# Proteome-wide Mendelian randomization identifies causal plasma proteins in Interstitial lung disease

**Kunrong Yu**
Guangzhou University of Chinese Medicine
**Wanying Li**
Guangzhou University of Chinese Medicine
**Wenjie Long**
The First Affiliated Hospital of Guangzhou University of Chinese Medicine
**Yijia Li**
Guangzhou University of Chinese Medicine
**Yanting Li**
Guangzhou University of Chinese Medicine
**Huili Liao**
The First Affiliated Hospital of Guangzhou University of Chinese Medicine
**Jianhong Liu**
ljh833@163.com

The First Affiliated Hospital of Guangzhou University of Chinese Medicine

**Research Article**

**Keywords:** Interstitial lung disease, subtype, Protein, Mendelian randomization, Biomarker, Drug target

**Posted Date:** April 4th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-4192254/v1



**Additional Declarations:** No competing interests reported.

## Abstract

# Background

Interstitial lung disease (ILD) has exhibited limited overall treatment advancements, with scant exploration into circulating protein biomarkers causally linked to ILD and its subtypes beyond idiopathic pulmonary fibrosis (IPF). Therefore, our study aims to investigate potential drug targets and circulating protein biomarkers for ILD and its subtypes.

# Methods

We utilized the most recent large-scale plasma protein quantitative trait loci (pQTL) data detected from the antibody-based method and ILD and its subtypes' GWAS data from the updated FinnGen database for Mendelian randomization analysis. To enhance the reliability of causal associations, we conducted external validation and sensitivity analyses, including Bayesian colocalization, bidirectional Mendelian randomization analysis, and phenotype scanning.

# Results

Genetic prediction levels of eight proteins were associated with the risk of ILD or its subtypes. Through a series of sensitivity analyses, three proteins were identified as priority proteins for circulating biomarkers and potential therapeutic targets. Specifically, CDH15(Cadherin-15)increased the risk of ILD OR = 1.32, 95%CI 1.16–1.49, P = $1.60\times10^{-6}$ , and LTBR Lymphotoxin-beta receptor increased the risk of sarcoidosis OR = 1.39, 95%CI 1.20–1.61, p = $9.38\times10^{-6}$ , while ADAM15 (A disintegrin and metalloproteinase 15) were protective proteins for ILD OR = 0.86, 95% CI 0.81–0.92, P = $1.59\times10^{-6}$ and IPF OR = 0.81, 95% CI 0.75–0.89). Moreover, no causal proteins for other ILD subtypes were found.

# Conclusion

This study identified several new circulating protein biomarkers associated with the risk of ILD and its subtypes. It offers a new perspective for future research on the diagnosis and treatment of ILD and its subtypes.


## Background

Interstitial lung diseases (ILD) constitute a diverse array of pulmonary disorders primarily characterized by interstitial proliferation and inflammatory cell infiltration. While these diseases are grouped based on shared clinical, radiographic, and pathological features, they represent distinct disease trajectories. Idiopathic pulmonary fibrosis (IPF) represents one of the most commonly encountered subtypes in the

diagnosis and management of ILD, constituting a chronic and incurable condition. More and more case reports suggest that ILD is a prevalent complication of connective tissue disease (CTD), primarily affecting patients with rheumatoid arthritis (RA) and systemic sclerosis (SSc). Additionally, other ILD subtypes, including idiopathic nonspecific interstitial pneumonia (iNSIP) and hypersensitivity pneumonitis (HP), are associated similarly with generally poor prognoses. Compared to the foregoing subtypes, patients with sarcoidosis generally have a better overall prognosis.

Epidemiological data from the PERSEIDS study(1), covering the years 2014–2018 and focusing on the European population, reveal an incidence of ILD ranging from 33.6-247.4 per $10^5$ persons, with IPF and systemic sclerosis-related ILD (SSc-ILD) ranging from 2.8–31.0 and 1.4–10.1 per $10^5$ persons, respectively. Another epidemiological investigation focused on sarcoidosis, which reported an annual incidence rate of 10–15 per $10^5$ persons in the European population(2). Furthermore, fibrotic pathological changes predominate over inflammation, with fibrosing interstitial lung diseases (F-ILD) and non-IPF fibrosing ILD (non-IPF F-ILD) occurring at rates of 26.7-236.8 and 22.3-205.8 per $10^5$ persons. Sarcoidosis is the most common subtype of non-IPF F-ILDs. The data on progressive fibrosing ILD (PF-ILD) are particularly noteworthy, with an incidence ranging from 2.1–14.5/$10^5$ person-years and a prevalence ranging from 6.9–78.0 per $10^5$ persons. Surprisingly, progressive fibrosis is also observed in one-third of non-IPF F-ILD cases. Even in sarcoidosis, approximately 10–30% of patients progress to progressive lung disease(3). These findings support the emphasis of the 2022 ATS/ERS/JRS/ALAT clinical practice guidelines on recognizing progressive pulmonary fibrosis (PPF) as a phenotypic manifestation of specific ILDs rather than a distinct diagnosis, regardless of underlying diseases(4). The prognosis may be similar to IPF when progression, usual interstitial pneumonia (UIP), or both are present in non-IPF fibrosing ILD (non-IPF F-ILD). Recent research on the trajectory of ILD has reported a higher incidence of progressive fibrosis (PF) behaviour occurring in up to 50% of ILD patients, which correlates with increased mortality rates(5).

Based on these latest epidemiological data on ILD, it is evident that the prognosis for ILD diagnosis and treatment progress is not optimistic. With the extension of lifespan, it is foreseeable that the incidence of ILD and its subtypes, predominantly influenced by age, will continue to rise gradually. Furthermore, the number of patients with progressive fibrosis among non-IPF individuals is equally staggering, thus elevating non-IPF research to a status equivalent to IPF is urgently required.

Multiple studies demonstrate that drugs developed with support from human genetics evidence have at least twice the likelihood of approval(6, 7), and combining high-throughput, population-scale proteomics with human genetics for clearer causal gene mediation will undoubtedly increase this likelihood even further. While previously conducted similar research(8), they primarily focused on IPF and did not systematically analyze ILD and its other subtypes. Moreover, a key distinction is that previous studies utilized proteomic genetic data based on aptamer-based detection methods, while this study employs antibody-based genetic research proteomic data. Comparative assessments indicate that the aptamer-based detection method offers broader biological coverage and higher accuracy, whereas the antibody-

based detection method exhibits greater protein target specificity, with stronger correlations to certain diseases and immune assays(9). Thus, researchers widely accept that these two technologies, capturing different features of protein chemistry, serve as complementary tools for biomarker discovery(10).

Overall, employing the latest antibody-based genetic research proteomic data with larger sample sizes for ILD and its subtypes, including IPF, sarcoidosis, and ILD related to systemic autoimmune disease (SAD-ILD), through Mendelian randomization may reveal diverse and more meaningful potential drug targets as well as subtype-specific circulating protein biomarkers for ILD.

## Methods

Figure 1 outlines the overall study design. In detail, we utilized the latest large-scale antibody-based detection method for protein quantitative trait loci (pQTL) data from a European population. We examined their associations with ILD and its subtypes using a two-stage (discovery and replication) proteome-wide Mendelian randomization (MR) framework. A series of sensitivity analyses, including Bayesian colocalization, bidirectional MR analysis, and phenotype scanning, were conducted to verify the authenticity of causal relationships. Finally, protein-protein interaction (PPI) and druggability evaluation of priority proteins were performed to explore the potential as therapeutic targets for ILD and its subtypes.

# Data Source

We obtained genome-wide summary-level statistics for cis-pQTL from the UK Biobank Pharma Proteomics Project (UKB-PPP) study(11). GWAS summary statistics for ILD and its subtypes can be retrieved from the GWAS catalog(12), FinnGen (R10), and UK Biobank (UKB)(13, 14). All GWASs were almost based on the European population to ensure homogeneity of the study population. FinnGen (R10) had no sample overlap with the UK Biobank and GWAS catalog. Plasma pQTL data for MR external validation was available and obtained from the original studies(15–19). Details of each GWAS were described in Additional file 1: Table S1.

# Instrumental variable selection

As global investigations into the genome of circulatory proteins continue to deepen, it found that genetic factors determining circulatory protein levels are typically located in cis (proximate) positions of coding genes rather than in trans (distant) single nucleotide polymorphisms (SNPs). These cis-acting SNPs closely associated with proteins are likely to influence gene transcription and translation, directly affecting circulatory protein levels. Consequently, employing such cis-acting SNPs for MR studies serves as a more proximal proxy, less likely to be mediated by other molecules, thereby reducing potential horizontal pleiotropy and enhancing the validity of MR assumptions.

The plasma pQTL data for preliminary MR analysis derived from the UKB-PPP study(11), which utilized an antibody-based detection method to comprehensively localize pQTLs for 2,922 proteins among 54,219 participants, identifying 14,287 pQTLs at a multiple testing correction threshold of $P < 1.7\times10^{-11}$. Within

this study, pQTLs located within 1 Mb of the transcription start site of protein-coding genes were defined as cis-pQTLs, while those outside were considered trans-pQTLs.

