

Original Research

Genetic Evidence for Causal Relationships Between Circulating Cathepsin Levels and Chronic Obstructive Pulmonary Disease: A Mendelian Randomization Study

Chao Duan, BS¹ Ao Zhang, MS² Suyan Tian, PhD³

Abstract

Background: Cathepsins, a family of lysosomal proteolytic enzymes, have been implicated in the pathogenesis of various complex diseases, including chronic obstructive pulmonary disease (COPD). However, the causal relationship between cathepsins and COPD remains unclear.

Methods: This study employed Mendelian randomization (MR) to investigate the potential causal effects of cathepsin levels on COPD risk. This MR analysis utilized genetic data from individuals of European ancestry in the INTERVAL study and FinnGen consortium. Specifically, summary-level genetic data for 9 cathepsins (B, E, F, G, H, O, L2, S, and Z) were obtained from the INTERVAL study, while COPD summary statistics were sourced from the FinnGen consortium. We conducted comprehensive MR analyses, including univariable MR, reverse MR, multivariable MR (MVMR), and MR least absolute shrinkage and selection operator, to assess causal relationships between cathepsin levels and COPD risk.

Results: Univariable MR analysis revealed no significant causal relationships (forward or reverse) between the 9 cathepsins and COPD risk. However, MVMR analysis identified cathepsins O and S as having direct causal effects on COPD. For cathepsins O and S, odds ratio was estimated as 1.130 ($p=0.022$, 95% confidence interval [CI]=1.018–1.255) and 1.068 ($p=0.025$, 95% CI=1.008–1.132), respectively. Furthermore, these 2 cathepsins were independent risk factors for COPD after adjusting for smoking.

Conclusions: To our knowledge, this is the first MR study to systematically evaluate the causal role of cathepsins in COPD. Further research, particularly clinical trials, is warranted to validate these associations and explore the therapeutic potential of targeting cathepsins in COPD management.

1. Intensive Care Unit, The First Hospital of Jilin University, Changchun, Jilin, China
2. Department of Neurology, The First Hospital of Jilin University, Changchun, Jilin, China
3. Division of Clinical Research, The First Hospital of Jilin University, Changchun, Jilin, China

Abbreviations:

CI=confidence interval; **COPD**=chronic obstructive pulmonary disease; **ECM**=extracellular matrix; **GWAS**=genome-wide association study; **IV**=instrumental variable; **IVW**=inverse-variance weighted; **LASSO**=least absolute shrinkage and selection operator; **MR**=Mendelian randomization; **MVMR**=multivariable MR; **PRESSO**=pleiotropy residual sum and outlier; **PTMs**=post-translational modifications; **RCT**=randomized controlled trial; **SNP**=single nucleotide polymorphism

Funding Support:

This study was supported by the Health Department of Jilin Province (2023LC008).

Citation:

Duan C, Zhang A, Tian S. Genetic evidence for causal relationships between circulating cathepsin levels and chronic obstructive pulmonary disease: a Mendelian randomization study. *Chronic Obstr Pulm Dis*. 2025;12(5):380-389. doi: <https://doi.org/10.15326/jcopdf.2025.0626>

Publication Dates:

Date of Acceptance: July 26, 2025

Published Online Date: August 6, 2025

Address correspondence to:

Suyan Tian, PhD
Division of Clinical Research
The First Hospital of Jilin University
Changchun, Jilin, China
Email: wmxt@jlu.edu.cn

Keywords:

causal inference; cathepsin; chronic obstructive pulmonary disease; Mendelian randomization

This article has an online supplement.

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive inflammatory lung disorder characterized by persistent airflow limitation.^{1,2} The disease clinically manifests through respiratory symptoms including dyspnea, chronic cough, excessive sputum production, and wheezing.² Pathologically, COPD encompasses 2 principal components: emphysema, involving alveolar sac destruction leading to impaired gas exchange, and chronic bronchitis, featuring persistent bronchial inflammation with mucus hypersecretion.³ The alveolar tissue degeneration in emphysema reduces lung elasticity, while chronic bronchitis causes airway narrowing through inflammatory thickening and mucus plugging. Despite therapeutic advances, COPD remains incurable, with treatment focusing primarily on symptom management and disease progression delay.¹ These clinical realities underscore the critical importance of early detection and sustained intervention, which can substantially improve patient outcomes and quality of life.

Cathepsins are a family of proteolytic enzymes predominantly found in all animals in addition to other organisms. There are several types of cathepsins, each categorized mainly by their enzymatic nature and substrate specificity, including cathepsins B, D, L, S, K, G, H, V, and C among others.⁴ These enzymes are known for their roles in lysosomal degradation, where they contribute to the breakdown of proteins inside the lysosomes, critical cellular organelles responsible for digesting various biomolecules. Their broad range of functions in cellular maintenance and regulation renders specific cathepsins crucial for cellular functioning and highlights their potential roles in the management and targeted treatments of various complex diseases.⁵⁻⁷

