



Mendelian randomisation highlights hypothyroidism as a causal determinant of idiopathic pulmonary fibrosis

Yanan Zhang^{a,b}, Meng Zhao^{c,d}, Ping Guo^{a,b}, Yanjun Wang^{a,b}, Lu Liu^{a,b}, Jinghua Zhao^e, Ling Gao^{c,d}, Zhongshang Yuan^{a,b,*}, Fuzhong Xue^{a,b,*}, Jiajun Zhao^{c,d,*}

^a Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, 250012, China

^b Institute for Medical Dataology, Shandong University, Jinan, Shandong, 250003, China

^c Department of Endocrinology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, 250021, China

^d Department of Endocrinology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, 250021, China

^e Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

ARTICLE INFO

Article History:

Received 28 July 2021

Revised 5 October 2021

Accepted 20 October 2021

Available online xxx

Keywords:

Idiopathic pulmonary fibrosis

Hypothyroidism

Mendelian randomisation

GWAS summary statistics

ABSTRACT

Background: Although the association between hypothyroidism and idiopathic pulmonary fibrosis (IPF) is found in observational studies, it remains uncertain whether hypothyroidism causally influences IPF.

Methods: Two-sample Mendelian randomisation (MR) was conducted with hypothyroidism genome-wide association study (GWAS) data in the UK Biobank from 289,307 individuals (18,740 cases and 270,567 controls) and the largest GWAS summary statistics of IPF from 11,259 individuals (2,668 cases and 8,591 controls). Findings were verified using an independent validation dataset, as well as through different MR methods with different model assumptions. A multivariable MR based on Bayesian model averaging was further performed to evaluate whether hypothyroidism, even given several other comorbidities of IPF, remained to be the true causal one of IPF.

Findings: A positive causal effect of hypothyroidism on IPF was revealed (MR inverse-variance weighted [MR-IVW], odds ratio [OR]=1.125, 95% confidence interval [CI] 1.028–1.231; P=0.011), which was further verified in an independent validation set (MR-IVW, OR=1.229, 95% CI 1.054–1.432; P=0.008). The results were consistent from a variety of MR methods. Bidirectional analyses also indicated no reverse causation. Multivariable MR analysis showed hypothyroidism had the strongest marginal evidence (marginal inclusion probability=0.397, false discovery rate=0.025) compared with other comorbidities of IPF.

Interpretation: Our results illustrate the significant causal effect of hypothyroidism on IPF, which holds even given several other comorbidities of IPF. These findings may have an important insight into pathogenesis and possible future therapies of IPF.

Funding: National Natural Science Foundation of China, the Natural Science Foundation of Shandong Province and the Young Scholars Program of Shandong University.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Idiopathic pulmonary fibrosis (IPF) is a condition in which the lungs become scarred and breathing is increasingly difficult, and occurs primarily in older adults with unknown etiology [1] and a median survival of 3.8 years [2]. In recent years, IPF has become more prevalent and the incidence is on the rise [3]. One consensus regarding IPF is that damage to the alveolar epithelium and an aberrant wound-healing response cause widespread deposition of dense fibrotic tissue, reduce the lung compliance, and inhibit gas transfer

[4]. Treatment options in IPF are limited thus far. Consequently, understanding of comorbidities in the IPF patient population, in particular their relationship, can greatly improve clinical practice including quality of life and potentially survival [5].

Comorbidities frequently occurring in IPF include both respiratory (e.g., chronic obstructive pulmonary disease (COPD), obstructive sleep apnea (OSA), pulmonary hypertension (PH), and lung cancer) and non-respiratory (e.g., ischemic heart disease (IHD), atherosclerosis (AS), type 2 diabetes (T2D), gastroesophageal reflux disease (GERD), and hypothyroidism) diseases [5, 6]. Among these, the association between hypothyroidism and IPF is of particular interest: a retrospective case-control study [7] showed that hypothyroidism was significantly associated with IPF even after adjusting for body mass index, smoking, diabetes, GERD, and chronic corticosteroid

* Corresponding authors.

E-mail addresses: yuanzhongshang@sdu.edu.cn (Z. Yuan), xuefzh@sdu.edu.cn (F. Xue), jjzhao@sdu.edu.cn (J. Zhao).

Research in context

Evidence before this study

We searched PubMed on October 3, 2021, with the search terms (“idiopathic pulmonary fibrosis “OR “pulmonary fibrosis”) AND (“hypothyroidism” OR “thyroid function” OR “thyroid hormones”) and without date restrictions. We identified 50 PubMed-indexed articles in total, including in particular a case-control study which reported the association between hypothyroidism and idiopathic pulmonary fibrosis (IPF) after adjusting for body mass index, smoking, diabetes, gastroesophageal reflux, and chronic corticosteroid therapy. Several observational studies also illustrated the relationship between thyroid function and pulmonary fibrosis. The relationship between hypothyroidism and IPF is bound to be controversial due to a range of known and unknown confounders, and the underlying mechanisms may be complex. The causality of hypothyroidism on IPF is yet to be established.

Added value of this study

We establish a causal relationship between hypothyroidism and IPF through a two-sample Mendelian randomisation (MR) framework using large-scale genome-wide association studies (GWASs) summary statistics. Findings were verified using an independent validation dataset, as well as through different MR methods with different model assumptions. Bidirectional analyses also indicated no reverse causation. Multivariable MR analysis showed hypothyroidism had the largest marginal inclusion probability compared with other comorbidities of IPF.

Implications of all the available evidence

Our study was the first attempt shedding light on the directional causal relationship between hypothyroidism and IPF, adding to existing knowledge in etiology of IPF. It will also facilitate revealing pathogenic pathways and planning future studies.

factors and disease outcomes [12]. Two-sample MR estimates and tests for the causal effect of an exposure on an outcome measured in two separate studies with no sample overlap [13].

In this study, we investigate the causal relationship between genetically predicted hypothyroidism and IPF in the MR framework using publicly available GWAS summary statistics. The results can provide additional evidence in the etiology of IPF.

