

Integrating Multi-Omics Data to Uncover Causal Links Between Mitochondria-Related Genes and Chronic Obstructive Pulmonary Disease: A Mendelian Randomization Study

Er Hong, Jia Mao, Zhicheng Ke, Yang Wu

Department of Respiratory, Ningbo Municipal Hospital of Traditional Chinese Medicine, Affiliated to Zhejiang Chinese Medical University, Ningbo, Zhejiang, 315010, People's Republic of China

Correspondence: Er Hong, Department of Respiratory, Ningbo Municipal Hospital of Traditional Chinese Medicine, Affiliated to Zhejiang Chinese Medical University, No. 819, Li Yuan North Road, Ningbo, Zhejiang, 315010, People's Republic of China, Tel +86 0574-83885201, Email 2024n061@zcmu.edu.cn

Background: As a relatively common respiratory disease, chronic obstructive pulmonary disease (COPD) has a high incidence and mortality rate. Mitochondrial dysfunction has been implicated in COPD pathogenesis, but the causal genes and underlying molecular mechanisms remain unclear.

Methods: We performed a summary data-based Mendelian Randomization (SMR) study integrating summary data from genome-wide association studies (GWAS) with blood-based methylation (mQTL), expression (eQTL), and protein quantitative trait loci (pQTL) to identify mitochondrial-related genes causally associated with COPD. Significant findings were validated using two-sample MR, independent lung tissue transcriptomic data (GSE76925), and weighted gene co-expression network analysis (WGCNA) to assess transcriptional consistency and functional convergence.

Results: Our integrative SMR and colocalization analyses identified 77 mitochondrial genes linked to COPD risk, including *GPXI*, *TUFM*, *COQ5*, *BPHL*, and *NAGS*. A methylation-to-expression regulatory cascade was observed, with hypermethylation at *GPXI* cg24011261 associated with increased gene expression and higher COPD risk—despite its role as an antioxidant enzyme. Two-sample MR confirmed robust causal effects of *GPXI*, *BPHL*, *TUFM* and *COQ5* expression on COPD. These findings were replicated in lung tissue: *GPXI*, *COQ5*, and *TUFM* were significantly upregulated in COPD patients (GSE76925). WGCNA revealed that these genes reside within a highly interconnected turquoise module strongly correlated with COPD status ($r = 0.43$, $p < 0.001$) and enriched in oxidative phosphorylation and mitochondrial energy metabolism pathways.

Conclusion: This study provides systematic genetic and transcriptomic evidence that mitochondrial-related genes, particularly *GPXI* and *TUFM*, exert causal effects on COPD risk through regulatory cascades and coordinated network dysregulation. The convergence of genetic, epigenetic, and co-expression evidence underscores mitochondrial dysfunction as a central mechanism in COPD pathogenesis and highlights potential targets for future therapeutic development.

Keywords: chronic obstructive pulmonary disease, mitochondrial dysfunction, multi-omics, mendelian randomization, colocalization

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a prevalent respiratory disorder characterized by the presence of chronic bronchitis and emphysema as its primary forms, featuring persistent airflow limitation coupled with progressive dyspnea.¹ The global prevalence of COPD is on the rise, making it one of the leading causes of mortality and morbidity.² While smoking is the predominant risk factor for COPD, genetic and environmental factors also play significant roles.³

Mitochondrial dysfunction has been implicated in COPD pathogenesis, but the mechanisms remain unclear. The persistent oxidative stress and chronic inflammatory characteristics of COPD lead to mitochondrial damage, resulting in reduced ATP

production, impaired mitochondrial autophagy, and an increase in reactive oxygen species (ROS) production.⁴ These disruptions exacerbate cellular damage, promote apoptosis, and contribute to the remodeling of airway structures.⁵ These mitochondrial perturbations are believed to directly fuel the hallmark pathologies of COPD. Additionally, an increase in the levels of free mitochondrial DNA (mtDNA) in the plasma of COPD patients has been observed,⁶ and mtDNA mutations further compromise mitochondrial function, leading to dysregulation of energy metabolism, weakened respiratory muscles, and exacerbation of symptoms such as dyspnea, skeletal muscle injury, and decreased exercise capacity.⁷ The translocation of Parkin, a ubiquitin-related degradation molecule associated with mitochondrial autophagy, is impaired following the activation of the Pink1 pathway, a critical process for clearing damaged mitochondria. This impairment exacerbates mitochondrial dysfunction and promotes cellular damage, thereby accelerating the progression of COPD.⁸ Morphological alterations of mitochondria in COPD patients are observed, including fragmentation, swelling, and damage to the cristae integrity, leading to impaired energy production and cellular functions.⁹ Although the important role of mitochondrial dysfunction in the pathogenesis of COPD has been recognized, it remains unclear which genes play a key role in this process. Thus, despite growing evidence from genomics, epigenomics, and proteomics implicating mitochondrial processes in COPD,^{10–12} there is a critical lack of causal integration across molecular layers. Most existing studies are associative and cannot establish directionality or distinguish primary drivers from secondary effects. To address this gap, we employed a multi-omics summary data-based Mendelian randomisation (SMR) approach to systematically evaluate the causal role of mitochondrial-associated genes in susceptibility to COPD.

MR is a causal inference method based on genetic variants that can be used to evaluate the causal association between exposure factors and diseases.¹³ Traditional MR studies primarily rely on single-omics data, which may not be robust enough to clarify the mechanisms of complex diseases. However, multi-omics MR integrates data from genomics, transcriptomics, epigenomics, and proteins, thereby enhancing the accuracy and robustness of causal inference.

Therefore, this study leverages a SMR framework to systematically investigate the causal roles of mitochondria-related genes in COPD pathogenesis. Specifically, we hypothesize that genetic variants influencing the epigenome (mQTLs), transcriptome (eQTLs), and proteome (pQTLs) of these genes can causally affect COPD risk. By integrating these different layers of biological data, we aim to identify high-confidence causal genes, elucidate their regulatory mechanisms, and provide robust evidence for new molecular biomarkers and therapeutic targets in COPD.

Materials and Methods

Study Design

This study was designed according to the STROBE-MR reporting guidelines and used a MR approach to explore the causal association between mitochondria-related genes and COPD. The approach included comprehensive analysis of genetic data, application of SMR and HEIDI tests based on multi-omics summary data to assess associations between mitochondrial-related genes and COPD, and colocalization analyses to identify shared genetic determinants. The study design, workflow for the selection of genetic variants, and analytical methodology are detailed in [Figure 1](#). This study is a secondary analysis of publicly available, anonymized summary-level data. Ethical approval has been sought and stamped. Furthermore, all data utilised by this research institute originate from publicly accessible databases, which obtained the requisite ethical approval and informed consent from participants during their original studies.

