DOI:10.16016/j.2097-0927.202412012

基于生物信息学和机器学习鉴定慢性阻塞性肺病合并肺动脉高压标志物及实验验证

杨 燕[1],陶春容[1],朱佑君[1],张 聪[2],李德峰[3] 陆军军医大学(第三军医大学)第二附属医院:[1]心内科,[3]临床医学研究中心,重庆;[2]陆军军医大学(第三军医大学)基础医学院实验动物学教研室,重庆

[摘要] **目的** 利用生物信息学筛选诊断慢性阻塞性肺病(chronic obstructive pulmonary disease,COPD)合并肺动脉高压(pulmonary arterial hypertension,PAH)的关键生物标志物,并验证其临床意义。**方法** 通过高通量测序数据分析鉴定了COPD合并PAH的差异表达基因(differentially expressed genes,DEGs),并通过功能富集分析探讨了这些基因的生物学功能。利用最小绝对收缩与选择算子(least absolute shrinkage and selection operator,LASSO)、随机森林(random forest,RF)、支持向量机递归特征消除(support vector machine-recursive feature elimination,SVM-RFE)机器学习方法筛选出5个潜在的生物标志物。通过单细胞分析揭示了关键基因在巨噬细胞中的表达模式。提取肺组织巨噬细胞进行细胞水平验证,观察肺PAH关键生物标志物表达差异。通过外周血单个核细胞(peripheral blood mononuclear cells,PBMC)数据验证关键生物标志物临床意义。利用缺氧方法建立COPD合并PAH小鼠模型,将16只8周龄普通级COPD小鼠(雌雄不限,体质量20~22 g)按随机数字抽样法分为(n=8):缺氧组(氧浓度为10%±0.5%,COPD合并PAH组)和常氧组(COPD组)。使用免疫荧光技术标记关键生物标志物,计算表达情况。**结果** COPD合并PAH筛选出28个DEGs($|Log_2FC|\geq2$,$P<0.05$)。功能富集分析显示,COPD合并PAH的DEGs与主要组织相容性复合体(majorhistocompatibility complex,MHC)Ⅱ和细胞分裂相关,参与溶酶体、氧化磷酸化及细胞周期通路($P<0.05$)。机器学习得到5个潜在的生物标志物(GRN、KLF4、SHTN1、LRP1和GPNMB),进一步的单细胞分析揭示了这些标志物在疾病发展中存在逆向表达模式。从PBMCs中构建的诺模图模型,用于诊断COPD合并PAH[受试者操作曲线下面积(area under the subject curve,AUC)为0.907]。COPD合并PAH组GRN、KLF4、SHTN1、LRP1和GPNMB蛋白和基因显著高表达($P<0.05$)。**结论** 本研究发现了GRN、KLF4、SHTN1、LRP1和GPNMB可作为COPD合并PAH预测和诊断的关键生物标志物,为COPD合并PAH的临床治疗提供了新的见解。

[关键词] 慢性阻塞性肺病合并肺动脉高压;诊断标志物;机器学习;生物信息学分析;肺巨噬细胞
[中图法分类号] R318.04;R544;R563.9 [文献标志码] A

Identification and experimental validation of biomarkers for chronic obstructive pulmonary disease complicated with pulmonary arterial hypertension based on bioinformatics and machine learning

YANG Yan[1], TAO Chunrong[1], ZHU Youjun[1], ZHANG Cong[2], LI Defeng[3] [1]Department of Cardiology, [3]Clinical Medical Research Center, Second Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing; [2]Animal Research Center, College of Basic Medical Sciences, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] **Objective** To identify the key biomarkers for diagnosing chronic obstructive pulmonary disease (COPD) complicated with pulmonary arterial hypertension (PAH) using bioinformatics, and validate

[基金项目] 重庆市自然科学基金博士后科学基金(cstc2021jcyjbshX0071)

[通信作者] 李德峰,E-mail:Ldf201407@tmmu.edu.cn

2097-0927/Copyright © 2025,陆军军医大学(第三军医大学)。这篇开放获取文章遵循CC BY许可协议(https://creativecommons.org/licenses/by/4.0/)

引用格式:杨燕,陶春容,朱佑君,等.基于生物信息学和机器学习鉴定慢性阻塞性肺病合并肺动脉高压标志物及实验验证[J].陆军军医大学学报,2025,47(09):948-958.DOI:10.16016/j.2097-0927.202412012.

Citation:YANG Y,TAO C R,ZHU Y J,*et al*.Identification and experimental validation of biomarkers for chronic obstructive pulmonary disease complicated with pulmonary arterial hypertension based on bioinformatics and machine learning[J].J Army Med Univ,2025,47(09):948-958.DOI:10.16016/j.2097-0927.202412012.

their clinical significance. **Methods** High-throughput sequencing data analysis was employed to identify differentially expressed genes (DEGs) in COPD-PAH. Functional enrichment analysis was then conducted to explore the biological functions of these DEGs. Machine learning methods, including least absolute shrinkage and selection operator (LASSO), random forest (RF), and support vector machine-recursive feature elimination (SVM-RFE), were utilized to screen 5 potential biomarkers. Single-cell analysis was performed to reveal the expression patterns of these key genes in macrophages. The clinical significance of these biomarkers was further validated using peripheral blood mononuclear cells (PBMC) data. A mouse model of COPD-PAH was established using hypoxia exposure. Sixteen mice (either sexes, 8 weeks old, weighing 20~22 g) were randomly divided into a hypoxia group [O_2 (10.0±0.5)%, COPD-PAH, n=8] and a normoxia group (COPD, n=8). Immunofluorescence assay was used to label the key biomarkers, and their expression levels were quantified. **Results** A total of 28 DEGs ($|Log_2FC| \geqslant 2$, $P<0.05$) were identified in COPD-PAH patients. Functional enrichment analysis indicated that DEGs in COPD were primarily associated with major histocompatibility complex (MHC) Ⅱ and cell division, and involved in lysosomes, oxidative phosphorylation, and cell cycle pathways ($P<0.05$). Machine learning identified 5 potential biomarkers (GRN, KLF4, SHTN1, LRP1, and GPNMB), and subsequent single-cell analysis revealed that these markers exhibited reverse expression patterns during disease progression. A nomogram model constructed based on PBMC data yielded an area under the curve (AUC) of 0.907 in diagnosing COPD-PAH. GRN, KLF4, SHTN1, LRP1 and GPNMB were significantly upregulated in the COPD-PAH group ($P<0.05$). **Conclusion** GRN, KLF4, SHTN1, LRP1 and GPNMB are identified as key biomarkers for the prediction and diagnosis of COPD-PAH, which providing new insights for the clinical and treatment of the condition.