Growing evidence implicates cathepsins in COPD development through multiple pathological mechanisms.⁸ Cathepsins S, L, and K contribute to disease progression by degrading extracellular matrix (ECM) components, including elastin and collagen, in lung tissue. This proteolytic activity drives the structural remodeling of airways and parenchyma characteristic of COPD.⁹⁻¹³ Beyond ECM degradation, cathepsins participate in COPD-associated inflammation by activating cytokines and chemokines that sustain pulmonary inflammatory cascades. Their pathogenic role is further supported by elevated levels detected in COPD patients' sputum and bronchoalveolar lavage fluid.^{13,14} Cathepsins also regulate immune responses frequently dysregulated in COPD. For example, cathepsin G modulates neutrophil function¹⁰—a critical defense mechanism in lungs. Abnormal cathepsin G activity may impair neutrophil responses, exacerbating inflammation and tissue damage.⁸ Furthermore, increased cathepsin E protein in the lung epithelium of COPD patients has been observed.¹⁵

While these observations derive primarily from in vitro and observational studies (which are susceptible to confounding and reverse causality), they collectively suggest cathepsins as potential therapeutic targets. However, the causal relationship between cathepsin levels and COPD risk requires validation through robust methods.

Mendelian randomization (MR) is a research method used in epidemiology to assess the causal relationship between potentially modifiable risk factors and health outcomes.¹⁵ This technique leverages genetic variation as a surrogate to examine the causal effect of a modifiable exposure on disease in observational data. In other words, it uses genetic variants as instrumental variables (IVs) to estimate the causal effect of an exposure on the outcome.^{16,17} MR relies on the natural random assortment of genes at conception, which obeys Mendel's laws of inheritance. This allocation mimics the randomization process in a randomized controlled trial (RCT), minimizing confounding factors that typically affect observational studies. While MR studies cannot replace RCTs, they offer insights that help bridge the gap between correlation and causation. In this study, we aimed to elucidate the causal relationship between cathepsin levels and COPD risk using MR analyses, and, therefore, to provide valuable insights on the prevention and early intervention for COPD.

Methods and Materials

Experimental Data

Genetic association statistics for 9 cathepsin (i.e., cathepsin B, E, F, G, H, O, L2, S, and Z) levels were derived from the INTERVAL study, comprising 3301 participants of European ancestry.¹⁸ COPD summary statistics were obtained from the FinnGen consortium,¹⁹ including 6915 COPD cases and 186,723 controls. Lastly, we used a summary genome-wide association study (GWAS) data for smoking from the Medical Research Council-Integrative Epidemiology Unit consortium,²⁰ which included 280,508 cases and 180,558 controls. Ethical approval for this study was waived by the First Hospital of Jilin University Institutional Review Board as no original research data were collected.

IVs for cathepsin levels were selected based on the following criteria: a linkage disequilibrium threshold of $R^2 < 0.001$ within a 10,000kb clumping window, and a genome-wide significance level of 5×10^{-6} .

Mendelian Randomization

For a single nucleotide polymorphism (SNP) to serve as a valid IV in MR analyses, 3 fundamental assumptions must be satisfied: (1) relevance assumption dictates that the SNP must be strongly correlated with the exposure, (2) independence assumption requires that the SNP must be independent of any confounders that affect the

relationship between the exposure and the outcome, and (3) exclusion restriction assumption requires that the SNP should not have a direct association with the outcome, nor should it be related to the outcome through any pathways other than the exposure.

The inverse-variance weighted (IVW) method served as our primary analytical approach for estimating causal effects, offering optimal statistical power when all genetic variants meet IV assumptions.²¹ However, recognizing that violations of these assumptions—particularly through horizontal pleiotropy—could bias IVW estimates, we implemented supplementary pleiotropy-robust methods to validate our findings. These included MR-Egger regression, which accounts for balanced pleiotropy²²; the weighted median approach, providing consistent estimates when even up to 50% of weights derive from invalid instruments²³; the weighted mode method, effective when the largest SNP cluster shares a common causal estimate²⁴; and MR-pleiotropy residual sum and outlier (PRESSO), which identifies and corrects for outlier variants.²⁵

Both the MR-PRESSO global test and MR-Egger intercept (MR-Egger intercept p -value <0.05) were used to identify outliers and detect horizontal pleiotropy. Additionally, the MR-PRESSO outlier test was conducted to mitigate or eliminate horizontal pleiotropy by removing outliers (p -value of the MR-PRESSO global test <0.05). Cochran's Q test was used to assess heterogeneity among SNPs, with a p -value of <0.05 indicating heterogeneity. If the p -value was less than 0.05, a random-effect model was used to estimate causal effect size. Otherwise, a fixed-effect model was applied instead. The R TwoSampleMR package²⁶ was utilized for conducting 2-sample MR analyses. The MR-PRESSO tests were carried out using the MR-PRESSO package.

Reverse MR analyses in which COPD was considered as the exposure and cathepsins as the outcome were performed to explore reverse causality, using the GWAS studies from the forward MR analyses. Next, multivariable IVW MR analysis involving 9 cathepsins as predictors was conducted using the R MendelianRandomization package. To address multicollinearity, least absolute shrinkage and selection operator (LASSO) analysis was performed to select relevant features and construct the final model, and the R MrLasso package²⁷ was employed for this analysis. Conditional F-statistics were calculated using the R multivariable MR (MVMR) package.

