scientific reports

OPEN

The relationship between genetic prediction of 486 blood metabolites and the risk of COPD: mendelian randomization study

Yang Zhang[1,4], Xinlai Ma[2,4], Fan Wu[2], Yuxiao Sun[2], Hongyu Mou[2], Xue Liu[1] & Wei Zhang[1,3]✉

Metabolic disorders are an important feature of chronic lung disease. Patients diagnosed with chronic obstructive pulmonary disease (COPD) have been found to experience metabolic disorders. Nonetheless, evidence on the causal role of circulating metabolites in promoting or preventing COPD is still lacking. Conducting a methodical examination on the causal connection between blood metabolites and COPD can aid in identifying fresh objectives for the screening and prevention of COPD. Therefore, we performed a two-sample Mendelian randomization (MR) analysis to evaluate the causal association between COPD and 486 blood metabolites. We used two-sample MR techniques and genome-wide association study (GWAS) data to evaluate the correlation between COPD and 486 serum metabolites. To evaluate the causal impact of serum metabolites on the risk of COPD, we predominantly employed inverse variance weighting (IVW) methodology. The MR-Egger regression test was employed to assess multiple validity, while the presence of heterogeneity was examined using the Cochran's Q test. To ensure the reliability of the findings, a leave-one-out analysis was conducted. The Bonferroni correction is used to adjust for multiple comparisons, ensuring rigorous validation of our results. After filtering by IVW and sensitivity analysis, we identified 10 known metabolites including fructose, margarate (17:0), guanosine, 2-stearoylglycerophosphocholine, hexadecanedioate, lactate, 5-oxoproline, paraxanthine, phenyllactate (PLA) and N-acetylglycine. Of these, fructose, margarate (17:0), guanosine, 2-stearoylglycerophosphocholine and hexadecanedioate are risk metabolites, and additionally, lactate, 5-oxoproline, paraxanthine phenyllactate(PLA) and N-acetylglycine are protective metabolites. In addition, the study identified five currently unknown chemical structures. Cochran's Q-test showed no significant heterogeneity, and MR Egger's intercept analysis confirmed the absence of horizontal multidirectionality. Leave-one-out analysis also proved the reliability of the MR analysis. We identified seven COPD-related risks and eight protective human serum metabolites. By combining genomics and metabolism, it provides new insights into the underlying mechanisms of COPD, with important implications for COPD screening and prevention.

Keywords COPD, Blood metabolites, Mendelian randomization

COPD is a diverse respiratory disorder that is identified by continuous and gradual restriction of airflow, which involves abnormalities in the airways (such as bronchiectasis) and chronic symptoms related to breathing (like dyspnea, cough, and sputum)[1]. Globally, COPD has emerged as the third most common reason for mortality, with its occurrence rising in correlation with advancing age[2]. A FEV1/FVC ratio below 70% after bronchodilator use is considered the gold standard for diagnosing COPD[3]. Decreased lung function in the general population is associated with systemic biomarkers[4]. Certain blood metabolites have been viewed as clinical markers of acute exacerbations of COPD[5], however alterations in their quantities are not consistent in COPD[6]. Thus, evidence of causation is particularly important in the case of biomarkers[7].

In recent years, the emergence of metabolomics as a component of systems biology has provided a new approach to studying the underlying mechanisms of disease. Metabolomics, in particular, has the ability

[1]Department of Pulmonary and Critical Care Medicine, Shandong University of Traditional Chinese Medicine Affiliated Hospital, Jinan, China. [2]The First Clinical Medical College, Shandong University of Traditional Chinese Medicine, Jinan, China. [3]Department of Pulmonary and Critical Care Medicine, Shandong University of Traditional Chinese Medicine Affiliated Hospital, No. 16369, Jingshi Road, Lixia District, Jinan 250013, Shandong, China. [4]These authors contributed equally to this work and are co-first authors. ✉email: huxzhijia@126.com

to offer understanding into the biological processes of disease development through the detection of altered metabolites or metabolic pathways[8,9]. Several metabolomics studies have shown metabolic dysregulation in patients with COPD[10,11]. Genetic epidemiologic studies of COPD show associations between COPD-related genetic polymorphisms and blood biomarkers[12]. Nevertheless, the number of biomarkers linked to COPD that have been discovered so far is limited. While many researchers have previously utilized proteomic methods to discover fresh indicators, like plasma sRAGE for detecting the existence and advancement of emphysema[13], others have employed metabolomic methods to identify potential markers of disease severity or therapeutic candidates. These markers include various types of lipids and derivatives (primarily phospholipids, but also ceramides, fatty acids, and arachidonoids), amino acids, coagulation factors, and nucleic acid components. These substances may play a role in their functions, proteolytic metabolism, energy production, oxidative stress, immune-inflammatory responses, and coagulation disorders[14]. Dysregulation of sphingolipid metabolism is frequently observed in individuals suffering from chronic obstructive pulmonary disease. There is increasing evidence indicating that sphingolipids have a crucial involvement in the development of several lung conditions, such as asthma, acute lung injury, emphysema, COPD, and cystic fibrosis[15,16]. Exacerbation of COPD is linked to decreased levels of sphingomyelin in the plasma, whereas elevated levels are connected to the rapid advancement of emphysema[17,18]. Furthermore, amino acids serve as the fundamental components of proteins and have a vital function in intermediary metabolism. Evidence suggests that dysregulated amino acid metabolism may be present in COPD patients, even during periods of rest. This indicates that an abnormal amino acid profile could potentially be a significant risk factor for COPD, as supported by research[19,20]. This suggests that targeting these metabolites may be a promising approach to treating COPD[21]. Regrettably, the precise connection between metabolites and COPD remains uncertain due to the absence of any prospective studies on metabolites and COPD thus far. The lack of clarity regarding the causal connection between metabolites and COPD arises from the inherent design flaws in conventional observational studies, including modifications in metabolites due to intentional lifestyle adjustments in patients following COPD diagnosis and alterations in metabolites caused by prolonged use of specific medications. Implementing rigorous randomized controlled trials (RCTs) is challenging due to ethical concerns, lengthy observation periods, costly funding, and various limitations; however, they hold the utmost credibility in showcasing causal effects in evidence-based medicine.

MR is a useful framework for studying causality through genetic tools (e.g., single nucleotide polymorphisms [SNPs])[22]. In the absence of randomized controlled trials, MR stands out as a very effective method for investigating causal relationships between exposures and outcomes of interest. In addition, MR helps to reduce confounding and reverse causality bias inherent in observational studies. In this study, we investigated the potential causality of genetically predicted 486 blood metabolites in COPD progression in an MR framework using publicly available GWAS summary statistics. The findings may provide additional evidence for the etiology of COPD.