[**Key words**] chronic obstructive pulmonary disease with concomitant pulmonary arterial hypertension; diagnostic biomarkers; machine learning; bioinformatics analysis; pulmonary macrophages

Supported by the Postdoctoral Science Foundation of Natural Science Foundation of Chongqing (cstc2021jcyjbshX0071). Corresponding author: LI Defeng, E-mail: Ldf201407@tmmu.edu.cn

慢性阻塞性肺病(chronic obstructive pulmonary disease, COPD)是一种全球性公共卫生挑战,临床特征表现为不完全可逆的气流受限,并伴有慢性全身性炎症[1]。但COPD的影响不仅限于呼吸系统,它还可能导致全身性的健康问题[2]。肺动脉高压(pulmonary arterial hypertension, PAH)是由于肺动脉压力异常升高导致的血流动力学异常状态,其特征包括内皮功能障碍和血管重塑[3]。PAH对全球约1%人口的生活质量产生显著影响,且在≥65岁老年群体中的患病率高达10%[4]。PAH是COPD的严重并发症,当COPD患者合并PAH时,更容易出现心力衰竭等严重临床症状,预后差、死亡率高[5]。目前尽管存在有效的治疗药物,但对于COPD合并PAH的患者长期治疗效果并不理想。因此,亟需早期筛查和干预以降低死亡率并改善患者预后。

单细胞测序技术近年来在生命科学研究中取得了重大突破,使研究人员能够在单细胞水平上探究复杂疾病的细胞异质性及其分子机制[6]。COPD合并PAH作为重要的肺部疾病,严重影响患者的生活质量和生存率。单细胞技术的研究为揭示这两种疾病在细胞类型、基因表达及信号通路方面提供了识别新型生物标志物和治疗靶点的机会。目前,单细胞RNA测序已被用于识别COPD患者肺组织中的特异性细胞亚群及其炎症反应[7];在PAH研究中,单细胞技术帮助解析出肺血管重塑过程中的关键细胞和分子机制[8],这些研究为疾病的早期诊断和个性化治疗提供了重要线索。本研究旨在通过结合COPD合并PAH患者的转录组和单细胞测序数据,在细胞水平上揭示这两种合并疾病的分子机制和疾病发展特征,为寻找潜在的生物标志物和治疗靶点提供新的视角。

1 材料与方法

1.1 主要实验试剂

4%多聚甲醛购自武汉博士德生物工程有限公司,Triton X-100、山羊血清、RIPA裂解液、脱脂奶粉、TRIzol试剂、辣根过氧化物酶标记山羊抗兔和辣根过氧化物酶标记山羊抗小鼠二抗购自上海碧云天生物科技有限公司,RT Master Mix反转录试剂盒和SYBR Green MasterMix购自美国MedChemExpress公司,兔抗GRN抗体购自北京博奥森生物技术有限公司,兔抗KLF4、兔抗LRP1抗体购自武汉爱博泰克生物科技有限公司,鼠抗GPNMB、兔抗SHTN1抗体购自武汉三鹰生物技术有限公司,免疫组化兔二步法试剂盒、小鼠

二步法试剂盒和DAB显色液购自北京中杉金桥生物技术有限公司。

1.2 数据来源

COPD及PAH转录组数据分别来源于基因表达数据库(Gene Expression Omnibus, GEO)数据库(http://www.ncbi.nlm.nih.gov/geo/) GSE212331、GSE217438和GSE33463数据集[9]。单细胞RNA-seq数据分别来源于GSE173896和GSE210248数据集[10-11],见表1。

表1 COPD及PAH转录组数据来源

数据集编号	平台	组织	健康组	疾病组
GSE212331	GPL10558	肺	15	72
GSE217438	GPL2112	心肺	5	5
GSE173896	GPL20795	肺	7	9
GSE210248	GPL20301	肺	3	3
GSE33463	GPL6947	外周血	41	30

1.3 数据预处理

RNA-seq数据进行如下处理:①利用R包将探针ID转换为基因名;②移除对应多个基因的探针;③对具有多个基因名的表达值取中位数作为表达值。单细胞RNA-seq数据作如下处理:①使用CreateSeuratObject建立Seurat对象,设置最小细胞数为3,最小特征数为200,并对基因和细胞进行首次筛选;②使用以下阈值再次筛选细胞:特征数在250~7 500内,计数>500,过表达双阈值为0.15,线粒体基因的比例<5%,红细胞读数的百分比<20%;③选取前2 000个高变异基因作为主成分分析(principal component analysis, PCA)输入。

1.4 差异表达基因分析

使用"线性模型用于微阵列数据"(linear models for microarray data analysis,limma)R软件包,基于批量RNA测序数据[12],识别病例组和正常样本之间的差异表达基因(differentially expressed genes, DEGs)。使用Benjamini-Hochberg校正P值。采用$P<0.05$和$|Log_2FC|\geq 2$作为阈值来识别DEGs。通过VennDiagram包获取COPA和PAH共同的DEGs。clusterProfiler R包用于富集分析[13]。

1.5 生物标志物与潜在靶点筛选

本研究采用3种机器学习方法:最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)、随机森林(random forest, RF)、支持向量机递归特征消除(support vector machine-recursive feature elimination, SVM-RFE)对得到的生物标志物进行筛选鉴定。LASSO回归分析用于统计模型中的变量选择和正则化。RF预测连续变量和生成稳健预测。SVM-RFE用于最大边距特征选择。

1.6 单细胞分析

使用FindClusters算法识别不同的细胞亚群。统一流形逼近与投影(uniform manifold approximation and projection, UMAP)用于PCA后降维,使用Harmony进行去批次处理。使用ACT(http://xteam.xbio.top/ACT/index.jsp)对细胞类型进行注释。已有的研究表明,免疫反应在COPD和PAH发展中有重要作用[14-15],因此,使用monocle2进行细胞轨迹及拟时分析,降维方法为"DDRtree"。

1.7 诺模图的构建与验证

为了评估模型的适用性能力,使用外周血数据集GSE33463进一步确认了关键生物标志物的适用性。受试者操作曲线下面积(area under the subject curve, AUC)被用来评估模型的区分能力。此外,校准曲线用于评估校准程度,决策曲线分析(decision curve analysis, DCA)对模型临床的有效性进行评价。

1.8 小鼠模型构建

利用缺氧方法建立COPD合并PAH小鼠模型。16只8周龄普通级COPD组小鼠(雌雄不限,体质量20~22 g,购自北京维通利华实验动物技术有限公司),按简单随机抽样分为(n=8):缺氧组[氧浓度为(10.0±0.5)%,COPD合并PAH组]、常氧组(COPD组)[16]。每周更换1次食物、水和垫料。每天光照与黑暗周期为12 h/12 h,持续14 d。14 d后,给予安乐死,收集小鼠肺组织,提取小鼠肺组织巨噬细胞[17]。动物实验经陆军军医大学实验动物福利伦理委员会审核(AMUWEC20245279),符合动物伦理和动物福利要求。

1.9 小鼠肺组织形态鉴定

取小鼠右肺叶及后叶病理学检查。在4%多聚甲醛溶液中固定,固定后的组织块经过标准的脱水、透明、浸蜡和石蜡包埋处理,将石蜡包埋的组织块切成4 μm厚的连续切片,使用苏木精-伊红染色(hematoxylin-eosin staining,HE)进行染色。采用测量平均线性截距和平均肺泡数评估COPD程度。

1.10 免疫组化实验

将组织切片或提取的肺部巨噬细胞用4%多聚甲醛固定15 min,随后用含0.2% Triton X-100的PBS进行渗透处理10 min。在室温下,样品用含5%正常山羊血清的PBS封闭1 h。样品在封闭液中与一抗(兔抗p-GRN、兔抗p-KLF4、兔抗p-LRP1、鼠抗p-GPNMB、兔抗p-SHTN1,均为1∶200)孵育,4 ℃过夜。用PBS洗涤后,样品与相应的兔/鼠二抗(1∶200)在室温下孵育2 h,PBS洗涤后,使用DAB显色液进行显色并在显微镜先观察拍照。图像采集使用Easyscaner软件,并使用MoticDSM V3.0.6进行了图像分析。此外,使用Image J软件计算了平均光密度(average optical density,AOD)。

1.11 Western blot检测

提取的肺巨噬细胞在4 ℃使用RIPA缓冲液进行匀浆和裂解。蛋白质通过10%SDS-PAGE分离,并将

分离后的蛋白质电转移至硝酸纤维素膜,使用5%脱脂奶封闭,并在4 ℃下与一抗孵育过夜。用TBST溶液洗涤3次,在室温下与二抗孵育1 h。经过进一步洗涤后,使用BIO-RAD的ChemiDoc MP成像系统进行蛋白质检测。

