DOI:10.13193/j.issn.1673-7717.2026.05.023

基于生物信息学探讨原发性干燥综合征和特发性肺间质纤维化的共病机制及潜在中药预测

叶晓滨[1,2],罗芳涛[3]
(1.厦门大学附属中山医院,福建 厦门 361004;2.厦门市思明区筼筜街道社区卫生服务中心,福建 厦门 361012;
3.厦门医学院附属第二医院,福建 厦门 361021)

摘要:目的 通过生物信息学方法获得原发性干燥综合征(Primary Sjögren's syndrome,pSS)-特发性肺间质纤维化(Idiopathic pulmonary fibrosis,IPF)的差异基因及所涉及途径,匹配预测治疗pSS-IPF的潜在中药。方法 从GEO数据库分别下载训练数据集和验证集的基因芯片数据,确定pSS-IPF样本中的差异基因、蛋白互作网络,并应用CytoHubba插件鉴定关键基因,MCODE插件聚类分析并挖掘疾病相关功能模块,利用TRRUST数据库预测共同差异表达基因的转录因子。对关键基因的GO功能和KEGG通路进行分析,阐明关键基因的生物学过程。同时采用验证集差异基因表达情况对前面分析的结果进行验证。最后利用Coremine Medical数据库预测对关键基因具有显著调控作用的中药。结果 从GEO数据库获得pSS-IPF差异表达基因136个,导入STRING数据库进行检索得到127个节点和474条边组。MCODE插件聚类分析并挖掘5个疾病相关功能模块,CytoHubba插件寻找关系网络数据中的25个Hub基因,上传至GeneMANIA数据库,得出*MX*1、*IFI*44*L*、*ISG*15、*RSAD*2等与核心基因最密切的10个基因,然后对以上基因进行GO注释和KEGG通路分析。富集分析结果显示,内质网蛋白质加工、病毒感染相关信号通路、脂质与动脉粥样硬化、氨基酸的生物合成和细胞质DNA传感途径信号通路等通路对pSS-IPF的发病至关重要。经TRRUST分析获得*IRF*1、*SP*1和*E*2*F*1以上3个主要的转录因子,还有*CASP*7、*TNFSF*10、*HSPA*5和*PHGDH*受转录因子调控的基因。采用COREMINE预测关键基因和通路,结果显示人参、菟丝子、雷公藤、苦参、枸杞、甘草、冬虫夏草、人参叶是治疗pSS-IPF的核心中药。结论 本研究明确了pSS-IPF的核心基因及基因所富集的通路,探讨了pSS-IPF主要涉及的生物学过程,并采用验证数据集验证了Hub基因,为pSS-IPF治疗的新靶点选择提供了支撑,预测了针对核心基因的治疗中药,为pSS-IPF的治疗和研究提供了依据。

关键词:生物信息学;pSS;IPF;差异基因分析;中药预测
中图分类号:R259.32;R256.3 **文献标志码**:A **文章编号**:1673-7717(2026)05-0119-13

基金项目:国家重点研发计划项目(2021YFC3312218);成都中医药大学大学生科研实践创新项目(ky-2023002);成都中医药大学临床医学院研究生科研创新实践项目(CXZD2021007)
作者简介:叶晓滨(1982-),男,福建厦门人,主管中药师,学士,研究方向:中药学。
通讯作者:罗芳涛(1964-),男,福建厦门人,主任医师,学士,研究方向:内分泌与骨质疏松等。E-mail:4283298@qq.com。

[21] 刘培,俞赟丰,杨欣雨,等. 基于"痰、瘀、毒"理论探讨冠心病炎症机制和中药干预策略[J]. 中国实验方剂学杂志,2023,29(12):185-191.
[22] KAPLAN M H, SCHINDLER U, SMILEY S T, et al. Stat6 is required for mediating responses to IL-4 and for development of Th2 cells[J]. Immunity, 1996, 4(3): 313-319.
[23] FAN Y G, ZHANG W X, HUANG X S, et al. Senescent-like macrophages mediate angiogenesis for endplate sclerosis *via* IL-10 secretion in male mice[J]. Nat Commun, 2024, 15(1): 2939.
[24] MEIZLISH M L, KIMURA Y, POPE S D, et al. Mechanosensing regulates tissue repair program in macrophages[J]. Sci Adv, 2024, 10(11): eadk6906.
[25] YAN L F, WANG J, CAI X, et al. Macrophage plasticity: signaling pathways, tissue repair, and regeneration[J]. MedComm, 2024, 5(8): e658.
[26] 朱玲玲,谭莹,佘云,等. 基于中医阴阳学说探讨巨噬细胞极化在糖尿病肾脏疾病中的作用[J]. 中国中医药信息杂志,2024,31(10):19-23.
[27] 杨虹,周世欣,李红梅,等. 基于STAT6/PPAR-γ通路探讨升陷化纤方及其拆方调控M2型巨噬细胞极化对肺纤维化的影响[J]. 中国中医药信息杂志,2025,32(1):113-119.
[28] 李红蓉,常丽萍,秘红英,等. 巨噬细胞对内皮细胞凋亡的影响及通心络的干预作用[J]. 中国实验方剂学杂志,2017,23(13):121-127.
[29] 李红蓉,常丽萍,刘玉金,等. 通心络对巨噬细胞极化的影响[J]. 中国药理学通报,2017,33(4):577-580.
[30] 李会玉,孔青,郝彦雷. 中药通心络胶囊对糖尿病足大鼠nephrin、podocin水平、巨噬细胞M1表型及VEGF蛋白的作用[J]. 中国老年学杂志,2024,44(1):137-142.
[31] 贾承明,马静,陈昌波,等. 补肾活血方对去卵巢小鼠冠脉微循环的保护作用及机制研究[J]. 中国中西医结合杂志,2019,39(2):230-236.
[32] 王亚萍,张金梅,周漫,等. 基于Notch通路探讨三七总皂苷抑制巨噬细胞M1型极化的机制研究[J]. 现代中西医结合杂志,2024,33(21):2921-2930.
[33] 孙颖,刘玲,施晓艳,等. 丹皮酚通过下调miR-155/JAK1-STAT1通过抑制巨噬细胞M1极化[J]. 中国中药杂志,2020,45(9):2158-2164.
[34] ZHANG X, LIU M H, QIAO L, et al. Ginsenoside Rb1 enhances atherosclerotic plaque stability by skewing macrophages to the M2 phenotype[J]. J Cell Mol Med, 2018, 22(1): 409-416.
[35] 何信用,张哲,贾连群,等. 黄芪甲苷调控MAP3K8介导的细胞焦亡及巨噬细胞极化交互作用防治动脉粥样硬化的机制[J]. 中华中医药杂志,2023,38(5):2311-2316.

Exploring Comorbidity Mechanism and Potential Traditional Chinese Medicine Prediction of Primary Sjogren's Syndrome and Idiopathic Pulmonary Interstitial Fibrosis Based on Bioinformatics

YE Xiaobin[1,2], LUO Fangtao[3]

(1. Zhongshan Hospital Affiliated to Xiamen University, Xiamen 361004, Fujian, China;
2. Community Health Service Center of Tongdang Street, Siming District, Xiamen 361012, Fujian, China;
3. The Second Affiliated Hospital of Xiamen Medical College, Xiamen 361021, Fujian, China)

Abstract: *Objective* To obtain the differential genes of primary Sjögren's syndrome (pSS) – idiopathic pulmonary fibrosis (IPF) and related pathways by bioinformatics methods, and to match and predict the potential Chinese medicines for treating pSS – IPF. *Methods* The gene chip data of training data set and validation data set were downloaded from GEO database, and the interaction networks of differential genes and proteins in pSS – IPF samples were determined. The key genes were identified by CytoHubba plug – in, and the disease – related functional modules were clustered and analyzed by MCODE plug – in. The transcription factors of common differentially expressed genes were predicted by TRRUST database. The GO function and KEGG pathway of key genes were analyzed to clarify the biological process of key genes. At the same time, the results of the previous analysis were verified by the differential gene expression in the verification set. Finally, the Coremine Medical database was used to predict the traditional Chinese medicine with significant regulatory effect on key genes. *Results* A total of 136 differentially expressed genes of pSS – IPF were obtained from GEO database, and 127 nodes and 474 edge groups were obtained by importing them into STRING database. The MCODE plug – in performs cluster analysis on the PPI network and mines five disease – related functional modules. The CytoHubba plug – in searched 25 Hub genes in the relational network data and uploaded them to the GeneMANIA database, and obtained 10 genes which were most closely related to Hub genes, such as *MX1*, *IFI44L*, *ISG15* and *RSAD2*. Then, the Hub genes were annotated by GO and analyzed by KEGG pathway. The results of enrichment analysis showed that the protein processing of endoplasmic reticulum, signal pathway related to virus infection, lipid and atherosclerosis, amino acid biosynthesis and cytoplasmic DNA sensing pathway were very important for the pathogenesis of pSS – IPF. TRRUST analysis showed that *IRF1*, *SP1* and *E2F1* were the three main transcription factors, and *CASP7*, *TNFSF10*, *HSPA5* and *PHGDH* were also regulated by transcription factors. COREMINE was used to predict the key genes and pathways. The results showed that Renshen (Ginseng Radix et Rhizoma), Tusizi (Cuscutae Semen), Leigongteng (Radix et Rhizoma Tripterygii), Kushen (Sophorae Flavescentis Radix), Gouqi (Lycii Fructus), Gancao (Glycyrrhizae Radix et Rhizoma), Dongchongxiacao (Ophiocordyceps) and Renshenye (Ginseng Folium) were the core Chinese medicines for pSS – IPF. *Conclusion* In this study, the core genes and gene – enriched pathways of pSS – IPF were identified, the biological processes involved in pSS – IPF were discussed, and the Hub gene was verified by the verification data set, which provided the support for the selection of new targets for pSS – IPF treatment, predicted the therapeutic Chinese medicine for Hub gene, and provided a basis for the treatment and research of pSS – IPF.