Methods
Data on 486 blood metabolites and COPD from GWAS
This study utilized a genome-wide association study (GWAS) conducted by Shin et al. in 2014, which included 7,824 adults from two European cohorts, TwinsUK and KORA, involving 2,163,597 associated SNPs[23]. Fasting serum samples were metabolically analyzed by untargeted mass spectrometry and comprehensive high-performance liquid chromatography (HPLC) and gas chromatography-mass spectrometry (GC-MS), and 486 serum metabolites were successfully identified. Table S1 details the names of 486 metabolites, where X represents an unidentified chemical composition. Of these, 309 were identified through the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and categorized into eight biochemical classes: amino acids, carbohydrates, cofactors and vitamins, energy, lipids, nucleotides, peptides, and exogenous.The remaining 177 unknown metabolites were not further analyzed due to their uncertain chemical nature.Complete statistics for these genome-wide association summary datasets are publicly accessible via the Metabolomics GWAS server (https://metabolomics.helmholtz-muenchen.de/gwas/).The FinnGen biospecimen repository (https://www.finngen.fi/en) was used as an outcome variable for GWAS analysis of COPD. The analysis included genotype data from 6915 COPD patients and 186,723 controls. The data used in this study is publicly available and has been anonymized to ensure that it is not personally identifiable.All participants had obtained local ethical approval and provided informed consent, so no additional ethical approval was required for this study.

IVs selection
To assess the possible causal relationship between circulating metabolites and COPD, we performed a two-sample MR analysis[24]. This study examined blood metabolites and COPD as exposure and outcome factors. According to our analysis, for instrumental variables to be valid, three basic assumptions must be met[25], as shown in Fig. 1. First, genetic variants used as instrumental variables must be strongly correlated with blood metabolites. In addition, the genetic mutation must be independent of any potential confounding variables. In addition, genetic variants must affect COPD only through blood metabolites and not through other pathways. To satisfy these three hypotheses, we set the significance threshold for autonomy at $P < 1 \times 10 - 5$[26]. Furthermore, in order to guarantee the independence of SNPs and remove linkage disequilibrium (LD)[27], we established the LD threshold as r2 < 0.001 and a distance of 10,000 kb. We excluded SNPs significantly associated with known risk factors, including palindromic SNPs, using the PhenoScanner (http://www.phenoscanner.medschl.cam.ac.uk/) website[28].In the present study, we removed SNPs associated with smoking. This measure aims to eliminate SNPs that may bias MR estimates and ensure more stable and reliable results.
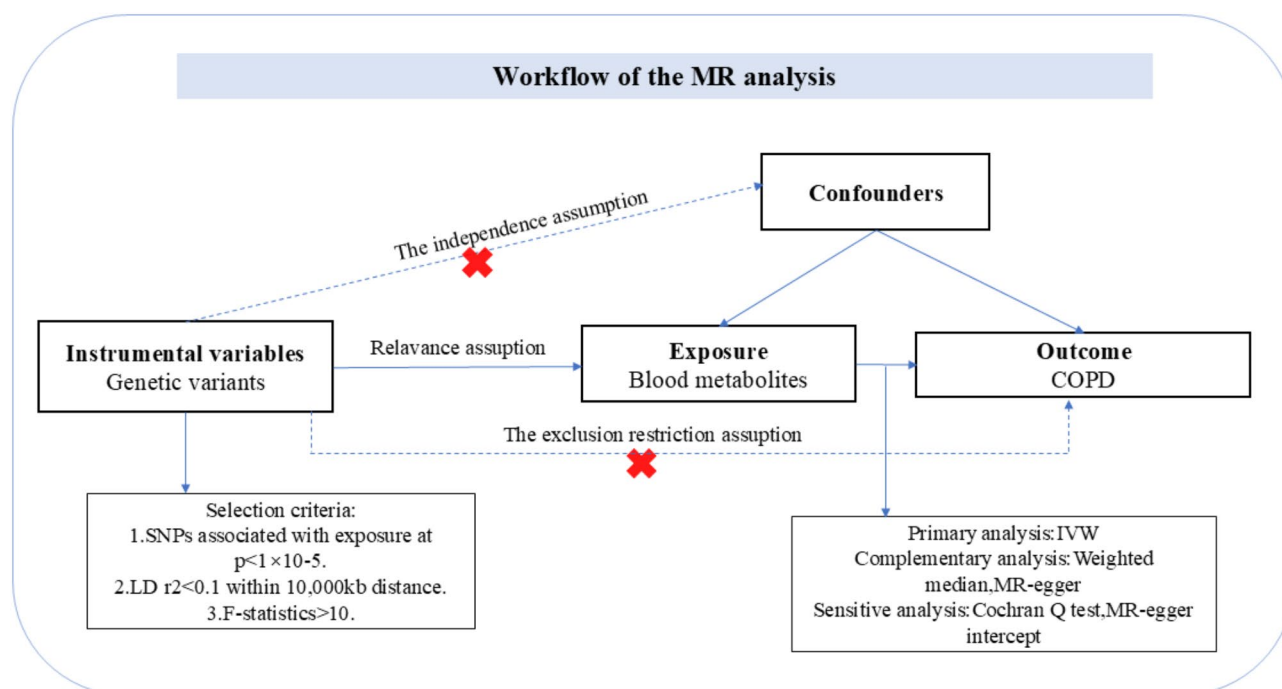
Fig. 1. Workflow of the MR analysis.

Statistical analysis

In order to guarantee the dependability and accuracy of the results, a variety of strong statistical techniques were utilized and sensitivity analyses were performed to evaluate the possible influence of different sources of bias. The MR inverse variance weighting (MR-IVW) method was used as the primary analytical method[29]. Subsequently, we applied the Bonferroni correction algorithm ($\alpha = 0.05$, $P_B = 0.05/486$, $P_B = 1e-4$) to correct for the type I error rate in multiple hypothesis testing and to identify metabolites with a significant causal association with COPD. metabolites with Bonferroni adjusted P-values lower than 1e-4 were considered statistically significant. However, due to the limited number of results, we adjusted the threshold to $1e-4 < P_B < 0.05$[30]. Fifteen metabolites with Bonferroni values between 1e-4 and 0.05 were categorized as potentially statistically significant. In order to evaluate the strength of the findings, we conducted supplementary sensitivity analyses using the weighted-median approach[31] and MR-Egger method[32]. Valid estimates were obtained using the weighted median approach, given that over 50% of the information was derived from valid instrumental variables (IV).

After Bonferroni correction, we used the MR-Egger method for assessing the presence of horizontal pleiotropy in the selected IVs[32], with $P > 0.05$ indicating no horizontal pleiotropy. We performed heterogeneity analyses using Cochran's Q-test to assess heterogeneity among the estimated SNPs effects, with $P > 0.05$ indicating no significant heterogeneity[33]. Furthermore, we conducted a sensitivity analysis by leaving out one SNP at a time to investigate if certain SNPs had a disproportionate impact on the overall estimation. In addition, to prevent the influence of weak IVs on the included IVs, we also computed the F-statistic ($F = beta^2/se^2$)[34,35]. All analyses were done in R software (version 4.3.2).