1.12　qRT-PCR检测

利用TRIzol法提取细胞总RNA,Nanodrop2000微量分光光度计检测总RNA浓度及纯度,取吸光度$D(260)/D(280)$在1.8~2.0的总RNA样本,置-80 ℃保存。取2 μg总RNA样本,去除基因组DNA并逆转录成cDNA,样本置-80 ℃保存。PCR引物由擎科生物公司合成,引物序列见表2。利用逆转录试剂盒逆转录为cDNA,随后按照SYBR Green MasterMix试剂盒说明书配置qRT-PCR反应体系,采用实时荧光定量PCR仪进行扩增和检测,结果采用$2^{-\Delta\Delta Ct}$计算分析目的基因的mRNA相对表达量。

1.13　统计学分析

实验数据利用R 4.3.3软件进行统计分析。采用Wilcoxon符号秩检验分析两组样本间的差异,多组间比较采用Kruskal-Wallis H检验,$P<0.05$表示差异具有统计学意义。

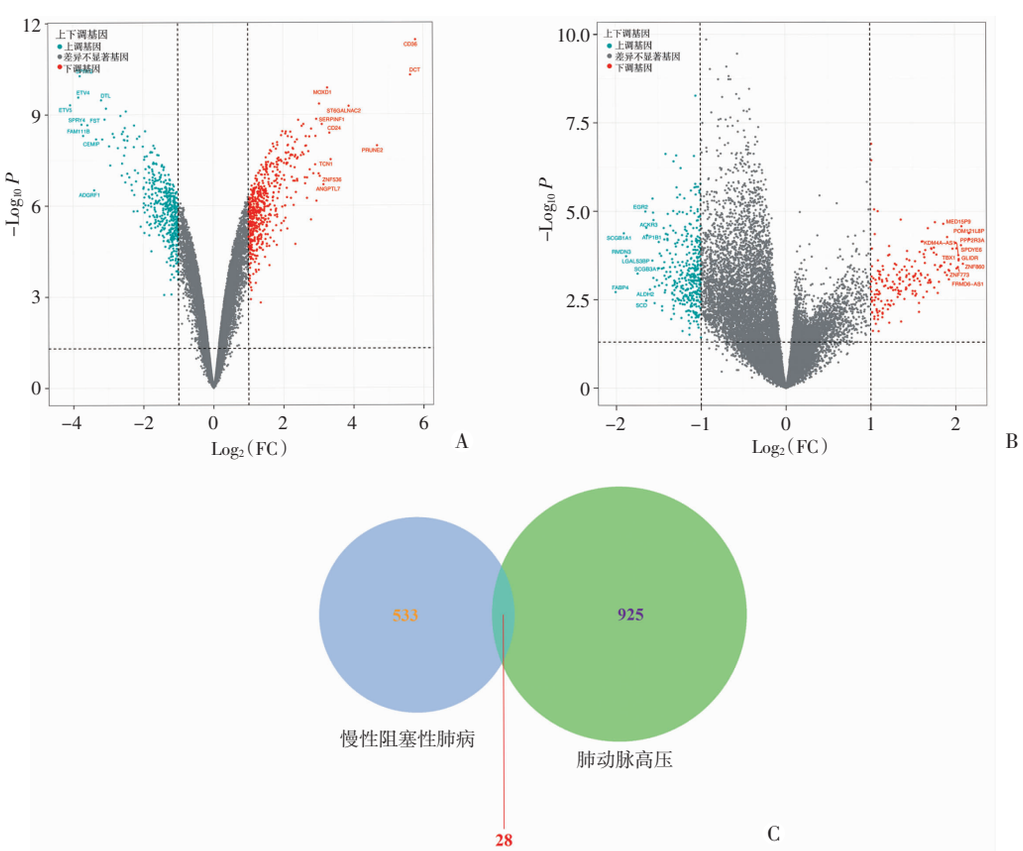
表2　目的基因引物序列

基因	引物序列(5′→3′)
GRN	上游:TCCCCTTCTGGACAAATGGC
	下游:TAAAGATGCAGGAGTGGCCG
KLF4	上游:GGGCCCAATTACCCATCCTT
	下游:GGCATGAGCTCTTGGTAATGG
SHTN1	上游:TTCCCTTTTATTCCCGGCCC
	下游:TCGCCTATTGCTTGCTCCTT
LRP1	上游:TGACATGGATGGGCAGAACC
	下游:CCCCCACAAAACCTCAGACA
GPNMB	上游:CGGCCAAAGCCATCATAACG
	下游:GGTTCGTCTCACAGTCAGCA

2　结果

2.1　鉴定差异表达基因

使用高通量基因表达数据集GSE212331对COPD患者组与健康对照组进行差异表达分析时,发现561个显著差异表达的基因,其中192个基因在COPD患者组表现为上调,而369个基因则表现为下调(图1A)。基因表达数据集GSE217438对PAH患者组与健康对照组进行对比分析,结果表明存在953个差异表达基因,其中520个基因上调,433个基因下调(图1B)。进



A:COPD差异表达基因火山图;B:PAH差异表达火山图;C:COPD和PAH差异表达基因韦恩图

图1　COPD和PAH的基因表达差异分析

一步利用这2个数据集差异表达基因取交集得到28个共同的差异表达基因(图1C)。COPD合并PAH的GO富集结果表明，共同差异表达基因(common differentially expressed genes，Co-DEGs)与主要组织相容性复合体(majorhistocompatibility complex，MHC)Ⅱ和细胞分裂相关(图2A、B)，KEGG通路分析显示COPD合并PAH差异表达基因主要参与溶酶体、氧化磷酸化及细胞周期等通路(图2C、D)。通过蛋白质-蛋白质相互作用网络(protein-protein interaction networks，PPI)分析，进一步揭示了*SHTN1*、*GPNMB*、*GRN*和*KLF4*等基因在网络中占有核心地位(图2E)，在COPD合并PAH病理过程中可能具有重要功能。

2.2 生物标志物的筛选与验证

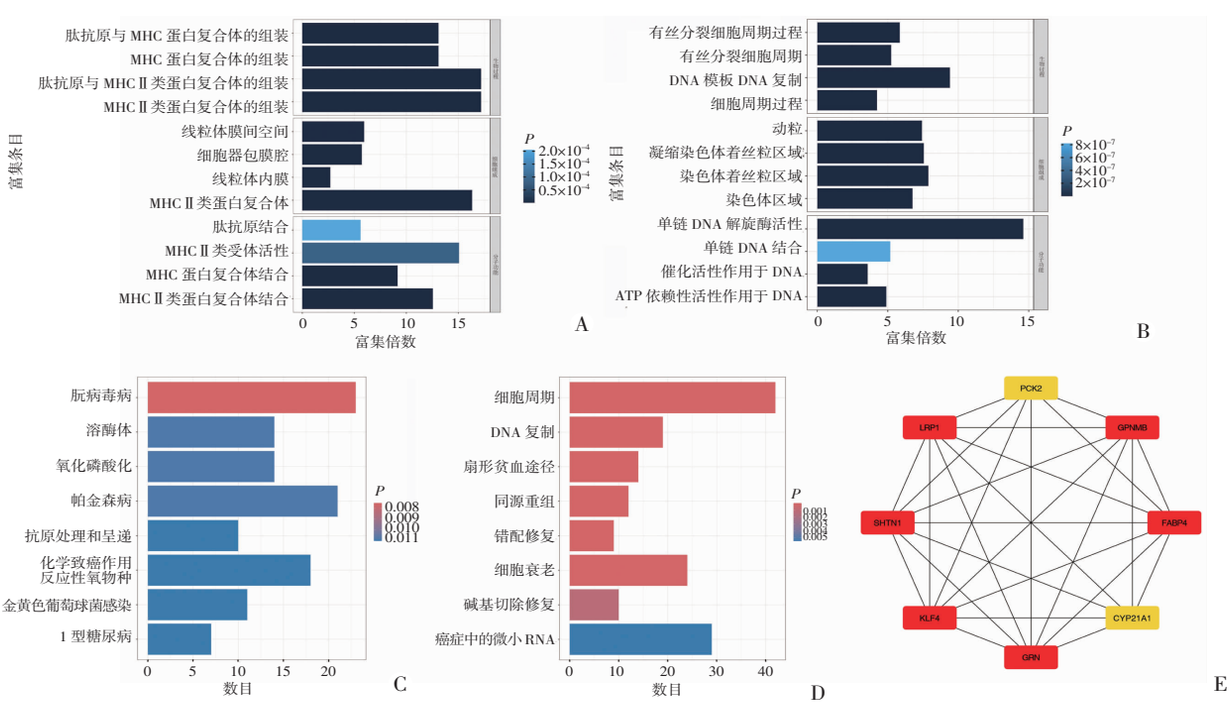
通过LASSO逻辑回归从28个Co-DEGs中挑选出变量，识别出7个在LASSO回归模型中具有非零系数的潜在预测因子(图3A)。RF方法被用来进一步验证Co-DEGs，并根据重要性排名选择前9个变量(图3B)。使用SVM-RFE选择变量，当特征变量数量为5时，关联的误差错误率达到了的最小值0.005(图3C)。最后，通过重要性得分，将排名前8的基因列出(图3D)。对3种机器学习结果取交集确定了5个关键生物标志物(*GRN*、*KLF4*、*SHTN1*、*LRP1*和*GPNMB*)。

