

基于慢性阻塞性肺疾病潜在治疗靶点的孟德尔随机化和 GEO 数据库识别分析

蒋先伟^{1,2}, 王明航^{1,2,3}, 李慧茹^{1,2}, 董晓升^{1,2}, 刘元元^{1,2}

(1. 河南中医药大学第一附属医院 国家区域中医(肺病)诊疗中心, 河南 郑州 450000; 2. 河南中医药大学第一临床医学院, 河南 郑州 450046; 3. 河南中医药大学第一附属医院呼吸科, 河南 郑州 450000)

[摘 要] **目的:** 通过使用微阵列数据集和孟德尔随机化 (MR) 方法筛选慢性阻塞性肺疾病 (COPD) 患者遗传、诊断及治疗的关键靶点, 为 COPD 患者的临床诊疗提供依据。**方法:** 通过基因表达综合数据库 (GEO) 获得 4 个 COPD 基因表达谱数据, 采用 R 软件对数据进行整理和归一化处理, 并筛选出差异表达基因 (DEGs)。MR 分析 COPD 与相关表达数量性状位点 (eQTL) 的因果关系, 再与 DEGs 取交集来确定潜在关键靶点。采用基因集合富集分析 (GSEA)、基因本体论 (GO) 功能富集分析和京都基因与基因组百科全书 (KEGG) 信号通路富集分析探讨关键靶点的功能作用和途径, 外部数据集对其表达进行验证。**结果:** 共筛选出 1 571 个 DEGs, 其中表达上调基因 820 个, 表达下调基因 751 个。MR 分析最终筛选出 286 个 COPD 相关基因, 通过与 DEGs 取交集得出 3 个上调基因, 包括二酰甘油激酶 γ (DGKG)、重肽神经丝蛋白 (NEFH) 和 Fc 受体样 B (FCRLB); 6 个下调基因, 包括金属还原酶 4 (STEAP4)、含普列克底物蛋白家族 F 成员 2 (PLEKHF2)、分化簇 3D (CD3D)、转胶蛋白 2 (TAGLN2)、三基序蛋白 22 (TRIM22) 和核糖体蛋白 L9 (RPL9)。生物学功能分析结果主要涉及铁输入细胞、氧化还原酶活性、原发性免疫缺陷症和辅助性 T (Th) 1 及 Th2 细胞分化等途径。MR 分析结果证实了上述靶点与 COPD 之间的因果关系。外部验证集分析, 与健康对照组比较, COPD 样本组中 FCRLB 基因表达水平升高 ($P < 0.01$), CD3D 和 RPL9 基因表达水平降低 ($P < 0.05$ 或 $P < 0.01$), 与 MR 分析结果一致, 证明了本研究的可靠性。**结论:** DGKG、NEFH、FCRLB、STEAP4、PLEKHF2、CD3D、TAGLN2、TRIM22 和 RPL9 可能是 COPD 患者疾病发生发展过程中的重要调控因子及临床诊疗靶点。

[关键词] 慢性阻塞性肺疾病; 孟德尔随机化; 基因本体论功能富集分析; 京都基因和基因组百科全书信号通路富集分析; 生物标志物; 生物信息学

[中图分类号] R563.9 **[文献标志码]** A

[收稿日期] 2024-08-19 **[录用日期]** 2024-10-04

[基金项目] 中国生物技术发展中心国家重点研发计划项目 (2023YFC3502602, 2023YFC3502600); 河南省科技厅科技攻关项目 (232102310472); 河南省卫健委国家中医临床研究基地科研专项 (2022JDZX046); 河南省卫健委中医药科学研究专项 (2022ZY1047); 河南省教育厅高校科技创新团队项目 (23IRTSTHN027)

[作者简介] 蒋先伟 (1995—), 男, 河南省周口市人, 医师, 在读博士研究生, 主要从事中医药防治呼吸系统疾病的基础和临床方面的研究。

[通信作者] 王明航, 主任医师, 博士研究生导师 (E-mail: wmh107hn@163.com)

©《吉林大学学报(医学版)》编辑部, 开放获取遵循 CC BY-NC-ND 协议。

© Editorial Board of Journal of Jilin University (Medicine Edition). Open access under CC BY-NC-ND license.

Mendelian randomization and GEO database identification analysis based on potential therapeutic targets for chronic obstructive pulmonary disease

JIANG Xianwei^{1,2}, WANG Minghang^{1,2,3}, LI Huiru^{1,2}, DONG Xiaosheng^{1,2}, LIU Yuanyuan^{1,2}

(1. First Affiliated Hospital, Henan University of Traditional Chinese Medicine, National Regional Chinese Medicine (Lung Disease) Diagnostic and Treatment Center, Zhengzhou 450000, China; 2. Henan University of Traditional Chinese Medicine, First School of Clinical Medicine, Zhengzhou 450046, China; 3. Department of Respiratory Medicine, First Affiliated Hospital, Henan University of Traditional Chinese Medicine, Zhengzhou 450000, China)

ABSTRACT Objective: To screen the key genetic, diagnostic and therapeutic targets of chronic obstructive pulmonary disease (COPD) patients by using microarray datasets and Mendelian randomization (MR) method, and to provide the evidence for clinical diagnosis and treatment of COPD. **Methods:** Four COPD gene expression profile datasets were obtained from the Gene Expression Omnibus (GEO) database. The data were processed and normalized using R software, and differentially expressed genes (DEGs) were screened. MR analysis was performed to explore the causal relationship between COPD and expression quantitative trait loci (eQTL), intersection with DEGs was taken to identify potential key targets. Gene Set Enrichment Analysis (GSEA), Gene Ontology (GO) functional enrichment analysis, and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway enrichment analysis were conducted to investigate the functional roles and pathways of the key targets, external datasets were used to validate their expression. **Results:** A total of 1 571 DEGs were screened, including 820 upregulated genes and 751 downregulated genes. MR analysis identified 286 COPD-related genes, and intersection with DEGs revealed 3 upregulated genes: diacylglycerol kinase gamma (*DGKG*), neurofilament heavy polypeptide (*NEFH*), and Fc receptor like B (*FCRLB*); and 6 downregulated genes: STEAP4 metalloredutase (*STEAP4*), pleckstrin homology domain containing family F member 2 (*PLEKHF2*), CD3d molecule (*CD3D*), transgelin 2 (*TAGLN2*), tripartite motif containing 22 (*TRIM22*), and ribosomal protein L9 (*RPL9*). The biological function analysis results indicated that these genes were mainly involved in pathways such as iron ion transport into the cells, oxidoreductase activity, primary immunodeficiency, and Th1 and Th2 cell differentiation. The MR analysis results confirmed the causal relationship between these targets and COPD. The external validation results showed that compared with healthy controls, the expression level of FCRLB in COPD samples was significantly increased ($P < 0.01$), while the expression levels of CD3D and RPL9 were significantly decreased ($P < 0.05$ or $P < 0.01$), which was consistent with the MR analysis results, highlighting the reliability of this study. **Conclusion:** *DGKG*, *NEFH*, *FCRLB*, *STEAP4*, *PLEKHF2*, *CD3D*, *TAGLN2*, *TRIM22*, and *RPL9* may serve as important regulatory factors and clinical diagnostic/therapeutic targets in the pathogenesis of COPD, providing clues for early screening, diagnosis, and targeted treatment of COPD.