Lastly, posthoc statistical power assessment of the MR analysis was conducted using the mRnd online tool.^{28,29} The current study was designed following the Strengthening the Reporting of Observational Studies in Epidemiology-MR checklist.³⁰

Results

Univariable Mendelian Randomization

The causal relationship between 9 cathepsins (cathepsin B, E, F, G, H, O, L2, S, and Z) and the risk of COPD was investigated using a 2-sample univariable MR analysis. Initially, all p -values from Cochran's Q statistics exceeded 0.05. In conjunction with the results from leave-one-out plots (Supplementary Figure 1 in the online supplement), it was concluded that significant heterogeneity was absent. Furthermore, the MR-Egger intercept test revealed no horizontal pleiotropy, with a p -value greater than 0.05. All F-statistics of single SNPs were larger than 10.

The results of univariable MR analysis are tabulated in Table 1, while the forest plot in Figure 1 graphically elucidates IVW results. For example, our analysis showed that the risk of developing COPD for a one-unit increment in abundance level of cathepsin O was estimated to be 1.078 (odds ratio [OR]=1.078, $p=0.081$, 95% confidence interval [CI]=0.991–1.174) by the IVW method. All 4 complementary MR methods supported in concordance with this null relationship. The OR was estimated as 1.076 ($p=0.478$, 95% CI=0.885–1.309) by MR-Egger, 1.033 ($p=0.596$, 95% CI=0.917–1.163) by weighted median, 1.011 ($p=0.900$, 95% CI=0.861–1.187) by weighted mode, and 1.044 ($p=0.401$, 95% CI=0.948–1.149) by MR-PRESSO. Regarding cathepsin S, which has been reported to associate with the development and progression of COPD by numerous prior studies, our analysis found a nonsignificant positive causal effect of this protein on COPD risk. Specifically, OR was estimated as 1.037 ($p=0.464$, 95% CI=0.940–1.145) by IVW, 1.064 ($p=0.556$, 95% CI=0.869–1.302) by MR-Egger, 1.132 ($p=0.053$, 95% CI=0.999–1.282) by weighted median, 1.140 ($p=0.150$, 95% CI=0.959–1.356) by weighted mode, and 1.037 ($p=0.471$, 95% CI=0.989–1.145) by MR-PRESSO, respectively. Overall, none of these 9 cathepsins were found to cause an increase or decrease in developing COPD. Furthermore, 4 MR methods comparison plots are shown in Supplementary Figure 2 in the online supplement, demonstrating robust MR results.

Subsequently, reverse MR analysis was performed. Null reverse causal relationship was identified between 9 cathepsin types and the risk of developing COPD, which were supported consistently by all MR methods based on corresponding adjusted p -values. For these MR analyses, neither heterogeneity (Cochran's Q p -value >0.05) nor horizontal pleiotropy (MR-Egger intercept p -value >0.05) was detected. The reverse MR results are shown in Supplementary Table 1 in the online supplement.

Multivariable Mendelian Randomization

MVMR analyses (including IVW, MR-Egger, and MR-PRESSO methods) were performed to analyze the genetic predisposition for 9 cathepsin types in relation to the risk of having COPD, which identified both cathepsins O and S as a risk factor for COPD (Table 2). Specifically, OR was

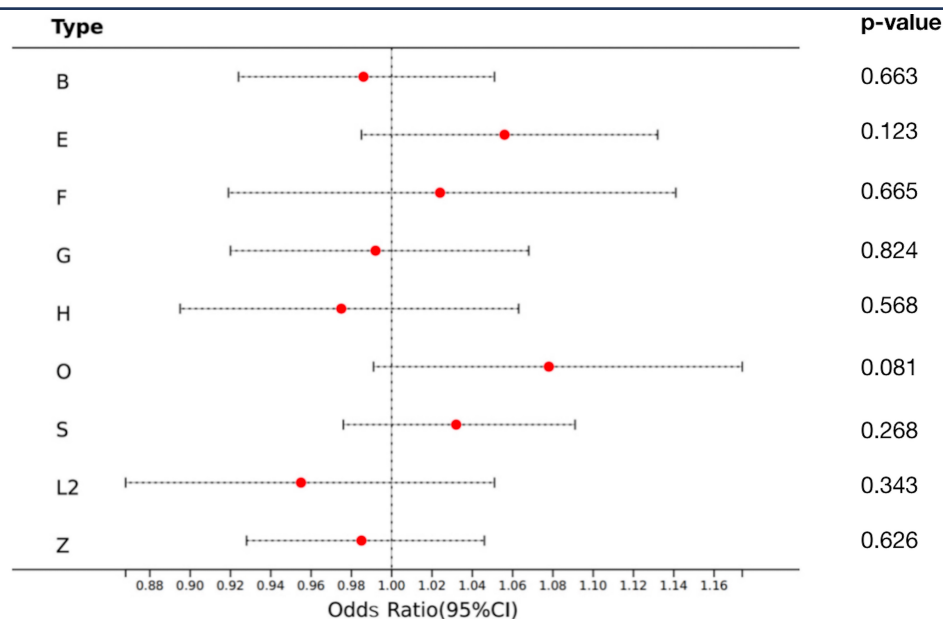
Table 1. Results of Univariable Mendelian Randomization Analysis Between Cathepsins and Chronic Obstructive Pulmonary Disease