2.3 单细胞分析

为了进一步研究COPD合并PAH在细胞水平上的发展机制，将GSE173896和GSE210248数据集纳入研究。COPD及PAH 2个数据集经过质量控制，分别有24 469和22 552个细胞被保留用于进一步分析(图4A、B)。观察到5个关键生物标志物在不同细胞中表达模式存在显著差异，然而在巨噬细胞中均有表达(图4C、D)。我们采用monocle2 R包对2组数据中的单核/巨噬/树突状细胞进行拟时序分析。结果表明，在COPD组中，关键基因在单核细胞向树突状细胞演化过程中表达先降低再升高(图5A)。而在PAH组中，关键基因在单核细胞向树突状细胞演化过程中表达呈相反的趋势(图5B)。

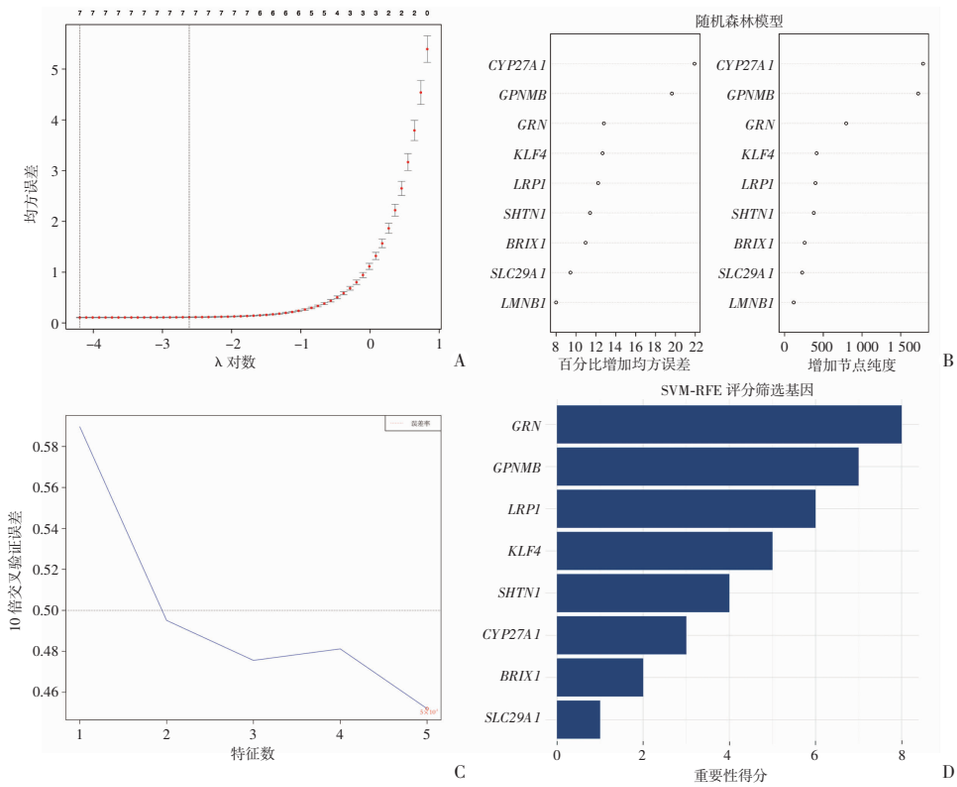
2.4 利用PBMC验证诊断模型

以上结果表明，生物标志物在COPD合并PAH中显著差异表达，且具有枢纽作用。基于这些发现，本研究利用外周血测序数据开发并验证了特征基因的诊断模型。通过对COPD合并PAH患者PBMC的测序数据分析，建立基于*GRN*、*KLF4*、*SHTN1*、*LRP1*和*GPNMB*的生物标志物列线图模型(图6A)。此外，DCA曲线显示，当列线图模型的阈值概率在0~0.96时，净收益>0，表明该模型在COPD合并PAH预测中具有更高的临床价值(图6B)。列线图模型的净收益曲线面积也超过了单一生物标志物的曲线面积(图6B)，进一步表明该模型在临床诊断中具有更高的诊断价值。校准曲线分析显示，模型的逻辑回归线与最优曲线一致，表



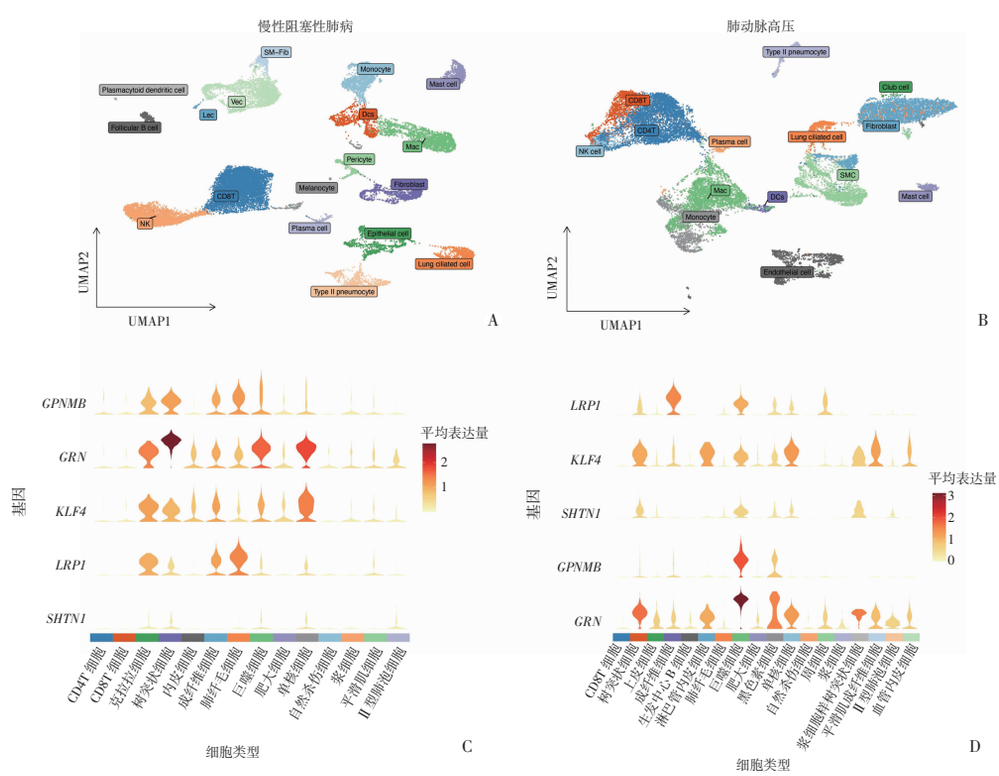
A：COPD差异基因GO富集分析；B：PAH差异基因GO富集分析；C：COPD差异基因KEGG富集结果；D：PAH差异基因KEGG富集结果；E：COPD合并PAH差异基因PPI分析结果图

