Title: Context-specific eQTLs reveal causal genes underlying shared genetic architecture of critically ill COVID-19 and idiopathic pulmonary fibrosis

Authors: Trisha Dalapati[1, 2], Liuyang Wang[2], Angela G. Jones[2, 3], Jonathan Cardwell[4], Iain R. Konigsberg[4], Yohan Bossé[5], Don D. Sin[6], Wim Timens[7], Ke Hao[8], Ivana Yang[4], Dennis C. Ko[2,3,9,10]

Affiliations:

[1]Medical Scientist Training Program, Duke University School of Medicine, Durham, NC, USA.

[2]Department of Molecular Genetics and Microbiology, Duke University School of Medicine, Durham, NC, USA.

[3]University Program in Genetics and Genomics, Duke University, Durham, NC, USA

[4]Department of Biomedical Informatics, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[5]Institut universitaire de cardiologie et de pneumologie de Québec – Université Laval, Department of Molecular Medicine, Québec City, Canada

[6]Center for Heart Lung Innovation, University of British Columbia and St. Paul's Hospital, Vancouver, BC, Canada

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

[8]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[9]Division of Infectious Diseases, Department of Medicine, Duke University School of

Medicine, Durham, NC, USA.

[10]Lead contact

To whom correspondence should be addressed: Dennis C. Ko, 0048B CARL Building

Box 3053, 213 Research Drive, Durham, NC 27710. 919-684-5834.

dennis.ko@duke.edu. @denniskoHiHOST

SUMMARY

Most genetic variants identified through genome-wide association studies (GWAS) are suspected to be regulatory in nature, but only a small fraction colocalize with expression quantitative trait loci (eQTLs, variants associated with expression of a gene). Therefore, it is hypothesized but largely untested that integration of disease GWAS with context-specific eQTLs will reveal the underlying genes driving disease associations. We used colocalization and transcriptomic analyses to identify shared genetic variants and likely causal genes associated with critically ill COVID-19 and idiopathic pulmonary fibrosis. We first identified five genome-wide significant variants associated with both diseases. Four of the variants did not demonstrate clear colocalization between GWAS and healthy lung eQTL signals. Instead, two of the four variants colocalized only in cell-type and disease-specific eQTL datasets. These analyses pointed to higher *ATP11A* expression from the C allele of rs12585036, in monocytes and in lung tissue from primarily smokers, which increased risk of IPF and decreased risk of critically ill COVID-19. We also found lower *DPP9* expression (and higher methylation at a specific CpG) from the G allele of rs12610495, acting in fibroblasts and in IPF lungs, and increased risk of IPF and critically ill COVID-19. We further found differential expression of the identified causal genes in diseased lungs when compared to non-diseased lungs, specifically in epithelial and immune cell types. These findings highlight the power of integrating GWAS, context-specific eQTLs, and transcriptomics of diseased tissue to harness human genetic variation to identify causal genes and where they function during multiple diseases.

Keywords: GWAS, eQTL, IPF, COVID-19, ATP11A, DPP9, COLOC, mQTL, scRNA-seq, iCPAGdb, macrophage, TLR4, efferocytosis

3

INTRODUCTION

Genome-wide association studies (GWAS) are a powerful approach to identify

relationships between disease phenotypes and genetic variants, most commonly single

nucleotide polymorphisms (SNPs). Elucidating associated SNPs can reveal

unanticipated underlying pathophysiology of disease, unbiased by preconceived notions

of plausible biology, and may lead to informed drug targets and therapies. Indeed,

success along the drug development pipeline increases with human genetic evidence.[1]

In 2021, two-thirds of the Food and Drug Administration approved drugs were supported

by human genetic evidence.[2] Beyond associations with individual diseases, multiple

independent GWAS can be integrated to identify pleiotropy – the phenomenon where a

single genetic locus affects multiple diseases or traits that are otherwise considered

unrelated.[3,4] Pleiotropy can reveal shared pathogenic mechanisms among different

diseases. Understanding these mechanisms and identifying potential therapeutic targets

require the identification of the causal genes associated with the genetic variant.

One method to identify putative casual genes is to measure the degree of

colocalization between disease GWAS and molecular quantitative trait loci (QTLs), such

as expression QTLs (eQTLs), using statistical frameworks that integrate summary

statistics to determine if overlapping signals are due to the same causal SNP. [5-7] While

the chain of causality implied by colocalization makes intuitive sense (a SNP affects

transcription of a gene to impact disease), there is surprisingly poor colocalization

between disease GWAS and eQTL signals.[8] One possible reason for lack of

colocalization includes using eQTL datasets from unsuitable contexts. For example,

cell- and disease-specific eQTLs may be required.[8] Alternatively, other kinds of

4

molecular QTLs, such as methylation, splicing, or protein QTLs (mQTLs, sQTLs,

pQTLs, respectively), may be more causally important in certain disease

pathogenesis.[8,9]

Here, we focused on pleiotropic SNPs associated with both critically ill COVID-19

and idiopathic pulmonary fibrosis (IPF) and formally tested colocalization with context-

specific eQTLs to identify the likely causal genes. We previously developed a software

called the interactive Cross-Phenotype Analysis of GWAS database (iCPAGdb) to study

pleiotropy by comprehensively identifying shared association signals between user-

uploaded GWAS and all publicly cataloged GWAS summary statistics.[10,11] Using the

first COVID-19 GWAS published by Ellinghaus et al.[12] and the highest-powered IPF

GWAS at the time,[13] we identified a shared signal in *DPP9*, rs12610495, associated at a

suggestive threshold with severe COVID-19 (p-value = 5.20×10^{-6})[12] and at genome-

wide significance with IPF (p = 2.92×10^{-12})[13].[11] The *DPP9* locus was later confirmed to

be associated with critically ill COVID-19 at genome-wide significance.[14-16] Others have

also confirmed and expanded on the shared genetic associations between critically ill

COVID-19 and IPF.[11,17-19]

Notably, post-COVID-19 pulmonary fibrosis (PCPF) is clinically reminiscent of

IPF.[20] PCPF, which occurs due to irreversible lung scarring and stiffening, causes

progressive difficulties in breathing and ultimately necessitates lung transplantation.[20,21]

Critical illness is a major risk factor for PCPF.[22] A meta-analysis reported that nearly 45

percent of COVID-19 survivors had long-term respiratory symptoms and fibrosis,

emphasizing the urgent need to understand the mechanisms and potential treatments of

PCPF.[22] In fact, drugs for IPF, including pirfenidone and nintedanib, have been co-opted

5

during emergency treatment of PCPF.[23] Thus, shared genetic associations between critically ill COVID-19 and IPF might reflect common pathophysiology that could be utilized for therapeutic benefit in both diseases.

In this study, we leveraged iCPAGdb again with the most current, highest-powered COVID-19 GWAS (COVID-19 HGI release 7)[24] and IPF meta-GWAS[25] to determine whether additional loci are associated with both diseases. We have identified five shared loci, including the *DPP9* locus containing rs12610495, that are likely due to the same causal variants based on colocalization analysis. For two of the variants, their risk alleles are reversed in COVID-19 and IPF, highlighting that a shared signal may have different functions in the two diseases and that the pulmonary damage from COVID-19 is likely distinct from IPF. We systematically identified the likely causal genes underlying the shared genetic architecture between critically ill COVID-19 and IPF by performing colocalization analysis using bulk, single-cell, and disease-specific eQTL datasets. For two genes, *ATP11A* and *DPP9,* we found colocalization of GWAS and eQTL signals only in a cell-type and disease-specific context. Therefore, context-specific eQTL are critical for identifying the causal genes underlying the shared genetic architecture of critically ill COVID-19 and IPF and may lead to new insights connecting the diseases.

6

MATERIALS AND METHODS

GWAS summary statistics and QTL datasets

All datasets used in this study and accession information are summarized in **Table 1** and **Table S1**. For analyses using GTEx v8, we obtained the "Tissue-Specific All SNP Gene Associations" files via the Google Cloud Platform, which included non-significant associations not searchable on the web browser.[26,27] In addition to the IPF eQTL and mQTL summary statistics, Borie et al. provided *DPP9* expression and cg07317664 methylation data.[28] The genome coordinates in datasets were lifted to GRCh38 using the R packages "liftOver"[29] and "rtracklayer"[30].

Table 1: Datasets used for present study

Dataset	Article Author (Year)	Type
Very severe respiratory confirmed (critically ill) COVID-19 vs. population	COVID-19 Host Genetics Initiative (2021)	GWAS
5-way meta-GWAS of IPF susceptibility	Allen (2022)	GWAS
GTEx Conditionally Independent eQTL	Lonsdale (2013); GTEx Consortium (2017, 2020)	eQTL
GTEx Tissue-Specific All SNP Gene Associations eQTLs	Lonsdale (2013); GTEx Consortium (2017, 2020)	eQTL
Human Lung Tissue eQTL Study	Hao (2012)	eQTL
BLUEPRINT	Chen (2016)	eQTL
Bulk IPF	Borie (2022)	eQTL; mQTL
Single-cell COVID-19 PBMCs	Aquino (2023)	eQTL
Single-cell IPF Lung	Natri (2023)	eQTL
IPF Lung Study 1	Jaffar (2022)	Bulk RNA-Seq
IPF Lung Study 2	Ammar (2019)	Bulk RNA-Seq
COVID-19 Whole Blood Study	Li (2021)	Bulk RNA-Seq
COVID-19 Lung Study	de Rooij (2022)	Bulk RNA-Seq
IPF Cell Atlas	Adams (2020)	scRNA-Seq
Lung Atlas of Lethal COVID-19	Melms (2021)	snRNA-Seq
Abbreviations: GTEx, Genotype-Expression Project; SNP, single nucleotide polymorphism; GWAS, genome-wide association studies; eQTL, expression quantitative trait loci; mQTL, methylation quantitative trait loci; RNA-seq, RNA-sequencing; scRNA, single-cell RNA; snRNA, single-nucleus RNA		

Discovery of pleiotropic SNPs using iCPAGdb

We previously reported the development of iCPAGdb which identified cross-phenotypic associations across 4400 traits loaded from the NHGRI-EBI GWAS Catalog.[10,11] For this study, critically ill COVID-19 summary statistics were formatted for the web browser (http://cpag.oit.duke.edu/explore/app/) and uploaded as the User Supplied GWAS for "GWAS source one". NHGRI was selected for "GWAS source two". P-threshold$_1$ (factor$_1$ X 10^{-x_1}) and P-threshold$_2$ (factor$_2$ X 10^{-x_2}) were set to genome-wide significant (5×10^{-8}; 5 for "factor$_1$" and "factor $_2$" and 8 for "x_1" and "x_2"). "European" was chosen for "LD 1000 Genomes population" as the linkage disequilibrium (LD) population as the IPF GWAS was based on European ancestry. Additionally, the command line version of iCPAGdb (https://github.com/tbalmat/iCPAGdb) was utilized to directly compare the critically ill COVID-19 and IPF meta-GWAS, which became publicly available after the publication of iCPAGdb. The full results from iCPAGdb after uploading the critically ill COVID-19 summary statistics are listed in **Table S2**.

Identification of candidate eGenes and LD proxies

The five SNPs shared between COVID-19 and IPF identified by iCPAGdb were queried in the eQTL datasets in **Table 1**. We checked whether each SNP was a conditionally independent eQTL in each of the 54 tissues in GTEx v8. Next, using the GTEx web browser, we downloaded all eGenes associated with each SNP. We then searched for conditionally independent eQTLs across the 54 tissues associated with the list of eGenes. We used LDmatrix to check if any of the conditionally independent eQTLs were in LD with our SNPs and could serve as LD proxies ($r^2 > 0.50$) (**Table S3**).

8

For rs12585036 and rs12610495, the only conditionally independent SNPs in LD were eQTLs for *ATP11A* and *DPP9*, respectively. Two conditionally independent SNPs were in perfect LD with rs2897075 and were eQTLs for *ZKSCAN1* and *COPS6*. rs1105569 was in strong LD ($r^2 > 0.90$) with 330 conditionally independent eQTL associated with 23 eGenes across 49 tissues, underscoring the challenges in identifying causal genes in the 17q21.31 inversion supergene.[31] For this locus, we further filtered conditionally independent eQTL in strong LD in tissues of interest for COVID-19 and IPF pathogenesis– fibroblasts, lung, lymphocytes, and whole blood. From literature review, we identified one additional eGene, *TRIM4*, associated with rs2897075[17] and two additional eGenes, *CRHR1* and *SPPL2C*, associated with rs1105569[31]. Altogether, we focused on 15 candidate protein-coding eGenes for our five SNPs of interest **(Table 2)**.

Table 2: Candidate protein-coding eGenes associated with SNPs of interest

SNP of Interest	Prioritized eGene	Identified via
rs35705950	*MUC5B*	GTEx
rs12585036	*ATP11A*	GTEx
rs12610495	*DPP9*	GTEx
rs2897075	*COPS6*	GTEx
	TRIM4	Literature
	ZKSCAN1	GTEx
rs1105569	*ARHGAP27*	GTEx
	ARL17A	GTEx
	ARL17B	GTEx
	CRHR1	Literature
	KANSL1	GTEx
	LRRC37A	GTEx
	LRRC37A2	GTEx
	MAPT	GTEx
	SPPL2C	Literature
Abbreviations: SNP, single nucleotide polymorphism; eGene, gene with an expression quantitative trait locus; GTEx, Genotype-Expression Project		

Colocalization analysis

We applied Giambartolomei et al.'s colocalization analysis (COLOC), using the R

package "coloc",[6] to determine if the associations identified by iCPAGdb were due to

the same causal SNP. COLOC uses a Bayesian framework to calculate the posterior

probabilities that two traits are not associated in the locus of interest (PP0), only one

trait is associated in the locus (PP1 and PP2), both traits are associated at the locus but

with different, independent causal variants (PP3), or both traits are associated with a

single causal variant in the locus (PP4). For COLOC using only GWAS summary

statistics, we filtered SNPs within a 1 megabase (Mb) window from the SNP of interest.

For COLOC using eQTL datasets, we filtered all eQTLs for a candidate eGene and then

filtered SNPs within a 1 Mb window from the SNP of interest. We ran the COLOC

"coloc.abf" function using the default prior parameters, $p1 = 1 \times 10^{-4}$, $p2 = 1 \times 10^{-4}$, and

$p12 = 1 \times 10^{-5}$ for all analyses. PP4 between 0.700 to 0.900 was interpreted as likely to

share a single causal variant, while PP4 > 0.900 was interpreted as sharing a single

causal variant. The PP4/PP3 measured the intensity of the colocalization signal with

values ≥ 5.00 indicating further support for colocalization and ≥ 3.00 suggesting likely

colocalization.[9,32]

IPF and COVID-19 transcriptomic analysis

We utilized publicly available bulk RNA datasets from lung or blood from patients

diagnosed with COVID-19 and IPF. Normalized counts were downloaded from GEO

(GSE213001, GSE134692, GSE172114).[33-35] When normalized counts were

unavailable, raw counts were downloaded (GSE159585).[36] Genes with counts less than

10

five were filtered out, and the remaining gene counts were normalized using DESeq2[37].

We utilized publicly available single-cell (sc) and single-nucleus (sn) RNA-sequencing

(RNA-Seq) datasets to investigate causal gene expression in IPF and COVID-19 lungs

compared to healthy controls. A scRNA-Seq dataset for IPF from Adams et al.[38]

containing 243,472 cells from 32 IPF lungs and 28 control donor lungs was downloaded

from GEO (GSE136831). A snRNA-Seq dataset for lethal COVID-19 from Melms et al.[39]

containing 116,314 nuclei from 19 lethal COVID-19 lungs and seven control lungs was

downloaded from the Broad Institute Single Cell Portal.[40] Previously annotated cell type

clusters were used to create expression dot plots and calculate cell type proportions

within each donor with the Python package "scanpy".[41] Only cell type clusters with at

least ten cells and cells with at least 500 total counts were retained for downstream

differential expression analyses. We then pseudobulked gene counts by donor and cell-

type using the Python package "decoupleR"[42]. Differential expression analysis between

diseased and healthy lungs was conducted in the Python package "pyDESeq2"[43] which

implements the traditional DESeq2[37] workflow. Cell type proportions calculated from the

datasets are in **Table S4-5**. Differential expression results are in **Table S6-7**.

