

“菌群-免疫-代谢”轴调控特发性肺纤维化的遗传证据：基于多组学孟德尔随机化的遗传标志物识别与干预中药预测

罗 成^{1,2,3}, 蒋子艺^{1,2}, 谭钧文^{1,2}, 叶远航⁴, 宁 博⁵, 宋幸铃⁶, 王 飞², 叶 莹¹, 柯 佳^{7,8,9,10}, 覃琬婷^{1,2*}

1. 成都中医药大学, 四川 成都 610075

2. 成都中医药大学附属医院, 四川 成都 610075

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, 广东 广州 510006

6. 成都中西医结合医院, 四川 成都 610041

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 基于多组学孟德尔随机化（Mendelian randomization, MR）框架，评估肠道菌群、免疫细胞、血清代谢物与特发性肺纤维化（idiopathic pulmonary fibrosis, IPF）的因果关联，筛选核心因果基因并预测干预中药及化合物，为 IPF 早期诊断与中药干预提供遗传学依据。方法 提取肠道菌群、血清代谢物及免疫细胞的遗传工具变量（instrumental variables, IVs），通过两样本 MR 分析其与 IPF 的因果关联；注释 IVs 邻近基因获得“菌群-免疫-代谢”因果基因集，整合 IPF 转录组数据进行差异表达基因（differentially expressed genes, DEGs）分析和加权基因共表达网络分析（weighted gene co-expression network analysis, WGCNA）筛选候选因果基因；经机器学习构建诊断模型并筛选核心基因，分析免疫细胞浸润与单细胞 RNA 测序，通过 CTD 和 ITCM 数据库预测干预中药及化合物，并进行分子对接验证。结果 MR 分析鉴定出 2 个肠道菌群分类群（乳杆菌科为保护因素，变形菌门为风险因素）、19 个免疫细胞表型（9 个保护因素、10 个风险因素）及 26 种血清代谢物（12 个保护因素、14 个风险因素）与 IPF 存在显著因果关联（ $P < 0.05$ ）。整合分析获得 21 个候选因果基因，机器学习筛选出 5 个核心基因：衔接蛋白复合体 5 亚基 $\mu 1$ （adaptor protein complex 5 subunit mu 1, AP5M1）、性别决定区 Y 框转录因子 11（SRY-box transcription factor 11）、锌指 MIZ 型包含蛋白 2（zinc finger and MIZ-type containing 2, ZMIZ2）、脆性组氨酸三联体（fragile histidine triad, FHIT）和细胞周期蛋白依赖性激酶 6（cyclin dependent kinase 6, CDK6），均具潜在诊断价值。免疫浸润显示 IPF 组织中活肥大细胞和单核细胞增多，M0/M2 巨噬细胞及活化自然杀伤细胞减少；单细胞分析显示 FHIT 在多种细胞中富集程度最高。筛选出 150 种潜在干预中药，高频中药包括桑白皮、月季花、延胡索等；活性成分包括山柰酚、柚皮素、汉黄芩素、血根碱等。分子对接证实核心化合物可与靶点稳定结合。结论 揭示了肠道菌群、免疫细胞、血清代谢物与 IPF 的因果关联，筛选出 AP5M1、SOX11、ZMIZ2、FHIT、CDK6 作为 IPF 的潜在候选诊断标志物；预测出山柰酚、柚皮素、非洲防己碱等活性化合物及桑白皮、月季花等中药的潜在干预价值，为 IPF 诊疗提供遗传学证据与理论支撑。

关键词：特发性肺纤维化；孟德尔随机化；肠道菌群；免疫细胞；血清代谢物；诊断标志物；桑白皮；月季花；山柰酚；柚皮素；汉黄芩素；血根碱

中图分类号：Q811.4；R285 文献标志码：A 文章编号：0253-2670(2026)18-7205-21

DOI: 10.7501/j.issn.0253-2670.2026.18.011

Genetic evidence for “microbiota-immunity-metabolism” axis in regulating idiopathic pulmonary fibrosis: Identification of genetic markers based on multi-omics Mendelian randomization and prediction of traditional Chinese medicine

收稿日期：2026-04-13

基金项目：中国科协青年科技人才培养工程博士生专项计划项目；国家自然科学基金青年科学基金项目（82505509）；湖北省自然科学基金资助项目（2023AFD173）；湖北省中医药管理局中医药科研项目（ZY2025M026）；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（2024MD753905）；成都市卫生健康委员会-成都中医药大学委校院联合创新基金（WXLH202403266）

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}, JIANG Ziyi^{1, 2}, TAN Junwen^{1, 2}, YE Yuanhang⁴, NING Bo⁵, SONG Xingling⁶, WANG Fei², YE Ying¹, KE Jia^{7, 8, 9, 10}, QIN Wanting^{1, 2}

1. Chengdu University of Traditional Chinese Medicine, Chengdu 610075, China

2. Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610075, China

3. China Joint Graduate School of Chinese Medicine, Suzhou 215105, China

4. Pengzhou Third People's Hospital, Chengdu 611931, China

5. Guangzhou University of Chinese Medicine, Guangzhou 510006, China

6. Chengdu Integrated Traditional Chinese and Western Medicine Hospital, Chengdu 610041, 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 evaluate the causal associations of gut microbiota, immune cells and serum metabolites with idiopathic pulmonary fibrosis (IPF) based on the multi-omics Mendelian randomization (MR) framework, screen core causal genes, and predict intervening traditional Chinese medicine (TCM) and compounds, so as to provide genetic evidence for the early diagnosis and TCM intervention of IPF. **Methods** Genetic instrumental variables (IVs) of gut microbiota, serum metabolites and immune cells were extracted, and their causal associations with IPF were analyzed by two-sample MR. Adjacent genes of IVs were annotated to obtain the "microbiota-immunity-metabolism" causal gene set. Integrated with IPF transcriptome data, candidate causal genes were screened by differentially expressed genes (DEGs) analysis and weighted gene co-expression network analysis (WGCNA). Diagnostic models were constructed by machine learning to screen core genes. Immune cell infiltration and single-cell RNA sequencing profiles were analyzed. Potential intervening TCM and compounds were predicted via Comparative Toxicogenomics Database (CTD) and Integrated Traditional Chinese Medicine Database (ITCM), followed by molecular docking verification. **Results** Two-sample MR identified 2 gut microbiota taxa (*Lactobacillaceae* as protective factor, *Proteobacteria* as risk factor), 19 immune cell phenotypes (nine protective, ten risk) and 26 serum metabolites (12 protective, 14 risk) that were significantly causally associated with IPF ($P < 0.05$). A total of 21 candidate causal genes were obtained by integrated analysis, and 5 core genes were screened by machine learning: adaptor protein complex 5 subunit mu 1 (AP5M1), SRY-box transcription factor 11 (SOX11), zinc finger and MIZ-type containing 2 (ZMIZ2), fragile histidine triad (FHIT) and cyclin dependent kinase 6 (CDK6), all with potential diagnostic value. Immune infiltration showed increased activated mast cells and monocytes, and decreased M0/M2 macrophages and activated natural killer cells in IPF tissues. Single-cell analysis revealed the highest enrichment of FHIT in multiple cell types. A total of 150 kinds of potential intervention Chinese medicine were screened. The high frequency Chinese medicine included Sangbaipi (*Mori Cortex*), Yuejihua (*Rosae Chinensis Flos*), Yanhusuo (*Corydalis Rhizoma*), etc. The active ingredients included kaempferol, naringenin, wogonin, sanguinarine, etc. Molecular docking confirmed that the core compound could stably bind to the target. **Conclusions** This study reveals the causal associations of gut microbiota, immune cells and serum metabolites with IPF, and identifies AP5M1, SOX11, ZMIZ2, FHIT and CDK6 as potential candidate diagnostic biomarkers for IPF. It also predicts the potential intervention value of active compounds such as kaempferol, naringenin and palmitate, as well as TCM including *Morus alba* and *Rosa chinensis*, providing genetic evidence and theoretical support for the diagnosis and treatment of IPF.

Key words: idiopathic pulmonary fibrosis; Mendelian randomization; gut microbiota; immune cells; serum metabolites; diagnostic markers; *Mori Cortex*; *Rosae Chinensis Flos*; kaempferol; naringenin; wogonin; sanguinarine

特发性肺纤维化 (idiopathic pulmonary fibrosis, IPF) 是肺纤维化中预后较差的类型, 以成纤维细胞异常增殖和细胞外基质过度沉积为特征^[1]。根据 2019 年全球疾病负担相关研究, IPF 在间质性肺疾病中的占比约为 12%, 其全球发病率为 (0.3~0.9) / 万, 患病率为 (0.2~2.9) / 万^[2]。尽管诊疗技术不断

进步, 但 IPF 发病机制复杂、早期诊断困难, 现有抗纤维化药物无法根治, 总体预后仍不理想^[3]。核心难题在于区分致病驱动因素与伴随分子, 因此识别免疫靶点及预测性生物标志物对改善预后至关重要。

现有证据表明肠道微生物群与 IPF 进展密切

相关。Bai 等^[4]整合分析 IPF 患者及博来霉素诱导的肺纤维化 (pulmonary fibrosis, PF) 大鼠模型, 发现瘤胃球菌丰度升高、粪便中色氨酸与甘氨酸水平下降、血清诱导型一氧化氮合酶上升; 动物实验证实补充这 2 种氨基酸可缓解 PF。IPF 与免疫细胞的复杂交互作用密切相关, Jiang 等^[5]构建靶向 FAP1⁺肺泡成纤维细胞的细胞毒性调节性 T 细胞 (regulatory T cells, Tregs), 在模型小鼠中成功逆转胶原沉积与肺泡上皮重塑, 降低转化生长因子- β 1 (transforming growth factor- β 1, TGF- β 1) 和白细胞介素 17a (interleukin-17a, IL-17a) 水平, 为难治性 IPF 提供了新的免疫清除策略。此外, 微囊藻毒素-亮氨酸-精氨酸 (leucine arginine, LR) 可通过抑制内质网未折叠蛋白反应阻断巨噬细胞 M2 极化和 TGF- β 1/Smad 通路, 缓解 PF^[6]。血清代谢物同样具有重要价值, Cai 等^[7]采用非靶向代谢组学发现 IPF 患者血清中棕榈酰乙醇酰胺及 2-氨基-1,3,4-十八烷三醇显著升高, 受试者工作特征 (receiver operating characteristic, ROC) 及生存分析证实二者可作为诊断与预后候选标志物。Wu 等^[8]利用两样本孟德尔随机化评估 486 种血清代谢物与 IPF 的因果关联, 鉴定出 21 种循环代谢物, 其中次黄嘌呤呈保护效应, 1-亚油酰甘油磷酸胆碱增加风险, 从遗传层面支持血清代谢物作为 IPF 致病关键节点及早期筛查靶点。

尽管上述研究分别揭示了肠道菌群、免疫细胞和血清代谢物在 IPF 发生发展中的潜在作用, 但三者间的交互关系及其共同调控 IPF 的分子机制尚不明确。为此, 本研究基于已发表的全基因组关联研究 (genome-wide association study, GWAS) 汇总数据, 采用孟德尔随机化 (Mendelian randomization, MR) 方法, 系统评估肠道菌群、血浆代谢物及免疫细胞与 IPF 的潜在因果关系。通过多维度整合分析, 旨在探究三者共同调控 IPF 的分子机制, 并预测干预中药及化合物, 为 IPF 的风险评估、早期诊断及靶向干预方向提供新的理论依据与候选靶点。

1 材料与方法

1.1 研究设计

整体分析流程如图 1 所示。首先, 从已发表的 GWAS 中获取肠道微生物群、血浆代谢物、免疫细胞及 IPF 的汇总统计资料。接着, 运用两样本 MR 方法评估上述暴露因素与 IPF 间的潜在因果关联。然后, 将 MR 分析所得工具变量单核苷酸多态性

(single nucleotide polymorphism, SNP) 导入 Ensembl 数据库进行基因注释, 获得与 IPF 相关的邻近基因集。将该基因集分别与 IPF 差异表达基因 (differentially expressed genes, DEGs) 及加权基因共表达网络分析 (weighted gene coexpression network analysis, WGCNA) 所得基因取交集, 得到候选因果基因。最后, 借助机器学习模型筛选核心基因, 并开展功能富集分析、ROC 曲线评估、列线图构建、决策曲线分析 (decision curve analysis, DCA)、免疫浸润分析、单细胞 RNA 测序 (single-cell RNA sequencing, scRNA-seq)、干预中药预测及分子对接验证。

1.2 数据收集和处理

本研究自基因表达综合数据库 (gene expression omnibus, GEO, <https://www.ncbi.nlm.nih.gov/geo/>) 检索并下载 4 个包含 IPF 患者的独立转录组表达谱数据集。其中, 数据集 GSE135099 与 GSE185691 被指定为 IPF 相关分析的核心训练集, 数据集 GSE124502 与 GSE195770 被用作 IPF 分析的独立外部验证集。各数据集所采用的检测平台、样本类型及样本量等详细信息见表 1。

1.3 MR 数据来源

1.3.1 肠道菌群数据来源 人类肠道微生物组数据来自当前最大规模的 GWAS 数据库 (<https://www.ebi.ac.uk/gwas/>), 包含荷兰微生物组项目的 7 738 名参与者, 详细信息见表 2。上述数据经 LifeLines 项目质量控制, 对应 GWAS 登记号 GCST90027446~GCST90027857。最终纳入 207 个类群 (5 门、10 纲、13 目、26 科、48 属、105 种) 及 205 条代谢通路。