Type	MR Method	Odds Ratio	p-Value	MR-Egger Intercept (p)	Heterogeneity Test (p)	MR-PRESSO Global (p)
Cathepsin B	IVW	0.986	0.663	0.707	0.844	0.841
	MR-Egger	0.960	0.602		0.802	
	Weighted Median	1.025	0.597			
	Weighted Mode	1.033	0.546			
	MR PRESSO	0.986	0.596			
Cathepsin E	IVW	1.056	0.123	0.420	0.650	0.651
	MR-Egger	1.014	0.817		0.630	
	Weighted median	1.054	0.256			
	Weighted mode	1.105	0.187			
	MR PRESSO	1.056	0.111			
Cathepsin F	IVW	1.024	0.665	0.442	0.569	0.572
	MR-Egger	1.136	0.388		0.536	
	Weighted Median	1.031	0.677			
	Weighted Mode	1.057	0.649			
	MR PRESSO	1.024	0.648			
Cathepsin G	IVW	0.992	0.824	0.580	0.627	0.651
	MR-Egger	0.956	0.558		0.572	
	Weighted Median	0.964	0.481			
	Weighted Mode	0.964	0.525			
	MR PRESSO	0.992	0.810			
Cathepsin H	IVW	0.975	0.568	0.723	0.064	0.085
	MR-Egger	0.946	0.572		0.048	
	Weighted Median	0.978	0.652			
	Weighted Mode	0.949	0.343			
	MR PRESSO	0.975	0.576			
Cathepsin O	IVW	1.078	0.081	0.982	0.683	0.179
	MR-Egger	1.076	0.478		0.596	
	Weighted Median	1.033	0.596			
	Weighted Mode	1.011	0.900			
	MR PRESSO	1.044	0.401			
Cathepsin S	IVW	1.032	0.268	0.742	0.312	0.265
	MR-Egger	1.046	0.377		0.268	
	Weighted Median	1.097	0.014			
	Weighted Mode	1.079	0.077			
	MR PRESSO	1.032	0.279			
Cathepsin L2	IVW	0.955	0.343	0.477	0.842	0.874
	MR-Egger	0.871	0.326		0.824	
	Weighted Median	0.933	0.246			
	Weighted Mode	0.936	0.421			
	MR PRESSO	0.955	0.236			
Cathepsin Z	IVW	0.985	0.626	0.565	0.850	0.832
	MR-Egger	1.007	0.883		0.818	
	Weighted Median	0.979	0.600			
	Weighted Mode	0.998	0.968			
	MR PRESSO	0.985	0.539			

MR=Mendelian randomization; PRESSO=pleiotropy residual sum and outlier; IVW=inverse-variance weighted

estimated as 1.130 ($p=0.022$, 95% CI=1.018–1.255) for cathepsin O and 1.068 ($p=0.025$, 95% CI=1.008–1.132) for cathepsin S by IVW, respectively. The forest plot in Figure 2 graphically elucidates MVMR IVW analysis results. To address potential multicollinearity between cathepsin types

and weak instrument bias (some cathepsin types have a small conditional F-statistic), we performed MR LASSO analysis to select more highly related subtypes, which identified cathepsins B, O, S, and Z out of 9 cathepsins. Based on these 4 cathepsins, we redid MVMR analyses (Table 3) and

Figure 1. Forest Plot of Univariable Mendelian Randomization Analysis Between the Abundance of 9 Cathepsins^a and the Risk of Having Chronic Obstructive Pulmonary Disease^a Cathepsins B, E, F, G, H, L2, O, S, and Z

CI=confidence interval

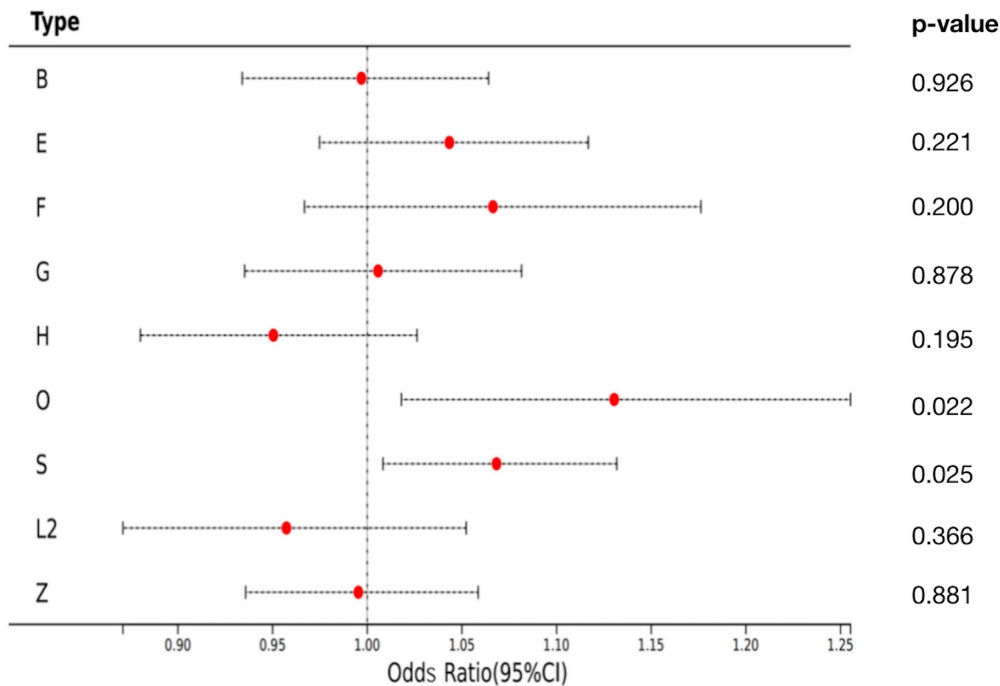
Table 2. Results of Multivariable Mendelian Randomization Analysis Between Cathepsins and Chronic Obstructive Pulmonary Disease