KEYWORDS Chronic obstructive pulmonary disease; Mendelian randomization; Gene Ontology functional enrichment analysis; Kyoto Encyclopedia of Genes and Genomes signaling pathway enrichment analysis; Biomarker; Bioinformatics

慢性阻塞性肺疾病 (chronic obstructive pulmonary disease, COPD) 是全球四大慢性疾病之一, 其具有迁延难愈、反复发作和持续进展的特

性, 2019年, 全球报告了2.123亿例COPD流行病例, 其中COPD导致330万人死亡^[1-3]。COPD是一项重要的公共卫生挑战, 如何从危险因素入手进

行早期预防和治疗，成为目前 COPD 研究的关键。

COPD 的危险因素具有多样性，为个体易感因素与环境因素共同作用的结果，其中遗传因素在个体易感因素中占据重要位置。近年来，测序技术的迅速发展和孟德尔随机化（mendelian randomization, MR）方法的广泛应用，为挖掘 COPD 的遗传相关靶点提供了机会，研究^[4-5]显示：编码基质金属蛋白酶 12（matrix metalloproteinase 12, MMP12）和谷胱甘肽 S-转移酶（glutathione s-transferase, GST）的基因多态性可能与肺功能的下降有关。全基因组扫描研究^[6]显示： α 尼古丁乙酰胆碱受体和刺猬因子相互作用蛋白等与 COPD 或者肺功能有关。国际 COPD 遗传学联盟最新的研究^[7]结果显示：有 82 个与 COPD 有关的基因位点，不同基因与 COPD 的不同病理或临床特征关联，从遗传基因的角度支持 COPD 存在异质性。一项 MR 分析^[8]显示：抗炎性肺蛋白棒细胞分泌蛋白 16 在血清中的高表达可增加 COPD 发病的风险并有助于疾病进展，可能成为 COPD 潜在的生物标志物。但上述研究缺乏大样本数据进行独立验证，本研究利用基因表达综合数据库（Gene Expression Omnibus, GEO）中多个独立的 COPD 数据集进行联合分析，整合了表达数量性状位点（eQTL-expression quantitative trait locus, eQTL）数据和全基因组关联研究（Genome-Wide Association Study, GWAS）数据，选用 MR 方法探讨相关基因与 COPD 间的因果关系，挖掘 COPD

潜在的治疗靶点，探讨其作用机制，为 COPD 患者的临床诊疗提供参考。

1 资料与方法

1.1 数据集来源 使用 GEO 数据库（<https://www.ncbi.nlm.nih.gov/>）获取 COPD 相关基因表达数据集，以“chronic obstructive pulmonary disease”为检索词检索后获得 GSE130928、GSE11906、GSE13896、GSE64614 和 GSE212331 数据集。为了扩大数据来源，将 GSE130928、GSE11906、GSE13896 和 GSE64614O 数据集作为训练集进行合并，使用 R 软件（版本：4.4.0）对上述数据集进行数据读取和校正，使用 sva 包（版本：3.52.0）进行归一化处理。利用 ggplot2（版本：3.5.1）和 ggpubr 包（版本：0.6.0）进行主成分分析（principal component analysis, PCA），消除批次效用。将 GSE212331 数据集作为外部验证集对筛选的基因进行验证。见表 1。

COPD 的结局数据来源于 GWAS 汇总数据集（Integrative Epidemiology Unit, IEU）的遗传关联数据库（<https://gwas.mrcieu.ac.uk/>）。所使用的 GWAS ID 为 ebi-a-GCST90018807，涉及 13 530 例 COPD 病例和 454 945 例欧洲血统对照者，共包括 24 180 654 个单核苷酸多态性（single nucleotide polymorphisms, SNPs）。eQTL 暴露数据来源于 eQTL 基因联盟（<https://www.eqtlgen.org/>）^[9]，包含 31 684 份全血样本的 eQTL 数据，用于后续 MR 分析。

表 1 GEO 数据集具体信息
Tab. 1 Specific informations of GEO datasets

GEO(ID)	<i>n</i>	Tissue (Homo sapiens)	Platform	Last updated	Attribute
GSE130928	88	Human Alveolar Macrophages	GPL570	April 17, 2020	Training
GSE11906	223	Human Airway Epithelium	GPL570	March 25, 2019	Training
GSE13896	70	Human Alveolar Macrophages	GPL570	March 25, 2019	Training
GSE64614	221	Human Airway Epithelium	GPL570	July 08, 2019	Training
GSE212331	87	Induced Sputum	GPL10558	December 14, 2023	Validation

1.2 差异表达基因(differentially expressed genes, DEGs)筛选和可视化 使用 R 软件中 limma 包（版本：3.58.1）对归一化后的数据集内的 DEGs 进行筛选，设置差异倍数（fold change, FC），筛选条件为 $|\log_2FC| > 0.585$ ，矫正后的 $P < 0.05$ ，筛选后获得 DEGs，并绘制火山图和热图，进行可视化

展示。
1.3 eQTL 数据筛选 获得的 eQTL 数据作为暴露因素，使用 R 软件中 TwoSampleMR 包（版本：0.6.2）筛选出强相关的 SNPs（ $P < 5 \times 10^{-8}$ ）作为工具变量（instrumental variable, IV），为了去除连锁不平衡，设置 $r^2 < 0.001$ ，遗传距离为 10 000 kb，

同时设置 $F > 10$ 排除弱关联或表型方差解释不足的SNP。

1.4 MR分析 本次研究中首先将eQTL数据作为暴露因素、SNPs作为工具变量、GWAS疾病数据集(ebi-a-GCST90018807)作为结局数据进行MR分析。采用R软件中TwoSampleMR”包和“forestplot”包进行数据分析及可视化。本次分析采用逆方差加权法(inverse-variance weighted, IVW)为主分析方法,MR-Egger法和加权中位数方法进行数据分析。为了更好地确定相关基因,设定该基因至少应满足IVW的结果有显著性差异($P < 0.05$),同时加权中位数和MR-Egger的结果与IVW法随机效应模型的结果必须方向相同。MR分析后,采用Cochran's Q检验评估IVW随机效应模型异质性,若检验的结果显示 $P < 0.05$,则表示存在异质性,将该基因剔除^[8]。采用MR-Egger回归确定潜在的水平多效性,结果中截距项的值接近于0且 $P > 0.05$ 时,则该MR结果不存在多效性^[9],最后,再通过漏斗图对结果进行评判,漏斗图对称则表示结果无潜在的异质性^[10]。

对“1.2”中得到的DEGs与结果进行交集,获得与疾病相关的交集基因,同时根据其在试验组和对照组的表达量差异,分为上调基因与下调基因,再对得到的交集基因作为暴露因素再次进行MR分析,评估其与疾病的因果关系。