图2 差异基因富集分析图



A:使用LASSO逻辑回归方法生成的交叉验证曲线;B:确定RF模型的最佳树后筛选出的前9个基因;C:显示变量选择错误率的曲线;D:前8个预测得到的关键基因

图3　机器学习识别COPD合并PAH诊断生物标记的应用



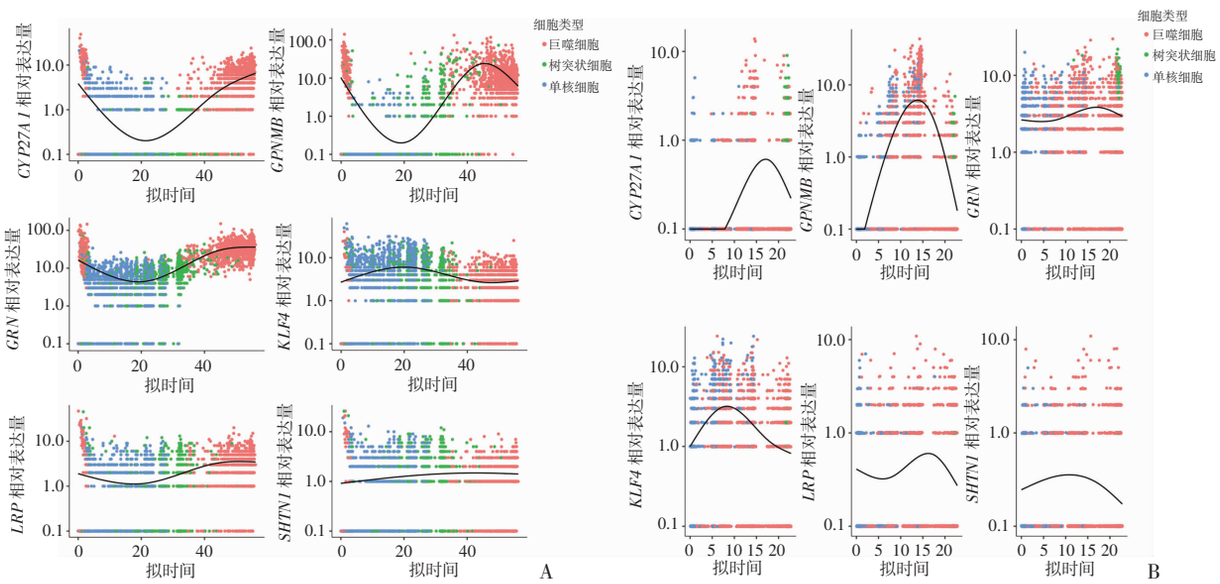
A:COPD细胞类型;B: PAH细胞类型; C:COPD组关键生物标志物在不同细胞类型中的表达情况;D:PAH组关键生物标志物在不同细胞类型中的表达情况

图4　COPD合并PAH单细胞分析

明模型的预测结果精确(图6C)。最后，通过ROC曲线分析评估了这些生物标志物及其列线图模型在疾病诊断中的敏感性和特异性，结果显示，GRN(AUC=0.935)、KLF4(AUC=0.938)、SHTN1(AUC=0.952)、LRP1(AUC=0.900)和GPNMB(AUC=0.864)均具有一定的诊断价值，而列线图模型(AUC=0.907)则表现出更高的诊断效率(图6D)。

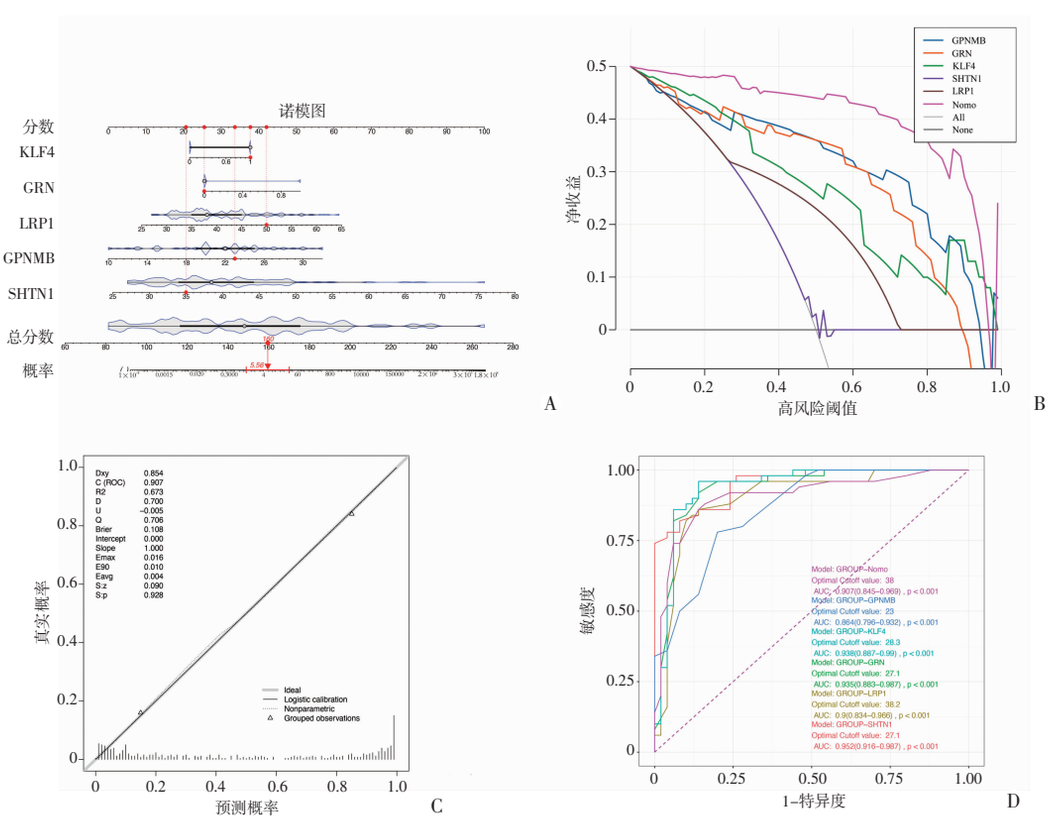
2.5　GRN、KLF4、SHTN1、LRP1和GPNMB在COPD合并PAH小鼠中表达显著增加

为了验证5个关键生物标志物结果的可靠性，通



A：关键生物标志物在COPD中的拟时表达谱；B：关键生物标志物在PAH中的拟时表达谱

图5　COPD合并PAH细胞拟时序分析



A：基于关键生物标志物的列线图模型；B：GSE33463决策分析曲线；C：GSE33463校准曲线；D：GSE33463 ROC曲线

图6　使用外周血数据集GSE33463开发和评估临床诊断模型

过qRT-PCR验证了标志物在肺巨噬细胞中的表达水平。结果显示,*GRN*、*KLF4*、*SHTN1*、*LRP1*和*GPNMB*肺巨噬细胞中表达显著升高($P<0.05$,图7A)。为了进一步验证5个关键生物标志物在肺巨噬细胞中的蛋白表达水平,进行了Western blot检测。结果显示,5个关键生物标志物蛋白的表达在肺巨噬细胞中显著增加($P<0.05$,图7B)。