1.3.2 血清代谢物数据来源 血清代谢物数据取自 GWAS 目录 (登记号 GCST90199621~GCST90201020), 涵盖 8 299 名欧洲个体, 详细信息见表 2。涉及 1 091 种血浆代谢物及 309 种代谢物比值, 分析约 15 万个 SNP 位点。从 1 091 种代谢物中筛选出 850 种已知代谢物, 分为 8 大类: 脂类 (395 种)、氨基酸类 (210 种)、外源物 (130 种)、核苷酸类 (33 种)、辅因子和维生素类 (31 种)、碳水化合物类 (22 种)、肽类 (21 种) 及能量类 (8 种); 其余为部分特征性分子 (21 种) 和未知分子 (220 种)。

1.3.3 免疫细胞数据来源 免疫细胞数据同样来自 GWAS 目录 (登记号 GCST0001391~

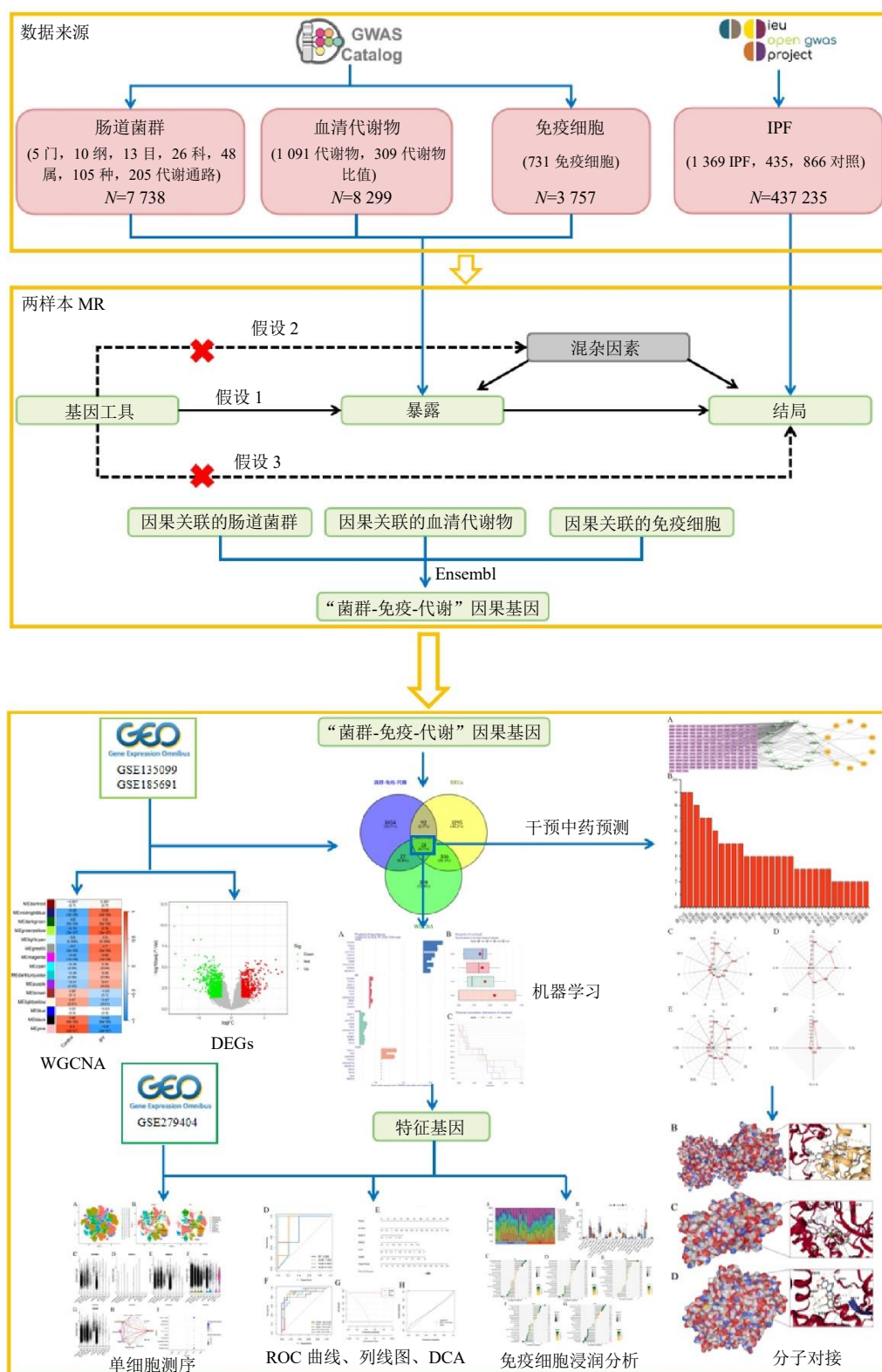


图 1 研究流程

Fig. 1 Research process

GCST0002121), 涵盖 731 种免疫表型的汇总统计, 包括绝对细胞计数 (118 项)、中位荧光强度 (389 项, 反映表面抗原水平)、形态学参数 (32 项) 及

相对细胞计数 (192 项), 详细信息见表 2。其中中位荧光强度 (median fluorescence intensity, MFI)、绝对细胞计数 (absolute cell counts, AC) 和相对细

表 1 微阵列数据

Table 1 Microarray data

数据集	数据平台	数据类型	组织	IPF	对照	生物	分析类型
GSE135099	GPL15207-17536	队列	肺	9	9	人	训练集
GSE185691	GPL570-55999	队列	肺	6	8	人	训练集
GSE124502	GPL21185-21174	队列	肺	6	2	人	测试集
GSE195770	GPL20844-88004	队列	肺	4	4	人	测试集

表 2 MR 的数据来源及特征

Table 2 Data sources and traits for MR analysis

样本	数据来源	ID	人群	样本量
肠道菌群	GWAS	GCST90027446~GCST90027857	欧洲人	7 738
血清代谢物	GWAS	GCST90199621~GCST90201020	欧洲人	8 299
免疫细胞	GWAS	GCST0001391~GCST0002121	欧洲人	3 757
IPF	GWAS	ebi-a-GCST90018120	欧洲人	1 369/435 866

胞计数 (relative cell counts, RC) 特征涉及 B 细胞、经典树突状细胞 (conventional dendritic cell, CDC)、T 细胞成熟阶段、单核细胞、髓系细胞、Treg 细胞及 T、B、NK 细胞组合 (T cell, B cell and natural killer cell, TBNK); 形态学参数 (morphological parameters, MP) 特征包含 CDC 和 TBNK 组合。基于 3 757 名欧洲人群样本, 在控制性别与年龄后, 利用高密度阵列对超过 2 200 万个 SNP 进行基因分型。

1.3.4 IPF 数据来源 IPF 遗传数据源自 GWAS 目录, 包含 1 369 例病例与 435 866 例对照, 所有样本均为欧洲人群, 确保暴露与结局种族背景一致。详细信息见表 2。原始 GWAS 均已获得伦理批准; 本研究仅使用汇总级别数据, 无需额外伦理审查。

1.4 工具变量筛选

从已发表 GWAS 汇总数据中提取暴露因素 (肠道菌群、血清代谢物、免疫细胞) 的遗传工具变量 (instrumental variables, IVs)。筛选标准如下: (1) 以全基因组显著性水平 $P < 5 \times 10^{-5}$ 为阈值, 提取与暴露显著关联的 SNP; (2) 基于欧洲人群 1 000 Genomes 参考数据, 设置 $r^2 < 0.001$ 且聚类距离 $> 100\ 00\text{ kb}$, 以消除连锁不平衡; (3) 剔除与结局存在直接关联的 SNP (结局关联 $P < 0.05$), 避免水平多效性干扰; (4) 计算各 SNP 的 F 统计量, 剔除 $F < 10$ 的弱工具变量。最后将筛选后的暴露 SNP 与结局数据协调, 去除等位基因不匹配或模糊链的 SNP, 保留效应方向一致的 SNP。

1.5 两样本 MR 分析

采用两样本 MR 方法评估各暴露与 IPF 间的

潜在因果关联。使用 R 包 TwoSampleMR 进行统计。以逆方差加权法 (inverse-variance weighted, IVW) 作为主要因果估计模型, 该方法在无水平多效性时提供最精确估计。同时辅以 MR-Egger 回归、加权中位数法和加权模式法, 这些方法对水平多效性具有不同敏感性, 用于检验结果稳健性。效应以比值比 (odds ratio, OR) 及其 95% 置信区间 (confidence interval, CI) 表示。仅当 IVW 方法 $P < 0.05$ 且所有方法效应方向一致时, 判定存在潜在因果关联。

1.6 敏感性分析与多效性检验

为验证 MR 结果的稳健性, 进行系列敏感性分析: (1) 采用 Cochran's Q 检验评估 SNP 间异质性 ($P > 0.05$ 表示无异质性); (2) 通过 MR-Egger 截距项检验水平多效性 ($P > 0.05$ 提示无多效性); (3) 留一法分析逐一剔除每个 SNP 后重新计算 IVW 估计, 评估单个 SNP 对因果效应的驱动作用。绘制散点图、森林图、漏斗图及留一法图, 可视化各 SNP 贡献及潜在偏倚。最终, 对满足 $IVWP < 0.05$ 、所有 MR 方法 OR 方向一致且多效性检验 $P > 0.05$ 的暴露认定为具有稳健因果关联的候选因素, 并提取其 SNP 用于后续分析。

1.7 基于 Ensembl 数据库的邻近基因注释

将从肠道菌群、血清代谢物及免疫细胞分析中获得的 IVs 对应 SNP 信息导入 Ensembl 数据库 (<https://www.ensembl.org/>), 获取各 SNP 的染色体位置。利用 BioMart 工具检索每个 SNP 上下游各 100 kb 范围内的基因; 其中肠道菌群 SNP 初始采用

100 kb 范围未检索到有效邻近基因, 为保证注释完整性, 扩大注释范围至上下游各 1 000 kb, 将检索所得基因作为该 SNP 的邻近基因, 纳入后续功能机制研究。

1.8 DEGs、WGCNA 与候选因果基因的整合分析

为合并 GSE135099 和 GSE185691 数据集并消除批次效应, 采用 R 包“limma”和“sva”处理。将不同数据集视为独立批次, 利用 sva 包的 ComBat 函数识别并移除批次效应。通过主成分分析 (principal component analysis, PCA) 图比较批次校正前后样本分布, 箱线图检查数据标准化前后表达量分布。DEGs 分析使用 limma 包: 对合并后经批次校正的表达矩阵构建线性模型, 采用经验贝叶斯方法估计方差, 通过 eBayes 函数计算差异显著性, 并以 Benjamini-Hochberg 法控制错误发现率 (false discovery rate, FDR)。以 $|\log_2FC| > 0.585$ [FC 表示差异倍数 (fold change)] 且校正后 $P < 0.05$ 为阈值筛选 DEGs。此外, 采用 WGCNA 包构建基因共表达网络。预处理表达矩阵后, 计算所有基因间的 Pearson 相关性并生成相似性矩阵。选取软阈值 $\beta = 6$ 构建邻接矩阵, 进而转换为拓扑重叠矩阵 (topological overlap matrix, TOM)。基于 TOM 相异性进行层次聚类, 利用动态树切割法识别基因模块。计算各模块特征基因, 与临床表型进行相关性分析, 筛选相关性最强且 $P < 0.05$ 的模块作为关键模块。在该模块中, 同时满足基因显著性、模块成员度及模块内连接度较高标准的基因视为关键基因。最后, 将 DEGs、WGCNA 关键基因与前述多组学邻近基因取交集, 获得同时满足遗传因果推断与组织表达差异的基因集, 作为后续功能机制研究的候选靶点。

1.9 因果基因的染色体分布与表达相关性可视化

从 Ensembl 获取差异基因的染色体位置, 基于 IPF 组与对照组表达谱计算 Z-score 归一化值。利用 R 包 circlize 绘制环形基因组图, 每条染色体轨道以热图形式展示基因在 2 组中的表达水平, 轨道外侧标注基因名称。进一步, 在 IPF 样本中计算差异基因两两间的 Pearson 相关系数, 构建相关性矩阵, 通过弦图展示基因间相关关系。

1.10 基于机器学习的枢纽基因筛选与模型解释

应用 4 种机器学习算法 [随机森林 (random forest, RF)、支持向量机 (support vector machine, SVM)、极限梯度提升树 (extreme gradient boosting,

XGB) 及广义线性模型 (generalized linear model, GLM)] 筛选 IPF 相关枢纽基因并构建预测模型。首先, 预处理表达矩阵, 提取样本类别信息作为响应变量, 按 7:3 随机划分为训练集与独立测试集。所有模型在 caret 包框架下训练, 采用五折交叉验证与重复抽样调参。具体配置: (1) RF 默认参数构建 500 棵树; (2) SVM 使用径向基核函数; (3) XGB 选用 xgbDART 方法增强正则化; (4) GLM 采用逻辑回归作为基线。模型评估与解释基于 DALEX 框架。在独立测试集上计算各模型预测残差, 绘制累积残差分布图与箱线图比较误差稳定性; 同时绘制 ROC 曲线并计算曲线下面积 (area under the curve, AUC)。为识别关键基因, 基于均方根误差损失函数计算并比较 4 种模型视角下的基因重要性排名, 重点关注被多个模型共同赋予高重要性的基因, 作为后续验证的候选枢纽基因。

1.11 单基因的 ROC 曲线诊断效能评估

从归一化表达矩阵中提取所有样本的基因表达数据, 根据样本 ID 命名规则提取分组标签, 将结局变量编码为二分类。读取最优模型包含的基因名称。对每个基因使用 R 包 pROC 的 roc 函数计算区分 IPF 与对照的 ROC 曲线及 AUC。将所有基因的 ROC 曲线叠加绘制于同一图中, 不同颜色区分基因, 图例标注基因名称及 AUC 值, 以筛选具有较高单独诊断潜能的标志物。