In this study, efforts were made to limit biases from horizontal pleiotropy and linkage disequilibrium and to ensure the strength of instrumental genetic tools. Only pQTLs meeting the following criteria were included: (1) genome-wide significance level $P < 5 \times 10^{-8}$; (2) located outside the major histocompatibility complex (MHC) region (chr6, 26–34 Mb); (3) exhibiting independent associations ($r^2 < 0.001$); (4) demonstrating cis-acting as pQTLs; and (5) F-statistic > 30. Ultimately, the test identified 1896 cis-pQTLs for 1896 proteins (instrumental variables see Additional file 1: Table S2).

# Mendelian randomization analysis

In pursuing potential therapeutic targets for ILD and its subtypes, Mendelian randomization (MR) analysis aids in ensuring the directionality from exposure to outcome in genetic studies. This study adhered strictly to three fundamental principles of Mendelian randomization analysis(20, 21): (1)Genetic variation must be reliably associated with exposure – Utilizing the most robust cis-pQTL variant associated with each plasma protein as instrumental exposure variable. (2)Genetic variation should not be associated with confounding factors of deterministic-outcome relationships – Enforcing strict inclusion criteria for instrumental variables and conducting multiple sensitivity analyses. (3) Genetic variation must not affect outcomes except through the benefits of exposure – Scrutinized via "Phenoscanner" and "LDlink"(22, 23). These principles ensure robustness and validity in identifying potential therapeutic targets through Mendelian randomization analysis.

In the primary MR analysis, we utilized summary statistical data from the FinnGen study (R10) of ILD and its subtypes as outcomes. The Wald ratio method from "TwoSampleMR" for proteins with a single cis-acting SNP to estimate the log odds change in ILD and its subtypes risk for per standard deviation (SD) increment or decrement of circulating protein levels as proxied by the instrumental variables (24, 25). In cases where a protein had two or more conditionally independent cis-pQTLs, considering the LD pattern among multiple cis-acting SNPs, the Inverse Variance Weighted (IVW) model was used to estimate MR effects and conduct pleiotropy and Cochran's Q tests for heterogeneity examination(26). In the preliminary analysis, stringent Bonferroni correction was applied to exclude multiple testing, using $P < 0.05/1896$ ($P < 2.64 \times 10^{-5}$) as the threshold to determine causality. Only proteins selected in the preliminary screening underwent MR external validation, with a replication MR P-value threshold set at 0.05, and outcome data sourced from the UK Biobank and GWAS catalog(27–29). The same or significant variant strategy was employed, where the former used the same SNPs as genetic instruments as in the preliminary analysis, and the latter used the most significant pQTL for that protein from the study as instrumental tools to validate the primary findings.

# Reverse causality detection

To ensure the directionality of the causal association between the proteins preliminarily selected and ILD and its subtypes, and to detect potential reverse causal relationships, we applied filtering criteria of p <

$5\times10^{-8}$, $r^2$ = 0.001, and clump_kb = 10000 to extract instrumental variables for bidirectional MR analysis from the summary statistical data of FinnGen (R10), with Steiger test results as further supplementary validation.

# Bayesian colocalization analysis

Although MR serves as a powerful tool for detecting causal effects and employs cis-pQTL to minimize horizontal pleiotropy, linkage disequilibrium (LD) still may confound its results. In other words, positive results indicating a causal association between exposure and outcome might be false positives when influenced by different genetic variants. Therefore, to minimize the risk of such false positives, we conducted a Bayesian colocalization analysis through the coloc.abf algorithm to assess the potential LD between the proteins preliminarily identified and ILD and its subtypes(30).

This analysis method utilizes the "coloc" package with default parameters (p1 = $1\times10^{-4}$, p2 = $1\times10^{-4}$, and p12 = $1\times10^{-5}$) to evaluate the probability of shared causal variants between the protein and trait(31). The “Locuscomparer” package was used to visualize the region results of colocalization(32). In this study, we primarily focused on the posterior probability of hypothesis 4 ($PPH_4$), which indicates strong evidence if PPH4 > 80% that the protein and trait share the same causal variant within the genomic locus(23).

# Phenotype scanning

To minimize the potential for horizontal pleiotropy in instrumental genetic variables, we conducted phenotype scans using "Phenoscanner" and "LDlink"(22, 23). SNPs were considered pleiotropic if they met the following criteria: (1) The association is significant at the genome-wide level ($P < 5 \times 10^{-8}$). (2) The SNP is associated with known risk factors for ILD and its subtypes, including metabolic traits, proteins, or clinical characteristics.

# PPI and druggability evaluation on the potentials of therapeutic targets

Furthermore, to explore whether potential drug targets for ILD and its subtypes are strongly associated with existing therapeutic drugs on the market and to unearth promising drug targets, we conducted a PPI analysis using the STRING database (V12.0). We set the minimum required interaction score threshold to 0.7. Additionally, we searched the PubMed database for literature evidence concerning these promising drug targets to assess their feasibility and credibility as drug targets.

# Results

### Screening the proteome for ILD causal proteins

After applying Bonferroni correction ($P < 2.64 \times 10^{-5}$), our primary MR analysis revealed significant associations between ILD and four plasma proteins, with IPF displaying similar associations. Notably, BRSK2 Serine/threonine-protein kinase BRSK2 , LRRC37A2(Leucine-rich repeat-containing protein37A2),

and ADAM15(A disintegrin and metalloproteinase 15) exhibited shared associations, while CDH15 (Cadherin-15) showed no significant association with IPF. In contrast, the results for sarcoidosis identified four candidate proteins that were entirely distinct from those associated with ILD and IPF after Benjamini-Hochberg correction: VEGFB (vascular endothelial growth factor B), ANXA11 (annexin A11), LTBR (Lymphotoxin-beta receptor), and ANGPTL4 (Angiopoietin-related protein 4). Furthermore, no significant association was found between SAD-ILD, allergic bronchopulmonary aspergillosis (ABPA), and any proteins in the study.

Specifically, the Wald ratio indicated that the genetically predicted increase in CDH15 was associated with an elevated risk of ILD, with an odds ratio (OR) of 1.32 (95%CI 1.16-1.49; $P = 1.09 \times 10^{-5}$). Similarly, elevating the expression levels of BRSK2 (OR=1.30; $P=1.70 \times 10^{-13}$) also increased the risk of ILD, while increasing the expression levels of ADAM15 (OR=0.86 95% CI 0.81-0.92; $P=1.59 \times 10^{-6}$) and LRRC37A2 (OR=0.82; $P=7.40 \times 10^{-8}$) decreased the risk of ILD. Similarly, the odds ratio (95% CI) of IPF per standard deviation increase in genetically predicted levels of protein was 1.40 (1.26-1.55) for BRSK2, whereas 0.81 (0.75–0.89) for ADAM15, 0.74 (0.66–0.82) for LRRC37A2. For sarcoidosis, another subtype of ILD, higher plasma levels of LTBR were associated with a significantly increased risk of developing sarcoidosis, with on average 39% increased risk (OR=1.39; $p=9.38\times10^{-6}$), while ANGPTL4 (OR=0.72; $p=1.68\times10^{-5}$), VEGFB (OR=0.21; $p=3.49\times10^{-13}$) and ANXA11 (OR=0.16; $p=1.09\times10^{-7}$) were associated with decreased risk, respectively(refer to Table 1 and Fig.2).

Furthermore, to explore specific circulating proteins for the non-IPF F-ILD subtype and non-IPF ILD subtype, we also obtained datasets labeled "Other interstitial pulmonary diseases with fibrosis" and "Other interstitial pulmonary diseases" from the UK Biobank and GWAS Catalog databases for MR analysis. Interestingly, we observed that KLRF1 had a relatively low p-value ($p = 3.10 \times 10^{-5}$) with the non-IPF F-ILD dataset from the UKB, although it exceeded the threshold of $P < 2.64 \times 10^{-5}$. Nevertheless, we deemed it necessary to report this finding. Furthermore, no other significant results were found. Additionally, some previously described biomarkers for IPF(8), specifically FUT3_FUT5, were also evaluated in this MR study (the protein quantification method in the UKB-PPP study categorized FUT3 and FUT5 under the same cis-pQTL) (refer to Table 1a).

[1]Table 1a-1b can be placed here.

## External validation of potential drug targets

To replicate the preliminary findings across various outcome GWAS datasets, we adopted both the same-variant and significant-variant approaches(Additional file 1: Table S3). Notably, significant associations for BRSK2 were identified in both ILD and IPF GWAS Catalog cohorts using two strategies. Specifically, employing independent significant pQTL documented by Zhang et al. as genetic instruments(15), elevated expression of BRSK2 was associated with an increased risk of ILD (OR=2.43; 95%CI 1.59-3.72; $p=4.37\times10^{-5}$). For ADAM15 and CDH15, significance was only in the ILD GWAS Catalog cohort using the same-variant strategy ($P < 0.05$). As no other independent significant pQTLs were for LRRC37A2, the

original pQTL continued, still significantly associated in both external cohorts for ILD and IPF. Furthermore, LTBR showed significant associations in the sarcoidosis UK Biobank cohort using the significant-variant strategy, while VEGFB and ANXA11 showed associations using the same-variant strategy. However, ANGPTL4 did not yield notable results upon external validation (Fig.3).