2. Methods

2.1. Study design

A two-sample MR was firstly performed on the bi-directional causal relationship between hypothyroidism and IPF, followed by verification using an independent validation dataset as well as using different MR methods with different model assumptions. A multivariable MR based on Bayesian model averaging (MR-BMA) [14] was further conducted to see whether hypothyroidism remains to be selected as the true causal factor of IPF, given several other comorbidities of IPF. Fig. 1 showed the overview diagram of our study design.

2.2. GWAS data sources and instrument selection in univariable MR

Two-sample MR analyses were conducted using GWAS summary data. In total, we obtained three GWASs of hypothyroidism and one GWAS of IPF. All GWAS studies were restricted to unrelated individuals of European ancestry. In discovery analysis, we used the hypothyroidism GWAS from UK Biobank (UKBB) (data field 20002_1226) [15], which had the largest sample size (N=289,307 with 18,740 cases and 270,567 controls). UKBB is a large-scale population-based cohort study of half a million people aged between 40 and 69 years from the United Kingdom [16]. The hypothyroidism genetic association data were obtained from UKBB release 2 data in the Atlas of GWAS Summary Statistics (<https://atlas.ctglab.nl/>), where hypothyroidism was defined with self-reported history of hypothyroidism/myxedema. In validation analysis, we used the hypothyroidism GWAS data from the large GWAS study on thyroid function and dysfunction [17], with 3,440 cases and 49,983 controls, where hypothyroidism was defined with thyroid stimulating hormone levels above the cohort-specific reference range. In sensitivity analysis, another hypothyroidism GWAS from UKBB (data field 41204_E03), diagnosed with secondary ICD-10 with 244,890 individuals, was further analyzed to assess the robustness against different illness code [15]. GWAS summary statistics on IPF was from the largest GWAS to date [18], derived from a meta-analysis on three previous IPF studies including the Chicago study [19], the Colorado Study [20,21], and the UK study [22]. The meta-analysis obtained a maximum sample size of up to 11,259 individuals (2,668 cases and 8,591 controls) for 10,790,934 well-imputed SNPs with minor allele count >10 in the three studies. Each study was separately conducted adjusting for the first 10 principal components to account for population structure. IPF cases from the studies are strictly diagnosed using guidelines American Thoracic Society and European Respiratory Society guidelines [1,23,24] and had appropriate institutional review board or ethics approval. In addition, sample overlap between two datasets could bias the estimated causal effect [25], thus we conducted linkage disequilibrium score regression (LDSC) to assess sample overlap [26].

The MR framework uses independent instrumental SNPs as instrumental variables (IVs) for the exposure (e.g., hypothyroidism) to estimate and test the causal effect on the outcome (i.e., IPF). A standard MR analysis requires three model assumptions to hold [12] (Fig. 1a): (i) instruments are associated with the exposure of interest; (ii) instruments are not associated with any other confounders that may be associated with both exposure and outcome; (iii) instruments only influence the outcome by the path of exposure. A primary step of MR is to determine appropriate genetic variants to serve as valid

therapy. In addition, several studies also showed the relationship between thyroid function and pulmonary fibrosis [8, 9]. Given the fact that inadequate adjustment for confounding factors can bias the association between hypothyroidism and IPF, a better approach to assess the evidence on their causal relationship is needed, especially when other comorbidities of IPF are considered.

Although randomised controlled trial is considered as a gold standard to evaluate causality, its use is limited with respect to practicality, cost, and ethical considerations [10]. IPF is relatively rare and it is not feasible to collect large samples in longitudinal studies to sufficiently study the relationship between hypothyroidism and IPF. Furthermore, a host of known and unknown confounders also bring one big challenge to investigate the causality of hypothyroidism on IPF. The advent of causal inference between the exposure and the outcome from observational studies offers such possibility. Notably, Mendelian randomisation (MR) is a useful framework to study causality via genetic instruments (e.g., single nucleotide polymorphisms [SNPs]) [11]. It is analogous to a “natural” randomised controlled trial in which the random allocation of exposure-influencing genetic alleles largely rules out the impact from unobserved confounding factors and avoids the reverse causation and measurement errors, which can plague other study designs [11]. Publicly available genome-wide association studies (GWASs) have produced lots of summary-level data, enabling two-sample MR as an efficient and cost-effective method to interrogate the causal relationships among health risk

