EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Research letter

Genetic overlap between idiopathic pulmonary fibrosis and COVID−19

Richard J Allen, Beatriz Guillen-Guio, Emma Croot, Luke M Kraven, Samuel Moss, Iain Stewart, R Gisli Jenkins, Louise V Wain

Please cite this article as: Allen RJ, Guillen-Guio B, Croot E, *et al*. Genetic overlap between idiopathic pulmonary fibrosis and COVID−19. *Eur Respir J* 2022; in press (https://doi.org/10.1183/13993003.03132-2021).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

Copyright ©The authors 2022. For reproduction rights and permissions contact permissions@ersnet.org

Genetic overlap between idiopathic pulmonary fibrosis and COVID−19

Richard J Allen[1]*, Beatriz Guillen-Guio[1]*, Emma Croot[2], Luke M Kraven[1], Samuel Moss[3], Iain Stewart[3], R Gisli Jenkins[3], Louise V Wain[1,4]

[1] Department of Health Sciences, University of Leicester, Leicester, UK
[2] Department of Genetics and Genome Biology, University of Leicester, Leicester, UK
[3] National Heart and Lung Institute, Imperial College London, London, UK
[4] National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK
* These authors contributed equally

To the editor,

Coronavirus disease 2019 (COVID-19) is an infectious disease potentially leading to long lasting respiratory symptoms and has resulted in over 4 million deaths worldwide. Idiopathic pulmonary fibrosis (IPF) is a chronic interstitial lung disease (ILD) characterised by an aberrant response to alveolar injury leading to progressive scarring of the lungs. Individuals with ILD are at a higher risk of death from COVID-19[1].

Large genome-wide association studies (GWAS) have identified multiple genetic signals associated with severe COVID-19[2], including a signal within the *DPP9* gene that is also associated with increased IPF risk[3]. GWAS have identified 20 genome-wide significant signals of association with IPF risk[4,5] with the largest genetic risk factor being a common variant located in the promoter region of *MUC5B* (rs35705950, odds ratio>4). Previous analyses suggest IPF is a causal risk factor for severe COVID-19 but noted that the effect of rs35705950 was in the opposite direction (i.e. the allele associated with increased risk of IPF was protective for severe COVID-19)[6].

We aimed to further explore the shared genetic architecture and identify novel shared genetic loci between the two diseases, using new enlarged GWAS of IPF and COVID-19 risk.

We used the largest GWAS of IPF risk which consisted of unrelated European individuals from across five studies[5]. Cases were selected from centres in the USA, UK and Spain diagnosed using American Thoracic Society and European Respiratory Society guidelines. This data is available to access from https://github.com/genomicsITER/PFgenetics.

For COVID-19, the summary statistics from version 6 of the COVID-19 Host Genetics Initiative (HGI_v6, available to access from https://www.COVID-19hg.org/results/r6/) were used. This analysis considered four different COVID-19 phenotypes according to the severity of the disease and the controls used; A2) Very severe respiratory confirmed COVID-19 vs. population, B1) Hospitalised COVID-19 vs. not hospitalised COVID-19, B2) Hospitalised COVID-19 vs. population and, C2) COVID-19 vs. population. The COVID-19 phenotypes A2, B1 and B2 capture both susceptibility and severity of COVID-19, while phenotype C2 captures only susceptibility to COVID-19 infection.

Using LD Score Regression[7], we calculated the genome-wide genetic correlation between IPF and the four COVID-19 phenotypes. There was a significant weak positive genome-wide correlation between

IPF and COVID-19 severity phenotypes (A2 r^2=0.274 p=0.0045, B1 r^2=0.279 p=0.0093, and B2 r^2=0.261 p=0.0005) but not with COVID-19 infection (C2 r^2=0.066 p=0.433).