The figures visually show the differences for specific proteins in OR and 95%CI between the primary and replicated analyses. Therefore, we conducted heterogeneity tests separately for the MR analysis results of each protein and found that almost all exhibited high heterogeneity ($I^2$>75%) (Additional file 2: Table S1). Given the larger sample size and higher credibility of evidence from the primary analysis, we relied on it for the risk ratio of proteins to the disease.

## Sensitivity analysis for causal proteins

Firstly, in the bidirectional MR analysis, we observed a genetically predicted significant inverse causal association between ILD and IPF and BRSK2 ILD:$\beta_{IVW}$=0.145 p=2.28×$10^{-4}$; IPF: $\beta_{IVW}$=0.083 p=2.56×$10^{-3}$ , Further Q-tests and pleiotropy tests revealed evidence of heterogeneity(IPF: $p_Q$=1.15×$10^{-27}$, $p_{pleiotropy}$=0.01; ILD: $p_Q$=1.15×$10^{-18}$, $p_{pleiotropy}$=0.03). Leave-one-out analysis indicated that rs35705950 drove this effect. After removing this genetic instrument variable, no reverse causal association or heterogeneity was observed. In addition to BRSK2, sarcoidosis also demonstrated a reverse causal association with ANGPTL4 ($\beta_{IVW}$=-0.03, p=0.016 , and the absence of heterogeneity($p_Q$=0.12, $p_{pleiotropy}$=0.86) further reduced the likelihood of false-positive results and suggested that this reverse causal effect was not contributed by individual SNP. Furthermore, ILD or its subtypes had no observed causal effects on the remaining six proteins. Steiger filtering further ensured the directionality of the associations (Table 1, Additional file 1: Table S4 and Additional file 2: Fig.S2-3).

Secondly, Bayesian colocalization based on pQTLs strongly indicated that ADAM15 coloc.abf-$PPH_4$=0.997 and CDH15 coloc.abf-$PPH_4$=0.863 share the same variants with ILD, and IPF also shares the same variants with ADAM15 (coloc.abf-$PPH_4$=0.966 . Additionally, three out of the four proteins preliminarily identified for sarcoidosis—VEGFB coloc.abf-$PPH_4$=0.907 , LTBR coloc.abf-$PPH_4$=0.947 , and ANGPTL4 coloc.abf-$PPH_4$=0.957 —received robust support by genetic colocalization analysis ($PPH_4$> 80%) under standard priors and window(± 1Mb). On the other hand, BRSK2, LRRC37A2, and ANXA11 did not share the same variants with ILD or its subtypes, suggesting that the causal associations of these proteins were likely driven by different SNPs within their respective genomic regions (Table1, Fig.4 and Additional file 2: Fig.S4). This phenomenon highlighted the possibility of LD confounding even with cis-pQTL.

Lastly, to mitigate bias stemming from horizontal l pleiotropy, we discovered significant associations between VEGFB (rs660442) and ANGPTL4 (rs2278236) with suspected risks of sarcoidosis through "Phenoscanner" and "LDlink" ($P < 5 \times 10^{-8}$). Specifically, rs660442 was associated with rheumatoid arthritis (RA), and previous clinical observational studies reported a considerable proportion of

sarcoidosis cases accompanied by RA(33). However, Mendelian randomization studies have not confirmed this association. Considering that both mechanisms may be immune-related, we highly suspect RA is one of the risk factors for sarcoidosis and that RA may act as a confounding factor for the positive association between VEGFB and sarcoidosis. Additionally, rs2278236 was found to be associated with high-density lipoprotein cholesterol levels. Previous epidemiological and genetic studies have reported abundant genetic overlap between lipid metabolism and immune-related diseases(34). Therefore, we cannot dismiss the possibility of false-positive results stemming from these two confounding factors. Apart from these specific findings, we did not identify any other proxy instruments, such as ADAM15 (rs11589479), associated with known risk factors for ILD or its subtypes (Table 1b).

Based on the evidence outlined above, we categorized preliminary identified proteins into priority and sub-priority groups. Three proteins (ADAM15, CDH15, and LTBR) passed all tests, making them priority candidates for further exploration as plasma protein markers and potential drug targets. Proteins that passed colocalization analysis but failed other sensitivity analyses, such as ANGPTL4 and VEGFB, or proteins that failed in colocalization analysis, such as LRRC37A2, BRSK2, and ANXA11, are classified as sub-priority.

### PPI and druggability evaluation on the potentials of therapeutic targets

The PPI network suggests potential associations of the priority proteins (Additional file 1: Table S5). Specifically, using STRING (Version12.0), a tool that evaluates the degree of association between proteins based on text mining, co-expression, experimental evidence, and other methods, we found that CDH1, associated with ADAM15, and CDH2, associated with CDH15, were closely related in previous studies related to pulmonary fibrosis.

We attempted to discover a causal association between ADAM15 and CDH15 through single-sample MR analysis using GWAS data from these proteins in the UKB-PPP dataset. We found that ADAM15 has a causal association with CDH15 IVW: $p=0.043$ , while the reverse is invalid. The MR-Egger intercept did not significantly deviate from zero in our study ($p_{pleiotropy}=0.444$) (Fig.5), suggesting no evidence of horizontal pleiotropy. Additionally, ADAM15 showed bidirectional causal associations with CDH1 (IVW: $p < 0.05$). These findings suggest potential causal associations between these proteins at the plasma level, and in particular the unidirectional causal association of ADAM15 with CDH15 is of interest. Due to the presence of heterogeneity ($p_Q<0.05$), we used IVW's random-effects model to minimize the effect on the results (Additional file 1: Table S6 and Additional file 2: Fig.S5-6).

Additionally, we discovered reliable interactions between ADAM15 and Src tyrosine kinase (SRC), HCK, and CDH15 and CTNNB1 (with a minimum required interaction score threshold >0.7). Some of these proteins have been previously identified as targets for current treatments of ILD and IPF. Specifically, SRC is one of the targets for Nintedanib(35), a drug recommended in the current IPF treatment guidelines. Furthermore, the mechanism by which increased expression levels of CDH1 delay the progression of pulmonary fibrosis may be explained by the targeted regulation of YTHDF2 by miR-494-3p to delay the

epithelial-mesenchymal transition (EMT)(36). Additionally, studies have reported that the anti-fibrotic function of CDH15 with the NRF2 activator sulforaphane (SFN) may be achieved by inhibiting the expression of CDH2(37). These findings suggest substantial evidence demonstrating a potential association and feasibility of ADAM15 and CDH15 with ILD in terms of molecular mechanisms and drug development. Furthermore, for LTBR, we found that it has the strongest correlation with LTB.

# Discussion

Recent observational reports have indicated an improvement in the survival rates of IPF over the past 20 years, possibly attributed to early diagnosis and the approval of pirfenidone and nintedanib as standard-of-care (SoC) treatments for IPF globally almost a decade ago(38, 39). However, due to the lack of curative treatments and therapies to improve symptoms, the prognosis and quality of life for patients with IPF and other subtypes of ILD, especially those with progressive disease, remain poor. For patients with sarcoidosis who require systemic treatment or do not respond to conventional therapy, more rational and effective treatment strategies are still being explored. Most recent ILD clinical trials are still in phases II and III, and preliminary reports have not shown any remarkable efficacy in improving survival rates and quality of life compared to previous treatment outcomes(40–42). Additionally, many studies had been halted or discontinued due to significant adverse reactions(43, 44). These circumstances reflect the limited understanding of the pathogenesis of ILD and the urgent need for interventions that can improve the prognosis of patients with ILD and its subtypes. It encourages us to continue to deepen our understanding of the mechanisms of these complex diseases and to develop drugs based on more reliable evidence.

Based on exploring potential drug targets and circulating protein biomarkers for ILD and its subtypes, we collected datasets for ILD and its subtypes as comprehensively as possible. We conducted a series of causal association and sensitivity analyses, including plasma proteomic two-sample MR analysis and Bayesian colocalization techniques. The primary MR analysis identified eight circulating proteins significantly causally associated with ILD or its subtypes, among which the causative proteins for sarcoidosis were utterly different from those for ILD and IPF. It undoubtedly provided evidence of heterogeneity in circulating protein levels between the sarcoidosis population and common subtypes of ILD, including IPF. Combining epidemiological data suggested that sarcoidosis accounted for a considerably small proportion of the ILD dataset compared to IPF. Elevated levels of one protein predicted by genetics and the lower level of two may be associated with increased susceptibility to ILD and IPF, along with the higher level of CDH15 possibly being linked to increased susceptibility to ILD. Meanwhile, the lower levels of three proteins and the higher level of one may be associated with an increased risk of sarcoidosis. The replication stage of MR analysis used different datasets and validated six of the eight candidate proteins identified in the primary. Bayesian colocalization of protein expression levels highlighted a minimal potential for confounding due to linkage disequilibrium for ADAM15, CDH15, LTBR, ANGPTL4, and VEGFB causal effects (coloc. abf- $PPH_4 > 0.80$). Furthermore, considering that the causal relationships identified by MR may also be due to reverse causality(45), bidirectional MR was conducted

for all eight potential causal proteins in this study. Only BRSK2 and ANGPTL4 showed reverse causality. Given the significance of reverse causality in assessing drug target potential, ANGPTL4 was categorized as a sub-priority despite passing colocalization analysis. However, it still holds promise as a supplementary circulating protein biomarker for distinguishing sarcoidosis.