Data Sources and Processing

Mitochondrial mechanism genes were identified from the MitoCarta3.0 database, totaling 1,136 genes. We utilized GWAS data for COPD from distinct cohorts. The FinnGen R10 dataset (FINNGEN R10_J10 COPD), serving as the primary discovery dataset, included 20,066 cases and 338,303 controls (<https://r10.risteys.finnngen.fi/>). Validation of our findings was conducted using the UK Biobank dataset (UKB-D-COPD_EXCL), which included 26,710 cases and 334,484 controls (<https://www.ukbiobank.ac.uk/>). The COPD phenotype in these GWAS datasets was analyzed as a composite outcome, without stratification into clinical subtypes.

Additionally, we obtained summary data for blood eQTLs from eQTLGen, which included genetic data from 31,684 individuals.¹⁴ Blood mQTL summary data were derived from a meta-analysis of two European cohorts: the Brisbane

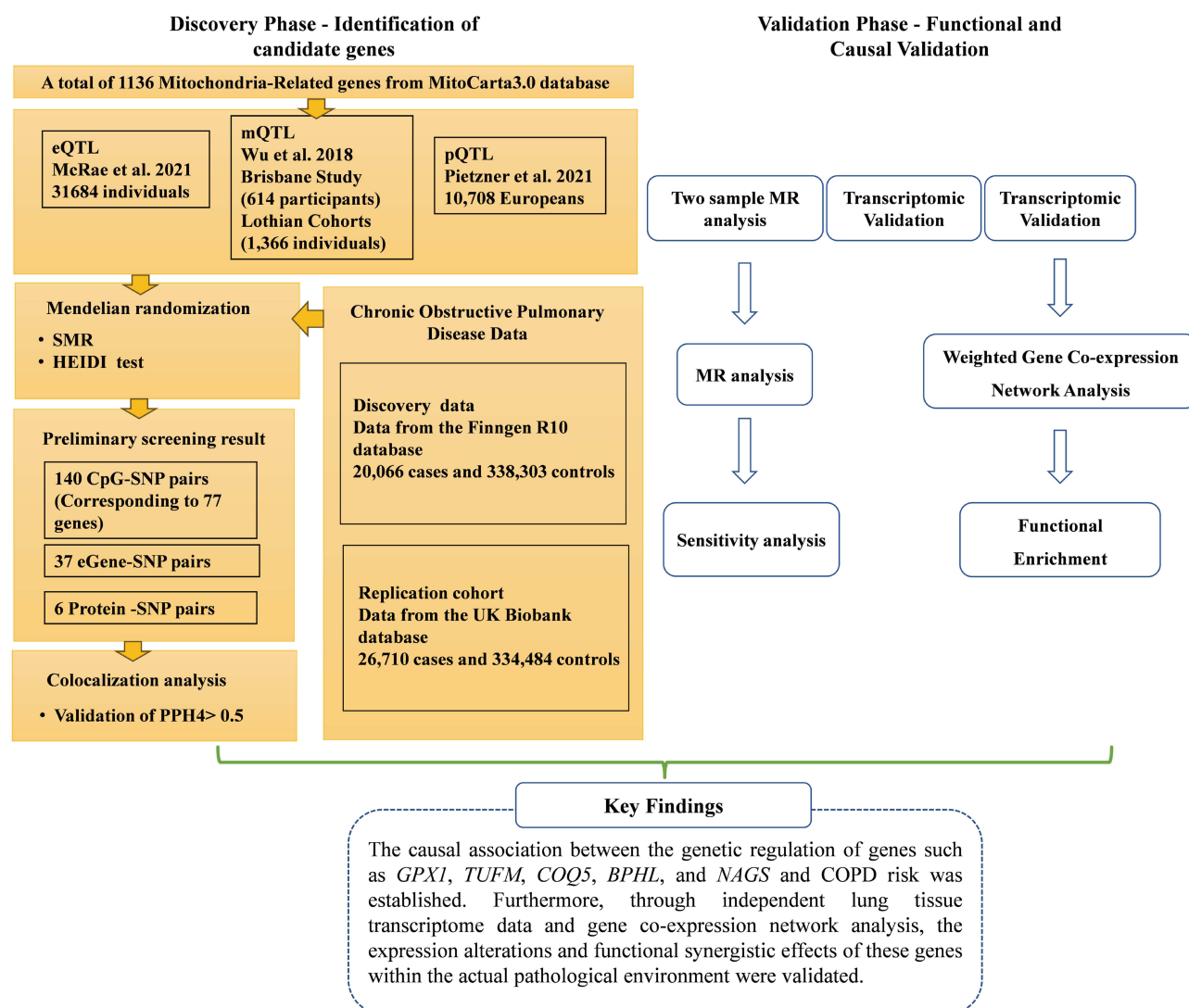


Figure 1 Study design and workflow The flowchart illustrates the multi-stage process of the study. The discovery phase involved identifying candidate mitochondria-related genes, followed by Mendelian Randomization (SMR), colocalization analysis, and validation to establish causal links with Chronic Obstructive Pulmonary Disease (COPD) using large-scale genetic databases. The validation phase included two-sample MR analysis, transcriptomic validation in an independent patient cohort, and functional network analysis to confirm findings and explore biological mechanisms.

Systems Genetics Study (n=614) and the Lothian Birth Cohorts (n=1366). Blood pQTL summary data were obtained from Pietzner et al,¹⁵ which included data from 10,708 Europeans.

To assess the tissue-specific expression of target genes and explore their potential causal impact on COPD, we used lung eQTL data from the GTEx v8 dataset, which included 838 donors and 17,382 samples from 52 tissues and two cell lines.

In the analysis of eQTL loci for key genes (*GPXI*, *BPHL*, *NAGS*, *TUFM*, *COQ5*) and their therapeutic targets for COPD (FINNGEN R10_J10 COPD), the eQTLs also originate from eQTLGen. Transcriptome expression profile data were downloaded from the Gene Expression Omnibus database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). The dataset name is GSE76925.

SMR Analysis

Utilizing the SMR software tool (version SMR v1.3.1), we conducted SMR and HEIDI tests to assess the associations between mitochondrial-related gene methylation, expression, protein abundance, and COPD. The SMR approach showed stronger statistical power compared to traditional MR analyses, especially when the exposure and outcome data were derived from two large, independent cohorts based on top-related cis-QTLs. We selected the top-related cis-QTLs with

a window centered around the corresponding gene ± 1000 kb and a significance threshold of $P = 5.0 \times 10^{-8}$.¹⁶ SNPs with allele frequency differences exceeding the specified threshold (set at 0.2 in this study) between any two datasets (including LD reference samples, QTL summary data, and outcome summary data) were excluded. For mQTL, eQTL, and pQTL, the maximum allowable proportion of SNPs with allele frequency differences was set at 0.05 for the discovery set and 0.1 for the validation set (the latter exceeding 5%).

In addition to investigating the causal associations between QTLs (mQTL, eQTL, and pQTL) and COPD, we further explored the causal associations between mQTL and eQTL by treating mQTL as the exposure and eQTL as the outcome. Similarly, we investigated the causal associations between eQTL and pQTL by considering eQTL as the exposure and pQTL as the outcome.