Results

After rigorous quality control of instrumental variables, 486 blood metabolites were obtained for MR analysis. In addition, LD analysis, echo sequence removal to improve accuracy, and reliability of the selected IVs were performed. Of the 486 blood metabolites, a total of 491 MR analyses were performed, and five of these metabolites were found to have two different forms. We set the threshold for the F-statistic to 10. The F-statistic value for all selected SNPs exceeded 10, suggesting that our IVs were sufficiently robust. 486 metabolites analyzed are shown in Table S2. After Bonferroni correction (P value < 1e-4), none of the metabolites were statistically significant. When the significance threshold was adjusted to 1e-4 < P value < 0.05, the IVW method analysis identified 15 circulating metabolites that may be causally related to COPD risk, with 10 known compounds and 5 unknown compounds. The relationship between the 15 metabolites and COPD is shown in the forest plot(Fig. 2).The metabolites lactate(OR = 0.224, 95%CI: 0.083-0.600,$P = 0.002$), fructose(OR = 5.374, 95%CI: 1.540-18.755,$P = 0.008$), margarate (17:0)(OR = 2.837, 95%CI: 1.049-7.870,$P = 0.039$), 5-oxoproline(OR = 0.320, 95%CI: 0.152 -0.675,$P = 0.002$), guanosine(OR = 1.347, 95%CI: 1.030–1.761,$P = 0.029$), paraxanthine(OR = 0.718, 95%CI: 0.551–0.935, $P = 0.013$), phenyllactate(PLA)(OR = 0.534, 95%CI: 0.296–0.962,$P = 0.036$), N-acetylglycine(OR = 0.725, 95%CI: 0.547–0.962, $P = 0.025$), X-10,810(OR = 0.611, 95%CI: 0.386–0.965,$P = 0.034$), X-02269(OR = 1.674, 95%CI: 1.062–2.638,$P = 0.026$), X-09789(OR = 1.478, 95%CI: 1.022–2.138,$P = 0.037$), X-11,552(OR = 0.664, 95%CI: 0.443–0.995,$P = 0.047$), 2-stearoylglycerophosphocholine (OR = 1.749, 95%CI: 1.023–2.990,$P = 0.040$), hexadecanedioate(OR = 1.340, 95%CI: 1.067–1.683,$P = 0.011$)

Metabolites	NSNP	Method	P.val	OR(95%CI)
5-oxoproline	22	Inverse variance weighted	0.002759178	0.32(0.15 to 0.68)
	22	MR Egger	0.061240569	0.04(0.00 to 0.76)
	22	Weighted median	0.064093449	0.28(0.07 to 1.08)
lactate	11	Inverse variance weighted	0.002931881	0.22(0.08 to 0.60)
	11	MR Egger	0.929894051	1.28(0.02 to 95.63)
	11	Weighted median	0.037382905	5.98(1.11 to 32.25)
X-14977	9	Inverse variance weighted	0.006858382	0.46(0.26 to 0.81)
	9	MR Egger	0.519259183	3.60(0.10 to 134.22)
	9	Weighted median	0.061112160	3.37(0.94 to 11.98)
fructose	3	Inverse variance weighted	0.008353094	5.37(1.54 to 18.76)
	3	MR Egger	0.016468454	0.19(0.06 to 0.66)
	3	Weighted median	0.001868103	0.21(0.08 to 0.56)
hexadecanedioate	24	Inverse variance weighted	0.011755594	1.34(1.07 to 1.68)
	24	MR Egger	0.263511836	1.59(0.73 to 3.46)
	24	Weighted median	0.009632391	1.61(1.12 to 2.31)
paraxanthine	13	Inverse variance weighted	0.013920970	0.72(0.55 to 0.94)
	13	MR Egger	0.088931323	0.47(0.21 to 1.04)
	13	Weighted median	0.104654439	0.74(0.52 to 1.06)
N-acetylglycine	14	Inverse variance weighted	0.025844850	0.73(0.55 to 0.96)
	14	MR Egger	0.283949924	0.41(0.09 to 1.97)
	14	Weighted median	0.113190065	0.53(0.24 to 1.16)
X-02269	8	Inverse variance weighted	0.026398576	1.67(1.06 to 2.64)
	8	MR Egger	0.037894272	0.52(0.31 to 0.90)
	8	Weighted median	0.024648416	0.66(0.45 to 0.95)
guanosine	13	Inverse variance weighted	0.029205092	1.35(1.03 to 1.76)
	13	MR Egger	0.103177667	0.55(0.28 to 1.08)
	13	Weighted median	0.104076607	0.60(0.33 to 1.11)
X-10810	16	Inverse variance weighted	0.034728397	0.61(0.39 to 0.97)
	16	MR Egger	0.014644141	8.05(2.41 to 26.86)
	16	Weighted median	0.178149298	1.43(0.85 to 2.39)
phenyllactate (PLA)	18	Inverse variance weighted	0.036829856	0.53(0.30 to 0.96)
	18	MR Egger	0.021880357	5.03(1.56 to 16.19)
	18	Weighted median	0.013592654	1.87(1.14 to 3.07)
X-09789	12	Inverse variance weighted	0.037818018	1.48(1.02 to 2.14)
	12	MR Egger	0.421583725	0.40(0.05 to 3.51)
	12	Weighted median	0.033749748	0.55(0.31 to 0.95)
margarate (17:0)	7	Inverse variance weighted	0.039981442	2.87(1.05 to 7.87)
	7	MR Egger	0.500305386	2.66(0.17 to 41.88)
	7	Weighted median	0.054310537	2.05(0.99 to 4.27)
2-stearoylglycerophosphocholine	13	Inverse variance weighted	0.040731447	1.75(1.02 to 2.99)
	13	MR Egger	0.044542576	1.46(1.03 to 2.06)
	13	Weighted median	0.020873593	1.45(1.06 to 1.99)
X-11552	16	Inverse variance weighted	0.047669679	0.66(0.44 to 1.00)
	16	MR Egger	0.467493441	0.60(0.16 to 2.20)
	16	Weighted median	0.135866058	0.57(0.27 to 1.19)