We investigated the twenty previously reported IPF genetic association signals[4,5] for their association in the four COVID-19 GWAS, and 26 variants reaching genome-wide significance in the COVID-19 GWAS were tested for their association with IPF (proxy variants – r^2>0.8 in European population – were investigated if the top associated variant was not included). At genetic loci showing an association with both traits (after Bonferroni correction for multiple testing), we investigated whether the same causal variant was driving both the IPF and COVID-19 associations using coloc[8]. Regions with a posterior probability>80% of there being a shared causal variant (assuming up to one causal variant for each trait in the region and that variant has been measured) were deemed to have colocalised. Four genetic association signals showed evidence of a shared causal variant between IPF and at least one COVID-19 phenotype (posterior probability>80%), namely loci at 7q22.1, near *MUC5B*, near *ATP11A* and near *DPP9* (**Table 1a**). The 7q22.1 locus has not previously been reported for association with COVID-19. Three additional IPF genetic signals (at 17q21.31, *DSP* and *DEPTOR*) showed an association with COVID-19 but did not colocalise, suggesting there are different causal variants between the two traits at these loci. Visual inspection of the 17q21.31 locus revealed extended linkage disequilibrium (due to the presence of a large inversion) meaning colocalisation analyses could not determine whether there were shared or distinct causal variants.

For the four genetic loci shared between IPF and at least one COVID-19 phenotype, we investigated whether shared genetic signals were associated with gene expression in lung tissue (GTEx_lung[9], n=515) and whole blood (eQTLGen[10], n=31,684). Where the variant met a false discovery rate of 5%, colocalisation analyses were performed using coloc and deemed to be linked to gene expression if the posterior probability of a shared causal variant was greater than 80%. Three of the four shared signals colocalised with expression of the single nearest gene in blood or lung (*MUC5B*, *ATP11A* and *DPP9*) (**Table 1b**). The IPF and COVID-19 risk increasing alleles at the 7q22.1 signal colocalised with decreased expression of *ZKSCAN1* and *TRIM4* in blood.

Finally, we performed a phenome-wide association study (PheWAS) to identify if the overlapping genetic signals had been previously reported for association with other traits ($p<10^{-5}$) using publicly available resources (PhenoScanner_v2, GWAS Catalog and Open Targets). Colocalisation analyses were performed to determine if the same causal variant was driving both traits. The signal on chromosome 7 was associated with a number of blood traits and the signal near *ATP11A* was associated with blood traits and HbA1c (average blood glucose levels, used in diagnosing diabetes) (**Table 1b**). The IPF and HbA1c signals did not colocalise, however, as diabetes is a risk factor for COVID-19[11], we further investigated the effects on gene expression for this signal in all GTEx tissues. The allele (rs423117_T, the sentinel variant from the Hb1AC GWAS) associated with higher Hb1AC levels was associated with increased *ATP11A* expression in liver and decreased expression in cultured fibroblasts, but there was no association with *ATP11A* expression in blood.

In summary, genetic association signals near *MUC5B*, *DPP9* and *ATP11A* have previously been reported for both COVID-19 severity and IPF risk; we show for the first time that these signals are likely due to the same underlying causal variant. In addition, we show the signal at 7q22.1 associated with IPF, also shows a novel association with COVID-19 and implicates *TRIM4* and *ZKSCAN1*.

Despite a positive genome-wide genetic correlation between IPF risk and the COVID-19 severity phenotypes (A2, B1 and B2), we show that two of the four shared signals (at *MUC5B* and *ATP11A*) have opposite directions of effect on risk for the two diseases. The allele associated with increased risk of IPF and increased *ATP11A* expression in blood (rs9577395_C) was associated with decreased risk of severe COVID-19. The lipid flippase ATP11A has been suggested to have an important role in

the innate immune response, and a depletion of this protein in human cells has been related to an increased inflammatory response[12]. Therefore, an increased expression of *ATP11A* may lead to better COVID-19 outcomes by attenuating chronic inflammation following initial infection. Our PheWAS highlighted a potential link with HbA1c and diabetes risk at this locus via *ATP11A* expression, although effects were tissue dependent. The IPF risk allele at *MUC5B* may have a protective effect in airway defence in patients with COVID-19[6]. These findings of opposite genetic and tissue effects potentially highlight important differences between development of long-term chronic disease and response to infection, which could have implications when considering new drug targets.