Keywords: bioinformatics; pSS; IPF; differential gene analysis; traditional Chinese medicine prediction

原发性干燥综合征(Primary Sjögren's syndrome, pSS)是一种自身免疫系统的慢性炎症性疾病,其特征在外分泌腺中存在淋巴细胞浸润介导。腺体功能的丧失可能与唾液腺纤维化有关,由成纤维细胞表达的一组独特的基因导致的,可作为原发性疾病或出现在其他结缔组织疾病中。肺富含血管和结缔组织,因此肺纤维化在pSS患者中很常见[1]。特发性肺间质纤维化(Idiopathic pulmonary fibrosis, IPF)是一种病因不明的慢性、进行性纤维化间质性肺病,其特征是反复的上皮细胞损伤、肺泡结构破坏和成肌纤维细胞的促纤维化介质诱导的细胞外基质沉积。IPF的发病率和病死率逐年增加,其预后极差,中位生存期仅为3~5年[2]。目前对pSS – IPF的共病机制文献报道相对较少,真实的患病率和严重程度都被严重低估。然而,一些证据表明两者可能有一定的联系。例如:pSS患者出现干咳和气短提示可能伴有肺纤维化[3];部分pSS合并肺结节的患者具有纤维化的性质[4]。

基于以上背景,我们采用生物信息学对pSS和IPF各两组数据集进行分析,通过训练数据集和验证数据集差异性表达基因筛选、基因本体(Gene ontology, GO)[5]和京都基因与基因组百科全书(Kyoto encyclopedia of genes and genomes, KEGG)分析[6]、蛋白质 – 蛋白质相互作用网络(Protein – protein interaction networks, PPI)[7]构建确定了对疾病的早期诊断和治疗可能产生显著影响的潜在枢纽基因和通路。最后,通过训练数据集和验证数据集所得结果的相互验证来提升分析的科学性,并寻找潜在的治疗靶点以及对应中药。中医药有多靶点、多途径的整体疗效优势,能拓展临床辨证思路,有利于理法方药的制定,为临床中药组方加减以及新药开发提供依据。具体流程图见插页XXⅣ图1。

1 材料与方法

1.1 GEO数据的下载与预处理 分别以"primary Sjögren's syndrome""idiopathic pulmonary fibrosis"为检索词,选择"Homo sapiens"为研究对象检索高通量基因表达(GENE EXPRESSION OMNIBUS, GEO)数据库[8],下载基因矩阵文件与平台注释文

件。获得所需芯片 GSE40611、GSE84844、GSE21369、GSE185691。随机选取 GSE84844 和 GSE185691 两个基因表达谱作为训练数据集，GSE40611 和 GSE21369 作为验证集。分析训练数据集差异性表达基因（differentially expressed genes，DEGs），为了提高 DEG 的可靠性，将 P 值 <0.05 同时 $|\log FC| \geq 0.263$（FC = 1.2 倍）为阈值进行筛选。根据平台注释文件提供信息将探针转换成对应的基因 ID，利用 perl 对基因重注释，对基因芯片进行校正、标准化、log2 转化处理[9]。

1.2 训练数据集差异表达基因的筛选 对训练数据集 GSE84844 和 GSE185691 中实验组与对照组间的 DEGs 进行分析，筛选参数 logFC > 0 为上调基因，logFC < 0 为下调基因，分别得到数据集中具有差异表达的上调基因和下调基因[10]，DEGs 通过绘制火山图和聚类热图进行可视化，同时，对数据集共同具有的上调基因和下调基因进行层次聚类，获取共差异表达上调基因（up - DEGs）和共差异表达下调基因（down - DEGs）[11]。

1.3 训练数据集共同 DEGs GO 和 KEGG 通路富集分析 对共同 DEGs 进行 GO 功能富集分析和 KEGG 通路富集分析。将所获得的数据导入 Sangerbox 平台并进行可视化，展示排名前十的生物学过程（Biological process，BP）[12]、细胞组分（Cellular composition，CC）[13] 和分子功能（Molecular function，MF）[14] 以及 KEGG 信号通路。

1.4 构建训练数据集共同 DEGs 蛋白互作网络，从中筛选功能模块和关键基因 将训练数据集共同 DEGs 上传至字符串（Search tool for the retrieval of interacting genes/proteins，STRING）数据库[15] 进行分析，筛选阈值设定为综合得分 ≥ 0.15。将关系网络数据导入 Cytoscape 3.7.2 软件，利用 MCODE 插件对该 PPI 网络进行聚类分析并挖掘疾病相关功能模块，其中设置参数为 Degree Cutoff ≥ 2，Node Score Cutoff ≥ 0.2，K - core ≥ 2，Max. Depth = 100。利用 cytoHubba 插件寻找关系网络数据中的 Hub 基因，Degree 算法筛选连接度最高的 25 个基因，作为表达基因中的 Hub 基因并进行绘图。将获得的 25 个 Hub 基因上传至 GeneMANIA 数据库，在线分析蛋白共表达、物理相关性、预测、共定位和基因相关性等蛋白互作方式，并绘制差异表达基因的蛋白质相互作用网络。然后对 Hub 基因进行 GO 注释和 KEGG 通路分析。

1.5 Hub 基因表达差异的验证 GSE40611 和 GSE21369 数据集分别是 GEO 数据库中 pSS 和 IPF 来源的数据集，包括实验组和对照组的样本。分析 GSE40611 数据集中核心基因的表达水平，与 GSE21369 数据集中核心基因的表达水平进行比较。

1.6 转录因子调控网络和差异分析 利用转录调控关系文本（Transcriptional regulatory relationships unraveled by sentence - based text mining，TRRUST）数据库[16] 预测共同 DEGs 的转录因子（TF），通过 Cytoscape 3.7.2 软件构建 Hub 基因 - TF 调控网络。读取其中一个疾病校正后的表达数据文件和 TF 列表文件绘制出 TF 差异分析小提琴图。

1.7 中药治疗靶点预测 Coremine Medical 数据库是一个集基因本体论、蛋白表达、解剖和中草药等医学综合信息的开放检索平台[17]，通过输入基因名称可检索出与该基因相关的疾病及潜在治疗药物。借助 Coremine Medical 数据库查询关键基因相关信息，以 $P < 0.05$ 的结果作为筛选标准，检索出相关潜在中药[18]。若药物数量过多，则根据中医药理论知识和临床常用为原则进一步筛选。然后于中药系统药理学数据库（TCMSP）中，以口服利用度（OB）≥30%，类药性（DL）≥0.18 为阈值筛选出相关中药的蛋白质有效成分。利用 Uniprot 数据库将已获得的有效蛋白质成分转换成对应的 gene symbol，删除无匹配结果的有效成分。最后，将数据导入 Cytoscape 3.7.2 软件，构建“中药 - 有效成分 - 靶点”网络并将其可视化，并对预测的中药进行性味、归经、功效的归纳。

2 结果

2.1 DEGs 分析结果 通过分析基因芯片 GSE185691 和 GSE84844 中的数据，结果 GSE185691 共筛选出 3478 个肺移植患者和 IPF 患者的 DEGs，其中上调基因 1658 个（47.67%），下调基因 1820 个（52.33%）。GSE84844 共筛选出 2266 个 pSS 患者和健康志愿者的 DEGs，包括上调基因 1561 个（68.89%），下调基因 705 个（31.11%）。使用 Venn 图对两个系列的 DEGs 进行取交集，最后筛选出 136 个基因显著差异表达，其中共差异表达上调基因（up - DEGs）62 个和共差异表达下调基因（down - DEGs）74 个，见插页ⅩⅩⅤ图 2。

2.2 DEGs 功能富集分析 通过“clusterProfiler”R 语言包对上述两数据集 136 个共同 DEGs 进行 GO 和 KEGG 富集分析，如插页ⅩⅩⅥ图 3 所示，结果显示 158 个 BP 条目、37 个 CC 条目、39 个 MF 条目和 9 个 KEGG 条目被显著富集。GO 分析结果显示，上调基因主要参与对病毒的反应、孕激素代谢过程等生物学过程；下调基因主要参与受体介导的胞吞作用、内质网中的蛋白质折叠等生物学过程。KEGG 通路富集分析结果显示，上调基因主要富集在细胞因子 - 细胞因子受体相互作用和人乳头瘤病毒感染通路；下调基因主要富集在细胞凋亡、造血细胞谱系、脂质与动脉粥样硬化等信号通路。

2.3 PPI 网络及关键模块和 Hub 基因筛选 通过 STRING 数据库检索上述 136 个共同 DEGs 的潜在互作关系，设置蛋白种类为“Homosapiens”，相互作用阈值为 0.15，隐藏没有联系的节点，包括 127 个节点和 474 条边组。将 PPI 网络图导入 Cytoscape 3.7.2 软件，其中 71 个基因下调和 56 个基因上调。通过 MCODE 插件对基因进行聚类分析，识别出 5 个重要的基因簇模块，以上获取的基因为关键基因，其主要子网络 MCODE1 基因簇模块中的基因是在孕酮代谢过程、对病毒的防御反应、对共生体的防御反应、C21 - 类固醇激素代谢过程、视黄酸生物合成过程、二萜生物合成过程、内质网中的蛋白质折叠等生物过程中发挥重要作用的关键基因。进一步使用 CytoHubba 插件根据 Degree 算法获取连接度前 25 的基因作为 Hub 基因并作图。应用 GeneMANIA 网站绘制 Hub 基因相互作用图，同时分析互作蛋白的功能。PPI 网络中的节点代表相应的蛋白，且节点中的圆饼图分别用不同颜色显示了基因所参与的生物学功能[19]，网络中的连线表示蛋白质间的相互作用类型。与 Hub 基因最密切的 10 个基因是 *MX1*、*IFI44L*、*ISG15*、*RSAD2*、*ADAR*、*DDX58*、*LRRFIP2*、*OAS2*、*OAS1*、*IFIT3*。在基因功能预测界面显示，其主要富集的生物学功能包括对Ⅰ型干扰素的反应、对Ⅰ型干扰素的细胞反应、对病毒的反应、病毒基因组复制、病毒基因组复制的调节、病毒过程的调节、共同表达、物理相互作用。其中主要的 Hub 基因有 *MAVS*、*BST2*、*MX2*、*PCBP2*、*XAF1*、