1.12 列线图构建与 DCA

为可视化最优模型并评估其临床价值, 基于最优模型基因集构建列线图, 并进行校准曲线与决策曲线分析。从归一化表达矩阵中提取最优模型基因列表对应的表达数据, 与样本分组信息合并构建分析数据集。利用 R 包 rms 的 lrm 函数拟合多变量逻辑回归模型 (以基因表达值为预测变量, IPF 患病状态为结局)。基于该模型, 使用 nomogram 函数绘制列线图, 将各基因表达水平映射至评分标尺, 直观计算每个样本的疾病风险概率。采用 bootstrap 法 (重复抽样 1 000 次) 进行内部验证, 绘制校准曲线, 评估模型预测概率与实际观察概率的一致性。进一步使用 rmda 包的 decision curve 函数进行 DCA, 计算不同阈值概率下的净获益, 与“全干预”和“不干预”策略比较, 通过 plot decision curve 函数绘制决策曲线, 综合评估模型临床实用性。

1.13 免疫细胞浸润分析

采用 CIBERSORT 反卷积算法, 以 LM22 免疫

细胞特征基因集为参考，通过支持向量回归估计 IPF 数据集每个样本中 22 种免疫细胞的相对丰度。基于归一化表达矩阵运行 CIBERSORT，仅保留置换 $P < 0.05$ 的样本。获得免疫细胞丰度矩阵后，绘制堆叠条形图展示每个样本的免疫组成，按样本类型分组；采用箱线图及 Wilcoxon 秩和检验比较 2 组间免疫细胞丰度差异。同时，为每个基因绘制棒棒糖图，以直观展示每个候选基因与各免疫细胞亚群的相关性强度及显著性。

1.14 scRNA-seq 数据处理与细胞通讯分析

基于 GEO 数据库中的 IPF 单细胞转录组数据集 GSE279404 展开分析，见表 3。使用 Seurat 包处理数据。质量控制：保留表达基因数 200~5 000 的细胞，排除线粒体基因占比 $> 15\%$ 或血红蛋白基因占比 $> 3\%$ 的细胞。通过 LogNormalize 标准化，利用

FindVariableFeatures 识别前 2 000 个高变基因。采用 PCA 线性降维，选取前 20 个主成分，基于共享最近邻图聚类，使用均匀流形逼近与投影（uniform manifold approximation and projection, UMAP）与 t-分布随机邻域嵌入（t-distributed stochastic neighbor embedding, t-SNE）可视化。为整合不同样本批次效应，应用 Harmony 算法校正，基于校正后前 10 个主成分重新聚类。细胞类型注释通过 SingleR 包完成（以 Human Primary Cell Atlas 为参考），对每个细胞簇自动标注。通过 FindAll Markers 识别各簇差异表达基因（阈值： $\log_2FC \geq 0.25$ ，且在不少于 25% 的簇内细胞中表达）。注释结果在 UMAP 与 t-SNE 图中展示。使用 CellChat 分析配体-受体介导的细胞间通讯，评估所有细胞类型间的相互作用强度，并分析各细胞簇在信号网络中的输出与输入影响。

表 3 IPF 的单细胞 RNA 测序数据信息 (GSE279404 数据集)
Table 3 Sc RNA-seq data information for IPF (GSE279404 dataset)

GEO	数据类型	组织	数据平台	IPF	对照
GSM8686573	scRNA-seq	肺	10X Genomics	0	1
GSM8686574	scRNA-seq	肺	10X Genomics	0	1
GSM8686575	scRNA-seq	肺	10X Genomics	0	1
GSM8686576	scRNA-seq	肺	10X Genomics	0	1
GSM8686577	scRNA-seq	肺	10X Genomics	0	1
GSM8686578	scRNA-seq	肺	10X Genomics	0	1
GSM8686579	scRNA-seq	肺	10X Genomics	0	1
GSM8686580	scRNA-seq	肺	10X Genomics	0	1
GSM8686555	scRNA-seq	肺	10X Genomics	1	0
GSM8686556	scRNA-seq	肺	10X Genomics	1	0
GSM8686557	scRNA-seq	肺	10X Genomics	1	0
GSM8686558	scRNA-seq	肺	10X Genomics	1	0
GSM8686559	scRNA-seq	肺	10X Genomics	1	0
GSM8686560	scRNA-seq	肺	10X Genomics	1	0
GSM8686561	scRNA-seq	肺	10X Genomics	1	0
GSM8686562	scRNA-seq	肺	10X Genomics	1	0

1.15 潜在干预中药预测

将 IPF 相关核心靶点提交至 CTD 数据库 (<https://ctdbase.org/>)，筛选文献支持数 > 1 的化合物。借助 ITCM 数据库 (<http://itcm.biotcm.net/>) 检索与上述化合物对应的中药，依据《中国药典》2025 年版及“十四五”规划教材《中药学》标准化中药名称，剔除无法查询的条目。通过古今医案云平台分析中药的有毒无毒及性味归经。将交集基因-化合物-中药相互作用关系导入 Cytoscape (3.10.1 版) 构建网络图，利用拓扑分析按度 (degree) 值筛选排名

前 3 的化合物。

1.16 分子对接验证

对上述化合物与关键因果蛋白进行分子对接验证。受体蛋白结构获取与筛选依据：首先从 UniProt (<https://www.uniprot.org/>) 检索并筛选人源、经试验验证的关键因果蛋白，获取对应 Entry ID 及标准氨基酸序列；优先从 PDB 数据库 (<http://www.rcsb.org/>) 下载分辨率 $> 2 \text{ \AA}$ ($1 \text{ \AA} = 0.1 \text{ nm}$) 的实验解析晶体结构，若不存在符合条件的晶体结构，则采用 AlphaFold 数据库 (<https://www.alphafold>).

ebi.ac.uk/) 中高精度预测蛋白结构模型。配体化合物结构处理: 从 PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) 下载 CTD 筛选所得化合物的 SDF 格式结构文件, 使用 Chem3D 22.0.0 转换为三维构象, 并采用 MM2 力场完成能量最小化处理。活性口袋与对接盒子设定: 采用全蛋白无偏覆盖策略, 将对接区域设置为完整包裹受体蛋白全部结构空间, 覆盖蛋白质所有潜在结合位点, 不局限于预设活性口袋, 以全面筛查配体与蛋白的潜在结合区域; 采用 AutoDock Vina 1.5.7 软件开展半柔性分子对接, 网格间距根据蛋白大小设置为 0.375~1.000 Å, 对接盒子尺寸根据受体蛋白三维结构范围精准调整, 确保完全包裹蛋白全长序列与所有结构区域。重复性验证: 为保证对接结果稳定性与可重复性, 每组蛋白-配体对接独立重复验证 10 次, 以 10 次重复结果的结合能平均值作为最终数据, 仅保留结合构象稳定一致的结果用于后续分析。分子对接完成后评估受体与配体间的结合亲和力, 采用 PyMOL 3.1 软件对最优对接结果进行可视化分析, 明确受体-配体最佳结合构象、氢键相互作用模式及结合位点特征。

2 结果

2.1 肠道菌群、免疫细胞及血清代谢物对 IPF 的因果效应

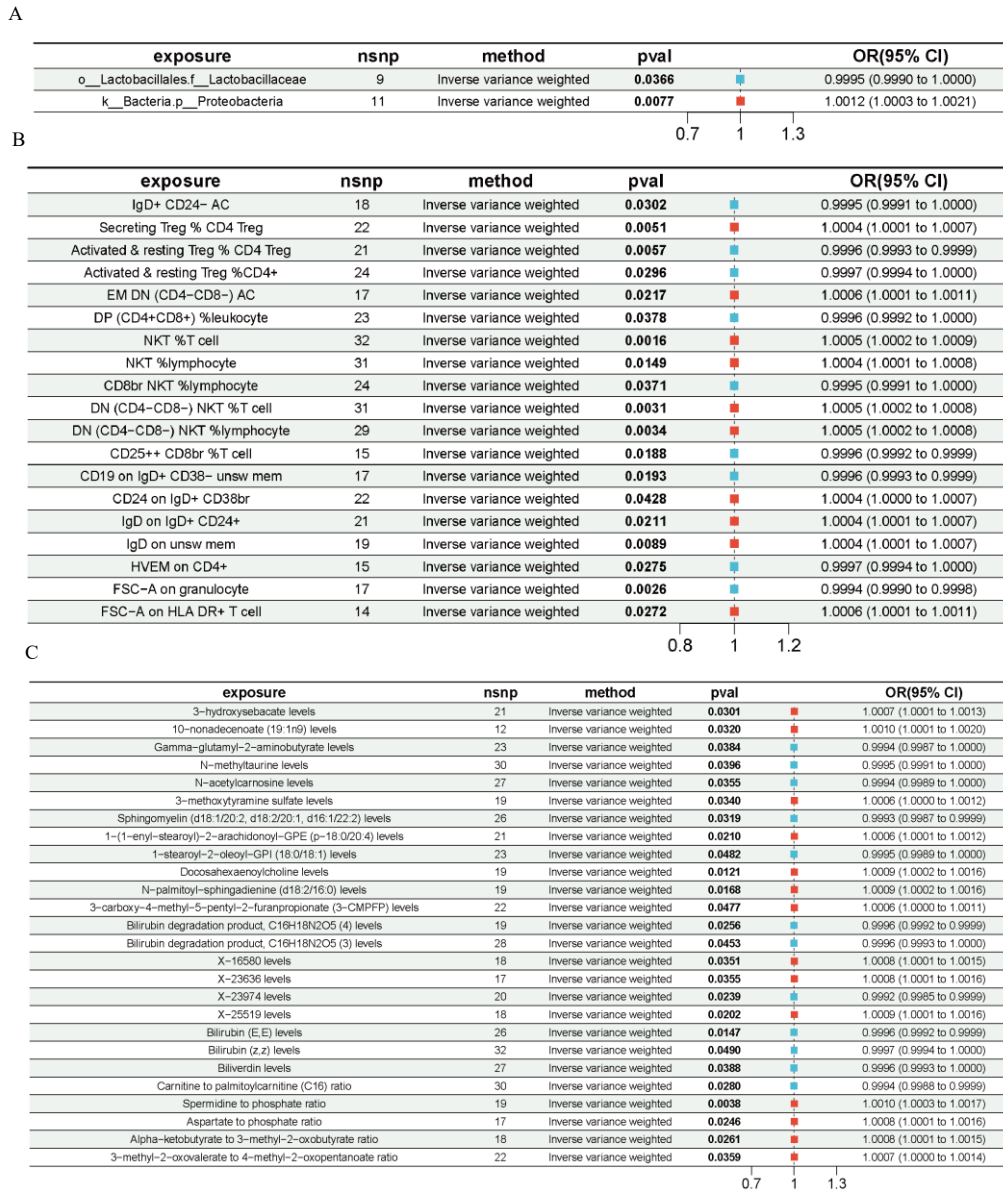
2.1.1 工具变量 SNP 筛选与质量控制 本研究基于双样本 MR 设计, 系统解析肠道菌群、免疫细胞、血清代谢物 3 类暴露与 IPF 的潜在因果关联, 严格遵循 MR 研究规范完成 IVs 的筛选、质控与等位基因协调。肠道菌群中初始纳入 SNP 共计 1 903 个, 经严格质控与协调后剔除 1 883 个, 最终纳入 20 个 SNP 开展 MR 分析, F 统计量范围为 19.57~60.93。免疫细胞中初始纳入 SNP 共计 14 696 个, 经严格质控与协调后剔除 14 284 个, 最终纳入 412 个 SNP 开展 MR 分析, F 统计量范围为 19.54~2 380.38。血清代谢物中初始纳入 SNP 共计 34 843 个, 经严格质控与协调后剔除 34 270 个, 最终纳入 573 个 SNP 开展 MR 分析, F 统计量范围为 19.50~5 308.35。所有纳入分析的 SNP 均满足工具变量三大核心假设: 与暴露因素显著相关、独立于结局混杂因素、仅通过暴露因素介导结局效应且无水平多效性, 为后续因果推断提供了稳健的工具变量支撑。

2.1.2 肠道菌群对 IPF 的因果关联 MR 分析结果显示, 共鉴定出 2 个与 IPF 存在显著因果关联的肠道菌群类群 ($P_{IVW}<0.05$, 图 2-A)。其中, 乳杆菌

科 (Lactobacillaceae) 为 IPF 保护因素 ($OR=0.999\ 5$, $95\%CI: 0.999\sim1.000$, $P=0.036\ 6$), 变形菌门 (Proteobacteria) 为危险因素 ($OR=1.001\ 2$, $95\%CI: 1.000\ 3\sim1.002\ 1$, $P=0.007\ 7$)。上述 OR 值均接近 1, 提示 2 类菌群并非 IPF 的强效应致病因素, 而是与 IPF 发病风险相关的遗传易感关联因素, 二者存在微弱但经统计学检验稳健的因果关联。