Type	MR Method	Odds Ratio	p-Value	MR-Egger Intercept (p)	Heterogeneity Test (p)	PRESSO Global (p)
Cathepsin B	IVW	0.997	0.926	0.436	0.495	0.705
	MR-Egger	0.984	0.666			
	MR PRESSO	0.997	0.929			
Cathepsin E	IVW	1.043	0.221			
	MR-Egger	1.034	0.354			
	MR PRESSO	1.035	0.338			
Cathepsin F	IVW	1.066	0.200			
	MR-Egger	1.068	0.193			
	MR PRESSO	1.070	0.180			
Cathepsin G	IVW	1.006	0.878			
	MR-Egger	1.011	0.777			
	MR PRESSO	1.009	0.810			
Cathepsin H	IVW	0.950	0.195			
	MR-Egger	0.940	0.133			
	MR PRESSO	0.949	0.186			
Cathepsin O	IVW	1.131	0.022			
	MR-Egger	1.148	0.012			
	MR PRESSO	1.139	0.018			
Cathepsin S	IVW	1.068	0.025			
	MR-Egger	1.069	0.023			
	MR PRESSO	1.068	0.027			
Cathepsin L2	IVW	0.957	0.366			
	MR-Egger	0.946	0.258			
	MR PRESSO	0.953	0.318			
Cathepsin Z	IVW	0.995	0.881			
	MR-Egger	0.994	0.851			
	MR PRESSO	0.995	0.864			

MR=Mendelian randomization; PRESSO=pleiotropy residual sum and outlier; IVW=inverse-variance weighted

For personal use only. Permission required for all other uses.

Figure 2. Forest Plot of Multivariable Mendelian Randomization Analysis Between the Abundance of 9 Cathepsins^a and the Risk of Having Chronic Obstructive Pulmonary Disease



^a Cathepsins B, E, F, G, H, L2, O, S, and Z

CI=confidence interval

the calculated conditional F-statistics for these 4 cathepsins indicated marginal weak instrument bias. The results are in line with MVMR analysis result with all 9 cathepsin types as covariates.

Furthermore, given that COPD is a heavily smoking-mediated disease, we also performed a MVMR using smoking and cathepsins B, O, S, and Z as covariates. The analysis results indicated that after adjusting for smoking status, both cathepsins O and S were genetically related to COPD risk. Specifically, OR was estimated as 1.217 ($p=0.033$, 95% CI=1.096–1.458) for cathepsin O and 1.130 ($p=0.003$, 95% CI=1.043–1.224) for cathepsin S while OR for smoking was as 7.614 ($p<0.001$, 95% CI=3.819–15.180) by IVW, respectively (Table 4).

Posthoc Power Calculation

Lastly, a posthoc power calculation was performed to assess the statistical power of the current MR study. At a significance level of 0.05, a sample size of 193,638 individuals including 69,15 cases (3.57%) exhibited a power of 0.80 and 0.56 to detect a 1.2- (OR=1.2) and 1.15-fold risk (OR=1.15) of developing COPD per one-unit increase in the genetically predicted specific cathepsin level. Here, we assumed that these SNPs accounted for approximately 3% of the variance in the cathepsin levels. The sample size of the current MR study was still inadequate to detect a subtle effect.

Discussion

In this study, we employed a comprehensive MR approach incorporating univariable, reverse, and multivariable analyses to investigate potential causal relationships between cathepsin levels and COPD risk. Our MVMR results identified cathepsins O and S as potential risk factors for COPD. Existing evidence suggests cathepsin O contributes to COPD pathology through several interconnected pathways. As a protease capable of degrading ECM components, cathepsin O may promote the tissue remodeling and destruction characteristic of COPD. This is supported by observations of elevated cathepsin O levels in COPD patient lung tissues compared to healthy controls.³¹ The enzyme's ECM-degrading activity, particularly targeting structural proteins in lung tissue, likely contributes to emphysema development - a hallmark feature of COPD involving alveolar wall destruction that impairs lung elasticity and gas exchange. Beyond its structural effects, cathepsin O appears to influence COPD progression through inflammatory modulation. The enzyme regulates cytokine and chemokine activity, potentially exacerbating the chronic inflammation that drives COPD pathogenesis. Furthermore, emerging evidence positions cathepsin O as a responder to oxidative stress,³² a key driver of COPD development in individuals exposed to cigarette smoke and environmental pollutants. In this context, cathepsin O may participate in processing damaged proteins and organelles resulting from oxidative stress in COPD patients.