The rs2897075_T allele at 7q22.1, associated with increased IPF and COVID-19 risk, was linked to decreased *TRIM4* and *ZKSCAN1* expression. TRIM4 is an important regulator of virus-induced interferon induction pathways and a proteomic study identified significant adjacency between SARS-CoV-2 M protein and TRIM4[13]. Viral infection-induced micro-injury to the alveolar epithelium is thought to be a trigger for development of IPF[14] , suggesting the interferon-mediated innate immune response could be central to both risk of chronic lung disease and worse outcomes due to SARS-CoV-2 infection. We also showed that the IPF and COVID-19-risk variant at *DPP9* was associated with a reduced *DPP9* expression. This serine dipeptidyl peptidase inhibits inflammasome activation[15] and has been related to antigen presentation[16]; having an important role in the immune response. Further functional studies are required to better understand the specific role of these genes in the development of IPF and in response to COVID-19 infection.

Loci previously implicated by IPF GWAS relating to telomere dysfunction (*TERT*, *TERC*, *RTEL1*) and mitotic spindle assembly (*KIF15*, *MAD1L1, SPDL1*, *KNL1*) were not associated with COVID-19.

The colocalisation analyses assume a single measured causal variant. Although conditional analyses found no evidence of multiple independent association signals at the regions studied, we cannot guarantee all causal variants were measured. Furthermore, we utilised whole blood and lung tissue for gene expression so we cannot rule out cell-specific effects. A limitation of our analysis are the population groups used. Given the difficulties in selecting controls for infection GWAS[17], we used all of the HGI COVID-19 GWAS which used four different COVID-19 phenotypes. We found that the genetic correlation results were almost identical across the three COVID-19 severity phenotypes (A2, B1 and B2). This suggests that variation in the colocalisation results may be due to variation in power as a consequence of different sample size and chance of misclassification in the COVID-19 GWAS. Secondly, to maximise the power of the analysis we utilised the largest GWAS of IPF and COVID-19 available. The IPF GWAS included only European individuals however the COVID-19 GWAS was performed as a multi-ancestry analysis with the majority of individuals being from European populations. Further analyses in non-European populations could help identify other overlapping ancestry-specific effects.

In conclusion, using the largest IPF and COVID-19 GWAS to date, we show there is a positive genome-wide genetic correlation between IPF and severe COVID-19 risk. However, some IPF-related pathways may have an opposite (e.g. *MUC5B* and *ATP11A* pathways) effect on severe COVID-19 risk.

Funding and conflicts of interest

R Allen is an Action for Pulmonary Fibrosis Mike Bray Research Fellow, and received registration fees for attendance of British Thoracic Society 2021 winter meeting from British Thoracic Society, outside the submitted work. B Guillen-Guio is supported by Wellcome Trust grant 221680/Z/20/Z. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author

Accepted Manuscript version arising from this submission. Luke Kraven is supported by Medical Research Council and GlaxoSmithKline (IMPACT iCASE PhD studentship (MR/N013913/1)). G Jenkins is a trustee of Action for Pulmonary Fibrosis and reports personal fees from Astra Zeneca, Biogen, Boehringer Ingelheim, Bristol Myers Squibb, Chiesi, Daewoong, Galapagos, Galecto, GlaxoSmithKline, Heptares, NuMedii, PatientMPower, Pliant, Promedior, Redx, Resolution Therapeutics, Roche, Veracyte and Vicore. L Wain holds a GSK/British Lung Foundation Chair in Respiratory Research (C17-1), is supported by Medical Research Council (Research grant MR/V00235X/1) and reports research funding from GSK, Orion, Genentech and AstraZeneca and consultancy for Galapagos, outside of the submitted work. All other authors have nothing to disclose. 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. This research used the SPECTRE High Performance Computing Facility at the University of Leicester.