*DHX*9，这 6 个基因编码的蛋白质，与其他相关基因的蛋白质存在共表达、共定位、共享蛋白质域等交互作用，一起参与调节，见插页ⅩⅩⅦ－插页ⅩⅩⅧ图 4。

2.4 Hub 基因功能富集分析 通过“clusterProfiler”R 语言包对上述 Hub 基因进行 GO 和 KEGG 富集分析，发现 Hub 基因被富集到 309 个不同的 GO 子集中，包括 BP 有 224 个，CC 有 26 个，MF 有 59 个。Hub 基因在 BP 中病毒的反应相关性高，其中对病毒和共生体的防御相对靠前；在 CC 中内质网、细胞核所在比例较高；在 MF 中主要富集在 RNA 结合。KEGG 主要富集通路为内质网蛋白质加工、病毒感染相关信号通路、脂质与动脉粥样硬化、氨基酸的生物合成和细胞质 DNA 传感途径信号通路等。见插页ⅩⅩⅨ图 5。

2.5 Hub 基因表达差异的验证 本研究选取 GSE21369 和 GSE40611 芯片作为验证集进行验证分析。GSE21369 含特发性肺纤维化（IPF）患者肺组织样本与非特异性间质性肺炎（NSIP）患者肺组织样本，GSE40611 含健康组织样本、pSS 患者腮腺组织样本和 SS 患者腮腺组织样本，验证 Hub 基因在两个验证集中的表达水平。分析结果显示，在 GSE21369 样本中，*AGO*2、*PCBP*2 表达水平差异均有统计学意义（$P<0.05$），并且 IPF 患者肺组织样本表达水平明显上调。为对 GSE21369 数据集的可靠性进行检验，提高 Hub 基因的本地特异性，并能够识别出贯穿于 pSS－IPF 的发生发展全过程的潜在生物标记，利用 GSE40611 数据集对上述 Hub 基因的表达水平进行验证。结果显示，在 GSE40611 样本中，*WDTC*1、*BRD*4、*CXCL*11、*IARS*、*XAF*1、*DHX*9 表达水平差异均有统计学意义（$P<0.05$）；*DNAJC*10、*CASP*7、*HSPA*5 表达水平有显著统计学意义（$P<0.01$）；*TNFSF*10、*DNAJA*1 表达水平有极显著统计学意义（$P<0.001$），见插页ⅩⅩⅧ图 6。

2.6 转录因子调控网络和差异分析 为了进一步验证交集基因和 pSS－IPF 的关系，我们将 Hub 基因上传到 TRRUST 数据库后经过分析获得转录因子共表达数据，然后将数据整理后导入到 Cytoscape 3.7.2 中，修饰过后获得转录因子共表达网络图[20]，如插页ⅩⅩⅨ图 7 所示。其中 *IRF*1、*SP*1 和 *E2F*1 是调控 Hub 基因表达的 3 个主要的转录因子，图中 *CASP*7、*TNFSF*10、*HSPA*5 和 *PHGDH* 是受转录因子调控的基因，*E2F*1 表达量差异有统计学意义（$P<0.05$），*SP*1 表达量差异有显著统计学意义（$P<0.01$）。分析结果显示，*IRF*1、*SP*1 和 *E2F*1 可以通过调控 Hub 基因的表达，与对病毒的防御反应、对共生体的防御反应、内质网中的蛋白质折叠、Ⅰ－kappaB 激酶/NF－kappaB 信号的正向调节的生物学过程和氨基酸的生物合成、碳代谢、凋亡等通路密切相关，所预测的生物途径或通路与 Hub 基因 GO 和 KEGG 富集分析研究结果一致。

2.7 中药预测与“中药－活性成分－靶基因”网络构建 通过 Coremine Medical 预测和挖掘潜在干预 pSS－IPF 的中药，以 $P<0.05$ 作为筛选条件，将得到的受转录因子调控 Hub 基因 *CASP*7、*TNFSF*10、*HSPA*5 和 *PHGDH* 导入，对所得中药进行标准化处理，预测得出 117 种中药，中药成分槲皮素、人参皂苷 RH_2 等与内质网中的蛋白质折叠、I－kappaB 激酶/NF－kappaB 信号的正向调节、I－kappaB 激酶/NF－kappaB 信号的调节、*IRE*1 介导的未折叠蛋白反应、I－κb 激酶/NF－κb 信号传导、蛋白质复性等生物过程相关，将结果导入 Cytoscape 软件中进行可视化展示，其中人参、菟丝子、雷公藤、苦参、枸杞、甘草、冬虫夏草、人参叶为常见中药，且度值较高，其治疗 pSS 和 IPF 的机制可能与通过作用于 *CASP*7、*TNFSF*10、*HSPA*5 和 *PHGDH* 进行干预。对预测的 117 种中药进行归经分类，主要集中在肝、肺、胃、脾和肾经，且与清热解毒、补虚、活血化瘀、化痰止咳平喘相关。在四气的分析中，使用频次最高的是温、寒、平，在五味的分析中苦、辛、甘使用最多，如插页ⅩⅩⅩ图 8、表 1 所示。

表 1 Hub 基因中药预测结果

Hub 基因	相关中药
*CASP*7	首乌藤、猫爪草、臭灵丹草、木香、菟丝子、独活、冬凌草、葱白、白花蛇舌草、青葙子、贯众、铁扫帚、枳壳、枳实、香橼、苦参、冬虫夏草、猪苓、莪术、藤黄、姜黄、郁金、龙葵、茜草、知母、喜树皮、喜树果、麻黄根、益智仁、沉香、半枝莲、苏木、麻黄、广藿香、枸杞、夏枯草、鱼腥草、金盏银盘、鬼针草、檀香、白术、苍耳子、石榴花、石榴皮、石榴子、方蟹、灵芝、补骨脂、姜皮、生姜、干姜、人参、人参叶、茶树根、啤酒花、沙棘
*TNFSF*10	猫眼草、藤黄、雷公藤、白首乌、旋覆花、金沸草、黄花夹竹桃叶、甘草、黄花夹竹桃、海藻、蛇蜕、瓦楞子、喜树皮、喜树果、蛇床子、黄荆叶、黄荆子、艾叶、干漆、石刁柏
*HSPA*5	蝉花、马勃、威灵仙、风轮菜、竹茹、天麻、紫竹根、黄芩、淡竹叶、紫草、藤黄、玄参、淫羊藿、高良姜、金樱根、金樱子、补骨脂、木蝴蝶、大黄、紫菀、生地黄、枸杞、三七叶、三七花、三七、赤芍、蒲公英、绞股蓝、地胆草、穿山龙、苦参、牡丹皮、太子参、雷公藤、茶树根、甘草、黄精、丹参、黄芪、菟丝子、泽泻、巴戟天、黄药子、车前草、小叶莲、车前子、荔枝核、防己、地骨皮、山楂、肉苁蓉
PHGDH	蝉花、冬虫夏草

3 讨论

本研究筛选 pSS－IPF 差异基因，探究其作用机制，并对其中药治疗靶点进行预测分析。

利用 GEO 数据库中的基因表达数据，系统地分析了 pSS－IPF 的共病机制，发现 pSS－IPF 筛选出的 136 个 DEGs 主要参与免疫反应、细胞因子信号传导、细胞增殖、细胞凋亡、细胞分化等生物学过程，主要富集在与病毒防御和免疫调节相关的信号通路，以上与两种疾病的发生发展密切相关，其结果为进一步探讨 pSS－IPF 发生发展中转录因子的作用机理奠定基础。

PPI 网络对上述进行分析并通过 MCODE 插件挖掘出 5 个具有生物学意义的模块，这些模块中的关键基因 *MX*2 通过介导多种免疫细胞发挥免疫作用；*DHX*9 是含有 *DEAH* 的 RNA 解旋酶家族的成员，在广泛的细胞过程中具有许多调节作用；*DNAJA*1 具有影响细胞生长和增殖的调节作用；*AGO*2 是 miRNA 成熟、稳定性和功能的主要调节因子；*ERC*1 是一种边缘结合蛋白。CytoHubba 插件鉴定出连接度较高的 Hub 基因，分别是 *DHX*9、*HSPA*5、*MX*2、*TNFSF*10、*PHGDH* 等。其中，*CASP*7、*TNFSF*10 等基因在 pSS－IPF 中都显著上调，可能通过调节病毒防御反应、内质网蛋白质折叠、I－kappaB 激酶/NF－kappaB 信号的正向调控、调节树突状细胞细胞因子的产生导致 pSS－IPF 发展。此外，*HSPA*5、*PHGDH* 等在 pSS－IPF 中显著下调，这些基因主要参与对 α 干扰素的反应、*IRE*1 介导的未折叠蛋白反应、核质运输的调节、α－干扰素产生的正调节、细胞对 dsRNA 的反应。结果表明，这些靶点主要分布于细胞质、膜结合囊泡、内质网、液泡等部位，具有与蛋白、受体、氧原子结合的功能，还具有蛋白激酶的活性，可参与机体对多种物质刺激的