Building on the SMR analysis, the authors developed a new multi-SNP-based SMR analysis method (-SMR-multi). This method considers all SNPs within a QTL probe window region that have a p-value below the default threshold of 5×10^{-8} and an LD R^2 value below the default threshold of 0.9 with the top-associated SNP. In this study, significance results from this method will be comprehensively considered for evaluation. Subsequently, the HEIDI test with $P > 0.05$ was applied to filter out pleiotropic effects. Results satisfying $P\text{-SMR} < 0.05$, $P\text{-SMR multi} < 0.05$, and $P\text{-HEIDI} > 0.05$ will be used for subsequent colocalization and integration analysis of eQTL, mQTL, and pQTL.

Using the R package “coloc”, we conducted colocalization analysis to identify mitochondrial gene-related cis-QTLs (including mQTLs, eQTLs, and pQTLs) that share causal variants with COPD.¹⁷ Specifically, when a GWAS signal and QTLs are found to colocalize, we infer that the locus in the GWAS signal may affect the phenotype by altering gene-related biological processes. In the colocalization analysis, we report five different posterior probabilities corresponding to five exclusive hypotheses: H0: No trait in the region is genetically associated with the SNP; H1: Only trait 1 is genetically associated with the SNP; H2: Only trait 2 is genetically associated with the SNP; H3: Both traits are associated with the SNP, but through different causal variants; H4: Both traits are associated with the SNP and share a common causal variant. For the colocalization analysis of mQTL-GWAS, eQTL-GWAS, and pQTL-GWAS, the colocalization region windows were set to ± 1000 kb. To allow for weaker QTLs to colocalize with the signal, a QTL signal was considered successfully colocalized with the GWAS signal when $P_{12} = 5 \times 10^{-5}$ and $PP_{H4} > 0.5$.¹⁸

Mendelian Randomization (MR) Analysis

IVs Selection

MR studies used single nucleotide polymorphisms (SNPs) associated with exposure as IVs. SNPs located within a ± 500 kb genomic range of the drug target gene will be utilised for IVs screening. The IVs included in this study met the following criteria: first, a threshold of 5×10^{-8} was employed to identify QTLs exhibiting genome-wide correlations with target gene expression.^{19,20} Subsequently, in the QTL data, the minor allele frequency (MAF) was greater than 1%.²¹ The F-statistic was employed to exclude weak IVs, with those yielding an F-statistic > 10 deemed strongly correlated with the exposure factor.²² Linkage disequilibrium (LD) between SNPs was then removed using the following criteria: $R^2 < 0.3$ and window size = 500kb.²³

Heterogeneity and Horizontal Pleiotropy Tests

Heterogeneity of instrumental variables was assessed using the Cochrane Q test based on inverse variance weighting (IVW) models,²⁴ with the Cochrane Q test based on Egger regression models employed as a supplementary indicator.^{25,26} MR PRESSO analysis and Egger regression-based methods were utilised to evaluate horizontal pleiotropy of IVs.²⁶

Instrumental Variability Sensitivity Test and Outlier Elimination

Sensitivity testing of IVs was conducted using the leave-one-out method.²⁷ Each time, a specific SNP within the IVs was excluded, and the MR analysis effect of the remaining IVs was calculated to assess and eliminate IVs exerting severe biased effects on MR results. For outliers, radial MR analyses were conducted using IVW models and Egger regression models.²⁸ Outliers were identified based on each locus's contribution to overall heterogeneity (quantified via Cochran's Q and Rucker's Q statistics).²⁸ SNPs identified as outliers in both sets of Radial MR analyses constituted Level 1 outliers.

These were prioritised for exclusion, followed by a renewed heterogeneity analysis. Should no heterogeneity be detected, the results after outlier removal were retained. Should heterogeneity persist, SNPs identified as outliers in both radial MR analyses (Level 2) will be excluded, and heterogeneity analysis repeated. If heterogeneity remains, the final result will be based on the outlier SNP excluded according to the leave-one-out analysis (Level 3). Additionally, MR PRESSO analysis was conducted to supplement the assessment of radial MR results.²⁶

MR Analysis

Inverse-variance weighted (IVW) was used as the primary method to assess the causal association between commonly used medicines and diabetic eye disease by calculating the odds ratio (OR) and 95% confidence interval (CI).²⁹ To ensure the robustness of our findings, we also utilized three additional MR methods: MR-Egger, weighted median (WM), and weighted mode.³⁰ Additionally, the Steiger filtering method was applied to support the robustness of the causal directionality from exposure to outcome, thereby mitigating the potential for horizontal pleiotropy to some extent.³¹

Transcriptomic Validation in Patient Tissues

To validate the expression patterns of candidate genes in lung tissue, we analysed RNA sequencing data from the Gene Expression Omnibus database.³² Standardised expression data and clinical information for 111 COPD samples and 40 normal lung tissue samples (GSE76925) were downloaded from the aforementioned database. Differential expression analysis was performed using the R package *limma* (v3.62.2), with $P < 0.05$ set as the threshold for statistical significance.

Weighted Gene Co-Expression Network Analysis (WGCNA) and Functional Enrichment

Weighted Gene Co-expression Network Analysis (WGCNA) was employed to construct co-expression networks.³³ First, sample clustering analysis was performed on the expression data, utilising hierarchical clustering to identify potential outlier samples. Subsequently, the *pickSoftThreshold* function was used to determine an appropriate soft-thresholding power, with power = 3 ultimately selected for constructing the weighted network. Following network construction, modules were identified using the dynamic tree cut method, with a minimum module size of 100 genes. Modules were then correlated with clinical phenotypes (specifically COPD phenotype and positive gene expression phenotype in this dataset) to select modules highly associated with traits. Candidate core genes within key modules were further selected using thresholds of $|\text{gene significance (GS)}| > 0.2$ and $|\text{module membership (MM)}| > 0.75$.³⁴

Functional enrichment of the analyzed gene sets (eQTL and pQTL levels combined) was performed using the R package “*clusterProfiler*”. In the GO enrichment analysis, three subcategories were analyzed: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC).³⁵ In addition, KEGG functional pathway enrichment analysis was performed without considering its disease pathway (Human Diseases).³⁶ The significance threshold for enriched pathways was set at $P < 0.05$ as the criterion for result selection.

Statistical Analysis

All statistical analyses were conducted using R (version 4.3.3). The R packages “*ggplot2*” and “*ggrepel*” were used for creating Manhattan plots, and “*forestplot*” was used for forest plots. The plotting codes for *SMRLocusPlot* and *SMREffectPlot* were sourced from Zhu et al.³⁷ The significance criterion for SMR analysis is $P\text{-SMR} < 0.05$, $P\text{-SMR multi} < 0.05$ and $P\text{-HEIDI} > 0.05$, while the significance criterion for colocalization results is $PP.H4 > 0.5$. The R packages “*TwoSampleMR*” and “*RadialMR*” were employed for MR analysis. Additionally, scatter plots, leave-one-out plots, and funnel plots were produced using built-in functions within “*TwoSampleMR*” and “*RadialMR*”.