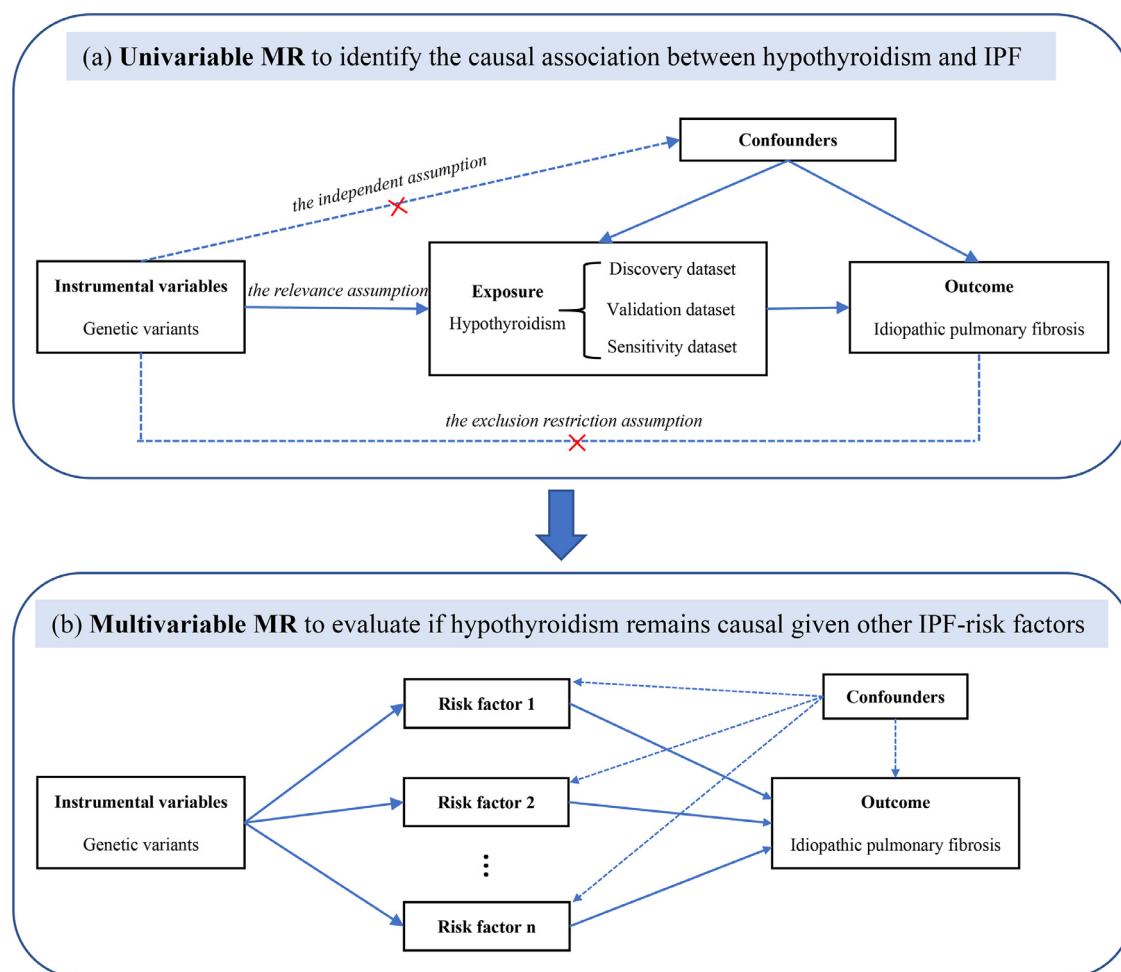


Fig. 1. Study design. (a) The causal diagram for standard Mendelian randomisation (MR) analysis, which involves instrumental variable (IV) and requires three assumptions. (b) Illustrative diagram of IV assumptions made in multivariable MR model. IPF, idiopathic pulmonary fibrosis; MR, Mendelian randomisation.

IVs for exposure. For UKBB hypothyroidism data used in discovery and sensitivity analysis, IVs were derived from lead SNPs reported by GWAS Atlas (<https://atlas.ctglab.nl/traitDB>). For non-UKBB hypothyroidism data used in validation analysis, IVs were obtained through the clumping procedure of PLINK (version 1.90) [27] ($r^2 < 0.01$), where r^2 is a measure of linkage disequilibrium (LD), a non-random association in the occurrence of alleles at two loci [28]. Then, independent instrumental SNPs were harmonised with the outcome IPF GWAS summary statistics.

2.3. Instrument and variable selection in multivariable MR

Multivariable MR can be considered as an extension of standard univariable MR including multiple variables (Fig. 1b). We selected the large-scale GWAS of potential IPF-related factors including several comorbidities and smoking from two commonly used GWAS platforms: GWAS Atlas, <https://atlas.ctglab.nl/traitDB>, and Neale lab, <http://www.nealelab.is/uk-biobank> (Table 1). MR-BMA basically requires the independence among instrumental SNPs, OSA and GERD were removed since there were no independent SNPs left after the LD pruning using PLINK (version 1.90) [27] ($r^2 < 0.001$). Among the combination SNPs of all retained IPF-related factors, LD clumping was further used to re-determine the final instruments to satisfy the independence. Given another model assumption of multivariable MR is that it only included variables strongly instrumented by at least one instrumental SNP ($P < 5 \times 10^{-8}$), COPD and lung cancer were

further removed. Finally, hypothyroidism, IHD, CAD, T2D, and smoking were included in the MR-BMA model.

2.4. Statistical analysis

The F statistic (also known as the Cragg-Donald statistic) [29] was employed to evaluate the issue of weak instrument in MR analysis. MR inverse-variance weighted (MR-IVW) method was used [30] as the main analysis. Estimator is represented the weighted regression slope of the SNP-outcome effect on the SNP-exposure effect when the intercept is limited to zero. Several other MR methods with different model assumptions were performed to relax the assumption and address the pleiotropy effect: (i) the weighted median method, which can provide consistent estimates when at least half of the instruments used in the analysis are valid [31]. (ii) MR Robust Adjusted Profile Score (MR-RAPS) [32], which is robust to both systematic and idiosyncratic pleiotropy. (iii) MR-lasso [33], which penalizes the number of candidate instrumental SNPs to be used in the model. (iv) MR-IVW method using robust regression (MR-Robust) [33], which reduces the standard error of estimates. (v) leave-one-out (LOO) cross-validation analysis [34] and Mendelian Randomisation Pleiotropy RESidual Sum and Outlier (MR-PRESSO) analysis [35], which makes causal inference as well as outlier detection. Furthermore, we depicted diagnostic plots (e.g., funnel plot) to illustrate the MR results, which can detect visually directional pleiotropy using symmetry of graphical representations. A reverse causation analysis

Table 1
Characteristics of exposures and outcome.

Variable	Source	GWAS platforms	Cases	Controls	Sample size	Population
Exposure						
Atherosclerosis	UKBB data field I9_CORATHER	Neale lab	14,334	346,860	361,194	European
Chronic obstructive pulmonary disease	UKBB data field 41204_J44	Neale lab	1,531	359,663	361,194	European
Gastroesophageal reflux	UKBB data field 41204_K21	Neale lab	10,743	350,451	361,194	European
Hypothyroidism	UKBB data field 20002_1226	GWAS Atlas	18,740	270,567	289,307	European
Ischemic heart disease	UKBB data field I9_IHD	Neale lab	20,857	340,337	361,194	European
Lung cancer	UKBB data field LUNG_CANCER_MESOT	Neale lab	2,007	359,187	361,194	European
Obstructive sleep apnea	UKBB data field G6_SLEEPAPNO	Neale lab	2,249	358,945	361,194	European
Smoking	UKBB data field 20160	GWAS Atlas	235,095	149,918	385,013	European
Type 2 diabetes	UKBB data field 41204_E11	GWAS Atlas	16,673	228,217	244,890	European
Outcome						
Idiopathic pulmonary fibrosis	PMID: 31710517	Collaborative Group of genetic studies of IPF	2,668	8,591	11,259	European

was carried out to exclude the possibility that IPF causally affected hypothyroidism using IPF-associated SNPs as IVs.