1.5 交集基因的基因本体论(Gene Ontology,GO)富集分析和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes,KEGG)信号通路富集分析 利用R软件中clusterProfiler包(版本:4.10.1)对共表达基因进行GO生物功能富集分析和KEGG信号通路富集分析,设置过滤条件为 $P < 0.05$,对富集显著的生物功能或通路以柱状图进行表示。为了明确交集基因的染色体位置,使用OmicCircos(版本:1.32.0)对交集基因的基因组位置进行可视化。

1.6 基因集富集分析(Gene Set Enrichment Analysis,GESA) 为了进一步挖掘上述基因的功能和通路富集情况,本课题组利用从分子特征数据库(<https://www.gsea-msigdb.org/gsea/msigdb>)下载所得的c2.cp.kegg.Hs.symbols.gmt基因集为参考,对交集基因进行分析,以 $P < 0.05$ 为筛选条件确定显著富集的通路,对其中归一化富集评分(normalized enrichment score, NSE) > 0 和NSE $<$

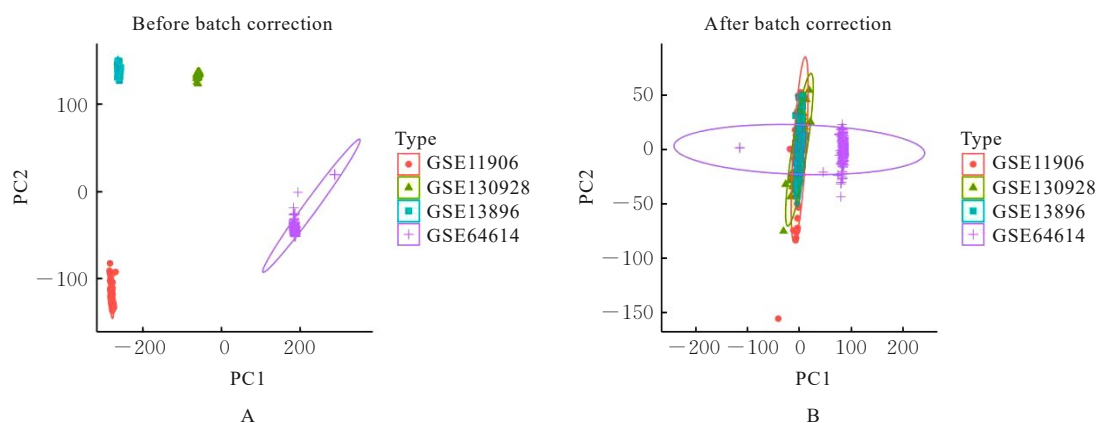
0的前5个通路进行可视化。

1.7 外部验证组差异分析 使用limma包(版本:3.58.1)读取GSE212331数据集中试验组与对照组的表达数据,以验证共表达基因是否在对照组与试验组之间表现出差异,将上述结果与MR分析结果进行比较,并进行可视化。

2 结果

2.1 DEGs鉴定结果 对GSE130928、GSE11906、GSE13896和GSE64614数据集中的每个基因的表达值进行校正及合并,并通过PCA分析消除批效应。经过校正后,数据集中的所有样本都达到了可接受的同质性(图1)。通过对校正后的合并数据集中对照组和试验组的表达数据进行分析,共获得1571个DEGs,其中820基因表达上调,751个基因表达下调(图2A)。按表达量进行排序,前10个表达上调的基因:细胞因子样蛋白1(cytokine-like 1, CYTL1)、安眠蛋白1 α (meprin alpha, MEP1A)、泛素羧基端酯酶L1(ubiquitin C-terminal hydrolase-L1, UCHL1)、髓系C型凝集素结构域家族5成员A(myeloid C-type lectin domain family 5 member A, CLEC5A)、脱氧核糖核酸酶2B(deoxyribonuclease 2 beta, DNASE2B)、基质金属蛋白酶12(matrix metalloproteinase 12, MMP12)、趋化因子2(chemokine 2, CCL2)、几丁质酶1(recombinant chitinase 1, CHIT1)、细胞色素P450家族成员1B1(cytochrome P450 1B1, CYP1B1)和细胞色素P450家族成员1A1(cytochrome P450 1A1, CYP1A1);前10个表达下调的基因:分泌球蛋白家族3A成员1(secretoglobulin family 3A member 1, SCGB3A1)、分泌珠蛋白家族1A成员1(secretoglobulin family 3A member 1, SCGB1A1)、维生素K依赖性蛋白S(vitamin K-dependent protein S, PROS1)、角蛋白19(keratin 19, KRT19)、人类白细胞抗原B(human leukocyte antigen-B, HLA-B)、半胱氨酸丰富蛋白1(cysteine rich protein 1, CRIP1)、视黄酸受体应答蛋白3(retinoic acid receptor response protein 3, RARRES3)、水通道蛋白3(aquaporin 3, AQP3)、四旋蛋白6(tetraspanin 6, TSPAN6)和溶质载体家族22成员4(solute carrier family 22 member 4, SLC22A4)(图2B)。

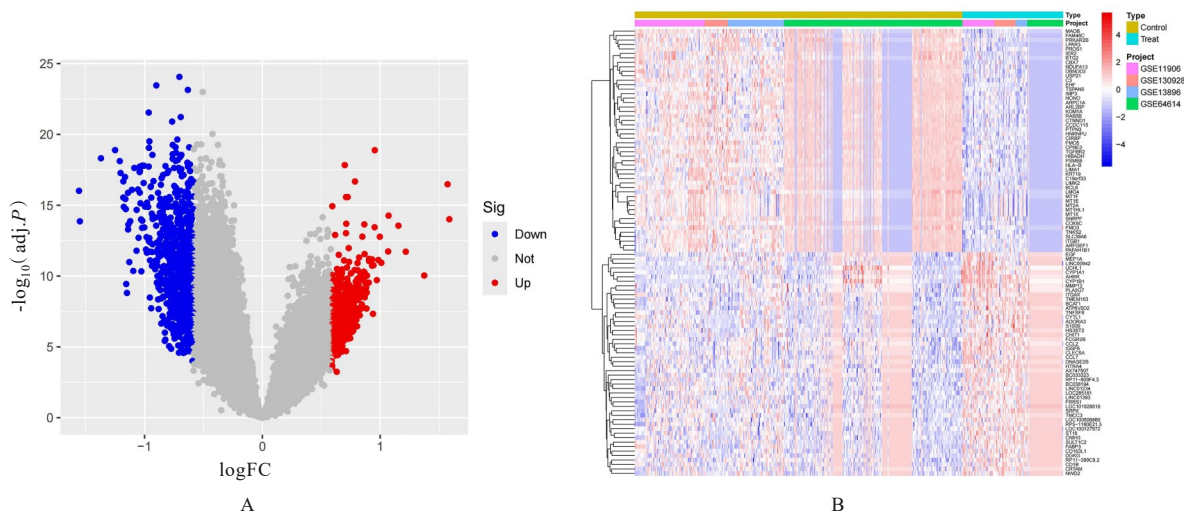
2.2 MR分析结果 经过筛选,最终获得25457个强关联SNPs作为工具变量,所有SNPs的F值



A: Before PCA analysis; B: After PCA analysis.

图1 PCA分析图

Fig. 1 PCA analysis diagram



A: Differential gene volcano map; B: Differential gene heat map.

