• 数据挖掘与循证医学 •

特发性肺纤维化的诊断生物标志物发现与中药靶向预测：基于孟德尔随机化与机器学习的多组学研究

罗　成[1, 2, 3]，叶远航[4]，宁　博[5]，李家劼[6]，谭钧文[1, 2]，王　飞[2]，柯　佳[7, 8, 9, 10]，覃琬婷[1, 2*]

1. 成都中医药大学，四川 成都　610032
2. 成都中医药大学附属医院，四川 成都　610072
3. 中国中医药联合研究生院，江苏 苏州　215105
4. 彭州市第三人民医院，四川 成都　611931
5. 广州中医药大学第二临床医学院，广东 广州　510006
6. 云南中医药大学，云南 昆明　650500
7. 湖北省中医院，湖北 武汉　430061
8. 中医肝肾研究及应用湖北省重点实验室，湖北中医药大学附属医院，湖北 武汉　430061
9. 湖北时珍实验室，湖北 武汉　430060
10. 湖北中医药研究院，湖北 武汉　430074

摘　要：目的　利用生物信息学、孟德尔随机化（Mendelian randomization，MR）和机器学习分析探索特发性肺纤维化（idiopathic pulmonary fibrosis，IPF）的潜在靶点，并初步预测可能的相关中药。**方法**　从 GEO 获得 IPF 微阵列数据集，并鉴定差异表达基因（differentially express genes，DEGs）。基于表达数量性状基因座（expression quantitative trait loci，eQTL）数据和全基因组关联研究（genome-wide association study，GWAS）数据，采用 MR 分析筛选与 IPF 相关的基因。将 MR 分析得出的风险基因与 DEGs 取交集，筛选出 IPF 相关的核心基因。利用功能富集分析、基因集富集分析（gene set enrichment analysis，GSEA）、免疫细胞浸润分析以及单细胞 RNA 测序进行评估。应用机器学习算法筛选最优诊断特征基因。使用独立的 GEO 队列进行差异表达验证及受试者工作特征（receiver operating characteristic，ROC）分析。此外，基于数据库挖掘与分子对接对潜在干预中药进行预测分析。**结果**　共识别出 916 个与 IPF 相关的差异表达基因。与 224 个 MR 风险基因取交集后，得到 7 个关键基因：IRF7、TTC32、IFI6、ISG15（风险基因）及 ZNF204P、ISOC1、CTSK（保护基因）。这些基因主要富集于干扰素-β 产生、视黄酸诱导基因-1（retinoic acid-inducible gene-1，RIG-1）样受体信号通路、Toll 样受体信号通路以及 I 型干扰素信号通路。免疫浸润分析显示，IPF 患者组织中 M1/M0 巨噬细胞及静息肥大细胞减少，而活化肥大细胞增加。单细胞 RNA 测序揭示了这些基因在上皮细胞亚群中的特异性表达模式。机器学习算法确定 ZNF204P 和 IRF7 为最优诊断基因。在数据集验证中证实了这 2 基因在 IPF 中存在显著差异表达，并具有较高的诊断准确性。预测到与关键基因相关的潜在中药 385 味，中药四气以寒、温、平为主，五味以苦、甘、辛为主，归经以肝、肺、胃、脾、肾经为主，分类以清热药、补虚药、活血化瘀药、解表药和利水渗湿药为主。分子对接揭示了潜在靶向中药关键活性成分能够与核心基因蛋白形成稳定的互相作用。**结论**　确定了 7 个与 IPF 相关的关键基因。机器学习筛选出 ZNF204P 和 IRF7 可作为稳健的诊断生物标志物，并具有治疗靶点潜力。并预测出栀子、柴胡、肉苁蓉、黄芪等可能是靶向 IPF 核心基因的潜在中药。

关键词：生物信息学；孟德尔随机化；表达数量性状基因座；机器学习；特发性肺纤维化；中药预测

中图分类号：Q811.4；R285　　**文献标志码：**A　　**文章编号：**0253 - 2670(2026)09 - 3474 - 21

DOI: 10.7501/j.issn.0253-2670.2026.09.018

Discovery of dagnostic bomarkers and trgeted traditional Chinese medicine prediction for idiopathic pulmonary fibrosis: A multi-omics study based on Mendelian randomization and machine learning

收稿日期：2026-01-01

基金项目：中国科协青年科技人才培育工程博士生专项计划项目；国家重点研发计划（2020YFC2003104）；国家自然科学基金面上项目（82174347）；国家自然科学基金青年科学基金项目（C 类）（82505509）；四川省自然科学基金青年基金项目（2025ZNSFSC1853）；中国博士后科学基金第 75 批面上资助（地区专项支持计划）（2024MD753905）

作者简介：罗　成，博士研究生，医师，从事中医药防治呼吸病与老年病的临床研究。E-mail: 1121174213@qq.com

***通信作者：**覃琬婷，博士，助理研究员，从事中医药防治呼吸共病的作用与机制研究。E-mail: m18974400757@163.com

LUO Cheng[1, 2, 3], YE Yuanhang[4], NING Bo[5], LI Jiajie[6], TAN Junwen[1, 2], WANG Fei[2], KE Jia[7, 8, 9, 10], QIN Wanting[1, 2]

1. Chengdu University of Traditional Chinese Medicine, Chengdu 610032, China
2. Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610072, China
3. China Joint Graduate School of Chinese Medicine, Suzhou 215105
4. Pengzhou Third People's Hospital, Chengdu 611931, China
5. The Second Clinical College of Guangzhou University of Chinese Medicine, Guangzhou 510006, China
6. Yunnan University of Chinese Medicine, Kunming 650500, China
7. Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan 430061, China
8. Hubei Key Laboratory of theory and application research of liver and kidney in traditional Chinese medicine, Affiliated Hospital of Hubei University of Chinese Medicine, Wuhan 430061, China
9. Hubei Shizhen Laboratory, Wuhan 430060, China
10. Hubei Province Academy of Traditional Chinese Medicine, Wuhan 430074, China

Abstract: Objective To identify potential therapeutic targets for idiopathic pulmonary fibrosis (IPF) and predict related herbal medicines by integrating bioinformatics, Mendelian randomization (MR), and machine learning approaches. **Methods** IPF microarray datasets were obtained from the GEO database to identify differentially expressed genes (DEGs). Using expression quantitative trait loci (eQTL) data and genome-wide association study (GWAS) data, MR analysis was conducted to screen for genes associated with IPF. The risk genes identified from MR analysis were intersected with DEGs to filter core IPF-related genes. Subsequent evaluations included functional enrichment analysis, gene set enrichment analysis (GSEA), immune cell infiltration analysis, and single-cell RNA sequencing. Machine learning algorithms were applied to select optimal diagnostic feature genes. An independent GEO cohort was used for differential expression validation and receiver operating characteristic (ROC) analysis. **Results** A total of 916 IPF-associated DEGs were identified. Intersection with 224 MR risk genes yielded seven key genes: IRF7, TTC32, IFI6, and ISG15 (risk genes), along with ZNF204P, ISOC1, and CTSK (protective genes). These genes were primarily enriched in pathways related to interferon-beta production, RIG-I-like receptor signaling, Toll-like receptor signaling, and type I interferon signaling. Immune infiltration analysis revealed a decrease in M1/M0 macrophages and resting mast cells, alongside an increase in activated mast cells in IPF tissues. Single-cell RNA sequencing demonstrated specific expression patterns of these genes within epithelial cell subpopulations. Machine learning algorithms identified ZNF204P and IRF7 as the optimal diagnostic genes. Validation in the dataset confirmed their significant differential expression in IPF and high diagnostic accuracy. A total of 385 traditional Chinese medicines (TCMs) related to the key genes were predicted. The primary properties of these TCMs were cold, warm, and neutral (four natures); their main flavors were bitter, sweet, and pungent (five flavors); the principal meridian tropisms were the liver, lung, stomach, spleen, and kidney meridians; and the major classifications were heat-clearing drugs, tonifying drugs, blood-activating and stasis-resolving drugs, exterior-releasing drugs, and dampness-draining diuretics. Molecular docking simulations revealed that these chemical components could form stable interactions with the core proteins. **Conclusion** This study identified seven key genes associated with IPF. Machine learning screened ZNF204P and IRF7 as robust diagnostic biomarkers with therapeutic target potential. Furthermore, TCMs such as Zhizi (*Gardeniae Fructus*), Chaihu (*Bupleuri Radix*), Roucongrong (*Cistanches Herba*), and Huangqi (*Astragali Radix*) might be potential TCMs that targets core genes associated with IPF.

Key words: bioinformatics; Mendelian randomization; expression quantitative trait loci; machine learning; idiopathic pulmonary fibrosis; traditional Chinese medicine prediction

特发性肺纤维化（idiopathic pulmonary fibrosis，IPF）是一种慢性、进行性、致死性的间质性肺疾病，其主要病理特征为肺组织进行性不可逆瘢痕形成，致使肺功能持续减退，最终发展为呼吸衰竭[1]。患者常出现持续性干咳、活动后气短等非特异性临床表现，导致诊断困难且误诊率居高不下[2]。从病理机制上看，IPF 表现为细胞外基质过度沉积及炎症介质异常积聚[3]。流行病学调查表明，该病全球发病率为（0.09～1.30）/万人，患病率（0.33～4.51）/万人[4]，确诊后患者中位生存期通常仅为 3～5 年[5]。目前除肺移植外尚无根治手段，而移植仅适用于少数严格筛选的患者[6]。临床以抗纤维化药物联合支持治疗为主，虽然达尼布、吡非尼酮等药物用于延缓疾病进展，但其仍无法逆转纤维化进程，且存在耐受性差、治疗费用高等局限[7]。因此，深入揭示 IPF 的分子发病机制、探索早期诊断标志物并发现新型

治疗靶点成为该领域亟待突破的关键科学问题。

表达数量性状基因座（expression quantitative trait loci，eQTL）作为调控基因表达变异的重要遗传片段，通过阐明遗传变异对基因活性的影响，为解析复杂性状与疾病的遗传关联提供了基础[8]。孟德尔随机化（Mendelian randomization，MR）利用遗传变异作为工具变量可有效推断暴露因素与疾病结局之间的因果关系，显著降低传统观察性研究中的混杂偏倚与反向因果干扰[9]。近年来，MR 方法已广泛用于探讨 IPF 与血清代谢物[10]、免疫细胞[11]、炎症介质[12]及肠道菌群[13]等多种暴露因素的因果关联。全基因组关联研究（genome-wide association studies，GWAS）在识别疾病相关遗传变异中发挥关键作用，整合 eQTL 与 GWAS 数据的 MR 分析有助于阐明基因表达影响 IPF 的潜在因果机制[14]。

本研究通过整合孟德尔随机化、生物信息学及机器学习等多组学技术，旨在系统识别与 IPF 发病机制相关的关键因果基因，并阐明其在调控免疫微环境及相关信号通路中的作用。研究成果有望深化对 IPF 疾病异质性的理解，推动早期诊断与风险分层体系的优化，并为开发个体化治疗策略与新型治疗靶点提供理论依据。

1 材料与方法

1.1 数据收集和处理

从 GEO 数据库（https://www.ncbi.nlm.nih.gov/geo/）中收集 IPF 患者的转录组数据。训练数据集由 2 个数据集组成：GSE135099 和 GSE209929。利用 R 语言中的“limma”和“sva”软件包对数据进行整合，并通过主成分分析（principal component analysis，PCA）的应用去除批处理效应。选择 GSE135065 数据集作为独立验证集。此外，该研究还获得了单细胞转录组数据集 GSE279404，这将对后续针对特定细胞类型的研究至关重要。数据集信息见表 1。

表 1 微阵列数据

Table 1 Microarray data

数据集	数据平台	IPF 组数据数量	对照组数据数量	数据类型
GSE209929	GPL21185-21174	3	3	训练集
GSE135099	GPL15207-17536	9	9	训练集
GSE135065	GPL15207	9	9	测试集
GSE279404	GPL21697	8	8	scRNA-seq

1.2 差异基因表达分析（differential gene expression analyses，DEGs）

采用 R 语言生物信息学工具，对 IPF 样本与对照组的基因表达数据进行差异分析。首先利用“limma”包结合“dplyr”包进行数据处理和线性模型拟合，并使用 Benjamini-Hochberg 方法对原始 *P* 值进行多重检验校正，得到调整后 *P* 值。将满足调整后 *P* 值＜0.05 且$|\log_2 FC|$＞0.585［FC 表示差异倍数（fold change）］的基因定义为显著差异表达基因。随后使用“ggplot2”包绘制火山图以展示差异表达基因的整体分布特征，并借助“pheatmap”包生成热图，直观呈现显著差异基因的表达模式。

1.3 MR 数据来源

基于 R 平台采用 2 样本 MR 方法，探究人类基因表达与 IPF 之间的潜在因果关联。分析所用遗传数据来源于 2 个公共数据库：作为暴露变量的 eQTL 数据来自 OpenGWAS 平台，涵盖 19 942 个基因的遗传变异信息；IPF 相关结局数据源自 FINNGEN 项目，包含 497 283 名欧洲裔个体的样本。为保证人群遗传背景一致性，所有 GWAS 摘要数据均限定于欧洲裔个体。需特别说明的是，本研究仅基于公开 GWAS 摘要统计数据进行二次分析，无需另行伦理审查或知情同意流程。

1.4 工具变量（instrumental variables，IVs）筛选

依据以下严格标准筛选 IVs：首先采用“TwoSampleMR”软件包全面解析各基因相关 eQTL 数据，筛选与基因表达水平显著相关（显著性阈值为 $P<5\times10^{-8}$）的单核苷酸多态性（single nucleotide polymorphism，SNP）。在质控方面，设定连锁不平衡参数为 $r^2<0.001$，并要求 SNP 间物理距离大于 1×10^4 kb，以确保所选 IVs 为暴露的强独立相关因素。为评估 IVs 强度，计算各暴露变量的 *F* 统计量，剔除 *F* 值低于 10 的弱 IVs，以降低潜在弱工具偏倚。

