

类风湿关节炎免疫特征基因及其与土茯苓黄酮类化合物的关系研究

杨欣¹, 黄聪¹, 姚血明²

摘要 目的 基于 Gene Expression Omnibus (GEO) 数据集结合 LASSO 逻辑回归、ssGSEA 和 WGCNA 等生物信息学及统计学方法, 筛选与特定免疫细胞浸润相关的类风湿关节炎 (RA) 诊断标志物, 并分析土茯苓黄酮类化合物与其结合情况。方法 通过 GEO 数据库获取正常对照组和 RA 组基因芯片。利用 R 4.3.0 “WGCNA” 软件包对数据集进行整合分析, 识别其中的共表达模块信息, 筛选出与 RA 密切相关的关键模块。基于 R 软件中 “glmnet” 包对差异表达基因进行 LASSO 回归分析, 筛选 RA 的特征基因。受试者工作特征 (ROC) 曲线下的面积用来评估特征基因对 RA 的诊断价值。通过 R 中的 “GSVA”、“limma” 和 “GSEABase” 程序包对正常对照组和 RA 组的基因表达数据进行量化免疫细胞浸润分析。基于超高效液相色谱-四极杆-静电场轨道阱高分辨质谱 (UHPLC-Q-Exactive/MS) 技术鉴定土茯苓黄酮类化合物。通过分子对接技术分析黄酮类化合物与特征基因的结合情况。结果 LASSO 回归算法共筛选出 5 个特征基因 (载脂蛋白 D, 含锌指和 BTB 域 16, 趋化因子 C-C 亚族受体 5, 基质金属蛋白酶-1 和冠蛋白 1A), 5 个特征基因的 ROC 曲线下面积均大于 0.85, 特征基因与多种免疫细胞呈正相关。通过 UHPLC-Q-Exactive/MS 鉴定出土茯苓黄酮类化合物共 20 个, 其中 Mulberrin 和 Neobavaisoflavone 化合物与 5 个免疫特征基因均有较好的结合。结论 土茯苓黄酮类化合物与 RA 免疫特征基因有较好的结合, 能够为 RA 免疫调节治疗和早期诊断提供科学依据。

关键词 类风湿关节炎; 土茯苓; 免疫细胞浸润分析; 特征基因; 分子对接; 黄酮类化合物

中图分类号 R 285

文献标志码 A **文章编号** 1000-1492(2024)03-0484-07

doi:10.19405/j.cnki.issn1000-1492.2024.03.018

2024-02-23 接收

基金项目: 贵州中医药大学国家与省级科技创新人才团队培育项目 (编号: 贵中医 TD 合字 [2022]004 号); 贵州省科技计划项目 (编号: 黔科合基础-ZK[2021]一般 537)

作者单位: ¹ 贵州中医药大学基础医学院, 贵阳 550025

² 贵州中医药大学第二附属医院风湿免疫科, 贵阳 550025

作者简介: 杨欣, 女, 博士, 副教授;

姚血明, 男, 博士, 教授, 责任作者, E-mail: yxming19@fox-mail.com

类风湿关节炎 (rheumatoid arthritis, RA) 是一种慢性的、高度致残的自身免疫性疾病。RA 患者关节滑膜液中存在大量的炎性细胞浸润, 如巨噬细胞、树突状细胞、T 细胞等^[1]。土茯苓是百合科植物光叶菝葜的干燥根茎, 其在抗炎、免疫系统和肿瘤等方面具有明显的药理活性。已有研究^[2]表明土茯苓具有免疫调节作用, 其主要成分落新妇苷具有抗痛风性关节炎的作用。目前, 生物信息学技术与高通量测序快速发展, 能快速获取疾病相关的生物分子信息, 并对相关的生物分子进行功能研究。同时, 基因共表达网络分析 (weighted correlation network analysis, WGCNA) 在寻找 RA 潜在生物标志物、疾病治疗特征基因等中发挥重要作用^[3]。该研究利用了 GEO 数据集结合最小绝对收缩和选择算法 (least absolute shrinkage and selection operator, LASSO)、单样本基因集富集分析 (single sample gene set enrichment analysis, ssGSEA) 和 WGCNA 等筛选与特定免疫相关的 RA 诊断标志物。并通过超高效液相色谱-四极杆-静电场轨道阱高分辨质谱 (UHPLC-Q-Exactive Orbitrap MS) 技术鉴定土茯苓黄酮类化合物, 分子对接筛选土茯苓黄酮类化合物与靶点的结合情况, 为本课题组后面的实验研究以及未来相关的科学研究提供依据。

1 材料与方法

1.1 材料

1.1.1 药材 2022 年 4 月土茯苓药材采自贵州省安顺市, 由贵州中医药大学魏升华教授鉴定为百合科植物光叶菝葜 (*Smilax glabra* Roxb.) 的干燥根茎, 植物标本存放在贵州中医药大学栋垣楼 417 (标本号: TFL2022.04)。