图2 DEGs筛选图

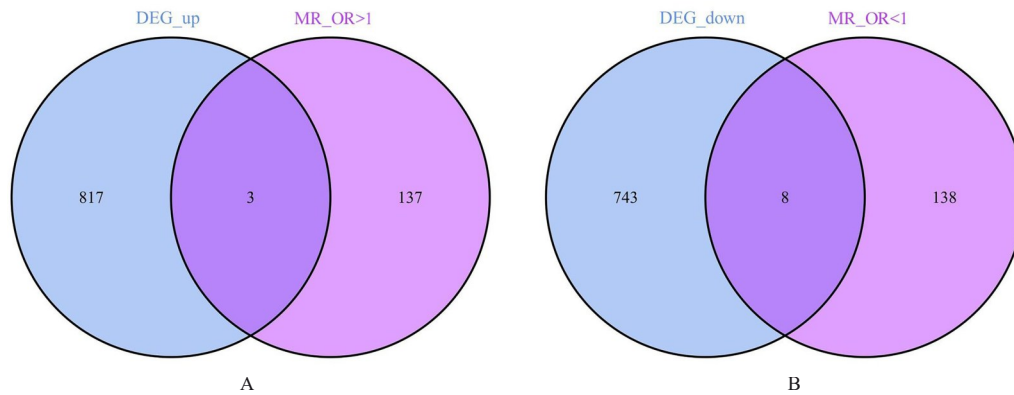
Fig. 2 DEGs screening diagram

均 >10 。同时鉴定出了286个COPD相关基因,通过与“2.1”中获得的DEGs取交集得出3个上调基因,包括二酰甘油激酶 γ (diacylglycerol kinase gamma, *DGKG*)、重肽神经丝蛋白 (neurofilament heavy polypeptide, *NEFH*) 和Fc受体样B (Fc receptor like B, *FCRLB*); 8个下调基因,包括HCP5 (HLA complex P5, *HCP5*)、PTGFR (prostaglandin F receptor, *PTGFR*)、金属还原酶4 (*STEAP4* metalloredutase, *STEAP4*)、含普列克底物蛋白家族F成员2 (pleckstrin homology domain containing family F member 2, *PLEKHF2*)、分化簇3D (CD3d molecule, *CD3D*)、转胶蛋白2 (transgelin 2,

TAGLN2)、三基序蛋白22 (tripartite motif containing 22, *TRIM22*) 和核糖体蛋白L9 (ribosomal protein L9, *RPL9*) (图3)。将上述11个交集基因作为暴露因素分别进行MR分析,评估上述基因与COPD的因果关系,结果显示:*HCP5*与*PTGFR*截距项的值接近于0且 $P>0.05$,存在多效性,将其剔除。最终结果显示:在采用IVW方法进行的MR分析中,3个上调的基因与COPD存在显著的正相关关系,*DGKG* [比值比 (odds ratio, OR) =1.184, 95%置信区间 (confidence interval, CI): 1.019~1.376, $P=0.027$]; *NEFH* (OR=1.078, 95%CI: 1.011~1.148, $P=0.021$); *FCRLB* (OR=1.085, 95%CI: 1.004~

1.172, $P=0.04$)。6个下调的基因降低COPD发生风险, *STEAP4* (OR=0.963, 95%CI: 0.934~0.993, $P=0.017$); *PLEKHF2* (OR=0.942, 95%CI: 0.892~0.995, $P=0.032$); *CD3D* (OR=0.855, 95%CI: 0.773~0.946, $P=0.002$); *TAGLN2* (OR=0.955, 95%CI: 0.912~1.000,

$P=0.049$); *TRIM22* (OR=0.921, 95%CI: 0.855~0.992, $P=0.03$); *RPL9* (OR=0.888, 95%CI: 0.823~0.958, $P=0.002$), 见图4。表明3个上调基因增加COPD发生风险 (OR>1), 6个下调基因降低COPD发生风险 (OR<1)。交集基因在染色体中分布图见图5。



A: Up-regulated genes; B: Down-regulated genes.

图3 交集基因图

Fig. 3 Intersecting gene maps

2.3 GO功能富集分析和KEGG信号通路富集分析结果 GO功能富集分析共得到64条结果, 其中生物过程 (biological process, BP) 43条, 主要涉及: 铁输入细胞、蛋白质同源三聚化和外周神经系统神经元的发育及分化等; 细胞组分 (cellular component, CC) 11条, 主要涉及 α - β T细胞受体复合体、突触后细胞骨架和早期内质体膜等; 分子功能 (molecular function, MF) 10条, 主要涉及: ATP依赖性二酰甘油激酶活性和氧化还原酶活性等。KEGG通路富集分析富集到11条通路, 主要涉及原发性免疫缺陷症和甘油酯代谢以及辅助性T (helper T, Th) 1和Th2细胞分化等 (图6)。

2.4 GESA MR分析中, 上调基因中的*NEFH*基因表达量升高增加了COPD发生风险 ($P=0.021$), 下调基因中*CD3D*和*RPL9*基因表达量升高降低了COPD发生风险 ($P=0.002$), 为了进一步挖掘上述3个基因的功能, 进行了GESA, 结果显示: *NEFH*基因高表达组中前5个通路是细胞因子-细胞因子受体相互作用、成熟期发病的青年糖尿病、神经活性配体受体相互作用、嗅觉传导和味觉传导; *NEFH*基因低表达组中前5个通路是亨廷顿氏症、氧化磷酸化、帕金森病、核糖体和鳞状细胞; *CD3D*基因高表达组中前5个通路是细胞因子-细

胞因子受体相互作用、Janus激酶 (Janus kinase, JAK)-信号转导和转录激活因子 (signal transducer and activator of transcription, STAT) 信号通路、成熟期发病的青年糖尿病、神经活性配体受体相互作用和嗅觉传导; *CD3D*基因低表达组中前5个通路是亨廷顿氏症、溶酶体、帕金森病、核糖体和鳞状细胞; *RPL9*基因高表达组中前5个通路是亨廷顿氏症、氧化磷酸化、帕金森病、核糖体和鳞状细胞; *RPL9*基因低表达组中前5个通路是细胞因子-细胞因子受体相互作用、成熟期发病的青年糖尿病、神经活性配体受体相互作用、嗅觉传导和味觉传导 (图7)。上述结果中*NEFH*基因为上调基因, *RPL9*基因为下调基因, 所富集到的通路在高表达组与低表达组中结果相反, 证实本研究结果的可靠性。

2.5 外部验证组差异分析 COPD样本组中*DGKG*、*NEFH*和*FCRLB*基因表达水平均高于健康对照组, 其中*FCRLB*表达量升高最明显 ($P<0.01$)。与健康对照组比较, COPD样本组中的*STEAP4*、*PLEKHF2*、*CD3D*、*TRIM22*和*RPL9*这5个基因表达量下调, 其中*CD3D*和*RPL9*基因表达量下调最显著 ($P<0.05$ 或 $P<0.01$), 以上结果与MR分析结果基本一致, 证明了本研究结果的可靠性。见图8。