Data Visualization

All visualizations were created in R version 4.3.0.[44] Lollipop plots, genotypic box plots

and -$\log_{10}$(p-value) scatterplots were generated using "ggplot2".[45] Manhattan plots and

LocusZoom plots were created using "fastman"[46] and "locuszoomr"[47], respectively.

While p-values from eQTL and mQTL summary statistics incorporated PEER factor

normalization, values plotted for gene expression and methylation were pre-PEER

11

normalization. For eQTL data from Borie et al.,[28] normalized RNA TPM values were plotted (transcripts per million after trimmed mean of M values (TMM) normalization across samples and inverse normal transformation on a per-gene basis). For mQTL data from Borie et al.,[28] normalized methylation beta referred to beta values of DNA methylation level measured on Illumina arrays (scale 0–1) after SeSame[48] data preprocessing and normalization.
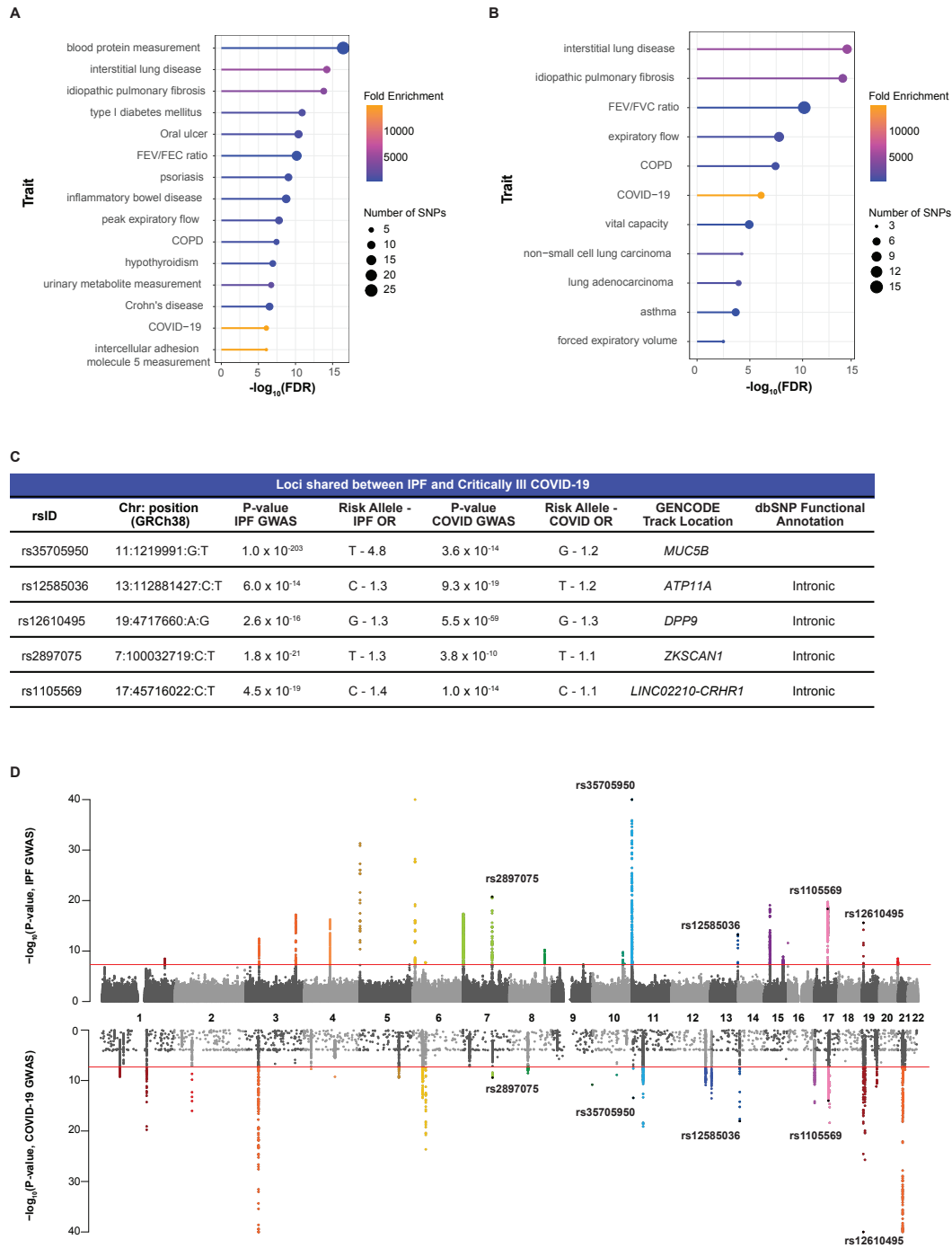
RESULTS

Identification of loci shared between critically ill COVID-19 and idiopathic

pulmonary fibrosis

The underlying causal genes and mechanisms for the shared loci associated with

critically ill COVID-19 and IPF have not been systematically identified. The COVID-19

Human Genetics Initiative integrated 82 studies from 35 countries to create the largest

cohort to date for risk, hospitalization, and critical illness.[24,49,50] Herein, we studied the

critically ill COVID-19 trait (cases = 18,152, controls = 1,145,546) as severe disease is a

risk factor for developing lung fibrosis. iCPAGdb revealed shared genetic associations

between critically ill COVID-19 and multiple human diseases (**Figure 1A; Table S2**).

Most interestingly, critically ill COVID-19 significantly overlapped with several pulmonary

traits (**Figure 1B)**, including interstitial lung disease (ILD) and IPF at five loci (ILD:

Fisher's exact test FDR = 6.05 x 10^{-15} and 3774.0-fold enrichment; IPF: FDR = 1.56 x

10^{-14} and 4780.0-fold enrichment) (**Figure 1C-D**). Importantly, the risk allele was

reversed for two of the five SNPs, rs35705950 and rs12585036. The odds ratio at each

shared locus was also generally larger for IPF, highlighting the complex disease

mechanisms linking COVID-19 and IPF (**Figure 1C**).

13

Figure 1: iCPAGdb reveals associations with critically ill COVID-19.

(A) Lollipop plot of the top 15 traits associated with COVID-19. Fold enrichment is calculated by dividing number of SNPs shared by number of SNPs expected to be shared.

(B) Plot of the pulmonary traits associated with COVID-19.

(C) Table of the five SNPs overlapping between critically ill COVID-19 and IPF. The location and functional annotation are from GENCODE on the UCSC Genome Browser and NCBI dbSNP, respectively.

(D) Miami plot of the IPF and critically ill COVID-19 GWAS highlights the five shared SNPs. Over 7.5 million SNPs were included in the IPF meta-GWAS summary statistics, while the top 10,000 SNPs were publicly available in the critically ill COVID-19 summary statistics from the COVID-19 Human Genetics Initiative.

P-values are -$\log_{10}$ transformed, and the red line indicates the genome-significant threshold at $p < 5.0 \times 10^{-8}$.

A

B

C

Loci shared between IPF and Critically Ill COVID-19							
rsID	Chr: position (GRCh38)	P-value IPF GWAS	Risk Allele - IPF OR	P-value COVID GWAS	Risk Allele - COVID OR	GENCODE Track Location	dbSNP Functional Annotation
rs35705950	11:1219991:G:T	1.0×10^{-203}	T - 4.8	3.6×10^{-14}	G - 1.2	*MUC5B*	
rs12585036	13:112881427:C:T	6.0×10^{-14}	C - 1.3	9.3×10^{-19}	T - 1.2	*ATP11A*	Intronic
rs12610495	19:4717660:A:G	2.6×10^{-16}	G - 1.3	5.5×10^{-59}	G - 1.3	*DPP9*	Intronic
rs2897075	7:100032719:C:T	1.8×10^{-21}	T - 1.3	3.8×10^{-10}	T - 1.1	*ZKSCAN1*	Intronic
rs1105569	17:45716022:C:T	4.5×10^{-19}	C - 1.4	1.0×10^{-14}	C - 1.1	*LINC02210-CRHR1*	Intronic

D

15

Although regional association plots visually suggested colocalizing signals in critically ill COVID-19 and IPF (**Figure 2A-E**), shared associations cannot be assumed to be driven by the same causal SNP. Overlapping GWAS signals may be due to a shared causal SNP or two independent signals driven by different causal SNPs with variable degrees of LD. To systematically reveal whether each signal was due to a single causal SNP, we performed formal colocalization testing using the critically ill COVID-19 GWAS and the IPF meta-GWAS (cases = 4,125, controls = 20,464). For all loci except rs1105569, the PP4 was greater than 0.900 (**Figure 2F; Table S9**). The exception, rs1105569 (PP4 = 0.725), is located within the 17q21.31 inversion supergene with extensive LD extending over nearly 900 kb,[31] which precluded our ability to determine whether an individual causal variant was driving the association in both diseases (**Figure 2E).** Overall, we determined that for the five shared loci identified by iCPAGdb, four demonstrated colocalizing signals likely driven by the same causal SNP in both critically ill COVID-19 and IPF.

Figure 2: IPF and critically ill COVID-19 GWAS colocalize at shared signals.

(A) LocusZoom plot highlights that rs35705950 is the lead variant (purple diamond) in both disease GWAS and is located upstream of *MUC5B* on chromosome 11.

(B) rs12585036 is a top variant in the IPF GWAS and is in strong LD with the lead variant (rs9577395). rs12585036 is the lead variant in the COVID-19 GWAS and is located within an intron of *ATP11A* on chromosome 13.

(C) rs12610495 is the lead variant in both disease GWAS and is located within an intron of *DPP9* on chromosome 19.

(D) rs2897075 is the lead variant in both disease GWAS and is located within an intron of *ZKSCAN1* on chromosome 7.

(E) rs1105569 is a top variant in the IPF GWAS and is in strong LD with the lead variant (rs3785884). rs1105569 is the lead variant in the COVID-19 GWAS. rs1105569 is located within an intron of *CRHR1* and in a supergene with high linkage disequilibrium between SNPs.

(F) COLOC indicates strong colocalization between the disease GWAS signals at four of the five shared SNPs with PP4 > 0.900, and PP4/PP3 > 5.00. COLOC is inconclusive for rs1105569 due to extensive linkage disequilibrium in the region.

P-values are -$\log_{10}$ transformed, and the dotted line indicates the genome-significant threshold at $p < 5.0 \times 10^{-8}$. Linkage disequilibrium information for European populations was obtained from LDlink and is relative to the lead variant in each plot.

17
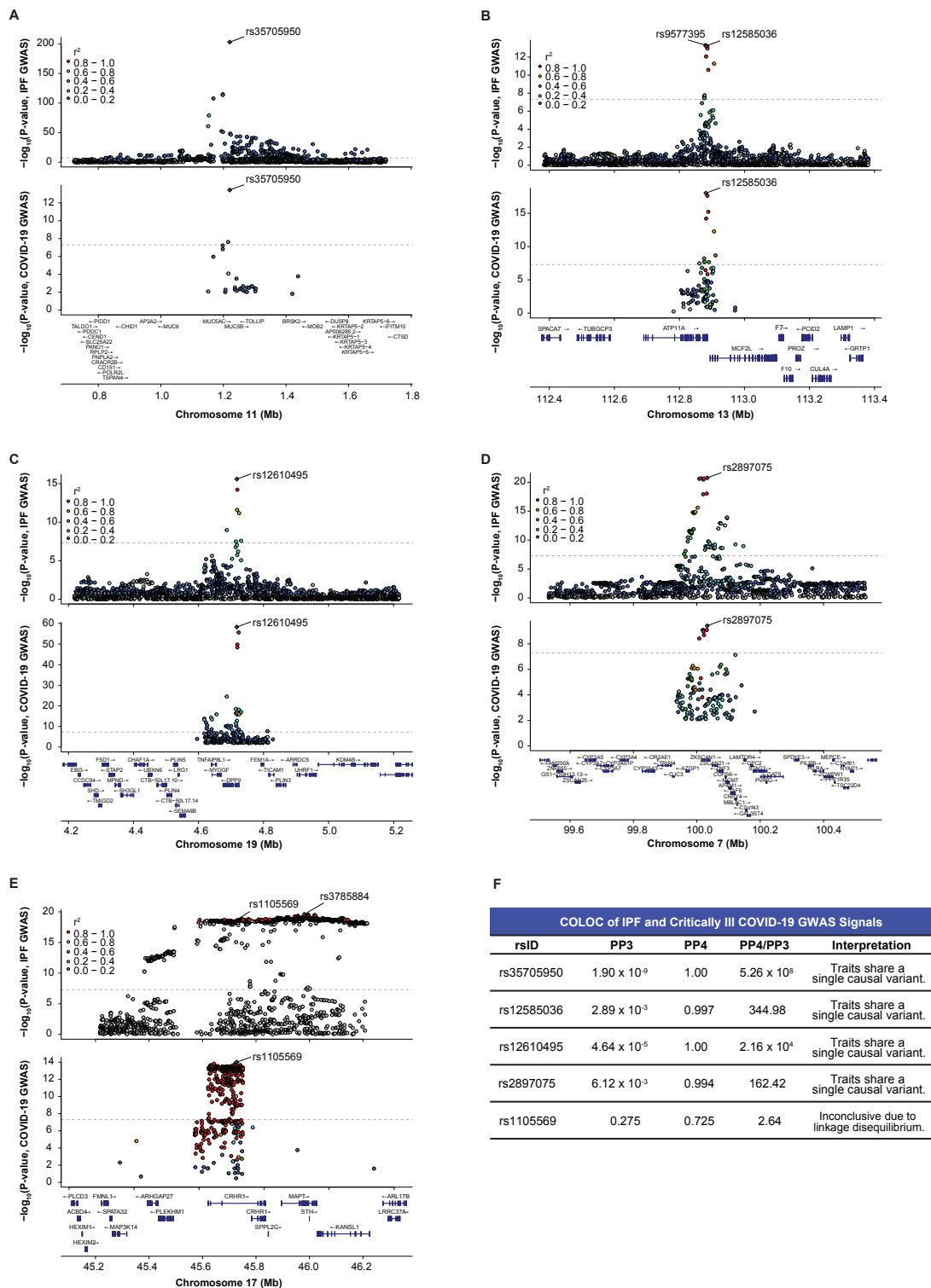
Identification of causal genes for shared loci in bulk tissue

SNPs can affect a nearby gene through altering function or regulation. To speculate how each SNP contributed to gene regulation during COVID-19 and IPF, we used HaploReg,[51] a publicly available tool that annotates the putative functional impact of queried SNPs and all SNPs in LD. We first explored whether any of our SNPs of interest were in strong LD with nonsynonymous variants, altering the amino acid sequence of the protein. For rs1105569, the lead variant was in strong LD (r^2>0.80) with 18 nonsynonymous variants (missense, frameshift, and nonsense) in *CRHR1*, *SPPL2C*, and *MAPT*.[51] These three genes, and several others in 17q21.31, were previously considered important for COVID-19.[14,24,50] No nonsynonymous variants were in strong LD with the other four SNPs of interest, indicating these signals were likely due to regulatory variants.

Next, we determined if the five SNPs of interest were associated with gene expression in bulk, healthy tissue. In GTEx,[26] only rs35705950 was a conditionally independent eQTL for *MUC5B* in lung (nominal p-value = 6.71 x 10^{-16}). We next assessed if our SNPs of interest were in LD with conditionally independent eQTLs in any tissue (**Table S2**). Through GTEx and literature review, we identified 15 plausible eQTL-protein-coding eGene pairs for COLOC analysis (**Table 2**).