反应过程[21]，参与蛋白、激素代谢[22]、脂质代谢、生物合成、免疫应答[23]和调控多种蛋白激酶活性[24]等作用。

对上述关键基因上传到 TRRUST，发现 *IRF*1、*SP*1 和 *E2F*13 在调控 pSS 和 IPF 发挥着重要的作用。*IRF*1[25]是一种介导干扰素效应的核转录因子，通过协调多种细胞信号级联，影响广泛的抗病毒反应，从脂质代谢的改变到蛋白质降解机制。*Sp*1[26]、*E2F*1[27]参与细胞增殖、凋亡和分化以及血管生成的调节，可以作为 pSS 的早期标志物，尤其是缺乏 *RF* 或 *Ro/La* 抗体的 pSS 患者。*E2F*1 转录因子在细胞周期和细胞凋亡的调节中起关键作用，表达异常可能导致 IPF 微环境中成纤维细胞增殖和血管生成减少，可能是 IPF 发病机制的关键调节因子。*CASP*7、*TNFSF*10、*HSPA*5 和 *PHGDH* 是受 *IRF*1、*SP*1 和 *E2F*13 调控的基因。其中，*TNFSF*10[28]在上皮细胞、内皮细胞、成纤维细胞和肺泡巨噬细胞表达水平明显上调，通过 *MID*1 的肿瘤坏死因子相关的凋亡诱导配体使 *PP*2*A* 失活并促进纤维化，伴随相应的肺功能下降，这可能为 IPF 提供新的治疗靶点。内质网应激介导的凋亡和衰老参与了肺纤维化的进程，*HSPA*5[29]是内质网稳态的主要调节者，其表达的基因主要是响应未折叠蛋白、黑素体、甲状腺激素合成和 N－聚糖生物合成，调节内质网应激为 IPF 的研究提供了新的途径。*PHGDH*[30]参与从头生物合成途径生成丝氨酸，胶原蛋白的合成需要丝氨酸和甘氨酸的从头生物合成，该途径可能是包括 IPF 在内的器官纤维化的治疗靶点。此外，Hub 基因 *DHX*9[31]可以诱导成纤维细胞和上皮细胞中 NF－κB 介导的抗病毒细胞因子的转录，在外分泌腺、多系统和器官表达量有差异，有针对病毒的防御反应和调节细胞核中的先天免疫反应。*MX*2 在干扰素应答信号通路中显著富集，该信号通路在 IPF 发展中起着重要作用。基于以上的研究结果，进一步探索调控 pSS－IPF 的关键因素和机制，并获得了转录因子在内质网、纤维细胞和肺泡巨噬细胞功能障碍诱导疾病发生发展的初步结论。

中医药较早认识了 pSS 和 IPF 在病因病机上具有相似性。清代医家沈明宗在《医征·燥病论》中亦提到“燥属次寒”之说，概括地道出了燥邪本性属阴的特征。而肺痹津少液涸，痹结不解，咽燥少唾并伴随持续性干咳。以上符合“燥毒致病”的疾病发展过程。《医经原旨》中云：“肺痿者，皮毛痿也，盖热乘肺金，在内则为叶焦，在外则皮毛虚弱而为急薄；若热气留着不去而及于筋脉骨肉，则病生痿躄。躄者，足弱不能行也。”与 pSS 可涉及肌肉骨骼、肾脏、肺部、神经和血液等多系统和器官，伴随慢性疲劳和疼痛的症状相似。中医认为 pSS 病机为肝肾阴虚，兼见气阴两虚、瘀毒互结等证，IPF 多与肺虚、邪毒、血瘀、痰结有关。二者的发生发展都离不开“燥”“虚”和“瘀”，临床应以益气滋阴润燥之法，结合活血解毒化瘀。本次预测的治疗中药性味多为温、寒、平，苦、辛、甘中药使用最多，归属于肝、肺、胃、脾和肾经，功效与清热解毒、补虚、活血化瘀、化痰止咳平喘相关，进一步说明本研究预测符合 pSS 与 IPF 的病因病机。

预测所获得的核心中药人参、菟丝子、雷公藤、苦参、枸杞、甘草、冬虫夏草、人参叶属于上述治疗范畴。人参、枸杞、冬虫夏草和人参叶中多糖成分具有抗炎、抗氧化[32]、抗疲劳、抗衰老和免疫调节[33]的药理作用，可以改善 T 细胞、B 细胞、巨噬细胞、淋巴细胞和树突细胞的功能，以上中药补益作用体现在抗氧化和免疫调节两个方面。《长沙药解》云“人参止燥证之渴”。人参中有效成分人参皂苷 *Rg*1 对博来霉素诱导的大鼠肺纤维化具有保护作用，这可能与 TGF－β1 的下调和 *Cav*－1 的上调有关[34]。甘草酸可抑制核心岩藻糖基化修饰调节的上皮－间质转化并减轻博莱霉素诱导的肺纤维化[35]。冬虫夏草提取物可减轻小鼠间质性肺炎的病理损伤，降低细胞凋亡及促炎因子的表达，改善肺脏组织纤维化，其机制可能与抑制 *mTORC*1 活性、促使转录因子 *EB* 参与肺泡上皮细胞自噬的发生有关[36]。雷公藤多苷缓解原发性干燥综合征伴间质性肺病患者症状、改善肺功能及影像学改变等总体疗效与环磷酰胺相当，而且耐受性良好[37]。中医中药治疗 pSS 与 IPF 优势明显，心得独到，疗效颇佳。其中，冯兴华教授采用杞菊地黄汤加石斛和北沙参等治疗肝肾阴虚型 pSS，疗效明显优于对照组[38]。宋教授提出以“风、毒”分期论治结缔组织病相关间质性肺病，采用石斛、枸杞、太子参、熟地黄、山茱萸等中草药配合雷公藤多苷片，病例均有不同程度的好转，取得了满意的效果[39]。马丽[40]在“金水相生论”的指导下，探讨了虫草类中药在防治肺病中的运用与科学内涵，并进一步证实了虫草类中药可通过调节 AQPs 而发挥抗肺纤维化效应。预测所获得的核心中药契合临床中治疗 pSS 与 IPF 的遣方用药，对临床具有一定的指导价值。

本研究以 pSS、IPF 基因芯片数据为基础，利用生物信息学方法，综合运用基因表达、功能注释、通路富集等方法，寻找与疾病发病密切相关的关键基因，揭示发病机制，为疾病的临床诊治及科学研究奠定基础。以上的研究说明预测所得部分中草药对于 pSS－IPF 的治疗有效，表明预测方法的合理性。但是，本次研究数据样本数量较少，且涉及平台比较单一，需要中医药学者及专家整合多学科优势深度钻研，为认识和挖掘传统方剂找到合适的方法和思路。

参考文献

[1] LI D F, LI HB, LI W, et al. Anti－Ro52 antibody as a protective factor for pulmonary fibrosis in primary sjögren's syndrome[J]. Iran J Immunol, 2022, 19(2): 161－171.

[2] YANAGIHARA T, CHONG S G, VIERHOUT M, et al. Current models of pulmonary fibrosis for future drug discovery efforts[J]. Expert Opin Drug Discov, 2020, 15(8): 931－941.

[3] LI D F, LI H B, WANG Y J, et al. Pulmonary fibrosis in primary sjögren syndrome: computed tomography, clinical features, and associated clinical factors[J]. Pol Arch Intern Med, 2023, 133(2): 16394.

[4] ARGYROPOULOU O D, CHATZIS L G, RONTOGIANNI D, et al. Autoimmune epithelitis beyond the exocrine glands: an unusual case of anti－ro/La and scl－70 lymphocytic interstitial pneumonia[J]. Clin Exp Rheumatol, 2019,118(3): 249－251.

[5] QIU L, SUN Y X, NING H M, et al. The scaffold protein AXIN1 gene ontology, signal network, and physiological function[J]. Cell Commun Signal, 2024, 22(1): 77.

[6] LI S L, LI X, JIANG S S, et al. Identification of sepsis－associated mitochondrial genes through RNA and single－cell sequencing approaches[J]. BMC Med Genomics, 2024, 17(1): 120.