Based on the comprehensive analysis, considering the solid independent causal association of ADAM15 with ILD and IPF at the protein quantity expression level, as well as the evidence suggesting no apparent confounding factors, we hypothesize that the genetic effects of ADAM15 span a wider spectrum among ILD patients compared to CDH15, suggesting the potential coverage of ILD subtypes beyond sarcoidosis. However, it remains plausible that the increased prominence of ADAM15 may be attributed to the higher prevalence of IPF within ILD in the FinnGen dataset utilized. In any case, ADAM15 could be prioritized as a potential therapeutic target for treating ILD, especially IPF. As for CDH15, which is associated with ILD but not with IPF, it is likely contributing to ILD subtypes other than IPF. However, this possibility still needs to be verified since there are currently no GWAS datasets subdividing ILD subtypes. Nevertheless, CDH15 still holds promise as a potential therapeutic target for treating ILD, particularly other subtypes aside from IPF. At the same time, LTBR could be prioritized as a potential drug target for sarcoidosis, another prognostically favourable and markedly heterogeneous subtype of ILD.

Additionally, GE, which did not pass colocalization analysis or other sensitive analysis, can serve as sub-priority as potential drug development targets and circulating protein markers for ILD, IPF or sarcoidosis. It is worth mentioning that whether LRRC37A2, BRSK2, and ADAM15 serve as specific indicators for IPF or are broadly present in patients across various ILD subtypes remains to be determined in our study, necessitating the emergence of more ILD subtype datasets in the future for clarification. This observation is particularly relevant given the ongoing clinical demand for simple and minimally invasive adjunct diagnostic approaches to differentiate IPF. Finally, we could not replicate previous findings from studies resembling IPF in our preliminary analysis. This discrepancy could be attributed to differences in the methods used to detect pQTL and variations in the IPF datasets.

A thorough literature review revealed that six out of the eight circulating proteins identified as causally linked in this study had not previously been directly associated with ILD or its subtypes. These proteins include ADAM15, CDH15, LRRC37A2, BRSK2, ANGPTL4, and VEGFB. While ANXA11 had previously been widely reported to be associated with sarcoidosis in terms of genetic polymorphism(46), our study provides additional evidence at the proteomic level. Furthermore, VEGFB belongs to the VEGF (vascular endothelial growth factor) family, with scarce individual descriptions of its association with sarcoidosis, whereas another member of the family, VEGFA, is quite the opposite. Multiple clinical observational studies have found significant differences in the levels of VEGF and VEGFA in the serum and bronchoalveolar lavage fluid of patients with sarcoidosis compared to healthy controls, and these levels are also associated with the progression of sarcoidosis(47, 48). Current research has reported that the severity of inflammation related to granulomatous mouse models can be improved by drug or gene inhibition of VEGFA(49). Strangely, our study results pointed towards VEGFB, rather than VEGFA, as significantly associated with sarcoidosis at the circulating protein level. This situation may be explained

by our research findings, which suggest that its adverse effects on sarcoidosis are associated with decreased levels of circulating proteins. Given the extensive research on VEGFA and the limited understanding of VEGFB(50), there may be overlooked implications, suggesting that future research on sarcoidosis could benefit from comparing VEGFB with VEGFA.

Given the priority of ADAM15, CDH15, and LTBR, we primarily conducted pharmacological assessments on these potential drug targets. ADAM15 is a transmembrane-anchored multi-domain protein belonging to the disintegrin metalloproteinase family(51). Increasing evidence suggests that Src family kinases (SFKs), including SRC and HCK, which are strongly correlated with ADAM15, are involved in the pathogenesis of pulmonary fibrosis. For instance, Src kinase can promote fibrosis by mediating fibroblast adhesion and invasion(52, 53), facilitating alveolar bronchialization(54), inducing capsule formation(55), and promoting extracellular matrix (ECM) deposition(35). The effects of specific Src kinase inhibitors - Nintedanib, Bosutinib, and Saracatinib have been validated in numerous in vitro or in vivo experiments(56, 57). ADAM15 has been found to regulate SRC in the inflammatory synovial tissue of RA, which can convert the apoptosis signal induced by FasL into triggering SRC/FAK phosphorylation to resist fibroblast apoptosis(58, 59). One of the characteristics of IPF is the abnormal accumulation of fibroblasts(60). Therefore, whether similar mechanisms occur in interstitial lung disease fibrosis or whether they are one of the mechanisms of Src kinase inhibition is worthy of further investigation. In addition, CDH1 strongly associated with ADAM15, and CDH2 strongly associated with CDH15 have also been found to play significant roles in IPF. CDH1 and CDH2 are both calcium-dependent cell adhesion proteins(60). However, whereas CDH1 promotes cell-cell adhesion and is a crucial molecule for maintaining epithelial cell phenotype(61), CDH2 acts oppositely. Numerous studies indicate that a hallmark of one of the primary pathological mechanisms in IPF, epithelial-mesenchymal transition (EMT), is the loss of E-cadherin (E-cad, encoded by CDH1 mRNA, an epithelial marker), along with increased expression of Vimentin (a mesenchymal marker) and N-cadherin (N-cad, encoded by CDH2 mRNA)(62–64). In this regard, targeting YTHDF2 with miR-494-3p can enhance the expression level of CDH1(35). On the other hand, the NRF2 activator sulforaphane (SFN) can inhibit the expression of CDH2(37), thus exerting anti-fibrotic functions. Our research also indicated a close connection between ADAM15 and the CDH family. Given the limited studies directly linking ADAM15 and CDH15 to ILD and IPF, we propose a novel hypothesis that ADAM15 may occupy an upstream position in the ILD cascade. It could exert its effects by regulating the expression of CDH family members, such as CDH15, CDH1, and CDH2, or phosphorylating the Src kinase family. Future studies should focus on and test this hypothesis.

LTBR is a unique receptor for LTB and belongs to the tumor necrosis factor (TNF) ligand family. Previous reports have shown significantly increased expression of LTB in major chronic inflammatory diseases characterized by lymphocytic infiltration, such as sarcoidosis(65). An intriguing finding previously reported is that LTBR is exclusively present in the epithelial cells lining the bronchial lumen, while its homolog, LTB, is predominantly distributed in granulomas and surrounding lymphocytes(66). Our study found that elevated circulating levels of LTBR lead to increased susceptibility to sarcoidosis. These associative findings are worth further investigation in the future. In summary, whether LTBR holds promise as a circulating biomarker for aiding in the differential diagnosis of sarcoidosis requires further

investigation. At the same time, LTB may also play a role in the sustained inflammation observed in sarcoidosis granulomas, making it a potential therapeutic target. Another related protein, TNF-α, is inhibited by Infliximab, which has been developed as a third-line treatment for clinical management of sarcoidosis, primarily used for severe cases where organ damage poses a threat to life and when glucocorticoids or other immunosuppressants are ineffective(67, 68).

Like other Mendelian randomization (MR) studies, our approach may also have inevitable limitations. MR results may be biased due to potential violations of its assumptions, such as the complex interplay between genetics and environment. However, our study design aimed to minimize explicit biases by using cis-pQTL to reduce potential horizontal pleiotropy and conducting rigorous sensitivity analyses to assess potential pleiotropic effects, excluding any initial screening proteins that did not meet these criteria. Additionally, we found high heterogeneity in the majority of the externally validated results. We attribute this to the difference in the datasets and methods of detective pQTL used in the primary and replication analyses. Given the current predominance of proteomic genetic data derived from the aptamer detection method and the scarcity of large-scale antibody-based detection methods, this hypothesis could not be thoroughly validated. The observed heterogeneity might also stem from the substantial diversity among ILD disease subtypes, leading to the inclusion of disparate populations across different datasets. At the same time, the lack of datasets for various ILD subtypes, especially iNSIP and CTD-ILD, and the small sample sizes in datasets such as SAD-ILD and non-IPF F-ILD limited our study results. As future studies incorporate more samples representing diverse ILD subtypes, refined subtype-specific GWAS can address these issues and enable more precise Mendelian randomization analyses. Such efforts are pivotal for advancing precision medicine, intending to optimize the efficacy of future therapies through targeted therapy. Lastly, research on ILD among non-European ethnic populations was limited, underscoring the need for further investigations to ascertain the generalizability of these findings.

# Conclusions

In conclusion, employing the latest characterization of genetic structure in plasma proteomics and conducting Mendelian randomization (MR) analyses on interstitial lung disease (ILD) and its multiple subtypes, along with comprehensive analyses including Bayesian colocalization, we identified several circulating proteins causally associated with ILD or its subtypes. We identified ADAM15, CDH15, and LTBR as priority biomarkers and potential therapeutic targets in circulation, particularly highlighting ADAM15. Further experimental studies are warranted to evaluate the efficacy of these candidate proteins.

# Declarations

### Ethics approval and consent to participate

Not applicable.