A previous study informed that rs35705950 is within an enhancer region of *MUC5B* that was differentially methylated and bound by the transcription factor FOXA2.[52] Subsequent colocalization analyses with eQTL and mQTL data from control and IPF lung tissue revealed that the T allele of rs35705950 was also associated with higher methylation within a repressor region of *MUC5B* and higher *MUC5B*

19

expression.[28] Similarly, we found that *MUC5B* was the only eGene for rs35705950 in the lung. Colocalization of the disease GWAS and *MUC5B*-eQTL signals resulted in a PP4 of 1.00, indicating that rs35705950 was the causal SNP driving both the GWAS and eQTL signals (**Figure 3A-B**).

An LD proxy of rs12585036, rs9577395 (r^2=0.99) was a conditionally independent eQTL for *ATP11A* in tissues of unclear relevance to IPF and COVID-19 (aorta, skin, small intestine). However, in tissues relevant to IPF and COVID-19 (EBV-transformed lymphocytes, cultured fibroblasts, lung, and whole blood), we found modest colocalization of the GWAS and *ATP11A*-eQTL signals in whole blood **(**PP4 = 0.621 with IPF GWAS; 0.749 with COVID-19 GWAS**; Figure 3C-D)**. In the remaining three tissues, we did not find colocalization of the GWAS and *ATP11A*-eQTL signals (PP4 < 0.408; **Table S8**), and the lead eQTLs were in weak LD with rs12585036 (r^2 < 0.20).

An LD proxy of rs12610495, rs2277732 (r^2 = 0.95) was a conditionally independent eQTL for *DPP9* in fibroblasts. Accordingly, the disease GWAS and *DPP9*-eQTL signals colocalized in fibroblasts (PP4 = 0.989), highlighting that this variant may affect *DPP9* expression in a cell-type specific manner (**Figure 3E-F**). In GTEx, the G allele was associated with lower *DPP9* expression in lungs and fibroblasts and was also the risk allele for both critically ill COVID-19 and IPF. In the remaining three tissues, we did not find colocalization of the GWAS and eQTL signals (PP4 < 0.225; **Table S8**), and the lead eQTLs were in weak LD with rs12610495 (r^2 < 0.20).

LD proxies of rs2897075, rs73158411 and rs13243708 were conditionally independent eQTLs in esophageal and breast tissue for *COPS6* and *ZKSCAN1,* respectively. However, we found no colocalization between the disease GWAS and
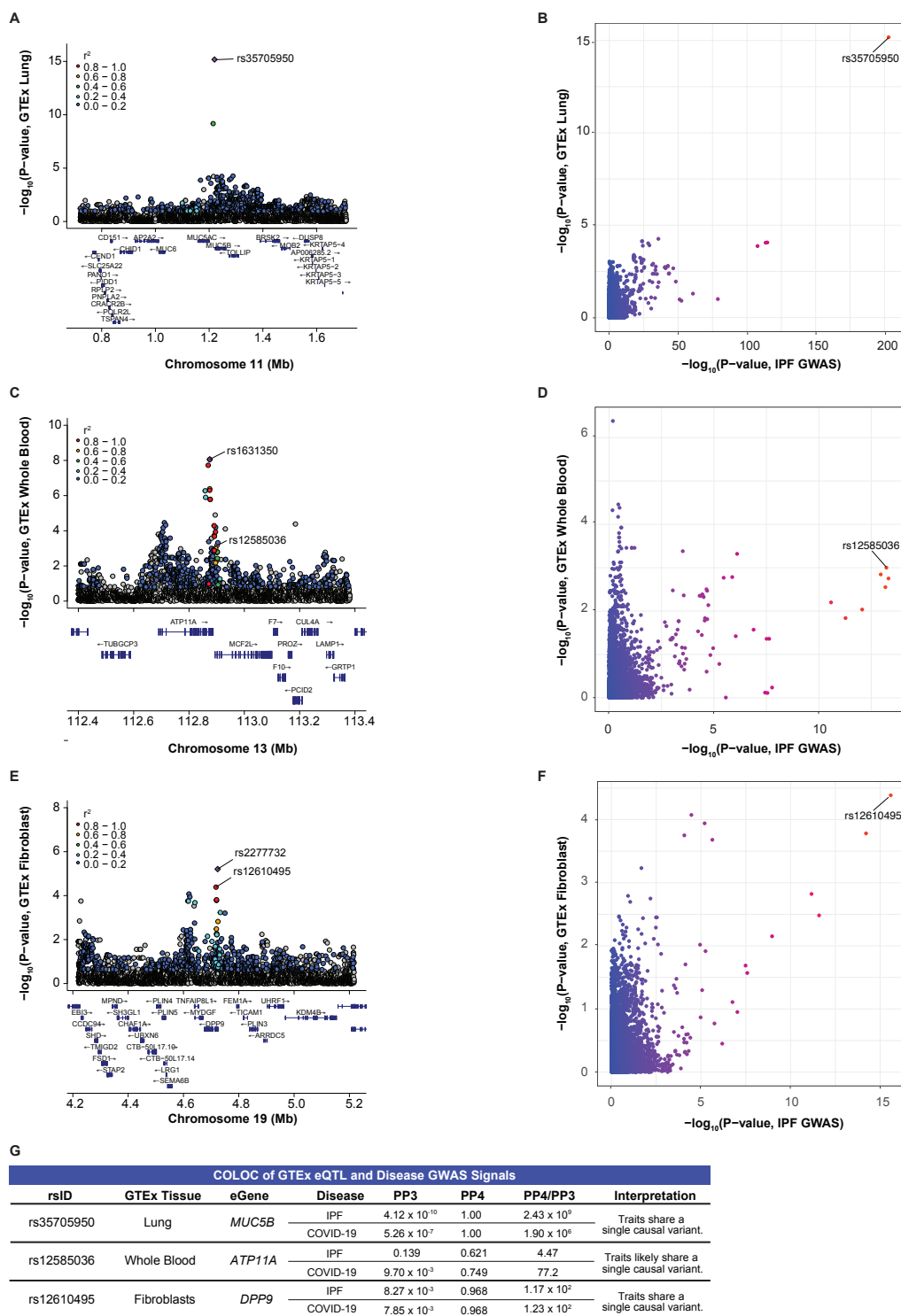
eQTL signals for *COPS6, TRIM4,* or *ZKSCAN1* in the tissues of interest (PP4 < 0.312;

Table S8). The lead eQTLs for *COPS6, TRIM4,* or *ZKSCAN1* were in weak with

rs2897075 (r^2 < 0.20).

For rs1105569, we tested 9 candidate protein-coding eGenes (**Table 2**) but did

not find clear evidence of colocalization for any eGene in the tissues of interest, which

was likely due to the extensive LD in the region (**Table S8**). Thus, out of five shared loci,

baseline bulk eQTL data in the most relevant healthy tissue (i.e. lung) only revealed a

clear causal gene (*MUC5B*) for rs35705950 (**Figure 3A-B**). Our focus on the shared

genetic variants between critically ill COVID-19 and IPF prompted us to broaden our

search to other relevant tissues, revealing that expression of *ATP11A* in whole blood

and *DPP9* in fibroblasts may be relevant for the effect of rs12585036 and rs12610495,

respectively, during pathogenesis (**Figure 3G**). Our results highlight a mystery in the

field of human genetics – lack of colocalization between GWAS and eQTL signals.[8,53,54]

To address this, we next examined context-specific eQTL datasets.

Figure 3: Colocalization of Disease GWAS and GTEx eQTL signals.

(A) LocusZoom plot shows rs35705950 is the lead eQTL (purple diamond) for *MUC5B* in GTEx lung tissue.

(B) Comparison of -log$_{10}$(p-values) from the IPF GWAS and *MUC5B*-eQTLs in GTEx lung tissue shows rs35705950 as the shared lead SNP.

(C) rs12585036 is not the lead eQTL nor in LD with the lead eQTL (rs1631350) for *ATP11A* in GTEx whole blood.

(D) Comparison of -log$_{10}$(p-values) from the IPF GWAS and *ATP11A*-eQTLs in GTEx lung tissue shows rs12585036 as the shared lead SNP.

(E) rs12610495 is in strong LD with the lead eQTL (rs2277732) for *DPP9* in GTEx fibroblasts.

(F) Comparison of -log$_{10}$(p-values) from the IPF GWAS and *DPP9*-eQTLs in GTEx fibroblasts show rs12610495 as the shared lead SNP.

(G) COLOC indicates strong colocalization between the disease GWAS and *MUC5B- and DPP9*-eQTL signals in lung and fibroblasts, respectively, with PP4 > 0.900 and PP4/PP3 > 5.00. COLOC suggests colocalization between the COVID-19 GWAS and *ATP11A*-eQTL signals in whole blood with PP4 > 0.700 and PP4/PP3 > 5.00.

P-values are -log$_{10}$ transformed. Linkage disequilibrium information for European populations was obtained from LDlink and is relative to the lead variant in each plot.

G

COLOC of GTEx eQTL and Disease GWAS Signals							
rsID	**GTEx Tissue**	**eGene**	**Disease**	**PP3**	**PP4**	**PP4/PP3**	**Interpretation**
rs35705950	Lung	*MUC5B*	IPF	4.12×10^{-10}	1.00	2.43×10^{9}	Traits share a single causal variant.
			COVID-19	5.26×10^{-7}	1.00	1.90×10^{6}	
rs12585036	Whole Blood	*ATP11A*	IPF	0.139	0.621	4.47	Traits likely share a single causal variant.
			COVID-19	9.70×10^{-3}	0.749	77.2	
rs12610495	Fibroblasts	*DPP9*	IPF	8.27×10^{-3}	0.968	1.17×10^{2}	Traits share a single causal variant.
			COVID-19	7.85×10^{-3}	0.968	1.23×10^{2}	

23

rs12585036 is an eQTL for ATP11A in lung tissue from a cohort of primarily ex- and current smokers

The Lung eQTL Study utilized lung samples from 1,111 patients, and most patients (85%) were either ex-smokers or current smokers (compared to 68% of all tissue donors in GTEx).[55-57] We were interested in the Lung eQTL Study due to its larger sample size compared to GTEx and because smoking is a risk factor for ILDs including IPF.[58-60] We hypothesized that this dataset might elucidate additional eQTL signals that were not detected in GTEx and better colocalize with the IPF and COVID-19 GWAS signals. Indeed, we found that a lead eQTL for *ATP11A* in the Lung eQTL Study, rs7998551 (p=7.68 x 10^{-8}), was in strong LD with rs12585036 (r^2 = 0.96) (**Figure 4A**). We subsequently found evidence of colocalization between the disease GWAS and the *ATP11A*-eQTL signals (PP4 = 0.996 with IPF GWAS; 0.995 with COVID-19 GWAS; **Figure 4B-C; Table S9**) but lack of colocalization between the *ATP11A*-eQTL signals from the Lung eQTL Study and GTEx lung (PP4 = 0.344; **Table S9**). This demonstrated that lungs exposed to a particular environmental stressor, such as smoking, can reveal eQTLs and causal genes that would otherwise be overlooked. The C allele of rs12585036 was associated with higher *ATP11A* expression and was the risk allele in IPF but the protective allele for critically ill COVID-19. Thus, *ATP11A* appeared to have opposite effects on the pathophysiology of these two diseases.

We found no colocalization at the *MUC5B* locus since the SNP array used in this study (Illumina Human1M-Duo BeadChip) did not include rs35705950, and there are no additional SNPs in LD with this SNP. We did not find colocalization at our other variants of interest.

Monocytes are a critical cell type for the eQTL effect of rs12585036

The colocalization of rs12585036 with *ATP11A* expression in lung tissue primarily from smokers revealed a plausible environmental context where an eQTL effect is revealed. However, as these data were from bulk lung tissue, it was unclear what specific cell type drove this difference. Smoking causes inflammation and infiltration of immune cells, specifically monocytes and neutrophils, into the lungs.[61-63] Therefore, we hypothesized the eQTL effect may be associated with a specific immune cell, which was also consistent with the modest colocalization we detected from bulk whole blood in GTEx (**Figure 3C-D**). The BLUEPRINT project[64] allowed us to interrogate *ATP11A* eQTLs in monocytes, neutrophils, and CD4+ T cells. rs12585036 was an eQTL only in monocytes (p-value = 3.58 x 10^{-12} in monocytes vs. 1.38 x 10^{-3} in neutrophils and 3.89 x 10^{-2} in T cells; **Figure 4D**) and demonstrated strong colocalization with the disease GWAS (PP4 = 0.994; **Figure 4C-E; Table S9**). Directionality of the signals from the BLUEPRINT project and Lung eQTL Study were consistent with the C allele of rs12585036 being associated with higher *ATP11A* expression. Thus, our analyses used two context-specific eQTL datasets to reveal that *ATP11A* was the causal gene associated with the IPF and COVID-19 GWAS signals at rs12585036. Further, the eQTL signal appeared to be important specifically in monocytes, possibly within the proinflammatory context of smoking.

Figure 4: Colocalization at rs12585036 in a proinflammatory context and cell-type specific manner reveals *ATP11A* as a causal gene.

(A) LocusZoom plot shows that rs12585036 is a top eQTL and in strong LD with the lead eQTL (rs7998551; purple diamond) for *ATP11A* in the Lung eQTL Study, which included primarily individuals with a smoking history.

(B) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and *ATP11A*-eQTLs in the Lung eQTL Study shows rs12585036 as a top shared SNP.

(C) COLOC indicates strong colocalization between the disease GWAS and the Lung eQTL Study's *ATP11A*-eQTL signals with PP4 > 0.900, and PP4/PP3 > 5.00.

(D) rs12585036 is a top eQTL and in strong LD with the lead eQTL (rs9577175) for *ATP11A* in monocytes in the BLUEPRINT project.

(E) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and monocytic *ATP11A*-eQTLs shows rs12585036 as a top shared SNP.

(F) COLOC indicates strong colocalization between the disease GWAS and monocytic *ATP11A*-eQTLs signals with PP4 > 0.900 and PP4/PP3 > 5.00.


P-values are -$\log_{10}$ transformed. Linkage disequilibrium information for European populations was obtained from LDlink and is relative to the lead variant in each plot.

A

B
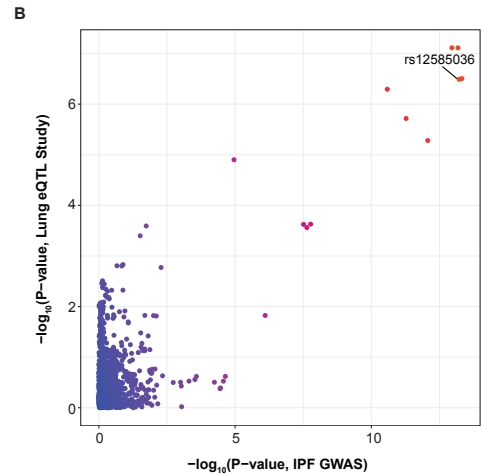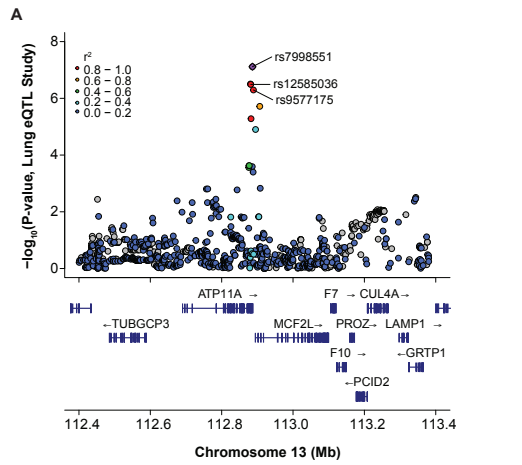
C

COLOC of the Lung eQTL Study and Disease GWAS Signals at *ATP11A*				
Disease	PP3	PP4	PP4/PP3	Interpretation
IPF	4.17×10^{-3}	0.996	2.39×10^{2}	Traits share a single causal variant.
COVID-19	4.88×10^{-3}	0.995	2.04×10^{2}	Traits share a single causal variant.