1.5 MR 分析

在确定合格工具变量后，采用两样本 MR 框架

评估因果关系。主要使用“TwoSampleMR”与“VariantAnnotation”软件包，综合运用 5 种统计方法：以逆方差加权法作为核心分析方法，辅以 MR-Egger 回归、加权中位数法、加权众数法及简单众数法进行验证。为保证分析结果有效性，重点检验 MR 分析的基本假设——IVs 仅通过暴露因素影响结局，为此采用 MR-Egger 截距检验（$P<0.05$）评估水平多效性，若截距显著偏离零则提示可能存在多效性偏倚。此外，针对不同研究人群与实验设计可能产生的异质性，通过 Cochran's Q 检验（$P<0.05$）进行异质性评估，并采用留一法敏感性分析验证结果稳健性，具体流程如图 1 所示。

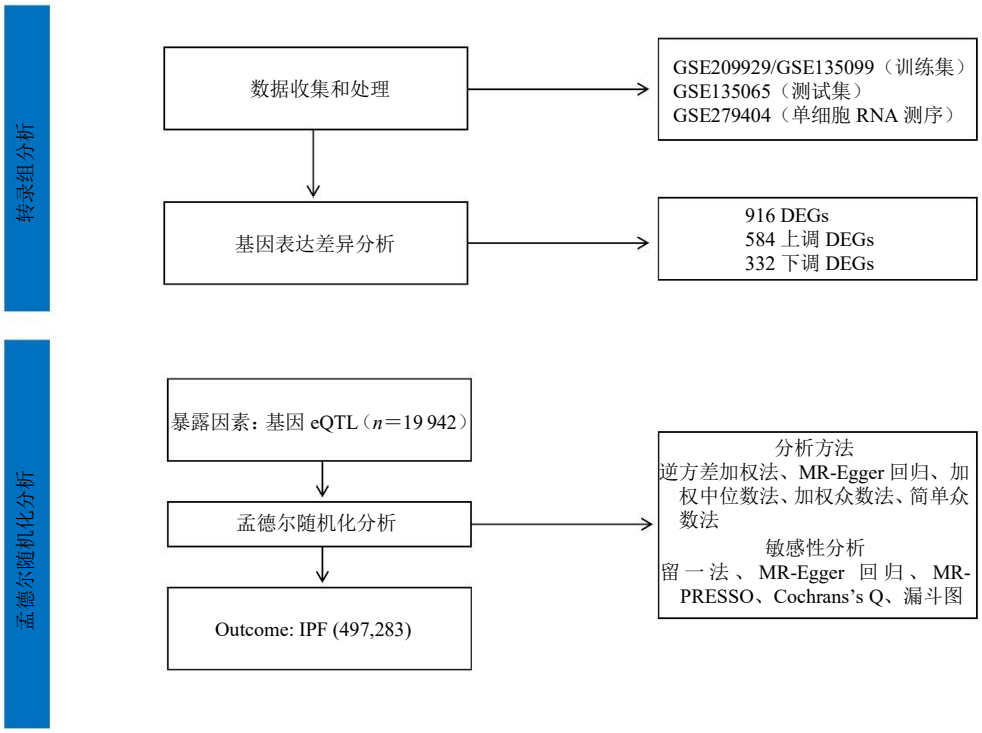


图 1　技术路线图

Fig. 1　Workflow diagram

1.6　IPF 关键基因筛选与验证

基于孟德尔随机化分析结果，首先排除存在显著异质性（Cochran's Q 检验 $P\leqslant1.0$）及比值比（odds ratio，OR）低于 1.0 的保护性基因。利用“VennDiagram”包对 MR 筛选所得关键基因与差异表达基因进行交集分析，可视化展示上调与下调基因的重叠情况。通过“forestploter”等软件包整合逆方差加权法与加权中位数法的结果，生成展示关键基因效应值及置信区间的森林图。使用“circlize”包绘制环状基因组图谱，直观呈现重要基因在染色体上的分布特征。为验证关键基因表达趋势，在独立数据集（GSE135099）中进行表达分析，采用“limma”包进行差异表达分析，结果通过“ggplot2”与“ggpubr”包可视化呈现。

1.7　关键基因功能富集分析

采用生物信息学方法系统分析关键基因的功能特征。利用“clusterProfiler”与“org.Hs.eg.db”等 R 包对筛选所得关键基因进行全面功能注释，包括基因本体论（gene ontology，GO）富集分析以及京都基因与基因组百科全书（Kyoto encyclopedia of genes and genomes，KEGG），并以气泡图形式进行可视化呈现。基于 GEO 训练集表达谱数据，采用基因集富集分析（gene set enrichment analysis，GSEA）方法，比较关键基因高/低表达组间的通路活性差异，使用 MSigDB 数据库中的“c2.cp.kegg.v2023.1.Hs.symbols”基因集进行分析。其中，根据其中位数将样本分为高表达组（表达量＞中位数）与低表达组（表达量≤中位数）

1.8　免疫细胞浸润与免疫相关性分析

采用“LM22”标志文件结合“CIBERSORT”算法分析数据集中对照组与实验组的免疫细胞浸润水平，设定 1 000 次置换检验，以 $P<0.05$ 为显著性阈值。通过“corrplot”与“ggplot2”包绘制箱线图及柱状图实现结果可视化。进一步使用“linkET”包探究

免疫浸润细胞与 IPF 关键基因间的关联性。

1.9 scRNA-seq 数据分析

首先进行严格质量控制，设定阈值过滤核糖体相关基因数低于 5 000 的样本；剔除线粒体基因比例＞15%或血红蛋白基因表达＞3%的低质量细胞。质控后采用 LogNormalize 算法进行数据标准化，通过 FindVariableFeatures 模块筛选前 2 000 个高变基因作为特征基因。降维阶段采用 PCA 提取主成分构建特征空间，基于图论的聚类方法识别不同细胞亚群。利用均匀流形近似与投影（uniform manifold approximation and projection，UMAP）与 t 分布随机邻域嵌入（t-distributed stochastic neighbor embedding，t-SNE）等非线性模型可视化细胞群体分布，揭示亚群空间结构。采用 Harmony 框架校正批次效应，在修正后的特征空间进行重聚类。通过 FindAllMarkers 模块进行差异基因筛选，绘制热图展示显著差异基因表达模式。SingleR 工具整合人类原发性细胞图谱参考数据库，对亚群进行全面功能注释，并将注释结果融入 UMAP/t-SNE 可视化。AUCell 算法构建基因表达排序矩阵，判定各细胞亚群中目标基因集的富集程度，通过 UMAP/t-SNE 降维图与小提琴图系统呈现目标基因集富集特征及其在细胞类型中的特异性分布规律。

1.10 受试者工作特征（receiver operating characteristic，ROC）曲线绘制与表达验证

基于 GSE135065 数据集，采用 Mann-Whitney U 检验分析标志基因在对照组与 IPF 患者组间的表达分布差异。随后利用 rms 包构建多因素逻辑回归模型，整合关键变量并绘制临床预测列线图。采用 pROC 工具绘制 ROC 曲线，以曲线下面积（area under urve，AUC）评估模型分类效能（AUC≥0.7 作为有效性阈值）。为深入探究区分基因的互作网络，本研究应用 Pearson 线性相关算法全面分析关键基因对的共表达模式，通过相关系数矩阵揭示基因调控网络的潜在关联特征。

1.11 机器学习算法

为有效识别 IPF 相关关键枢纽基因，本研究联合应用 3 种机器学习方法：最小绝对收缩与选择算子回归（least absolute shrinkage and selection operator，LASSO）、随机森林（random forest，RF）及支持向量机递归特征消除（support vector machine recursive feature elimination，SVM-RFE）。选择这 3 种算法旨在整合其互补优势：LASSO 通过 L1 正则化实现特征稀疏化与模型可解释性，适合筛选少量核心特征；RF 通过集成决策树捕捉基因间的非线性与交互作用，并评估特征重要性；SVM-RFE 通过递归剔除特征，优化分类边界，以识别最具判别力的稳定特征子集。这种组合策略旨在通过算法交叉验证，降低单一方法的偏倚，提高筛选结果的稳健性与生物学可靠性。最终，取 3 种算法结果的交集作为最优候选特征基因。这一策略旨在以最高的一致性标准筛选基因，最大限度地提高特征的特异性，确保所筛选的基因在不同算法原理下均表现稳健，从而为后续的生物标志物验证提供高置信度的候选靶点。

1.12 潜在干预中药预测与分析

将 IPF 相关的核心靶点提交至比较毒理基因组学数据库（comparative toxicogenomics database，CTD），从中筛选出文献支持数量大于 1 的化学成分。随后，借助 ITCM 数据库（http://itcm.biotcm.net/）检索与上述化学成分相对应的中药，并依据《中国药典》2020 年版及“十四五”规划教材《中药学》对中药名称进行标准化处理，剔除无法查询的中药条目。接着，通过古今医案云平台分析这些中药的性能与功效。最后，根据“十四五”规划教材《中药学》的分类标准对药物进行系统归类。

1.13 潜在中药靶点筛选

根据筛选结果确定了使用频次最高的 3 种中药，并基于 TCMSP 数据库（https://www.tcmsp-e.com）对其潜在活性成分进行了二次筛选。筛选标准：①分子类药性（drugability，DL）≥0.18；②口服生物利用度（oral bioavailability，OB）≥30%；③Caco-2 细胞渗透性（Caco-2 permeability，Caco-2）＞1×10^{-5}；④相对分子质量（molecule weight，M_W）≤500。在此基础上，进一步结合国内外已发表的基础实验研究文献，对所筛选出的活性成分进行药理学证据支持与验证。获取满足条件的活性成分后，进一步从 TCMSP 数据库中提取各成分对应的预测靶点，去除重复值后，使用 UniProt 数据库（https://www.uniprot.org/）进行标准化处理，得到中药潜在作用靶点。

1.14 分子对接验证

利用分子对接方法探索蛋白质与配体的相互作用。首先，从 UniProt 数据库（https://www.uniprot.rg/）和 PDB 数据库（http://www.rcsb.org/）获取并下载目标受体的三维结构。若相应晶体结构尚未解

析，则改用 AlphaFold 蛋白质结构数据库（https://ww.alphafold.ebi.ac.uk/）中预测的结构模型。同时，从 PubChem 数据库（https://pubchem.ncbi.nlm.nih.ov/）获取预测中药的活性化合物的二维分子形式，然后使用 Chem3D 22.0.0 软件将其转化为三维结构，并采用 MM2 力场进行能量最小优化，以获取最稳定的构象用于后续对接。随后，使用 AutoDock Vina 程序进行受体-配体复合物的对接模拟，并通过 PyMOL 软件可视化结合构象。

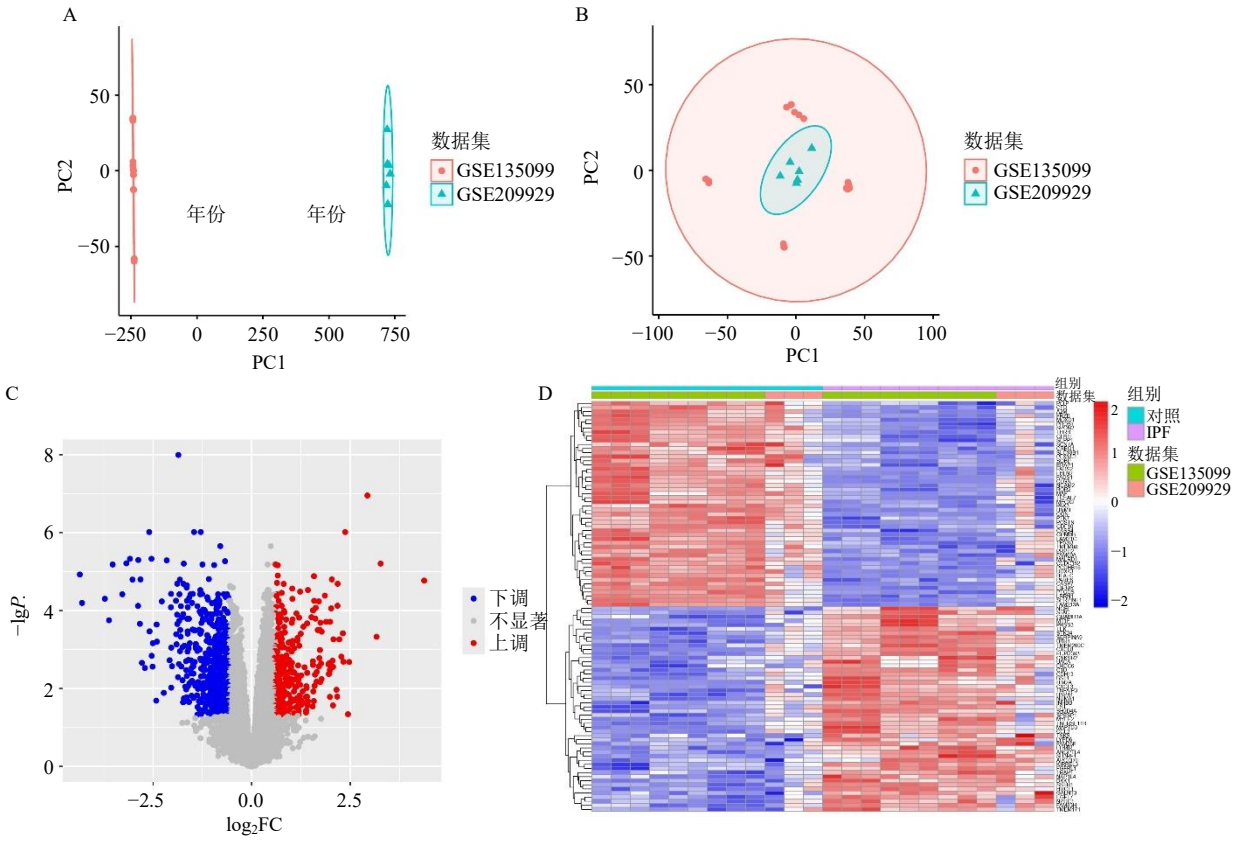
1.15 统计学分析

符合正态分布的数据以 $\bar{x} \pm s$ 表示，非正态分布数据则以中位数四分位间距表示。计量资料以 $\bar{x} \pm s$ 表示。两组间比较采用非配对 t 检验或 Mann-Whitney U 检验；多组间比较采用单因素方差分析及相应的多重比较检验。为探讨关键基因与浸润免疫细胞之间的关联性，采用 Pearson 相关性分析。所有统计分析均使用 SPSS 26.0 软件完成。