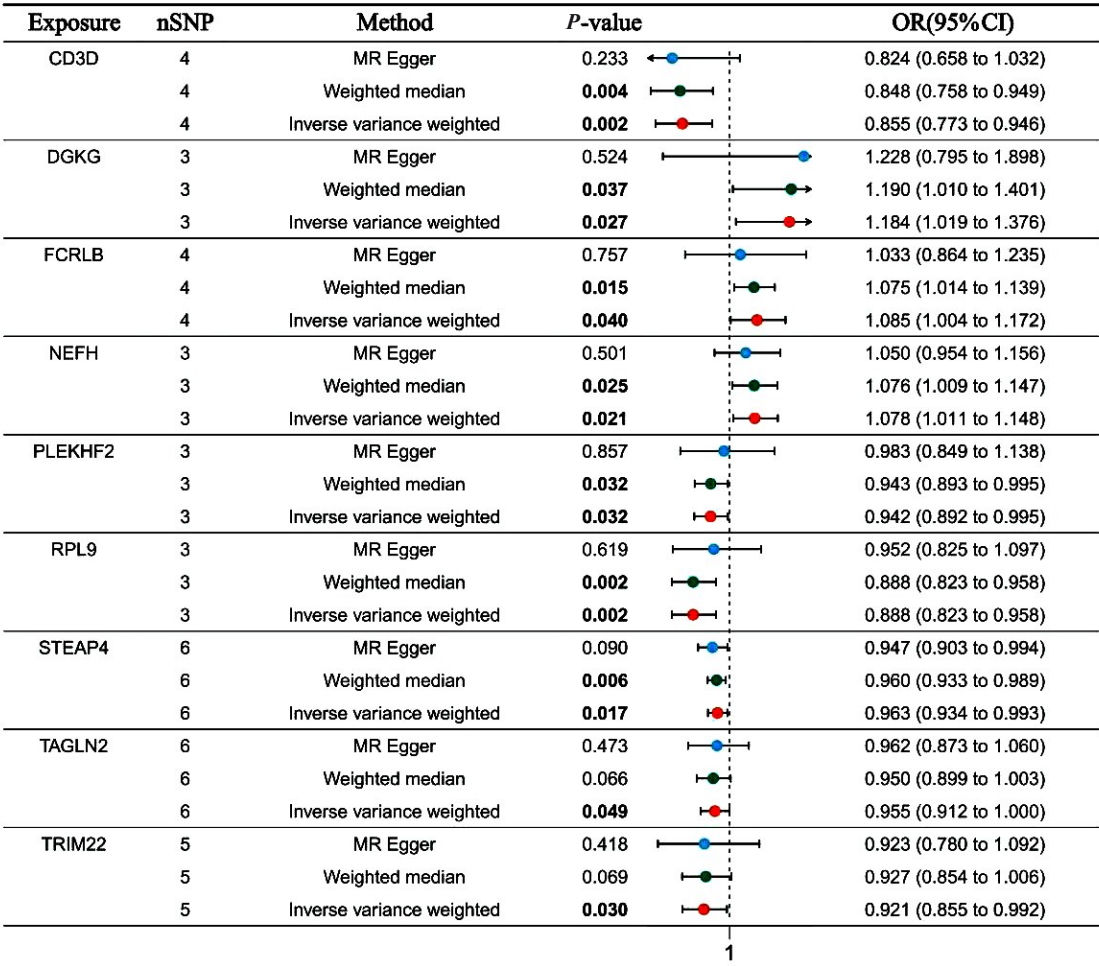


图 4 交集基因的 MR 分析森林图
Fig. 4 Forest plot of intersecting genes by MR analysis

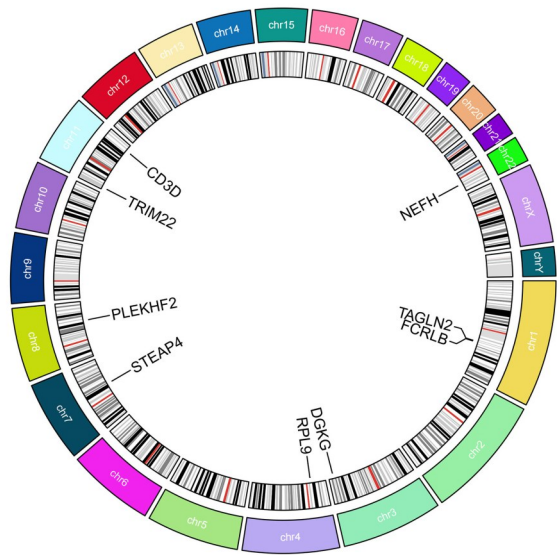


图 5 交集基因染色体位置图
Fig. 5 Chromosome location map of intersecting genes

3 讨论

COPD是呼吸系统常见病，具有遗传易感性，目前相关研究一直在致力于COPD早期诊断或疾病前诊断，因此利用MR方法筛选合适的诊断标志物或治疗靶点，是目前较为合适且可靠的方法^[11-13]。COPD的药物治疗主要包括支气管扩张剂、吸入性糖皮质激素、磷酸二酯酶4（phosphodiesterase 4，PDE-4）抑制剂和祛痰药及抗氧化剂等^[1, 14]，而相关生物制剂或靶向药物的使用也为COPD治疗提供了新选择^[15]，因此，研发相关生物制剂或靶向药物，可能是COPD未来治疗的主要发展方向^[16]。

本研究利用GEO数据库中的多个数据集和GWAS数据库中COPD遗传数据及eQTL数据进行MR分析，鉴定出9个共表达基因，包括3个上调基因：DGKG、NEFH和FCRLB基因，其表达量升高增加了COPD的发生风险；6个下调基因：STEAP4、PLEKHF2、CD3D、TAGLN2、TRIM22