Table 3. Results of Multivariable Mendelian Randomization Analysis After Least Absolute Shrinkage and Selection Operator Feature Selection Between Cathepsins and Chronic Obstructive Pulmonary Disease

Type	MR Method	Odds Ratio	p-Value	MR-Egger Intercept (p)	Heterogeneity Test (p)	PRESSO Global (p)
Cathepsin B	IVW	1.005	0.890	0.646	0.609 0.572	0.726
	MR-Egger	0.993	0.880			
	MR PRESSO	1.005	0.886			
Cathepsin O	IVW	1.107	0.036			
	MR-Egger	1.109	0.033			
	MR PRESSO	1.107	0.035			
Cathepsin S	IVW	1.075	0.018			
	MR-Egger	1.073	0.020			
	MR PRESSO	1.075	0.019			
Cathepsin Z	IVW	0.987	0.695			
	MR-Egger	0.986	0.658			
	MR PRESSO	0.987	0.684			

MR=Mendelian randomization; PRESSO=pleiotropy residual sum and outlier; IVW=inverse-variance weighted

Table 4. Results of Multivariable Mendelian Randomization Analysis Between Smoking Status, Cathepsins B, O, S, Z, and Chronic Obstructive Pulmonary Disease

Type	MR Method	Odds Ratio	p-Value	MR-Egger Intercept (p)	Heterogeneity Test (p)	PRESSO Global (p)
Cathepsin B	IVW	1.047	0.391	0.944	0.221 0.199	0.590
	MR-Egger	1.049	0.444			
	MR PRESSO	1.047	0.394			
Cathepsin O	IVW	1.217	0.033			
	MR-Egger	1.217	0.035			
	MR PRESSO	1.217	0.036			
Cathepsin S	IVW	1.130	0.003			
	MR-Egger	1.129	0.003			
	MR PRESSO	1.130	0.004			
Cathepsin Z	IVW	1.004	0.920			
	MR-Egger	1.004	0.920			
	MR PRESSO	1.004	0.920			
Smoking (ever smoker)	IVW	7.612	<0.001			
	MR-Egger	7.594	<0.001			
	MR PRESSO	7.612	<0.001			

MR=Mendelian randomization; PRESSO=pleiotropy residual sum and outlier; IVW=inverse-variance weighted

Cathepsin S, another key member of the cathepsin protease family, has been consistently shown to be elevated in both lung tissues and serum of COPD patients compared to healthy individuals.^{7,33} As a potent protease, cathepsin S efficiently degrades critical ECM components including elastin and collagen - structural proteins essential for maintaining normal lung architecture.³⁴ The excessive proteolytic activity of overexpressed cathepsin S may drive the pathological tissue destruction characteristic of emphysema, a defining feature of COPD that involves alveolar wall breakdown and progressive loss of lung elasticity. Beyond its direct effects on parenchymal destruction, cathepsin S-mediated ECM degradation likely contributes to the pathological airway remodeling observed in COPD. By altering the composition and integrity of airway connective tissue,

cathepsin S may promote structural changes that lead to airway narrowing and increased stiffness.^{9-12,35} These alterations can significantly worsen airflow limitation and contribute to disease progression. Together with cathepsin O, cathepsin S appears to play a multifaceted role in COPD pathogenesis through interconnected mechanisms involving ECM degradation, inflammatory modulation, tissue remodeling, and oxidative stress responses.³⁶⁻³⁸ Their combined actions may create a self-perpetuating cycle of tissue damage and functional decline that characterizes COPD development and progression.

Several limitations should be considered when interpreting our findings. First, the relatively small sample size of the exposure GWAS required us to adopt

a more lenient genetic significance threshold. Although this enabled the inclusion of additional SNPs, it may have increased susceptibility to weak instrument bias and horizontal pleiotropy. While our comprehensive sensitivity analyses helped mitigate these concerns, residual pleiotropic effects could still influence the results, as is inherent in all MR studies. Second, our analysis was constrained by the availability of GWAS summary data, which included only 9 cathepsins and missed several major cathepsin types, including cathepsins A, C, D, K, L, and W. Notably, we were unable to evaluate several cathepsins implicated in COPD pathogenesis by previous research, for example, cathepsins D and C.^{14,31,39} This limitation not only restricts the comprehensiveness of our investigation but may also introduce biases in the MVMR analysis results, as these omitted cathepsins could potentially confound the observed relationships. Another key limitation is that our findings derive from European populations, potentially limiting their applicability to other ethnic groups. Future research must validate these associations in diverse cohorts to determine their broader relevance. Additionally, while MR LASSO was applied to address multicollinearity, the potential for residual collinearity remains. Lastly, while MR identifies genetic associations between cathepsins and COPD, it cannot assess tissue-specific post-translational modifications (PTMs) regulating cathepsin activation, smoke-induced epigenetic and post-translational regulation, extracellular versus intracellular cathepsin activity differences, and redox modifications altering protease function. Therefore, future studies should combine proteomics, redox biochemistry, and single-cell analyses to fully elucidate how PTMs and environmental factors (e.g., smoking) modulate cathepsin-driven COPD pathogenesis.