MR-BMA was a recently developed multivariable MR method whose major advantage over other methods lies in its ability to select the true causal determinants of an outcome from a set of candidate variables, no matter whether these variables are correlated or not [14]. The model posterior probability (PP) was calculated for each possible set of variables, which quantifies the possibility of the specific variables being a causal determinant of IPF risk. In addition, we computed the marginal inclusion probability (MIP) for the variable by summing up the models PP including the variable, representing the probability of causal association between the variable and IPF. The model-averaged causal effect of each variable on the IPF was finally obtained. Note that it was non-trivial to obtain the P-value using standard statistical techniques, we alternatively use the permutations (200 times) to get the empirical P-values [36] and adjustment for multiple testing via the Benjamini and Hochberg false discovery rate (FDR) procedure [37]. In the MR-BMA analysis, Cochran's Q statistic [38] was used to identify outliers and Cook's distance (Cd) [39] to detect influential observations, respectively. A diagnosis of outlier and influential observations was performed for the top model.

We performed MR analyses using R packages "MendelianRandomization" [40], "MRPRESSO" [35] and "mr.raps" [32]. The statistical analyses were conducted within the R (version 4.0.0) environment. The statistical significance level was set to be 0.05.

2.5. Ethics

Summary data was used and as such ethical approval was not required.

2.6. Role of funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, and writing of the report.

3. Results

3.1. Univariable two-sample MR analysis

3.1.1. Discovery analysis

To address the concern over bias in the estimated causal effect [25] due to overlapping samples in the hypothyroidism GWAS and IPF GWAS, LDSC was used to obtain an approximately zero intercept of genetic covariance 0.0013 (z-test, se=0.0052, P=0.80), indicating little or no impact from sample overlap on the two datasets [26]. A total of 84 IVs were identified in the hypothyroidism GWAS dataset (UKBB data field 20002_1226) (Supplementary Table 1), which explained about 2.15% phenotypic variation of hypothyroidism. The F statistics of all these IVs are above ten with an overall F statistic of 75.62,

indicating that the weak instrument bias would not substantially influence the estimations of causal effects. The overall causal estimate supported the causality of hypothyroidism on IPF (Fig. 2), with an odds ratio (OR) of 1.125 (MR-IVW, 95% confidence interval [CI] 1.028-1.231; P=0.011), suggesting that self-reported history of hypothyroidism/ myxedema can lead to an average of 12.5% increase in the risk of IPF. Overall, the causal effect estimate of hypothyroidism on IPF from several MR methods were quite consistent. The relevant OR estimate was 1.101 (MR-lasso, 95%CI 1.012-1.198; P=0.025), 1.111 (MR-Robust, 95%CI 0.997-1.237; P=0.057), 1.104 (MR-RAPS, 95%CI 1.003-1.215; P=0.042), 1.125 (MR-PRESSO, 95%CI 1.028-1.231; P=0.013) and 1.149 (weighted median method, 95%CI 1.010-1.308; P=0.034), respectively (Fig. 2). The MR-PRESSO analysis showed no significant horizontal pleiotropic outliers. The diagnostic funnel plot shows a visually apparent symmetry, which suggested that the causal estimates would possibly not suffer from the influence of directional pleiotropy (Fig. 3a). In addition, LOO analysis also showed that no single instrumental SNP can influence the causal effect estimation (Fig. 3b).

3.1.2. Validation analysis

The intercept of the LDSC between two hypothyroidism GWASs from discovery and validation analysis is 0.0067 (z-test, se=0.0069, P=0.33), and it is -0.0003 (z-test, se=0.0049, P=0.95) between the validation hypothyroidism GWAS and the IPF GWAS, both indicating approximately no sample overlap [40]. Hence, this non-UKBB hypothyroidism dataset could serve as independent validation dataset and as well satisfy the requirement of exposure and outcome datasets with no sample overlap in two-sample MR. A total of nine independent IVs were finally screened in the independent validation hypothyroidism GWAS dataset [17] (Supplementary Table 2), which explained about 0.77% phenotypic variation of hypothyroidism. The F statistics of all these SNPs are above ten with an overall F statistic of 46.32, indicating less weak instrument bias. No significant causal effect can be detected by MR-IVW method. Further investigation showed that, among the nine SNPs, there were three SNPs having the pleiotropy (rs12449792, rs2983514, rs597808), which has been successfully detected by MR-PRESSO method. Consistently, all MR methods that were not robust to the pleiotropy effect failed to detect the causal associations, while the MR method that was robust to the pleiotropy effect successfully detected the causal effect of hypothyroidism on IPF, with the OR estimate being 1.215 (weighted median method, 95%CI 1.011-1.461; P=0.038), 1.229 (MR-PRESSO, 95%CI 1.102-1.369; P=0.014) (Supplement Figure 1). After removing the pleiotropic SNPs, the OR estimate was 1.229 (MR-IVW, 95%CI 1.054-1.432; P=0.008), 1.229 (MR-lasso, 95%CI 1.054-1.430; P=0.008), 1.226 (MR-Robust, 95%CI 1.095-1.373; P<0.001), 1.231 (MR-RAPS, 95%CI 1.046-1.449; P=0.013), 1.229 (MR-PRESSO, 95%CI 1.103-1.369; P=0.014) and 1.229 (weighted median method, 95%CI 1.016-1.484; P=0.034), respectively (Fig. 2).