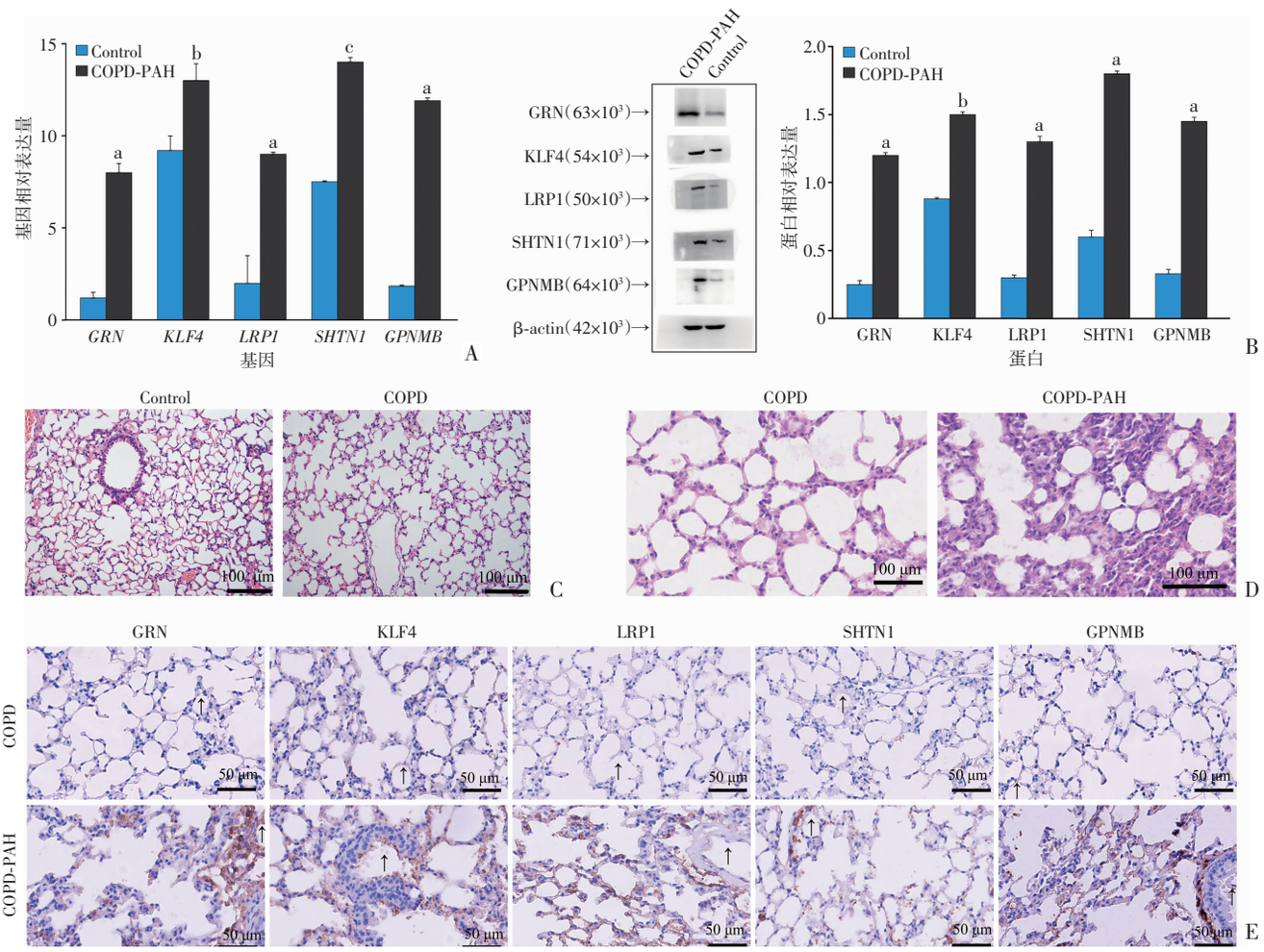
与野生型对照组相比,COPD组小鼠肺泡腔大小不一,部分肺泡壁破裂、融合,肺泡间隔增厚(图7C)。通过将COPD小鼠置于缺氧环境中诱导的COPD合并PAH的小鼠模型,与处于正常氧条件下的COPD组小鼠相比,缺氧的COPD合并PAH组肺动脉血管壁显著增厚(图7D)。免疫组化结果显示,COPD合并PAH小鼠肺组织中的GRN、KLF4、SHTN1、LRP1和GPNMB表达水平显著高于常氧组小鼠(图7E)。

3　讨论

3.1　COPD合并PAH生物标志物差异表达与功能富集

本研究揭示了COPD合并PAH的DEGs,并通过生物信息学方法进一步确定了这些基因在疾病中的生物学作用。已有研究表明,COPD和PAH病理进程涉及多个信号通路调控[18-19]。在对COPD患者和健康对照组的比较分析中,本研究识别出了561个DEGs,在PAH患者中发现了953个DEGs。筛选出28个COPD合并PAH的DEGs,

这些基因可能在两种疾病中扮演关键角色。这些基因除了参与免疫调节和血管重塑外[20-21],还参与



a:$P<0.001$,b:$P<0.05$,c:$P<0.01$,与Control组比较

A:关键基因*GRN*、*KLF4*、*SHTN1*、*LRP1*和*GPNMB*在肺巨噬细胞中的mRNA表达水平;B:常氧组和COPD合并PAH组肺巨噬细胞中的5个关键生物标志物蛋白水平;C:野生型对照组和COPD组肺组织HE染色结果;D:COPD组和COPD合并PAH组肺组织HE染色结果;E:COPD合并PAH组和COPD组小鼠肺组织免疫组化染色图　↑:示肺动脉

图7　5个关键生物标志物在COPD合并PAH组和常氧组小鼠中的表达情况

对外界刺激反应的调控及细胞-基质连接等生物过程。GO和KEGG通路分析证实，COPD相关DEGs主要参与免疫相关生物过程，如肽抗原呈递，而PAH相关DEGs则多与细胞周期调控和DNA修复过程有关，这与之前的研究结果相似[22-23]。

3.2 关键诊断标志物的筛选和生物学功能

本研究将机器学习应用于COPD与PAH共同生物标志物的筛选，识别出了*GRN*、*KLF4*等关键基因作为潜在的诊断标志物。之前的研究较少将机器学习技术应用于COPD合并PAH的联合研究中，多采用LASSO、RF等算法对单一一种疾病生物标志物进行筛选和验证[24-25]。通过结合LASSO、RF和SVM-RFE等多种机器学习方法，本研究确认了5个具有显著预测能力的诊断标志物。此外，通过单细胞分析进一步揭示了单核细胞经巨噬细胞向树突细胞分化过程中关键诊断标志物的表达模式。COPD组中关键标志物的表达先降低后升高，而PAH组呈现相反趋势，这些动态变化提示了不同时期细胞的状态和功能，强调了在疾病发展中免疫细胞塑造组织微环境的关键作用[26]。