Conclusion

To our knowledge, this study represents the first MR study to systematically investigate the causal relationship between circulating cathepsin levels and COPD risk. Our findings suggest that elevated levels of cathepsins O and S may serve as independent risk factors for COPD development, though these observations require further validation.

These findings may provide a novel therapeutic direction for COPD management through targeted modulation of specific cathepsin pathways. However, further investigation - particularly through RCTs - will be crucial to confirm these causal associations and evaluate the clinical potential of cathepsin-focused interventions for COPD patients.

Acknowledgements

Author contributions: ST conceived and designed the experiment. ST and CD ran the analysis and verified the underlying data. CD and ST wrote the original manuscript. AZ, CD, and ST were involved in data interpretation. All authors have read and approved the final version of the manuscript.

Data sharing statement: The GWAS data of cathepsins, smoking, and COPD were downloaded from the MRC Integrative Epidemiology Unit Open GWAS Project (<https://gwas.mrcieu.ac.uk>). The corresponding GWAS ID numbers are prob-a-718, prob-a-720, prob-a-721, prob-a-723, prob-a-724, prob-a-726, prob-a-727, prob-a-728, prob-a-729, ukb-b-20261, and finn-b-J10_COPD.

Declaration of Interest

The authors have no conflicts of interest to declare.

References

1. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive pulmonary disease. *Lancet*. 2022;399(10342):2227-2242. [https://doi.org/10.1016/S0140-6736\(22\)00470-6](https://doi.org/10.1016/S0140-6736(22)00470-6)
2. MacLeod M, Papi A, Contoli M, et al. Chronic obstructive pulmonary disease exacerbation fundamentals: diagnosis, treatment, prevention and disease impact. *Respirology*. 2021;26(6):532-551. <https://doi.org/10.1111/resp.14041>
3. Janjua S, Carter D, Threapleton CJD, Prigmore S, Disler RT. Telehealth interventions: remote monitoring and consultations for people with chronic obstructive pulmonary disease (COPD). *Cochrane Database Syst Rev*. 2021;7:CD013196. <https://doi.org/10.1002/14651858.CD013196.pub2>
4. Nakanishi H. Cathepsin regulation on microglial function. *Biochim Biophys Acta Proteins Proteom*. 2020;1868(9):140465. <https://doi.org/10.1016/j.bbapap.2020.140465>
5. Soond SM, Kozhevnikova MV, Frolova AS, et al. Lost or forgotten: the nuclear cathepsin protein isoforms in cancer. *Cancer Lett*. 2019;462:43-50. <https://doi.org/10.1016/j.canlet.2019.07.020>
6. Gao S, Zhu H, Zuo X, Luo H. Cathepsin G and its role in inflammation and autoimmune diseases. *Arch Rheumatol*. 2018;33(4):498-504. <https://doi.org/10.5606/ArchRheumatol.2018.6595>
7. Brown R, Small DM, Doherty DF, et al. Therapeutic inhibition of cathepsin S reduces inflammation and mucus plugging in adult β ENaC-Tg mice. *Mediators Inflamm*. 2021;2021(1):6682657. <https://doi.org/10.1155/2021/6682657>
8. Pandey KC, De S, Mishra PK. Role of proteases in chronic obstructive pulmonary disease. *Front Pharmacol*. 2017;8:512. <https://doi.org/10.3389/fphar.2017.00512>
9. Andraut PM, Schamberger AC, Chazeirat T, et al. Cigarette smoke induces overexpression of active human cathepsin S in lungs from current smokers with or without COPD. *Am J Physiol Lung Cell Mol Physiol*. 2019;317(5):L625-L638. <https://doi.org/10.1152/ajplung.00061.2019>
10. Effah CY, Drokow EK, Agboyibor C, et al. Neutrophil-dependent immunity during pulmonary infections and inflammations. *Front Immunol*. 2021;12:689866. <https://doi.org/10.3389/fimmu.2021.689866>
11. Hahn I, Klaus A, Janze AK, et al. Cathepsin G and neutrophil elastase play critical and nonredundant roles in lung-protective immunity against *Streptococcus pneumoniae* in mice. *Infect Immun*. 2011;79(12):4893-4901. <https://doi.org/10.1128/IAI.05593-11>
12. Golovatch P, Mercer BA, Lemaitre V, Wallace A, Foronjy RF, D'Armiento J. Role for cathepsin K in emphysema in smoke-exposed guinea pigs. *Exp Lung Res*. 2009;35(8):631-645. <https://doi.org/10.3109/01902140902822304>
13. Xu X, Yu T, Dong L, et al. Eosinophils promote pulmonary matrix destruction and emphysema via cathepsin L. *Signal Transduct Target Ther*. 2023;8:390. <https://doi.org/10.1038/s41392-023-01634-x>
14. Bchir S, Boumiza S, ben Nasr H, et al. Impact of cathepsin D activity and C224T polymorphism (rs17571) on chronic obstructive pulmonary disease: correlations with oxidative and inflammatory markers. *Clin Exp Med*. 2021;21:457-465. <https://doi.org/10.1007/s10238-021-00692-1>
15. Zhang X, Shan P, Homer R, et al. Cathepsin E promotes pulmonary emphysema via mitochondrial fission. *Am J Pathol*. 2014;184(10):2730-2741. <https://doi.org/10.1016/j.ajpath.2014.06.017>
16. Bowden J, Holmes MV. Meta-analysis and Mendelian randomization: a review. *Res Synth Methods*. 2019;10(4):486-496. <https://doi.org/10.1002/jrsm.1346>
17. Larsson SC, Butterworth AS, Burgess S. Mendelian randomization for cardiovascular diseases: principles and applications. *Eur Heart J*. 2023;44(47):4913-4924. <https://doi.org/10.1093/eurheartj/ehad736>
18. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558:73-79. <https://doi.org/10.1038/s41586-018-0175-2>
19. FinnGen. FinnGen: an expedition into genomics and medicine. FinnGen website. Updated 2025. Accessed April 2025. <https://www.finnngen.fi/en>
20. Integrative Epidemiology Unit Open Genome-Wide Association Study Project (GWAS). GWAS summary data ecosystem. GWAS website. Updated 2025. Accessed April 2025. <https://gwas.mrcieu.ac.uk/about/#background>
21. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. <https://doi.org/10.1002/gepi.21758>
22. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. <https://doi.org/10.1093/ije/dyv080>
23. Bowden J, Smith GD, 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. <https://doi.org/10.1002/gepi.21965>
24. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985-1998. <https://doi.org/10.1093/ije/dyx102>
25. 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;50:693-698. <https://doi.org/10.1038/s41588-018-0099-7>

26. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. <https://doi.org/10.7554/eLife.34408>
27. Maity S. MrLasso. GitHub website. Published 2021. Accessed April 2025. <https://github.com/smaityumich/MrLasso>
28. CNS Genomics. mRnd: power calculations for Mendelian randomization. CNS Genomics website. Access April 2025. <https://cnsgenomics.com/shiny/mRnd>
29. Brion MJA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*. 2013;42(5):1497-1501. <https://doi.org/10.1093/ije/dyt179>
30. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233. <https://doi.org/10.1136/bmj.n2233>
31. Singanayagam A, Glanville N, Cuthbertson L, et al. Inhaled corticosteroid suppression of cathelicidin drives dysbiosis and bacterial infection in chronic obstructive pulmonary disease. *Sci Transl Med*. 2019;11(507):eaav3879. <https://doi.org/10.1126/scitranslmed.aav3879>
32. Kos J, Mitrović A, Perišić Nanut M, Pišlar A. Lysosomal peptidases-intriguing roles in cancer progression and neurodegeneration. *FEBS Open Bio*. 2022;12(4):708-738. <https://doi.org/10.1002/2211-5463.13372>
33. Small DM, Brown RR, Doherty DE, et al. Targeting of cathepsin S reduces cystic fibrosis-like lung disease. *Eur Respir J*. 2019;53(3):1801523. <https://doi.org/10.1183/13993003.01523-2018>
34. Brown R, Nath S, Lora A, et al. Cathepsin S: investigating an old player in lung disease pathogenesis, comorbidities, and potential therapeutics. *Respir Res*. 2020; 21:111. <https://doi.org/10.1186/s12931-020-01381-5>
35. Smyth P, Sasiwachirangkul J, Williams R, Scott CJ. Cathepsin S (CTSS) activity in health and disease - a treasure trove of untapped clinical potential. *Mol Aspects Med*. 2022;88:101106. <https://doi.org/10.1016/j.mam.2022.101106>
36. Rasmussen DGK, Sand JMB, Karsdal MA, Genovese F. Development of a novel enzyme-linked immunosorbent assay targeting a neo-epitope generated by cathepsin-mediated turnover of type III collagen and its application in chronic obstructive pulmonary disease. *PLoS One*. 2017;12(1):e0170023. <https://doi.org/10.1371/journal.pone.0170023>
37. Lee GH, Cheng NW, Yu HH, et al. A novel zebrafish model to emulate lung injury by folate deficiency-induced swim bladder defectiveness and protease/antiprotease expression imbalance. *Sci Rep*. 2019;9:12633. <https://doi.org/10.1038/s41598-019-49152-7>
38. Bigot P, Chesseron S, Saidi A, et al. Cleavage of occludin by cigarette smoke-elicited cathepsin S increases permeability of lung epithelial cells. *Antioxidants (Basel)*. 2022;12(1):5. <https://doi.org/10.3390/antiox12010005>
39. Chitsamankhun C, Siritongtaworn N, Fournier BPJ, et al. Cathepsin C in health and disease: from structural insights to therapeutic prospects. *J Transl Med*. 2024;22:777. <https://doi.org/10.1186/s12967-024-05589-7>