[7] ZHU Q M, HSU Y H, LASSEN F H, et al. Protein interaction net-

works in the vasculature prioritize genes and pathways underlying coronary artery disease[J]. Commun Biol, 2024, 7(1): 87.
[8] XUE Y, ZHOU L, WANG J Q. Classification of distinct osteoarthritis subtypes with different knee joint tissues by gene expression profiles[J]. Bone Joint Res, 2023, 12(12): 702-711.
[9] 郑文江, 彭紫荆, 严倩, 等. 基于网络药理学和生物信息学探讨培土生金方对慢性阻塞性肺疾病的作用机制[J]. 世界科学技术-中医药现代化, 2020, 22(9): 3191-3205.
[10] LI X, NIAN B B, TAN C P, et al. Deep-frying oil induces cytotoxicity, inflammation and apoptosis on intestinal epithelial cells [J]. J Sci Food Agric, 2022, 102(8): 3160-3168.
[11] 贾颖, 卓频, 赵爽, 等. 基于GEO数据库的产后抑郁症关键差异表达基因分析[J]. 检验医学与临床, 2022, 19(22): 3094-3097.
[12] XU C L, ILYAS M K, MAGWANGA R O, et al. Transcriptomics for drought stress mediated by biological processes in-relation to key regulated pathways in *Gossypium darwinii*[J]. Mol Biol Rep, 2022, 49(12): 11341-11350.
[13] DANENBERG E, BARDWELL H, ZANOTELLI V R T, et al. Breast tumor microenvironment structures are associated with genomic features and clinical outcome[J]. Nat Genet, 2022, 54(5): 660-669.
[14] 卻丹华, 陈王焕, 姜飞鹏, 等. 基于网络药理学和分子对接研究当归四逆汤治疗原发性痛经的机制[J]. 中国中药杂志, 2021, 46(4): 855-864.
[15] LIU X Y, REN Y L, QIN S L, et al. Exploring the mechanism of 6-Methoxydihydrosanguinarine in the treatment of lung adenocarcinoma based on network pharmacology, molecular docking and experimental investigation[J]. BMC Complement Med Ther, 2024, 24(1): 202.
[16] XIAO L, LIN S D, ZHAN F. Identification of hub genes and transcription factors in patients with primary gout complicated with atherosclerosis[J]. Sci Rep, 2024, 14(1): 3992.
[17] WEI L W, YIN F Q, CHEN C X, et al. Expression of integrin α-6 is associated with multi drug resistance and prognosis in ovarian cancer[J]. Oncol Lett, 2019, 17(4): 3974-3980.
[18] 贠张君, 王慧静, 俞仪萱, 等. 基于生物信息学技术筛选结直肠癌的差异基因和中药预测研究[J]. 中国中药杂志, 2022, 47(6): 1666-1676.
[19] ZAINAL-ABIDIN R A, AFIQAH-ALENG N, ABDULLAH-ZAWAWI M R, et al. Protein-protein interaction (PPI) network of zebrafish oestrogen receptors: a bioinformatics workflow[J]. Life (Basel), 2022, 12(5): 650.
[20] LI J M, ZHANG Z, GUO K, et al. Identification of a key glioblastoma candidate gene, *FUBP*3 based on weighted gene co-expression network analysis[J]. BMC Neurol, 2022, 22(1): 139.
[21] 陈李乙, 李佳欣, 张美晴, 等. 升麻药材化学成分及药理作用研究进展[J]. 中草药, 2023, 54(5): 1685-1704.
[22] XU C, HE Z, SONG Y F, et al. Atypical pituitary hormone-target tissue axis[J]. Front Med, 2023, 17(1): 1-17.
[23] XIA L Z, OYANG L, LIN J G, et al. The cancer metabolic reprogramming and immune response [J]. Mol Cancer, 2021, 20(1): 28.
[24] MATHIEN S, TESNIÈRE C, MELOCHE S. Regulation of mitogen-activated protein kinase signaling pathways by the ubiquitin-proteasome system and its pharmacological potential[J]. Pharmacol Rev, 2021, 73(4): 263-296.
[25] TAN G, HUANG C Y, CHEN J Y, et al. An IRF1-dependent pathway of TNFα-induced shedding in intestinal epithelial cells [J]. J Crohns Colitis, 2022, 16(1): 133-142.
[26] JIANG J F, ZHOU Z Y, LIU Y Z, et al. Role of Sp1 in atherosclerosis[J]. Mol Biol Rep, 2022, 49(10): 9893-9902.
[27] SHEU C C, CHANG W A, TSAI M J, et al. Gene expression changes associated with nintedanib treatment in idiopathic pulmonary fibrosis fibroblasts: a next-generation sequencing and bioinformatics study[J]. J Clin Med, 2019, 8(3): 308.
[28] COLLISON A M, LI J Y, DE SIQUEIRA A P, et al. TRAIL signals through the ubiquitin ligase MID1 to promote pulmonary fibrosis [J]. BMC Pulm Med, 2019, 19(1): 31.
[29] 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.
[30] NIGDELIOGLU R, HAMANAKA R B, MELITON A Y, et al. Transforming growth factor (TGF)-β promotes *de novo* serine synthesis for collagen production[J]. J Biol Chem, 2016, 291(53): 27239-27251.
[31] LINDÉN M, RAMÍREZ SEPÚLVEDA J I, JAMES T, et al. Sex influences eQTL effects of SLE and Sj? gren's syndrome-associated genetic polymorphisms[J]. Biol Sex Differ, 2017, 8(1): 34.
[32] MU S, YANG W J, HUANG G L. Antioxidant activities and mechanisms of polysaccharides[J]. Chem Biol Drug Des, 2021, 97(3): 628-632.
[33] XIA L J, ZHANG C L, LV N Y, et al. AdMSC-derived exosomes alleviate acute lung injury *via* transferring mitochondrial component to improve homeostasis of alveolar macrophages[J]. Theranostics, 2022, 12(6): 2928-2947.
[34] LIU M J, ZHANG T Y, ZANG C, et al. Preparation, optimization, and *in vivo* evaluation of an inhaled solution of total saponins of *Panax notoginseng* and its protective effect against idiopathic pulmonary fibrosis[J]. Drug Deliv, 2020, 27(1): 1718-1728.
[35] SAMAREH FEKRI M, POURSALEHI H R, SHARIFIFAR F, et al. The effects of methanolic extract of *Glycyrrhiza glabra* on the prevention and treatment of bleomycin-induced pulmonary fibrosis in rat: experimental study[J]. Drug Chem Toxicol, 2021, 44(4): 365-371.
[36] 邢双丽, 赵明娟, 金荣. 冬虫夏草提取物对肺炎小鼠肺泡上皮细胞的自噬作用及机制[J]. 西北药学杂志, 2023, 38(1): 72-77.
[37] 熊琴, 杨彦瑜, 谭敏, 等. 原发性干燥综合征相关肺间质病变治疗的研究进展[J]. 临床荟萃, 2019, 34(10): 948-952.
[38] 孔祥民, 孙娜, 庞秀, 等. 冯兴华教授中医治疗原发性干燥综合征的经验研究[J]. 中医临床研究, 2016, 8(1): 12-14.
[39] 张艳, 徐丽萍, 关天容, 等. 从"风、毒"探讨宋欣伟分期诊治结缔组织病相关间质性肺病[J]. 浙江中医药大学学报, 2021, 45(8): 829-834.
[40] 马丽, 李昌煜, 钱正明, 等. 冬虫夏草金水相生治疗肺脏疾病的作用及其生理基础[J]. 菌物研究, 2020, 18(2): 80-85.

基于生物信息学探讨原发性干燥综合征和特发性肺间质纤维化的共病机制及潜在中药预测

(正文见 119－124 页)

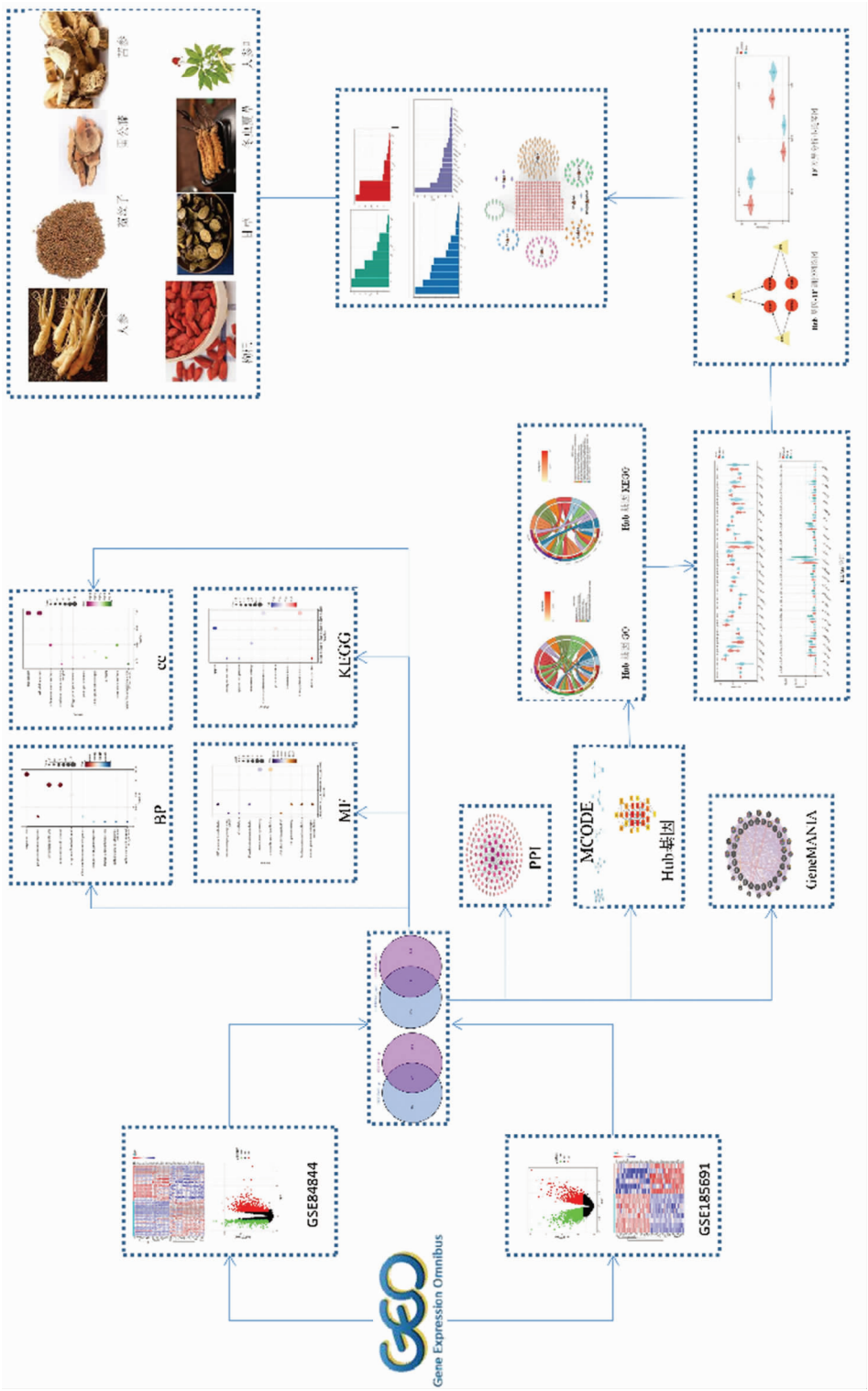
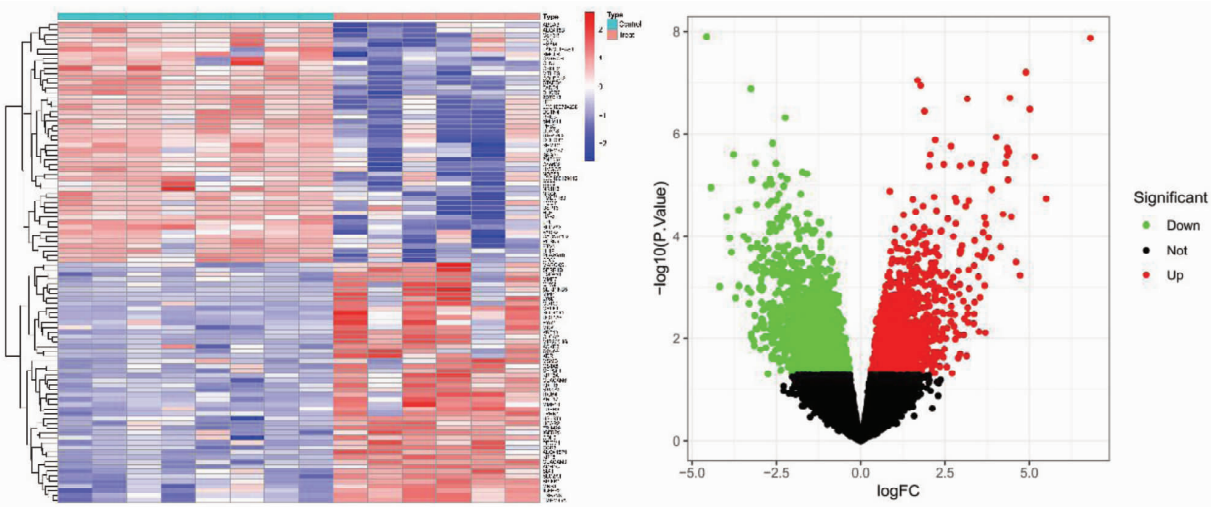
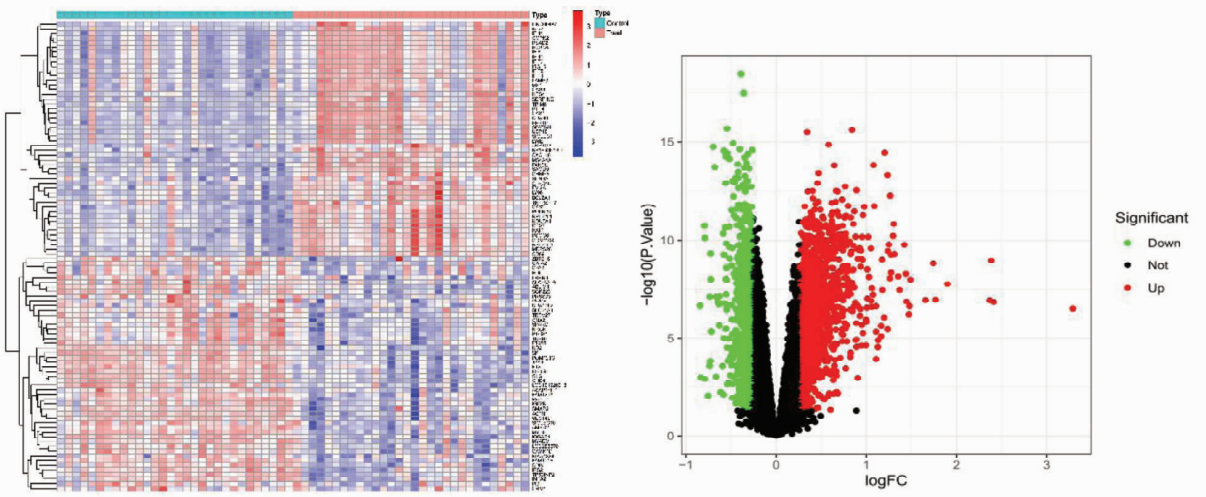