1.1.2 主要试剂 甲醇、乙腈和甲酸 (货号: 34860-1L-R、34851-1L、F0507-1L, 美国 sigma 公司), L-2-氯苯丙氨酸 (货号: C2001, 上海恒柏生物科技有限公司)。

1.1.3 主要仪器 离心机 (型号: Heraeus Fresco 17)、高分辨质谱仪 (型号: Q Exactive Focus) 和超高

效液相色谱(型号:Vanquish)均购自赛默飞世尔科技中国有限公司,电子天平(型号:BSA124S-CW,德国赛多利斯科学仪器有限公司),研磨仪(型号:JXFSTPRP-24,上海净信科技有限公司),超声仪(型号:YM-080S,深圳市方奥微电子有限公司),色谱柱(型号:ACQUITY UPLC BEH C18,沃特世科技上海有限公司)。

1.2 方法

1.2.1 基因芯片处理 借助美国国立生物技术信息中心(Gene Expression Omnibus(GEO)数据库检索“rheumatoid arthritis”样本。将R 4.3.0软件的“surrogate variable analysis(SVA)”软件包用于消除批间差异后使用二维主成分分析(principal component analysis, PCA)群集图显示样本间校正的效果^[4]。正常对照组和RA组差异基因识别采用(Linear Models for Microarray Data, Limma)程序包,差异基因筛选条件为 $|\log_2 \text{Fold change}| > 1$ 及校正 $P < 0.05$ 被认为差异有统计学意义。

1.2.2 加权共表达网络构建和模块识别 WGCNA分析基因网络服从无尺度分布的假设,根据基因的表达相似性将其划分为不同的模块,并与表型进行关联分析,最后识别感兴趣的基因集。基于R语言“WGCNA”软件包构建了基于差异基因表达情况的加权共表达网络^[5]。分别定义了基因显著性(gene significance, GS)及基因与模块的相关性(module membership, MM)。GS和MM的绝对值越大,代表基因与其所属模块以及临床表型的相关性越大。

1.2.3 筛选特征基因 基于R语言软件中“glmnet”包对差异表达基因进行LASSO回归分析,筛选RA的特征基因。lambda增加,自由度和残差减少,因此选取最小lambda值。然后通过glmnet函数进行交叉检验,选择均方误差最小时的lambda值并输出图形。将lambda带入模型,筛选最终输出的变量作为RA的特征基因^[6]。通过R语言软件在数据集中绘制受试者工作特征曲线(receiver operator characteristic curve, ROC)曲线,为了进一步验证这5个特征基因,计算ROC曲线下的面积(area under curve, AUC)评估特征基因用于RA诊断的价值,且使用标准正态分布计算了AUC的95%置信区间(confidence interval, CI), $AUC > 0.80$ 且 $P < 0.05$ 的基因作为后续分析的重点。

1.2.4 特征基因与免疫细胞的相关性分析 在TISIDB(<http://cis.hku.hk/TISIDB/downl>)下载免疫细胞的数据集,包括28种免疫细胞,分为适应性

免疫细胞类和先天性免疫细胞类。通过R软件相关程序输入免疫细胞浸润文件、4个GEO数据集合并文件,绘制免疫特征基因与TISIDB数据库中免疫细胞的可视化图。

1.2.5 UHPLC-Q-Exactive Orbitrap MS 鉴定分析土茯苓黄酮类化合物 按照文献中的方法进行操作^[7],超高效液相色谱条件:UPLC BEH C18色谱柱(100 mm × 2.1 mm, 1.7 μm)。进样体积为5 μl。流速单位为μl/min, A和B相中均加入0.1%甲酸。洗脱程序如下:0 min, 5% B相;3.5 min, 15% B相;6 min, 30% B相;12 min, 70% B相;18 min, 100% B相;26 min, 5% B相;30 min, 5% B相,流速为0.3 ml/min,进样量为1 μl,柱温为40℃。质谱条件:采用电喷雾离子源(ESI),正离子与负离子两种扫描模式,扫描模式为全扫描/数据依赖的二级扫描(full/ddMS2)。质量扫描范围m/z为100~1200,毛细管温度350℃,鞘气流速30 arb,辅助气流速10 arb, MS2采用低、中、高3种碰撞能。

1.2.6 分子对接分析 采用分子对接软件(SYBYL 2.1.1)中的Surflex-Docking模块分析土茯苓黄酮类化合物与5个特征基因的结合情况,结果以总分值(total_score, T_Score) > 5为阈值进行分析^[8]。从蛋白质晶体结构数据库RCSB(<http://www.rcsb.org/pdb>)中获取载脂蛋白D(apolipoprotein D, AP-OD, PDB ID:2HZQ),含锌指和BTB域16(zinc finger and BTB domain containing 16, ZBTB16, PDB ID:1BUO),趋化因子C-C亚族受体5(C-C chemokine receptor type 5, CCR5, PDB ID:7F1R),基质金属蛋白酶-1(matrix metalloproteinase 1, MMP1, PDB ID:3SHI)和冠蛋白1A(coronin-1A, CORO1A, PDB ID:2AQF)晶体结构。