2 结果

2.1 差异表达基因鉴定

在 IPF 训练数据集预处理阶段，采用批次效应校正技术以消除技术变异带来的干扰（图 2-A、B）。基于校正后的表达谱进行差异表达分析，共鉴定出 916 个差异表达基因（图 2-C、D），其中包括 332 个上调基因和 584 个下调基因。



A-批次校正前样本分布；B-批次校正后样本分布；C-差异表达基因火山图；D-差异表达基因热图。

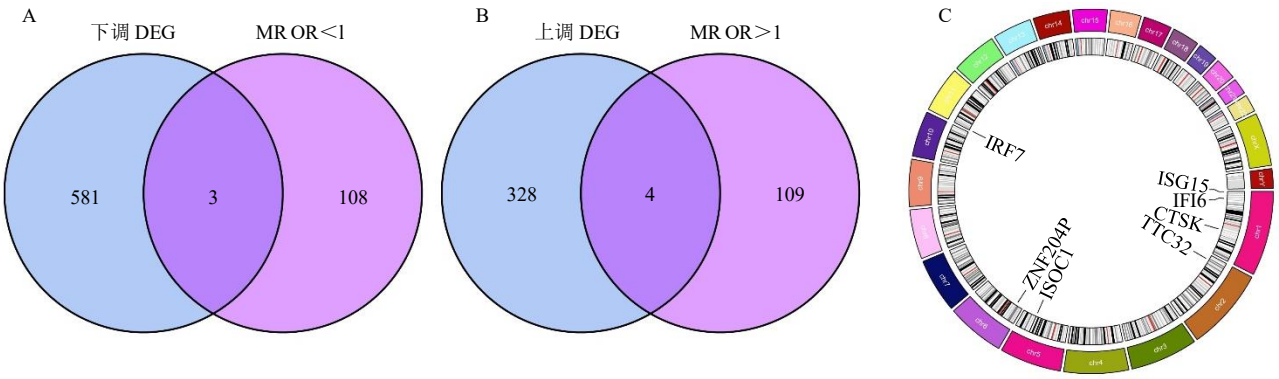
A-before the batch correction; B-after the batch correction; C-volcano plot of differential expression genes; D-heatmap of differential expression genes.

图 2 数据集批次校正与差异表达基因分析

Fig. 2 Dataset batch correction and differential expression gene analysis

2.2 IVs 筛选与 MR 分析

经过连锁不平衡调整并剔除弱 IVs 后，共获得 26 152 个与目标基因显著相关的 SNP 作为可靠 IVs。采用 2 样本 MR 框架评估各 SNP 对 IPF 发病风险的因果效应。以逆方差加权法 $P<0.05$ 为显著性阈值，共筛选出 224 个与 IPF 存在因果关联的基因，其中高风险基因 113 个，低风险基因 111 个（图 3-A、B）。将 MR 分析确定的高、低风险基因集与转录组 DEGse 取交集，最终鉴定出 7 个关键共表达基因（图 3-C）：干扰素调节因子 7（interferon regulatory factor 7，IRF7）、四肽重复结构域 32（tetratricopeptide repeat domain 32，TTC32）、干扰素



D

exposure	nsnp	method	pval	OR(95% CI)
ISOC1	8	Weighted median	**<0.001**	0.830 (0.746 to 0.923)
	8	Inverse variance weighted	**<0.001**	0.848 (0.775 to 0.928)
IFI6	8	Weighted median	**0.042**	1.317 (1.010 to 1.717)
	8	Inverse variance weighted	**0.027**	1.243 (1.025 to 1.507)
CTSK	4	Weighted median	**0.029**	0.880 (0.785 to 0.987)
	4	Inverse variance weighted	**0.021**	0.880 (0.789 to 0.981)
TTC32	4	Weighted median	**0.027**	1.110 (1.012 to 1.217)
	4	Inverse variance weighted	**0.019**	1.115 (1.018 to 1.222)
IRF7	5	Weighted median	**<0.001**	1.398 (1.150 to 1.701)
	5	Inverse variance weighted	**0.014**	1.361 (1.064 to 1.741)
ISG15	5	Weighted median	0.112	1.189 (0.961 to 1.472)
	5	Inverse variance weighted	**0.048**	1.225 (1.002 to 1.498)
ZNF204P	3	Weighted median	**0.004**	0.783 (0.663 to 0.925)
	3	Inverse variance weighted	**0.004**	0.786 (0.668 to 0.926)

A-疾病上调差异表达基因与孟德尔随机化结果中 OR 值小于 1 的基因取交集；B-疾病下调差异表达基因与孟德尔随机化结果中 OR 值大于 1 的基因取交集；C-疾病关键基因在人类染色体上的定位分布；D-IPF 疾病相关关键基因森林图。

A-disease upregulated DEGs are intersected with genes with OR values less than one in the MR results; B-disease downregulated DEGs are intersected with genes with OR values greater than one in the MR results; C-position of disease-critical genes on human chromosomes; D-forest map of critical genes related to IPF diseases.

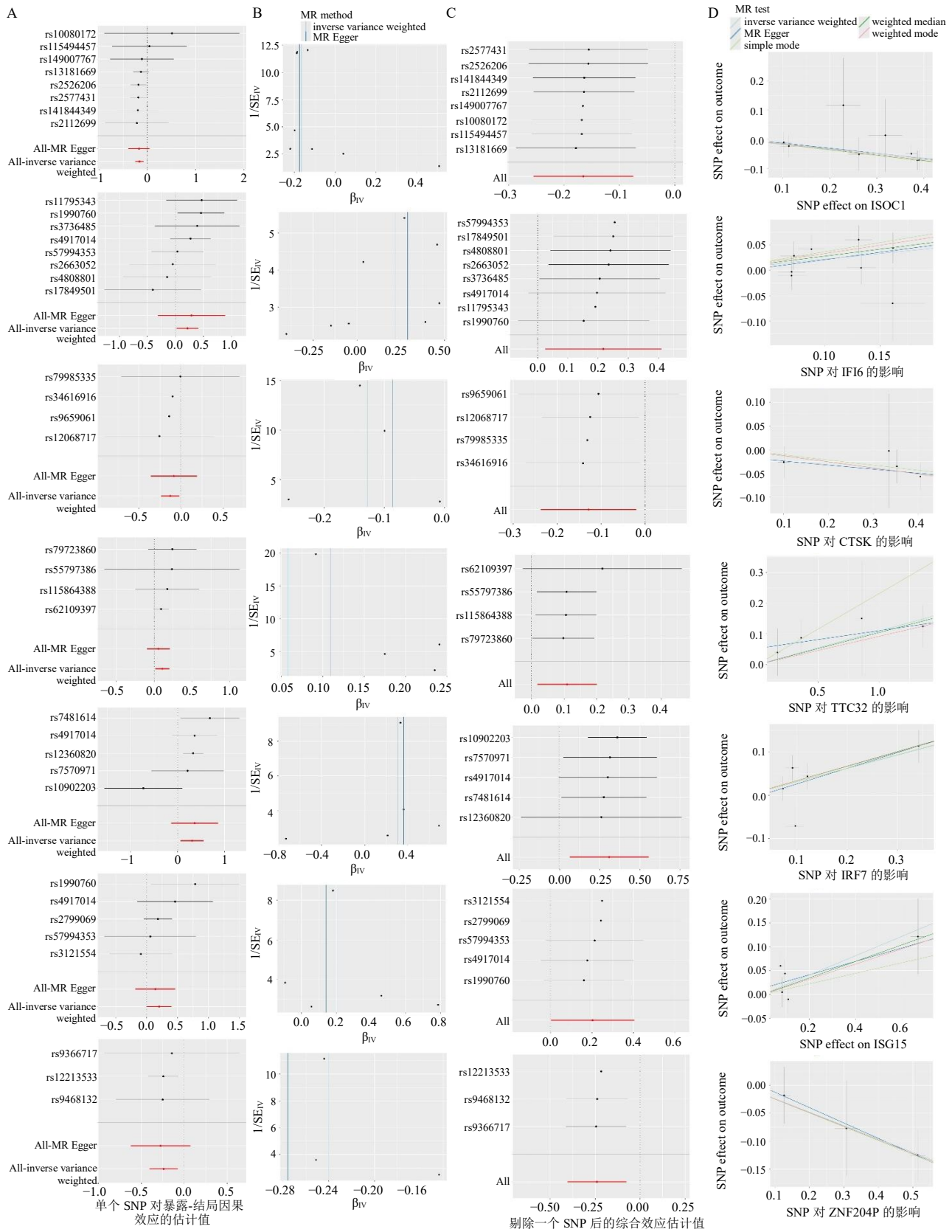
图 3　关键基因筛选与定位

Fig. 3　Screening and localization of critical genes

α 诱导蛋白 6（interferon alpha inducible protein 6，IFI6）、干扰素刺激基因 15（interferon stimulated gene 15，ISG15）、锌指蛋白 204 假基因（zinc finger protein 204 pseudogene，ZNF204P）、异分支酸酶结构域包含蛋白 1（isochorismatase domain containing 1，ISOC1）和组织蛋白酶 K（cathepsin K，CTSK）。MR 因果效应（图 3-C）显示，IRF7 表达增加使 IPF 发生风险升高 36.1%（OR＝1.361，95% CI：1.064～1.741，P＝0.014）；TTC32 表达增加使 IPF 发生风险升高 11.5%（OR＝1.115，95% CI：1.018～1.222，P＝0.019）；IFI6 表达增加使 IPF 发生风险升高 24.3%（OR＝1.243，95% CI：1.025～1.507，P＝0.027）；ISG15 表达增加使 IPF 发生风险升高 22.5%（OR＝1.225，95% CI：1.002～1.498，P＝0.048）。相比之下，ZNF204P 表达下降使 IPF 发生风险降低 21.4%（OR＝0.786，95% CI：0.668～0.826，P＝0.004）；ISOC1 表达下降使 IPF 发生风险降低 15.2%（OR＝0.848，95% CI：0.775～0.928，P＜0.001）；CTSK 表达下降使 IPF 发生风险降低 12%（OR＝0.880，95% CI：0.789～0.981，P＝0.021）。

2.3　基因敏感性分析

为评估 7 个 IPF 关键基因的可靠性，进行了系统敏感性分析（图 4-A～D）。首先采用 MR-Egger 回归与 Cochran's Q 检验分别评估水平多效性与异质性，统计结果显示均未达到显著性（P＞0.05），表明研究结果具有较高可信度（表 2）。漏斗图分析



A-7 个关键基因的森林图；B-7 个关键基因的漏斗图；C-7 个关键基因的留一法分析图；D-7 个关键基因的散点图。

A-forestplot of MR analysis of these seven critical genes; B-funnel plot of MR analysis of these seven critical genes; C-leave one out plot of MR analysis of these seven critical genes; D-scatterplot of MR analysis of these seven critical genes.

图 4　IPF 关键基因与 IPF 相关性的 MR 分析

Fig. 4　MR analysis of correlation between critical genes of IPF and IPF

表 2　IPF 的 7 个关键基因的敏感性分析

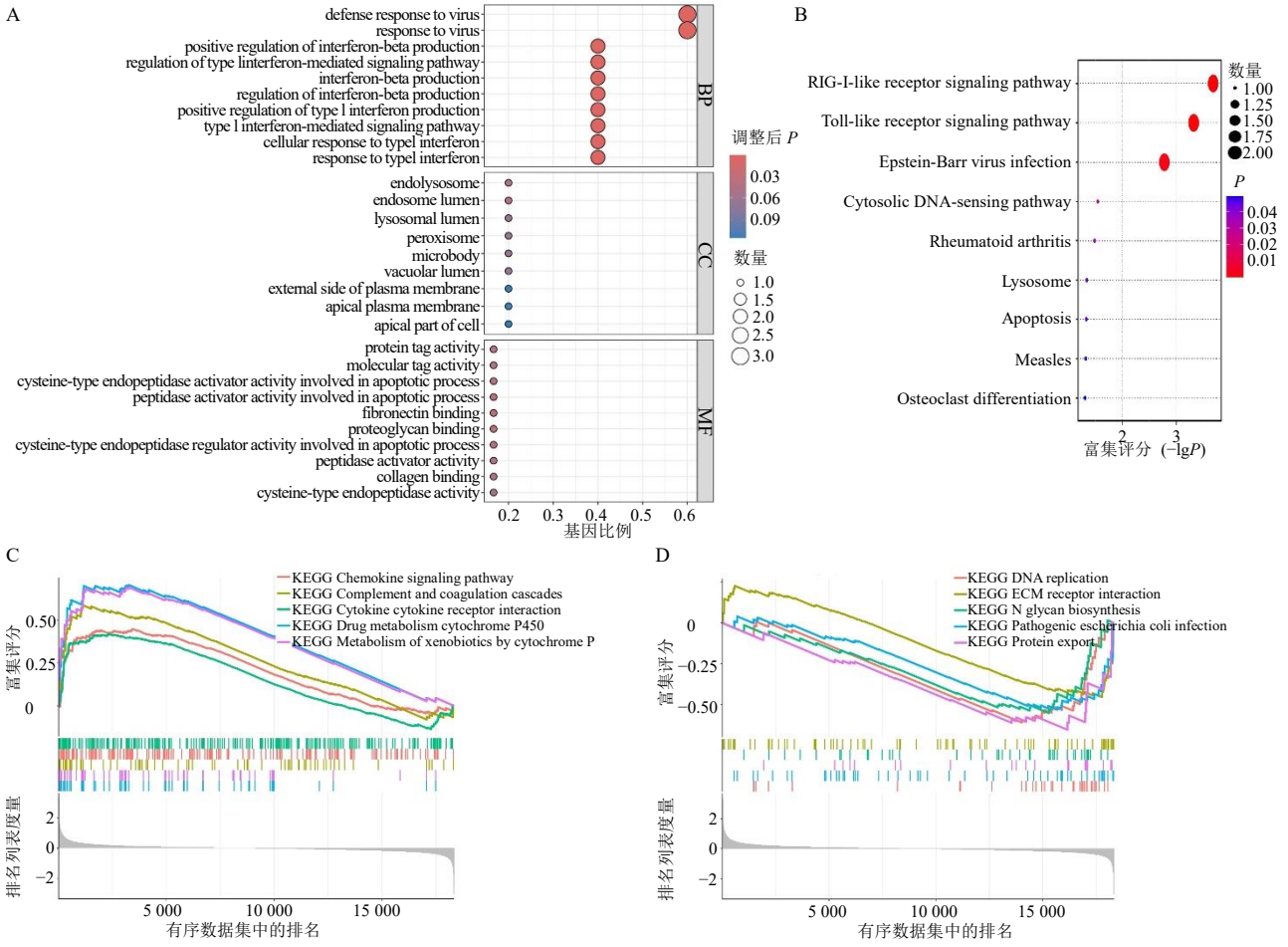
Table 2　Sensitivity analysis of seven critical genes associated with IPF