P<0.05 was considered statistically significant
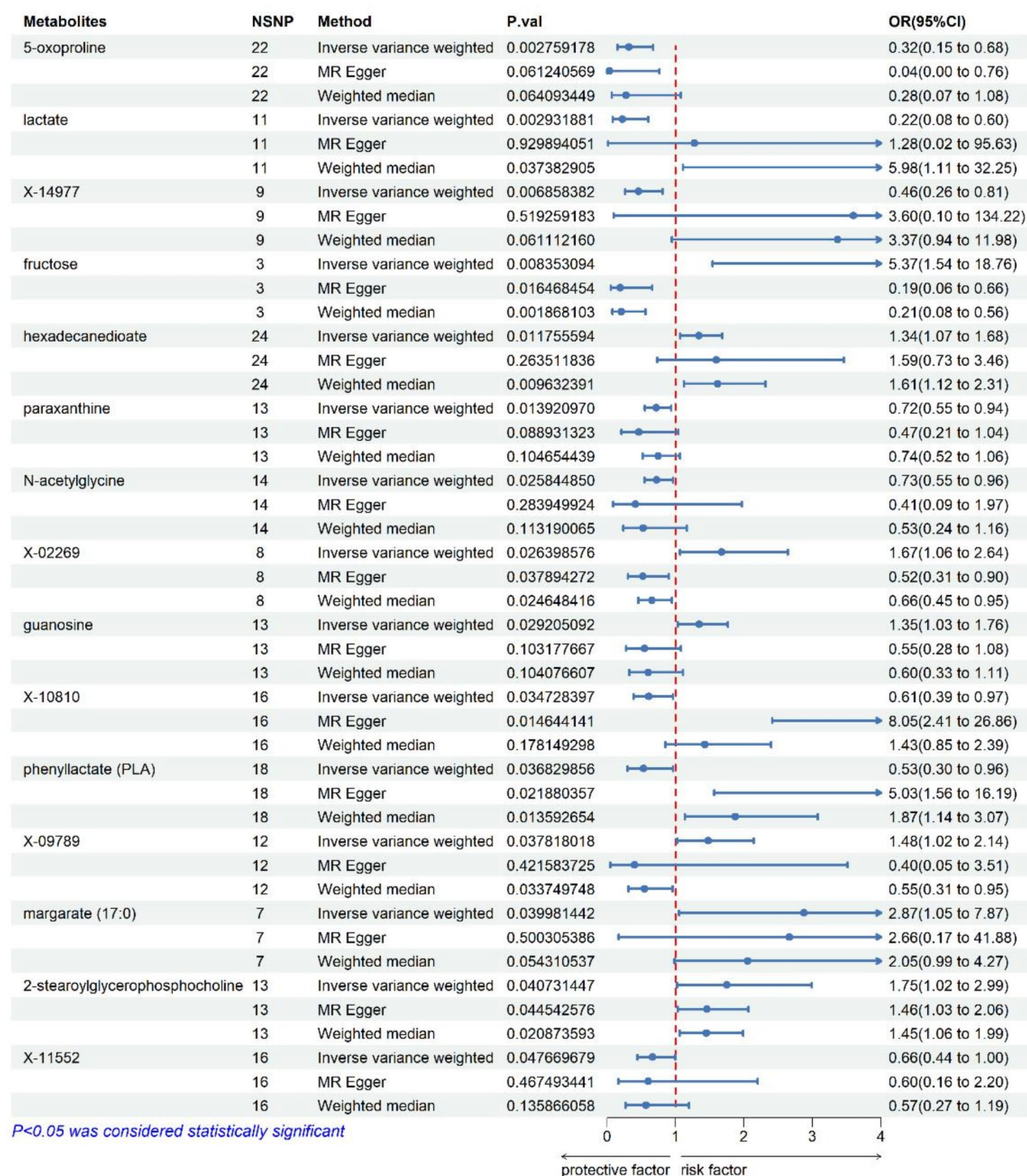
protective factor risk factor

Fig. 2. Forest plot of MR-analyzed blood metabolites on COPD causality. CI, confidence interval; OR, odds ratio; SNPs, single nucleotide polymorphisms.

and X-14,977(OR = 0.460, 95%CI: 0.262–0.808,P = 0.006) were analyzed using MR analyses, which included Cochran's Q-test and MR-Egger's intercept analysis(Table S3). Cochran's Q test assessed potential heterogeneity, especially for the metabolites studied, and showed no significant heterogeneity between IVs; furthermore, no outliers were found by MR-PRESSO. Although horizontal pleiotropy (MR-Egger $P<0.05$) was found in the MR study of the unknown compound X-02269, the results of the Cochrane's Q-test showed that no heterogeneity was found.This suggests that some confounding factors may still exist. Moreover, the leave-one-out method, depicted in Figure S1, validates the durability and consistency of the MR analysis. It demonstrates that removing any SNPs does not have a substantial impact on the overall results, thus reinforcing the trustworthiness and

stability of the MR analysis. The linear skewness in the scatterplot shows the degree of causality, as shown in Figs. 3 and 4.

Discussion

Our study aimed to investigate the possible causal impacts of 486 blood metabolites on the development of COPD by utilizing genetic proxies. Our analyses identified genetic metabolites linked to decreased risk of COPD by integrating two extensive GWAS datasets and utilizing a rigorous MR design. These metabolites include lactate, 5-oxoproline, paraxanthine, PLA, N-acetylglycine, X-10,810, X-11,552, X-14,977. Decreased chances of developing COPD are linked to higher levels of these metabolites. On the other hand, we discovered that individuals with a genetic predisposition to increased amounts of specific substances, such as fructose, margarate (17:0), guanosine, X-02269, X-09789, 2-stearoylglycerophosphocholine, and hexadecanedioate had a higher likelihood of developing COPD.

Early intervention is crucial in preventing and treating COPD, a chronic condition that has a lengthy progression. However early vague clinical symptoms make the diagnosis of COPD difficult. Hence, investigating the metabolites linked to the progression of COPD will not just aid in the initial detection and prevention of COPD, but also enhance comprehension of the biological mechanisms involved in treating COPD. Although there are associations between genetic, smoking and environmental factors and different clinical manifestations of COPD, the exact pathogenesis remains unclear. Timely detection is crucial in order to avoid permanent harm to the organs, highlighting the importance of dependable and pertinent biomarkers for enhancing the treatment of COPD. The use of metabolomics technology has generated curiosity in investigating the possible importance of metabolites in COPD. Metabolomics analysis is the systematic examination of small-sized biochemicals in biological samples, encompassing sugars, amino acids (AA), organic acids, nucleotides, and lipids[36]. According to the ECLIPSE cohort[37], a recent study using plasma metabolic profiling (PMP) revealed the correlation between cachexia and emphysema with various amino acids. Another study examined the correlation between 34 specific amino acids and dipeptides among various subcategories of individuals with COPD (including emphysema, airway disease, and cachexia)[21]. Bowler and his team discovered a connection between sphingolipids and ceramides with airflow blockage and emphysema in the group of patients from the COPDGene study. They found that five sphingolipids were linked to emphysema, while seven ceramides were associated with worsening of COPD symptoms[38]. This implies that plasma metabolomics may have a role in the clinical treatment of COPD, potentially aiding in outcome prediction, serving as an intervention tool, and indicating treatment response. Therefore, we carried out a significant MR investigation with the objective of clarifying the causal connection between blood metabolites and COPD, as well as the related metabolic routes. This study was designed to provide guidance for screening and management of COPD.