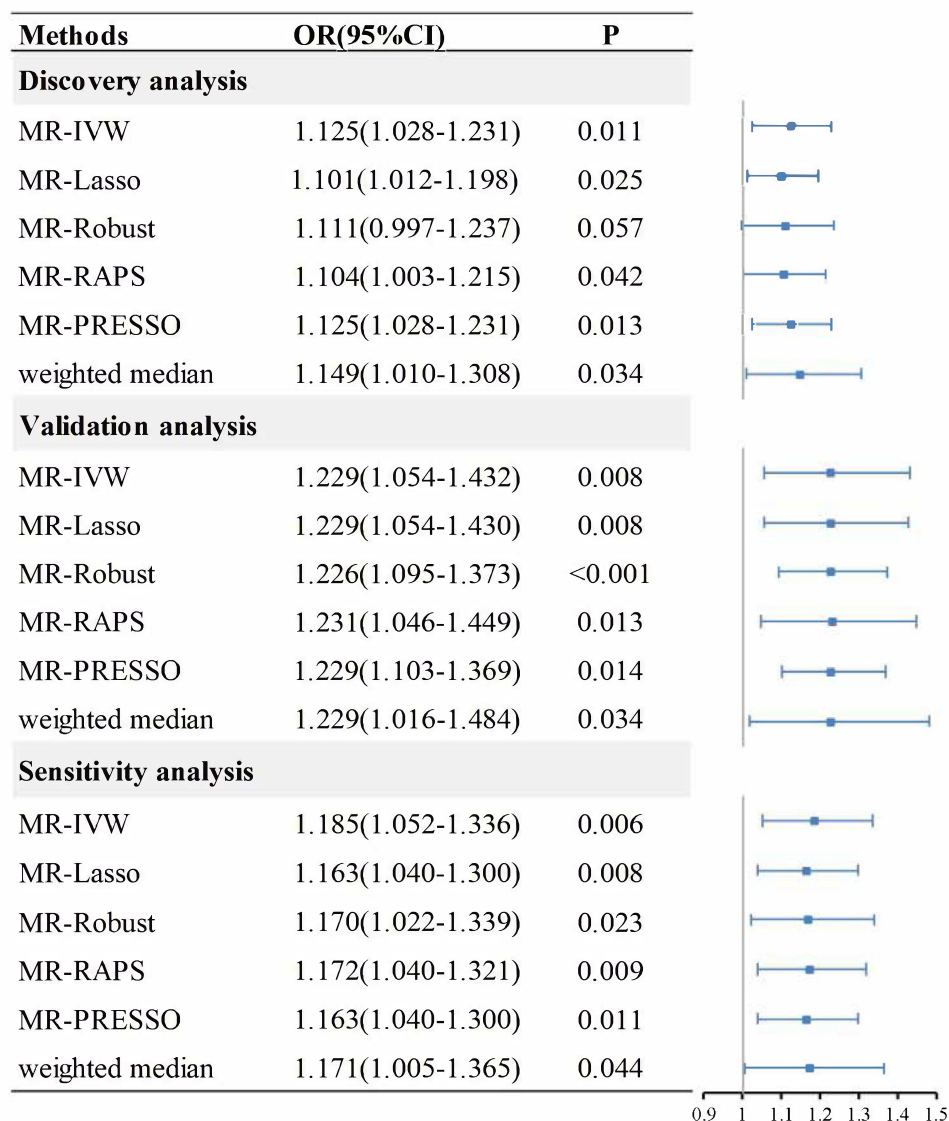


Fig. 2. The causal effect estimates from various Mendelian randomisation methods. The odds ratios of hypothyroidism on IPF are displayed as blue solid box. The 95% confidence intervals are shown as horizontal blue lines. MR, Mendelian randomisation; MR-IVW, MR inverse-variance weighted; MR-RAPS, Robust Adjusted Profile Score; MR-PRESSO, Mendelian Randomisation Pleiotropy RESidual Sum and Outlier; OR, odds ratio; 95%CI, The 95% confidence intervals.

3.1.3. Sensitivity analysis

The results are robust to the hypothyroidism GWAS dataset with different illness code (UKBB data field 41204_E03) (Supplementary Table 3), with 45 valid instrumental SNPs being selected. All findings also supported the causal role of hypothyroidism in the development of IPF (Fig. 2). In bidirectional analyses, the IPF-hypothyroidism causal model was not significant (MR-IVW, OR=1.014; 95%CI 0.983-1.046; P=0.371), indicating no reverse causation.

3.2. Multivariable MR analysis based on BMA

In the MR-BMA analysis, 183 independent genome-wide significant SNPs were identified as valid IVs (Supplementary Table 4). Through model diagnostics, no influential but an outlying instrumental SNP, rs763053, was detected (Fig. 4). Our analysis was performed after removing the outlier (n=182). Table 2 presented the top ten models (i.e., sets of IPF-related factors) according to their model PP, as well as the MIP and the model-averaged causal effect estimate of

the five IPF-related factors. Hypothyroidism had the strongest marginal evidence (MIP=0.397, FDR=0.025). Note that similar results were obtained when 183 IVs were included in the analysis (Supplementary Table 5).

4. Discussion

In the present study, we illustrated the causal effect of hypothyroidism on IPF under the two-sample MR framework. The results were verified using an independent validation dataset as well as through different MR methods with different model assumptions, suggesting the findings were robust and convincing. In addition, bidirectional analyses also indicated no reverse causation. The study took great advantage of an MR analysis using the genetic instruments and benefited from large sample size in the GWAS, containing 2,668 cases and 8,591 controls as the largest on IPF so far [18]. More importantly, the causal role of hypothyroidism in IPF remained even given other comorbidities of IPF in the multivariable MR model. Our study