基因	$P_{MR\text{-}Egger}$	$P_{MR\text{-}Egger.Q}$	$P_{IVW.Q}$
ISOC1	0.172	0.960	0.983
IFI6	0.383	0.427	0.537
CTSK	0.603	0.874	0.944
TTC32	0.526	0.896	0.815
IRF7	0.252	0.059	0.108
ISG15	0.456	0.232	0.321
ZNF204P	0.363	0.922	0.970

显示 SNP 效应值呈对称分布，未发现异常值，排除了个别 SNP 对总体估计产生偏倚的可能性。此外，留一法敏感性分析进一步证实结果的稳定性，即使逐次剔除单个 SNP，剩余 SNP 的合并效应值仍保持一致性，验证了分析方法的稳健性。这些综合验证结果共同证实了本研究发现结论的可靠性。

2.4　功能富集分析

GO 功能富集分析（图 5-A）显示，筛选出的关键基因在多个生物过程中显著富集，包括病毒应答反应、干扰素-β 生成增强以及 I 型干扰素介导的信



A-IPF 关键基因 GO 富集分析；B-IPF 关键基因 KEGG 通路富集分析；C-IPF 关键基因高表达组 GSEA 富集结果；D-IPF 关键基因低表达组 GSEA 富集结果。

A-GO enrichment analysis of IPF critical genes; B-KEGG enrichment analysis of IPF critical genes; C-GSEA enrichment results of the high-expression group of the critical genes of IPF; D-GSEA enrichment results of the low-expression group of the critical genes of IPF.

图 5　关键基因功能富集与 GSEA 富集分析

Fig. 5　Functional enrichment analysis and GSEA enrichment analysis of critical genes

号通路调控。在细胞组分层面，这些基因主要富集于内溶酶体、内体腔、溶酶体腔及过氧化物酶体等结构。分子功能分析结果表明，这些基因参与蛋白质标记活性、分子标签活性以及半胱氨酸型内肽酶活化等过程，后者在细胞凋亡进程中发挥作用。KEGG 通路富集分析揭示，这 7 个关键基因主要参

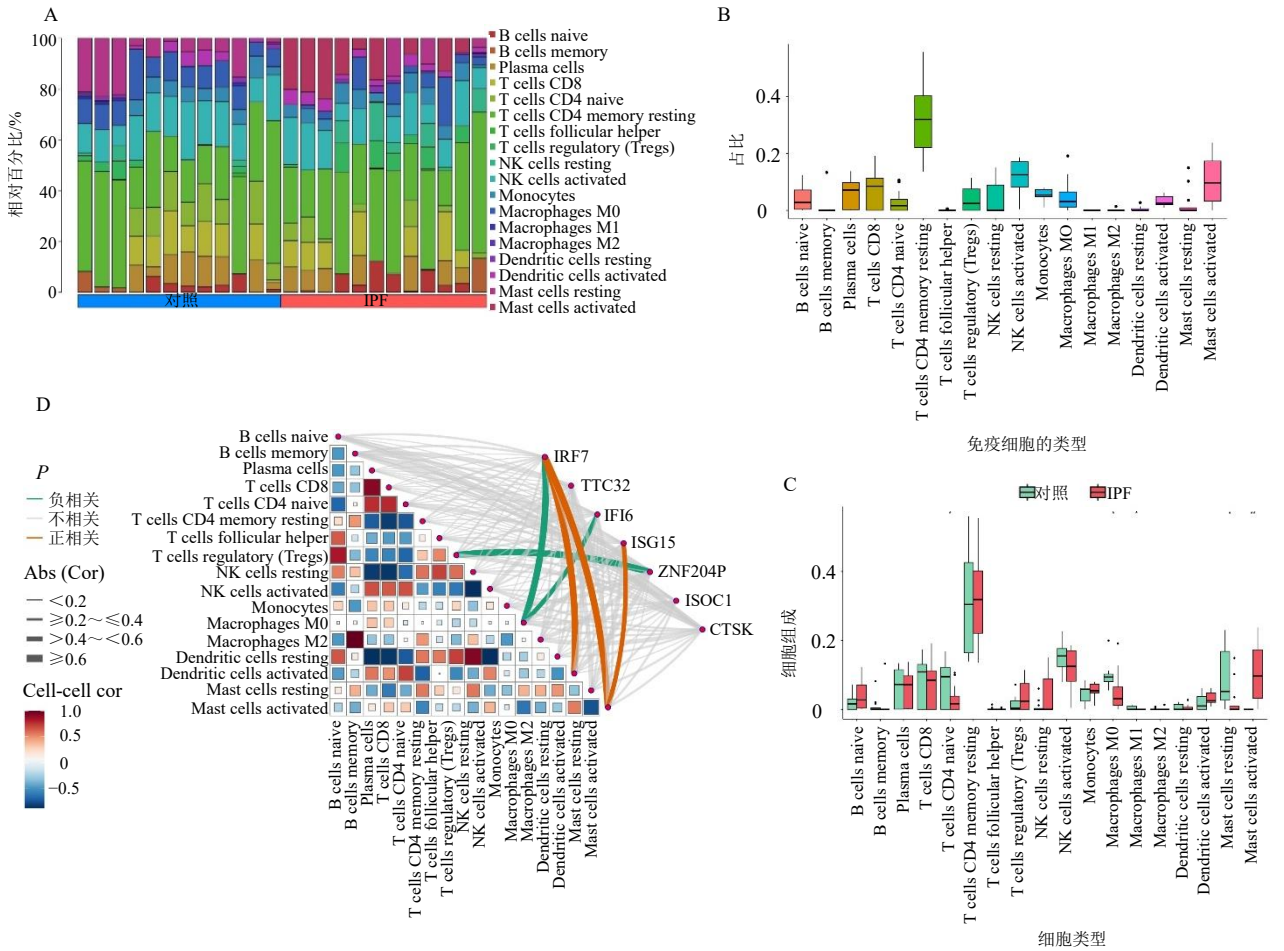
与视黄酸诱导基因 1（retinoic acid-inducible gene-I，RIG-1）样受体信号通路、Toll 样受体信号通路、EB 病毒感染及胞质 DNA 感知通路等（图 5-B）。

2.5 GSEA

GSEA 分析揭示了关键基因在高、低表达组间的功能差异。结果显示，高表达组中富集程度最高的通路为细胞色素 P450 药物代谢，其次为细胞色素 P450 介导的外源性物质代谢、趋化因子信号通路、补体与凝血级联反应以及细胞因子-受体相互作用通路（图 5-C）。而在低表达组中，关键基因主要富集于蛋白质外泌通路，其次是 DNA 复制、N-聚糖生物合成、致病性大肠杆菌感染及细胞外基质-受体相互作用通路（图 5-D）。这些结果表明关键基因可能通过调控蛋白质合成、代谢调节、蛋白互作水平、细胞周期控制以及免疫应激反应等生物学过程影响 IPF 进展。

2.6 免疫细胞浸润水平及其与关键基因的关联分析

免疫细胞浸润分析揭示了 IPF 微环境的显著特征。通过 CIBERSORT 算法量化分析，检测到 18 种免疫细胞类型与基因表达谱数据的关联性。在所有识别的免疫细胞中，静息记忆 $CD4^+$ T 细胞与活化肥大细胞占比最高（图 6-A、B）。与对照组相比，IPF 样本中初始 $CD4^+$ T 细胞、M1 巨噬细胞、M0 巨噬细胞及静息肥大细胞的比例显著降低，而活化肥大细胞比例明显升高，提示其可能参与疾病进程（图 6-C）。为探究关键基因的免疫学特征，进一步分析了关键基因与免疫浸润细胞的相关性（图 6-D）。结果显示，IRF7、IFI6、ISG15 和 ZNF204P 与



A-对照组与 IPF 组免疫浸润细胞比例堆叠柱状图；B-18 类免疫细胞箱线图；C-对照组与 IPF 组免疫细胞浸润水平比较箱线图；D-疾病关键基因与浸润免疫细胞相关性热图。

A-stacked histogram of the proportions of immune infiltration cells between control and IPF groups; B-box plot of 18 types of immune cells; C-box plot of the infiltration level of immune cells between control and IPF groups; D-correlation analysis between disease critical genes and infiltrated immune cells.

图 6 免疫浸润分析

Fig. 6 Immune infiltration analyses

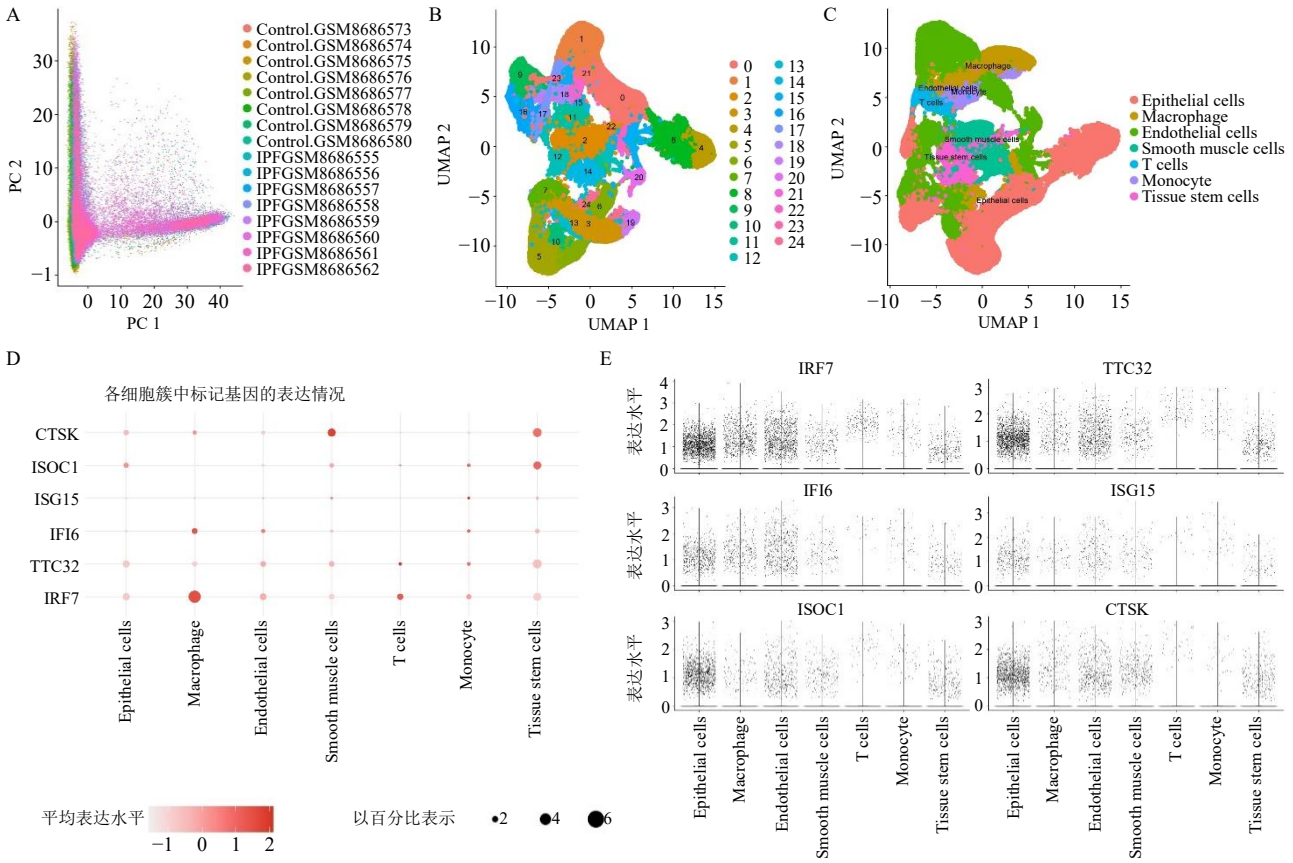
多种免疫浸润细胞存在关联：IRF7 和 IFI6 水平与 M0 巨噬细胞呈显著负相关；调节性 T 细胞与 ZNF204P 表达呈显著正相关；活化肥大细胞与 IRF7、ISG15 表达水平呈负相关；活化树突状细胞亦与 IRF7 表达呈负相关。值得注意的是，IRF7 展现出最广泛的免疫调节特性，与多种免疫细胞亚群的浸润水平均存在显著关联。这些结果提示 IPF 关键基因可能通过影响特定免疫细胞的募集与活化参与疾病进程。