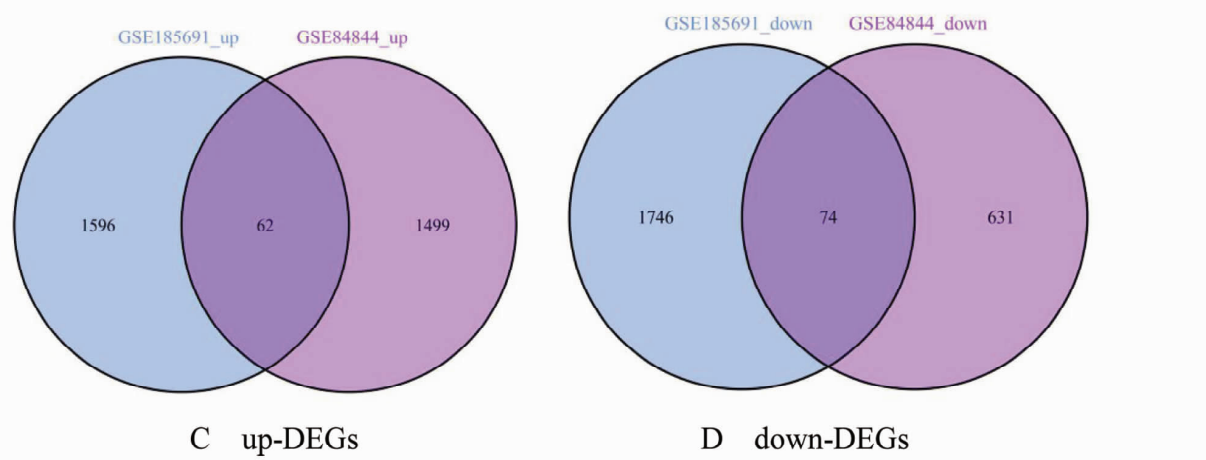
图 1　流程图



A 数据集 GSE185691



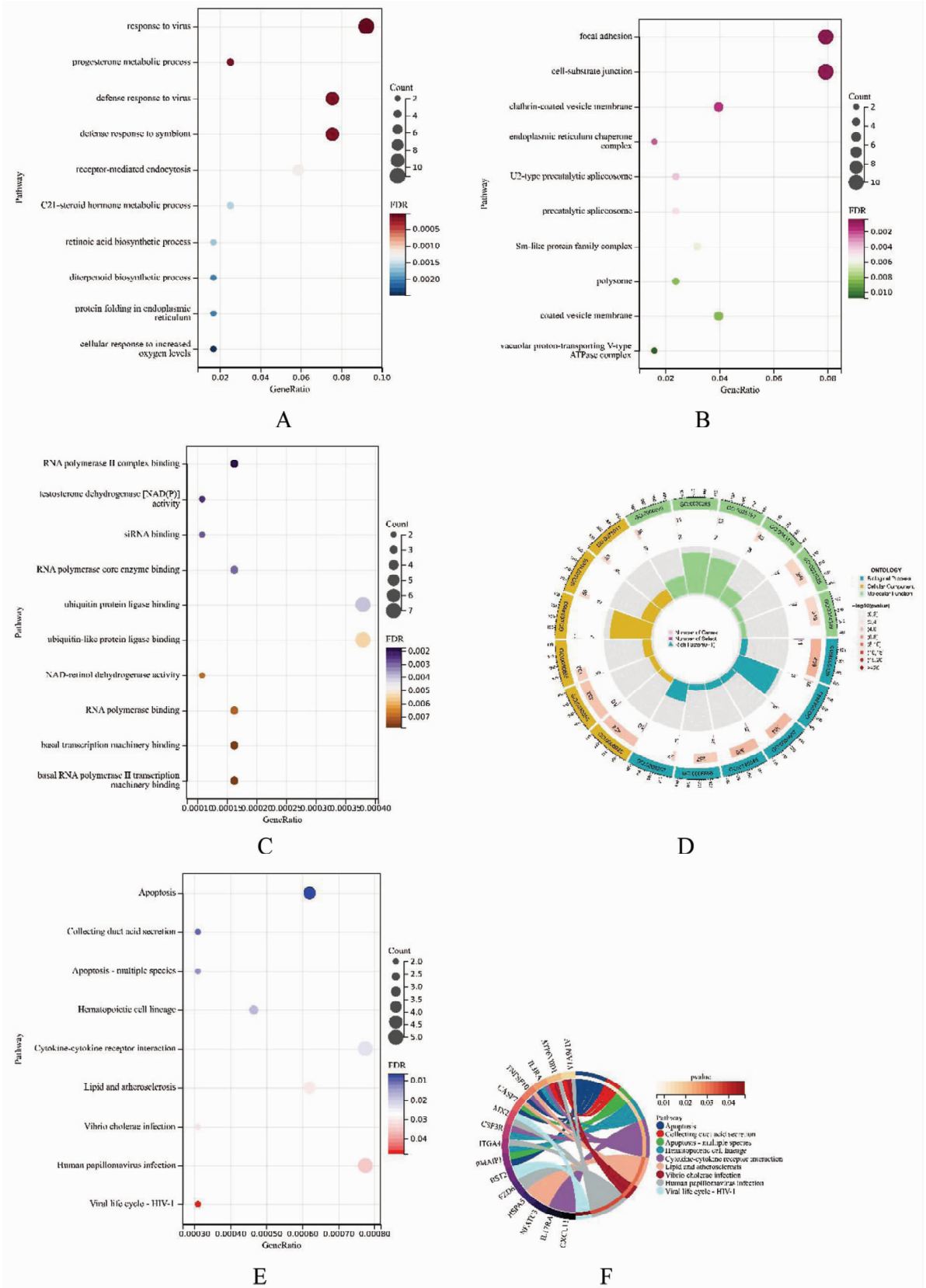
B 数据集 GSE84844



C up-DEGs D down-DEGs

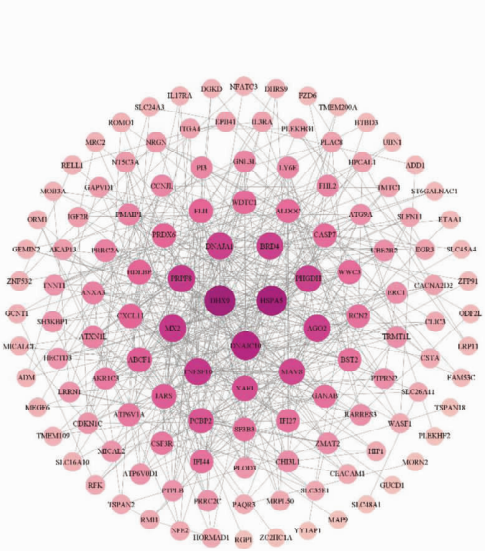
注：A 为数据集 GSE185691 聚类热图和火山图；B 为数据集 GSE84844 聚类热图和火山图，聚类热图类型中绿色代表对照组，橙色代表实验组，红色代表上调，蓝色代表下调；火山图中黑点代表无差异基因，红点代表上调基因，绿点代表下调基因；C 为 up－DEGs－Venn 图；D 为 down－DEGs－Venn 图。

图 2 两数据集中差异基因表达情况

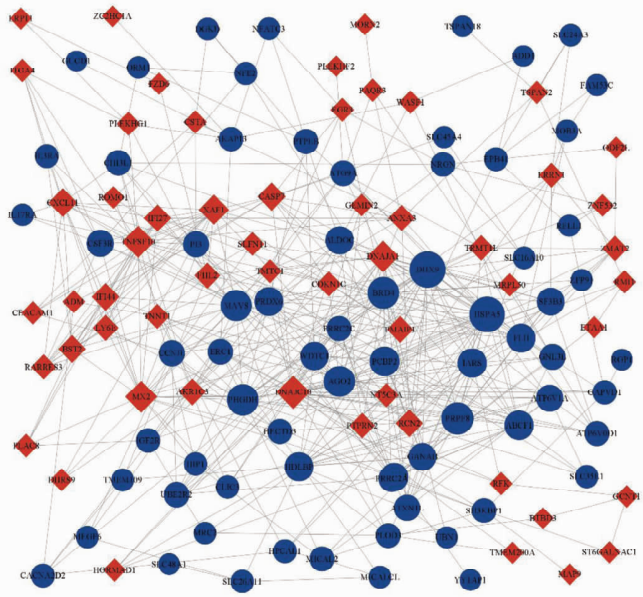


注：A 为 136 个基因富集分析中前 10 项生物学过程（BP）条目；B 为 136 个基因富集分析中前 10 项细胞组分（CC）条目；C 为 136 个基因富集分析中前 10 项分子功能（MF）条目；D 为 136 个基因富集分析中前 6 项 BP、CC、MF 条目的圆圈图；E 为 KEGG 通路条目的气泡图；F 为 KEGG 条目和基因之间的关系弦图。