2.1.3 免疫细胞对 IPF 的因果关联 在免疫细胞亚群层面, 共识别 19 个与 IPF 存在提示性因果关联的免疫细胞亚群 ($P_{IVW}<0.05$, 图 2-B)。其中 9 个亚群呈保护效应, 包括 IgD⁺CD24⁻ 活化细胞 (activated cell, AC)、活化态及静息态调节性 T 细胞 (regulatory T cells, Treg) 占 CD4⁺Treg 百分比、活化态及静息态 Treg 占 CD4⁺T 细胞百分比、CD4⁺CD8⁺ 细胞占白细胞百分比、CD8 高表达 (bright, br) 自然杀伤 T 细胞 (natural killer T cells, NKT) 细胞占淋巴细胞百分比、CD25⁺CD8^{br} 细胞占 T 细胞百分比、CD19 分子在 IgD⁺CD38⁻ 未类别转换记忆 B 细胞 (unswitched memory B cell, unsw mem) 上的表达水平、疱疹病毒入侵介质 (herpesvirus entry mediator, HVEM) 在 CD4⁺T 细胞上的表达水平和粒细胞的前向散射光面积 (forward scatter-area, FSC-A) ($P=0.002\ 6\sim0.037\ 8$, $OR=0.999\ 0\sim1.000\ 0$)。其余 10 种亚群与 IPF 风险升高相关, 包括分泌型 Treg 占 CD4⁺Treg 百分比、效应记忆 (effector-memory, EM) 双阴性 (double-negative, DN) (CD4⁻CD8⁻) 活化细胞、NKT 细胞占 T 细胞百分比、NKT 细胞占淋巴细胞百分比、双阴性 (CD4⁻CD8⁻) NKT 细胞占 T 细胞百分比、双阴性 (CD4⁻CD8⁻) NKT 细胞占淋巴细胞百分比、CD24 分子在 IgD⁺CD38^{br}B 细胞上的表达水平、IgD 分子在 IgD⁺CD24⁺B 细胞上的表达水平、IgD 分子在 unsw mem 上的表达水平和人类白细胞抗原 DR⁺ (human leukocyte antigen-DR⁺, HLA-DR⁺) T 细胞的 FSC-A ($P=0.001\ 6\sim0.042\ 8$, $OR=1.000\ 4\sim1.000\ 6$)。该效应量范围提示免疫细胞表型与 IPF 为微弱的遗传易感关联, 虽因果效应强度较低, 但经异质性与多效性检验验证, 该关联结果具有统计学稳健性。

2.1.4 血清代谢物对 IPF 的因果效应 血清代谢物分析共筛选出 26 个与 IPF 存在显著因果关联的代谢物 ($P_{IVW}<0.05$, 图 2-C)。其中 12 个代谢物为保护因素, 包括 γ -谷氨酰-2-氨基丁酸、*N*-甲基牛磺酸、*N*-乙酰肌肽、鞘磷脂 (d18:1/20:2, d18:2/20:1,



A-肠道微生物群与 IPF 的因果关联；B-免疫细胞特征与 IPF 的因果关联；C-血清代谢物与 IPF 之间的因果关联。

A-causal association analysis between gut microbiota and IPF; B-causal association analysis between immune cell characteristics and IPF; C-causal relationship between serum metabolites and IPF.

图 2 基于 MR 分析结果绘制的森林图
Fig. 2 Forest plot based on results of MR Analysis

d16:1/22:2)、1-硬脂酰-2-油酰-糖基磷脂酰肌醇(18:0/18:1)、胆红素降解产物[C₁₆H₁₈N₂O₅(4)]、胆红素降解产物[C₁₆H₁₈N₂O₅(3)]、X-23974、胆红素(E,E)、胆红素(Z,Z)、胆绿素和肉碱/棕榈酰肉碱(C16)比值($P=0.0147\sim0.0490$, OR=0.9992~0.9997)。14 个代谢物与 IPF 发病风险呈正相关,包括 3-羟基癸二酸、10-十九碳烯酸(19:1n9)、硫酸-3-甲氧酪胺、1-(1-烯基-硬脂酰)-2-花生四烯酰-甘油磷脂酰乙醇胺(p-18:0/20:4)、二十二碳六烯酰胆碱、N-棕榈酰-鞘二烯碱(d18:2/16:0)、3-羧基-4-甲

基-5-戊基-2-咪喃丙酸酯(3-CMPFP)、X-16580、X-23636、X-25519、亚精胺/磷酸盐比值、天冬氨酸/磷酸盐比值、 α -酮丁酸/3-甲基-2-氧代丁酸比值和 3-甲基-2-氧代戊酸/4-甲基-2-氧代戊酸比值($P=0.0038\sim0.0477$, OR=1.0006~1.0010)。OR 值接近 1 提示血清代谢物与 IPF 之间为微弱的遗传易感关联,本研究因果关联判定综合 P 值、异质性检验及水平多效性检验多重结果,并非仅依据 P 值,证实该关联具有可靠的统计学稳健性。

对所有阳性结果进行异质性及水平多效性检

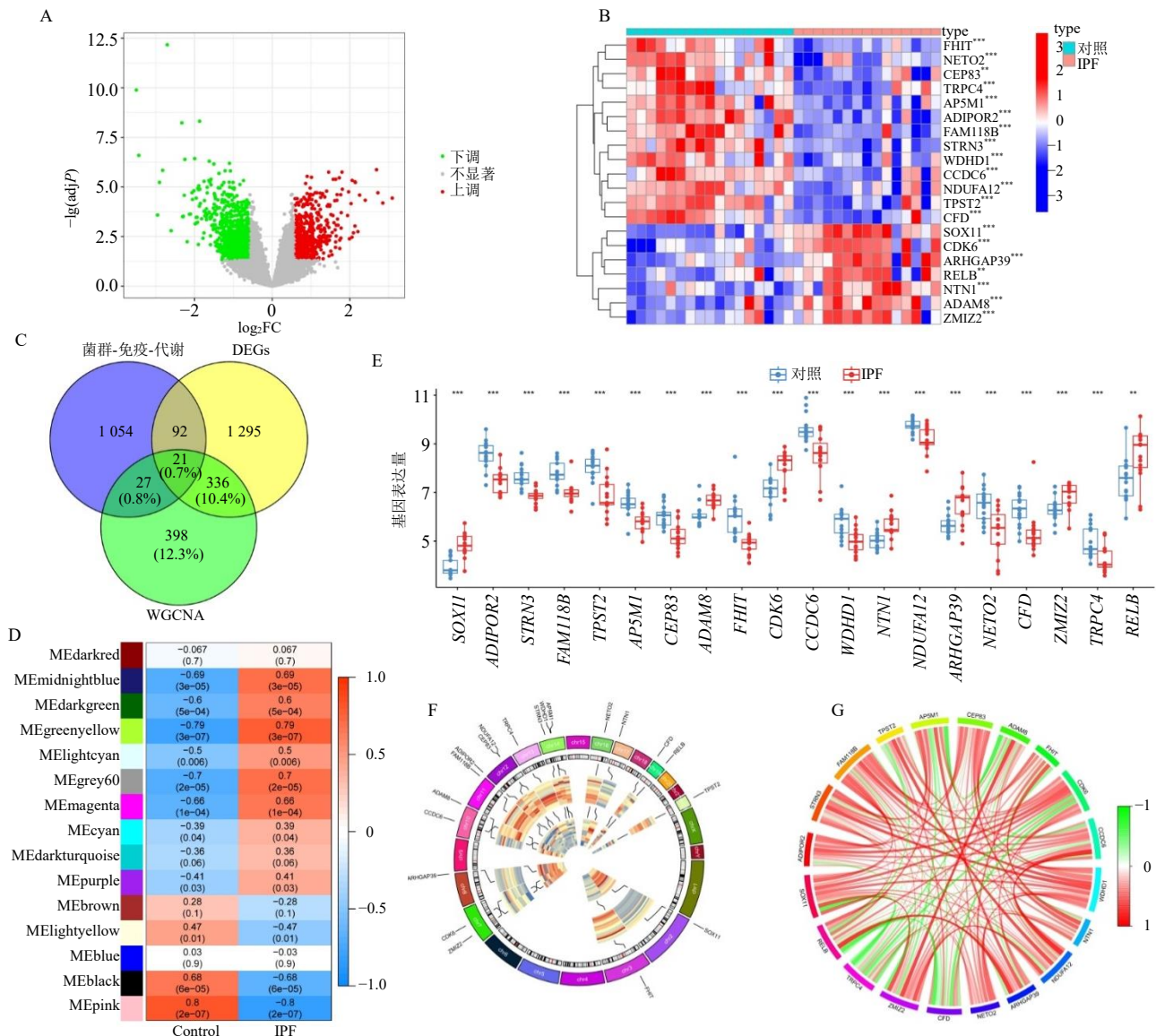
验, Cochran's Q 检验 ($P>0.05$) 和 MR-Egger 截距检验 ($P>0.05$) 均未发现显著异质性或多效性, 表明结果稳健。

2.2 “菌群-免疫-代谢” IVs 邻近基因的注释与整合

从肠道微生物群、免疫细胞及血清代谢物中分别筛选获得 IVs, 对相应 SNP 进行基因 ID 注释与整合, 得到肠道菌群邻近基因 27 个、免疫细胞邻近基因 703 个、代谢物邻近基因 813 个。去重后共获得 1 194 个与 IPF 相关的“菌群-免疫-代谢”因果基因。

2.3 DEGs、WGCNA 模块基因与“菌群-免疫-代谢”因果基因的交集分析

合并 2 个 GEO 数据集并经批次校正后, 获得 32 例样本的表达谱。差异表达分析识别出 1 744 个 DEGs (图 3-A)。WGCNA 将合并数据划分为 15 个模块 (图 3-D): MEpink 模块与 IPF 负相关性最强 ($r=-0.8$, $P=2\times 10^{-7}$), MEgreenyellow 模块正相关性最强 ($r=0.79$, $P=3\times 10^{-7}$)。两模块合并去重后得到 782 个基因, 再与 1 744 个 DEGs 合并去重,



A-火山图; B-21 个因果基因的表达热图; C-DEGs、WGCNA 与“菌群-免疫-代谢”因果基因的 Venn 图; D-模块特征基因与 IPF 表型直接的相关性热图; E-21 个因果基因的差异表达箱线图; F-差异基因的染色体圈图; G-差异基因的相关性圈图。

A-map of the volcano; B-expression heatmap of 21 causal genes; C-Venn plot of DEGs, WGCNA and “microbiota-immune-metabolism” causal genes; D-heatmap of direct correlation between module signature genes and IPF phenotypes; E-boxplots of differential expression of 21 causal genes; F-chromosome circle diagram of differentially expressed genes; G-correlation circle diagram of differentially expressed genes.

图 3 DEGs、WGCNA 模块基因与“菌群-免疫-代谢”因果基因的交集结果

Fig. 3 Intersection results of DEGs, WGCNA module genes and “microflora immune-metabolism” causal genes

最终获得 1 467 个关键表达基因。将 DEGs、WGCNA 模块基因与前述 1 194 个因果基因三者取交集,共得到 21 个与 IPF 关联的因果基因(图 3-B、C)。这些基因在 IPF 组织中表达差异显著(图 3-E),且遗传因果推断层面亦与 IPF 相关,可作为后续功能机制研究的核心候选靶点。

2.4 因果基因的染色体定位特征及其表达相关性

为解析 21 个因果基因的染色体分布及其在 IPF 组织中的共表达模式,绘制了染色体圈图与相关性圈图。如图 3-F 所示,染色体圈图展示了各基因的精确位置,并以热图轨道呈现 IPF 与对照组的差异表达水平。结果显示,这些基因分散于多条染色体,在 IPF 样本中呈现明显表达紊乱。相关性圈图(图 3-G)揭示了 21 个基因在 IPF 样本内的表达关联网络,提示可能存在协同或拮抗的调控关系。

2.5 基于机器学习的诊断模型构建及其性能评估

采用 RF、SVM、XGB 及 GLM 4 种算法构建 IPF 诊断模型,并通过 2 个独立外部验证集(GSE124502、GSE195770)验证模型泛化能力,以排除过拟合风险(图 4-A)。本研究通过残差分析(评估预测偏差与稳定性)和 ROC 曲线(评估诊断区分度)多维度综合评估模型性能。训练集内部评估显示,残差箱线图提示 SVM 模型残差中位数最低,预测偏差最小(图 4-B);残差反向累积分布曲线表明 GLM 模型曲线更靠左且陡峭,预测稳定性更优(图 4-C);训练集 ROC 曲线分析显示,SVM、RF、XGB、GLM 模型的 AUC 分别为 1.000、0.900、0.850、0.725,其中 SVM 模型的诊断区分度与内部拟合效果最优(图 4-D)。

为进一步验证模型可靠性并降低过拟合疑虑,采用 GSE124502 和 GSE195770 作为独立外部验证集开展验证,外部验证集 ROC 曲线结果显示(图 4-E),SVM、XGB、GLM 模型的 AUC 均达到 1.000,RF 模型 AUC 为 0.667,证实 SVM、XGB、GLM 模型在独立外部样本中仍具备优异的诊断效能,模型泛化性良好,无明显过拟合现象。

综合内部拟合效果、预测偏差、诊断区分度三大核心评估维度,确定 SVM 为最优模型;基于该模型按特征重要性评分排序,筛选出接头相关蛋白复合物 5 μ 1 亚基(adaptor-related protein complex 5 subunit mu-1, AP5M1)、SRY 盒转录因子 11(SRY box transcription factor 11, SOX11)、含 MIZ 型锌指

蛋白 2(zinc finger MIZ type containing 2, ZMIZ2)、脆性组氨酸三联体(fragile histidine triad, FHIT)、细胞周期蛋白依赖性激酶 6(cyclin dependent kinase 6, CDK6)为排名前 5 的核心特征基因(图 4-A),上述基因可作为 IPF 发生与进展的潜在候选诊断标志物。