2.7　scRNA-seq 数据分析

选取 GSE279404 数据集中的 18 个样本进行 scRNA-seq 分析。经过严格质控流程，最终获得包含 30 396 个基因的标准化表达矩阵。通过 PCA 与 UMAP 技术将细胞划分为 24 个亚群（图 7-A、B），并将所有细胞注释为 7 种主要类型：上皮细胞、巨噬细胞、内皮细胞、平滑肌细胞、T 细胞、单核细胞及组织干细胞（图 7-C）。为深入探究细胞异质性，重点分析了 IPF 关键基因的表达谱特征（图 7-D、E）。结果显示这些基因在上皮细胞中显著富集，提示上皮细胞可能在 IPF 进展中发挥关键作用。

2.8　机器学习算法筛选最优特征基因

为识别 IPF 相关关键基因并确定最优特征基因，本研究采用 LASSO、RF 及 SVM-RFE 3 种算法进行基因筛选。通过 RF 算法建立了 IPF 关键基因的重要性排序（图 8-A、D）；SVM-RFE 算法经过 5 折交叉验证迭代后筛选出 3 个特征基因（图 8-B～

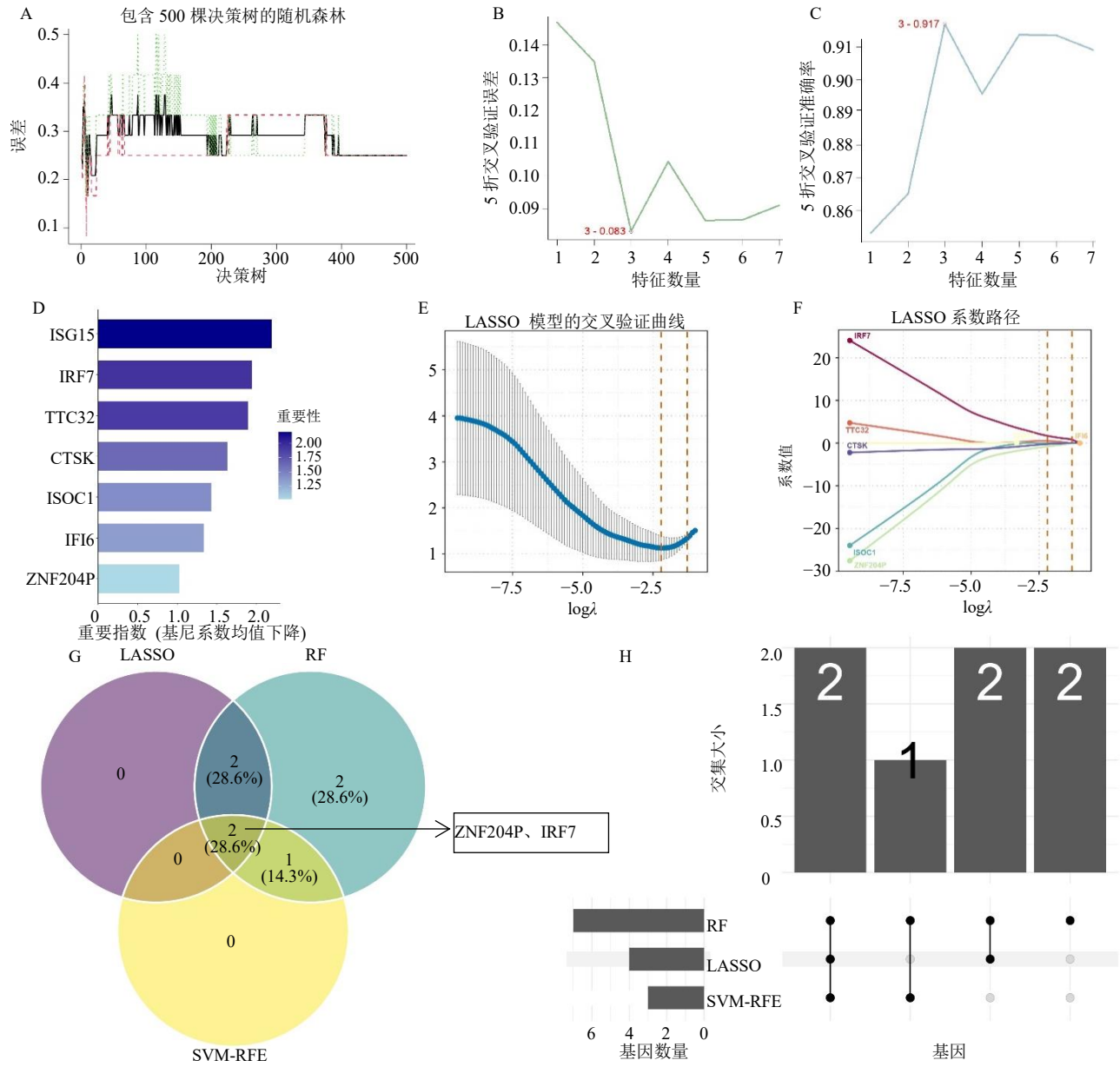


A-基于 UMAP 降维与聚类后形成的细胞群体可视化（每个点代表 1 个细胞，不同颜色表示不同细胞亚群）；B-基于 t-SNE 降维与聚类后形成的细胞群体可视化（每个颜色代表 1 个细胞亚群）；C-IRF7、TTC32、IFI6、ISG15、ZNF204P、ISOC1 和 CTSK 7 个关键基因在已注释细胞群体中的表达水平小提琴图；D-关键基因在 t-SNE 可视化图中的表达分布（颜色梯度表示基因表达丰度）；E-关键基因在 UMAP 可视化图中的表达分布（颜色梯度表示基因表达丰度）。

A-visualization of cell populations based on UMAP dimensionality reduction and clustering (each dot represents a single cell, with different colors indicating distinct cell subpopulations); B-visualization of cell populations based on t-SNE dimensionality reduction and clustering (each color represents a cell subpopulation); C-violin plots showing the expression levels of the seven key genes IRF7, TTC32, IFI6, ISG15, ZNF204P, ISOC1, and CTSK across annotated cell subpopulations; D-expression distribution of key genes in the t-SNE visualization (color gradient indicates gene expression abundance); E-expression distribution of key genes in the UMAP visualization (color gradient indicates gene expression abundance).

图 7　IPF 患者单细胞数据细胞亚群注释

Fig. 7　Annotation of cell subpopulations in single-cell data from IPF patients



A-最优决策树的 RF 误差曲线；B-SVM-RFE 算法 5 折交叉验证误差率曲线；C-SVM-RFE 算法 5 折交叉验证准确率曲线；D-采用 RF 计算重叠候选基因相对重要性，*x* 轴为重要性指数，*y* 轴为基因变量；E-交叉验证曲线；F-潜在自变量的系数路径图；G-LASSO、RF 与 SVM-RFE 算法取交集的韦恩图；H-集合图展示 3 种算法共同筛选的 2 个最优特征基因。

A-RF error curve of the best tree; B-error rate curve of the 5-fold cross-validation of the SVM-RFE algorithm; C-accuracy curve graph of the 5-fold cross-validation of the SVM-RFE algorithm; D-relative importance of overlapping candidate genes was calculated using RF, importance indices are plotted on the *x*-axis, and genetic variables are plotted on the *y*-axis; E-coefficient path of potential independent variables; F-cross-validation curve; G-venn diagram showing the two optimal feature genes shared by LASSO, RF, and SVM-REF algorithms; H-upset diagram showing the two optimal feature genes shared by LASSO, RF, and SVM-REF algorithms.

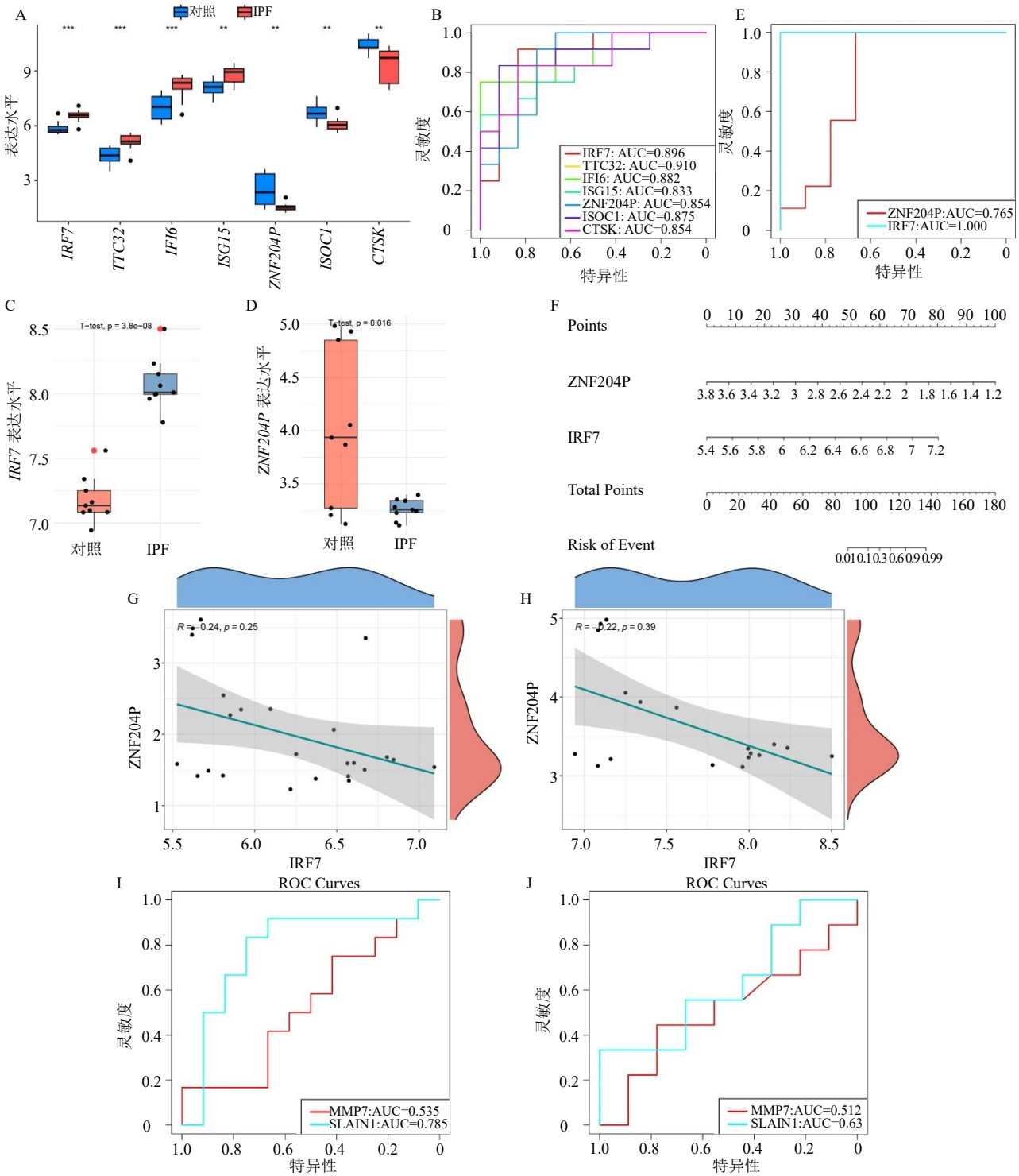
图 8　机器学习算法集成筛选最优特征基因

Fig. 8　Integration and screening of optimal characteristic genes using machine learning algorithms

C)；最小绝对收缩和选择算子（logistic least absolute shrinkage and selection operator，LASSO）回归确定了具有最小二项偏差的 4 个关键基因（图 8-E、F）。最终，为获取高特异性且稳健的特征基因，取 3 种算法结果的交集，精准鉴定出 ZNF204P 和 IRF7 2 个最优特征基因（图 8-G、H）。

2.9　差异表达验证与 ROC 曲线

基于原本训练集探究关键基因在 IPF 中的表达水平，结果显示其表达量存在显著差异（图 9-A）。进一步通过 ROC 曲线分析评估这些基因的诊断效



A-训练集中 7 个 IPF 关键基因表达水平；B-训练集中 IPF 关键基因 ROC 曲线，AUC 代表其诊断价值；C-测试集中 *IRF7* 表达水平验证；D-测试集中 *ZNF204P* 表达水平验证；E-测试集中 IRF7 和 ZNF204P 的 ROC 曲线；F-基于 ZNF204P 和 IRF7 构建的列线图；G-训练集中 IRF7 与 ZNF204P 表达呈显著负相关；H-测试集中 IRF7 与 ZNF204P 表达呈显著负相关；I-训练集中 MMP7 和 SLAIN1 的 ROC 曲线；J-测试集中 MMP7 和 SLAIN1 的 ROC 曲线。

A-expression levels of seven IPF key genes in the training set; B-ROC curve of IPF key genes in the training set, and the area under the curve represents its diagnostic value; C-verification of IRF7 expression level in the test set; D-validation of ZNF204P expression level in the test set; E-ROC curves of IRF7 and ZNF204P in the test set; F-ZNF204P and IRF7 based nomogram; G-significant negative correlation between IRF7 and ZNF204P expression in the training set; H-significant negative correlation between IRF7 and ZNF204P expression in the test set; I-ROC curves of MMP7 and SLAIN1 in the training set; J-ROC curves of MMP7 and SLAIN1 in the test set.