Result

Our study systematically investigated the causal links between mitochondria-related genes and Chronic Obstructive Pulmonary Disease (COPD) through a multi-omics approach, following the workflow detailed in [Figure 1](#).

SMR Analysis Identifies Causal Associations Across Multiple Omics Layers

We first conducted Summary data-based Mendelian Randomization (SMR) to identify causal associations between molecular QTLs and COPD. Our analysis revealed several significant causal links between mitochondrial-related genes and COPD risk at both the methylation and expression levels (Table 1 and Figure 2). Among the key findings, we identified a paradoxical risk-promoting role for *GPXI*. Specifically, higher methylation at site cg24011261 was associated with increased COPD risk (OR = 1.18, 95% CI: 1.07–1.31), and similarly, increased *GPXI* expression was also linked to a higher risk of COPD (OR=1.44, 95% CI: 1.18–1.76), with both associations supported by strong colocalization evidence (PPH4 > 0.75). Other key causal associations for genes such as *BPHL* and *NAGS* are also detailed in Table 1.

Transcriptomic Validation Confirms Upregulation of Key Causal Genes in COPD Lung Tissue

To validate these findings in the primary disease tissue, we analyzed an independent lung transcriptomic dataset (GSE76925). As shown in Figure 3, the expression of three key causal genes—(A) *TUFM*, (B) *COQ5*, and (C) *GPXI*—was significantly upregulated in COPD patients compared to controls, consistent with the direction of effect from our SMR analysis.

Our discovery SMR analysis identified a total of 140 significant mQTLs (Figure S1 and Table S2), 37 eQTLs (Figure S2A and Table S1), and 6 pQTLs (Figure S2B and Table S5). The pQTL associations, however, lacked robust colocalization evidence. A genome-wide overview of all SMR results is provided in the Manhattan plots in Figure S3.

Integration of Omics Data Reveals Complete Methylation-to-Disease Causal Pathways

To elucidate the underlying regulatory mechanisms, we integrated the multi-omics data. LocusCompare plots for *GPXI* confirmed strong colocalization between its mQTL/eQTL signals and COPD GWAS signals, suggesting a shared genetic architecture (Figure S4).

By formally testing the link between methylation and gene expression (Table 2 and Table S3), we uncovered complete, multi-layered causal pathways for three key genes. For *GPXI*, we established a paradoxical risk-promoting cascade: hypermethylation at cg24011261 causally drives higher *GPXI* expression, which in turn increases COPD risk (Figure S5). In contrast, for *BPHL* and *NAGS*, we identified protective pathways where increased methylation led to higher gene expression, which was subsequently associated with a decreased risk of COPD (Figure S6–S8). These analyses provide robust evidence for gene-specific regulatory cascades where DNA methylation acts as an upstream regulator of COPD susceptibility. We also explored the potential causal link from gene expression to protein abundance, but found no significant effects (Table S4).

Table 1 Causal Associations of Molecular QTLs with COPD Risk Identified by SMR Analysis

Gene	Exposure Type	Exposure ID	OR (95% CI)	P-SMR	PPH4
GPXI	Methylation	cg24011261	1.18 (1.07–1.31)	1.13×10^{-3}	0.92
	Expression	eQTL	1.44 (1.18–1.76)	4.03×10^{-4}	0.76
BPHL	Methylation	cg26651049	0.80 (0.70–0.91)	1.03×10^{-3}	0.92
	Expression	eQTL	0.82 (0.69–0.97)	0.02	0.25
NAGS	Methylation	cg19579216	0.90 (0.82–0.98)	0.02	0.20
	Methylation	cg07803108	0.90 (0.82–0.98)	0.02	0.20
	Expression	eQTL	0.68 (0.48–0.95)	0.02	0.26

Abbreviations: SMR, Summary data-based Mendelian Randomization; QTL, Quantitative Trait Locus; OR, Odds Ratio; CI, Confidence Interval; PPH4, Posterior Probability of Hypothesis 4 (colocalization).

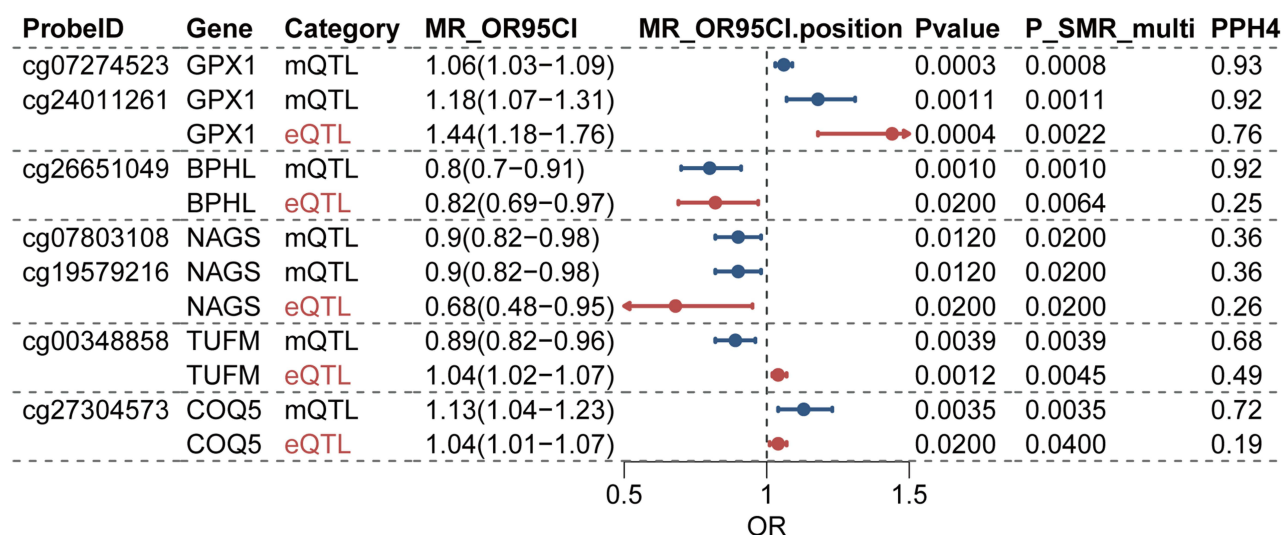


Figure 2 Summary of Mendelian Randomization (SMR) analysis for key mitochondria-related genes. The forest plot displays the odds ratios (ORs) and 95% confidence intervals (CIs) for the causal associations between genetic variants and COPD risk. Blue points represent methylation quantitative trait loci (mQTLs) and red points represent expression quantitative trait loci (eQTLs). The vertical dashed line indicates an OR of 1.0 (no effect). PPH4 indicates the posterior probability of colocalization, with higher values suggesting a shared causal variant. Red text indicates that this data represents gene expression (eQTL) results.