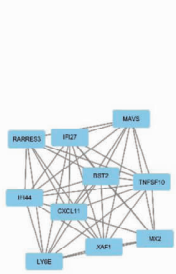
图 3 训练数据集 136 个共同 DEGs 的 GO 分析和 KEGG 通路功能富集分析



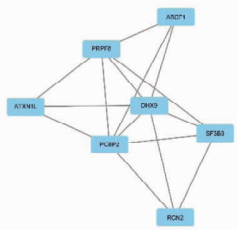
A



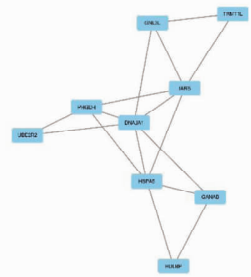
B



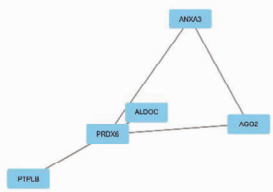
MCODE1



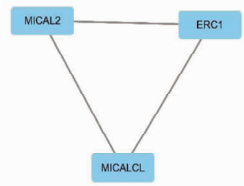
MCODE2



MCODE3

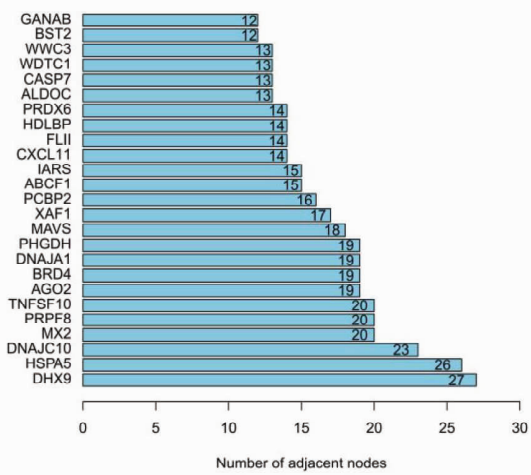


MCODE4

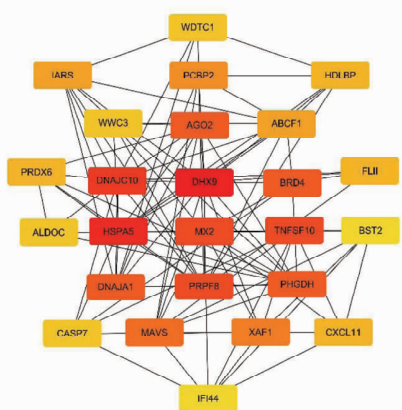


MCODE5

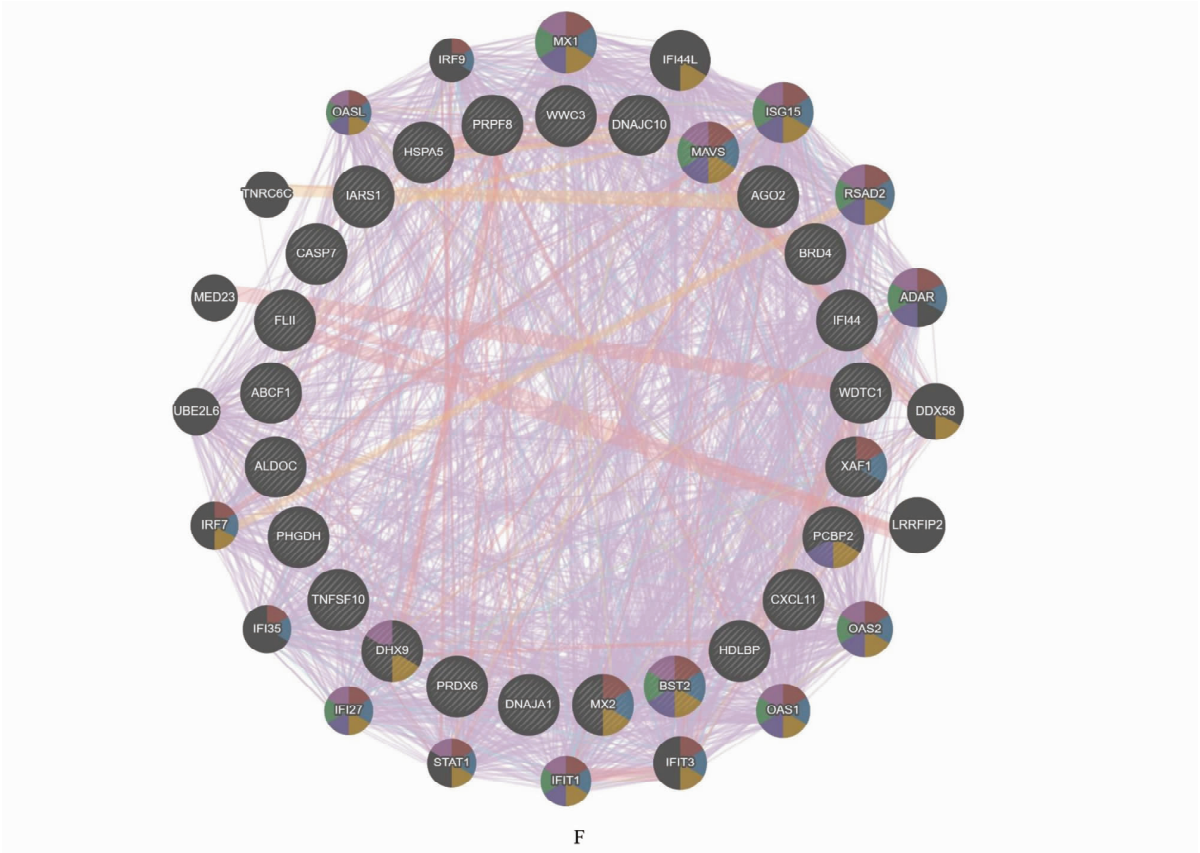
C



D

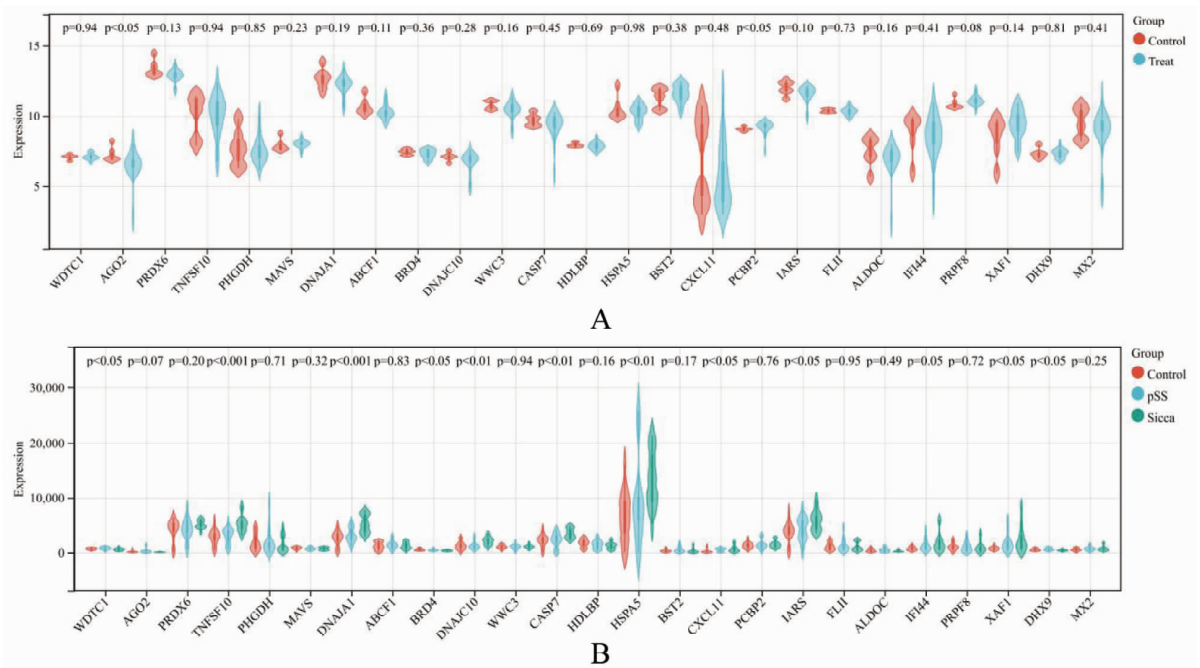


E



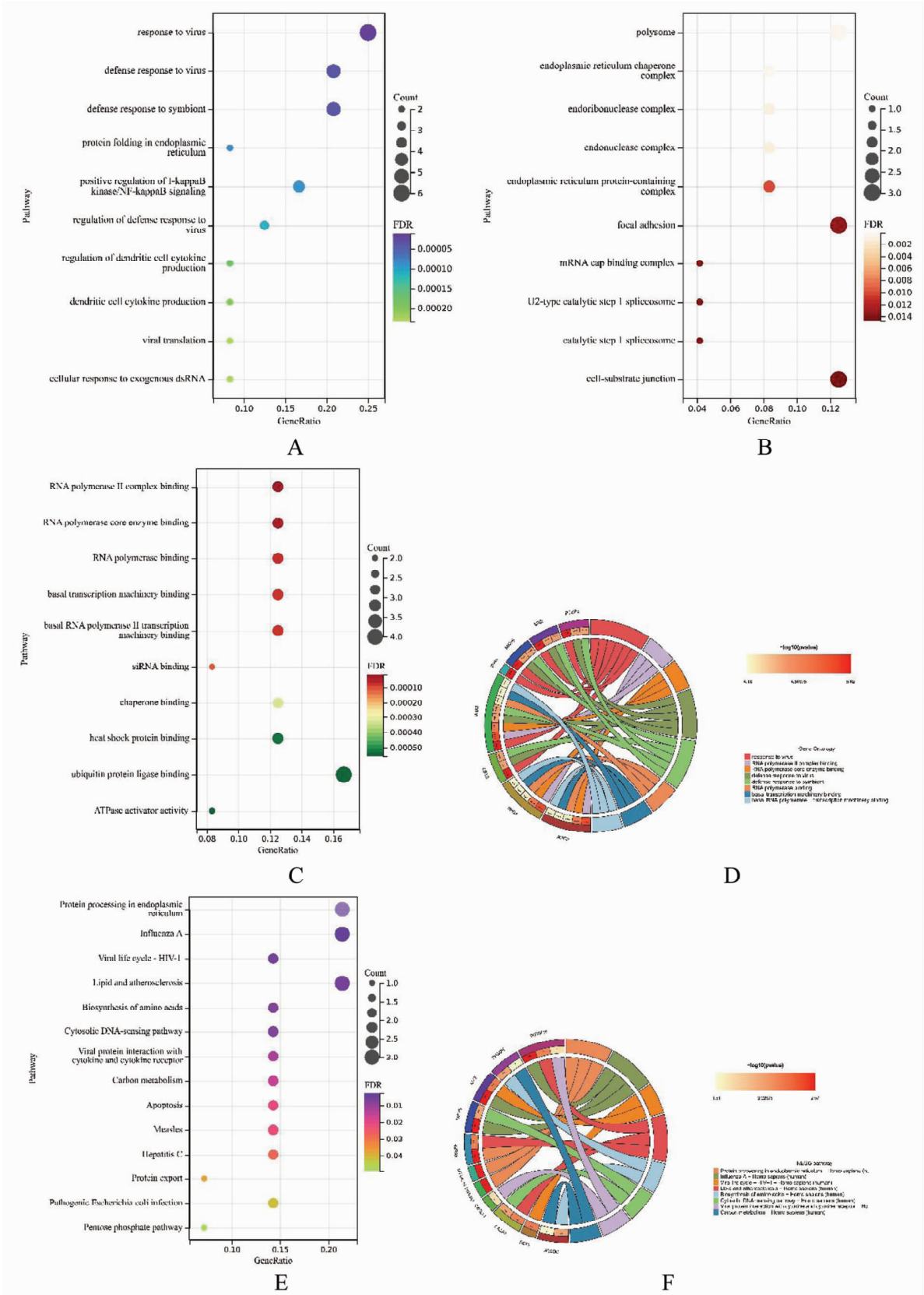
注：A 为训练数据集共同 DEGs 的 PPI 网络；B 为 PPI 网络中上调和下调基因，其中 71 个基因下调（蓝色）和 56 个基因上调（红色）；C 中 MCODE1（Score = 9.333，10 个节点，42 条边，*MX2* 为该模块核心基因），MCODE2（Score = 5，7 个节点，15 条边，*DHX9* 为该模块核心基因），MCODE3（Score = 4，9 个节点，16 条边，*DNAJA1* 为该模块核心基因），MCODE4（Score = 