本研究筛选的5个标志物中，*GRN*基因的产物作为一种分泌蛋白，是溶酶体功能的关键调节因子，并参与炎症、伤口愈合和细胞增殖等过程[27-29]。*KLF4*编码一种属于Kruppel转录因子家族的蛋白质，参与上皮细胞分化，有助于调节p53/TP53的转录[30]。*SHTN1*编码一种包含SH3和PTK结构域的蛋白质，这些结构域参与PPI和信号传导[31]。*SHTN1*在细胞骨架重构、细胞迁移和黏附以及细胞信号传导中发挥重要作用。*LRP1*编码低密度脂蛋白受体1。编码的前蛋白经蛋白酶加工生成515×10^3和85×10^3的亚基，形成成熟的内吞受体[32]，参与内吞作用和凋亡细胞的吞噬作用[33]。*GPNMB*基因的产物是一种Ⅰ型跨膜糖蛋白，在骨髓单核细胞中由G-CSF/CSF3和M-CSF/CSF1上调，因此GPNMB上调可能与单核细胞的分化有关[34]。

3.3 诊断标志物的验证及临床意义

通过qRT-PCR和Western blot实验验证，关键标志物在肺组织巨噬细胞中的表达与疾病状态具有显著相关性。在COPD合并PAH模型鼠中，肺组织的IHC实验也验证了关键生物标志物表达显著升高，此外，外周血的数据也表明，这5个关键生物标志物具有较高的临床预测效能。尽管实验验证表明GRN、KLF4、LRP1、GPNMB和SHTN1在病变细胞及组织中显著上调，但仍需要更大规模的临床研究来确认这些发现。

综上所述，本研究不仅加深了对这两种复杂疾病的理解，也为个性化治疗和早期诊断提供了新的可能性。未来的研究需要进一步验证这些标志物在更大病患群体中的应用潜力，并探讨尚未明示的分子机制及可能的治疗靶点。

利益冲突 本文所有作者声明不存在利益冲突

参考文献：

[1] CELLI B, FABBRI L, CRINER G, *et al*. Definition and nomenclature of chronic obstructive pulmonary disease: time for its revision[J]. Am J Respir Crit Care Med, 2022, 206(11): 1317-1325. DOI:10.1164/rccm.202204-0671PP.

[2] REHMAN A U, AHMAD HASSALI M A, MUHAMMAD S A, *et al*. The economic burden of chronic obstructive pulmonary disease (COPD) in the USA, Europe, and Asia: results from a systematic review of the literature [J]. Expert Rev Pharmacoecon Outcomes Res, 2020, 20(6): 661-672. DOI: 10.1080/14737167.2020.1678385.

[3] 杨媛华.《中国肺动脉高压诊断与治疗指南(2021版)》解读：肺动脉高压的诊断[J]. 中国实用内科杂志, 2021, 41(8): 696-699. DOI: 10.19538/j.nk2021080107.
YANG Y H. Interpretation of the guidelines for the diagnosis and treatment of pulmonary hypertension in China (2021 edition): diagnosis of pulmonary hypertension [J]. Chin J Pract Intern Med, 2021, 41(8): 696-699. DOI: 10.19538/j.nk2021080107.

[4] 万钧, 翟振国. 肺动脉高压临床诊治和管理中需要关注的热点问题：基于《2022 ESC/ERS肺动脉高压诊治指南》与《中国肺动脉高压诊断与治疗指南(2021版)》的比较与解读[J]. 中国全科医学, 2023, 26(3): 255-261. DOI: 10.12114/j.issn.1007-9572.2022.0692.
WAN J, ZHAI Z G. Hot issues in clinical diagnosis, treatment and management of pulmonary hypertension: Comparison and interpretation based on "2022 ESC/ERS Guidelines for Diagnosis and Treatment of Pulmonary Hypertension" and "China Guidelines for Diagnosis and Treatment of Pulmonary Hypertension (2021 Edition)" [J]. Chin Gen Pract, 2023, 26(3): 255-261. DOI: 10.12114/j.issn.1007-9572.2022.0692.

[5] 徐明艳, 韩校鹏, 刘英丽, 等. 慢性阻塞性肺疾病急性加重期并发肺动脉高压的危险因素[J]. 实用医学杂志, 2022, 38(19): 2467-2471. DOI: 10.3969/j.issn.1006-5725.2022.19.017.
XU M Y, HAN X P, LIU Y L, *et al*. Risk factors of pulmonary hypertension in acute exacerbation of chronic obstructive pulmonary disease[J]. J Pract Med, 2022, 38(19): 2467-2471. DOI: 10.3969/j.issn.1006-5725.2022.19.017.

[6] 张凌瑞, 胡龙飞, 黄万翔, 等. 基于单细胞转录组数据的

疾病表型预测研究进展[J/OL]. 生物信息学, 2024: 1-11. (2024-06-21). https://kns.cnki.net/kcms/detail/23.1513.Q.20240621.1429.002.html.
ZHANG L R, HU L F, HUANG W X, *et al*. Research progress of disease phenotype prediction based on single cell transcriptome data[J/OL]. Chin J Bioinform, 2024: 1-11. (2024-06-21). https://kns.cnki.net/kcms/detail/23.1513.Q.20240621.1429.002.html.

[7] 杨泽华, 丁毅鹏. 单细胞测序技术在慢性阻塞性肺疾病研究中的应用进展[J]. 海南医学院学报, 2024, 30(8): 617-622. DOI: 10.13210/j.cnki.jhmu.20230816.001.
YANG Z H, DING Y P. Advances in the application of single-cell sequencing technology in the study of chronic obstructive pulmonary disease[J]. J Hainan Med Univ, 2024, 30(8): 617-622. DOI: 10.13210/j.cnki.jhmu.20230816.001.

[8] 赵飞龙. 基于多组学数据的肺动脉高压相关生物标志物的挖掘[D]. 长沙: 中南大学, 2023. DOI: 10.27661/d.cnki.gzhnu.2023.004823.
ZHAO F L. Mining biomarkers related to pulmonary hypertension based on multiomics data[D]. Changsha: Central South University, 2023. DOI: 10.27661/d.cnki.gzhnu.2023.004823.

[9] NEGEWO N A, GIBSON P G, SIMPSON J L, *et al*. Severity of lung function impairment drives transcriptional phenotypes of COPD and relates to immune and metabolic processes[J]. Int J Chron Obstruct Pulmon Dis, 2023, 18: 273-287. DOI: 10.2147/COPD.S388297.

[10] WATANABE N, FUJITA Y, NAKAYAMA J, *et al*. Anomalous epithelial variations and ectopic inflammatory response in chronic obstructive pulmonary disease[J]. Am J Respir Cell Mol Biol, 2022, 67(6): 708-719. DOI: 10.1165/rcmb.2021-0555OC.

[11] CRNKOVIC S, VALZANO F, FLIEßER E, *et al*. Single-cell transcriptomics reveals skewed cellular communication and phenotypic shift in pulmonary artery remodeling[J]. JCI Insight, 2022, 7(20): e153471. DOI: 10.1172/jci.insight.153471.

[12] RITCHIE M E, PHIPSON B, WU D, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Res, 2015, 43(7): e47. DOI: 10.1093/nar/gkv007.

[13] YU G C, WANG L G, HAN Y Y, *et al*. Cluster Profiler: an R package for comparing biological themes among gene clusters[J]. OMICS, 2012, 16(5): 284-287. DOI: 10.1089/omi.2011.0118.

[14] 王亮亮, 宋春雪, 丁琦, 等. T淋巴细胞免疫在COPD肺气肿发病机制中的作用[J]. 中外医学研究, 2024, 22(16): 179-184. DOI: 10.14033/j.cnki.cfmr.2024.16.045.
WANG L L, SONG C X, DING Q, *et al*. Role of T lymphocyte immunity in the pathogenesis of COPD emphysema[J]. Chin Foreign Med Res, 2024, 22(16): 179-184. DOI: 10.14033/j.cnki.cfmr.2024.16.045.