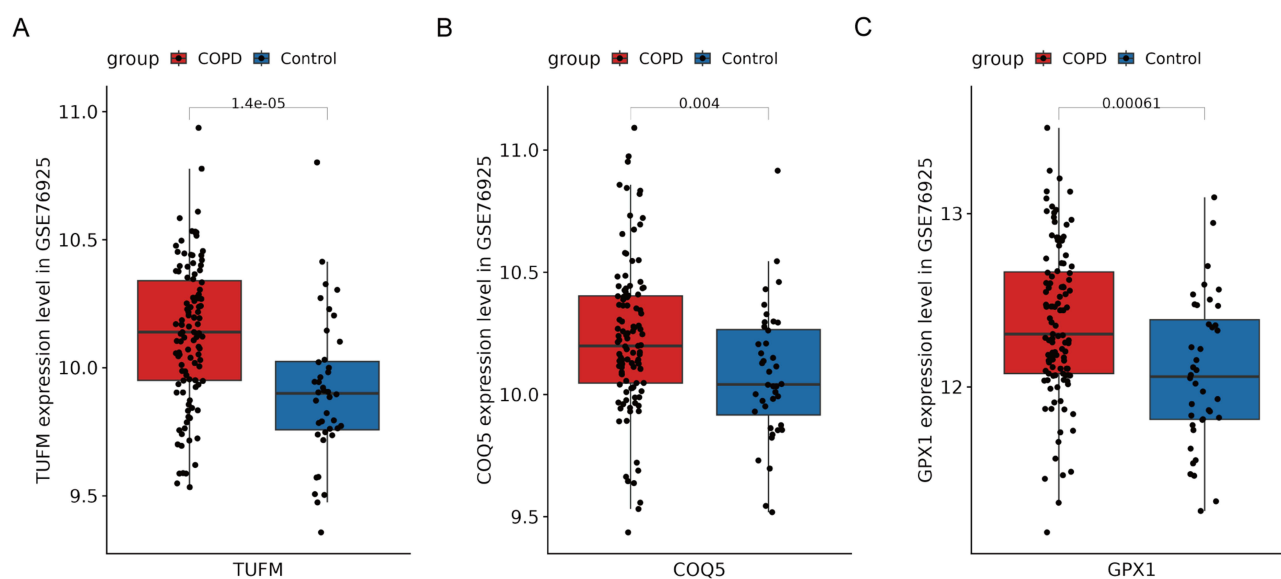


Figure 3 Transcriptomic validation of key causal genes in lung tissue. Box plots showing the differential expression of (A) *TUFM*, (B) *COQ5*, and (C) *GPX1* in lung tissue samples from COPD patients (red) and control subjects (blue). Expression data were obtained from the GEO dataset GSE76925. P-values from the comparison between the COPD and Control groups are shown for each gene, indicating significant upregulation in the COPD group.

Two-Sample MR and Replication Analyses Validate Key Findings

We further validated our top candidates using a formal two-sample Mendelian Randomization (MR) analysis (Figure S9). The IVW method confirmed significant causal associations for the expression of *GPX1*, *BPHL*, *TUFM*, and *COQ5* with COPD (Table 3). These findings were robust across multiple sensitivity analyses, which showed no evidence of significant horizontal pleiotropy (Figure S10; instrumental variables in Table S10 and Table S11; full sensitivity results in Table S12 and Table S13). Replication of our SMR findings in an independent cohort provided consistent, albeit limited, support (Table S7–Table S9). A summary of genes with multi-omics evidence is presented in Table 1, with robust validation from two-sample MR, lung tissue-specific SMR, and replication analyses detailed in Table 3 and Table S6.

Table 2 Multi-Layered Regulatory Pathway Analysis from Methylation to Gene Expression

Methylation Site (mQTL)	Target Gene (eQTL)	OR (95% CI)	P-SMR	P-SMR Multi	P-HEIDI
cg24011261	GPXI	1.57 (1.35–1.82)	1.94×10^{-9}	1.94×10^{-9}	0.11
cg26651049	BPHL	1.74 (1.43–2.13)	4.52×10^{-8}	4.52×10^{-8}	0.34
cg19579216	NAGS	1.25 (1.14–1.36)	5.83×10^{-7}	5.83×10^{-7}	0.07
cg07803108	NAGS	1.25 (1.14–1.36)	7.47×10^{-7}	7.47×10^{-7}	0.30

Notes: This table demonstrates the causal effect of DNA methylation on gene expression levels. OR represents the change in gene expression per standard deviation increase in methylation.

Table 3 Validation and Replication of Causal Associations for Key Genes

Gene	Validation Method	Outcome	OR (95% CI)	P-value	Notes
GPXI	Two-Sample MR (IVW)	COPD	1.41 (1.22–1.63)	4.54×10^{-8}	Primary MR analysis
BPHL	Two-Sample MR (IVW)	COPD	0.82 (0.76–0.89)	7.31×10^{-8}	Primary MR analysis
NAGS	Two-Sample MR (IVW)	COPD	0.85 (0.70–1.02)	0.082	Primary MR analysis
TUFM	Two-Sample MR (IVW)	COPD	1.04 (1.02–1.06)	2.11×10^{-5}	Primary MR analysis
COQ5	Two-Sample MR (IVW)	COPD	1.03 (1.01–1.05)	0.003	Primary MR analysis
TUFM	Lung Tissue SMR	COPD	1.17 (1.06–1.29)	1.78×10^{-3}	Tissue-specific validation
COQ5	Lung Tissue SMR	COPD	1.08 (1.01–1.15)	2.43×10^{-2}	Tissue-specific validation
NAGS	Lung Tissue SMR	COPD	0.86 (0.77–0.97)	1.51×10^{-2}	Tissue-specific validation
GPXI	UK Biobank SMR (mQTL)	COPD	1.002 (1.00–1.004)	1.95×10^{-2}	Replication in UKB cohort
TUFM	UK Biobank SMR (mQTL)	COPD	0.994 (0.99–0.999)	1.26×10^{-2}	Replication in UKB cohort

Notes: This table summarizes results from multiple validation approaches. Two-Sample MR (IVW) provides robust causal estimates. Lung Tissue SMR confirms tissue-specific effects. UK Biobank SMR provides replication in an independent cohort.

Functional Network Analysis Links Causal Genes to a Core Mitochondrial Dysfunction Module

Finally, to understand the functional context of the identified genes, we performed a Weighted Gene Co-expression Network Analysis (WGCNA). This analysis identified a turquoise module strongly correlated with COPD status and the expression of *GPXI*, *COQ5*, and *TUFM* (Figure S11 and Table S14 and Table S15). Genes within this module were significantly enriched in key mitochondrial pathways, including “Oxidative phosphorylation” and “Respiratory chain complex” (Figure S12 and Table S16 and Table S17), functionally implicating our identified causal genes in a coordinated network of mitochondrial dysfunction in COPD.