Studies have shown that elevated levels of fructose, margarate (17:0), guanosine, 2-stearoylglycerophosphocholine, and hexadecanedioate are associated with increased susceptibility to COPD. COPD patients are often associated with insulin resistance or abnormal glucose metabolism, which has also been validated by metabolomics techniques.It has been found that COPD patients have higher plasma levels of certain glucose metabolites, suggesting possible alterations in energy metabolism[39].It has been revealed that prolonged intake of fructose leads to the deterioration and restructuring of lung tissue and negatively affects breathing function. Long-term fructose exposure can lead to harmful effects such as airway hyperresponsiveness, chronic bronchitis, alveolar remodeling, and emphysema. Fructose ingestion causes an increase in the concentration of cytokines such as IL-10, IL-6, IL-1β, and TNFα, inducing chronic inflammatory responses that propagate to the lung parenchyma and lead to the accumulation of monocytes in the lung tissue[40]. Neopurine (one of the metabolites of guanosine) levels may dose-dependently increase oxidative stress in endothelial cells in hypoxic environments, leading to endothelial dysfunction[41]. Several studies have found high levels of lipid substances such as sphingomyelin and eicosadienoic acid in the alveoli and blood of COPD patients, suggesting that lipid metabolism may play an important role in chronic inflammation and airway remodeling in COPD[42]. Margarate is a long-chain saturated fatty acid that is an important component of cell membrane phospholipids. Excessive intake of saturated fatty acids (e.g., heptadecanoic acid) may exacerbate COPD by activating inflammatory signaling pathways and promoting the release of inflammatory factors. An animal study found that saturated fatty acids increased the number of lung macrophages and enhanced high-fat diet-induced neutrophil airway inflammation in a high-fat diet mouse model[43]. 2-stearoilglycerophosphocholine is a phospholipid molecule, which belongs to the class of phospholipids that are important components of biological membranes, especially in lipid bilayer structures.In addition, as a diacid fatty acid derivative, it may play a role in lipid metabolism.Our findings that 2-stearoilglycerophosphocholine and Hexadecanedioate may exacerbate chronic inflammation and airway remodeling in the lungs through their involvement in lipid metabolism are consistent with previous studies. Moreover, lipids have a significant impact on the functioning of the lungs. Recent studies have convincingly demonstrated the binding of lipids to inflammation. The oxidation of different biomolecules, such as lipids, resulting in epithelial cell damage and death, which is a crucial factor in the development of COPD, may be caused by both exogenous and endogenous reactive oxygen species produced by inflammation and mitochondrial dysfunction[44]. Apoptosis of endothelial cells is a crucial occurrence in the progression of emphysema. According to the research, it is indicated that these mechanisms could be linked to increased concentrations of ceramides in the respiratory system of individuals with COPD, which are regarded as an indicator of the condition. It is important to mention that there could be multiple routes for ceramide production, which can involve the activity of acidic sphingomyelinase and the synthesis of palmitate upon exposure to cigarette smoke[45]. This suggests that this taxonomic group could potentially be utilized as a novel target for therapeutic purposes. Nevertheless, additional investigation is required to further understand its particular function.

Besides the unidentified blood metabolites, this MR investigation discovered five blood metabolites (lactate, 5-oxoproline, paraxanthine, PLA, and N-acetylglycine) that are thought to be protective against COPD. The
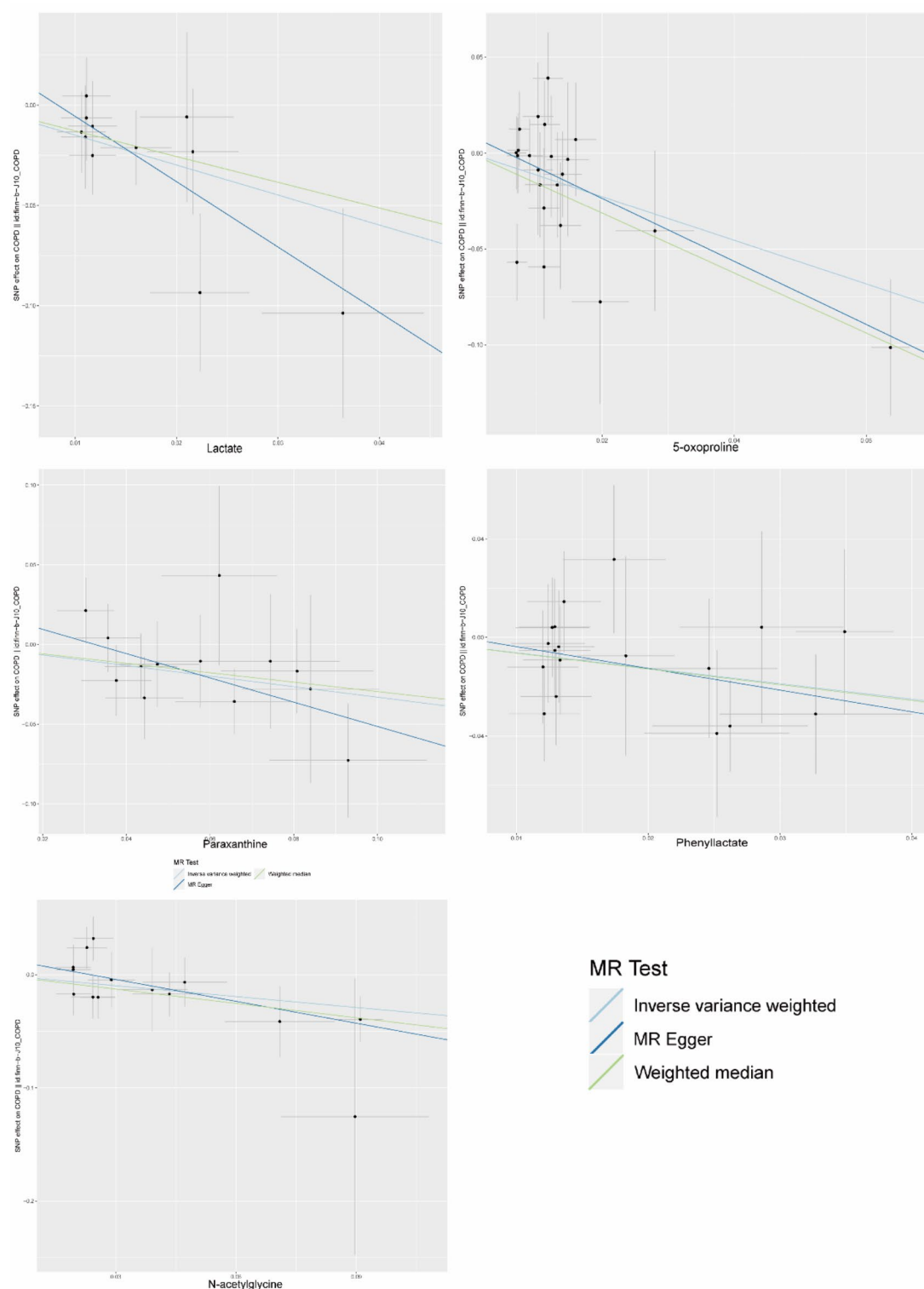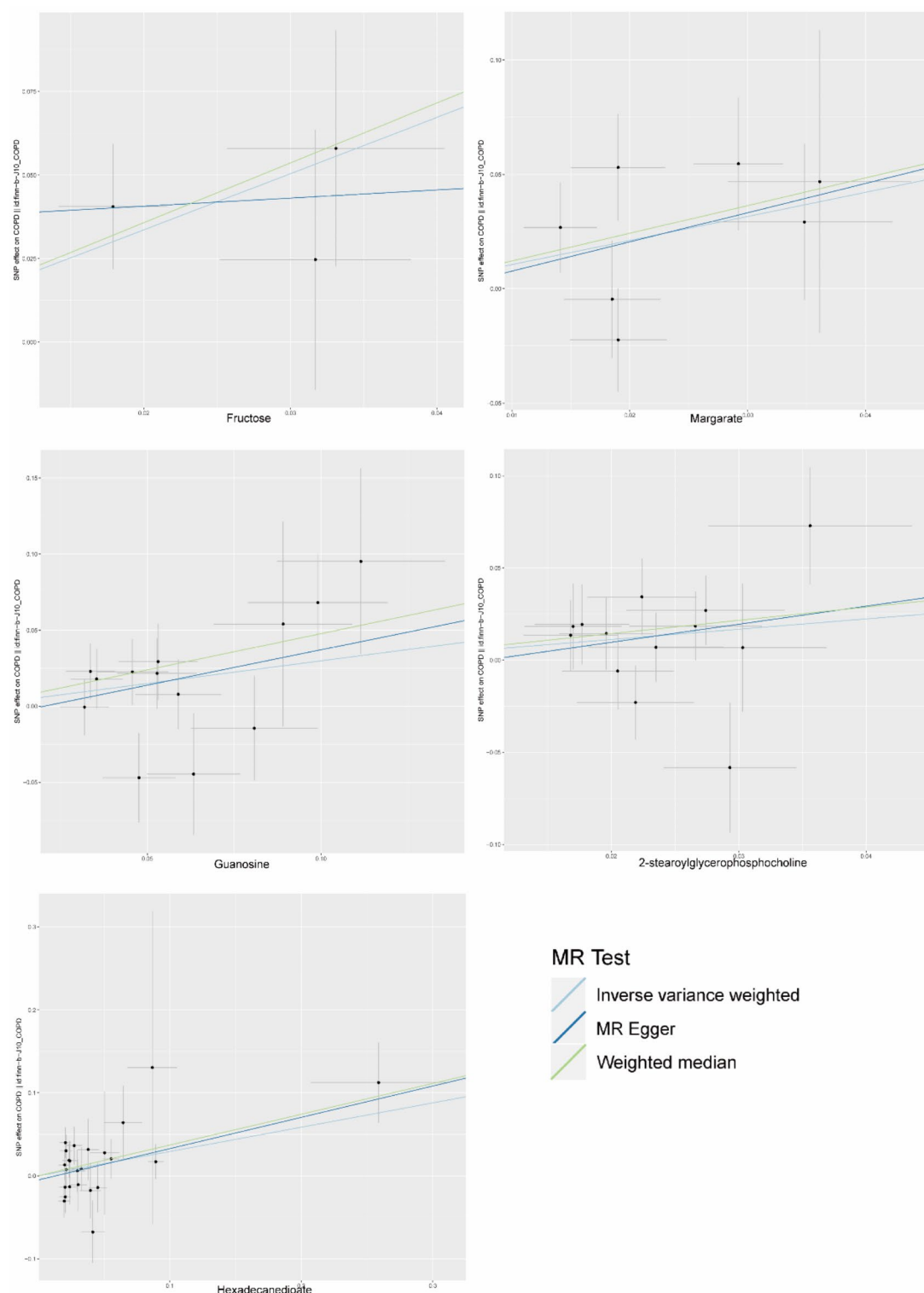
Fig. 3. Scatterplot of MR analysis of metabolites negatively associated with COPD risk. The extent of the cause-and-effect relationship is demonstrated by the incline of the linear graph. The x-axes represent genetic instrument- metabolites associations, and the y-axes represent genetic instrument-COPD associations . Black dots denote the genetic instruments included in the MR analysis.