D

E
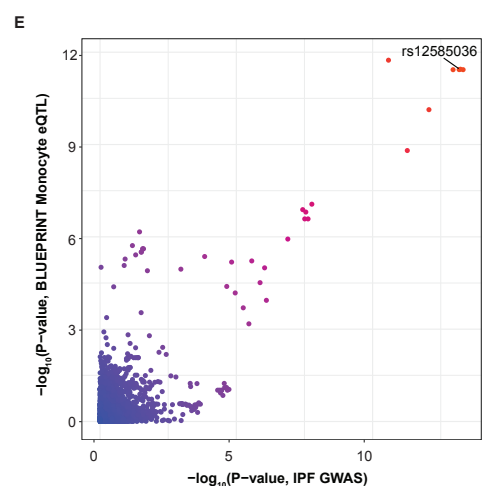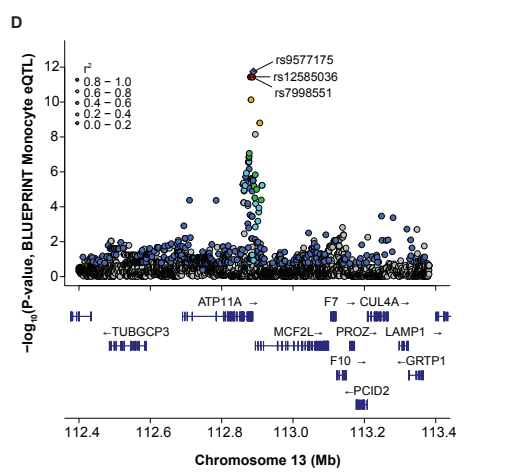
F

COLOC of the BLUEPRINT Monocyte eQTL and Disease GWAS Signals at *ATP11A*				
Disease	PP3	PP4	PP4/PP3	Interpretation
IPF	5.78×10^{-3}	0.994	1.72×10^{2}	Traits share a single causal variant.
COVID-19	5.60×10^{-3}	0.994	1.76×10^{2}	Traits share a single causal variant.

27

rs12610495 is an eQTL for DPP9 in IPF lung

We sought to determine if there was colocalization with eQTLs specifically identified in lungs from individuals with IPF. We utilized the Borie et al. eQTL dataset which included lung samples from healthy controls (n = 188) and individuals with IPF who also reported a more extensive smoking history (n = 234).[28] Borie et al. previously demonstrated colocalization of the eQTL, mQTL, and GWAS signals at rs35705950 within the *MUC5B* locus in both controls and IPF cases.[28]

We did not find evidence of colocalization between disease GWAS and control/IPF eQTL signals at rs12585036, rs2897075, and rs1105569. However, we found colocalization at rs12610495 for *DPP9* expression in IPF cases. Although the eQTL signals were in similar genomic locations for both controls and IPF cases, fine mapping further revealed that the lead variants between the two groups were distinct signals (**Figure 5A-D**). The lead variant in controls, rs758510 (p-value = 2.91×10^{-6} from FastQTL following PEER normalization), and the lead variant in the IPF cases, rs12462642 (6.65×10^{-7}), exhibited weak LD ($r^2 = 0.15$). In contrast, rs12462642 and rs12610495 (the lead GWAS variant; 6.52×10^{-5}) were in stronger LD ($r^2 = 0.68$). COLOC demonstrated moderately strong colocalization between the disease GWAS and IPF cases *DPP9*-eQTL signals (PP4 = 0.874 with IPF GWAS; 0.873 with COVID-19 GWAS) and weaker colocalization between the GWAS and control eQTL signals (PP4 = 0.661; 0.640; **Figure 5E-F; Table S9**). Plotting *DPP9* expression by rs12610495 genotype confirmed a more robust association of this lead GWAS SNP with expression in IPF cases (slope = -0.212; p-value = 1.96×10^{-2} using linear regression of pre-PEER normalized values) compared to controls (slope = -0.074; p-value = 0.526) and was

28

consistent with the G allele being associated with lower *DPP9* expression (**Figure 5G-H**). Thus, two independent context-specific eQTL datasets, from fibroblasts and IPF lung, pointed to *DPP9* as the causal gene underlying rs12610495. These discoveries could be related as IPF lungs are known to have high levels of activated, and potentially abnormal, fibroblasts.[65-68]

In summary, we used context-specific eQTL datasets of fibroblasts, monocytes, and lungs affected by smoking and IPF to show that *ATP11A* and *DPP9* were likely causal genes in critically ill COVID-19 and IPF. These results also revealed that *ATP11A* and *DPP9* were likely important in monocytes and fibroblasts, respectively. We also examined recent sc-eQTL datasets from Aquino et al. for COVID-19[69] and Natri et al. for IPF[70] but found no further evidence of colocalization, which could be due to low power (n = 80 Western European individuals in Aquino et al.; n = 48 controls, 66 IPF cases in Natri et al.).

Figure 5: Colocalization at rs12610495 in IPF cases reveals *DPP9* as a causal gene.

(A) LocusZoom plot shows that rs12610495 is not the lead eQTL nor in LD with the lead
eQTL (rs758510; purple diamond) for *DPP9* in controls in the Borie et al. dataset.

(B) rs12610495 is a top eQTL and in LD with the lead eQTL (rs12462642) for *DPP9* in IPF
cases.

(C) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and *DPP9*-eQTLs in controls shows
rs12610495 as a top shared SNP.

(D) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and *DPP9*-eQTLs in IPF cases
shows rs12610495 as the lead shared SNP.

(E) COLOC does not support colocalization between the disease GWAS and *DPP9*-eQTL
signals in controls with PP4 < 0.700.

(F) COLOC indicates colocalization between the disease GWAS and *DPP9*-eQTL signals in
IPF cases with PP4 > 0.700 and PP4/PP3 > 5.00.

(G) Box plot for *DPP9* expression in controls by genotype.

(H) *DPP9* expression in IPF cases by genotype.

P-values in LocusZoom plots are -$\log_{10}$ transformed. Linkage disequilibrium information for

European populations was obtained from LDlink and is relative to the top variant in each plot.

Slopes and p-values for genotype expression boxplots were obtained through linear regression

modeling.
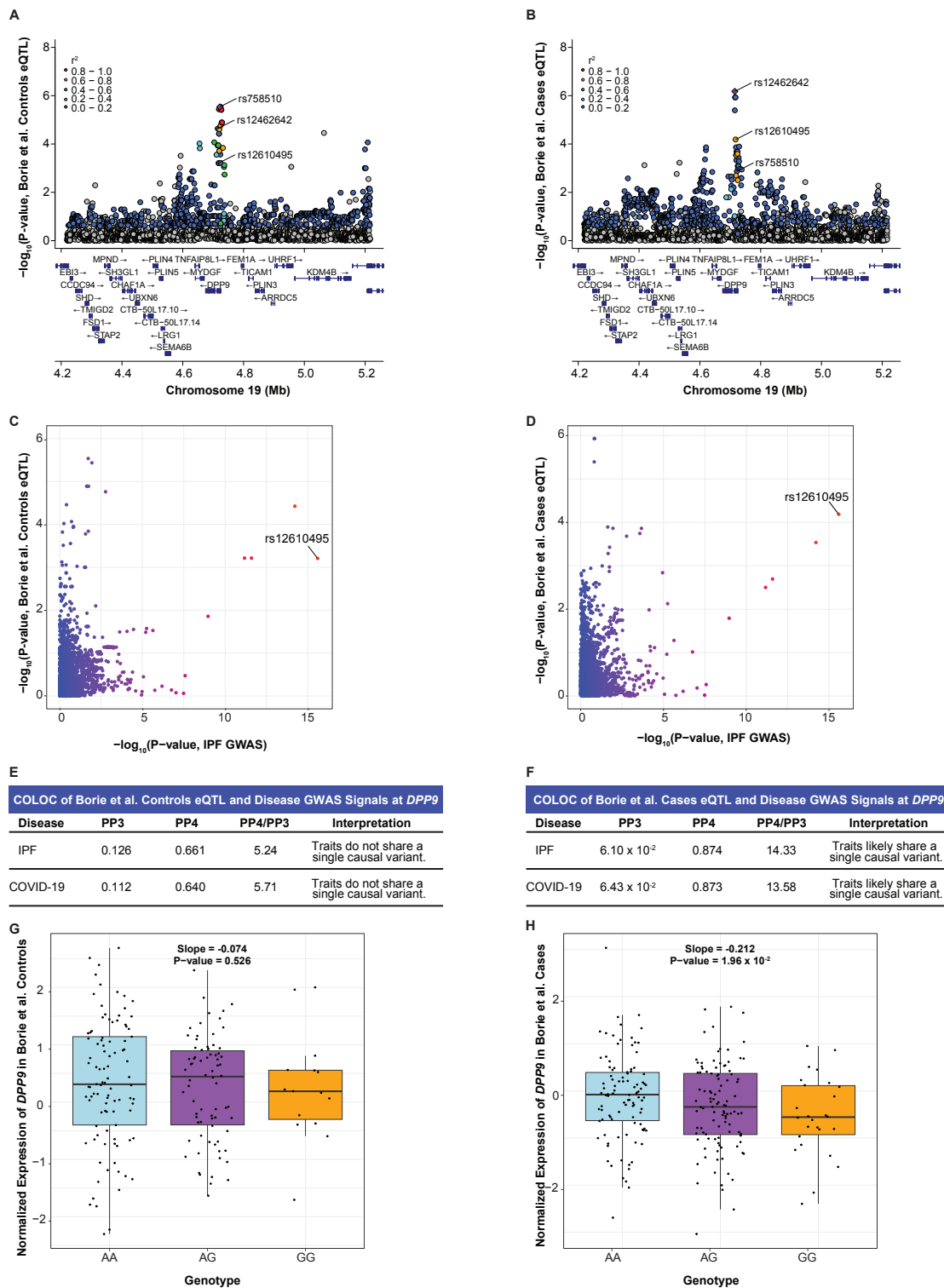
rs12610495 is an mQTL affecting methylation of the DPP9 promoter in IPF lungs

The stronger colocalization of GWAS and eQTL signals at rs12610495 in IPF

cases suggested that there may be underlying gene regulatory differences between

healthy and IPF lungs. Differences in DNA methylation can influence gene expression,

and mQTL may reveal locations in the genome where patterns of DNA methylation are

regulated by genetic variation.[71-73] We utilized the Borie et al. mQTL dataset which

included lung samples from healthy controls (n = 202) and IPF cases (n = 345). We

found that rs12610495 was not in strong LD with rs10420225 (r^2 = 0.25), the lead mQTL

for cg07317664 in controls, but was in strong LD with rs2277732 (r^2 = 0.95), the lead

mQTL for cg07317664 in cases (**Figure 6A-B**). We did not observe colocalization

between the disease GWAS and control cg07317664-mQTL signals (PP4 = 1.33 x 10^{-2}

with IPF GWAS; 9.47 x 10^{-3} with COVID-19 GWAS) (**Figure 6C, 6E**). Like the eQTL

colocalization at rs12610495, we found strong colocalization between the disease

GWAS and IPF cases cg07317664-mQTL signals (PP4 = 0.980 with IPF GWAS; 0.763

with COVID-19 GWAS; **Figure 6D, 6F**). Plotting methylation of cg07317664 by

rs12610495 genotype indicated a more robust association in IPF cases (slope = 1.9 x

10^{-2}; p-value = 3.20 x 10^{-9} by linear regression using pre-PEER methylation values)

compared to controls (slope = 1.2 x 10^{-2}; p-value = 6.99 x 10^{-4}) (**Figure 6G-H**).

Figure 6: Colocalization at rs12610495 reveals a differentially methylated site (cg07317664) near the *DPP9* transcription start site in IPF cases.

(A) LocusZoom plot shows that rs12610495 is not the lead mQTL nor in LD with the lead mQTL (rs10420225; purple diamond) for cg07317664 in controls in the Borie et al. dataset.

(B) rs12610495 is a top mQTL and in LD with the lead mQTL (rs2277732) for cg07317664 in IPF cases.

(C) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and cg07317664-mQTLs in controls shows rs12610495 as a top shared SNP.

(D) Comparison of -$\log_{10}$(p-values) from the IPF GWAS and cg07317664-mQTLs in IPF cases shows rs12610495 as the lead shared SNP.

(E) COLOC does not support colocalization between the disease GWAS and cg07317664-mQTL signals in controls with PP4 < 0.700.

(F) COLOC indicates colocalization between the disease GWAS and cg07317664-mQTL signals in cases with PP4 > 0.700.

(G) Box plot for normalized methylation beta in controls by genotype.

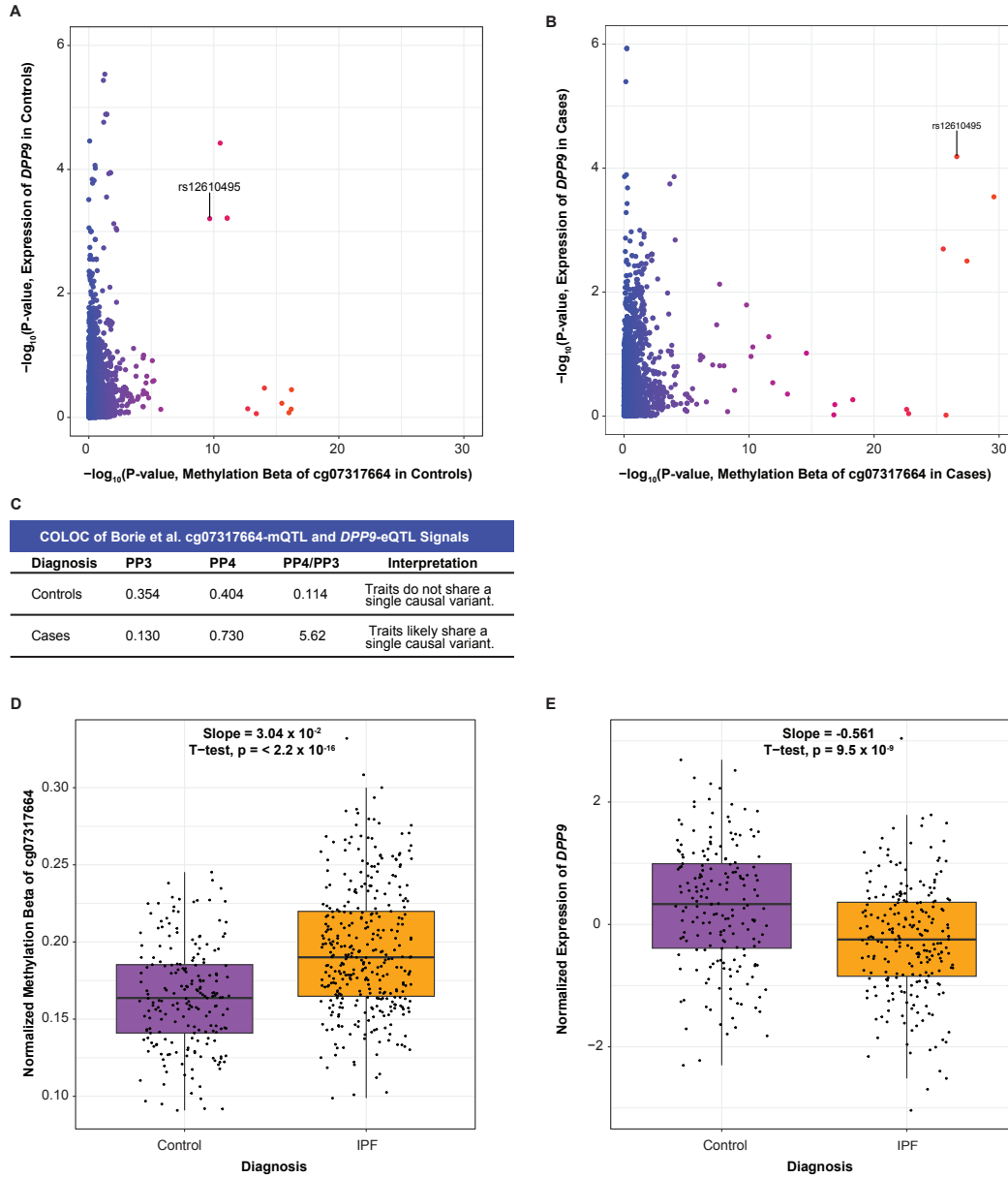(H) Normalized methylation beta in IPF cases by genotype.