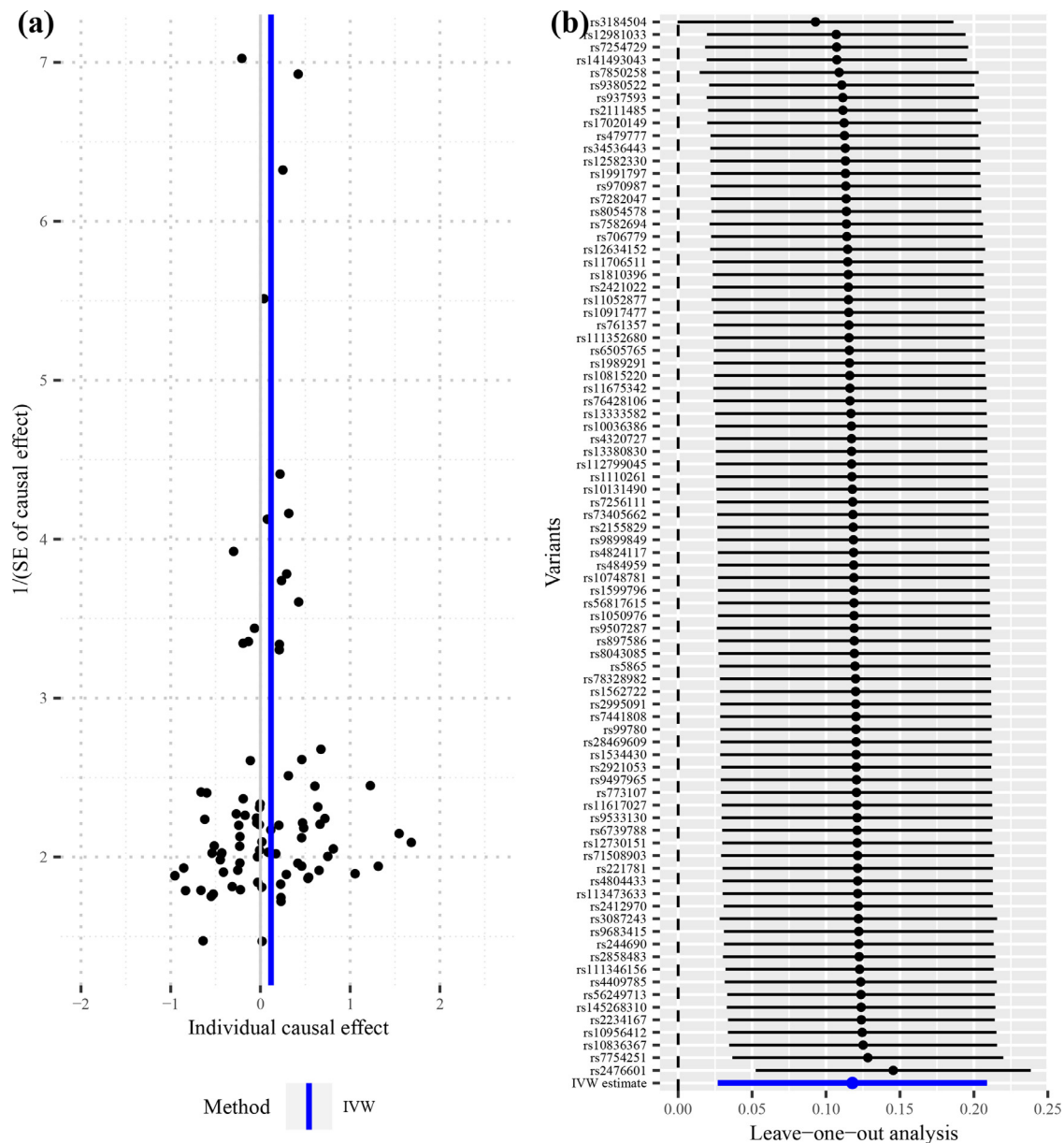


Fig. 3. Diagnostic plots for the two-sample Mendelian randomisation analysis. (a) Funnel plot for individual causal effect estimate. (b) Forest plot for leave-one-out analysis, with each point denoting the causal effect by IVW after removing the specific SNP on the left side. IVW, inverse variance weighted method.

corroborated results from previous observational studies showing a relationship between hypothyroidism and pulmonary fibrosis [7–9].

The underlying mechanisms in the causative role of hypothyroidism in IPF may be quite complex. Several biological mechanisms might explain the effect. Previous studies have suggested that the state of hypothyroidism would reinforce oxidative stress [41,42], which is implicated to affect apoptosis of structural and inflammatory cells as well as alter the cytokine microenvironment balance and, hence, involving in the pathogenic pathways of IPF [43,44]. In addition, mitochondrial dysfunction and metabolic aberrations are possibly involved in pulmonary fibrosis [45]. Thyroid hormone, which regulates fundamental biological functions, is critical for maintenance of cellular homeostasis during stress responses [46]. In an animal study [47], anti-fibrosis properties and protection of alveolar epithelial cells of thyroid hormones suggest a pathway on the restoration of mitochondrial function, so thyroid hormone may be a potential treatment for pulmonary fibrosis. Furthermore, TGF- β would contribute

to the development of IPF in expansion of the fibroblast and myofibroblast populations [48] by activating SMAD (mothers against decapentaplegic) transcription factors. An experimental study suggested that thyroid hormones could inhibit TGF- β signaling by antagonizing transcriptional activation of TGF- β /SMAD, which leads to attenuation of fibrotic responses [49], and thus may play a critical role in blocking the progression of IPF. In conclusion, hypothyroidism with a decrease in the synthesis and secretion of thyroid hormones [50], creates metabolic conditions that lead to failure in restoration of epithelium mitochondrial function and antagonism transcriptional activation of TGF- β /SMAD, thus contributing to IPF pathogenesis.

Some limitations of this study should be considered. First, it should be noted that we used the summary data of self-reported hypothyroidism with non-cancer illness code from UKBB field 20002 in discovery analysis, rather than the summary data of hypothyroidism with diagnosis from main ICD-10 code. Indeed, in total we