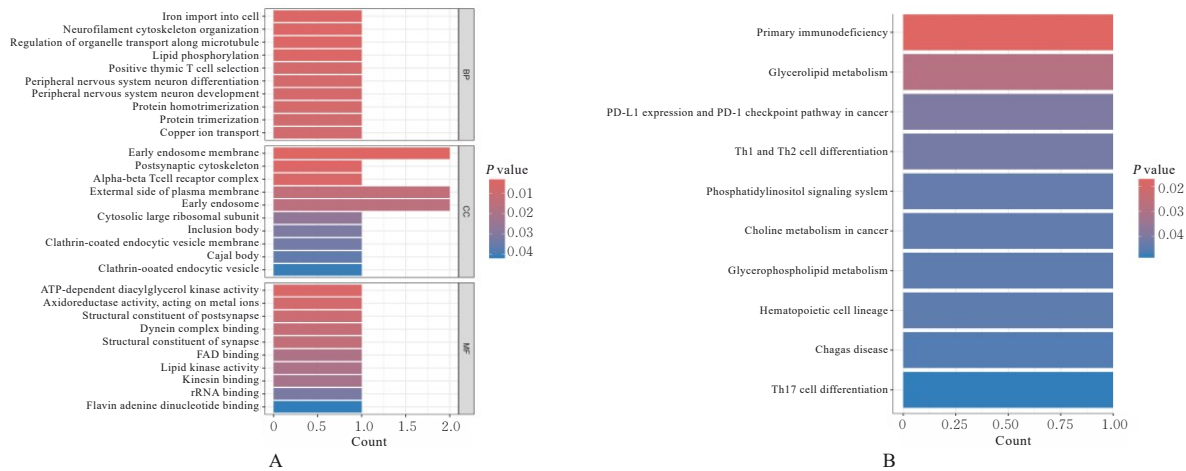


图6 交集基因的GO功能富集分析和KEGG信号通路富集分析

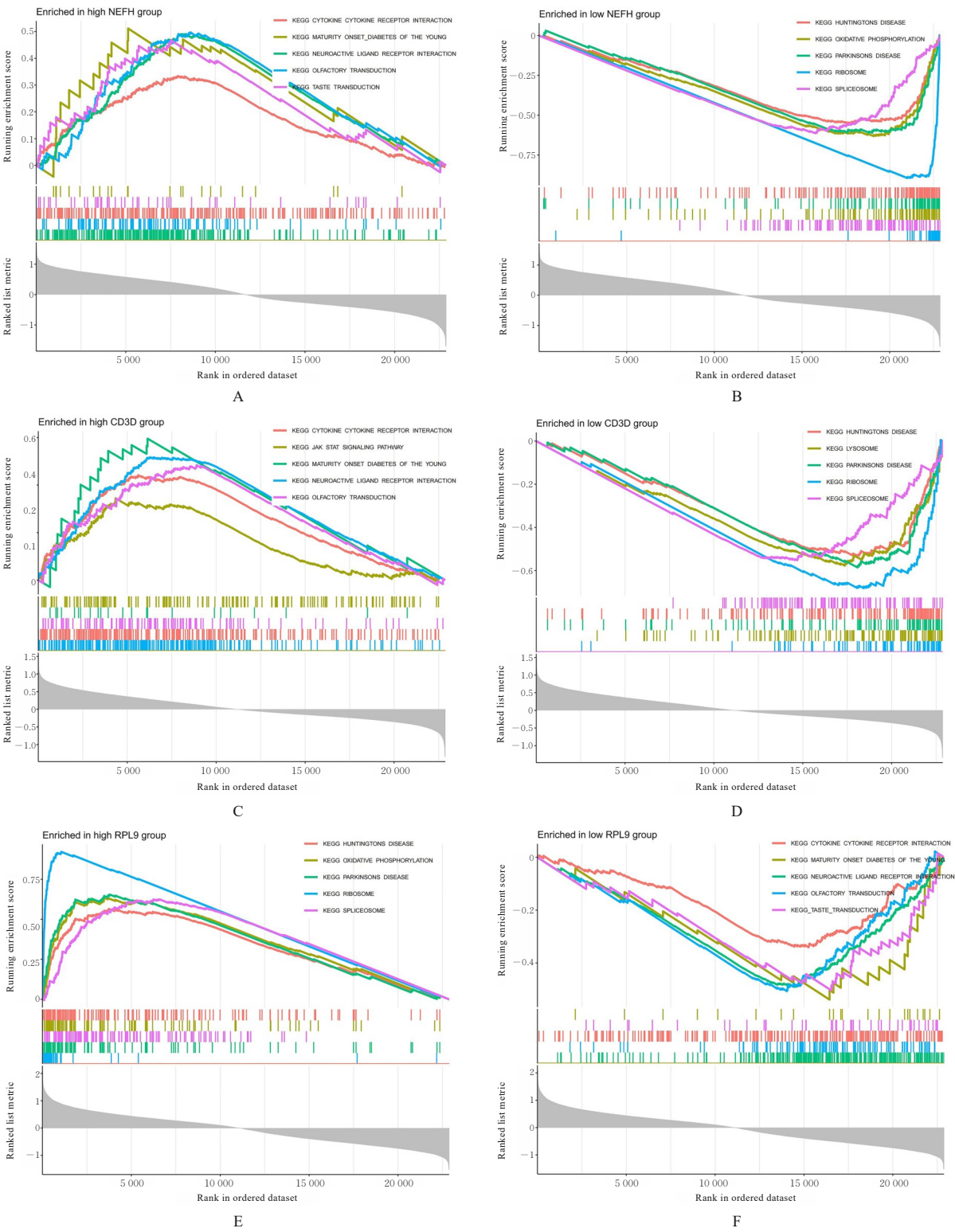
Fig. 6 GO functional enrichment analysis and KEGG signaling pathway enrichment analysis of intersecting genes

和 *RPL9* 基因, 其表达量减少也与疾病风险的降低有关。上述基因的表达模式为 COPD 的遗传基础提供了新的见解, 可能反映了疾病病理中的关键分子变化, 如缺氧诱导的肿瘤血管内皮细胞中 *DGKG* 过度表达可通过 E 盒结合锌指蛋白 2 (zinc finger E-box-binding protein, ZEB2) / 转化生长因子 $\beta 1$ (transforming growth factor- $\beta 1$, TGF- $\beta 1$) 轴促进肿瘤血管生成和免疫抑制调节性 T 淋巴细胞分化, 与 COPD 中的气道重塑与免疫调节机制基本一致^[17-18]。SONG 等^[19] 利用机器学习算法以 *FCRLB* 等 7 个核心基因构建肺腺癌焦亡相关的预后风险评估模型, 具有较好的准确度和灵敏度, 而细胞焦亡在 COPD 的发生发展中起关键作用^[20]。WANG 等^[21] 发现: *CD3D* 基因与儿童哮喘密切相关, 并在哮喘患者的肺泡灌洗液细胞中进行了验证。免疫细胞浸润分析^[22] 结果显示: *CD3D* 基因与静息肥大细胞和嗜酸性粒细胞增加呈负相关关系, 与 Th1、Th2 和 Th17 细胞中的几种细胞标志物高度相关, 与目前 COPD 的免疫炎症反应机制结果一致。*RPL9* 是一种新型晚期糖基化终末产物结合蛋白, 可以减少脂多糖和高迁移率族蛋白 B1 (high mobility group box 1, HMGB1) 刺激下的巨噬细胞中 TNF- α 的表达, 从而减轻炎症^[23-24]。另一项针对运动机能损伤的研究^[25] 显示: COPD 患者股四头肌中微小 RNA542-3p/5p (microRNA 542-3p/5p, miR-542-3p/5p) 水平升高, 抑制核糖体蛋白合成, 并与其肺部疾病的严重程度成正比, 该结果表明核糖体蛋白在 COPD 进程中可能发挥着重要作用。目前, 关于上述基因在 COPD 发病机制