[15] 缪美琪, 冯晓岚, 尹小龙, 等. 免疫炎症在肺动脉高压中的研究进展[J]. 临床内科杂志, 2022, 39(3): 148-151. DOI: 10.3969/j.issn.1001-9057.2022.03.002.
MIAO M Q, FENG X L, YIN X L, *et al*. Research progress of immune inflammation in pulmonary hypertension[J]. J Clin Intern Med, 2022, 39(3): 148-151. DOI: 10.3969/j.issn.1001-9057.2022.03.002.

[16] LIU J, WANG W, WANG L, *et al*. IL-33 initiates vascular remodelling in hypoxic pulmonary hypertension by up-regulating HIF-1α and VEGF expression in vascular endothelial cells[J]. EBioMedicine, 2018, 33: 196-210. DOI: 10.1016/j.ebiom.2018.06.003.

[17] 杨顺心, 周远群, 向鑫明, 等. ATP6V1A缺失导致脓毒症心脏驻留巨噬细胞清除能力损伤[J]. 陆军军医大学学报, 2024, 46(23): 2599-2607. DOI: 10.16016/j.2097-0927.202406060.
YANG S X, ZHOU Y Q, XIANG X M, *et al*. ATP6V1A deletion leads to impaired clearance of septic cardiac resident macrophages[J]. J Army Med Univ, 2024, 46(23): 2599-2607. DOI: 10.16016/j.2097-0927.202406060.

[18] 吴磊, 刘日慧, 彭悠, 等. 慢性阻塞性肺病诊断生物标志物及其靶向中药预测[J/OL]. 中国医院药学杂志, 2024: 1-11. (2024-07-30). https://kns.cnki.net/kcms/detail/42.1204.r.20240729.1615.004.html.
WU L, LIU R H, PENG Y, *et al*. Diagnostic biomarkers of chronic obstructive pulmonary disease and prediction of targeted traditional Chinese medicine[J/OL]. Chin J Hosp Pharm, 2024: 1-11. (2024-07-30). https://kns.cnki.net/kcms/detail/42.1204.r.20240729.1615.004.html.

[19] 彭威, 张泽盈, 肖云彬. 生物信息学在肺动脉高压中的研究进展[J]. 中国当代儿科杂志, 2024, 26(4): 425-431. DOI: 10.7499/j.issn.1008-8830.2310076
PENG W, ZHANG Z Y, XIAO Y B. Research progress on bioinformatics in pulmonary arterial hypertension[J]. Chin J Contemp Pediatr, 2024, 26(4): 425-431. DOI: 10.7499/j.issn.1008-8830.2310076

[20] WEI X L, XU H, KUFE D. Human mucin 1 oncoprotein represses transcription of the p53 tumor suppressor gene[J]. Cancer Res, 2007, 67(4): 1853-1858. DOI: 10.1158/0008-5472.CAN-06-3063.

[21] MAY P, HERZ J. LDL receptor-related proteins in neurodevelopment[J]. Traffic, 2003, 4(5): 291-301. DOI: 10.1034/j.1600-0854.2003.00086_4_5.x.

[22] 张梅, 马燕, 康淑琴, 等. COPD患者高分辨率CT分型与

气道炎症、免疫功能和临床疗效的关系研究[J]. 现代生物医学进展, 2024, 24(16): 3123-3127. DOI: 10.13241/j.cnki.pmb.2024.16.024.
ZHANG M, MA Y, KANG S Q, *et al*. Study on the relationship between high-resolution CT typing and airway inflammation, immune function, and clinical efficacy in COPD patients[J]. Prog Mod Biomed, 2024, 24(16): 3123-3127. DOI: 10.13241/j.cnki.pmb.2024.16.024.

[23] 陈倩冰, 李逸轩, 刘怡曼, 等. 肺动脉高压血管重塑过程中的巨噬细胞核心作用机制研究探讨[J]. 天津中医药, 2024, 41(10): 1325-1334.
CHEN Q B, LI Y X, LIU Y M, *et al*. Study on the mechanism of macrophages promoting pulmonary hypertension by affecting pulmonary vascular remodeling[J]. Tianjin J Tradit Chin Med, 2024, 41(10): 1325-1334.

[24] 孟小暄, 张永媚, 王昕宇, 等. 基于随机森林算法的COPD病人功能障碍的影响因素[J]. 护理研究, 2024, 38(5): 788-795.
MENG X X, ZHANG Y M, WANG X Y, *et al*. Influencing factors of dysfunction in patients with COPD based on random forest algorithm[J]. Chin Nurs Res, 2024, 38(5): 788-795.

[25] 吴洁, 马虹. 基于生物信息学对肺动脉高压差异基因的筛选研究[J]. 中国社区医师, 2023, 39(23): 59-61. DOI: 10.3969/j.issn.1007-614x.2023.23.020.
WU J, MA H. Differential gene screening study of pulmonary arterial hypertension based on bioinformatics[J]. Chin Community Dr, 2023, 39(23): 59-61. DOI: 10.3969/j.issn.1007-614x.2023.23.020.

[26] 陈建军. 炎症因子和纤维蛋白原及免疫细胞在COPD病情监测和判断预后中的价值[J]. 中国卫生工程学, 2019, 18(2): 256-258. DOI: 10.19937/j.issn.1671-4199.2019.02.034.
CHEN J J. The value of inflammatory factors, fibrinogen and immune cells in COPD disease monitoring and prognosis judgment[J]. Chin J Public Health Eng, 2019, 18(2): 256-258. DOI: 10.19937/j.issn.1671-4199.2019.02.034.

[27] ZHU J, NATHAN C, JIN W W, *et al*. Conversion of proepithelin to epithelins roles of SLPI and elastase in host defense and wound repair[J]. Cell, 2002, 111(6): 867-878. DOI:10.1016/S0092-8674(02)01141-8.

[28] BEEL S, MOISSE M, DAMME M, *et al*. Progranulin functions as a cathepsin D chaperone to stimulate axonal outgrowth *in vivo*[J]. Hum Mol Genet, 2017, 26(15): 2850-2863. DOI:10.1093/hmg/ddx162.

[29] ZHOU X L, SUN L R, BRACKO O, *et al*. Impaired prosaposin lysosomal trafficking in frontotemporal lobar degeneration due to progranulin mutations[J]. Nat Commun, 2017, 8: 15277. DOI:10.1038/ncomms15277.

[30] YET S F, MCA'NULTY M M, FOLTA S C, *et al*. Human EZF, a Krüppel-like zinc finger protein, is expressed in vascular endothelial cells and contains transcriptional activation and repression domains[J]. J Biol Chem, 1998, 273(2): 1026-1031. DOI:10.1074/jbc.273.2.1026.

[31] TORIYAMA M, SHIMADA T, KIM K B, *et al*. Shootin1: a protein involved in the organization of an asymmetric signal for neuronal polarization[J]. J Cell Biol, 2006, 175(1): 147-157. DOI:10.1083/jcb.200604160.

[32] WILLNOW T E, MOEHRING J M, INOCENCIO N M, *et al*. The low-density-lipoprotein receptor-related protein (LRP) is processed by furin *in vivo* and *in vitro*[J]. Biochem J, 1996, 313 (Pt 1)(Pt 1): 71-76. DOI:10.1042/bj3130071.

[33] MAY P, KRISHNA REDDY Y, HERZ J. Proteolytic processing of low density lipoprotein receptor-related protein mediates regulated release of its intracellular domain[J]. J Biol Chem, 2002, 277(21): 18736-18743. DOI: 10.1074/jbc.M201979200.

[34] BANDARI P S, QIAN J, YEHIA G, *et al*. Hematopoietic growth factor inducible neurokinin-1 type: a transmembrane protein that is similar to neurokinin 1 interacts with substance P[J]. Regul Pept, 2003, 111(1/2/3): 169-178. DOI:10.1016/s0167-0115(02)00288-4.

(收稿:2024-12-04;修回:2024-12-22)
(编辑 严 涵)