2.6 特征基因的 ROC 曲线、列线图构建和 DCA 评估

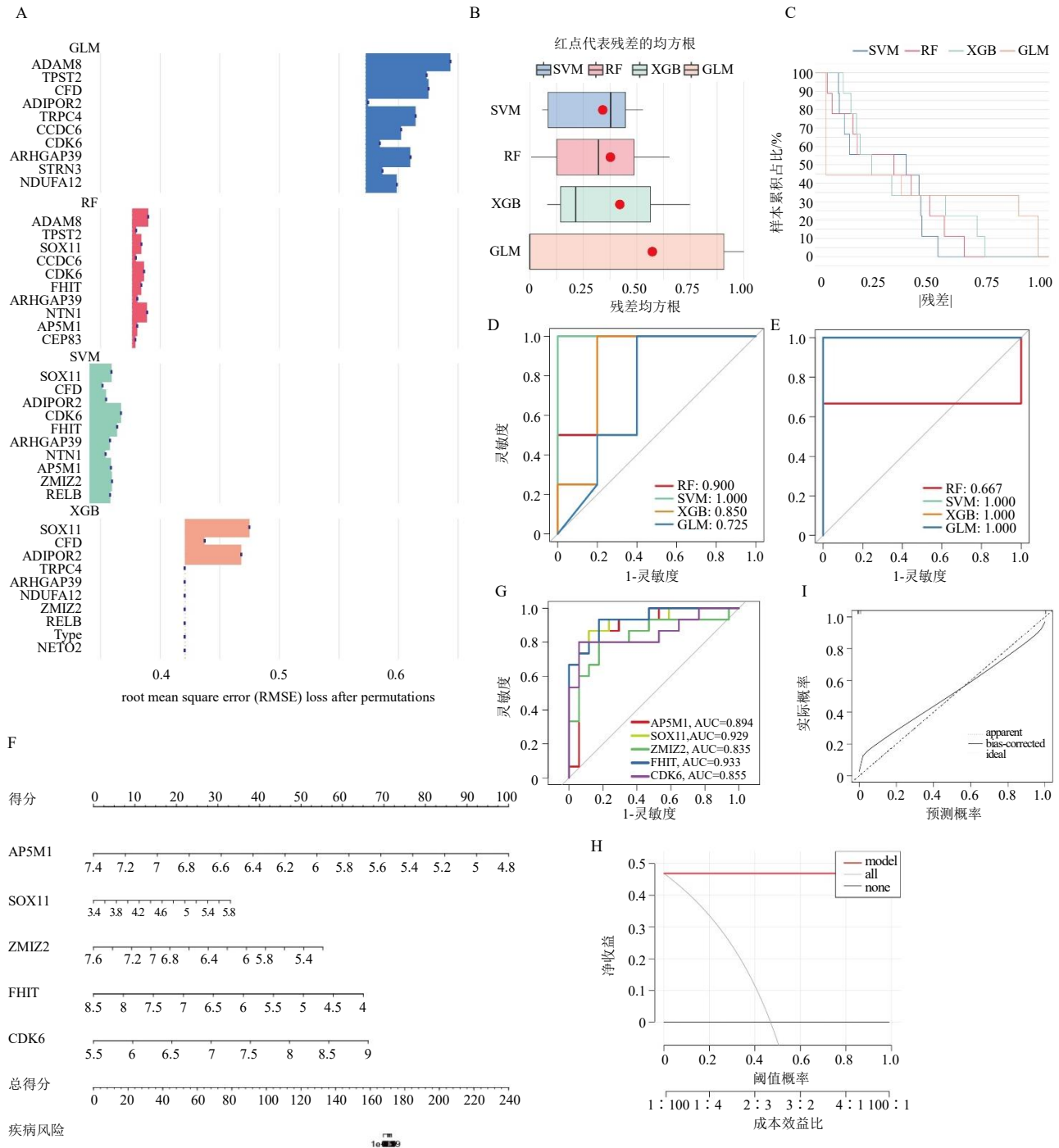
为评价上述 5 个基因单独作为诊断标志物的效能,分别绘制 ROC 曲线(图 4-G)。结果显示,AP5M1 的 AUC 为 0.894,SOX11 为 0.929,ZMIZ2 为 0.835,FHIT 为 0.933,CDK6 为 0.855,均具有潜在诊断价值。为便于临床使用,基于多因素逻辑回归构建列线图(图 4-F),将各基因表达水平映射至评分标尺,累加得分以直观预测个体 IPF 风险概率。采用 Bootstrap 重抽样进行内部验证并绘制校准曲线(图 4-I),模型预测概率与实际观测概率一致性良好,校准度较高。决策曲线分析(图 4-H)显示,在较宽阈值概率范围内,基于 5 个基因的模型净获益优于“全部干预”或“全部不干预”策略,提示该模型具有良好的临床净收益和应用价值,可为 IPF 早期识别与风险分层提供参考。

2.7 免疫细胞浸润分析

为解析 IPF 免疫微环境组成,采用 CIBERSORT 算法对 22 种免疫细胞浸润丰度进行反卷积。结果显示,IPF 组织中主要免疫细胞亚群为 CD4⁺静息记忆 T 细胞、活化肥大细胞、活化自然杀伤细胞和浆细胞(图 5-A)。如图 5-B 所示,与对照组相比,IPF 组中活化自然杀伤细胞和 M0 型巨噬细胞及 M2 型巨噬细胞的比例显著下降($P<0.05$);而单核细胞和活化肥大细胞的比例显著上升($P<0.05$)。进一步分析 5 个特征基因与免疫细胞的相关性(图 5-C~G):CDK6 表达与 Tregs 呈正相关,与 M1 型巨噬细胞呈负相关(图 5-C);FHIT 与活化树突状细胞呈负相关(图 5-D);SOX11 与静息自然杀伤细胞呈正相关,与 M1 型巨噬细胞呈负相关(图 5-F);ZMIZ2 与 Tregs 呈正相关,与静息自然杀伤细胞及 M1 型巨噬细胞呈负相关(图 5-G)。

2.8 单细胞分析

为深入探究 5 个特征基因的功能,利用 IPF 单细胞转录组数据(GSE279404)进行分析。降维聚类识别出 37 个细胞亚群(图 6-A),基于经典标记基因注释为上皮细胞、脂肪细胞、巨噬细胞、成纤



A-4 种机器学习模型的特征重要性; B-残差的箱线图; C-残差的反向累积分布; D-训练集 4 种机器学习模型的 ROC 曲线; E-测试集 4 种机器学习模型的 ROC 曲线; F-特征基因的列线图; G-特征基因的 ROC 曲线; H-DCA 结果; I-校准曲线。

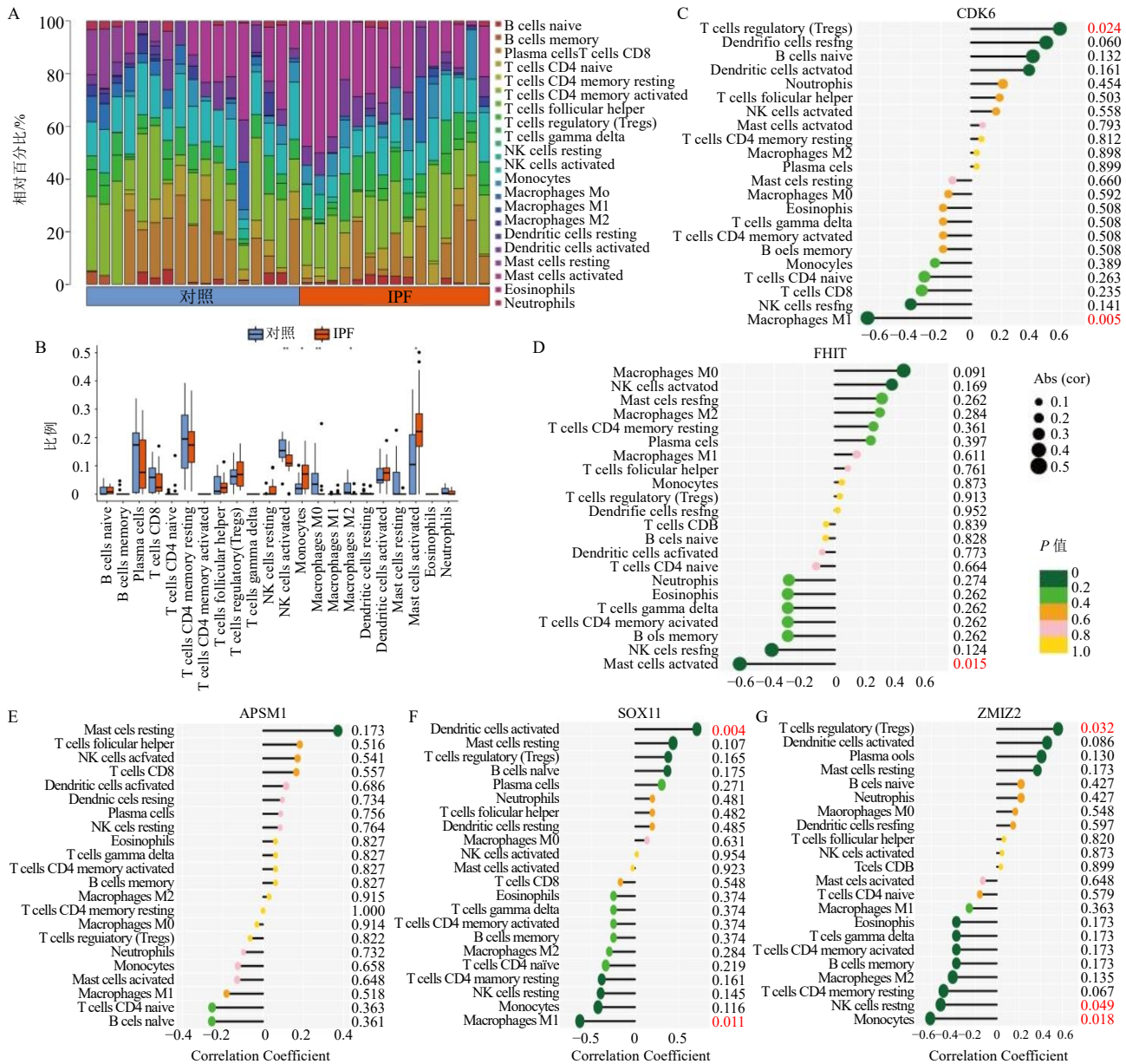
A-feature importance of four machine learning models; B-box plot of residuals; C-inverse cumulative distribution of residuals; D-ROC curves of four machine learning models in the training set; E-ROC curves of four machine learning models in the test set; F-nomogram of featured genes; G-ROC curves of characteristic genes; H-DCA results; I-calibration curve.

图 4 基于机器学习的诊断模型构建及其性能评估

Fig. 4 Machine learning-based diagnostic model construction and its performance evaluation

维细胞、内皮细胞、单核细胞、嗜碱粒细胞、CD4⁺T 细胞、B 细胞和 T 细胞 (图 6-B)。与对照组相比, IPF 组中成纤维细胞和 B 细胞增多, 内皮细胞、单

核细胞及脂肪细胞减少。细胞间通讯分析显示各亚群间存在广泛相互作用 (图 6-H)。值得注意的是, AP5M1、ZMIZ2、FHIT 和 CDK6 在上述细胞中广



A-IPF 组与对照组免疫细胞丰度柱状图；B-IPF 组与对照组免疫细胞浸润的箱线图；C~G-5 个特征基因与免疫细胞的相关性分析。

A-bar graph of immune cell abundance between IPF and Control groups; B-box plot of immune cell infiltration between IPF and Control groups; C—G—correlation analysis between five characteristic genes and immune cells.

图 5 免疫细胞浸润分析结果

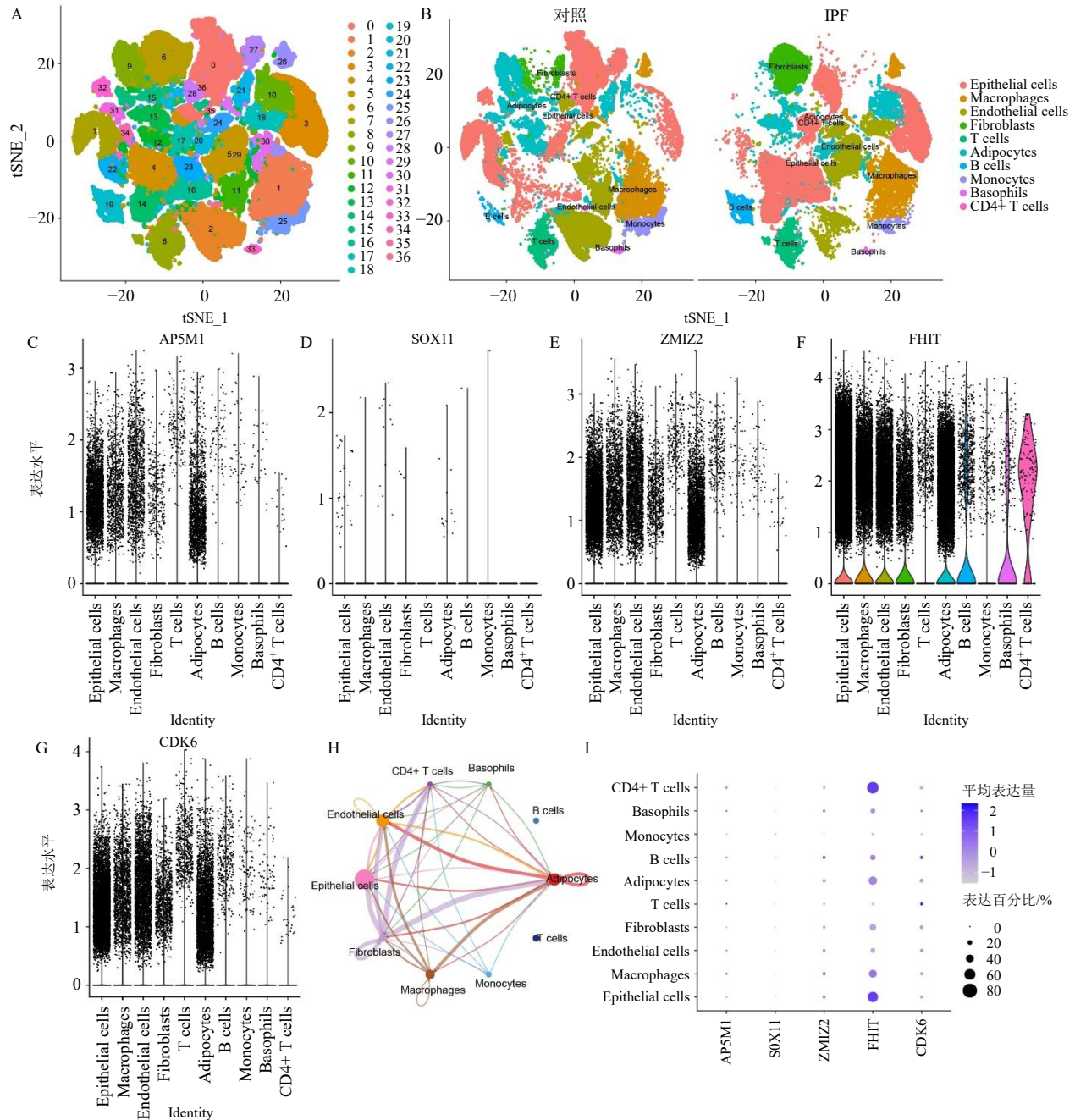
Fig. 5 Results of immune cell infiltration analysis