图 9　IPF 关键基因在 GEO 验证集中的表达验证

Fig. 9　Expression validation of IPF critical genes in GEO testing group

能，发现所有 IPF 关键基因的 AUC＞0.8，表明其具有较高的预测能力（图 9-B）。为评估 3 种机器学习技术筛选的特征基因的诊断效能，采用外部测试集 GSE135065 进行验证分析。基因表达水平评估表明，与对照组相比，IPF 患者样本中 IRF7 表达显著升高（P＜0.05），而 ZNF204P 表达显著降低（P＜0.05，图 9-C、D）。ROC 曲线分析显示 ZNF204P 和 IRF7 的 AUC 分别为 0.765 和 1.000（图 9-E）。综合 AUC 结果显示 ZNF204P 和 IRF7 均展现出优异的疾病鉴别能力，印证了前期分析结果的有效性。基于此，构建了整合 ZNF204P 和 IRF7 表达数据的诊断列线图模型（图 9-F）。为探究这 2 个基因间的潜在关联，在原始训练集和外部测试集上进行了 Pearson 相关性分析，结果显示 IRF7 与 ZNF204P 表达水平在 2 个独立数据集中均呈显著负相关（图 9-G、H）。

为评估 ZNF204P 和 IRF7 作为诊断生物标志物的潜在价值，将其诊断性能 ROC 曲线与已知的 IPF 相关标志物 MMP7 和 SLAIN1 进行了比较。在相同的训练集与外部测试集上，ZNF204P 与 IRF7 模型的 ROC 曲线显著优于现有标志物，表明其具有作为新型、稳健诊断生物标志物的潜力与补充价值（图 9-I、J）。

2.10 靶向中药预测与分析

将 IPF 相关的 7 个核心靶基因提交至 CTD 数据库进行检索，共获得 1 518 种相关化合物，各基因对应的化合物数量分别为：ISOC1（127 种）、IFI6（173 种）、CTSK（397 种）、TTC32（69 种）、IRF7（357 种）、ISG15（368 种）、ZNF204P（27 种）。经筛选并去重后，得到以槲皮素、白藜芦醇、二氧化硅、染料木黄酮及睾酮等为代表的潜在活性化合物。进一步借助 ITCM 数据库筛选潜在中药，并进行标准化处理、删除重复记录，获得 385 种潜在相关中药，其中高频药物包括栀子、肉苁蓉、蒲公英、大枣、柴胡、金银花、黄芪、沙棘和沙苑子等（图 10-A）。将这些中药导入中医药数据挖掘平台分析其用药规律，结果显示药性以寒、温、平居多（图 10-B）；药味以苦、甘、辛为主（图 10-C）；归经主要集中在肝、肺、胃、脾、肾经（图 10-D）。根据《中药学》分类，这些中药主要涉及 21 类，其中以清热药、补虚药、活血化瘀药、解表药和利水渗湿药为最主要类别（图 10-E）。

2.11 中药活性成分与和核心靶点的分子对接验证

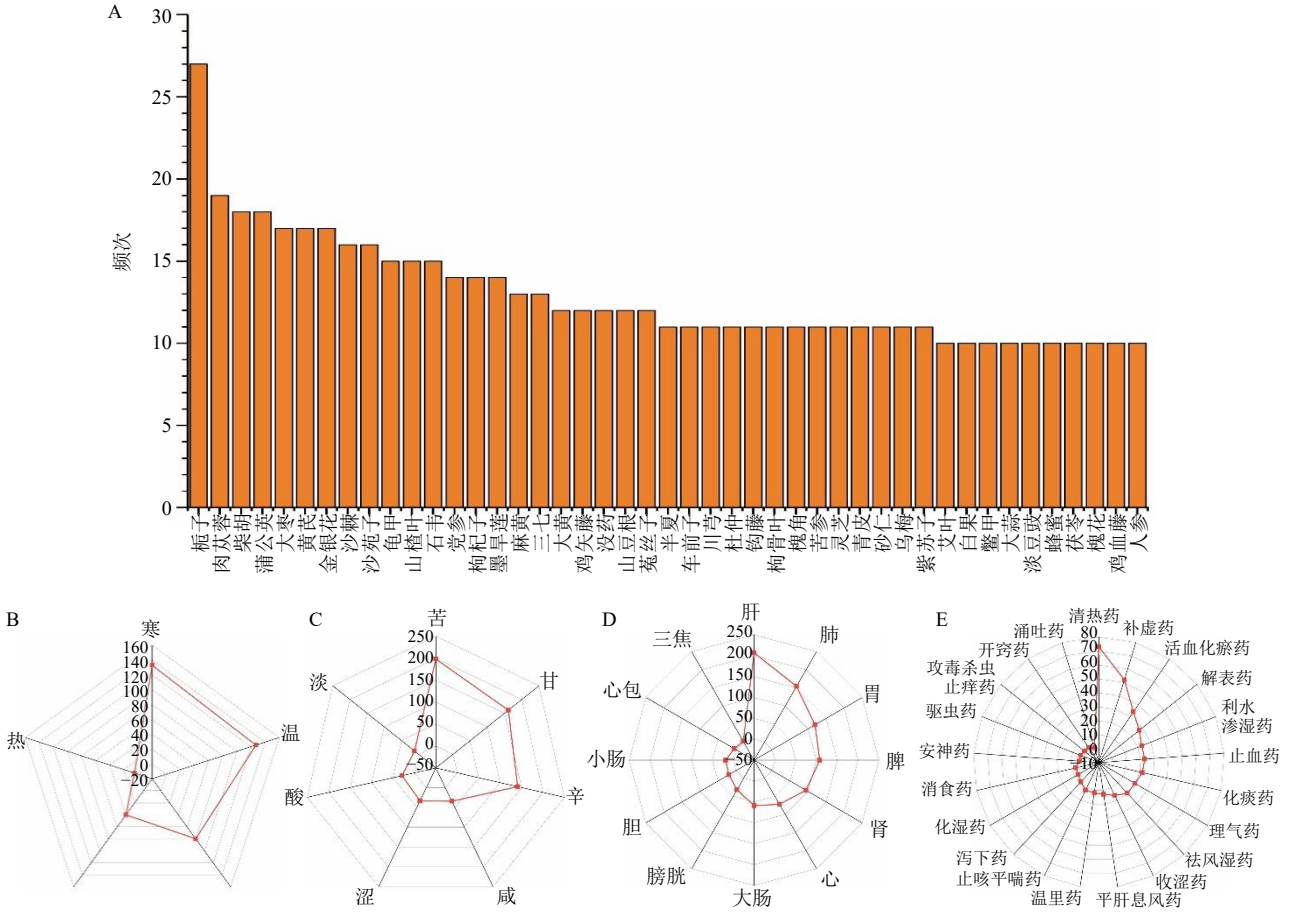
选择频次出现最多的前 3 个中药，基于满足 TCMSP 数据库的筛选条件以及相关文献证据，最终确定了 28 种核心中药的关键成分。柴胡 11 种：草澄茄素（cubebin）、异鼠李素（isorhamnetin）、岗松素（areapillin）、长管贝壳杉素 A（longikaurin A）、槲皮素（quercetin）、豆甾醇（stigmasterol）、山柰酚（kaempferol）α-菠甾醇（α-spinasterol）、乙酸亚油酯（linoleyl acetate）、曲克芦丁（troxerutin）和矮牵牛素（petunidin）；栀子 12 种：藏花酸（crocetin）、3-表齐墩果酸（3-epioleanolic acid）、欧前胡素（ammidin）、苏丹 III（sudan III）、亚油酸乙酯（mandenol）、异欧前胡素（isoimperatorin）、伞房花耳草素（corymbosin）、槲皮素、β-谷甾醇（β-sitosterol）、山柰酚、3-甲基山柰酚（3-methylkempferol）和豆甾醇；肉苁蓉 5 种：花生四烯酸（corymbosinarachidonate）、槲皮素、β-谷甾醇、苏齐内酯（suchilactone）和杨安木脂素（yangambin）。

靶点蛋白对应的配体结构则从 PubChem 平台获取其平面二维结构，并利用 Chem3D 22.0.0 软件中的 MM2 力场进行能量最小化处理，以优化构象用于后续对接分析。针对目标蛋白 ISOC1（UniProt ID: Q96CN7）、IFI6（P09912）、CTSK（P43235）、TTC32（Q5I0X7）、IRF7（Q92985）和 ISG15（P05161），依据其 UniProt 标识在 PDB 数据库中查询晶体结构信息。其中，CTSK、IRF7 和 ISG15 对应的 PDB ID（分辨率）分别为 7QBM（1.88 Å，1 Å＝0.1 nm）、2O61（2.8 Å）和 1Z2M（2.5 Å）。对于未能获得实验晶体结构的 ISOC1、IFI6 和 TTC32，其三维构象通过 AlphaFold 数据库预测获取；而 ZNF204P 由于缺乏可用结构信息，未纳入后续分析。

基于分子对接能量参数结合强度的分类分为 3 个评估标准：−4.25 kcal/mol（确定，1 kcal＝4.2 kJ）、−5 kcal/mol（良好）和−7 kcal/mol（强）。图 11-A～F 结果显示，柴胡、栀子及肉苁蓉的活性成分均能与核心靶点形成稳定结合，其结合能均优于−4.3 kcal/mol。值得注意的是，栀子中的苏丹 III 与 IRF7、柴胡中的岗松素与 CTSK，以及肉苁蓉中的槲皮素与 CTSK 之间的结合能均为最低。上述结果表明，这 3 种中药的活性成分很可能通过直接作用于关键靶点基因，从而在 IPF 的发生与发展过程中发挥调控作用。

3 讨论

本研究通过整合孟德尔随机化、转录组差异分



A-频次≥2 的潜在干预中药；B-中药四气分布雷达图；C-中药五味分布雷达图；D-中药归经分布雷达图；E-中药分类分布雷达图。

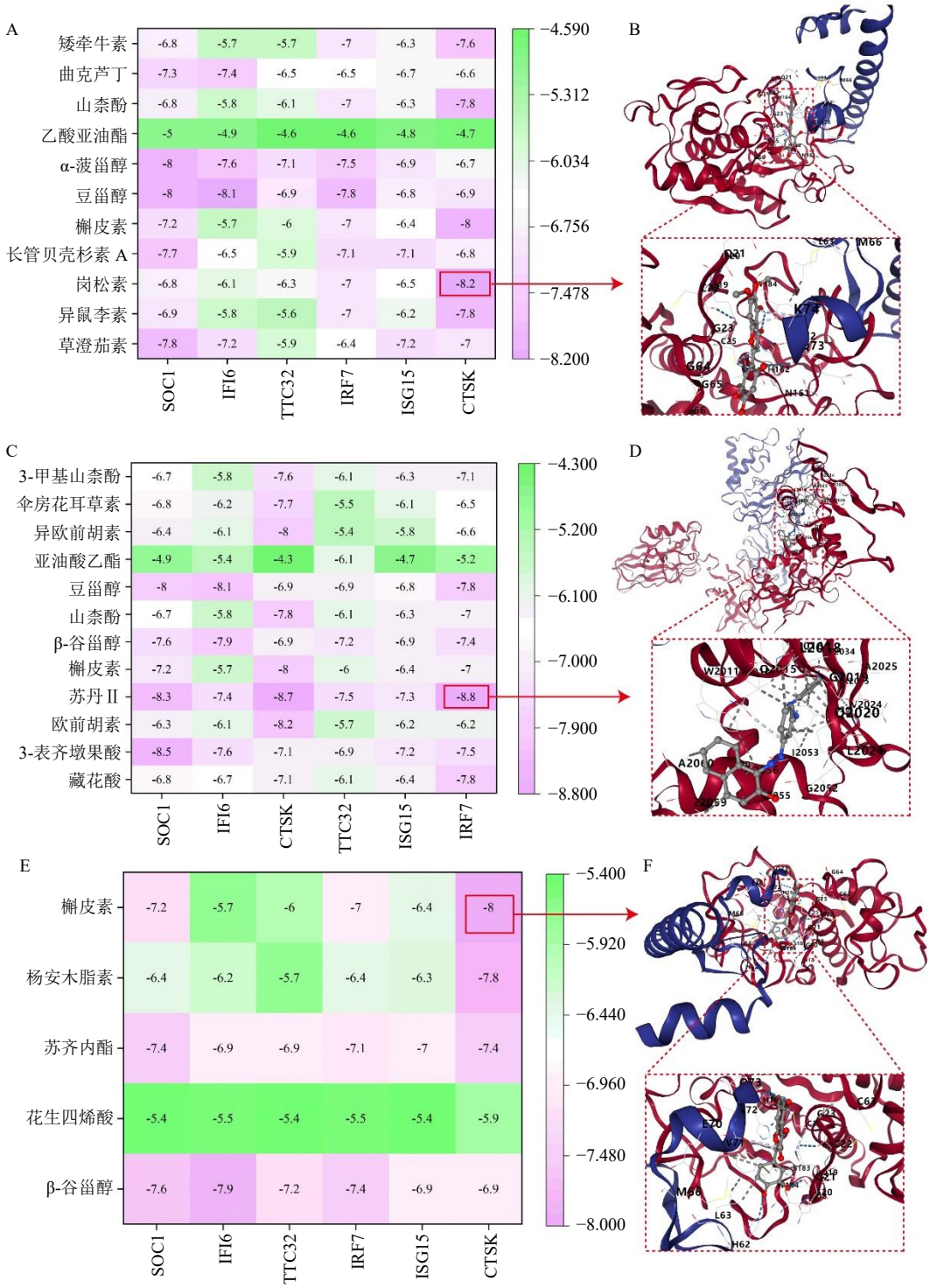
A-potential intervention traditional Chinese medicines (TCMs) with a frequency of ≥ 2; B-radar chart of the distribution of the four *qi* in TCMs; C-radar chart of the distribution of the five ingredients in TCMs; D-radar chart of the meridian distribution of TCMs; E-radar chart of classification and distribution of TCMs.