P-values in LocusZoom plots are -$\log_{10}$ transformed. Linkage disequilibrium information for European populations was obtained from LDlink and is relative to the top variant in each plot. Slopes and p-values for genotype methylation beta boxplots were obtained through linear regression modeling.
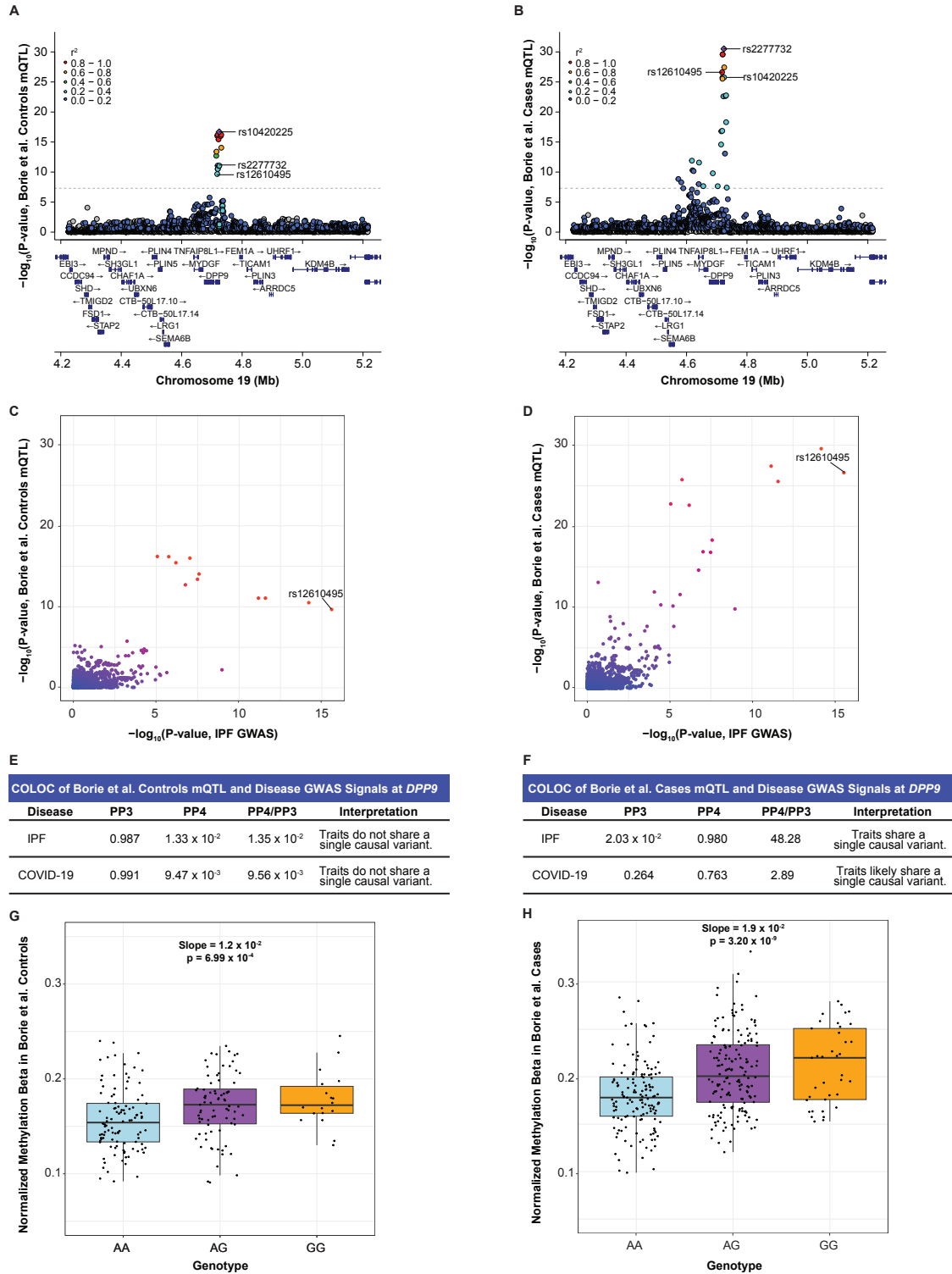
Cg07317664 is 7319 base pairs upstream of the transcription start site of *DPP9,*
leading us to hypothesize that methylation at this location may regulate *DPP9* gene
expression. This hypothesis was consistent with the genetic and eQTL data as the G
allele of rs12610495 was associated with greater methylation at a regulatory site of
DPP9 and decreased expression of *DPP9*, while the A allele was associated with less
methylation and increased expression. Comparing the mQTL and eQTL signals in
controls and cases separately demonstrated that colocalization occurred only in cases
(PP4 = 0.404 in controls; 0.730 in cases; **Figure 7A-C**;). Further, we found that IPF
cases, regardless of rs12610495 genotype, had significantly higher methylation levels at
cg07317664 (slope = 0.0304, p-value < 2.2×10^{-16}) and significantly lower expression of
DPP9 (slope = -0.561, p-value = 9.5×10^{-9}; **Fig 7D-E)**, suggesting that genotype was
just one factor contributing to reduced expression in IPF and underscoring the
importance of considering epigenetic regulation.

Figure 7: Colocalization at rs12610495 confirms the association between the cg07317664-mQTL and *DPP9*-eQTL signals in IPF cases.

(A) Comparison of -log$_{10}$(p-values) of cg07317664-mQTLs and *DPP9*-eQTLs in controls in the Borie et al. datasets shows that rs12610495 in not the lead SNP for either dataset.

(B) Comparison of -log$_{10}$(p-values) in IPF cases shows rs12610495 as a top shared SNP.

(C) COLOC does not support colocalization between cg07317664-mQTL and *DPP9*-eQTL signals in controls with PP4 < 0.700 but supports colocalization in IPF cases with PP4 > 0.700.

(D) Box plot for *DPP9* expression in lung by diagnosis.

(E) Box plot for normalized methylation beta in lung by diagnosis.

C

COLOC of Borie et al. cg07317664-mQTL and *DPP9*-eQTL Signals				
Diagnosis	PP3	PP4	PP4/PP3	Interpretation
Controls	0.354	0.404	0.114	Traits do not share a single causal variant.
Cases	0.130	0.730	5.62	Traits likely share a single causal variant.

Causal genes in bulk diseased tissue

The putative causal genes we identified also demonstrated altered expression in IPF and COVID-19. Consistent with the genetic associations and as previously reported, we found that *MUC5B* expression was significantly increased in lung tissue from individuals with IPF when compared to tissue from controls in two independent studies (GSE213001 and GSE34692) (**Figure 8A-B**). In contrast, *ATP11A* and *DPP9* expression were significantly decreased in IPF lung tissue when compared to control tissue (**Figure 8A-B)**. Lower *DPP9* expression in IPF lung tissue was concordant with the disease GWAS and eQTL association as the G allele of rs12610495 was associated with lower *DPP9* and increased risk of IPF. In contrast, the directionality of the observation for *ATP11A* was discordant with the GWAS and eQTL association, which suggested the C allele of rs12585036 was associated with higher *ATP11A* and increased risk of IPF. However, two reasons may reconcile this observation for *ATP11A*. Firstly, *ATP11A* may be higher during the onset of disease and decrease in later stages, which was when most samples in these studies were collected. Later changes in *ATP11A* levels may reflect an effect of the disease on gene expression. Secondly, a single cell type may be underlying the genetic signal which is confounded when looking at bulk tissue and may also be depleted during later stages of disease.

We did not observe significant differences in *MUC5B* or *DPP9* expression in lungs from patients who succumbed to COVID-19 when compared to control lungs, which could be due to the limited number of samples in the study we examined (GSE159585; **Figure 8C**). We found significantly lower *ATP11A* expression in COVID-19 lungs when compared to control lungs which was concordant with the GWAS and
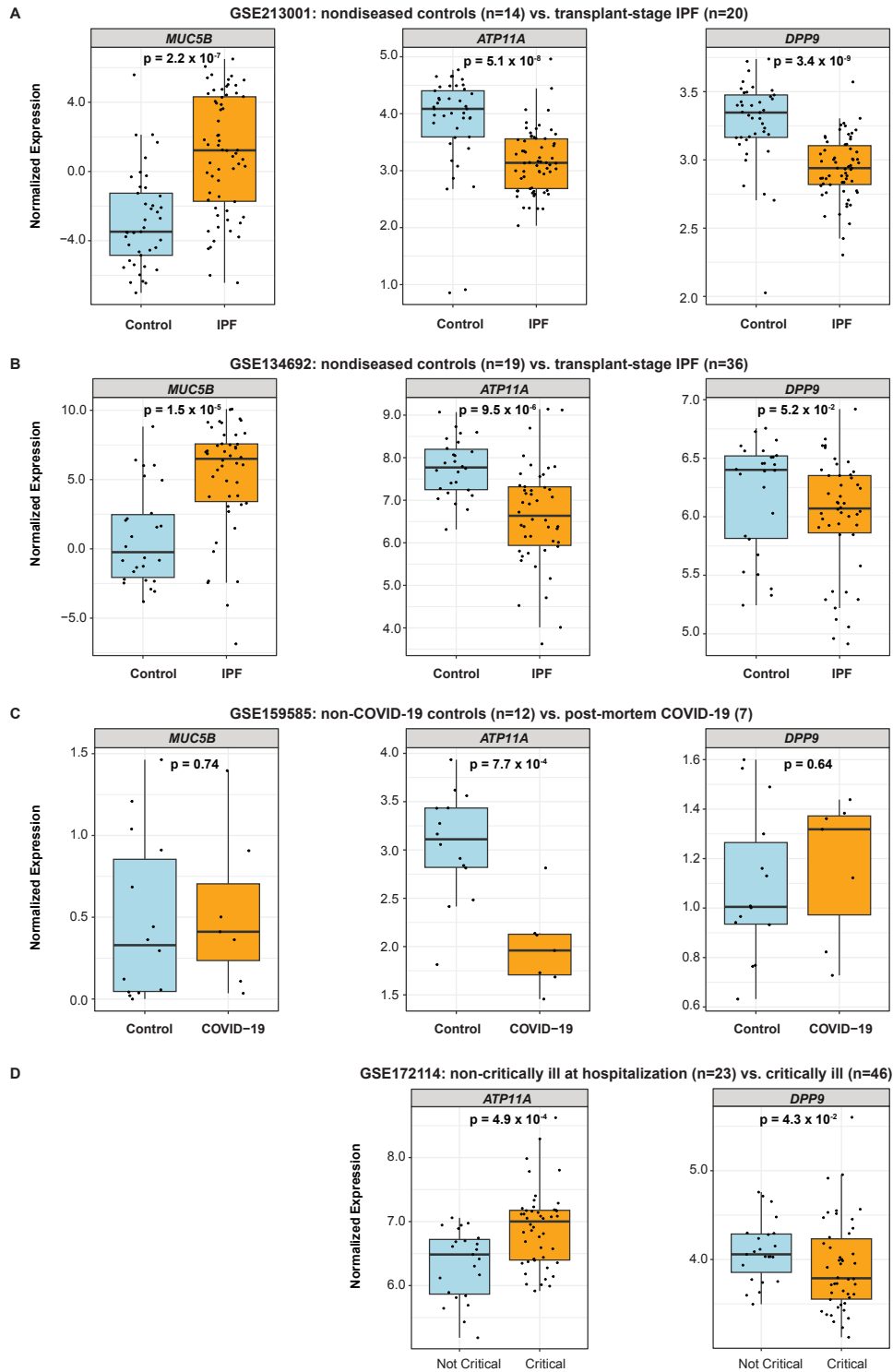
38

eQTL association as the T allele of rs12585036 was associated with low *ATP11A* and increased risk of COVID-19 (**Figure 8C**). However, in another study which measured gene expression in whole blood (GSE172114; **Figure 8D**), we found that *ATP11A* expression was higher in patients diagnosed with critically ill COVID-19 when compared to those with non-critically ill COVID-19, which could be due to an increase in myeloid lineage cells. As observed in the other datasets, *DPP9* is significantly lower in the blood of critically ill patients compared to the non-critically ill.

Altogether, these findings support the relevance of *ATP11A* and *DPP9* in IPF and COVID-19 pathogenesis, but directionality of how associated SNPs impact gene expression may differ in bulk diseased tissues possibly due to the impact of reverse causation of disease on gene expression, changes in cell type proportions, and/or the lack of temporal and spatial resolution. We therefore examined scRNA-Seq datasets.

Figure 8: Assessment of causal genes in bulk control and diseased tissue from

independent studies. Box plots depict normalized expression of causal genes by diagnosis

from:

(A) GSE213001, which included lung tissue from non-diseased controls (n=14) and IPF

patients undergoing lung transplant (n=20).

(B) GSE134692, which included lung tissue from non-diseased controls (n=19) and IPF

patients undergoing lung transplant (n=36).

(C) GSE159585, which included lung tissue from non-diseased controls (n=12) and from

patients who succumbed to COVID-19 (n=7).

(D) GSE172114, which included whole blood samples from hospitalized patients with non-

critically ill COVID-19 (n=23) and patients in the intensive care unit with critically ill

COVID-19 (n=46).

P-values were calculated using the Mann-Whitney U test.

A GSE213001: nondiseased controls (n=14) vs. transplant-stage IPF (n=20)

B GSE134692: nondiseased controls (n=19) vs. transplant-stage IPF (n=36)

C GSE159585: non-COVID-19 controls (n=12) vs. post-mortem COVID-19 (7)

D GSE172114: non-critically ill at hospitalization (n=23) vs. critically ill (n=46)

41

Causal genes in single cell types from diseased tissue

Using a scRNA-Seq dataset for IPF[38] and snRNA-Seq dataset for lethal COVID-19[39], we compared where the identified causal genes were expressed in healthy and diseased lung tissue. As differences in the expression of our identified causal genes could be due to either changes in cell type number or expression levels within a cell type, we first determined how cellular proportions differed between healthy and IPF or COVID-19 lungs. IPF lungs displayed decreases in type 1 and type 2 alveolar pneumocytes (AT1s and AT2s), and increases in other epithelial cell types (basal, ciliated, club, and goblet cells), myofibroblasts, peribronchial vascular endothelial cells, consistent with fibrosis and reduced healthy lung tissue (**Table S4)**. COVID-19 lungs displayed a significant increase only in macrophages but also notable increases in monocytes and fibroblasts, and decreases in AT1s, AT2s, and other airway epithelial cells (**Table S5**).

We subsequently determined what cell types expressed our causal genes. As expected, *MUC5B* was highly expressed in goblet cells and modestly in AT2s, other airway epithelial cells, and macrophages (**Table S6-7**). *ATP11A* was notably expressed in AT1s, AT2s, and macrophages and modestly expressed in fibroblasts, endothelial cells, monocytes, and other airway epithelial cells (**Table S6-7**). The expression in macrophages and monocytes was particularly intriguing given the colocalization of monocyte eQTL and GWAS signals at the *ATP11A* locus. *DPP9* was notably expressed in AT1s, AT2s, macrophages, and fibroblasts (**Table S6-7**).

Next, we examined differences in causal gene expression during IPF and COVID-19 in single cell types. Concordant with the GWAS and eQTL directionality, we

42

found that *MUC5B* was significantly increased during IPF in goblet cells, AT2s, other airway epithelial cells, and macrophages (**Figure 9A**). In COVID-19, we found decreased, although statistically nonsignificant, *MUC5B* in airway epithelial cells, which includes goblet cells (**Figure 9B**), consistent with our genetic associations having opposite effects in these two diseases.