泛表达；基于单细胞气泡图综合评估基因平均表达水平与阳性表达细胞比例，FHIT 的综合富集程度最高（图 6-C~G、I）。

2.9 靶向中药预测与分析

将 21 个交集基因提交至 CTD 数据库检索化合物，仅 10 个基因 [波形蛋白 (vimentin, VIM)、C-C 基序趋化因子配体 2 (C-C motif chemokine ligand 2, CCL2)、紧密连接蛋白 1 (tight junction protein 1, TJP1)、胶质纤维酸性蛋白 (glial fibrillary acidic

protein, GFAP)、早期生长反应因子 1 (early growth response 1, EGR1)、桥蛋白 (gephyrin, GPHN)、钙结合蛋白 1 (calbindin 1, CALB1)、神经丝轻链 (neurofilament light chain, NEFL)、速激肽前体 1 (tachykinin precursor 1, TAC1)、促肾上腺皮质激素释放激素 (corticotropin releasing hormone, CRH)] 有对应化合物，共 540 种。剔除非中药活性成分后，筛选出 15 种潜在活性化合物，代表物质包括柚皮素、山柰酚、血根碱、橙皮苷、圣草酚、



A-IPF 单细胞数据集的细胞亚群数量的 UMAP 图; B-IPF 和对照中识别出的主要细胞群的 UMAP 图; C~G-各细胞群中特征基因表达的小提琴图; H-不同细胞群之间推断的细胞通讯网络; I-各细胞群中特征基因表达的气泡图。

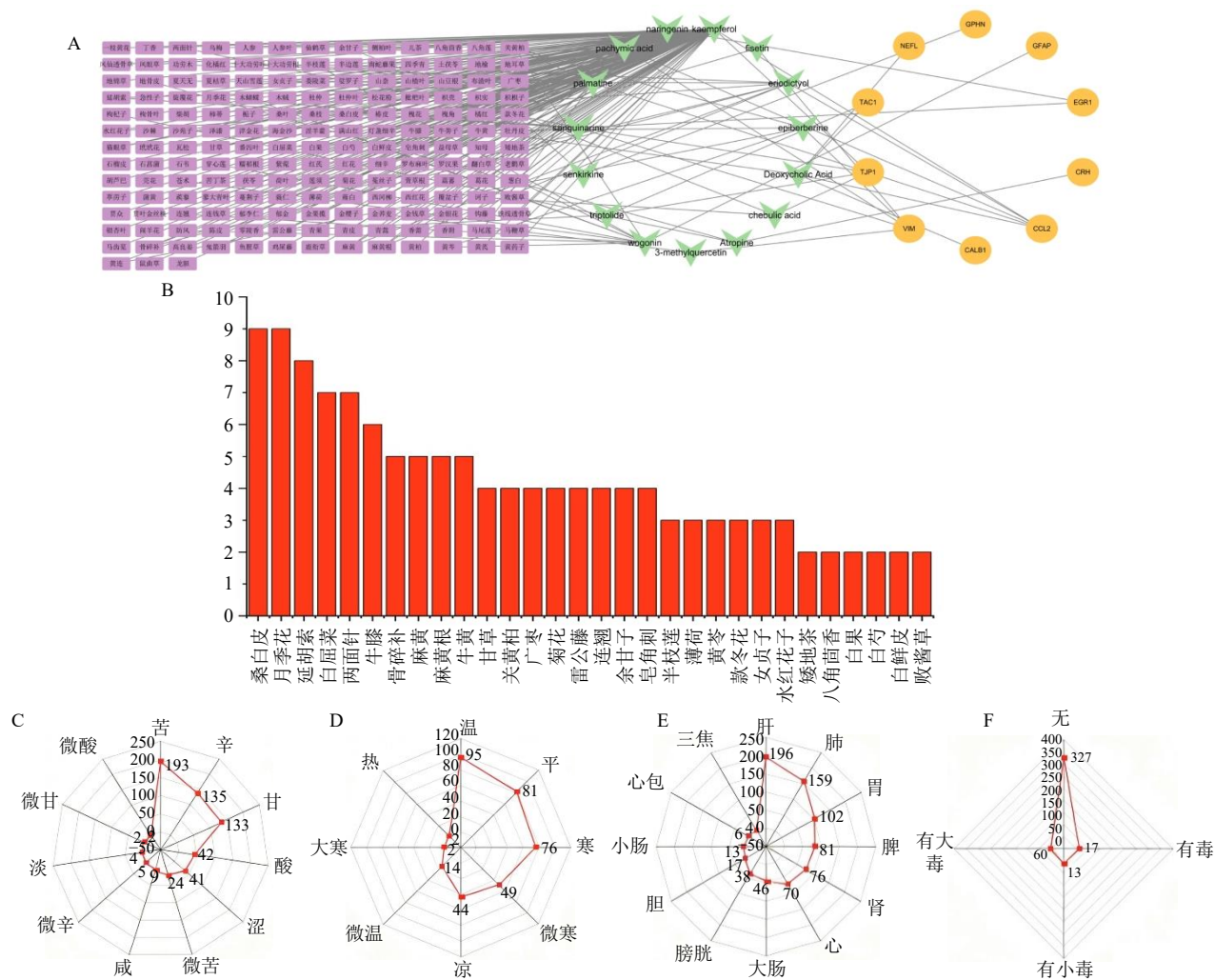
A-UMAP plot of the number of cell subsets in the IPF single-cell dataset; B-UMAP plots of major cell populations identified in IPF and controls; C—G-violin plot of characteristic gene expression in each cell population; H-inferred cellular communication networks between different cell populations; I-bubble plot of signature gene expression in each cell population.

图 6 单细胞转录组分析结果

Fig. 6 Results of single-cell transcriptome analysis

非洲防己碱及汉黄芩素等。进一步利用 ITCM 数据库标准化筛选, 获得 409 种相关中药, 去重及剔除未收录后得到 150 种潜在中药, 高频药物有桑白皮、月季花、延胡索、白屈菜、两面针、牛膝及骨

碎补等 (图 7-B)。将因果基因、化合物与中药导入 Cytoscape 构建“基因-成分- 中药”网络, 包含 197 个节点与 226 条边 (图 7-A)。通过拓扑分析, 按度值 (degree) 排序选取前 3 位化合物 (山柰酚、



A-“IPF 因果基因-成分-中药”网络图; B-频次排名前 30 的潜在干预中药; C-中药五味分布雷达图; D-中药四气分布雷达图; E-中药归经分布雷达图; F-中药毒性分布雷达图。

A-network diagram of “IPF causal gene-component-traditional Chinese medicine (TCM)””; B-top 30 TCMs with potential interventions; C-radar chart of the distribution of five flavors of TCM; D-radar chart of the distribution of the four natures of TCM; E-radar chart of the meridian tropism distribution of TCM; F-radar chart of the toxicity distribution of TCM.

图 7 靶向预测中药与分析

Fig. 7 Targeted prediction and analysis of TCM

柚皮素、非洲防己碱)用于后续分子对接(表 4)。苦、辛、甘居多(图 7-C); 归经集中于肝、肺、胃、脾、肾经(图 7-E); 大多数中药无毒性记载(图 7-F)。

表 4 前 5 位化合物的关键信息

Table 4 Key information of top 5 compounds

PubChem CID	化合物名称	中文名称	分子式	相对分子质量	度值
5280863	kaempferol	山柰酚	C ₁₅ H ₁₀ O ₆	286.24	136
439246	naringenin	柚皮素	C ₁₅ H ₁₂ O ₅	272.25	23
19009	palmatine	非洲防己碱	C ₂₁ H ₂₂ NO ₄ ⁺	352.40	12
5281703	wogonin	汉黄芩素	C ₁₆ H ₁₂ O ₅	284.26	11
5154	sanguinarine	血根碱	C ₂₀ H ₁₄ NO ₄ ⁺	332.30	8

2.10 分子对接验证结果

选择度值前 3 的化合物（山柰酚、柚皮素、非洲防己碱）分别与 10 个核心基因蛋白（VIM、CCL2、TJP1、GFAP、EGR1、GPHN、CALB1、NEFL、TAC1、CRH）进行分子对接。配体结构从 PubChem 获取二维结构，经 Chem3D 22.0.0 软件 MM2 力场能量最小化处理。受体晶体结构：除 NEFL 外（其三维构象来自 AlphaFold 数据库），其余 9 个靶点均从 PDB 下载（VIM:3SSU、CCL2:4ZK9、TJP1:2H2C、GFAP:6A9P、

EGR1:4R2D、GPHN:1JLJ、CALB1:6FIE、TAC1:4HOM、CRH:3EHU）。结合能评估标准：<-4.25 kcal/mol（1 kcal=4.2 kJ）为确定、<-5 kcal/mol 为良好和<-7 kcal/mol 为强。如表 5 和图 8-A 显示，所有化合物与靶点的结合能均低于-4.25 kcal/mol，形成稳定结合。尤其 CRH 和 TAC1 与 3 种化合物的结合能均低于-8 kcal/mol（图 8-B~D）。上述分子对接结果为计算预测结论，提示 3 种化合物与核心靶点结合稳定，具有针对 IPF 的潜在干预价值。

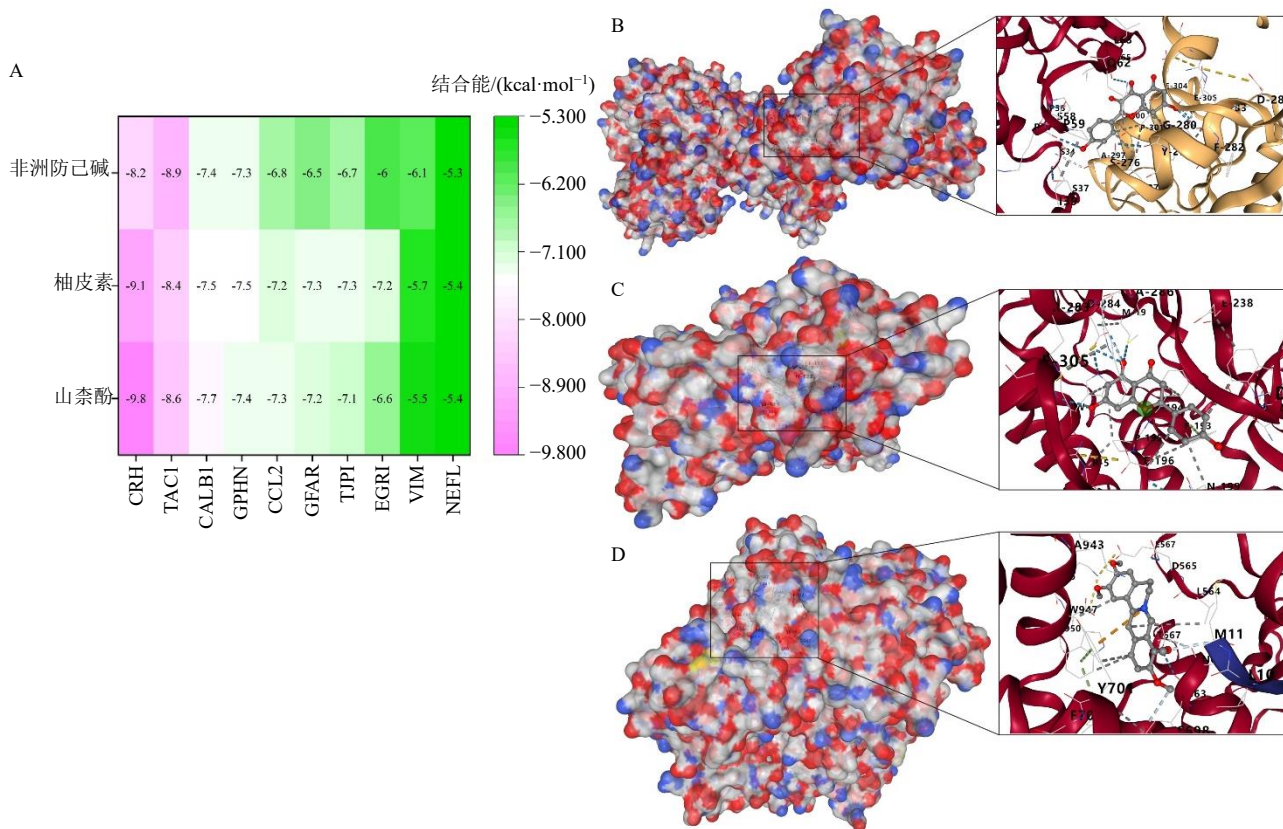
表 5 候选化合物与核心基因蛋白的分子对接结合能
Table 5 Molecular docking binding energy between candidate compounds and core genes