Discussion

This study, by integrating multi-omics genetic evidence, has for the first time systematically identified the key roles of multiple mitochondrial-associated genes in the pathogenesis of COPD at the causal level. We not only revealed the causal association between the genetic regulation of genes such as *GPXI*, *TUFM*, *COQ5*, *BPHL*, and *NAGS* and COPD risk, but also confirmed, through independent lung tissue transcriptome data and gene co-expression network analysis, the expression changes and functional synergy of these genes in the real pathological environment. This provides multi-level, verifiable biological evidence for the pathophysiological mechanisms of mitochondrial dysfunction in COPD.

One of the strongest pieces of evidence in this study points to the critical role of epigenetic modifications in the pathogenesis of COPD in concert with the synergistic regulation of gene transcription. Among them, the *GPXI* gene provides

a typical example: hypermethylation of the cg24011261 site in its promoter region was significantly associated with the up-regulation of *GPXI* mRNA expression, whereas the increase of *GPXI* expression was closely associated with the elevated risk of COPD. A clear regulatory chain of “high methylation promotes high GPX1 expression, which in turn increases the risk of COPD” was formed. *GPXI* encodes Glutathione Peroxidase 1, a key antioxidant enzyme involved in scavenging Reactive Oxygen Species (ROS). Study have shown that the concentration of *GPXI* is significantly reduced in COPD patients.³⁸ Notably, this study identified a significant association between genetically driven overexpression of the *GPXI* gene and increased COPD risk. This finding appears paradoxical, as *GPXI*, a key member of the glutathione peroxidase family,³⁹ is typically recognised for its antioxidant protective role in scavenging ROS. However, we hypothesise that this “risk-associated upregulation” may reflect a decompensatory compensatory mechanism. Under prolonged oxidative stress (such as smoking or air pollution), lung tissue cells may respond to persistent ROS assault by upregulating GPX1 transcription. However, this compensation may prove dysfunctional or ultimately fail: on the one hand, highly expressed *GPXI* protein may undergo misfolding or inactivation due to oxidative damage; on the other, the reduced glutathione (GSH) essential for its catalytic reaction⁴⁰ may be substantially depleted, limiting enzymatic activity. Consequently, despite elevated mRNA levels, actual antioxidant capacity remains markedly diminished. Against this backdrop, elevated *GPXI* expression no longer signifies enhanced antioxidant defence but instead serves as a biomarker for persistent oxidative damage and mitochondrial dysfunction, indicating the disease has progressed to an irreversible pathological stage. This interpretation aligns with our observation of *GPXI* upregulation in an independent lung tissue dataset (GSE76925), further supporting its biological relevance in authentic pathological processes.

Furthermore, the causal association between *TUFM* and *COQ5* provides additional support for the pivotal role of mitochondrial energy metabolism in COPD pathogenesis. *TUFM* (Tu Translation Elongation Factor, Mitochondrial) encodes a key translation elongation factor localised to the mitochondrial matrix,⁴¹ participating in the translation and assembly of respiratory chain subunits encoded by mitochondrial DNA.⁴² This study revealed through SMR analysis that elevated *TUFM* gene expression levels are significantly associated with increased COPD risk, a correlation validated by eQTL data from lung tissue. This finding suggests that *TUFM* upregulation is not merely a compensatory response but may reflect disruption within the mitochondrial protein synthesis system. Under sustained oxidative stress, mitochondrial DNA becomes vulnerable to damage, leading to the accumulation of misfolded proteins. Cells may attempt to maintain respiratory chain subunit synthesis efficiency by upregulating *TUFM*; however, this “overdrive” may instead exacerbate mitochondrial protein homeostasis imbalance, promoting a vicious cycle of mitochondrial dysfunction.

Similarly, *COQ5*, as a key methyltransferase in the coenzyme Q10 (*CoQ10*) biosynthetic pathway, is responsible for the hydroxylation modification of polyisoprenyl diene and is crucial for the integrity of the electron transport chain.⁴³ Our analysis indicates that genetically driven overexpression of *COQ5* similarly correlates positively with COPD risk. *CoQ10* functions not only as an electron carrier within the electron transport chain⁴⁴ but also as a lipophilic antioxidant.⁴⁵ Impairment of its synthesis reduces electron transfer efficiency, increasing electron leakage and ROS production; the abnormal overexpression of *COQ5* may thus indicate feedback dysregulation or failed functional compensation within this pathway.⁴⁶ Notably, analysis of the GEO dataset (GSE76925) confirmed significant upregulation of both *COQ5* and *TUFM* in COPD patient lung tissue. WGCNA further revealed their co-enrichment with *GPXI* within a highly interconnected turquoise co-expression module, which showed a significant positive correlation with the COPD phenotype ($r = 0.43$, $p < 0.001$). This module exhibited significant enrichment in KEGG pathways related to mitochondrial energy metabolism, such as “oxidative phosphorylation”.

More importantly, multiple genes within this module (such as *NDUFAF2*, *UQCRCQ*, *COX7A2*) are directly involved in the assembly and function of respiratory chain complexes I, III, and IV, indicating that this co-expression network represents a systemic mitochondrial dysfunction programme. Combined with DNA methylation regulation revealed by SMR, we hypothesise that epigenetic reprogramming may drive the coordinated dysfunction of this entire functional module. Consequently, *GPXI*, *TUFM*, and *COQ5* are not isolated risk factors but critical nodes within the mitochondrial “function-defence-regulation” network. Their concurrent upregulation may constitute a positive feedback loop: disrupted energy metabolism leads to increased ROS, which in turn induces upregulation of antioxidant genes (*GPXI*) and repair genes (*TUFM*, *COQ5*). However, due to substrate depletion or protein damage, this compensatory mechanism ultimately

fails, exacerbating mitochondrial injury and cellular dysfunction. This ultimately drives characteristic COPD pathologies such as airway inflammation and alveolar destruction.

Moreover, key genes identified in this study, such as *GPXI* and *TUFM*, are enriched in pathways related to oxidative stress and mitochondrial function. Environmental factors, particularly exposure to fine particulate matter like PM_{2.5}, have been extensively demonstrated as significant risk factors for COPD,⁴⁷ with their core pathogenic mechanism being the induction of pulmonary oxidative stress and mitochondrial damage. Consequently, we may reasonably hypothesise that individuals harbouring the *GPXI* risk allele may exhibit heightened oxidative stress responses and more severe lung function decline upon PM_{2.5} exposure, indicating a potential gene-environment interaction. For instance, environmental toxins may further deplete GSH,⁴⁸ exacerbate *GPXI* dysfunction, or amplify its dysregulation through epigenetic mechanisms such as DNA methylation. This provides crucial insights for future combined environmental epidemiology and genetic susceptibility studies, suggesting that early screening and intervention for individuals carrying specific genetic risks may hold significant clinical importance in highly polluted regions.