Table 1: Variants reaching Bonferroni-corrected significance for both IPF and COVID-19.
chr=chromosome. REF=reference allele. EFF=effect allele (i.e. the variant the effect estimates are in relation to). A2=Very severe respiratory confirmed COVID-19 (8,779 cases) vs population (1,001,875 controls. B1=Hospitalised COVID-19 (14,408 cases) vs not hospitalised COVID-19 (73,191 controls). B2=Hospitalised COVID-19 (24,274 cases) vs population (2,061,529 controls). C2=COVID-19 (112,612 cases) vs population (2,474,079 controls). OR=odds ratio. CI=confidence interval. FEV_1=forced expiratory volume in 1 second. FVC=forced vital capacity. HbA1c=Haemoglobin Type A1c. The coloc values give the posterior probability there is a shared causal variant between IPF and that COVID-19 phenotype at that genetic loci. Colocalisation analyses were only performed on signals showing a possible association with both traits after correcting for multiple testing. Signals that colocalise (i.e. posterior probability>80%) between IPF and COVID-19 are highlighted in grey. Percentages shown in the gene expression column are the posterior probability of colocalisation between the IPF risk signal and the gene expression eQTL signal in the tissue stated (only genes with posterior probability>80% are presented in the table). For the PheWAS results, phenotypes where the variant had $p<10^{-5}$ and which colocalised with the IPF signal (posterior probability>80%) are presented. Only non-ILD and non-COVID-19 phenotypes were investigated in PhenoScanner, Open Targets and GWAS Catalog for the PheWAS analysis. Proxy variants (with $r^2>0.8$) were also investigated in PhenoScanner. For Open Targets only traits with genome-wide summary statistics from GWAS Catalog were investigated.

chr:position rsid	REF/EFF	IPF	COVID-19 Phenotypes				Gene expression (tissue, coloc)	PheWAS
			A2	B1	B2	C2		
		OR [95% CI] p	OR [95% CI] p coloc	OR [95% CI] p coloc	OR [95% CI] p coloc	OR [95% CI] p coloc		
chr7:99630342 rs2897075	C/T	1.30 [1.23, 1.37] $p = 1.77\times10^{-21}$	1.07 [1.04, 1.12] $p = 1.63\times10^{-4}$ coloc = 88.2%	1.02 [0.99, 1.05] p = 0.238	1.04 [1.02, 1.06] $p = 5.55\times10^{-4}$ coloc = 48.1%	1.01 [1.00, 1.02] p = 0.014	Decreased *ZKSCAN1* (blood, 99.4%) Decreased *TRIM4* (blood, 85.4%)	• Lung function (FEV_1/FVC, Peak expiratory flow) • Chronic obstructive pulmonary disease • Blood traits (mean corpuscular haemoglobin and volume, red cell distribution width, red blood cell count, mean corpuscular volume, mean corpuscular haemoglobin concentration, platelet count, mean platelet volume) • Impedance of leg right • Low density lipoprotein cholesterol levels
chr11:1241221 rs35705950	G/T	5.06 [4.67, 5.47] $p = 9.09\times10^{-418}$	0.83 [0.77, 0.89] $p = 1.17\times10^{-7}$ coloc = 100%	0.89 [0.84, 0.94] $p = 2.20\times10^{-5}$ coloc = 98.5%	0.89 [0.86, 0.93] $p = 1.22\times10^{-8}$ coloc = 100%	0.99 [0.98, 1.01] p = 0.448	Increased *MUC5B* (lung, 100%)	-
chr13:113534984 rs9577395	G/C	1.29 [1.21, 1.38] $p = 4.78\times10^{-14}$	0.90 [0.87, 0.94] $p = 4.38\times10^{-6}$ coloc = 99.0%	0.94 [0.90, 0.97] $p = 8.76\times10^{-4}$ coloc = 52.1%	0.94 [0.91, 0.96] $p = 8.67\times10^{-7}$ coloc = 99.5%	0.99 [0.98, 1.00] p = 0.037	Increased *ATP11A* (blood, 99.6%)	• Blood traits (mean corpuscular volume, mean corpuscular haemoglobin, red cell distribution width, platelet count, red blood cell count) • HbA1c • Lung function (FEV_1/FVC)
chr19:4717672 rs12610495	A/G	1.28 [1.21, 1.36] $p = 2.56\times10^{-16}$	1.20 [1.15, 1.26] $p = 1.64\times10^{-15}$ coloc = 97.9%	1.08 [1.04, 1.11] $p = 1.73\times10^{-5}$ coloc = 98.5%	1.11 [1.09, 1.14] $p = 6.09\times10^{-18}$ coloc = 97.9%	1.03 [1.02, 1.04] $p = 5.10\times10^{-10}$ coloc = 98.1%	Decreased *DPP9* (blood, 88.3%)	• Appendicular lean mass