基因	化合物	结合能/(kcal·mol ⁻¹)	空腔体积/Å ³	中心 (x, y, z)	对接尺寸 (x, y, z)
CALB1	山柰酚	-7.7	340	-13, 8, -19	21, 21, 21
	柚皮素	-7.5	340	-13, 8, -19	21, 21, 21
	非洲防己碱	-7.4	395	-33, 3, -21	22, 22, 22
CCL2	山柰酚	-7.3	195	-4, -18, 5	21, 21, 21
	柚皮素	-7.2	195	-4, -18, 5	21, 21, 21
	非洲防己碱	-6.8	195	-4, -18, 5	22, 22, 22
CRH	山柰酚	-9.8	3 557	-38, -12, 66	21, 35, 30
	柚皮素	-9.1	4 992	-39, -38, 110	28, 35, 35
	非洲防己碱	-8.2	4 992	-39, -38, 110	28, 35, 35
EGR1	山柰酚	-6.6	1 773	4, -4, -16	21, 21, 21
	柚皮素	-7.2	1 773	4, -4, -16	21, 21, 21
	非洲防己碱	-6.0	1 773	4, -4, -16	22, 22, 22
GFAR	山柰酚	-7.2	630	-18, -235, 841	21, 27, 21
	柚皮素	-7.3	630	-18, -235, 841	21, 27, 21
	非洲防己碱	-6.5	630	-18, -235, 841	22, 22, 22
GPHN	山柰酚	-7.4	318	24, -7, 27	21, 21, 21
	柚皮素	-7.5	318	24, -7, 27	21, 21, 21
	非洲防己碱	-7.3	318	24, -7, 27	22, 22, 22
NEFL	山柰酚	-5.4	30	21, 4, -3	21, 21, 21
	柚皮素	-5.4	30	21, 4, -3	21, 21, 21
	非洲防己碱	-5.3	71	132, 43, -54	22, 22, 22
TAC1	山柰酚	-8.6	3 137	20, 26, 50	28, 21, 28
	柚皮素	-8.4	1 291	-4, 13, 67	21, 21, 21
	非洲防己碱	-8.9	1 291	-4, 13, 67	22, 22, 22
TJP1	山柰酚	-7.1	650	25, 0, -7	21, 21, 21
	柚皮素	-7.3	650	25, 0, -7	21, 21, 21
	非洲防己碱	-6.7	244	16, -3, 0	22, 22, 22
VIM	山柰酚	-5.5	31	-8, 3, 29	21, 21, 21
	柚皮素	-5.7	31	-8, 3, 29	21, 21, 21
	非洲防己碱	-6.1	31	-8, 3, 29	22, 22, 22

3 讨论

本研究采用多组学 MR 框架，系统评估肠道菌群、免疫细胞、血清代谢物与 IPF 之间的因果关联，

旨在为 IPF 的早期诊断标志物发现及干预中药筛选提供遗传学依据。基于 GWAS 数据开展两样本 MR 分析，对作为工具变量的 SNP 进行基因注释；整合



A-候选化合物与核心靶点的分子对接热图；B-山柰酚与 CRH 的分子对接图；C-柚皮素与 CRH 的分子对接图；D-非洲防己碱与 TAC1 的分子对接图。

A-heatmap of molecular docking between candidate compounds and core targets; B-molecular docking map of kaempferol and CRH; C-molecular docking map of naringenin and CRH; D-molecular docking map of tetrandrine and TAC1.

图 8 分子对接结果

Fig. 8 Molecular docking results

IPF 转录组数据，通过 DEGs 和 WGCNA 筛选候选因果基因；进而运用机器学习算法构建诊断模型，结合免疫细胞浸润评估、scRNA-seq、中药筛选及分子对接技术，识别关键遗传标志物及潜在干预中药与活性化合物。

两样本 MR 分析提示，肠道菌群中特定分类群的丰度变化与 IPF 风险存在因果关联。具体而言，乳杆菌科（Lactobacillaceae）丰度升高与 IPF 风险降低相关，而变形菌门（Proteobacteria）丰度增加则与 IPF 风险升高相关。乳杆菌科可通过“肠-肺轴”发挥抗炎效应；已有研究证实，补充该菌可减轻博来霉素诱导的肺纤维化，其机制与调节短链脂肪酸水平及抑制 Th17 细胞分化有关^[9-10]。变形菌门包含多种革兰阴性致病菌，其丰度增加与慢性炎症状态密切相关^[11-12]；在 IPF 患者的肺部菌群中，变形菌门丰度显著升高，可能通过脂多糖/Toll 样受体 4 通路激活肺泡巨噬细胞，驱动促纤维化因子的释放^[13-14]。本

研究从遗传学层面为上述菌群在 IPF 中的作用提供了新的证据。

MR 分析共识别出 19 种与 IPF 存在显著因果关联的免疫细胞表型。其中，9 种表现为保护因素（如 IgD⁺ CD24⁻ AC 和活化态及静息态 Treg 占 CD4⁺ Treg 百分比），10 种表现为风险因素（如分泌型 Treg 占 CD4⁺ Treg 百分比和 NKT 细胞占 T 细胞百分比）。本研究发现 Treg 在 IPF 中发挥双向调控作用，活化/静息态 Treg 为保护因素，而分泌型 Treg 则为风险因素，提示 Treg 功能存在异质性。已有研究证实，IPF 患者肺组织中静息态 Treg 有助于维持免疫稳态，而分泌型 Treg 高表达转化生长因子-β（transforming growth factor-β, TGF-β）、白细胞介素-10（interleukin-10, IL-10）等纤维化相关因子，可直接促进成纤维细胞活化^[15]。此外，NKT 细胞可通过分泌 Th2 型细胞因子促进胶原沉积，加重 PF 进程^[16]。本研究从遗传学角度验证了这些免疫细胞

的致病或保护作用。

本研究共发现 26 种血清代谢物与 IPF 显著相关, 其中 12 种为保护性因素、14 种为风险性因素。在保护性代谢物中, 胆红素及其代谢产物 [如胆红素 (*E,E*)、胆红素 (*Z,Z*)、胆绿素] 表现出显著的保护效应。胆红素具有强效的抗氧化及抗炎特性。临床研究显示, 其水平与 IPF 患者肺功能下降速率呈负相关, 高胆红素水平与更长的生存期相关^[17]。鞘磷脂水平升高同样与 IPF 风险降低相关, 其代谢产物 1-磷酸鞘氨醇 (1-phosphatidylserine, S1P) 可通过 S1PR2/S1PR3 受体调节成纤维细胞活化^[18]。在风险性代谢物中, 多种比值类代谢物 (如亚精胺/磷酸盐比值) 升高与 IPF 风险增加相关, 这可能反映了多胺代谢与磷酸盐代谢之间的失衡状态。尽管亚精胺在动物模型中展现出肺纤维化保护作用 (机制涉及诱导自噬及抑制内质网应激^[19]), 但其代谢异常 (如与磷酸盐的比值异常升高) 仍可能通过其他途径间接提示多胺代谢通路的紊乱。本研究从遗传学角度进一步支持了上述血清代谢物与 IPF 风险之间的因果关系。

通过 Ensembl 数据库对 MR 工具变量 SNP 进行邻近基因注释, 获得“菌群-免疫-代谢”因果基因集。将该基因集与 IPF 转录组数据的 DEGs 及 WGCNA 模块基因取交集, 共获得 21 个候选因果基因。进一步采用机器学习算法从中筛选出 5 个核心基因 AP5M1、SOX11、ZMIZ2、FHIT 和 CDK6, 可作为 IPF 潜在的诊断标志物。

AP5M1 作为衔接蛋白复合体 5 (adaptor protein complex 5, AP-5) 的 1 个亚基, 主要负责细胞内内体运输过程, 对于维持细胞稳态及保障溶酶体通路正常运行具有不可或缺的作用^[20]。已有报道指出, AP5M1 的变异可引发遗传性视网膜病变^[21], 并能够诱导胶质母细胞瘤发生凋亡^[20]。目前尚无直接实验证据证实 AP5M1 与 IPF 相关, 本研究基于其细胞内转运功能推测 AP5M1 可能通过影响成纤维细胞的自噬或凋亡途径, 进而导致肌成纤维细胞持续活化及细胞外基质过度积聚, 但其具体机制仍有待实验验证。SOX11 是 SOX 家族转录因子, 调控胚胎发育、细胞命运决定、增殖与分化等生物学过程。现有研究表明, SOX11 是 TGF- β 信号通路下游的促纤维化转录因子, 与皮肤及肺部的纤维化进程密切相关^[22]。在系统性硬化症患者的皮肤成纤维细胞中, SOX11 呈高表达状态, 并可被 TGF- β 诱导上

调^[22]。在人肺成纤维细胞中, SOX11 受 IL-13 信号通路调控。敲低 SOX11 可显著抑制骨膜蛋白及 C-C 基序趋化因子配体 26 (C-C motif chemokine ligand 26, CCL26) 等促纤维化分子的表达^[23]。此外, SOX11-胰岛素样生长因子 2 (insulin-like growth factor 2, IGF2) 信号轴在肺早期发育过程中调控肺泡上皮细胞的增殖与凋亡^[24]。基于上述证据, SOX11 作为 IPF 的一个风险基因, 可能通过促进成纤维细胞活化及影响上皮细胞稳态 2 种途径参与 IPF 的进展。ZMIZ2 是一种转录共调节因子, 参与 TGF- β /Smad 信号通路的调控, 能够增强 Smad3 的转录活性, 从而放大促纤维化信号^[25], 提示 ZMIZ2 可能在 IPF 中扮演转录调控放大器的角色。同时, ZMIZ2 也是经典 Wnt/ β -连环蛋白 (β -catenin) 信号通路的 1 个新型正调控因子, 可增强 β -catenin 介导的转录活性, 促进细胞增殖^[26]。该通路的异常激活已被证实可推动成纤维细胞活化、增殖及其向肌成纤维细胞的转分化^[27]。因此, ZMIZ2 可能通过激活 Wnt/ β -catenin 通路在 IPF 中发挥作用。FHIT 是 1 个抑癌基因, 作为肿瘤抑制因子, 调控细胞增殖、凋亡及细胞周期进程。Uematsu 等^[28]报道在 IPF 患者肺组织的化生细胞及异常增生的细支气管上皮中, FHIT 蛋白表达水平显著降低, 甚至出现等位基因缺失。在大鼠 PF 模型中, 同样观察到肺泡 II 型上皮细胞中 FHIT 表达随时间推移而下调^[29]。这些发现提示, FHIT 缺失可能通过增强 TGF- β 信号转导及促进成纤维细胞增殖, 参与 IPF 的发生与发展。CDK6 是细胞周期蛋白依赖性激酶家族成员, 通过与细胞周期蛋白 D (cyclin D) 结合, 驱动细胞增殖。在 PF 小鼠模型中, CDK4/6 抑制剂帕博西尼可显著降低肺组织胶原沉积, 从而抑制纤维化进展^[30]。在二氧化硅诱导的矽肺模型中, 巨噬细胞来源的外泌体 miR-107 可通过靶向抑制 CDK6, 促进肺成纤维细胞向肌成纤维细胞转分化, 进而引发 PF^[31]。此外, 长链非编码 RNA lncSIL 可通过 zeste 基因增强子同源物 2 (enhancer of zeste 2 polycomb repressive complex 2 subunit, EZH2) /P21/CDK6 信号通路, 负向调控 TGF- β 1 诱导的肺泡上皮细胞间质转化^[32]。上述研究表明, CDK6 可能通过驱动成纤维细胞增殖、促进 TGF- β 介导的上皮-间质转化以及调节肺部免疫微环境等多重机制参与 IPF 的病理过程。