### Competing interests

All authors declare no competing interests.

## Consent for publication

Not applicable.


## Acknowledgements

We would like to thank the investigators and participants of all GWAS from which we used summary statistics data.


## Author contribution

K.Y., W.L, and J.L. designed the study.

K.Y., W.L., Y.L., Y.L., and H.L. analyzed the data.

K.Y., H.L., and J.L. interpreted the data.

K.Y. and W.L wrote the manuscript.

All authors read and approved the final manuscript.


## Funding

This study was supported by the Scientific Research Project of Guangdong Provincial Bureau of Traditional Chinese Medicine (grant number 2022ZYYJ01)


## Data Availability

GWAS catalog(https://www.ebi.ac.uk/gwas/), FinnGen(https://www.finngen.fi/en/), and UK Biobank (http://www.nealelab.is/uk-biobank), the original studies for plasma pQTL with the PMID listed in Additional file 1: Table S1.

## Authors' information


Kunrong Yu ykgoose@163.com

Wanying Li 821090419@qq.com

Wenjie Long lwj5699@gzucm.edu.cn

Yijia Li liyijia@stu.gzucm.edu.cn

Yanting Li liyanting2000216@163.com

Huili Liao Lilitcm@163.com

Jianhong Liu ljh833@163.com

# Abbreviations

| | |
|---|---|
| ABPA | allergic bronchopulmonary aspergillosis |
| ADAM15 | A disintegrin and metalloproteinase 15 |
| ANGPTL4 | Angiopoietin-related protein 4 |
| ANXA11 | annexin A11 |
| BRSK2 | Serine/threonine-protein kinase BRSK2 |
| CDH1 | Cadherin-1 |
| CDH2 | Cadherin-2 |
| CDH15 | Cadherin-15 |
| CI | confidence interval |
| cis | proximate |
| CTD | connective tissue disease |
| CTD-ILD | connective tissue disease associated ILD |
| CTNNB1 | Catenin beta-1 |
| EMT | epithelial-mesenchymal transition |
| F-ILD | fibrosing interstitial lung diseases |
| FUT3 | Galactoside 3(4)-L-fucosyltransferase |
| FUT5 | Alpha-(1,3)-fucosyltransferase 5 |
| GWAS | Genome-wide association study |
| HCK | Tyrosine-protein kinase HCK |
| HP | hypersensitivity pneumonitis |
| ILD | Interstitial lung disease |
| iNSIP | idiopathic nonspecific interstitial pneumonia |
| IPF | Idiopathic pulmonary fibrosis |
| IVW | Inverse Variance Weighted |
| KLRF1 | Killer cell lectin-like receptor subfamily F member 1 |
| LD | linkage disequilibrium |
| LRRC37A2 | Leucine-rich repeat-containing protein37A2 |
| LTB | Lymphotoxin-beta |
| LTBR | Lymphotoxin-beta receptor |

| | |
|---|---|
| MHC | major histocompatibility complex |
| MR | Mendelian randomization analysis |
| non-IPF F-ILD | non-IPF fibrosing ILD |
| OR | risk ratio |
| PF | progressive fibrosis |
| PF-ILD | progressive fibrosing ILD |
| PPF | progressive pulmonary fibrosis |
| $PPH_4$ | posterior probability of hypothesis 4 |
| PPI | protein-protein interaction |
| pQTL | protein quantitative trait loci |
| RA | rheumatoid arthritis |
| SAD-ILD | ILD related to systemic autoimmune disease |
| SD | standard deviation |
| SFKs | Src family kinases |
| SNP | single nucleotide polymorphism |
| SoC | standard-of-care |
| SRC | Src tyrosine kinase |
| SSc | systemic sclerosis |
| SSc-ILD | systemic sclerosis-related ILD |
| TNF | tumor necrosis factor |
| trans | distant |
| UKB | The UK Biobank |
| UKB-PPP | UK Biobank Pharma Proteomics Project |
| VEGF | vascular endothelial growth factor |
| VEGFA | vascular endothelial growth factor |
| VEGFB | vascular endothelial growth factor B |

# References

1. Hilberg O, Hoffmann-Vold AM, Smith V, Bouros D, Kilpeläinen M, Guiot J, et al. Epidemiology of interstitial lung diseases and their progressive-fibrosing behaviour in six European countries. ERJ

Open Res. 2022 Jan;8(1):00597–2021.
2. M R, P D, S K, Ev A. Sarcoidosis: Epidemiology and clinical insights. J Intern Med. 2023 Jun;293(6).
3. Belperio JA, Shaikh F, Abtin FG, Fishbein MC, Weigt SS, Saggar R, et al. Diagnosis and Treatment of Pulmonary Sarcoidosis: A Review. JAMA. 2022 Mar 1;327(9):856–67.
4. G R, M RJ, L R, Cc T, Y I, T J, et al. Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2022 Jan 5;205(9).
5. Buschulte K, Kabitz HJ, Hagmeyer L, Hammerl P, Esselmann A, Wiederhold C, et al. Disease trajectories in interstitial lung diseases - data from the EXCITING-ILD registry. Respir Res. 2024 Mar 6;25(1):113.
6. Mr N, H T, Jl P, J S, P N, Y S, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015 Aug;47(8).
7. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet. 2019 Dec;15(12):e1008489.
8. Nakanishi T, Cerani A, Forgetta V, Zhou S, Allen RJ, Leavy OC, et al. Genetically increased circulating FUT3 level leads to reduced risk of idiopathic pulmonary fibrosis: a Mendelian randomisation study. Eur Respir J. 2022 Feb;59(2):2003979.
9. Katz DH, Robbins JM, Deng S, Tahir UA, Bick AG, Pampana A, et al. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. Sci Adv. 2022 Aug 19;8(33):eabm5164.
10. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. Nat Commun. 2021 Nov 24;12(1):6822.
11. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023 Oct;622(7982):329–38.
12. E S, A M, A A, A B, M C, L G, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023 Jun 1;51(D1).
13. UK Biobank — Neale lab. http://www.nealelab.is/uk-biobank. Accessed 1 August 2018.
14. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023 Jan;613(7944):508–18.
15. Zhang J, Dutta D, Köttgen A, Tin A, Schlosser P, Grams ME, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. Nat Genet. 2022 May;54(5):593–602.
16. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018 Jun;558(7708):73–9.

17. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun. 2018 Aug 15;9(1):3268.
18. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, et al. Mapping the proteo-genomic convergence of human diseases. Science. 2021 Nov 12;374(6569):eabj1541.
19. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021 Dec;53(12):1712–21.
20. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. JAMA. 2021 Oct 26;326(16):1614–21.
21. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018 Jul 12;362:k601.
22. Lin SH, Brown DW, Machiela MJ. LDtrait: An Online Tool for Identifying Published Phenotype Associations in Linkage Disequilibrium. Cancer Res. 2020 Aug 15;80(16):3443–6.
23. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res. 2019;4:186.
24. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife. 2018 May 30;7:e34408.
25. Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. Genet Epidemiol. 2017 Dec;41(8):714–25.
26. Deng YT, Ou YN, Wu BS, Yang YX, Jiang Y, Huang YY, et al. Identifying causal genes for depression via integration of the proteome and transcriptome from brain and blood. Mol Psychiatry. 2022 Jun;27(6):2849–57.
27. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021 Oct;53(10):1415–24.
28. Duckworth A, Gibbons MA, Allen RJ, Almond H, Beaumont RN, Wood AR, et al. Telomere length and risk of idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease: a mendelian randomisation study. Lancet Respir Med. 2021 Mar;9(3):285–94.
29. Jiang L, Zheng Z, Fang H, Yang J. A generalized linear mixed model association tool for biobank-scale data. Nat Genet. 2021 Nov;53(11):1616–21.
30. Folkersen L, Fauman E, Sabater-Lleal M, Strawbridge RJ, Frånberg M, Sennblad B, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS Genet. 2017 Apr;13(4):e1006706.
31. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet.