图 10　靶向中药预测分析

Fig. 10　Prediction and analysis of targeted traditional Chinese medicines

析和机器学习算法，系统筛选出 7 个与 IPF 存在因果关联的关键基因（IRF7、TTC32、IFI6、ISG15、ZNF204P、ISOC1 和 CTSK），并进一步确定 ZNF204P 与 IRF7 为最具诊断潜力的特征基因。这一发现为 IPF 的早期诊断和靶向治疗提供了新的候选分子。功能富集分析显示，上述基因主要富集于干扰素相关信号通路（如 I 型干扰素信号通路、RIG-I 样受体和 Toll 样受体通路），提示先天免疫应答失调在 IPF 发病中具有重要作用。既往研究表明，I 型干扰素信号通路的异常激活可促进肺上皮细胞衰老和成纤维细胞活化，加速纤维化进程[15-16]。

免疫微环境分析显示，IPF 患者肺组织中 M0/M1 巨噬细胞比例下降，活化肥大细胞比例升高。巨噬细胞表型转换是纤维化进程的关键环节，炎症阶段以 M1 为主，随着疾病进展向促纤维化的 M2 表型转变[17]。这一发现与既往研究 H1N1 流感病毒感染后肺损伤模型数据一致，M0 巨噬细胞首先在损伤部位极化为 M1 表型以促进急性炎症反应，随后可进一步向 M2 表型转变，促进肌成纤维细胞分化，最终推动 PF 发展[18]。因此，M0 与 M1 巨噬细胞比例下降与 IPF 向促纤维化状态的病理转变相符。肥大细胞作为炎症反应与组织重塑的关键调节因子，其活化与多器官纤维化密切相关[19-20]。本研究证实 IPF 中肥大细胞异常活化，与既往研究结果一致[21-22]，提示肥大细胞可能通过募集免疫细胞、促进成纤维细胞增殖、细胞外基质（extracellular matrix，ECM）重塑及降低上皮-间质转化（epithelial-mesenchymal transition，EMT）相关蛋白及白细胞介



A-柴胡活性成分与核心靶点的分子对接热图；B-CTSK 与岗松素的分子对接图；C-栀子活性成分与核心靶点的分子对接热图；D-IRF7 与苏丹 III 的分子对接图；E-肉苁蓉活性成分与核心靶点的分子对接热图；F-CTSK 与槲皮素的分子对接图。

A-molecular docking heatmap of the active components of *Radix Bupleuri* and core targets; B-molecular docking map of CTSK and areapillin; C-molecular docking heat map of the active components of *Gardeniae Fructus* and core targets; D-molecular docking map of IRF7 and sudan III; E-molecular docking heat map of the active components of *Cistanches Herba* and core targets; F-molecular docking diagram of CTSK and quercetin.

图 11　分子对接结果

Fig. 11　Molecular docking results

素-13（interleukin-13，IL-13）表达参与 IPF 进展[19,23-24]。深入解析巨噬细胞亚群失衡与肥大细胞活化机制，有望为开发新型免疫调节疗法提供线索。

单细胞分析进一步显示，上述关键基因主要富

集于上皮细胞亚群，与 IPF 作为“上皮细胞驱动的纤维化疾病”这一认识高度吻合。II 型肺泡上皮细胞（type II alveolar epithelial cells，AT2）功能障碍是 IPF 发生发展的核心环节[25-26]。AT2 细胞反复微损伤可启动氧化应激、内质网应激等病理过程，导致细胞衰老及分泌衰老相关表型因子，进而促进成纤维细胞活化与 ECM 过度沉积[27-29]。值得关注的是，Lu 等[30]构建的可吸入基因编辑纳米平台通过靶向下调促衰老基因赖氨酸乙酰转移酶 7［K(lysine) acetyltransferase 7，KAT7］，成功减少衰老相关分泌表型因子产生，逆转 AT2 细胞衰老特征，减轻炎症与纤维化进展，最终促进 IPF 缓解。因此，靶向 AT2 细胞衰老过程可能成为延缓 IPF 进展的新型治疗策略。

为确定最有效的调控因子，本研究联合 LASSO、随机森林及支持向量机递归特征消除 3 种算法进行关键基因筛选，最终确定 ZNF204P 与 IRF7 为最重要的特征基因。ZNF204P 属于锌指蛋白相关非编码 RNA 家族。尽管不具备编码蛋白质能力，其可通过表观遗传过程参与基因表达调控，其中作为竞争性内源 RNA 发挥“分子海绵”功能，通过吸附 microRNA 间接影响下游靶基因。既往研究发现该基因已被确定为多种癌症[31]和精神分裂症[32]的潜在预后生物标志物或易感位点。Hwang 等[33]研究表明 ZNF204P 缺失会导致细胞活力下降并对迁移侵袭能力产生负面影响；在分子层面，胞质中 ZNF204P 与关键干细胞转录因子八聚体结合转录因子 4（octamer-binding transcription factor 4，OCT4）、SRY 盒转录因子 2（SRY-box transcription factor 2，SOX2）共享 miRNA-145-5p 结合位点，其下调可引起 OCT4 与 SOX2 水平相应降低，提示其参与调控 miRNA-145-5p/干细胞转录因子通路。Rastegari 等[34]报道乳腺癌样本中 ZNF204P 表达显著降低，而较高表达水平与较好患者预后相关；该基因构建的风险模型能强力预测患者生存率。此外，还有研究发现结直肠癌与胃癌组织中 ZNF204P 表达显著下降[34]。肝细胞癌研究表明敲低 ZNF204P 可抑制肿瘤细胞增殖、迁移与侵袭，其机制可能与 ZNF204P 在胞质中与多能性调控因子 OCT4、SOX2 共享 miRNA 结合位点有关—ZNF204P 通过竞争性结合 miR-145-5p 激活 OCT4/SOX2 信号轴促进肿瘤发生发展[33]。本研究结果与前期发现一致，ZNF204P 在多种疾病中表达上调，敲低该基因可有效抑制细胞增殖、迁移与侵袭，从而对 IPF 产生治疗作用。因此，靶向 ZNF204P 及其信号网络可能为 IPF 治疗提供新策略。

IRF7 是 IFN-I 的核心转录调控因子，临床实践中广泛用作抗病毒药物。Zhang 等[35]研究表明 IRF7 作为人类免疫缺陷病毒 1 型（human immunodeficiency virus type 1，HIV-1）的宿主调控因子，可通过调节细胞相关 DNA 水平及 RNA/DNA 值影响病毒库规模与转录活性，进而参与治疗反应，提示 IRF7 可能成为 HIV-1 治疗的潜在靶点。研究报道叉头框蛋白（forkhead box O3，FOXO3）/IRF7 通路上调可增强 I 型干扰素应答，从而抑制肠道病毒 71 型复制，为新型抗病毒策略开发提供方向[36]。此外，IRF7 已被探索作为 IPF[37]与骨髓纤维化[38]的潜在治疗靶点。研究表明 IRF7 参与纤维化进程，部分机制与转化生长因子-β（transforming growth factor-β，TGF-β）信号通路调控相关，系统性硬化症患者皮肤组织中 IRF7 表达水平较正常对照显著升高，且 IRF7 能够与 Smad3 蛋白相互作用增强 TGF-β 介导的促纤维化效应[39]。在肾纤维化模型中，TGF-β/Smad3 通路正向调控 IRF7 表达，通过 IRF7-组织蛋白酶 S 信号通路促进巨噬细胞向肌成纤维细胞转分化，从而加剧肾纤维化进展[40]。相反，IRF7 也可能通过改变炎症微环境影响纤维化结局。Park 等[41]证实抑制 I 型干扰素信号通路可影响巨噬细胞信号转导和转录激活因子 3（signal transducer and activator of transcription 3，STAT3）信号传导，增强抗炎反应并促进组织修复再生，最终显著降低纤维化程度。Chen 等[42]报道间歇性缺氧-复氧刺激通过缺氧反应因子-1α（hypoxia-inducible factor-1α，HIF1A）触发特定 miRNA 表达，促进人气道上皮细胞中 NF-κB 依赖性促炎促纤维化标志基因上调，同时下调 IRF5 和 IRF7 介导的 I/II 型干扰素表达，提示 IRF 通路可能成为干预气道炎症与肺纤维化的潜在靶点。综上所述，基于前述研究及本研究发现，纤维化组织中 IRF7 显著上调，靶向调控 IRF7 信号网络有望成为 IPF 治疗新策略。

IPF 在中医学中可归属于“肺痿”“肺痹”等范畴。其病机特点多属本虚标实、虚实夹杂。肺虚为其本，痰、瘀、热阻滞肺络为其标[43]。病理关键在于“肺虚络痹”，多因肺气不足，宣降失常，气机郁滞，津液输布受阻而聚湿成痰；气虚不能推动血行，血流涩滞则凝结成瘀。痰瘀久结，蕴而化热，痹阻

肺络，致使肺体失养、肺用受损，形成进行性加重的纤维化病变[44]。诚如《张氏医通》所言：“肺失所养，转枯转燥，然后成之。于是肺火日炽，肺热日深，肺中小管日窒，咳声以渐不扬。”《金匮要略·肺痿肺痈咳嗽上气病脉证治七》载：“肺痿之病，从何得之？……热在上焦者，因咳为肺痿。”说明内热灼伤肺津，可导致肺叶枯萎，发为本病。临床观察可见，IPF 在起病或急性加重阶段，常因外邪引动，表现为痰热壅肺或痰瘀热互结的实证[45]。《素问·痿论》中“肺热叶焦”之论，进一步提示热邪是肺痿形成的重要病机基础。因此，热邪不仅是 IPF 发生发展的致病因素，也是在病理过程中产生的重要病理性产物。本研究筛选出与 IPF 核心靶点相关的潜在中药 385 味，其中高频药物包括栀子、肉苁蓉、蒲公英、大枣、柴胡、黄芪、沙棘、沙苑子等。药物类别以清热药、补虚药、活血化瘀药为主，性味归经以苦、甘、辛及肝、肺、脾、胃、肾经多见。清热药可清解肺中热邪；补虚药旨在益气养阴、补益肺脾肾，扶助正气以固其本；活血化瘀药针对“久病入络”之瘀血阻滞，可能干预纤维化形成的关键环节。从药性分布来看，苦寒之品功擅清化热痰；甘味药物多为补益之品，尤宜用于改善 IPF 本虚的病理基础；辛味药物善行气活血、宣通肺络，针对“瘀血阻络”核心病机形成干预。归经方面，肝、肺、脾、胃、肾经药物分布最为集中，体现了从疏肝理气、培土生金、金水相生等多角度调治 IPF 的配伍思路[43]。以上药物分布特征与 IPF“虚、痰、瘀、热”交织的核心病机高度吻合。

现代药理研究为上述高频中药的抗纤维化作用提供了证据。栀子所含的栀子苷具有抗炎与抗氧化特性，研究表明其可能通过抑制 TGF-β/Smad 及 p38 丝裂原活化蛋白激酶（mitogen-activated protein kinase，MAPK）信号通路，减轻博来霉素诱导的 IPF 炎症反应[46]。肉苁蓉中提取的苯乙醇苷类成分松果菊苷，在免疫调节、抗氧化及抗衰老方面表现出活性，Zhang 等[47]发现其可通过抑制 M2 型巨噬细胞极化，下调 Janus 激酶 2（Janus kinase 2，JAK2）/STAT3 通路，从而发挥抗纤维化效应。黄芪作为补气要药，长于补益肺脾，相关实验提示黄芪注射液可能通过上调 miR-29a-3p 抑制 I 型胶原蛋白 α1 链（collagen type I alpha 1 chain，COL1A1）表达，减少肺组织胶原沉积，改善肺纤维化大鼠的肺功能[48]。研究发现，柴胡含药血清能够抑制 TGF-β1 诱导的 HFL1 细胞增殖及肌成纤维细胞转化，并促进细胞凋亡，其机制可能与调控 Smad3/脑组织富含的 Ras 同源物（Ras homolog enriched in brain，Rheb）轴有关[49]。蒲公英甾醇是从蒲公英提取的有效成分，有研究发现，蒲公英的有效成分蒲公英甾醇，则可通过调节 Wnt/β-连环蛋白（β-catenin）信号通路，抑制 TGF-β1 诱导的 EMT，降低纤维化标志物与炎症因子表达，从而减缓 IPF 进展[50]。这为中医药辨证论治 IPF 提供了新的线索与候选药物参考。然而，网络预测结果仍需后续通过体外体内实验进行严格验证，其具体的药效物质基础、最优配伍规律及深层作用机制，有待进一步深入探索。

尽管在理解 IPF 分子机制方面已取得显著进展，但本研究仍存在若干固有局限性。首先，数据来源于公共数据库，可能存在队列异质性和偏倚。其次，孟德尔随机化分析基于欧洲人群，可能限制研究结论在其他族裔中的适用性，需进一步验证方可推广。本研究的孟德尔随机化分析在工具变量筛选中，未执行针对已知 IPF 风险因素（如吸烟、身体质量指数等）的 SNP 关联性排除步骤，这构成了一个方法学上的局限性。再次，对单细胞转录组测序技术的依赖存在局限，该方法可能无法完全捕捉 IPF 复杂免疫微环境中的细胞间相互作用网络，可能导致测量误差。此外，从基因到中药的关联预测完全基于 CTD、TCMSP 等公共数据库的文献挖掘，这些关联可能受到已发表文献范围和偏倚的影响。分子对接模拟所提示的成分-靶点结合可能性，仅为计算机预测结果，其结合能与实际的生物活性或体内药效之间不能直接划等号，无法替代必要的体外与体内实验验证。最后，尽管本研究初步揭示 ZNF204P 与 IRF7 在 IPF 中的潜在诊断价值，仍需在真正的大型前瞻性多中心队列研究中验证二者诊断功能是否延伸至早期 IPF 识别或 IPF 与其他间质性肺疾病的鉴别。未来将联合临床机构检测患者样本中这 2 个基因的表达评估其作为诊断生物标志物的临床可行性，并通过体外培养肺成纤维细胞或上皮细胞，验证预测所得的高频活性成分或中药能否调节干扰素调节因子 7（Interferon Regulatory Factor，IRF7）/ZNF204P 表达及纤维化相关表型，推动基础研究向临床转化。

4 结论

通过整合遗传学分析、免疫学评估与计算生物学技术，本研究成功筛选出 7 个与 IPF 遗传学存在

因果关联的重要基因：IRF7、TTC32、IFI6、ISG15、ZNF204P、ISOC1 和 CTSK，系统阐明了这些基因在 IPF 发生发展中的关键作用。进一步利用机器学习方法进行基因筛选，最终确定 ZNF204P 与 IRF7 为最具潜力的候选生物标志物或治疗靶点，凸显了其作为 IPF 临床诊断标志物或治疗干预点的应用前景。此外，通过计算数据库挖掘，初步预测出栀子、柴胡、肉苁蓉等可能干预上述核心基因的潜在中药，为探索 IPF 的中医药治疗策略提供了新的研究方向与假设。尽管本研究尚处于初步阶段，但研究结果为开发 IPF 新型诊断工具及个体化精准治疗策略提供了重要理论基础，并为深入探究该疾病的分子机制奠定了坚实基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] Luo C, Wu X H, Zhang S P, *et al*. Cuproptosis: A novel therapeutic mechanism in lung cancer [J]. *Cancer Cell Int*, 2025, 25(1): 231.