中的研究较少, 但本文作者发现多数蛋白与 COPD 发病机制密切相关, 如干预慢性炎症、减轻细胞焦亡以及免疫调节等方面, 但上述机制仍需进一步深入探讨和验证。

本研究中, GO 功能富集分析结果显示: 交集基因主要涉及铁输入细胞、 α - β T 细胞受体复合体和氧化还原酶活性等; KEGG 信号通路富集分析结果显示: 共表达基因主要涉及甘油酯代谢以及 Th1 和 Th2 细胞分化等。上述结果与既往研究^[26-29] 一致, NAJAFINOBAR 等^[26] 利用飞行时间二次离子质谱技术 (time-of-flight secondary ion mass spectrometry, ToF-SIMS) 法对 COPD 患者肺组织中的铁积累进行量化, 结果显示: 与对照组比较, COPD 患者肺组织切片中铁含量升高 12 倍。另一项研究^[27] 也证实: 气道铁含量升高会导致更频繁的 COPD 急性加重。COPD 是一种异质性慢性炎症性疾病, Th1 和 Th2 细胞分化在 COPD 的发病机制中扮演了重要角色, 其产生的多种细胞因子或介质, 可直接介导 COPD 炎症的发生^[28-29]。基于上述研究, 本文作者推测: 交集基因可能是通过干预 Th1 和 Th2 细胞分化影响自身免疫, 以及抑制铁离子流入细胞来减轻 COPD 慢性炎症, 影响疾病进程。

为了进一步探讨共表达基因在染色体中的分布, 本文作者确定了基因位点, 发现共表达基因与多条染色体相关。相关研究^[30] 显示: X 染色体与 COPD 表型有关, *TMSB4X* 基因附近的 1 个变体 rs5979771 与肺功能相关并呈现全基因组显著性。本研究结果显示: 位于 1 号和 11 号染色体上的基因



A: NEFH high-expression group; B: NEFH low-expression group; C: CD3D high-expression group; D: CD3D low-expression group; E: RPL9 high-expression group; F: RPL9 low-expression group.

图 7 GESA 结果图

Fig. 7 Diagrams of results of GESA

最多，该结果与既往的 GWAS 分析^[31-32]具有一致性。上述染色体可能与 COPD 显著相关，对 COPD

相关 SNPs 的鉴定也具有一定的意义。

基于 MR 分析结果，本文作者选择了 3 个显著

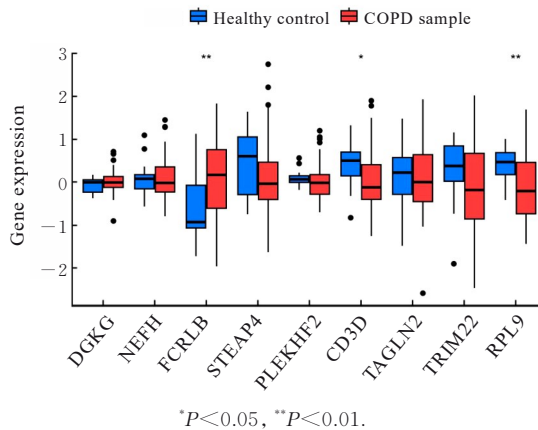


图8 验证组差异分析箱线图

Fig. 8 Boxplot of variance analysis of validation group

关联的 *NEFH*、*CD3D* 和 *RPL9* 基因, 利用 GESA 富集分析探讨其潜在影响通路, 同时基于外部验证数据集对交集基因进行验证, 结果显示: *CD3D* 和 *RPL9* 基因在 COPD 样本组中下调; GESA 富集分析结果显示: 3 个基因主要富集于细胞因子-细胞因子受体相互作用、氧化磷酸化、核糖体和鳞状细胞等通路, 上述通路均与 COPD 的发生发展密切相关^[33-35]。YEW-BOOTH 等^[33] 利用蛋白印迹或免疫组织化学技术对吸烟者、COPD 患者和健康人肺组织中 JAK-STAT 信号通路蛋白进行检测, 结果表明: JAK-STAT 信号通路在 COPD 患者中被异常激活, 证实了 JAK-STAT 信号通路在 COPD 发生发展中的作用。另一项动物实验^[36-37] 结果显示: 过氧化物酶体增殖物激活受体- γ 激动剂可以通过抑制 JAK-STAT 信号通路明显改善香烟烟雾或脂多糖刺激的巨噬细胞的极化障碍和功能活性, 减轻气道炎症, 恢复肺功能。研究^[38] 显示: 受环境因素等影响, COPD 患者外源性活性氧和肺部炎症及结构细胞产生的内源性活性氧增多, 造成线粒体功能障碍, 引起气道上皮细胞损伤, 诱导平滑肌细胞增生, 导致肺功能下降, 与氧化还原及氧化磷酸化反应密切相关。目前, 临床中抗氧化剂的应用是 COPD 临床治疗的关键措施, 如 N-乙酰半胱氨酸等, 可以改善并延缓 COPD 的急性发作。上述研究表明: *NEFH*、*CD3D* 和 *RPL9* 基因可能通过 JAK-STAT 信号通路影响 COPD 进程中的氧化应激及线粒体功能, 进一步影响 COPD 发生发展, 该推论可能是 COPD 的潜在机制, 为后续治疗及机制探讨提供了参考。

本研究基于 MR 和 GEO 数据库挖掘出在

COPD 疾病过程中与遗传因素高度相关的 9 个基因, 通过后续的深入分析基因的 MF 和 BP 等, 明确出 *NEFH*、*CD3D* 和 *RPL9* 基因为 3 个核心基因, 较完整地揭示了基因与疾病的相关性以及未来向临床应用转化的可能性, 但仍需要后期进一步实验验证。本研究中相关步骤虽然较为完整, 但仍有分析技术存在局限性, 相关数据集选择存在主观性等不足。但该研究筛选出的基因, 可为 COPD 精准医疗的发展提供新参考, 为后续研究和相关药物的研发提供了一定的方向, 具有一定的临床意义。

利益冲突声明:

所有作者声明不存在利益冲突。

作者贡献声明:

蒋先伟参与研究设计、数据分析和论文撰写, 王明航参与论文审校, 李慧茹参与网络构建和关键基因验证分析, 董晓升参与数据收集和统计学分析, 刘元元参与图片整合和论文校对。

[参考文献]

- [1] 易 琼, 李 方, 雷 思, 等. 自主研发的藏四味清肺合剂联合常规治疗对慢性阻塞性肺疾病急性加重期患者的临床疗效和安全性[J]. 中南大学学报(医学版), 2024, 49(6): 921-931.
- [2] SIN D D, DOIRON D, AGUSTI A, et al. Air pollution and COPD: gold 2023 committee report[J]. Eur Respir J, 2023, 61(5): 2202469.
- [3] SAFIRI S, CARSON-CHAHHOUD K, NOORI M, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019[J]. BMJ, 2022, 378: e069679.
- [4] DING Z, WANG K, LI J, et al. Association between glutathione S-transferase gene M1 and T1 polymorphisms and chronic obstructive pulmonary disease risk: a meta-analysis[J]. Clin Genet, 2019, 95(1): 53-62.
- [5] HUNNINGHAKE G M, CHO M H, TESFAIGZI Y, et al. MMP12, lung function, and COPD in high-risk populations[J]. N Engl J Med, 2009, 361(27): 2599-2608.
- [6] PILLAI S G, GE D L, ZHU G H, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci[J]. PLoS Genet, 2009, 5(3): e1000421.
- [7] SAKORNSAKOLPAT P, PROKOPENKO D, LAMONTAGNE M, et al. Genetic landscape of chronic

- obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations [J]. *Nat Genet*, 2019, 51(3): 494-505.
- [8] MILNE S, LI X, HERNANDEZ CORDERO A I, et al. Protective effect of club cell secretory protein (CC-16) on COPD risk and progression: a Mendelian randomisation study[J]. *Thorax*, 2020, 75(11): 934-943.
- [9] VÖSA U, CLARINGBOULD A, WESTRA H J, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression[J]. *Nat Genet*, 2021, 53(9): 1300-1310.
- [10] BOWDEN J, DEL GRECO M F, MINELLI C, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption[J]. *Int J Epidemiol*, 2019, 48(3): 728-742.
- [11] HARTWIG F P, DAVEY SMITH G, BOWDEN J. Robust inference in summary data Mendelian randomization *via* the zero modal pleiotropy assumption[J]. *Int J Epidemiol*, 2017, 46(6): 1985-1998.
- [12] BOWDEN J, DAVEY SMITH G, BURGESS S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression[J]. *Int J Epidemiol*, 2015, 44(2): 512-525.
- [13] CHOI J Y, RHEE C K. Diagnosis and treatment of early chronic obstructive lung disease (COPD)[J]. *J Clin Med*, 2020, 9(11): 3426.
- [14] 中华医学会, 中华医学杂志社, 中华医学会全科医学分会, 等. 中国慢性阻塞性肺疾病基层诊疗与管理指南(2024年)[J]. *中华全科医师杂志*, 2024, 23(6): 578-602.
- [15] OHNISHI H, EITOKU M, YOKOYAMA A. A systematic review and integrated analysis of biologics that target Type 2 inflammation to treat COPD with increased peripheral blood eosinophils [J]. *Heliyon*, 2022, 8(6): e09736.
- [16] WANG C X, ZHOU J D, WANG J Q, et al. Progress in the mechanism and targeted drug therapy for COPD[J]. *Signal Transduct Target Ther*, 2020, 5(1): 248.
- [17] ZHANG L R, XU J L, ZHOU S Q, et al. Endothelial DGKG promotes tumor angiogenesis and immune evasion in hepatocellular carcinoma [J]. *J Hepatol*, 2024, 80(1): 82-98.
- [18] TAN L F, YANG X F, ZHANG J X, et al. Correlation between HIF1- α expression and airway remodeling in COPD [J]. *Int J Chron Obstruct Pulmon Dis*, 2024, 19: 921-931.
- [19] SONG Y B, QU Z, FENG H, et al. Genomic and immunological characterization of pyroptosis in lung adenocarcinoma[J]. *J Oncol*, 2022, 2022: 6905588.
- [20] LIU X L, HUANG X L, XU F. The influence of pyroptosis-related genes on the development of chronic obstructive pulmonary disease [J]. *BMC Pulm Med*, 2023, 23(1): 167.
- [21] WANG Z L, HE Y, CUN Y P, et al. Identification of potential key genes for immune infiltration in childhood asthma by data mining and biological validation [J]. *Front Genet*, 2022, 13: 957030.
- [22] SUN J, LIU T, YAN Y, et al. The role of Th1/Th2 cytokines played in regulation of specific CD4⁺ Th1 cell conversion and activation during inflammatory reaction of chronic obstructive pulmonary disease [J]. *Scand J Immunol*, 2018, 88(1): e12674.
- [23] WATANABE M, TOYOMURA T, WAKE H, et al. Cationic ribosomal proteins can inhibit pro-inflammatory action stimulated by LPS+HMGB1 and are hindered by advanced glycation end products [J]. *Biotechnol Appl Biochem*, 2024, 71(2): 264-271.
- [24] WATANABE M, TOYOMURA T, WAKE H, et al. Identification of ribosomal protein L9 as a novel regulator of proinflammatory damage-associated molecular pattern molecules[J]. *Mol Biol Rep*, 2022, 49(4): 2831-2838.
- [25] FARRE-GARROS R, LEE J Y, NATANEK S A, et al. Quadriceps miR-542-3p and -5p are elevated in COPD and reduce function by inhibiting ribosomal and protein synthesis[J]. *J Appl Physiol (1985)*, 2019, 126(6): 1514-1524.
- [26] NAJAFINOBAR N, VENKATESAN S, VON SYDOW L, et al. ToF-SIMS mediated analysis of human lung tissue reveals increased iron deposition in COPD (GOLD IV) patients [J]. *Sci Rep*, 2019, 9: 10060.
- [27] ZHANG W Z, OROMENDIA C, KIKKERS S A, et al. Increased airway iron parameters and risk for exacerbation in COPD: an analysis from SPIROMICS[J]. *Sci Rep*, 2020, 10(1): 10562.
- [28] VARRICCHI G, POTO R. Towards precision medicine in COPD: targeting type 2 cytokines and alarmins[J]. *Eur J Intern Med*, 2024, 125: 28-31.
- [29] PROBOSZCZ M, GORYCA K, NEJMAN-GRYZ P, et al. Phenotypic variations of mild-to-moderate obstructive pulmonary diseases according to airway inflammation and clinical features [J]. *J Inflamm Res*, 2021, 14: 2793-2806.

- [30] HAYDEN L P, HOBBS B D, BUSCH R, et al. X chromosome associations with chronic obstructive pulmonary disease and related phenotypes: an X chromosome-wide association study [J]. *Respir Res*, 2023, 24(1): 38.
- [31] LUTZ S M, CHO M H, YOUNG K, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry[J]. *BMC Genet*, 2015, 16: 138.
- [32] ALMUNTASHIRI S, ZHU Y, HAN Y, et al. Club cell secreted protein CC16: potential applications in prognosis and therapy for pulmonary diseases[J]. *J Clin Med*, 2020, 9(12): 4039.
- [33] YEW-BOOTH L, BIRRELL M A, LAU M S, et al. JAK-STAT pathway activation in COPD [J]. *Eur Respir J*, 2015, 46(3): 843-845.
- [34] LERNER C A, SUNDAR I K, RAHMAN I. Mitochondrial redox system, dynamics, and dysfunction in lung inflammaging and COPD[J]. *Int J Biochem Cell Biol*, 2016, 81(Pt B): 294-306.
- [35] BOZINOVSKI S, VLAHOS R, ANTHONY D, et al. COPD and squamous cell lung cancer: aberrant inflammation and immunity is the common link[J]. *Br J Pharmacol*, 2016, 173(4): 635-648.
- [36] HE S R, TIAN R Y, ZHANG X Y, et al. PPAR γ inhibits small airway remodeling through mediating the polarization homeostasis of alveolar macrophages in COPD[J]. *Clin Immunol*, 2023, 250: 109293.
- [37] KUME H, YAMADA R, SATO Y, et al. Airway smooth muscle regulated by oxidative stress in COPD[J]. *Antioxidants (Basel)*, 2023, 12(1): 142.
- [38] KOLAROV V, STEVULJEVIĆ J K, ILIĆ M, et al. Factorial analysis of N-acetylcysteine and propolis treatment effects on symptoms, life quality and exacerbations in patients with Chronic Obstructive Pulmonary Disease (COPD): a randomized, double-blind, placebo-controlled trial [J]. *Eur Rev Med Pharmacol Sci*, 2022, 26(9): 3192-3199.