2014 May;10(5):e1004383.

32. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. Nat Genet. 2019 May;51(5):768–9.
33. Suzuki A, Kamio K, Takeno M, Terasaki Y, Taniuchi N, Sato J, et al. Pulmonary sarcoidosis complicated by rheumatoid arthritis in a patient presenting with progressive fibrosing interstitial lung disease and treated with nintedanib: a case report and literature review. Ther Adv Respir Dis. 2023;17:17534666231158279.
34. Andreassen OA, Desikan RS, Wang Y, Thompson WK, Schork AJ, Zuber V, et al. Abundant genetic overlap between blood lipids and immune-mediated diseases indicates shared molecular genetic mechanisms. PloS One. 2015;10(4):e0123057.
35. Li H, Zhao C, Tian Y, Lu J, Zhang G, Liang S, et al. Src family kinases and pulmonary fibrosis: A review. Biomed Pharmacother Biomedecine Pharmacother. 2020 Jul;127:110183.
36. Ning J, Du H, Zhang Y, Liu Q, Jiang T, Pang Y, et al. N6-Methyladenosine Modification of CDH1 mRNA Promotes PM2.5-Induced Pulmonary Fibrosis via Mediating Epithelial Mesenchymal Transition. Toxicol Sci Off J Soc Toxicol. 2022 Jan 24;185(2):143–57.
37. Liu P, Luo G, Dodson M, Schmidlin CJ, Wei Y, Kerimoglu B, et al. The NRF2-LOC344887 signaling axis suppresses pulmonary fibrosis. Redox Biol. 2021 Jan;38:101766.
38. Guenther A, Krauss E, Tello S, Wagner J, Paul B, Kuhn S, et al. The European IPF registry (eurIPFreg): baseline characteristics and survival of patients with idiopathic pulmonary fibrosis. Respir Res. 2018 Jul 28;19(1):141.
39. Jegal Y, Park JS, Kim SY, Yoo H, Jeong SH, Song JW, et al. Clinical Features, Diagnosis, Management, and Outcomes of Idiopathic Pulmonary Fibrosis in Korea: Analysis of the Korea IPF Cohort (KICO) Registry. Tuberc Respir Dis. 2022 Apr;85(2):185–94.
40. Raghu G, van den Blink B, Hamblin MJ, Brown AW, Golden JA, Ho LA, et al. Long-term treatment with recombinant human pentraxin 2 protein in patients with idiopathic pulmonary fibrosis: an open-label extension study. Lancet Respir Med. 2019 Aug;7(8):657–64.
41. Kolb M, Crestani B, Maher TM. Phosphodiesterase 4B inhibition: a potential novel strategy for treating pulmonary fibrosis. Eur Respir Rev Off J Eur Respir Soc. 2023 Mar 31;32(167):220206.
42. Richeldi L, Azuma A, Cottin V, Hesslinger C, Stowasser S, Valenzuela C, et al. Trial of a Preferential Phosphodiesterase 4B Inhibitor for Idiopathic Pulmonary Fibrosis. N Engl J Med. 2022 Jun 9;386(23):2178–87.
43. Maher TM, van der Aar EM, Van de Steen O, Allamassey L, Desrivot J, Dupont S, et al. Safety, tolerability, pharmacokinetics, and pharmacodynamics of GLPG1690, a novel autotaxin inhibitor, to treat idiopathic pulmonary fibrosis (FLORA): a phase 2a randomised placebo-controlled trial. Lancet Respir Med. 2018 Aug;6(8):627–35.
44. Palmer SM, Snyder L, Todd JL, Soule B, Christian R, Anstrom K, et al. Randomized, Double-Blind, Placebo-Controlled, Phase 2 Trial of BMS-986020, a Lysophosphatidic Acid Receptor Antagonist for the Treatment of Idiopathic Pulmonary Fibrosis. Chest. 2018 Nov;154(5):1061–9.

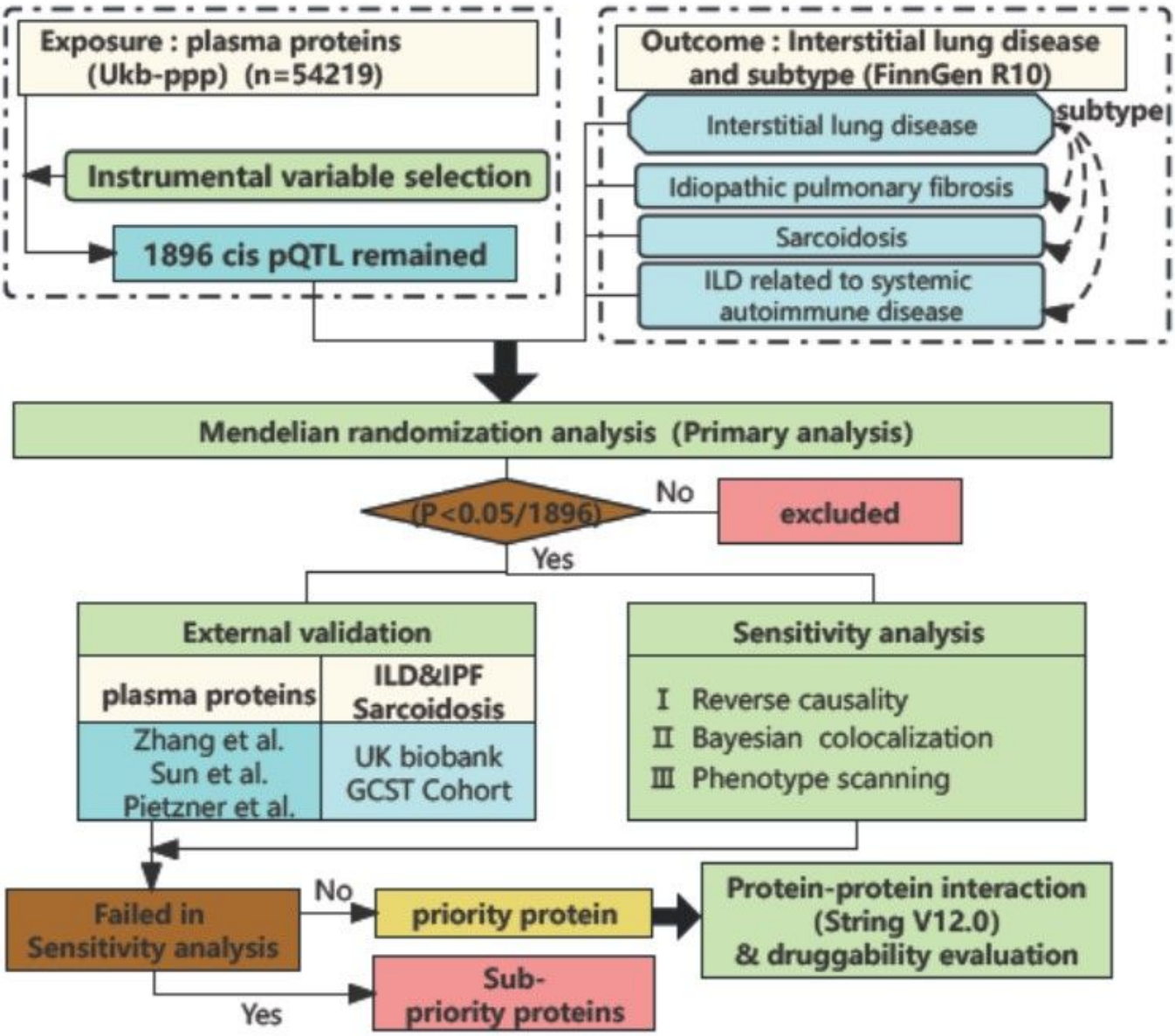
45. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet. 2020 Oct;52(10):1122–31.
46. Karakaya B, van der Vis JJ, Veltkamp M, Biesma DH, Grutters JC, van Moorsel CHM. ANXA11 rs1049550 Associates with Löfgren's Syndrome and Chronic Sarcoidosis Patients. Cells. 2022 May 5;11(9):1557.
47. Yamashita M, Mouri T, Niisato M, Kowada K, Kobayashi H, Chiba R, et al. Heterogeneous characteristics of lymphatic microvasculatures associated with pulmonary sarcoid granulomas. Ann Am Thorac Soc. 2013 Apr;10(2):90–7.
48. Tuleta I, Biener L, Pizarro C, Nickenig G, Skowasch D. Proangiogenic and Profibrotic Markers in Pulmonary Sarcoidosis. Adv Exp Med Biol. 2018;1114:57–66.
49. Harding JS, Herbath M, Chen Y, Rayasam A, Ritter A, Csoka B, et al. VEGF-A from Granuloma Macrophages Regulates Granulomatous Inflammation by a Non-angiogenic Pathway during Mycobacterial Infection. Cell Rep. 2019 May 14;27(7):2119-2131.e6.
50. Lee C, Chen R, Sun G, Liu X, Lin X, He C, et al. VEGF-B prevents excessive angiogenesis by inhibiting FGF2/FGFR1 pathway. Signal Transduct Target Ther. 2023 Aug 18;8(1):305.
51. Edwards DR, Handsley MM, Pennington CJ. The ADAM metalloproteinases. Mol Aspects Med. 2008 Oct;29(5):258–89.
52. Okutani D, Lodyga M, Han B, Liu M. Src protein tyrosine kinase family and acute inflammatory responses. Am J Physiol Lung Cell Mol Physiol. 2006 Aug;291(2):L129-141.
53. Oyaizu T, Fung SY, Shiozaki A, Guan Z, Zhang Q, dos Santos CC, et al. Src tyrosine kinase inhibition prevents pulmonary ischemia-reperfusion-induced acute lung injury. Intensive Care Med. 2012 May;38(5):894–905.
54. Jaeger B, Schupp JC, Plappert L, Terwolbeck O, Artysh N, Kayser G, et al. Airway basal cells show a dedifferentiated KRT17highPhenotype and promote fibrosis in idiopathic pulmonary fibrosis. Nat Commun. 2022 Sep 26;13(1):5637.
55. Barbayianni I, Kanellopoulou P, Fanidis D, Nastos D, Ntouskou ED, Galaris A, et al. SRC and TKS5 mediated podosome formation in fibroblasts promotes extracellular matrix invasion and pulmonary fibrosis. Nat Commun. 2023 Sep 21;14(1):5882.
56. Beyer C, Distler JHW. Tyrosine kinase signaling in fibrotic disorders: Translation of basic research to human disease. Biochim Biophys Acta. 2013 Jul;1832(7):897–904.
57. Wang J, Zhuang S. Src family kinases in chronic kidney disease. Am J Physiol Renal Physiol. 2017 Sep 1;313(3):F721–8.
58. Janczi T, Böhm BB, Fehrl Y, DeGiacomo P, Kinne RW, Burkhardt H. ADAM15 in Apoptosis Resistance of Synovial Fibroblasts: Converting Fas/CD95 Death Signals Into the Activation of Prosurvival Pathways by Calmodulin Recruitment. Arthritis Rheumatol Hoboken NJ. 2019 Jan;71(1):63–72.
59. Böhm BB, Freund I, Krause K, Kinne RW, Burkhardt H. ADAM15 adds to apoptosis resistance of synovial fibroblasts by modulating focal adhesion kinase signaling. Arthritis Rheum. 2013