[2] Denis A, Tsiri P, Guiot J, *et al*. A new era in the treatment of progressive fibrosing interstitial lung diseases [J]. *Breathe*, 2025, 21(2): 240259.

[3] Wang X Y, Wu X Y, Zhou L Q, *et al*. Demethyleneberberine ameliorates pulmonary fibrosis by inhibiting the NLRP3 inflammasome and the epithelial–mesenchymal transition [J]. *Int Immunopharmacol*, 2025, 161: 115003.

[4] Maher T M, Bendstrup E, Dron L, *et al*. Global incidence and prevalence of idiopathic pulmonary fibrosis [J]. *Respir Res*, 2021, 22(1): 197.

[5] Majewski S, Górska K, Lewandowska K B, *et al*. Real-world treatment persistence and predictive factors for discontinuation of antifibrotic therapies in patients with idiopathic pulmonary fibrosis: A post-hoc analysis of two multicenter observational cohort studies in Poland [J]. *Front Pharmacol*, 2025, 16: 1586197.

[6] Jones N, Rahar B, Bernau K, *et al*. Mechanisms on how matricellular microenvironments sustain idiopathic pulmonary fibrosis [J]. *Int J Mol Sci*, 2025, 26(11): 5393.

[7] Divolis G, Synolaki E, Tringidou R, *et al*. Transcriptomic analysis reveals shared deregulated neutrophil responses in COVID-19 and idiopathic pulmonary fibrosis [J]. *Respir Res*, 2025, 26(1): 213.

[8] Pan H, Jing C Q. Exploring druggable targets and inflammation-mediated pathways in cancer: A Mendelian randomization analysis integrating transcriptomic and proteomic data [J]. *Inflamm Res*, 2025, 74(1): 46.

[9] Chen Z, Tang M Y, Wang N, *et al*. Genetic variation reveals the therapeutic potential of BRSK2 in idiopathic pulmonary fibrosis [J]. *BMC Med*, 2025, 23(1): 22.

[10] Shi Y, Chen S, Zhou Z K, *et al*. Causal effects between genetically determined human serum metabolite levels on the risk of idiopathic pulmonary fibrosis: A Mendelian randomization study [J]. *Clin Respir J*, 2025, 19(6): e70087.

[11] Gong P, Lu Y M, Chai X, *et al*. Exploring the causal relationship between immune cells and idiopathic pulmonary fibrosis: A Mendelian randomization analysis [J]. *J Clin Lab Anal*, 2025, 39(8): e70026.

[12] Fan Q L, Meng Y, Nie Z H, *et al*. The role of inflammatory factors in mediating the causal effects of type 1 diabetes mellitus on idiopathic pulmonary fibrosis: A two-step Mendelian randomization study [J]. *Medicine*, 2025, 104(4): e41320.

[13] 吴宣諭, 肖祥, 陈嘉靖, 等. 基于“肠-肺轴”探讨肠道菌群与特发性肺纤维化的遗传因果关联及干预中药预测 [J]. 中草药, 2024, 55(17): 5921-5937.

[14] Zhu Z H, Zhang F T, Hu H, *et al*. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets [J]. *Nat Genet*, 2016, 48(5): 481-487.

[15] Guo-Parke H, Cappa O, Linden D A, *et al*. IFN-mediated bronchial epithelium cellular senescence in chronic obstructive pulmonary disease [J]. *Am J Respir Cell Mol Biol*, 2025, 73(6): 871-883.

[16] Yin Y, Zhao S J, Li W, *et al*. *In situ* reprogramming of fibroblasts into antigen-presenting pseudo-dendritic cells via IFN-β-engineered protoplast-derived exosomes delivered by microneedle arrays to enhance adaptive immunity [J]. *Theranostics*, 2025, 15(17): 9179-9199.

[17] Kim J Y, Cho D W, Choi J Y, *et al*. CXCL11 reprograms M2-biased macrophage polarization to alleviate pulmonary fibrosis in mice [J]. *Cell Biosci*, 2024, 14(1): 140.

[18] Wang C G, Liu S J, Li C Y, *et al*. Monitoring the cascade of monocyte-derived macrophages to influenza virus infection in human alveolus chips [J]. *ACS Appl Mater Interfaces*, 2024, 16(44): 60045-60055.

[19] Luo Y, Zhang H K, Yu J, *et al*. Stem cell factor/mast cell/CCL2/monocyte/macrophage axis promotes Coxsackievirus B3 myocarditis and cardiac fibrosis by

increasing Ly6C^{high} monocyte influx and fibrogenic mediators production [J]. *Immunology*, 2022, 167(4): 590-605.

[20] Pimpalwar N, Celik S, Karbalaei Sadegh M, *et al*. Analysis of genetic variant associated with heart failure mortality implicates thymic stromal lymphopoietin as mediator of strain-induced myocardial fibroblast-mast cell crosstalk and fibrosis [J]. *FASEB J*, 2024, 38(4): e23510.

[21] Andersson C K, Andersson-Sjöland A, Mori M, *et al*. Activated MCTC mast cells infiltrate diseased lung areas in cystic fibrosis and idiopathic pulmonary fibrosis [J]. *Respir Res*, 2011, 12(1): 139.

[22] Tan C, Zhou H, Xiong Q F, *et al*. Cromolyn sodium reduces LPS-induced pulmonary fibrosis by inhibiting the EMT process enhanced by MC-derived IL-13 [J]. *Respir Res*, 2025, 26(1): 3.

[23] Cardamone C, Parente R, De Feo G, *et al*. Mast cells as effector cells of innate immunity and regulators of adaptive immunity [J]. *Immunol Lett*, 2016, 178: 10-14.

[24] Galli S J, Gaudenzio N, Tsai M. Mast cells in inflammation and disease: Recent progress and ongoing concerns [J]. *Annu Rev Immunol*, 2020, 38: 49-77.

[25] Zepp J A, Morrisey E E. Cellular crosstalk in the development and regeneration of the respiratory system [J]. *Nat Rev Mol Cell Biol*, 2019, 20(9): 551-566.

[26] Liberti D C, Kremp M M, Liberti W A, *et al*. Alveolar epithelial cell fate is maintained in a spatially restricted manner to promote lung regeneration after acute injury [J]. *Cell Rep*, 2021, 35(6): 109092.

[27] Gonzalez-Gonzalez F J, Chandel N S, Jain M, *et al*. Reactive oxygen species as signaling molecules in the development of lung fibrosis [J]. *Transl Res*, 2017, 190: 61-68.

[28] Liu Q, Ren Y P, Jia H M, *et al*. Vanadium carbide nanosheets with broad-spectrum antioxidant activity for pulmonary fibrosis therapy [J]. *ACS Nano*, 2023, 17(22): 22527-22538.

[29] Zmijewski J W, Thannickal V J. Autophagy in idiopathic pulmonary fibrosis: Predisposition of the aging lung to fibrosis? [J]. *Am J Respir Cell Mol Biol*, 2023, 68(1): 3-4.

[30] Lu Q L, Ye C W, Mao W, *et al*. Targeting senescent alveolar type 2 cells with a gene-editable FePt dual-atom catalyst for mitigating idiopathic pulmonary fibrosis [J]. *ACS Nano*, 2025, 19(25): 23162-23176.

[31] Lin H J, Qiu X K, Zhang B, *et al*. Identification of the predictive genes for the response of colorectal cancer patients to FOLFOX therapy [J]. *OncoTargets Ther*, 2018, 11: 5943-5955.

[32] Shi J X, Levinson D F, Duan J B, *et al*. Common variants on chromosome 6p22.1 are associated with schizophrenia [J]. *Nature*, 2009, 460(7256): 753-757.

[33] Hwang J H, Lee J, Choi W Y, *et al*. ZNF204P is a stemness-associated oncogenic long non-coding RNA in hepatocellular carcinoma [J]. *BMB Rep*, 2022, 55(6): 281-286.

[34] Rastegari M, Sazegar H, Doosti A. Prognostic significance of *CHCHD2P9* and ZNF204P in breast cancer: Exploring their expression patterns and associations with malignancy-related genes [J]. *Mol Biol Rep*, 2024, 51(1): 707.

[35] Zhang Z H, Trypsteen W, Blaauw M, *et al*. IRF7 and RNH1 are modifying factors of HIV-1 reservoirs: A genome-wide association analysis [J]. *BMC Med*, 2021, 19(1): 282.

[36] Yang D K, Wang X W, Gao H L, *et al*. Downregulation of miR-155-5p facilitates enterovirus 71 replication through suppression of type I IFN response by targeting FOXO3/IRF7 pathway [J]. *Cell Cycle*, 2020, 19(2): 179-192.

[37] King T E Jr, Albera C, Bradford W Z, *et al*. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): A multicentre, randomised, placebo-controlled trial [J]. *Lancet*, 2009, 374(9685): 222-228.

[38] Sørensen A L, Mikkelsen S U, Knudsen T A, *et al*. Ruxolitinib and interferon-α2 combination therapy for patients with polycythemia vera or myelofibrosis: A phase II study [J]. *Haematologica*, 2020, 105(9): 2262-2272.

[39] Wu M H, Skaug B, Bi X J, *et al*. Interferon regulatory factor 7 (IRF7) represents a link between inflammation and fibrosis in the pathogenesis of systemic sclerosis [J]. *Ann Rheum Dis*, 2019, 78(11): 1583-1591.

[40] Ren C N, Mi T, Zhang Z X, *et al*. Targeting interferon regulatory factor 7 alleviates renal fibrosis by inhibiting macrophage-to-myofibroblast transition [J]. *Life Sci*, 2025, 376: 123755.

[41] Park S J, Garcia Diaz J, Comlekoglu T, *et al*. Type I IFN receptor blockade alleviates liver fibrosis through macrophage-derived STAT3 signaling [J]. *Front Immunol*, 2025, 16: 1528382.

[42] Chen S T, Jheng C Y, Lee Y C, *et al*. Intermittent hypoxia-reoxygenation-induced miRNAs inhibit expression of *IRF* and interferon genes but activate NF-κB and expression of pulmonary fibrosis markers in human small airway epithelial cells [J]. *Life Sci*, 2025, 370: 123569.

[43] 叶远航, 罗成, 宁博, 等. 基于“虚气留滞”探讨上皮间质转化对特发性肺纤维化的影响 [J]. 国际中医中药杂志, 2024, 46(12): 1543-1548.

[44] 茆春阳, 杜燕, 雷伟伟, 等. 基于“燥伤肺络”理论探析肺纤维化 [J]. 陕西中医, 2025, 46(4): 522-526.

[45] 刘学, 李博洋, 吴凡, 等. “瘀毒”理论指导下从“肺热络瘀”论治特发性肺纤维化 [J]. 山东中医杂志, 2025, 44(2): 141-144.

[46] Yin J B, Wang Y X, Fan S S, *et al*. Geniposide ameliorates bleomycin-induced pulmonary fibrosis in mice by inhibiting TGF-β/Smad and p38MAPK signaling pathways [J]. *PLoS One*, 2024, 19(9): e0309833.

[47] Zhang Y F, Fan L M, Wang M N. Echinacoside ameliorates bleomycin-induced idiopathic pulmonary fibrosis by regulating macrophage polarization [J]. *J Mol Histol*, 2025, 57(1): 4.

[48] 周飘, 杜婧, 吴程, 等. 黄芪注射液通过 miR-29a-3p/COL1A1 信号轴干预肺纤维化的机制研究 [J]. 中华中医药学刊, 2023, 41(3): 107-110.

[49] 李妲, 沈枭, 吴趋荟, 等. 柴胡含药血清通过 Smad3/Rheb 轴调节 HFL1 细胞凋亡和肌成纤维细胞转化 [J]. 中国实验方剂学杂志, 2023, 29(11): 89-96.

[50] 邱慧萍, 樊丹丹, 张莉. 蒲公英甾醇通过调节 Wnt/β-catenin 信号通路抑制转化生长因子 β1 对肺纤维化的改善作用 [J]. 中国当代医药, 2024, 31(35): 16-22.

[责任编辑　潘明佳]