lancet Respiratory Medicine*. 2017.

9. Zhu QQ, Zhang XL, Zhang SM, 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(43):e1901.

10. Coghlan MA, Shifren A, Huang HJ, et al. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. *BMJ open respiratory research*. 2014;1(1):e000057.

11. Petrovski S, Todd JL, Durheim MT, et al. An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2017;196(1):82-93.

12. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426.

13. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. 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, et al. The MUC5B promoter polymorphism is associated with specific interstitial lung abnormality subtypes. *Eur Respir J*. 2017;50(3):10.1183/13993003.00537-2017. Print 2017 Sep.

16. Allen RJ, Oldham JM, Fingerlin TE, et al. Polygenicity of idiopathic pulmonary fibrosis [abstract]. *Genetic Epidemiology*. 2018;42:684.

17. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-wide meta-analysis of susceptibility to idiopathic pulmonary fibrosis [abstract]. *American Society of Human Genetics*. 2018.

18. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *bioRxiv*. 2019:636761.

19. Sudlow C, Gallacher J, Allen N, 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(3):e1001779.

20. Dressen A, Abbas AR, Cabanski C, 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. *The Lancet Respiratory Medicine*. 2018.

21. Travis WD, King TE, Bateman ED, et al. American thoracic society/european respiratory society international multidisciplinary consensus classification of the idiopathic interstitial pneumonias. *American journal of respiratory and critical care medicine*. 2002;165(2):277-304.

22. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine*. 2011;183(6):788-824.

23. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of idiopathic pulmonary fibrosis. an official ATS/ERS/JRS/ALAT clinical practice guideline. *American journal of respiratory and critical care medicine*. 2018;198(5):e44-e68.

24. McCarthy S, Das S, Kretzschmar W, Durbin R, Abecasis G, Marchini J. A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv*. 2016:035170.

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(7):906-913.

26. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122.

27. Hao K, Bossé Y, Nickle DC, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet*. 2012;8(11):e1003029.

28. Lamontagne M, Couture C, Postma DS, et al. Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One*. 2013;8(7):e70220.

29. Obeidat M, Miller S, Probert K, et al. GSTCD and INTS12 regulation and expression in the human lung. *PLoS One*. 2013;8(9):e74630.

30. Jansen R, Hottenga J, Nivard MG, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet*. 2017;26(8):1444-1451.

31. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204.

32. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature methods*. 2015;12(10):931.

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. *F1000Research*. 2015;4.

34. Iotchkova V, Ritchie GR, Geihs M, et al. GARFIELD-GWAS analysis of regulatory or functional information enrichment with LD correction. *bioRxiv*. 2016:085738.

35. Slowikowski K, Hu X, Raychaudhuri S. SNPsea: An algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics*. 2014;30(17):2496-2497.

36. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*. 2019:1.

37. Hobbs BD, Putman RK, Araki T, et al. Overlap of genetic risk between interstitial lung abnormalities and idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2019((in press)).

38. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic risk score software. *Bioinformatics*. 2015;31(9):1466-1468.

39. Moore C, Blumhagen R, Yang I, et al. Resequencing study confirms host defence and cell senescence gene variants contribute to the risk of idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2019.

40. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426.

41. Hunninghake GM, Hatabu H, Okajima Y, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. *N Engl J Med*. 2013;368(23):2192-2200.

42. Putman RK, Hunninghake GM, Dieffenbach PB, et al. Interstitial lung abnormalities are associated with acute respiratory distress syndrome. *American journal of respiratory and critical care medicine*. 2017;195(1):138-141.

43. Das F, Bera A, Ghosh-Choudhury N, Abboud HE, Kasinath BS, Choudhury GG. TGFβ-induced deptor suppression recruits mTORC1 and not mTORC2 to enhance collagen I (α2) gene expression. *PloS one*. 2014;9(10):e109608.

44. Woodcock HV, Eley JD, Guillotin D, et al. The mTORC1/4E-BP1 axis represents a critical signaling node during fibrogenesis. *Nature communications*. 2019;10(1):6.

45. 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(6):e0128814.

46. Coe BP, Lee EH, Chi B, 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 and Cancer*. 2006;45(1):11-19.

47. Kang JU, Koo SH, Kwon KC, Park JW, 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):1-11.

48. Tanenbaum ME, Macůrek L, Janssen A, Geers EF, Alvarez-Fernández M, Medema RH. Kif15 cooperates with eg5 to promote bipolar spindle assembly. *Current biology*. 2009;19(20):1703-1711.

49. Alder JK, Chen JJ, Lancaster L, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc Natl Acad Sci U S A*. 2008;105(35):13051-13056.

50. Stuart BD, Lee JS, Kozlitina J, et al. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: An observational cohort study with independent validation. *The lancet Respiratory medicine*. 2014;2(7):557-565.

51. McDonough JE, Martens DS, Tanabe N, et al. A role for telomere length and chromosomal damage in idiopathic pulmonary fibrosis. *Respiratory research*. 2018;19(1):132.

52. Stuart BD, Choi J, Zaidi S, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. *Nat Genet*. 2015;47(5):512-517.

53. Kropski JA, Loyd JE. Telomeres revisited: RTEL1 variants in pulmonary fibrosis. *Eur Respir J*. 2015;46(2):312-314.

54. Kannengiesser C, Borie R, Menard C, et al. Heterozygous RTEL1 mutations are associated with familial pulmonary fibrosis. *Eur Respir J*. 2015;46(2):474-485.

55. Deng Y, Li Z, Liu J, et al. Targeted resequencing reveals genetic risks in patients with sporadic idiopathic pulmonary fibrosis. *Hum Mutat*. 2018;39(9):1238-1245.

56. Coon TA, McKelvey AC, Lear T, et al. The proinflammatory role of HECTD2 in innate immunity and experimental lung injury. *Science translational medicine*. 2015;7(295):295ra109-295ra109.

57. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS genetics*. 2013;9(3):e1003348.

58. Calafato MS, Thygesen JH, Ranlund S, et al. Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders. *The British Journal of Psychiatry*. 2018;213(3):535-541.

59. Mullins N, Power R, Fisher H, et al. Polygenic interactions with environmental adversity in the aetiology of major depressive disorder. *Psychol Med*. 2016;46(4):759-770.

60. Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): A genetic association study in UK biobank. *The Lancet Respiratory Medicine*. 2015;3(10):769-781.