The primary strength of this study lies in its rigorous application of SMR framework, integrating genomic, epigenomic, transcriptomic, and proteomic data to enhance the reliability of causal inference. Colocalization analysis excluded false-positive associations and revealed a regulatory cascade where DNA methylation modulates gene expression, thereby influencing COPD risk (eg, *GPXI*). More importantly, we achieved multi-level validation from genetic association to tissue expression and biological pathways through dual-sample MR, independent lung tissue expression data (GSE76925) validation, and WGCNA functional network analysis, significantly enhancing the biological credibility of our findings. Nevertheless, several limitations remain. Firstly, primary QTL data primarily derive from blood samples, potentially failing to fully reflect lung tissue-specific regulatory mechanisms. Although we partially mitigated this issue through lung tissue eQTL validation and GSE76925 transcriptomic data, tissue-specific regulatory differences warrant further exploration in future studies. Secondly, the relatively limited sample size of the current lung tissue eQTL dataset constrained statistical power; replication and expansion should be pursued in larger cohorts or at single-cell resolution. Finally, this study did not directly model gene-environment interactions. It is well established that smoking and PM_{2.5} exposure constitute core environmental risk factors for COPD, with their pathogenic mechanisms centring on inducing oxidative stress and mitochondrial damage - precisely the key pathways identified in this study. Consequently, individuals harbouring risk alleles for *GPXI* or *TUFM* may exhibit heightened genetic susceptibility when exposed to these environmental factors, suggesting a potential “gene-environment interaction”. Future research urgently requires integrating detailed environmental exposure data to systematically evaluate such interactions. This will enable the identification of high-risk subpopulations, thereby advancing precision prevention and personalised intervention strategies.

Despite these limitations, our findings have significant implications for future research and clinical translation. The robustly identified causal genes, particularly *TUFM* and *COQ5* which were validated in lung tissue, represent high-priority targets for the development of novel therapeutics aimed at restoring mitochondrial function. Moreover, specific methylation sites like cg24011261 in *GPXI* could be developed as potential biomarkers for early disease detection or risk stratification. Future studies should focus on validating these targets in experimental models and exploring the gene-smoking interactions in large, well-phenotyped patient cohorts to pave the way for personalized medicine in COPD.

Conclusions

In summary, this study systematically identified the causal roles of multiple mitochondrial-associated genes (*GPXI*, *TUFM*, *COQ5*, *BPHL*, *NAGS*) in COPD pathogenesis by integrating multi-omics genetic evidence. We found that DNA methylation may participate in disease development by regulating gene expression, with the upregulation of *GPXI* potentially representing a dysregulated antioxidant compensatory mechanism. These causal associations were validated in dual-sample MR analyses and confirmed by differential expression of corresponding genes in independent COPD lung tissue transcriptomic data (GSE76925). Further WGCNA analysis revealed that these key genes exhibit high co-expression, clustering within a functional module significantly associated with the COPD phenotype. This module demonstrated marked enrichment in core mitochondrial pathways, including oxidative phosphorylation. Consequently, this study not only establishes causality at the genetic level but also reveals biological coherence at the transcriptional

regulation and gene network levels. It provides systematic evidence for understanding the mitochondrial mechanisms of COPD and lays a robust foundation for future functional studies and therapeutic target development.

Abbreviations

COPD, Chronic obstructive pulmonary disease; GWAS, Genome-wide association study; eQTLs, Expression quantitative trait loci; mQTLs, Methylation QTLs; pQTLs, Protein QTLs; ROS, Reactive oxygen species; mtDNA, Mitochondrial DNA; MR, Mendelian Randomization.

Data Sharing Statement

All data generated or analyzed during this study are included in this article and [supplementary information files](#).

Ethics Approval and Informed Consent

This study utilizes publicly available, anonymized summary-level data from GWAS catalogs, eQTLGen, and other public repositories. The original studies obtained ethical approval and informed consent from all participants. This study was confirmed to be exempt from ethical review by the Medical Ethics Committee of Ningbo Municipal Hospital of Traditional Chinese Medicine (Approval No. LW-2025-043), in accordance with Article 32, Item 1 and Item 2 of the “Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects” issued by the National Health Commission of the People’s Republic of China (February 18, 2023).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by NINGBO Medical & Health Leading Academic Discipline Project (2022-Z08).

Disclosure

The authors have declared that no competing interests exist in this work.

References

1. Labaki WW, Rosenberg SR. Chronic obstructive pulmonary disease. *Ann Intern Med.* 2020;173(3):Itc17–itc32. doi:10.7326/AITC202008040
2. Rabe KF, Watz H. Chronic obstructive pulmonary disease. *Lancet.* 2017;389(10082):1931–1940. doi:10.1016/S0140-6736(17)31222-9
3. Lytras T, Kogevinas M, Kromhout H, et al. Occupational exposures and 20-year incidence of COPD: the European Community Respiratory Health Survey. *Thorax.* 2018;73(11):1008–1015. doi:10.1136/thoraxjnl-2017-211158
4. Antunes MA, Lopes-Pacheco M, Rocco PRM. Oxidative stress-derived mitochondrial dysfunction in chronic obstructive pulmonary disease: a concise review. *Oxid Med Cell Longev.* 2021;2021:6644002. doi:10.1155/2021/6644002
5. Wiegman CH, Michaeloudes C, Haji G, et al. Oxidative stress-induced mitochondrial dysfunction drives inflammation and airway smooth muscle remodeling in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol.* 2015;136(3):769–780. doi:10.1016/j.jaci.2015.01.046
6. Giordano L, Gregory AD, Pérez Verdaguer M, et al. Extracellular release of mitochondrial DNA: triggered by cigarette smoke and detected in COPD. *Cells.* 2022;11(3):369. doi:10.3390/cells11030369
7. Meyer A, Zoll J, Charles AL, et al. Skeletal muscle mitochondrial dysfunction during chronic obstructive pulmonary disease: central actor and therapeutic target. *Exp Physiol.* 2013;98(6):1063–1078. doi:10.1113/expphysiol.2012.069468
8. Ng Kee Kwong F, Nicholson AG, Harrison CL, Hansbro PM, Adcock IM, Chung KF. Is mitochondrial dysfunction a driving mechanism linking COPD to non-small cell lung carcinoma? *Eur Respir Rev.* 2017;26(146):170040. doi:10.1183/16000617.0040-2017
9. Caldeira DAF, Weiss DJ, Rocco PRM, Silva PL, Cruz FF. Mitochondria in focus: from function to therapeutic strategies in chronic lung diseases. *Front Immunol.* 2021;12:782074. doi:10.3389/fimmu.2021.782074
10. Karim L, Kosmider B, Bahmed K. Mitochondrial ribosomal stress in lung diseases. *Am J Physiol Lung Cell Mol Physiol.* 2022;322(4):L507–L517. doi:10.1152/ajplung.00078.2021
11. Peng H, Yang M, Chen ZY, et al. Expression and methylation of mitochondrial transcription factor a in chronic obstructive pulmonary disease patients with lung cancer. *PLoS One.* 2013;8(12):e82739. doi:10.1371/journal.pone.0082739