We hypothesized that *ATP11A* would be increased in epithelial cells and/or myeloid lineage cells based on the GWAS and eQTL colocalization in lung tissue primarily from smokers and in monocytes. However, we found significantly lower expression in AT2s (**Figure 9C; Table S6**), which may be due to the depletion of epithelial cells during IPF. Consistent with our hypothesis, *ATP11A* was significantly increased during IPF in macrophages, a large proportion of which are derived from monocytes recruited to the lung during IPF,[74] as well as in other airway epithelial cells and myofibroblasts, which are a key driver of fibrosis (**Figure 9C; Table S6**).[65] During COVID-19, *ATP11A* was significantly decreased in AT1s, AT2s, other airway epithelial cells, concordant with the GWAS and eQTL data (**Figure 9D; Table S7)**.

We hypothesized that *DPP9* would be decreased during IPF and COVID-19 in fibroblasts and other cell types that increase in proportion during IPF based on our GWAS and eQTL colocalization in fibroblasts and IPF lung. *DPP9* was significantly decreased in AT2s and monocytes during IPF (**Figure 9E; Table S6**). However, we found that *DPP9* was significantly increased in myofibroblast during IPF, which may be due to the hyperproliferation of this cell type during IPF (**Figure 9E; Table S6)**. During COVID-19, we found that *DPP9* was significantly decreased in AT1s and AT2s (**Figure 9F; Table S7)**.

43

Figure 9: Assessment of causal genes at the single-cell resolution in control and diseased tissue. Box plots depict normalized expression of (A) *MUC5B*, (C) *ATP11A*, and (E) *DPP9* in an IPF single-cell RNA sequencing study by Adams et al. which included controls (n=28) and IPF patients (n=32). Box plots depict normalized expression of (B) *MUC5B*, (D) *ATP11A*, and (F) *DPP9* in a lethal COVID-19 single-nucleus RNA sequencing study by Melms et al. which included controls (n=7) and COVID-19 patients (n=19).

P-values were calculated using Wald test with P values adjusted for multiple comparisons using the Benjamini-Hochberg method (*Padj < .05; **Padj < .001; ***Padj < .0001).
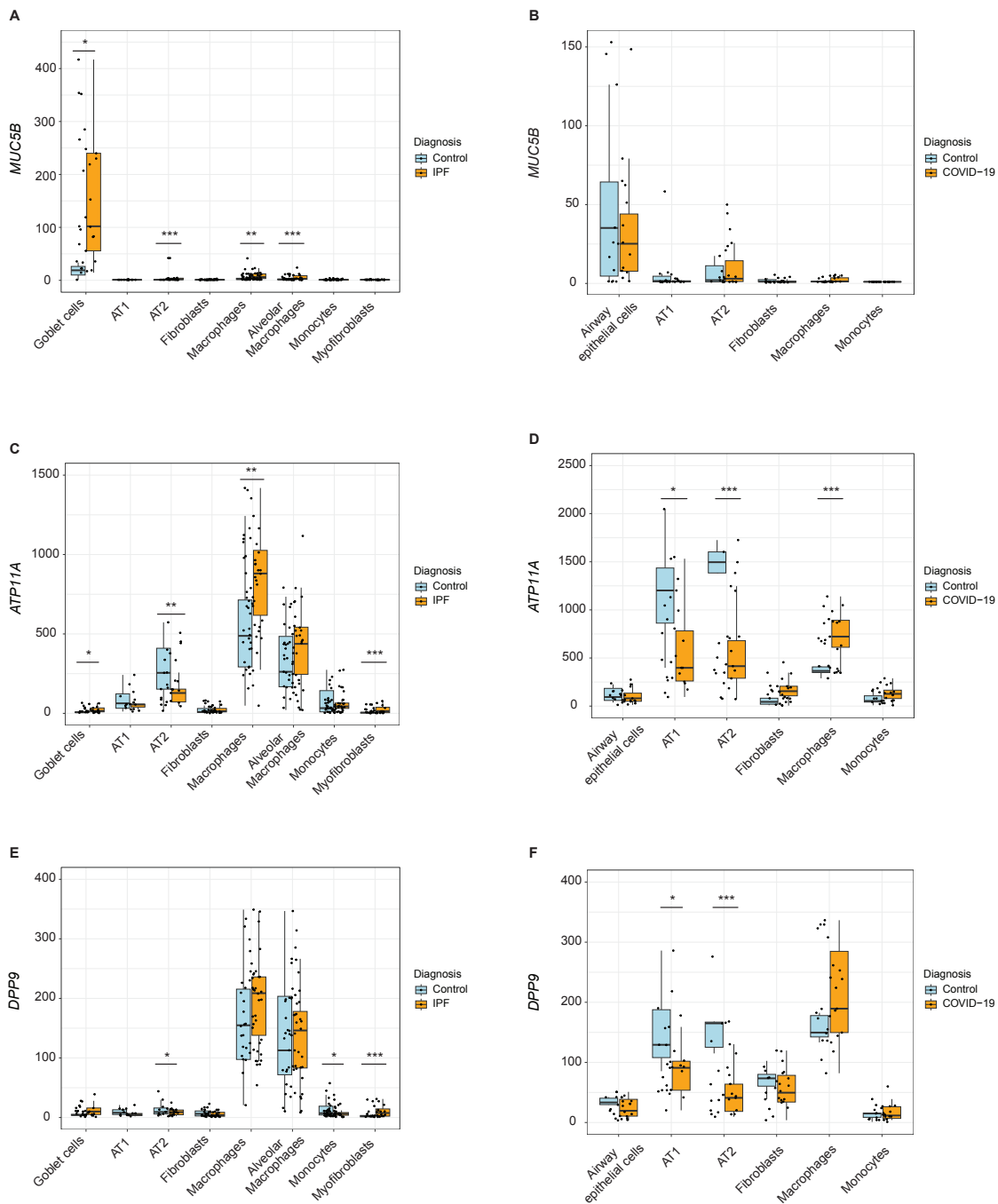
DISCUSSION

GWAS provide a powerful approach to gather human genetic evidence to increase success in drug development, but such efforts require the identification of the underlying causal genes.[1] Molecular QTL studies were designed to bridge the gap between GWAS variants and causal genes, revealing molecular effects on gene expression, splicing, and methylation. However, studies have found limited colocalization between eQTL and human disease GWAS signals. This has been speculated to be due to multiple factors including lack of power in detecting GWAS and eQTL signals; inappropriate context with regards to cell type, developmental stage, and response; nonlinear relationships between eQTL effect and phenotype; overlap of eQTL signals due to multiple signals; and strong LD between variants limiting deconvolution of causality (e.g. rs1105569 within the 17q21.31 inversion supergene). Here, we found evidence that consideration of cell type and disease can improve colocalization of GWAS and QTL signals to reveal causal genes and where they may function during pathogenesis. Whereas eQTLs from healthy lung revealed *MUC5B* as the causal gene underlying the association of rs35705950 with critically ill COVID-19 and IPF, *MUC5B* may be atypical in that its expression is largely restricted to mucus-secreting cells and the effect size of the eQTL is large,[28] making accurate mapping of this eQTL feasible with smaller numbers and in bulk tissue. Examination of eQTL data from isolated cell types and lung disease states was necessary to reveal *DPP9* and *ATP11A* as two additional causal genes. Thus, well-powered tissue- and disease-specific eQTL datasets will continue to reveal additional causal genes for human diseases.

For each of the five shared loci between critically ill COVID-19 and IPF, identification of the likely causal gene leads to testable hypotheses for how each gene impacts these two diseases. This has been demonstrated for the *MUC5B* locus in IPF. The variant rs35705950 is within an enhancer that is differentially regulated and binds the transcription factor FOXA2 to regulate *MUC5B* expression.[52] Additional colocalization studies at rs35705950 between IPF GWAS, control and IPF eQTL, and mQTL signals suggested that higher methylation at a repressor region further contributes to increased *MUC5B* expression in cases of IPF.[28] *In vivo* studies have functionally demonstrated that excessive MUC5B production impairs mucociliary clearance and increases the fibroproliferative response, while reducing mucin production may ameliorate fibrosis.[75] Though increased mucus plugging was observed in COVID-19 lung autopsies,[76] the same allele associated with increased *MUC5B* is protective against severe COVID-19, suggesting increased mucus production in this acute context is protective. The opposite risk alleles of *MUC5B* (and *ATP11A*) for IPF and critically ill COVID-19 also underscores that though these two diseases share five of the same genetic determinants, the underlying pathophysiology is unique—critically ill COVID-19 and PCPF is not simply IPF due to a known etiology. For the remaining four loci overlapping between COVID-19 and IPF, identification of causal genes has now begun to similarly elucidate mechanisms of diseases.

The causal gene associated with rs2897075 has been difficult to identify due to lack of colocalization between GWAS and eQTL signals. We speculate that the most likely causal genes are *TRIM4* and *ZKSCAN1.* rs2897075 is an eQTL for *TRIM4* in several tissues, including the lung, and decreased expression is associated with the risk

47

allele of critically ill COVID-19 and IPF. TRIM4 has been reported to polyubiquitinate

viral pattern recognition receptors including RIG-I and MDA5 and promoting type I

interferon production during Sendai virus[77] and SARS-CoV-2[78] infections, respectively.

rs2897075 is also an eQTL for *ZKSCAN1* in fibroblasts, and decreased expression is

associated with the risk allele of IPF and critically ill COVID-19. Interestingly, *ZKSCAN1*

mRNA can be backspliced into a circular RNA (circRNA), a type of noncoding RNA that

can act as sponges for microRNA and proteins and may contribute to fibrosis.[79,80]

Whether rs2897075 is associated with levels of *ZKSCAN1* circRNA is unknown, though

both *ZKSCAN1* and *circZKSCAN1* have been implicated in liver fibrosis.[81] Although

several studies have reported rs2897075 as a genome-wide significant hit in IPF

GWAS, its mechanism and phenotypic effect has not been reported to our knowledge.

Importantly, we noted that although rs2897075 was not an eQTL in the Borie et al.

eQTL dataset, a variant in strong LD rs6963345 (r^2 = 1.00) was associated with

numerous mQTLs, suggesting that this SNP may influence DNA methylation.[28]

The directionality of the rs12585036 alleles indicates that increased *ATP11A*

expression is associated with increased risk of IPF while decreased expression is

associated with increased risk of critically ill COVID-19. Our colocalization data suggest

that the association at the *ATP11A* locus may be due to expression in monocytes and

derived cells in the lung. Cycles of alveolar damage, recruitment of monocytes, and

their differentiation into monocyte-derived alveolar macrophages occur repeatedly

during IPF.[74] Specifically, M2 macrophages promote fibrosis, possibly through secretion

of TGF-beta[82] and other profibrotic factors.[74] ATP11A, or ATPase Phospholipid

Transporting 11A, is a member of the P4-ATPase family of lipid flippases, responsible

48

for translocating aminophospholipids on the cellular membrane.[83,84] We propose two

mechanisms that could explain ATP11A's involvement in the pathogenesis of IPF and

COVID-19. Firstly, ATP11A is cleaved during apoptosis and results in

phosphatidylserine's translocation to the outer leaflet of the cell membrane, which is an

"eat me" signal during efferocytosis.[83-86] Increased apoptosis of epithelial cells and

resistance to apoptosis in myofibroblasts are hallmarks of IPF, correlating with aberrant

collagen deposition.[87,88] During IPF, we speculate that increased ATP11A may

contribute to impaired efferocytosis, an observation reported in murine models[89] and

IPF patient samples[90]. Secondly, ATP11A has also been reported to promote

internalization of toll-like receptor 4 (TLR4), which is highly expressed on both alveolar

pneumocytes and myeloid lineage cells, preventing hypersecretion of proinflammatory

cytokines after stimulation with lipopolysaccharide.[91] This may skew macrophages

toward profibrotic M2 polarization. Additionally, TLR4 signaling contributes to proper

AT2 renewal and lung repair.[92] In the case of COVID-19, the ectodomain of the SARS-

CoV-2 spike protein has been consistently reported to stimulate TLR4.[93,94] Without

adequate internalization of TLR4 by ATP11A, downstream immune signaling may be

overactivated, leading to hypersecretion of proinflammatory cytokines. Thus, we

propose that ATP11A may affect IPF and COVID-19 by functioning in both epithelial

and myeloid cells and eliciting impaired cell death and immune pathways (**Figure 10A**).

Additionally, the opposite effects of rs12585036 and rs35705950 during IPF and

COVID-19 highlight the trade-offs of specific genetic variants in the face of different

threats.

Colocalization of disease GWAS and eQTL signals indicates that decreased *DPP9* expression is associated with increased risk of COVID-19 and IPF and that fibroblasts may drive the association at least in IPF lung. DPP9, Dipeptidyl Peptidase 9, is a serine peptidase that is in the same family as DPP4, which is well-known as a target of hypoglycemic drugs and as a receptor for the coronavirus causing Middle Eastern Respiratory Syndrome. DPP9 helps maintain the inactive conformation of the NLRP1 inflammasome, inhibiting activation and downstream pyroptosis.[95-97] In addition to regulating inflammasome activity, DPP9 also regulates survival, proliferation, migration, and adhesion, including in dermal fibroblasts.[98] Interestingly, fibroblast activating protein (FAP) belongs to the same serine protease subfamily as DPP9. However, unlike the other DPP proteins, FAP is largely expressed by cancerous cells, inducing EMT. In IPF, FAP is specifically localized to fibrotic foci and not in normal tissue.[99] DPP9 interacts with FAP directly in oral squamous cell carcinoma, and *DPP9* overexpression can inhibit the EMT activity induced by FAP.[100] We previously reported that based on transcriptomics from patients infected with COVID-19 that *DPP9* decreased concordantly with resolution of infection, and therefore, may be dampening the inflammatory response.[11] We suspect that lower DPP9 may exacerbate pathogenesis of IPF and COVID-19 by preventing excess NLRP1-mediated inflammation and weakening cell-to-cell adhesion, promoting a dysregulated cycle of inflammation and fibrosis (**Figure 10B**).

Figure 10: Hypothesized mechanisms linking disease-associated SNPs to causal genes,

***ATP11A* and *DPP9*.**

(A) The C allele of rs12585036 is associated with increased *ATP11A*, leading to 1) defective

efferocytosis and impaired lung tissue repair and 2) increased internalization of toll-like

receptor 4 and increased M2-macrophage polarization. These mechanisms lead to

increased risk of IPF and protection from COVID-19.