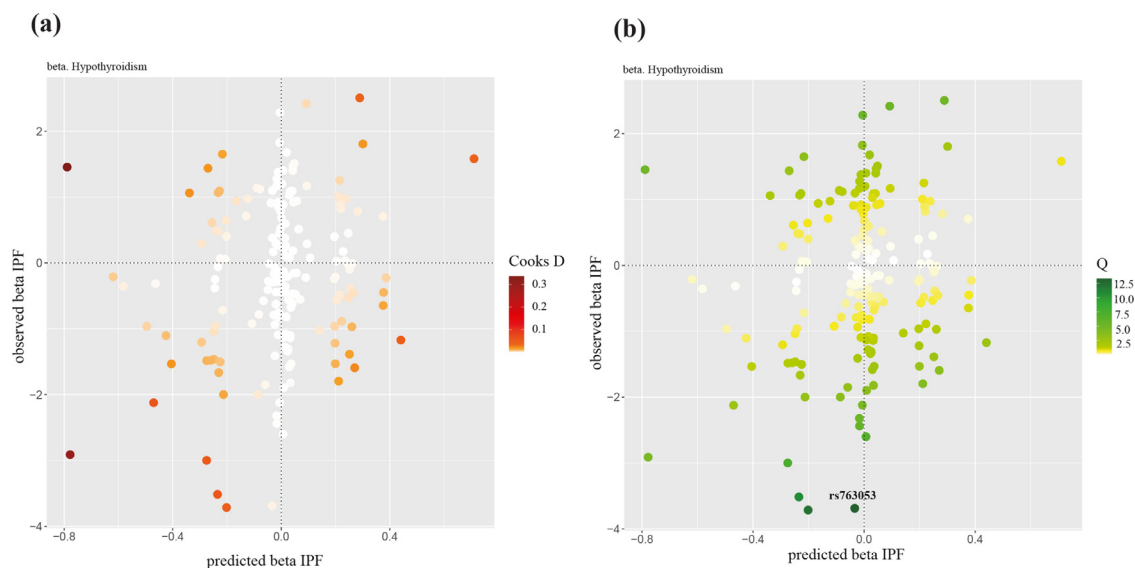


Fig. 4. Model diagnostics in multivariable MR based on BMA. The predicted associations with IPF based on the model including hypothyroidism (x-axis) are plotted against the observed associations with IPF (y-axis). This is the top model when keeping outliers and influential genetic variants in the analysis. (a) Cook's distance for the influential points; (b) the q -statistic for outliers. Any genetic variant with q value larger than threshold (q -statistic = 13.246) or Cook's distance larger than the median of the relevant F -distribution ($Cd = 0.457$) is marked by a label indicating the gene region. IPF, idiopathic pulmonary fibrosis.

Table 2

The ranking of models for idiopathic pulmonary fibrosis after model diagnostics in the multivariable MR analysis.

	Model	PP	casual effect	Risk factor	MIP	Model-averaged causal effect	Empirical p-value	FDR
1	hypothyroidism	0.290	0.111	hypothyroidism	0.397	0.044	0.005	0.025
2	IHD	0.256	0.503	IHD	0.355	0.176	0.780	0.995
3	AS	0.210	0.364	AS	0.300	0.106	0.905	0.995
4	smoking	0.055	0.037	smoking	0.087	0.003	0.995	0.995
5	hypothyroidism, IHD	0.045	0.109,0.466	T2D	0.039	0.002	0.920	0.995
6	hypothyroidism, AS	0.038	0.110,0.340					
7	AS, IHD	0.032	0.318,0.473					
8	T2D	0.024	0.043					
9	hypothyroidism, smoking	0.010	0.111,0.047					
10	IHD, smoking	0.009	0.505,0.039					

Top ten models (combination of risk factors) ranked by the model posterior probability and all risk factors ranked by the marginal inclusion probability in the analysis after model diagnostics based on 182 genetic variants. Causal effects are log odds ratios for idiopathic pulmonary fibrosis. Empirical P-values are computed using 200 permutations and adjustment for multiple testing via the Benjamini and Hochberg false discovery rate (FDR) procedure. PP=posterior probability, MIP=Marginal inclusion probability, IHD= ischemic heart disease, AS=atherosclerosis, T2D= type 2 diabetes.

screened out 1,690 SNPs from the hypothyroidism GWAS data with diagnosis from main ICD-10, but only two SNPs remained after merging with these SNPs with 1000 Genome reference panel data, and both SNPs were not matched with the GWAS of IPF. Second, some comorbidities GWAS of IPF (i.e., OSA, GERD, COPD, PH, and lung cancer), failed to be adjusted in our multivariable MR model, which may have little impact on the overall causal estimates, but can increase the residual error and thus reduce the statistical efficiency. However, here we would like to follow the original multivariable MR work to emphasize that the main goal of MR-BMA method is, through model averaging procedure, to prioritize and select causal risk factors in a Bayesian framework from a set of candidate risk factors. Since the variable selection procedure often shrinks estimates towards the null, MR-BMA can lead to causal effect estimates being biased towards the null when there is a causal effect and unbiased estimates when there is no causal effect. Third, IPF is known to dramatically increase in prevalence with age [51], the use of summary rather than individual-level data from GWAS prevents direct estimation of the causal effect between hypothyroidism and age-stratified IPF. In addition, the IPF GWAS in this study was unable to adjust for age and gender due to these information not being available for all individuals in each dataset. Finally, the result from MR reflects the change in IPF risk due to a genetically predisposed (lifetime) change in

hypothyroidism status, hence the short-term effect of hypothyroidism status on IPF risk merits additional investigation.

To conclude, our study rendered strong evidence of self-reported history of hypothyroidism/myxedema as a causal determinant on IPF risk, even when several other comorbidities of IPF were taken into account. Importantly, it is biologically plausible that hypothyroidism is embedded in the critical pathway for patients with IPF, for whom hypothyroidism should be considered and monitored closely. However, translating the findings into the treatment of IPF may be not straightforward, and experimental studies are certainly required for further validation.

Data Sharing Statement

All data are available on the aforementioned public repository and are accessible with permission from the corresponding data committee. No restrictions on data availability other than those imposed by the corresponding data committee.

Contributors

ZY, FX, and JJZ conceived the study. ZY obtained the genetic data. ZY, FX, JJZ and YZ verified all the data in the study. YZ performed the

analyses and interpreted the results. All authors contributed to the initial draft; JHZ and ZY made numerous revisions. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (81872712, 81673272, and 82173624), the Natural Science Foundation of Shandong Province (ZR2019ZD02) and the Young Scholars Program of Shandong University (2016WLJH23). We would like to thank Collaborative Group of genetic studies of IPF for providing us with the IPF GWAS summary data.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103669.