12. Selemidis S. Proteomic and other ‘-omic’ analyses to develop disease stage-specific platforms and therapeutic strategies for COPD: it is about time. *Respirology*. 2021;26(10):904–905. doi:10.1111/resp.14133
13. Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27(11):3253–3265. doi:10.1681/ASN.2016010098
14. Vösa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–1310. doi:10.1038/s41588-021-00913-z
15. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374(6569):eabj1541. doi:10.1126/science.abj1541
16. Qi T, Wu Y, Zeng J, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun*. 2018;9(1):2282.
17. Hukku A, Pividori M, Luca F, Pique-Regi R, Im HK, Wen XJTAJo HG. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am J Human Genetics*. 2021;108(1):25–35. doi:10.1016/j.ajhg.2020.11.012
18. Pairo-Castineira E, Rawlik K, Bretherick AD, et al. GWAS and meta-analysis identifies 49 genetic variants underlying critical COVID-19. *Nature*. 2023;617(7962):764–768. doi:10.1038/s41586-023-06034-3
19. Chen H, Zhang Y, Li S, et al. The association between genetically predicted systemic inflammatory regulators and polycystic ovary syndrome: a mendelian randomization study. *Front Endocrinol*. 2021;12:731569. doi:10.3389/fendo.2021.731569
20. Li J, Tang M, Gao X, Tian S, Liu W. Mendelian randomization analyses explore the relationship between cathepsins and lung cancer. *Commun Biol*. 2023;6(1):1019. doi:10.1038/s42003-023-05408-7
21. Kim JY, Song M, Kim MS, et al. An atlas of associations between 14 micronutrients and 22 cancer outcomes: mendelian randomization analyses. *BMC Med*. 2023;21(1):316. doi:10.1186/s12916-023-03018-y
22. Palmer TM, Lawlor DA, Harbord RM, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res*. 2012;21(3):223–242. doi:10.1177/0962280210394459
23. Clarke L, Zheng-Bradley X, Smith R, et al. The 1000 Genomes Project: data management and community access. *Nat Methods*. 2012;9(5):459–462. doi:10.1038/nmeth.1974
24. Kulinskaya E, Dollinger MB. An accurate test for homogeneity of odds ratios based on Cochran’s Q-statistic. *BMC Med Res Methodol*. 2015;15:49. doi:10.1186/s12874-015-0034-x
25. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol*. 2016;45(6):1961–1974. doi:10.1093/ije/dyw220
26. 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):693–698. doi:10.1038/s41588-018-0099-7
27. Dong R, Zhang Q, Peng H. Gastroesophageal reflux disease and the risk of respiratory diseases: a Mendelian randomization study. *J Transl Med*. 2024;22(1):60. doi:10.1186/s12967-023-04786-0
28. Spiller W, Bowden J, Sanderson E. Estimating and visualising multivariable Mendelian randomization analyses within a radial framework. *PLoS Genet*. 2024;20(12):e1011506. doi:10.1371/journal.pgen.1011506
29. 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):304–314. doi:10.1002/gepi.21965
30. Minelli C, Del Greco MF, van der Plaats DA, et al. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int J Epidemiol*. 2021;50(5):1651–1659. doi:10.1093/ije/dyab084
31. Liu X, Yu H, Yan G, Sun M. Role of blood lipids in mediating the effect of dietary factors on gastroesophageal reflux disease: a two-step mendelian randomization study. *Eur J Nutr*. 2024;63(8):3075–3091. doi:10.1007/s00394-024-03491-y
32. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol*. 2016;1418:93–110.
33. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
34. Feng T, Lai C, Zhong D, et al. Weighted gene co-expression network analysis reveals prognostic and diagnostic significance of PAQR4 in patients with early and late hepatocellular carcinoma. *J Gastrointest Oncol*. 2022;13(2):768–779. doi:10.21037/jgo-22-168
35. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium Nat Genet*. 2000;25(1):25–29.
36. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 2023;51(D1):D587–d592. doi:10.1093/nar/gkac963
37. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genet*. 2016;48(5):481–487. doi:10.1038/ng.3538
38. Zinellu E, Zinellu A, Pau MC, et al. Glutathione peroxidase in stable chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Antioxidants*. 2021;10(11).
39. Pei J, Pan X, Wei G, Hua Y. Research progress of glutathione peroxidase family (GPX) in redoxiation. *Front Pharmacol*. 2023;14:1147414. doi:10.3389/fphar.2023.1147414
40. Handy DE, Loscalzo J. The role of glutathione peroxidase-1 in health and disease. *Free Radic Biol Med*. 2022;188:146–161. doi:10.1016/j.freeradbiomed.2022.06.004
41. Liu N, Pang B, Kang L, Li D, Jiang X, Zhou CM. TUFM in health and disease: exploring its multifaceted roles. *Front Immunol*. 2024;15:1424385. doi:10.3389/fimmu.2024.1424385
42. Zhan J, Jin K, Xie R, et al. AGO2 protects against diabetic cardiomyopathy by activating mitochondrial gene translation. *Circulation*. 2024;149(14):1102–1120. doi:10.1161/CIRCULATIONAHA.123.065546
43. Dawidziuk M, Podwysocka A, Jurek M, et al. Congenital coenzyme Q5-linked pathology: causal genetic association, core phenotype, and molecular mechanism. *J Appl Genet*. 2023;64(3):507–514. doi:10.1007/s13353-023-00773-9
44. Al Saadi T, Assaf Y, Farwati M, et al. Coenzyme Q10 for heart failure. *Cochrane Database Syst Rev*. 2021;(2)(2):Cd008684. doi:10.1002/14651858.CD008684.pub3
45. Rauchová H. Coenzyme Q10 effects in neurological diseases. *Physiol Res*. 2021;70(Suppl4):S683–s714. doi:10.33549/physiolres.934712

46. Widmeier E, Yu S, Nag A, et al. ADCK4 deficiency destabilizes the coenzyme q complex, which is rescued by 2,4-dihydroxybenzoic acid treatment. *J Am Soc Nephrol.* 2020;31(6):1191–1211. doi:10.1681/ASN.2019070756
47. Fan X, Dong T, Yan K, Ci X, Peng L. PM2.5 increases susceptibility to acute exacerbation of COPD via NOX4/Nrf2 redox imbalance-mediated mitophagy. *Redox Biol.* 2023;59:102587. doi:10.1016/j.redox.2022.102587
48. Xu Y, Li Y, Li J, Chen W. Ethyl carbamate triggers ferroptosis in liver through inhibiting GSH synthesis and suppressing Nrf2 activation. *Redox Biol.* 2022;53:102349. doi:10.1016/j.redox.2022.102349

International Journal of Chronic Obstructive Pulmonary Disease

Dovepress
Taylor & Francis Group

Publish your work in this journal

The International Journal of COPD is an international, peer-reviewed journal of therapeutics and pharmacology focusing on concise rapid reporting of clinical studies and reviews in COPD. Special focus is given to the pathophysiological processes underlying the disease, intervention programs, patient focused education, and self management protocols. This journal is indexed on PubMed Central, MedLine and CAS. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-chronic-obstructive-pulmonary-disease-journal>