specific mechanism of their protective effect on COPD needs further in-depth study. Lactate is an important product of glycolysis and plays a key role in energy metabolism, inflammation regulation and signaling.Lactic acid regulates the function of immune cells and influences the inflammatory response.For example, lactic acid can inhibit macrophage and T-cell activity, thereby reducing inflammation. An experimental study found that

Fig. 4. Scatterplot of MR analysis of metabolites positively associated with COPD risk. The extent of the cause-and-effect relationship is demonstrated by the incline of the linear graph. The x-axes represent genetic instrument- metabolites associations, and the y-axes represent genetic instrument-COPD associations . Black dots denote the genetic instruments included in the MR analysis.

lactic acid produced by alveolar type II cells converted alveolar macrophages to an anti-inflammatory phenotype and inhibited excessive inflammation in mice with lung injury[46]. 5-Oxoproline is an important intermediate in glutathione metabolism and plays a key role in oxidative stress and inflammation regulation. Some animal studies have found that glutathione supplementation improves oxidative stress and reduces lung damage[47].As

an amino acid, proline plays a crucial part in controlling the synthesis and metabolism of arginine, as well as the immune response of the body. The Nitric Oxide Synthase (NOS) family produces Nitric Oxide (NO), using L-Arginine as a substrate. NO production in the lungs and airways can play a variety of roles in lung development, regulating airway and vascular smooth muscle tone and participating in inflammatory processes and host defense. It has been shown that the exacerbation of COPD can be inhibited by inhibiting arginase or supplementing arginine to increase the substrate availability of NOS[48]. In addition to this, arginine is involved in wound healing, control of asthma and certain cardiovascular related diseases[49]. Amino acids are present in large quantities in certain fish and marine oils. An examination of data from the National Health and Nutrition Examination Survey (NHANES) revealed that adults who consumed a lot of fish had a higher 1-second force expiratory volume (FEV1) compared to those with low fish consumption[50]. A metabolomics study identified the differential expression of 23 serum metabolomic biomarkers in healthy smokers and COPD smokers, revealing metabolite biomarkers of early COPD, which are primarily associated with inflammatory responses and caffeine metabolism[51].Paraxanthine is a metabolite of caffeine and belongs to the methylated xanthine class of compounds. It is produced through the metabolism of caffeine in the body and has antioxidant and anti-inflammatory properties[52]. PLA is an intermediate product of phenylalanine metabolism and plays an important role in amino acid metabolism and inflammation regulation. In an infected mouse model, subinhibitory concentrations of PLA were effective in clearing pathogenic bacteria from the lungs of mice, suggesting that PLA may modulate the inflammatory response by influencing the function of immune cells[53].

However, our findings are not entirely consistent with those of previous studies, which may be due to the small sample size of the previous study, which may affect the broad applicability of the findings.The clinical presentation of COPD patients varies widely, and the different subtypes may affect the changes in metabolic markers; therefore, it may be possible that there are certain metabolic markers that are more specific to a particular patient group. In a metabolomic analysis of current and former smokers in a white cohort, lower serum amino acid concentrations were found to be associated with a higher incidence of respiratory deterioration. In particular, there was a consistent and substantial association between reduced tryptophan concentrations and worse lung function, extent of emphysema on chest CT, and work capacity[54]. In addition, a study based on an Asian population found that elevated serum uric acid levels measured at two different times were consistently and significantly associated with an accelerated decline in FEV1 and a decrease in the percent predicted FEV1 value in nonsmoking individuals[55].

This MR analysis has several significant advantages. Our study utilizes MR to determine causality and relies on publicly available data, thereby eliminating the need for individual-level datasets and effectively circumventing the limitations of traditional observational studies, such as insufficient sample size and potential confounders. In addition, we conducted multiple sensitivity analyses to ensure that our study was not confounded by horizontal pleiotropy and heterogeneity, confirmed that there was no evidence of violation of the assumption of instrumental variable independence in the analyses, and ruled out potential outliers through the MR-PRESSO methodology, which further solidified the robustness and consistency of our findings.