References

- [1] Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011;183(6):788–824.
- [2] Neumark N, CC Jr., K-A Rose, N Kaminski. The Idiopathic Pulmonary Fibrosis Cell Atlas. *Am J Physiol Lung Cell Mol Physiol* 2020;319(6):L887–L92.
- [3] Harari S, Davi M, Biffi A, Caminati A, Ghirardini A, Lovato V, et al. Epidemiology of idiopathic pulmonary fibrosis: a population-based study in primary care. *Intern Emerg Med* 2020;15(3):437–45.
- [4] Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N Engl J Med* 2018;378(19):1811–23.
- [5] Raghu G, Amatto VC, Behr J, Stowasser S. Comorbidities in idiopathic pulmonary fibrosis patients: a systematic literature review. *European Respiratory Journal* 2015;46(4):1113–30.
- [6] Oldham JM, Collard HR. Comorbid Conditions in Idiopathic Pulmonary Fibrosis: Recognition and Management. *Front Med* 2017;4:123.
- [7] Oldham JM, Kumar D, Lee C, Patel SB, Takahashi-Manns S, Demchuk C, et al. Thyroid Disease Is Prevalent and Predicts Survival in Patients With Idiopathic Pulmonary Fibrosis. *Chest* 2015;148(3):692–700.
- [8] Sato Y, Tanino Y, Nikaido T, Togawa R, Kawamata T, Wang X, et al. Clinical significance of thyroid hormone and antibodies in patients with idiopathic interstitial pneumonia. *J Thorac Dis* 2020;12(3):522–37.
- [9] Awano N, Izumo T, Fukuda K, Ton M, Yamada D, Takemura T, et al. Is hypothyroidism in idiopathic pleuroparenchymal fibroelastosis a novel lung-thyroid syndrome? *Respir Investig* 2018;56(1):48–56.
- [10] Bothwell LE, Podolsky SH. The Emergence of the Randomized, Controlled Trial. *N Engl J Med* 2016;375(6):501–4.
- [11] Smith GD, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;32(1):1–22.
- [12] Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 2016;45(5):1600–16.
- [13] Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int J Epidemiol* 2016;45(3):908–15.
- [14] Zuber V, Colijn J, Klaver C, Burgess S. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat Commun* 2020;11(1):29.
- [15] Watanabe K, Stringer S, Frei O, Umicvic Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;51(9):1339–48.
- [16] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562(7726):203–9.
- [17] Teumer A, Chaker L, Groeneweg S, Li Y, Di Munno C, Barbieri C, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat Commun* 2018;9(1):4455.
- [18] Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2020;201(5):564–74.
- [19] Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1(4):309–17.
- [20] Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45(6):613–20.
- [21] Fingerlin TE, Zhang W, Yang IV, Ainsworth HC, Russell PH, Blumhagen RZ, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016;17(1):74.
- [22] Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med* 2017;5(11):869–80.
- [23] American Thoracic Society/European Respiratory Society. International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002;165(2):277–304.
- [24] Raghu G, Rochwerf B, Zhang Y, Garcia CA, Azuma A, Behr J, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. *Am J Respir Crit Care Med* 2015;192(2):e3–19.
- [25] Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016;40(7):597–608.
- [26] Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015;47(11):1236–41.
- [27] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559–75.
- [28] Canzoniero J, Rosenberg N. Mathematical properties of the measure of linkage disequilibrium. *Theor Popul Biol* 2008;74:130–7.
- [29] Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res* 2017;26(5):2333–55.
- [30] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;37(7):658–65.
- [31] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40(4):304–14.
- [32] Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *The Annals of Statistics* 2020;48(3):1742–69.
- [33] Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. *Genet Epidemiol* 2020;44(4):313–29.
- [34] Noyce AJ, Kia DA, Hemani G, Nicolas A, Price TR, De Pablo-Fernandez E, et al. Estimating the causal influence of body mass index on risk of Parkinson disease: A Mendelian randomisation study. *PLoS Med* 2017;14(6):e1002314.
- [35] Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;50(5):693–8.
- [36] Zuber V, Gill D, Ala-Korpela M, Langenberg C, Butterworth A, Bottolo L, et al. High-throughput multivariable Mendelian randomization analysis prioritizes apolipoprotein B as key lipid risk factor for coronary artery disease. *Int J Epidemiol* 2021;50(3):893–901.
- [37] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 1995;57(1):289–300.
- [38] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557–60.
- [39] Cook RD. Influential Observations in Linear Regression. *Journal of the American Statistical Association* 1979;74(365):169–74.
- [40] Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46(6):1734–9.
- [41] Santi A, Duarte MM, Moresco RN, Menezes C, Bagatini MD, Schetinger MR, et al. Association between thyroid hormones, lipids and oxidative stress biomarkers in overt hypothyroidism. *Clin Chem Lab Med* 2010;48(11):1635–9.
- [42] Mancini A, Di Segni C, Raimondo S, Olivieri G, Silvestrini A, Meucci E, et al. Thyroid Hormones, Oxidative Stress, and Inflammation. *Mediators Inflamm* 2016;2016:6757154.
- [43] Fois AG, Paliogiannis P, Sotgia S, Mangoni AA, Zinellu E, Pirina P, et al. Evaluation of oxidative stress biomarkers in idiopathic pulmonary fibrosis and therapeutic applications: a systematic review. *Respir Res* 2018;19(1):51.
- [44] Mastruzzo C, Crimi N, Vancheri C. Role of oxidative stress in pulmonary fibrosis. *Monaldi Arch Chest Dis* 2002;57(3–4):173–6.
- [45] Mora AL, Bueno M, Rojas M. Mitochondria in the spotlight of aging and idiopathic pulmonary fibrosis. *J Clin Invest* 2017;127(2):405–14.
- [46] Muller R, Liu Y-Y, Brent GA. Thyroid Hormone Regulation of Metabolism. *Physiol Rev* 2014;94(2):355–82.
- [47] Yu G, Tzouveleakis A, Wang R, Herazo-Maya JD, Ibarra GH, Srivastava A, et al. Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function. *Nat Med* 2018;24(1):39–49.

- [48] Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers* 2017;3:17074.
- [49] Alonso-Merino E, Martín Orozco R, Ruíz-Llorente L, Martínez-Iglesias OA, Velasco-Martín JP, Montero-Pedrazuela A, et al. Thyroid hormones inhibit TGF- β signaling and attenuate fibrotic responses. *Proc Natl Acad Sci U S A* 2016;113(24):E3451–60.
- [50] Chaker L, Bianco AC, Jonklaas J, Hypothyroidism Peeters RP. *Lancet* 2017;390(10101):1550–62.
- [51] Wolters PJ, Collard HR, Jones KD. Pathogenesis of Idiopathic Pulmonary Fibrosis. *Annu Rev Pathol* 2014;9:157–79.