为探索潜在干预中药, 将 21 个候选因果基因中可从 CTD 数据库检索到的 10 个基因进行化合物

筛选, 获得 15 种潜在活性化合物及 150 种潜在干预中药。活性化合物主要包括柚皮素、山柰酚、血根碱、橙皮苷、圣草酚、非洲防己碱及汉黄芩素等。现代药理研究提示上述筛选的化合物具有潜在抗纤维化干预价值。研究发现, 山柰酚治疗减轻 IPF 小鼠肺部炎症浸润和胶原沉积, 并降低 α -平滑肌肌动蛋白 (α -smooth muscle actin, α -SMA) 和纤连蛋白表达, 下调肺组织中磷酸化核因子- κ B (phosphorylated nuclear factor- κ B, p-NF- κ B)、肿瘤坏死因子- α (tumor necrosis factor- α , TNF- α)、磷酸化 NF- κ B 抑制蛋白- α (phosphorylated inhibitory of NF- κ B- α , p-I κ B- α) 和 IL-6 蛋白表达, 发挥缓解小鼠 PF 进展的作用^[33]。研究表明, 柚皮素能降低 PF 小鼠中肺组织中羟脯氨酸 (hydroxyproline, HYP)、IL-1 β 、IL-18、TGF- β 1 含量, 下调核苷酸结合寡聚化结构域样受体蛋白 3 (nucleotide-binding oligomerization domain-like receptor protein 3, NLRP3)、含 CARD 结构域凋亡相关斑点样蛋白 (apoptosis associated speck like protein containing a CARD, ASC)、半胱氨酸天冬氨酸蛋白酶-1 (cysteine aspartate protease-1, Caspase-1) mRNA 和蛋白表达水平以及 I 型胶原 (Collagen I)、 α -SMA 和波形蛋白 (Vimentin) 蛋白表达, 上调 E-钙黏蛋白 (epithelial-cadherin, E-cadherin) 蛋白表达, 这表明其能够抑制 PF 进程^[34]。张艳等^[35]研究发现, 橙皮苷通过上调 E-cadherin 蛋白表达及下调 CollagenI、纤连蛋白 (Fibronectin)、 α -SMA 蛋白表达, 从而改善 PF 进展。汉黄芩素下调 DNA 损伤标志物 γ -组蛋白 H2AX (gamma-histone 2A variant X, γ -H2AX) 和衰老相关标志物磷酸化 p53、p21、视网膜母细胞瘤蛋白和衰老相关 β -半乳糖苷酶的蛋白表达。此外, 汉黄芩素作为 CDK9 的抑制剂, 还可直接抑制 CDK9 和 p53/p21 信号通路表现出抗 PF 的能力^[36]。汪志斌等^[37]通过体内外实验发现, 血根碱能上调 E-cadherin 表达, 下调 Vimentin 及 α -SMA 表达, 并通过改变 TGF- β 1R 构象来抑制 PF。

尽管本研究取得上述发现, 但仍需具有以下局限性。其一, MR 分析所采用的肠道菌群、免疫细胞及血清代谢物 GWAS 数据均来源于欧洲人群, 而 IPF 的遗传数据亦基于欧洲人群, 因此研究结论外推至其他种族或地区人群时需谨慎。不同人群间的遗传背景、生活方式及菌群结构差异可能影响因果关联的稳健性。其二, 本研究虽通过多组学 MR 识

别菌群、免疫细胞及代谢物与 IPF 的因果关联, 但三者之间的中介关系尚未通过中介 MR 或实验研究加以验证。“菌群-免疫-代谢”轴上下游调控网络的具体机制仍需进一步阐明。其三, 机器学习模型在训练集中表现出极高的诊断效能, 但合并后样本经历了多轮特征筛选, 存在过拟合风险。模型缺乏独立的外部验证集, 其泛化能力需在大规模、前瞻性、多中心队列中严格验证。其四, 分子对接结果仅基于计算模拟, 虽显示部分化合物与靶点蛋白具有较强结合亲和力, 但其实际生物活性、药动学特性、体内疗效及潜在毒性仍需通过细胞实验、动物模型及临床试验进行验证。其五, 本研究筛选出的核心基因中, 部分基因如 AP5M1 等尚未有其与 IPF 的直接研究, 其在疾病发生发展中的具体作用机制有待深入探索。未来研究可结合多组学中介分析、功能实验、前瞻性队列及多中心临床试验, 进一步解析“菌群-免疫-代谢”轴在 IPF 中的作用机制, 推动诊断标志物的临床转化及靶向治疗策略的研发。

4 结论

本研究基于多组学 MR 框架, 整合肠道菌群、免疫细胞及血清代谢物的遗传数据, 结合转录组与机器学习, 探讨了 IPF 的潜在诊断标志物及分子机制。两样本 MR 分析识别出与 IPF 显著因果关联的 2 个菌群 (乳杆菌科为保护因素、变形菌门为风险因素)、免疫细胞表型 (9 个保护、10 个风险) 及血清代谢物 (12 个保护、14 个风险), 为“菌群-免疫-代谢”轴提供了遗传学证据。通过 DEGs、WGCNA 筛选出 21 个候选基因, 并基于 SVM 机器学习模型构建高效诊断组合, 识别出 AP5M1、SOX11、ZMIZ2、FHIT、CDK6 共 5 个具有诊断效能的因果基因。药物筛选与分子对接显示, 山柰酚、柚皮素、非洲防己碱等化合物与 CRH、TAC1 等靶点结合稳定, 提示其作为 IPF 干预的潜在候选药物。本研究为 IPF 的风险评估、候选标志物筛选及中医药潜在干预提供遗传学证据与理论依据, 所有候选标志物与化合物均需后续实验验证转化价值。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Wanas H, Khalifa A K, Elbadawy H M, et al. Targeting inflammatory and oxidative pathways in idiopathic pulmonary fibrosis: Synergistic effects of quercetin and pirfenidone in modulating Nrf2/PPAR- γ and TLR-4/NF- κ B pathways [J]. *Pharmacol Res Perspect*, 2026, 14(2):

- e70229.
- [2] Vos T, Lim S S, Abbafati C, *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019 [J]. *Lancet*, 2020, 396(10258): 1204–1222.
 - [3] Arvanitaki A, Diller G P, Lawrence A, *et al.* Deep-learning networks accurately detect pulmonary hypertension in patients with idiopathic pulmonary fibrosis [J]. *ERJ Open Res*, 2026, 12(2). DOI:10.1183/23120541.00672-2025.
 - [4] Bai B, Li F Z, Zheng P Y, *et al.* Gut microbiota dysbiosis and metabolic abnormalities promote oxidative stress and fibrosis in idiopathic pulmonary fibrosis [J]. *Sci Rep*, 2026, 16(1): 12305.
 - [5] Jiang Y H, Zhou M, Cheng M D, *et al.* CAR-engineered cytolytic Tregs reverse pulmonary fibrosis and remodel the fibrotic niche with limited CRS [J]. *JCI Insight*, 2025, 10(15): e182050.
 - [6] Wang J, Xu L Z, Xiang Z, *et al.* Microcystin-LR ameliorates pulmonary fibrosis via modulating CD206⁺ M2-like macrophage polarization [J]. *Cell Death Dis*, 2020, 11(2): 136.
 - [7] Cai W J, Zhang H Y, Li Z Z, *et al.* Potential biomarkers of idiopathic pulmonary fibrosis: Metabonomics driven lipid profiling [J]. *J Transl Med*, 2025, 23(1): 1010.
 - [8] Wu F, Li B Y, Li J Q, *et al.* Association between genetic prediction of 486 blood metabolites and the risk of idiopathic pulmonary fibrosis: A Mendelian randomization study [J]. *Biomed Rep*, 2025, 22(3): 52.
 - [9] Budden K F, Gellatly S L, Wood D L A, *et al.* Emerging pathogenic links between microbiota and the gut-lung axis [J]. *Nat Rev Microbiol*, 2017, 15(1): 55–63.
 - [10] Quan Y Y, Yin Z J, Chen S L, *et al.* The gut-lung axis: Gut microbiota changes associated with pulmonary fibrosis in mouse models induced by bleomycin [J]. *Front Pharmacol*, 2022, 13: 985223.
 - [11] Hiippala K, Kainulainen V, Kalliomäki M, *et al.* Mucosal prevalence and interactions with the epithelium indicate commensalism of *Sutterella spp* [J]. *Front Microbiol*, 2016, 7: 1706.
 - [12] Zhou Y L, Xu Z Z, He Y, *et al.* Gut microbiota offers universal biomarkers across ethnicity in inflammatory bowel disease diagnosis and infliximab response prediction [J]. *mSystems*, 2018, 3(1): e00188–e00117.
 - [13] Larsen J M, Musavian H S, Butt T M, *et al.* Chronic obstructive pulmonary disease and asthma-associated Proteobacteria, but not commensal *Prevotella spp.*, promote Toll-like receptor 2-independent lung inflammation and pathology [J]. *Immunology*, 2015, 144(2): 333–342.
 - [14] He Z Y, Zhu Y S, Jiang H. Inhibiting toll-like receptor 4 signaling ameliorates pulmonary fibrosis during acute lung injury induced by lipopolysaccharide: An experimental study [J]. *Respir Res*, 2009, 10(1): 126.
 - [15] McCubbrey A L, Janssen W J. Using a single-cell atlas of peripheral blood mononuclear cells to understand disease trajectories in idiopathic pulmonary fibrosis [J]. *Am J Respir Crit Care Med*, 2024, 210(4): 385–387.
 - [16] Kumar V, Hertz M, Agro A, *et al.* Type 1 invariant natural killer T cells in chronic inflammation and tissue fibrosis [J]. *Front Immunol*, 2023, 14: 1260503.
 - [17] Shi S Y, Liu Y, Qiu X H, *et al.* Correlation between serum bilirubin levels and the severity as well as the prognosis of idiopathic pulmonary fibrosis [J]. *Chron Respir Dis*, 2020, 17: 1479973120957676.
 - [18] Cartier A, Hla T. Sphingosine 1-phosphate: Lipid signaling in pathology and therapy [J]. *Science*, 2019, 366(6463): eaar5551.
 - [19] Baek A R, Hong J S, Song K S, *et al.* Spermidine attenuates bleomycin-induced lung fibrosis by inducing autophagy and inhibiting endoplasmic reticulum stress (ERS)-induced cell death in mice [J]. *Exp Mol Med*, 2020, 52(12): 2034–2045.
 - [20] Shin J, Lee Y M, Jung S, *et al.* MUDENG, a component of adaptor complex 5, mediates TRAIL- and TMZ-triggered apoptosis in glioblastoma (GBM) via multiple pathways [J]. *Apoptosis*, 2026, 31(4). DOI:10.1007/s10495-026-02293-3.
 - [21] Kaminska K, Cancellieri F, Quinodoz M, *et al.* Bi-allelic variants in three genes encoding distinct subunits of the vesicular AP-5 complex cause hereditary macular dystrophy [J]. *Am J Hum Genet*, 2025, 112(4): 808–828.
 - [22] Nanri Y, Nunomura S, Honda Y, *et al.* A positive loop formed by SOX11 and periostin upregulates TGF- β signals leading to skin fibrosis [J]. *J Invest Dermatol*, 2023, 143(6): 989–998.
 - [23] Mitamura Y, Nunomura S, Nanri Y, *et al.* Hierarchical control of interleukin 13 (IL-13) signals in lung fibroblasts by STAT6 and SOX11 [J]. *J Biol Chem*, 2018, 293(38): 14646–14658.

- [24] Zhang J, Sun X Y, Lai J Y, *et al.* Unraveling the SOX11-IGF2 signaling axis in early lung development: Implications for lung disease [J]. *Commun Biol*, 2025, 8(1): 880.
- [25] Lomelí H. ZMIZ proteins: Partners in transcriptional regulation and risk factors for human disease [J]. *J Mol Med*, 2022, 100(7): 973-983.
- [26] Lee S H, Zhu C F, Peng Y, *et al.* Identification of a novel role of ZMIZ2 protein in regulating the activity of the Wnt/ β -catenin signaling pathway [J]. *J Biol Chem*, 2013, 288(50): 35913-35924.
- [27] Ma R, Xie Y P, Dong Y, *et al.* Shensiqigui tablets alleviate bleomycin-induced pulmonary fibrosis by inhibiting the hedgehog/Wnt- β -catenin pathway [J]. *Histol Histopathol*, 2025, 40(10): 1639-1649.
- [28] Uematsu K, Yoshimura A, Gemma A, *et al.* Aberrations in the fragile histidine triad (FHIT) gene in idiopathic pulmonary fibrosis [J]. *Cancer Res*, 2001, 61(23): 8527-8533.
- [29] 谷雨, 段斐, 党志勇, 等. 肺泡 II 型上皮细胞染色体脆性位点基因 FHIT 与肺间质纤维化的关系 [J]. *解剖学杂志*, 2013, 36(4): 720-723.
- [30] Birnhuber A, Egemnazarov B, Biasin V, *et al.* CDK4/6 inhibition enhances pulmonary inflammatory infiltration in bleomycin-induced lung fibrosis [J]. *Respir Res*, 2020, 21(1): 167.
- [31] Xia J R, Wang D, Guo W, *et al.* Exposure to micron-grade silica particles triggers pulmonary fibrosis through cell-to-cell delivery of exosomal miR-107 [J]. *Int J Biol Macromol*, 2024, 266: 131058.
- [32] 张万方, 王琳, 潘鹏涛, 等. lncSIL 通过 EZH2/P21/CDK6 信号通路负向调控 TGF- β 1 诱导的肺泡上皮细胞间质转化 [J]. *安徽医科大学学报*, 2024, 59(4): 600-604.
- [33] 张欣欣, 谢铨子, 谢依, 等. 系统药理学探讨山柰酚治疗特发性肺纤维化的分子机制 [J]. *世界中医药*, 2025, 20(4): 538-546.
- [34] 蒋怡芳, 范晓杰, 刘晓, 等. 柚皮素对博来霉素诱导的小鼠肺纤维化的改善作用及其作用机制 [J]. *安徽医科大学学报*, 2021, 56(2): 202-207.
- [35] 张艳, 马文, 杨卉, 等. 橙皮苷对模型小鼠肺纤维化的改善作用 [J]. *中国药业*, 2025, 34(4): 71-76.
- [36] Wang L B, Lin F, Liu Y L, *et al.* Wogonin protects against bleomycin-induced mouse pulmonary fibrosis via the inhibition of CDK9/p53-mediated cell senescence [J]. *Front Pharmacol*, 2024, 15: 1407891.
- [37] 汪志斌. 槲皮素与血根碱治疗肺纤维化作用的分子机制的初步研究 [D]. 辽宁: 辽宁大学, 2024.

[责任编辑 潘明佳]