References

References

1. Drake TM, Docherty AB, Harrison EM, Quint JK, Adamali H, Agnew S, Babu S, Barber CM, Barratt S, Bendstrup E. Outcome of hospitalization for COVID-19 in patients with interstitial lung disease. an international multicenter study. American Journal of Respiratory and Critical Care Medicine 2020;202(12):1656-65.

2. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD. Genetic mechanisms of critical illness in covid-19. Nature 2021;591(7848):92-8.

3. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 2013;45(6):613-20.

4. Dhindsa RS, Mattsson J, Nag A, Wang Q, Wain LV, Allen R, Wigmore EM, Ibanez K, Vitsios D, Deevi SV. Identification of a missense variant in SPDL1 associated with idiopathic pulmonary fibrosis. Communications Biology 2021;4(1):1-8.

5. Allen RJ, Stockwell A, Oldham JM, Guillen-Guio B, Flores C, Noth I, Yaspan BL, Jenkins RG, Wain LV, International IPF Genetics Consortium. Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. medRxiv 2021(12.06.21266509).

6. Fadista J, Kraven LM, Karjalainen J, Andrews SJ, Geller F, Baillie JK, Wain LV, Jenkins RG, Feenstra B, COVID-19 Host Genetics Initiative. Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. EBioMedicine 2021;65:103277.

7. Bulik-Sullivan BK, Loh P, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47(3):291-5.

8. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genetics 2014;10(5):e1004383.

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

10. Võsa U, Claringbould A, Westra H, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Kasela S. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. BioRxiv 2018:447367.

11. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2020;14(4):303-10.

12. van der Mark, Vincent A, Ghiboub M, Marsman C, Zhao J, van Dijk R, Hiralall JK, Ho-Mok KS, Castricum Z, de Jonge WJ, Oude Elferink RP. Phospholipid flippases attenuate LPS-induced TLR4 signaling by mediating endocytic retrieval of toll-like receptor 4. Cellular and Molecular Life Sciences 2017;74(4):715-30.

13. Meyers JM, Ramanathan M, Shanderson RL, Beck A, Donohue L, Ferguson I, Guo MG, Rao DS, Miao W, Reynolds D. The proximal proteome of 17 SARS-CoV-2 proteins links to disrupted antiviral signaling and host translation. PLoS Pathogens 2021;17(10):e1009412.

14. John AE, Joseph C, Jenkins G, Tatler AL. COVID-19 and pulmonary fibrosis: A potential role for lung epithelial cells and fibroblasts. Immunol Rev 2021;302(1):228-40.

15. Griswold AR, Ball DP, Bhattacharjee A, Chui AJ, Rao SD, Taabazuing CY, Bachovchin DA. DPP9's enzymatic activity and not its binding to CARD8 inhibits inflammasome activation. ACS Chemical Biology 2019;14(11):2424-9.

16. Geiss-Friedlander R, Parmentier N, Möller U, Urlaub H, Van den Eynde, Benoit J, Melchior F. The cytoplasmic peptidase DPP9 is rate-limiting for degradation of proline-containing peptides. J Biol Chem 2009;284(40):27211-9.

17. Mozzi A, Pontremoli C, Sironi M. Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches. Infection, Genetics and Evolution 2018;66:286-307.