Nov;65(11):2826–34.
60. Black M, Milewski D, Le T, Ren X, Xu Y, Kalinichenko VV, et al. FOXF1 Inhibits Pulmonary Fibrosis by Preventing CDH2-CDH11 Cadherin Switch in Myofibroblasts. Cell Rep. 2018 Apr 10;23(2):442–58.
61. Marmai C, Sutherland RE, Kim KK, Dolganov GM, Fang X, Kim SS, et al. Alveolar epithelial cells express mesenchymal proteins in patients with idiopathic pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol. 2011 Jul;301(1):L71-78.
62. Giannelli G, Koudelkova P, Dituri F, Mikulits W. Role of epithelial to mesenchymal transition in hepatocellular carcinoma. J Hepatol. 2016 Oct;65(4):798–808.
63. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest. 2003 Dec;112(12):1776–84.
64. Tang W, Du L, Sun W, Yu Z, He F, Chen J, et al. Maternal exposure to fine particulate air pollution induces epithelial-to-mesenchymal transition resulting in postnatal pulmonary dysfunction mediated by transforming growth factor-β/Smad3 signaling. Toxicol Lett. 2017 Feb 5;267:11–20.
65. Agyekum S, Church A, Sohail M, Krausz T, Van Noorden S, Polak J, et al. Expression of lymphotoxin-beta (LT-beta) in chronic inflammatory conditions. J Pathol. 2003 Jan;199(1):115–21.
66. Boussaud V, Soler P, Moreau J, Goodwin RG, Hance AJ. Expression of three members of the TNF-R family of receptors (4-1BB, lymphotoxin-beta receptor, and Fas) in human lung. Eur Respir J. 1998 Oct;12(4):926–31.
67. Baughman RP, Grutters JC. New treatment strategies for pulmonary sarcoidosis: antimetabolites, biological drugs, and other treatment approaches. Lancet Respir Med. 2015 Oct;3(10):813–22.
68. Belperio JA, Fishbein MC, Abtin F, Channick J, Balasubramanian SA, Lynch Iii JP. Pulmonary sarcoidosis: A comprehensive review: Past to present. J Autoimmun. 2023 Oct 19;103107.

# Tables

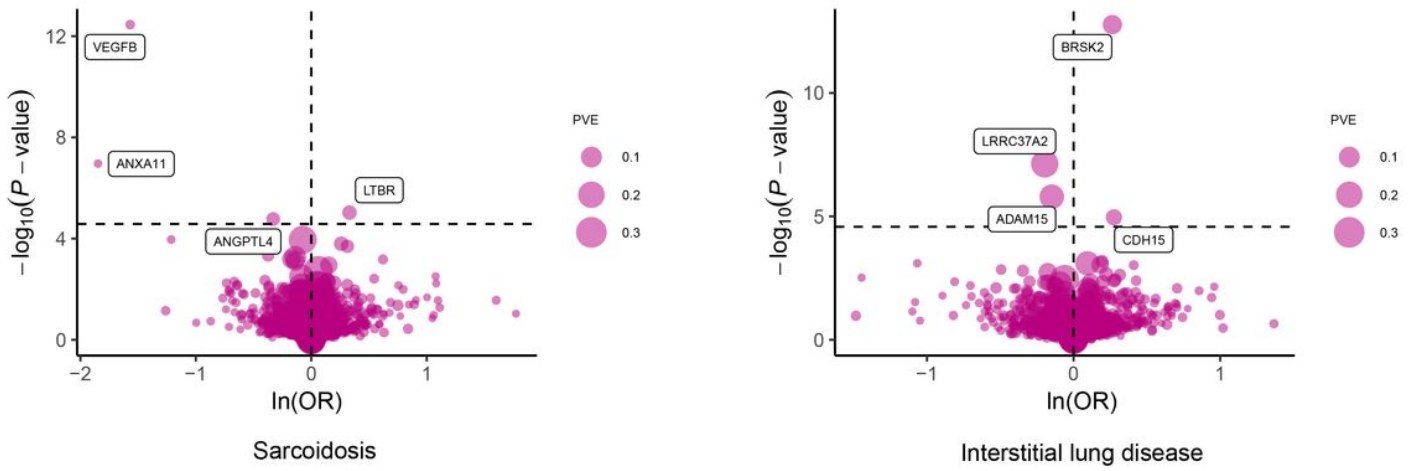
Tables 1a and 1b are available in the Supplementary Files section.

# Figures




## Figure 1

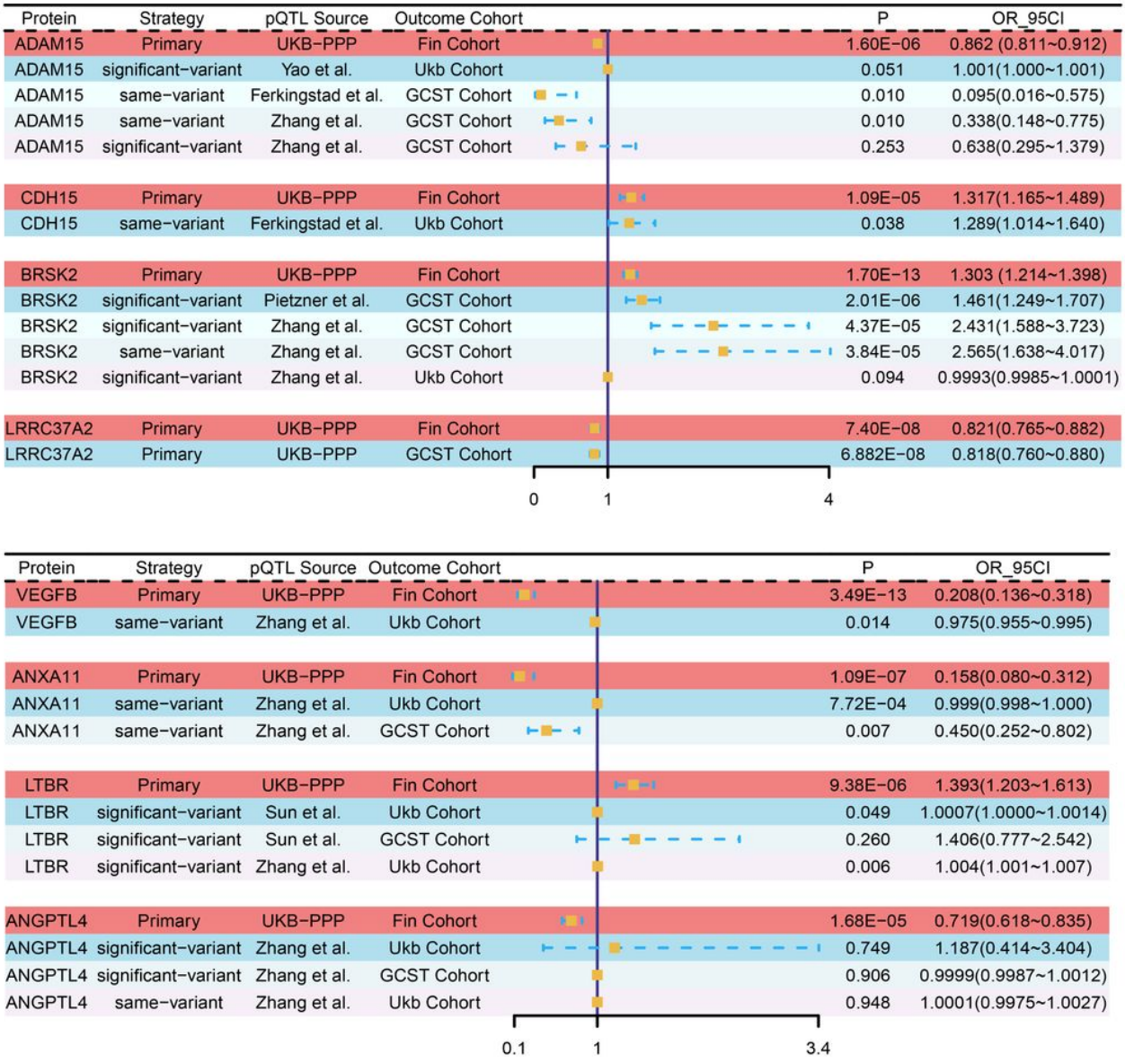
Study design for identification of plasma proteins causally associated with ILD and its subtypes