3，5 个节点，6 条边，*AGO2* 为该模块核心基因），MCODE5（Score = 3，3 个节点，3 条边，*ERC1* 为该模块核心基因）；D 为 Hub 基因前 25 名 degree 值排序；图 E 为 TOP25Hub 基因图；F 为 GeneMANIA 网站绘制 Hub 基因相互作用图。

图 4　训练数据集 136 个共同 DEGs 的 PPI 网络的构建与子网络模块图



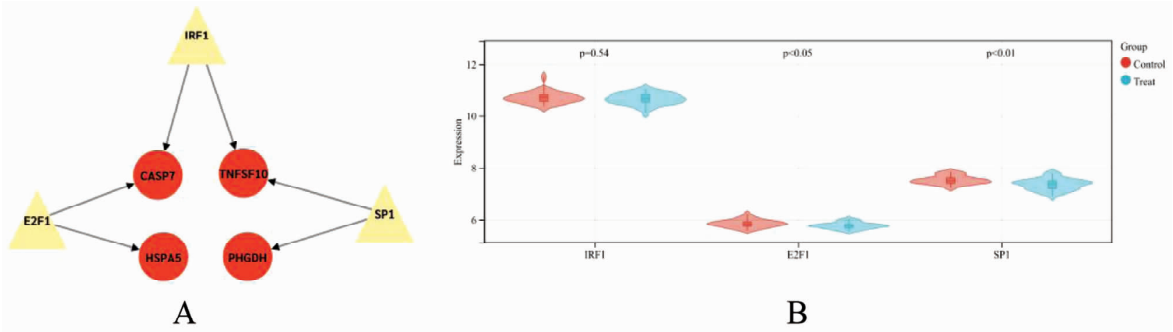
注：A 为 GSE21369 中 Hub 基因表达水平的小提琴图；B 为 GSE40611 中 Hub 基因表达水平的小提琴图。

图 6　Hub 基因在 GSE21369 和 GSE40611 组织样本中的表达差异



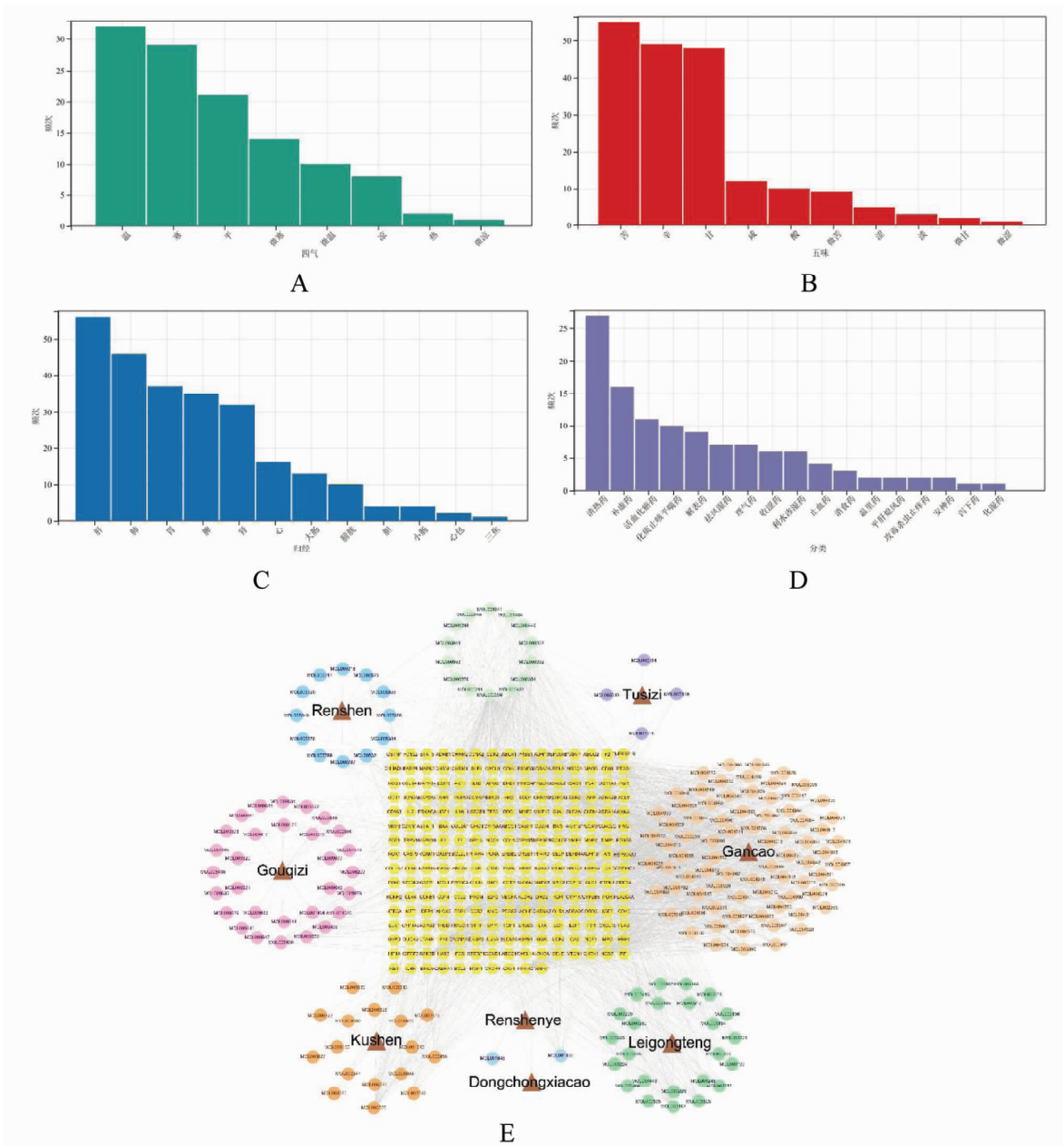
注：A 为 136 个基因富集分析中前 10 项生物学过程（BP）条目；B 为 136 个基因富集分析中前 10 项细胞组分（CC）条目；C 为 136 个基因富集分析中前 10 项分子功能（MF）条目；D 为富集最显著的前 8 个 GO 和基因之间的关系弦图；E 为 KEGG 通路条目的气泡图；F 为富集最显著的前 8 个通路和基因之间的关系弦图。

图 5 Hub 基因功能富集分析



注：A 为 Hub 基因 – TF 调控网络图；B 为 TF 差异分析小提琴图。

图 7 转录因子调控网络和差异分析



注：A 为 Hub 基因预测相关中药四气；B 为 Hub 基因预测相关中药五味；C 为 Hub 基因预测相关的中药归经；D 为 Hub 基因预测相关的中药功效；E 为中药 – 活性成分 – 作用靶点网络。

图 8 Hub 基因中药预测结果