(B) The G allele of rs12610495 is associated with increased *DPP9*, leading to 1) impaired

cell adhesion and migration affecting lung tissue repair, and 2) excessive NLRP1-

inflammasome activity and proinflammatory cytokine production. These mechanisms

lead to increased risk of both IPF and COVID-19.
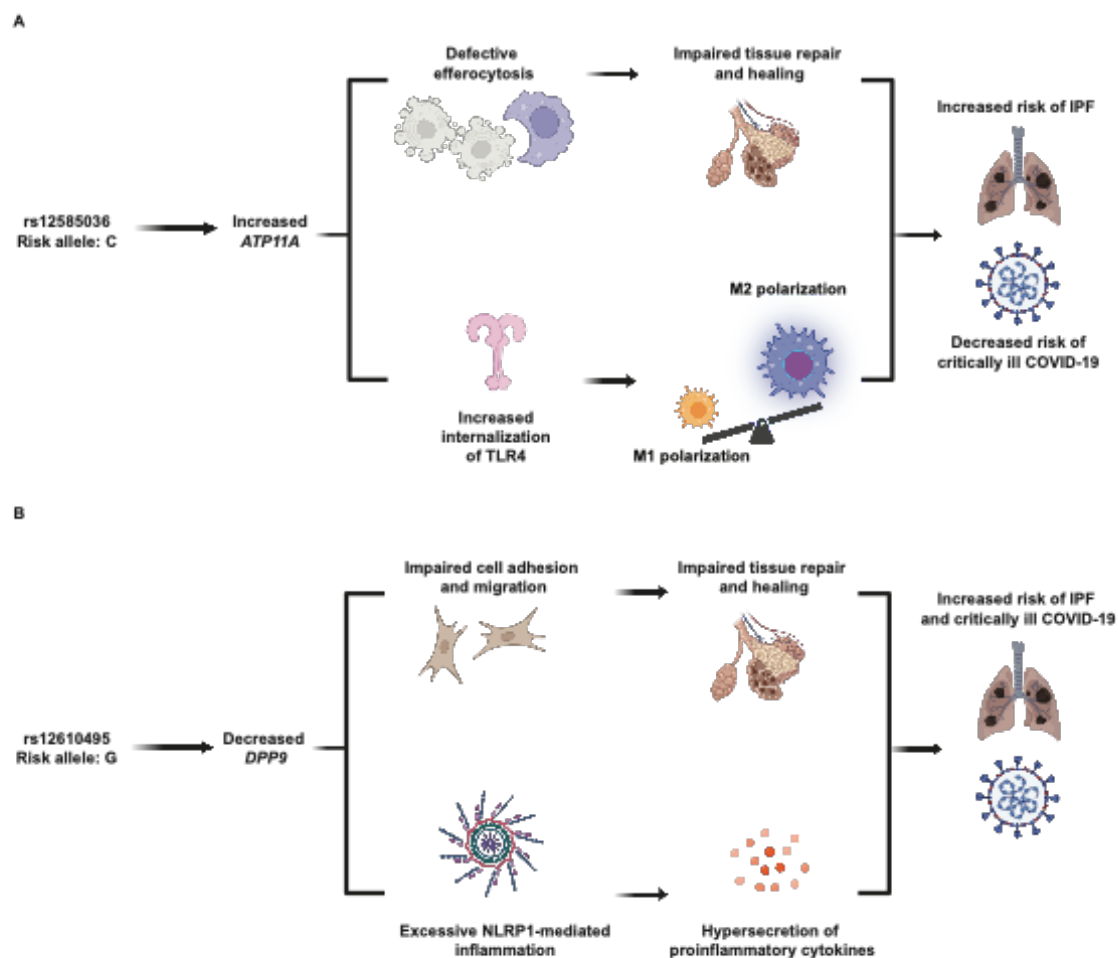
While the evidence we have presented suggests that variants in *ATP11A* and *DPP9* impact COVID-19 and IPF pathogenesis by altering expression of these genes, recent evidence from Nakanishi et al. demonstrated that the sQTL data for these loci also colocalize with severe COVID-19.[101] Future studies will be necessary to understand the effect of these risk variants on splicing, but we note that each sQTL is predicted to only impact a single non-canonical isoform based on Ensembl annotation. The isoform associated with *DPP9* (ENST00000599248) is not protein-coding; the associated isoform for *ATP11A* (ENST00000415301) is highly truncated. Thus, *ATP11A* and *DPP9* demonstrate a new "problem" in human genetics; the same human disease signal may colocalize with multiple molecular QTL signals, expanding the list of possible causal genes that underlie a GWAS signal. eQTL and sQTL data may also demonstrate two consequences of the same SNP as altering splicing can certainly impact overall gene expression.

Identifying shared genetic associations and their causal genes can reveal new insight into the pathogeneses of COVID-19 and IPF that could spur further therapeutic development. Indeed, MUC5B has been proposed as a target for IPF treatment since the genetic association was first reported.[102] For the causal genes with new evidence presented here, *DPP9* may be particularly attractive, and inhibitors in development have been speculated to be therapeutic for pulmonary fibrosis.[103] However, for treatment of COVID-19 and IPF, increasing *DPP9* expression or activity seems to be the desired outcome. It remains to be tested if recombinant DPP9 or increasing expression in fibroblasts is beneficial in critically ill COVID-19 or IPF.

ACKNOWLEDGEMENTS

REFERENCES

1. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. Nat Genet *47*, 856-860. 10.1038/ng.3314.
2. Ochoa, D., Karim, M., Ghoussaini, M., Hulcoop, D.G., McDonagh, E.M., and Dunham, I. (2022). Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. Nat Rev Drug Discov *21*, 551. 10.1038/d41573-022-00120-3.
3. Gratten, J., and Visscher, P.M. (2016). Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. Genome Med *8*, 78. 10.1186/s13073-016-0332-x.
4. Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nature Genetics *51*, 1339-1348. 10.1038/s41588-019-0481-0.
5. Hormozdiari, F., van de Bunt, M., Segrè, A.V., Li, X., Joo, J.W.J., Bilow, M., Sul, J.H., Sankararaman, S., Pasaniuc, B., and Eskin, E. (2016). Colocalization of GWAS and eQTL Signals Detects Target Genes. Am J Hum Genet *99*, 1245-1260. 10.1016/j.ajhg.2016.10.003.
6. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet *10*, e1004383. 10.1371/journal.pgen.1004383.
7. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet *17*, e1009440. 10.1371/journal.pgen.1009440.
8. Mostafavi, H., Spence, J.P., Naqvi, S., and Pritchard, J.K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat Genet *55*, 1866-1875. 10.1038/s41588-023-01529-1.
9. Pan, S., Kang, H., Liu, X., Li, S., Yang, P., Wu, M., Yuan, N., Lin, S., Zheng, Q., and Jia, P. (2024). COLOCdb: a comprehensive resource for multi-model colocalization of complex traits. Nucleic Acids Res *52*, D871-d881. 10.1093/nar/gkad939.
10. Wang, L., Oehlers, S.H., Espenschied, S.T., Rawls, J.F., Tobin, D.M., and Ko, D.C. (2015). CPAG: software for leveraging pleiotropy in GWAS to reveal similarity between human traits links plasma fatty acids and intestinal inflammation. Genome Biol *16*, 190. 10.1186/s13059-015-0722-1.
11. Wang, L., Balmat, T.J., Antonia, A.L., Constantine, F.J., Henao, R., Burke, T.W., Ingham, A., McClain, M.T., Tsalik, E.L., Ko, E.R., et al. (2021). An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. Genome Med *13*, 83. 10.1186/s13073-021-00904-z.
12. Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., et al. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med *383*, 1522-1534. 10.1056/NEJMoa2020283.
13. Allen, R.J., Guillen-Guio, B., Oldham, J.M., Ma, S.F., Dressen, A., Paynton, M.L., Kraven, L.M., Obeidat, M., Li, X., Ng, M., et al. (2020). Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med *201*, 564-574. 10.1164/rccm.201905-1017OC.
14. Degenhardt, F., Ellinghaus, D., Juzenas, S., Lerga-Jaso, J., Wendorff, M., Maya-Miles, D., Uellendahl-Werth, F., ElAbd, H., Rühlemann, M.C., Arora, J., et al. (2022). Detailed stratified GWAS analysis for severe COVID-19 in four European populations. Human Molecular Genetics *31*, 3945-3966. 10.1093/hmg/ddac158.
15. Kousathanas, A., Pairo-Castineira, E., Rawlik, K., Stuckey, A., Odhams, C.A., Walker, S., Russell, C.D., Malinauskas, T., Wu, Y., Millar, J., et al. (2022). Whole-genome

54

sequencing reveals host factors underlying critical COVID-19. Nature *607*, 97-103. 10.1038/s41586-022-04576-6.

16. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., et al. (2021). Genetic mechanisms of critical illness in COVID-19. Nature *591*, 92-98. 10.1038/s41586-020-03065-y.

17. Allen, R.J., Guillen-Guio, B., Croot, E., Kraven, L.M., Moss, S., Stewart, I., Jenkins, R.G., and Wain, L.V. (2022). Genetic overlap between idiopathic pulmonary fibrosis and COVID-19. Eur Respir J *60*. 10.1183/13993003.03132-2021.

18. Fadista, J., Kraven, L.M., Karjalainen, J., Andrews, S.J., Geller, F., Baillie, J.K., Wain, L.V., Jenkins, R.G., and Feenstra, B. (2021). Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. EBioMedicine *65*, 103277. 10.1016/j.ebiom.2021.103277.

19. Verma, A., Minnier, J., Wan, E.S., Huffman, J.E., Gao, L., Joseph, J., Ho, Y.L., Wu, W.C., Cho, K., Gorman, B.R., et al. (2022). A MUC5B Gene Polymorphism, rs35705950-T, Confers Protective Effects Against COVID-19 Hospitalization but Not Severe Disease or Mortality. Am J Respir Crit Care Med *206*, 1220-1229. 10.1164/rccm.202109-2166OC.

20. Davis, H.E., McCorkell, L., Vogel, J.M., and Topol, E.J. (2023). Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol *21*, 133-146. 10.1038/s41579-022-00846-2.

21. Meyerholz, D.K. (2023). Rigid respiration: fulminant pulmonary fibrosis after COVID-19. EBioMedicine *87*, 104428. 10.1016/j.ebiom.2022.104428.

22. Hama Amin, B.J., Kakamad, F.H., Ahmed, G.S., Ahmed, S.F., Abdulla, B.A., Mohammed, S.H., Mikael, T.M., Salih, R.Q., Ali, R.K., Salh, A.M., and Hussein, D.A. (2022). Post COVID-19 pulmonary fibrosis; a meta-analysis study. Ann Med Surg (Lond) *77*, 103590. 10.1016/j.amsu.2022.103590.

23. Lassan, S., Tesar, T., Tisonova, J., and Lassanova, M. (2023). Pharmacological approaches to pulmonary fibrosis following COVID-19. Front Pharmacol *14*, 1143158. 10.3389/fphar.2023.1143158.

24. Initiative, C.-H.G. (2023). A second update on mapping the human genetic architecture of COVID-19. Nature *621*, E7-e26. 10.1038/s41586-023-06355-3.

25. Allen, R.J., Stockwell, A., Oldham, J.M., Guillen-Guio, B., Schwartz, D.A., Maher, T.M., Flores, C., Noth, I., Yaspan, B.L., Jenkins, R.G., and Wain, L.V. (2022). Genome-wide association study across five cohorts identifies five novel loci associated with idiopathic pulmonary fibrosis. Thorax *77*, 829-833. 10.1136/thoraxjnl-2021-218577.

26. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nature Genetics *45*, 580-585. 10.1038/ng.2653.

27. Portal, G.-T.E.P.G. GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2).

28. Borie, R., Cardwell, J., Konigsberg, I.R., Moore, C.M., Zhang, W., Sasse, S.K., Gally, F., Dobrinskikh, E., Walts, A., Powers, J., et al. (2022). Colocalization of Gene Expression and DNA Methylation with Genetic Risk Variants Supports Functional Roles of MUC5B and DSP in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med *206*, 1259-1270. 10.1164/rccm.202110-2308OC.

29. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res *34*, D590-598. 10.1093/nar/gkj144.

30. Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841-1842. 10.1093/bioinformatics/btp328.

31.   Campoy, E., Puig, M., Yakymenko, I., Lerga-Jaso, J., and Cáceres, M. (2022). Genomic
        architecture and functional effects of potential human inversion supergenes. Philos
        Trans R Soc Lond B Biol Sci *377*, 20210209. 10.1098/rstb.2021.0209.

32.   Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A., and Wallace, C. (2015).
        Integration of disease association and eQTL data using a Bayesian colocalisation
        approach highlights six candidate causal genes in immune-mediated diseases. Human
        Molecular Genetics *24*, 3305-3313. 10.1093/hmg/ddv077.

33.   Jaffar, J., Wong, M., Fishbein, G.A., Alhamdoosh, M., McMillan, L., Gamell-Fulla, C., Ng,
        M., Wilson, N., Symons, K., Glaspole, I., and Westall, G. (2022). Matrix
        metalloproteinase-7 is increased in lung bases but not apices in idiopathic pulmonary
        fibrosis. ERJ Open Res *8*. 10.1183/23120541.00191-2022.

34.   Sivakumar, P., Thompson, J.R., Ammar, R., Porteous, M., McCoubrey, C., Cantu, E., 3rd,
        Ravi, K., Zhang, Y., Luo, Y., Streltsov, D., et al. (2019). RNA sequencing of transplant-
        stage idiopathic pulmonary fibrosis lung reveals unique pathway regulation. ERJ Open
        Res *5*. 10.1183/23120541.00117-2019.

35.   Carapito, R., Li, R., Helms, J., Carapito, C., Gujja, S., Rolli, V., Guimaraes, R., Malagon-
        Lopez, J., Spinnhirny, P., Lederle, A., et al. (2022). Identification of driver genes for
        critical forms of COVID-19 in a deeply phenotyped young patient cohort. Sci Transl Med
        14, eabj7521. 10.1126/scitranslmed.abj7521.

36.   de Rooij, L., Becker, L.M., Teuwen, L.A., Boeckx, B., Jansen, S., Feys, S., Verleden, S.,
        Liesenborghs, L., Stalder, A.K., Libbrecht, S., et al. (2023). The pulmonary vasculature in
        lethal COVID-19 and idiopathic pulmonary fibrosis at single-cell resolution. Cardiovasc
        Res *119*, 520-535. 10.1093/cvr/cvac139.

37.   Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
        dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-
        0550-8.

38.   Adams, T.S., Schupp, J.C., Poli, S., Ayaub, E.A., Neumark, N., Ahangari, F., Chu, S.G.,
        Raby, B.A., DeIuliis, G., Januszyk, M., et al. (2020). Single-cell RNA-seq reveals ectopic
        and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci Adv *6*,
        eaba1983. 10.1126/sciadv.aba1983.

39.   Melms, J.C., Biermann, J., Huang, H., Wang, Y., Nair, A., Tagore, S., Katsyv, I.,
        Rendeiro, A.F., Amin, A.D., Schapiro, D., et al. (2021). A molecular single-cell lung atlas
        of lethal COVID-19. Nature *595*, 114-119. 10.1038/s41586-021-03569-1.

40.   Tarhan, L., Bistline, J., Chang, J., Galloway, B., Hanna, E., and Weitz, E. (2023). Single
        Cell Portal: an interactive home for single-cell genomics data. bioRxiv.
        10.1101/2023.07.13.548886.

41.   Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene
        expression data analysis. Genome Biol *19*, 15. 10.1186/s13059-017-1382-0.

42.   Badia, I.M.P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S.,
        Taus, P., Dugourd, A., Holland, C.H., Ramirez Flores, R.O., and Saez-Rodriguez, J.
        (2022). decoupleR: ensemble of computational methods to infer biological activities from
        omics data. Bioinform Adv *2*, vbac016. 10.1093/bioadv/vbac016.

43.   Muzellec, B., Teleńczuk, M., Cabeli, V., and Andreux, M. (2023). PyDESeq2: a python
        package for bulk RNA-seq differential expression analysis. Bioinformatics *39*.
        10.1093/bioinformatics/btad547.

44.   R Core Team (2023). (R Foundation for Statistical Computing).

45.   Valero-Mora, P.M. (2010). ggplot2: Elegant Graphics for Data Analysis. Journal of
        Statistical Software, Book Reviews *35*, 1 - 3. 10.18637/jss.v035.b01.

46.   Paria, S.S., Rahman, S.R., and Adhikari, K. (2022). fastman: A fast algorithm for
        visualizing GWAS results using Manhattan and Q-Q plots. bioRxiv,
        2022.2004.2019.488738. 10.1101/2022.04.19.488738.