Nevertheless, there are constraints to our study. While the F-statistics of chosen SNPs showed satisfactory resilience, the limitation arises from the restricted number of SNPs available for a thorough analysis of genome-wide exposure. Expanding the sample size in future studies could improve the reliability of the results and the accuracy of the causal effect estimates. In addition, the results of the study are not based on the entire population, which may limit their generalizability to different ethnic groups.Therefore, the generalizability of our findings to other populations warrants further exploration and validation. We suggest that future multi-ethnic MR analyses be conducted to further validate our findings.Second, given the limitations of the database information, we were unable to assess the potential impact of disease-specific classifications on study outcomes, such as severity or acute exacerbations of COPD.In addition, the impact of clinical interventions could not be determined by MR analysis. While MR analysis provides valuable insights into etiology, it is important to note that our findings should be validated by targeted metabolomics in a COPD cohort before applying them to the clinic.

Conclusion

This MR study used genetic proxies to determine a causal relationship between blood metabolites and COPD. Furthermore, the research identified five distinct blood metabolites that could potentially be linked to the progression of COPD. The discoveries offer valuable perspectives on possible approaches for early detection, avoidance, and management of COPD. By combining genomics and metabolomics, this analysis provides a reference point for studying the causes and development of COPD.

Data availability

All data in this study are available in publicly available databases. No additional ethical approval was required. If you need all the data used in this study, please contact the corresponding author.

References

1. Celli, B. et al. Definition and nomenclature of chronic obstructive pulmonary disease: time for its revision. *Am. J. Respir Crit. Care Med.* **206** (11), 1317–1325. https://doi.org/10.1164/rccm.202204-0671PP (2022).
2. Raherison, C. & Girodet, P. O. Epidemiology of COPD. *Eur. Respir Rev.* **18** (114), 213–221. https://doi.org/10.1183/09059180.00003609 (2009).

3. HalpinDMG et al. Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease. *Am. J. Respir Crit. Care Med.* **203** (1), 24–36. https://doi.org/10.1164/rccm.202009-3533SO (2021).

4. Dahl, M., Tybjaerg-Hansen, A., Vestbo, J., Lange, P. & Nordestgaard, B. G. Elevated plasma fibrinogen associated with reduced pulmonary function and increased risk of chronic obstructive pulmonary disease. *Am. J. Respir Crit. Care Med.* **164** (6), 1008–1011. https://doi.org/10.1164/ajrccm.164.6.2010067 (2001).

5. Pavord, I. D. et al. Mepolizumab for eosinophilic chronic obstructive pulmonary disease. *N Engl. J. Med.* **377** (17), 1613–1629. https://doi.org/10.1056/NEJMoa1708208 (2017).

6. Hastie, A. T. et al. Association of sputum and blood eosinophil concentrations with clinical measures of COPD severity: an analysis of the SPIROMICS cohort. *Lancet Respir Med.* **5** (12), 956–967. https://doi.org/10.1016/S2213-2600(17)30432-0 (2017).

7. Bafadhel, M., Pavord, I. D. & Russell, R. E. K. Eosinophils in COPD: just another biomarker? [published correction appears in Lancet Respir Med. ;5(8):e28. doi: 10.1016/S2213-2600(17)30260-6.]. Lancet Respir Med. 2017;5(9):747–759. (2017). https://doi.org/10.1016/S2213-2600(17)30217-5

8. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell. Biol.* **17** (7), 451–459. https://doi.org/10.1038/nrm.2016.25 (2016).

9. Laíns, I. et al. Metabolomics in the study of retinal health and disease. *Prog Retin Eye Res.* **69**, 57–79. https://doi.org/10.1016/j.preteyeres.2018.11.002 (2019).

10. Labaki, W. W. et al. Serum amino acid concentrations and clinical outcomes in smokers: SPIROMICS metabolomics study. *Sci. Rep.* **9** (1), 11367. https://doi.org/10.1038/s41598-019-47761-w (2019).

11. Yu, B. et al. Metabolomics Identifies Novel Blood Biomarkers of Pulmonary Function and COPD in the General Population. Metabolites. ;9(4):61. (2019). https://doi.org/10.3390/metabo9040061

12. Regan, E. A. et al. Omics and the search for blood biomarkers in chronic obstructive pulmonary disease. Insights from COPDGene. *Am. J. Respir Cell. Mol. Biol.* **61** (2), 143–149. https://doi.org/10.1165/rcmb.2018-0245PS (2019).

13. Yonchuk, J. G. et al. Circulating soluble receptor for advanced glycation end products (sRAGE) as a biomarker of emphysema and the RAGE axis in the lung. *Am. J. Respir Crit. Care Med.* **192** (7), 785–792. https://doi.org/10.1164/rccm.201501-0137PP (2015).

14. Gea, J., Enríquez-Rodríguez, C. J., Agranovich, B. & Pascual-Guardia, S. Update on metabolomic findings in COPD patients. *ERJ Open. Res.* **9** (5), 00180–2023. https://doi.org/10.1183/23120541.00180-2023 (2023).

15. Oskeritzian, C. A., Milstien, S. & Spiegel, S. Sphingosine-1-phosphate in allergic responses, asthma and anaphylaxis. *Pharmacol. Ther.* **115** (3), 390–399. https://doi.org/10.1016/j.pharmthera.2007.05.011 (2007).

16. Koike, K. et al. Bioactive sphingolipids in the pathogenesis of chronic obstructive pulmonary disease. *Ann. Am. Thorac. Soc.* **15** (Suppl 4), S249–S252. https://doi.org/10.1513/AnnalsATS.201809-592MG (2018).

17. Ahmed, F. S. et al. Plasma sphingomyelin and longitudinal change in percent emphysema on CT. The MESA lung study. *Biomarkers* **19** (3), 207–213. https://doi.org/10.3109/1354750X.2014.896414 (2014).

18. Hojjati, M. R. & Jiang, X. C. Rapid, specific, and sensitive measurements of plasma sphingomyelin and phosphatidylcholine. *J. Lipid Res.* **47** (3), 673–676. https://doi.org/10.1194/jlr.D500040-JLR200 (2006).

19. Pouw, E. M., Schols, A. M., Deutz, N. E. & Wouters, E. F. Plasma and muscle amino acid levels in relation to resting energy expenditure and inflammation in stable chronic obstructive pulmonary disease. *Am. J. Respir Crit. Care Med.* **158** (3), 797–801. https://doi.org/10.1164/ajrccm.158.3.9708097 (1998).

20. Hofford, J. M. et al. The nutritional status in advanced emphysema associated with chronic bronchitis. A study of amino acid and catecholamine levels. *Am. Rev. Respir Dis.* **141** (4 Pt 1), 902–908. https://doi.org/10.1164/ajrccm/141.4_Pt_1.902 (1990).

21. Ubhi, B. K. et al. Targeted metabolomics identifies perturbations in amino acid metabolism that sub-classify patients with COPD. *Mol. Biosyst.* **8** (12), 3125–3133. https://doi.org/10.1039/c2mb25194a (2012).

22. Sekula, P., Del Greco, M. F., Pattaro, C. & Köttgen, A. Mendelian randomization as an approach to assess causality using observational data. *J. Am. Soc. Nephrol.* **27** (11), 3253–3265. https://doi.org/10.1681/ASN.2016010098 (2016).

23. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40** (Database issue), D109–D114. https://doi.org/10.1093/nar/gkr988 (2012).

24. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA* **326** (16), 1614–1621. https://doi.org/10.1001/jama.2021.18236 (2021).

25. Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44** (4), 313–329. https://doi.org/10.1002/gepi.22295 (2020).

26. Yang, X. et al. Causal relationship between gut microbiota and serum vitamin D: evidence from genetic correlation and Mendelian randomization study. *Eur. J. Clin. Nutr.* **76** (7), 1017–1023. https://doi.org/10.1038/s41430-021-01065-3 (2022).

27. Roze, D. Causes and consequences of linkage disequilibrium among transposable elements within eukaryotic genomes. *Genetics* **224** (2), iyad058. https://doi.org/10.1093/genetics/iyad058 (2023).

28. Kamat, M. A. et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35** (22), 4851–4853. https://doi.org/10.1093/bioinformatics/btz469 (2019).

29. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37** (7), 658–665. https://doi.org/10.1002/gepi.21758 (2013).

30. Zhi-Gang, Y. & Han-Dong, W. A causal link between Circulating leukocytes and three major urologic cancers: a Mendelian randomization investigation. *Front. Genet.* **15**, 1424119. https://doi.org/10.3389/fgene.2024.1424119 (2024).

31. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40** (4), 304–314. https://doi.org/10.1002/gepi.21965 (2016).

32. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect Estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44** (2), 512–525. https://doi.org/10.1093/ije/dyv080 (2015).

33. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of Pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27** (R2), R195–R208. https://doi.org/10.1093/hmg/ddy163 (2018).

34. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40** (3), 755–764. https://doi.org/10.1093/ije/dyr036 (2011).

35. Palmer, T. M. et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21** (3), 223–242. https://doi.org/10.1177/0962280210394459 (2012).

36. Psychogios, N. et al. The human serum metabolome. *PLoS One.* **6** (2), e16957. https://doi.org/10.1371/journal.pone.0016957 (2011). Published 2011 Feb 16.

37. Ubhi, B. K. et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur. Respir J.* **40** (2), 345–355. https://doi.org/10.1183/09031936.00112411 (2012).

38. Bowler, R. P. et al. Plasma sphingolipids associated with chronic obstructive pulmonary disease phenotypes. *Am. J. Respir Crit. Care Med.* **191** (3), 275–284. https://doi.org/10.1164/rccm.201410-1771OC (2015).

39. Sauerwein, H. P. & Schols, A. M. Glucose metabolism in chronic lung disease. *Clin. Nutr.* **21** (5), 367–371. https://doi.org/10.1054/clnu.2002.0561 (2002).

40. Suehiro, C. L. et al. A possible association between fructose consumption and pulmonary emphysema [published correction appears in Sci Rep. ;10(1):16158. (2020). https://doi.org/10.1038/s41598-020-69892-1.]. Sci Rep. 2019;9(1):9344. doi:10.1038/s41598-019-45594-1.

41. Liu, Y. et al. Neopterin in patients with COPD, asthma, and ACO: association with endothelial and lung functions. *Respir Res.* **25** (1), 171. https://doi.org/10.1186/s12931-024-02784-4 (2024).
42. Kotlyarov, S. & Kotlyarova, A. Molecular mechanisms of lipid metabolism disorders in infectious exacerbations of chronic obstructive pulmonary disease. *Int. J. Mol. Sci.* **22** (14), 7634. https://doi.org/10.3390/ijms22147634 (2021). Published 2021 Jul 17.
43. Tashiro, H. et al. Saturated fatty acid increases lung macrophages and augments house dust Mite-Induced airway inflammation in mice fed with High-Fat diet. *Inflammation* **40** (3), 1072–1086 (2017).
44. Boukhenouna, S., Wilson, M. A., Bahmed, K. & Kosmider, B. Reactive oxygen species in chronic obstructive pulmonary disease. *Oxid. Med. Cell. Longev.* **2018**, 5730395. https://doi.org/10.1155/2018/5730395 (2018).
45. Gong, J. et al. Cigarette smoke reduces fatty acid catabolism, leading to apoptosis in lung endothelial cells: implication for pathogenesis of COPD. *Front. Pharmacol.* **10**, 941. https://doi.org/10.3389/fphar.2019.00941 (2019).
46. Roy, R. M. et al. Lactate produced by alveolar type II cells suppresses inflammatory alveolar macrophages in acute lung injury. *FASEB J.* **37** (12), e23316 (2023).
47. Lavoie, J. C., Mohamed, I. & Teixeira, V. Dose-Response effects of glutathione supplement in parenteral nutrition on pulmonary oxidative stress and alveolarization in newborn Guinea pig. *Antioxid. (Basel).* **11** (10), 1956 (2022).
48. Scott, J. A., Maarsingh, H., Holguin, F. & Grasemann, H. Arginine therapy for lung diseases. *Front. Pharmacol.* **12**, 627503. https://doi.org/10.3389/fphar.2021.627503 (2021).
49. Ling, Z. N. et al. Amino acid metabolism in health and disease. *Signal. Transduct. Target. Ther.* **8** (1), 345. https://doi.org/10.1038/s41392-023-01569-3 (2023).
50. Schwartz, J. Role of polyunsaturated fatty acids in lung disease. *Am. J. Clin. Nutr.* **71** (1 Suppl). https://doi.org/10.1093/ajcn/71.1.393s (2000). 393S-6S.
51. Chen, Q. et al. Serum metabolite biomarkers discriminate healthy smokers from COPD smokers. *PLoS One.* **10** (12), e0143937. https://doi.org/10.1371/journal.pone.0143937 (2015).
52. Purpura, M., Jäger, R. & Falk, M. An assessment of mutagenicity, genotoxicity, acute-, subacute and subchronic oral toxicity of Paraxanthine (1,7-dimethylxanthine). *Food Chem. Toxicol.* **158**, 112579. https://doi.org/10.1016/j.fct.2021.112579 (2021).
53. Chifiriuc, M. et al. Drastic Attenuation of Pseudomonas aeruginosa pathogenicity in a holoxenic mouse experimental model induced by subinhibitory concentrations of phenyllactic acid (PLA). *Int. J. Mol. Sci.* **8** (7), 583–592 (2007).
54. Labaki, W. W. et al. Serum amino acid concentrations and clinical outcomes in smokers: SPIROMICS metabolomics study. *Sci. Rep.* **9** (1), 11367 (2019).
55. Ahn, K. M., Lee, S. Y., Lee, S. H., Kim, S. S. & Park, H. W. Lung function decline is associated with serum uric acid in Korean health screening individuals. *Sci. Rep.* **11** (1), 10183 (2021).

Author contributions

YZ and XM designed the study and drafted the manuscript. Data collection and analysis were conducted by FW, YS, HM and XL. WZ provided assistance with methodology guidance and revising the manuscript. The final manuscript was read and approved by all authors.

Declarations

Ethics approval and consent to participate

The data used in this study is publicly available and has been anonymized to ensure that it is not personally identifiable.All participants had obtained local ethical approval and provided informed consent, so no additional ethical approval was required for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Founding

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92216-0.

Correspondence and requests for materials should be addressed to W.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.