## Figure 2

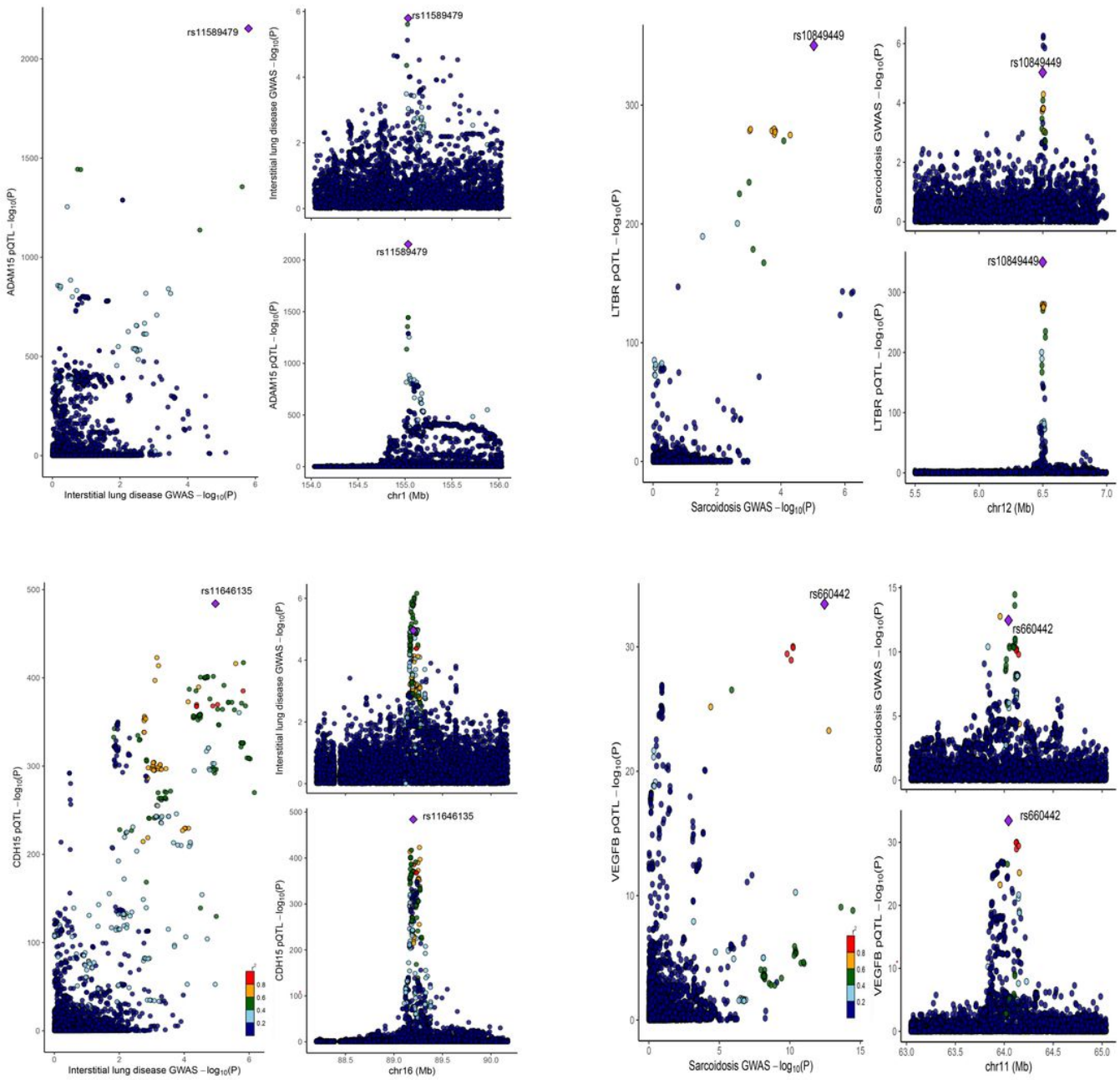
Volcano plots of the MR results for proteins on the risk of (A) Interstitial lung disease and (B) Sarcoidosis.

OR for increased risk of ILD and Sarcoidosis were expressed as per SD increase in plasma protein levels. The dashed horizontal black line corresponded to P = $2.64 \times 10^{-5}$ (0.05/1896). ln = natural logarithm; PVE = proportion of variance explained




## Figure 3

External validation of the causal relationship between potential causal proteins and (A) Interstitial lung disease or (B) Sarcoidosis




## Figure 4

Bayesian colocalization analysis of potential causal proteins and Interstitial lung disease or its subtypes.

Colocalization analysis of ADAM15-ILD(A), CDH15-ILD(B), LTBR-Sarcoidosis(C) and VEGFB-Sarcoidosis(D). Diamond purple points represented the SNP that with the minimal sum of P value in corresponded protein GWAS and ILD or its subtypes GWAS.

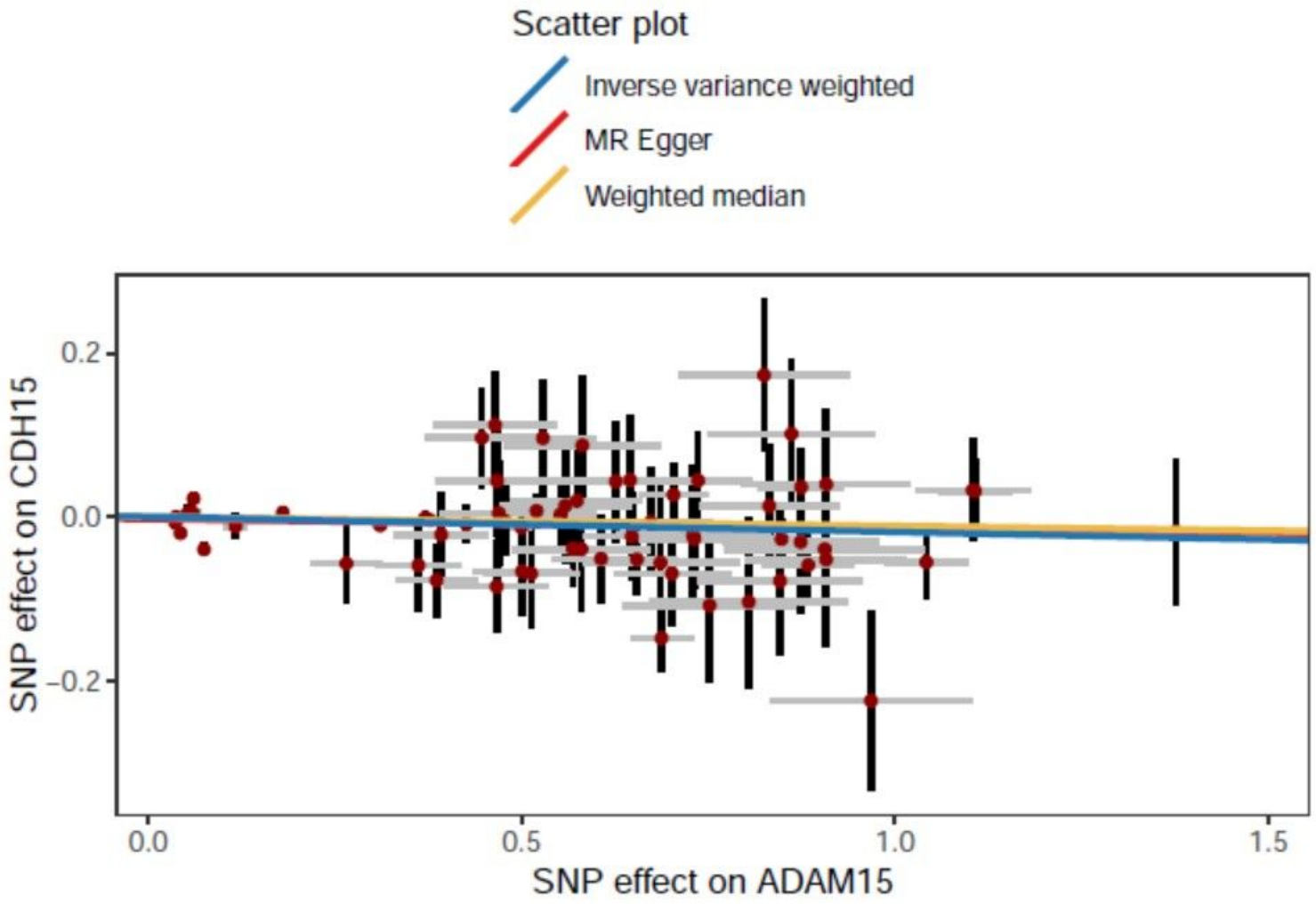



**Figure 5**

Scatter plot to visualize the causal effect of ADAM15 on CDH15.

The slope of the straight line indicates the magnitude of the causal association. IVW indicates inverse-variance weighted.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1a1b.docx
- Additionalfile1.xlsx
- Figure.S1.pdf
- Figure.S2.pdf
- Figure.S3.pdf
- Figure.S4.pdf

- Figure.S5.pdf
- Figure.S6.pdf
- Additionalfile2.docx