47. Lewis, M. (2024). locuszoomr: Gene Locus Plot with Gene Annotations.
48. Zhou, W., Triche, T.J., Jr., Laird, P.W., and Shen, H. (2018). SeSAMe: reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. Nucleic Acids Res *46*, e123. 10.1093/nar/gky691.
49. Initiative, C.-H.G. (2022). A first update on mapping the human genetic architecture of COVID-19. Nature *608*, E1-e10. 10.1038/s41586-022-04826-7.
50. Initiative, C.-H.G. (2021). Mapping the human genetic architecture of COVID-19. Nature *600*, 472-477. 10.1038/s41586-021-03767-x.
51. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res *40*, D930-934. 10.1093/nar/gkr917.
52. Helling, B.A., Gerber, A.N., Kadiyala, V., Sasse, S.K., Pedersen, B.S., Sparks, L., Nakano, Y., Okamoto, T., Evans, C.M., Yang, I.V., and Schwartz, D.A. (2017). Regulation of MUC5B Expression in Idiopathic Pulmonary Fibrosis. American journal of respiratory cell and molecular biology *57*, 91-99. 10.1165/rcmb.2017-0046OC.
53. Connally, N.J., Nazeen, S., Lee, D., Shi, H., Stamatoyannopoulos, J., Chun, S., Cotsapas, C., Cassa, C.A., and Sunyaev, S.R. (2022). The missing link between genetic association and regulatory function. Elife *11*. 10.7554/eLife.74970.
54. Willett, J.D.S., Lu, T., Nakanishi, T., Yoshiji, S., Butler-Laporte, G., Zhou, S., Farjoun, Y., and Richards, J.B. (2023). Colocalization of expression transcripts with COVID-19 outcomes is rare across cell states, cell types and organs. Hum Genet *142*, 1461-1476. 10.1007/s00439-023-02590-w.
55. Hao, K., Bossé, Y., Nickle, D.C., Paré, P.D., Postma, D.S., Laviolette, M., Sandford, A., Hackett, T.L., Daley, D., Hogg, J.C., et al. (2012). Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet *8*, e1003029. 10.1371/journal.pgen.1003029.
56. dbGaP (2019). Common Fund (CF) Genotype-Tissue Expression Project (GTEx), dbGaP Study Accession: phs000424.v8.p2. 2024-04-22 ed.
57. Bossé, Y., Postma, D.S., Sin, D.D., Lamontagne, M., Couture, C., Gaudreault, N., Joubert, P., Wong, V., Elliott, M., van den Berge, M., et al. (2012). Molecular signature of smoking in human lung tissues. Cancer Res *72*, 3753-3763. 10.1158/0008-5472.Can-12-1160.
58. Zhu, J., Zhou, D., Yu, M., and Li, Y. (2024). Appraising the causal role of smoking in idiopathic pulmonary fibrosis: a Mendelian randomization study. Thorax *79*, 179-181. 10.1136/thorax-2023-220012.
59. Margaritopoulos, G.A., Vasarmidi, E., Jacob, J., Wells, A.U., and Antoniou, K.M. (2015). Smoking and interstitial lung diseases. Eur Respir Rev *24*, 428-435. 10.1183/16000617.0050-2015.
60. Baumgartner, K.B., Samet, J.M., Stidley, C.A., Colby, T.V., and Waldron, J.A. (1997). Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. Am J Respir Crit Care Med *155*, 242-248. 10.1164/ajrccm.155.1.9001319.
61. Elisia, I., Lam, V., Cho, B., Hay, M., Li, M.Y., Yeung, M., Bu, L., Jia, W., Norton, N., Lam, S., and Krystal, G. (2020). The effect of smoking on chronic inflammation, immune function and blood cell composition. Sci Rep *10*, 19480. 10.1038/s41598-020-76556-7.
62. Lugg, S.T., Scott, A., Parekh, D., Naidu, B., and Thickett, D.R. (2022). Cigarette smoke exposure and alveolar macrophages: mechanisms for lung disease. Thorax *77*, 94-101. 10.1136/thoraxjnl-2020-216296.
63. Wohnhaas, C.T., Baßler, K., Watson, C.K., Shen, Y., Leparc, G.G., Tilp, C., Heinemann, F., Kind, D., Stierstorfer, B., Delić, D., et al. (2024). Monocyte-derived alveolar macrophages are key drivers of smoke-induced lung inflammation and tissue remodeling. Front Immunol *15*, 1325090. 10.3389/fimmu.2024.1325090.

64. Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell *167*, 1398-1414.e1324. 10.1016/j.cell.2016.10.026.

65. Willis, B.C., duBois, R.M., and Borok, Z. (2006). Epithelial origin of myofibroblasts during fibrosis in the lung. Proc Am Thorac Soc *3*, 377-382. 10.1513/pats.200601-004TK.

66. Phan, S.H. (2012). Genesis of the myofibroblast in lung injury and fibrosis. Proc Am Thorac Soc *9*, 148-152. 10.1513/pats.201201-011AW.

67. Lin, Y., and Xu, Z. (2020). Fibroblast Senescence in Idiopathic Pulmonary Fibrosis. Front Cell Dev Biol *8*, 593283. 10.3389/fcell.2020.593283.

68. Peyser, R., MacDonnell, S., Gao, Y., Cheng, L., Kim, Y., Kaplan, T., Ruan, Q., Wei, Y., Ni, M., Adler, C., et al. (2019). Defining the Activated Fibroblast Population in Lung Fibrosis Using Single-Cell Sequencing. Am J Respir Cell Mol Biol *61*, 74-85. 10.1165/rcmb.2018-0313OC.

69. Aquino, Y., Bisiaux, A., Li, Z., O'Neill, M., Mendoza-Revilla, J., Merkling, S.H., Kerner, G., Hasan, M., Libri, V., Bondet, V., et al. (2023). Dissecting human population variation in single-cell responses to SARS-CoV-2. Nature *621*, 120-128. 10.1038/s41586-023-06422-9.

70. Natri, H.M., Del Azodi, C.B., Peter, L., Taylor, C.J., Chugh, S., Kendle, R., Chung, M.I., Flaherty, D.K., Matlock, B.K., Calvi, C.L., et al. (2024). Cell-type-specific and disease-associated expression quantitative trait loci in the human lung. Nat Genet *56*, 595-604. 10.1038/s41588-024-01702-0.

71. Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K., et al. (2016). Systematic identification of genetic influences on methylation across the human life course. Genome Biol *17*, 61. 10.1186/s13059-016-0926-z.

72. McRae, A.F., Marioni, R.E., Shah, S., Yang, J., Powell, J.E., Harris, S.E., Gibson, J., Henders, A.K., Bowdler, L., Painter, J.N., et al. (2018). Identification of 55,000 Replicated DNA Methylation QTL. Sci Rep *8*, 17605. 10.1038/s41598-018-35871-w.

73. Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. Neuropsychopharmacology *38*, 23-38. 10.1038/npp.2012.112.

74. Perrot, C.Y., Karampitsakos, T., and Herazo-Maya, J.D. (2023). Monocytes and macrophages: emerging mechanisms and novel therapeutic targets in pulmonary fibrosis. Am J Physiol Cell Physiol *325*, C1046-C1057. 10.1152/ajpcell.00302.2023.

75. Hancock, L.A., Hennessy, C.E., Solomon, G.M., Dobrinskikh, E., Estrella, A., Hara, N., Hill, D.B., Kissner, W.J., Markovetz, M.R., Grove Villalon, D.E., et al. (2018). Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. Nat Commun *9*, 5363. 10.1038/s41467-018-07768-9.

76. Kato, T., Asakura, T., Edwards, C.E., Dang, H., Mikami, Y., Okuda, K., Chen, G., Sun, L., Gilmore, R.C., Hawkins, P., et al. (2022). Prevalence and Mechanisms of Mucus Accumulation in COVID-19 Lung Disease. Am J Respir Crit Care Med *206*, 1336-1352. 10.1164/rccm.202111-2606OC.

77. Yan, J., Li, Q., Mao, A.P., Hu, M.M., and Shu, H.B. (2014). TRIM4 modulates type I interferon induction and cellular antiviral response by targeting RIG-I for K63-linked ubiquitination. J Mol Cell Biol *6*, 154-163. 10.1093/jmcb/mju005.

78. Zhang, X., Yang, Z., Pan, T., Sun, Q., Chen, Q., Wang, P.H., Li, X., and Kuang, E. (2023). SARS-CoV-2 Nsp8 suppresses MDA5 antiviral immune responses by impairing TRIM4-mediated K63-linked polyubiquitination. PLoS Pathog *19*, e1011792. 10.1371/journal.ppat.1011792.

79.   Li, C., Wang, Z., Zhang, J., Zhao, X., Xu, P., Liu, X., Li, M., Lv, C., and Song, X. (2019). Crosstalk of mRNA, miRNA, lncRNA, and circRNA and Their Regulatory Pattern in Pulmonary Fibrosis. Mol Ther Nucleic Acids *18*, 204-218. 10.1016/j.omtn.2019.08.018.

80.   Jiang, X., and Ning, Q. (2019). Circular RNAs as novel regulators, biomarkers and potential therapies in fibrosis. Epigenomics *11*, 1107-1116. 10.2217/epi-2019-0001.

81.   Yao, Z., Luo, J., Hu, K., Lin, J., Huang, H., Wang, Q., Zhang, P., Xiong, Z., He, C., Huang, Z., et al. (2017). ZKSCAN1 gene and its related circular RNA (circZKSCAN1) both inhibit hepatocellular carcinoma cell growth, migration, and invasion but through different signaling pathways. Mol Oncol *11*, 422-437. 10.1002/1878-0261.12045.

82.   Khalil, N., Bereznay, O., Sporn, M., and Greenberg, A.H. (1989). Macrophage production of transforming growth factor beta and fibroblast collagen synthesis in chronic pulmonary inflammation. J Exp Med *170*, 727-737. 10.1084/jem.170.3.727.

83.   Takatsu, H., Tanaka, G., Segawa, K., Suzuki, J., Nagata, S., Nakayama, K., and Shin, H.W. (2014). Phospholipid flippase activities and substrate specificities of human type IV P-type ATPases localized to the plasma membrane. J Biol Chem *289*, 33543-33556. 10.1074/jbc.M114.593012.

84.   Segawa, K., Kurata, S., and Nagata, S. (2016). Human Type IV P-type ATPases That Work as Plasma Membrane Phospholipid Flippases and Their Regulation by Caspase and Calcium. J Biol Chem *291*, 762-772. 10.1074/jbc.M115.690727.

85.   Nagata, S., Suzuki, J., Segawa, K., and Fujii, T. (2016). Exposure of phosphatidylserine on the cell surface. Cell Death Differ *23*, 952-961. 10.1038/cdd.2016.7.

86.   Segawa, K., and Nagata, S. (2015). An Apoptotic 'Eat Me' Signal: Phosphatidylserine Exposure. Trends Cell Biol *25*, 639-650. 10.1016/j.tcb.2015.08.003.

87.   Drakopanagiotakis, F., Xifteri, A., Polychronopoulos, V., and Bouros, D. (2008). Apoptosis in lung injury and fibrosis. Eur Respir J *32*, 1631-1638. 10.1183/09031936.00176807.

88.   Uhal, B.D. (2008). The role of apoptosis in pulmonary fibrosis. European Respiratory Review *17*, 138-144. 10.1183/09059180.00010906.

89.   Kim, K.K., Dotson, M.R., Agarwal, M., Yang, J., Bradley, P.B., Subbotina, N., Osterholzer, J.J., and Sisson, T.H. (2018). Efferocytosis of apoptotic alveolar epithelial cells is sufficient to initiate lung fibrosis. Cell Death Dis *9*, 1056. 10.1038/s41419-018-1074-z.

90.   Morimoto, K., Janssen, W.J., and Terada, M. (2012). Defective efferocytosis by alveolar macrophages in IPF patients. Respir Med *106*, 1800-1803. 10.1016/j.rmed.2012.08.020.

91.   van der Mark, V.A., Ghiboub, M., Marsman, C., Zhao, J., van Dijk, R., Hiralall, J.K., Ho-Mok, K.S., Castricum, Z., de Jonge, W.J., Oude Elferink, R.P., and Paulusma, C.C. (2017). Phospholipid flippases attenuate LPS-induced TLR4 signaling by mediating endocytic retrieval of Toll-like receptor 4. Cell Mol Life Sci *74*, 715-730. 10.1007/s00018-016-2360-5.

92.   Liang, J., Zhang, Y., Xie, T., Liu, N., Chen, H., Geng, Y., Kurkciyan, A., Mena, J.M., Stripp, B.R., Jiang, D., and Noble, P.W. (2016). Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. Nat Med *22*, 1285-1293. 10.1038/nm.4192.

93.   Sahanic, S., Hilbe, R., Dünser, C., Tymoszuk, P., Löffler-Ragg, J., Rieder, D., Trajanoski, Z., Krogsdam, A., Demetz, E., Yurchenko, M., et al. (2023). SARS-CoV-2 activates the TLR4/MyD88 pathway in human macrophages: A possible correlation with strong pro-inflammatory responses in severe COVID-19. Heliyon *9*, e21893. 10.1016/j.heliyon.2023.e21893.

94.   Zhao, Y., Kuang, M., Li, J., Zhu, L., Jia, Z., Guo, X., Hu, Y., Kong, J., Yin, H., Wang, X., and You, F. (2021). SARS-CoV-2 spike protein interacts with and activates TLR41. Cell Res *31*, 818-820. 10.1038/s41422-021-00495-9.

95.   Hollingsworth, L.R., Sharif, H., Griswold, A.R., Fontana, P., Mintseris, J., Dagbay, K.B., Paulo, J.A., Gygi, S.P., Bachovchin, D.A., and Wu, H. (2021). DPP9 sequesters the C terminus of NLRP1 to repress inflammasome activation. Nature *592*, 778-783. 10.1038/s41586-021-03350-4.

96.   Okondo, M.C., Rao, S.D., Taabazuing, C.Y., Chui, A.J., Poplawski, S.E., Johnson, D.C., and Bachovchin, D.A. (2018). Inhibition of Dpp8/9 Activates the Nlrp1b Inflammasome. Cell Chem Biol *25*, 262-267.e265. 10.1016/j.chembiol.2017.12.013.

97.   Zhong, F.L., Robinson, K., Teo, D.E.T., Tan, K.Y., Lim, C., Harapas, C.R., Yu, C.H., Xie, W.H., Sobota, R.M., Au, V.B., et al. (2018). Human DPP9 represses NLRP1 inflammasome and protects against autoinflammatory diseases via both peptidase activity and FIIND domain binding. J Biol Chem *293*, 18864-18878. 10.1074/jbc.RA118.004350.

98.   Gabrilovac, J., Čupić, B., Zapletal, E., Kraus, O., and Jakić-Razumović, J. (2017). Dipeptidyl peptidase 9 (DPP9) in human skin cells. Immunobiology *222*, 327-342. 10.1016/j.imbio.2016.09.007.

99.   Acharya, P.S., Zukas, A., Chandan, V., Katzenstein, A.L., and Puré, E. (2006). Fibroblast activation protein: a serine protease expressed at the remodeling interface in idiopathic pulmonary fibrosis. Hum Pathol *37*, 352-360. 10.1016/j.humpath.2005.11.020.

100.   Wu, Q.Q., Zhao, M., Huang, G.Z., Zheng, Z.N., Chen, Y., Zeng, W.S., and Lv, X.Z. (2020). Fibroblast Activation Protein (FAP) Overexpression Induces Epithelial-Mesenchymal Transition (EMT) in Oral Squamous Cell Carcinoma by Down-Regulating Dipeptidyl Peptidase 9 (DPP9). Onco Targets Ther *13*, 2599-2611. 10.2147/ott.S243417.

101.   Nakanishi, T., Willett, J., Farjoun, Y., Allen, R.J., Guillen-Guio, B., Adra, D., Zhou, S., and Richards, J.B. (2023). Alternative splicing in lung influences COVID-19 severity and respiratory diseases. Nat Commun *14*, 6198. 10.1038/s41467-023-41912-4.

102.   Mahida, R., and Turner, A.M. (2013). MUC5B: a good target for future therapy in pulmonary fibrosis? Thorax *68*, 401. 10.1136/thoraxjnl-2012-202957.

103.   Benramdane, S., De Loose, J., Filippi, N., Espadinha, M., Beyens, O., Rymenant, Y.V., Dirkx, L., Bozdag, M., Feijens, P.B., Augustyns, K., et al. (2023). Highly Selective Inhibitors of Dipeptidyl Peptidase 9 (DPP9) Derived from the Clinically Used DPP4-Inhibitor Vildagliptin. J Med Chem *66*, 12717-12738. 10.1021/acs.jmedchem.3c00609.