1.3 统计学处理 采用R 4.3.0软件对数据进行统计分析。连续变量以 $\bar{x} \pm s$ 表示,用glmnet包进行Lasso回归分析,选择变量和调整模型复杂度,从而避免过度拟合;使用ROC曲线及AUC评价特征基因诊断RA的价值, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 GEO数据集处理及差异基因分析 4个数据集合并作为训练数据集(表1),正常对照组32个样本,RA组44个样本。以PCA聚类图的形式展示消除批次之间的差异前和后,PCA结果显示,图1A为矫正前的GEO数据集的分布情况,图1B为矫正后的GEO数据集的分布情况,在数据矫正前4个数据

集有明显差异,而批次矫正后基本上消除了4个数据集的批处理效应。

表1 RA 基因芯片基本信息

数据集	平台	正常对照组(n)	RA组(n)	实验类型
GSE1919	GPL	5	5	Array
GSE55235	GPL	10	10	Array
GSE55457	GPL	10	13	Array
GSE77298	GPL	7	16	Array

2.2 关键模块与 RA 的相关性 借助 R 软件的“WGCNA”软件包对 4 个 GEO 数据集的表达谱数据构建加权共表达网络,通过 WGCNA 分析将差异表达的基因构建 4 个不同的模块。图 2A 显示了蓝色模块、蓝绿色模块、棕色模块、黄色模块与 RA 的相关性,其中蓝色模块和蓝绿色模块与 RA 的相关性较高($r = 0.75, 0.73$),且蓝色模块和蓝绿色模块中基因与 RA 表型呈正相关($r = 0.87, 0.91, P < 0.05$)(图 2B、C),由于蓝色模块中基因与 RA 表型

相关性低于蓝绿色模块,说明这些基因在蓝绿色模块中具有重要角色,因此本研究选择蓝绿色模块中的基因进行深入研究。

2.3 类风湿关节炎的特征基因 将蓝绿色模块中连接性最高的 69 个基因作为候选中心基因。4 个数据集合并处理后鉴定差异表达基因共 63 个,图 3A 显示有 9 个交集基因。图 3B 为 LASSO 回归的 10 折交叉验证图,显示基于 LASSO 回归算法对 9 个基因进行筛选,图中的每一条曲线代表了每个自变量系数的变化轨迹,纵坐标是系数的值,下横坐标是 L1 范数,上横坐标是此时模型中非零系数的个数。参数系数绝对值随着 L1 范数值变小而变大。当 L1 范数达到一定值以后,一部分不重要的变量(特征基因)将被压缩为 0,代表该变量已被剔除。图 3C 为最佳调谐参数选择图,得到交叉验证曲线和最小化平均交叉验证误差的 lambda 的值。共筛选出 5 个对 RA 疾病具有诊断价值的特征基因,分别为 APOD、ZBTB16、CCR5、MMP1 和 CORO1A。

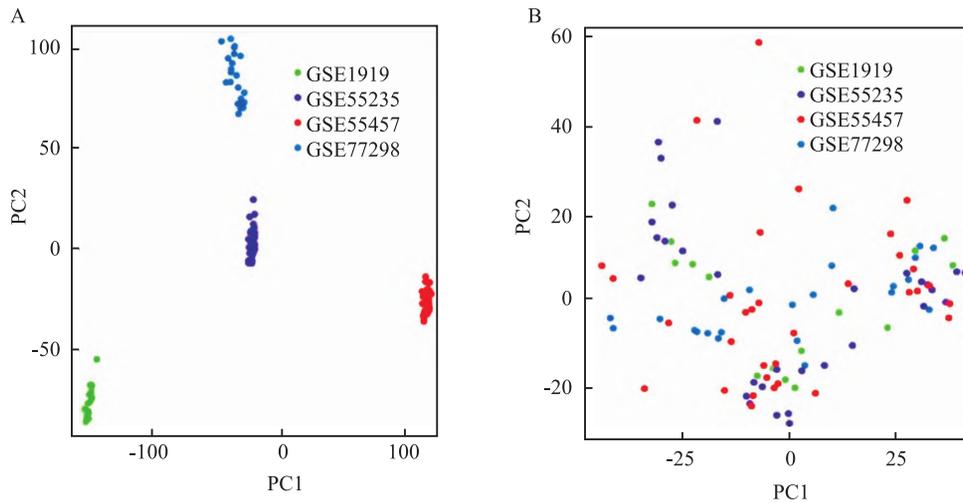


图1 批次矫正前后 PCA 结果
A:批次矫正前 PCA 结果;B:批次矫正后 PCA 结果

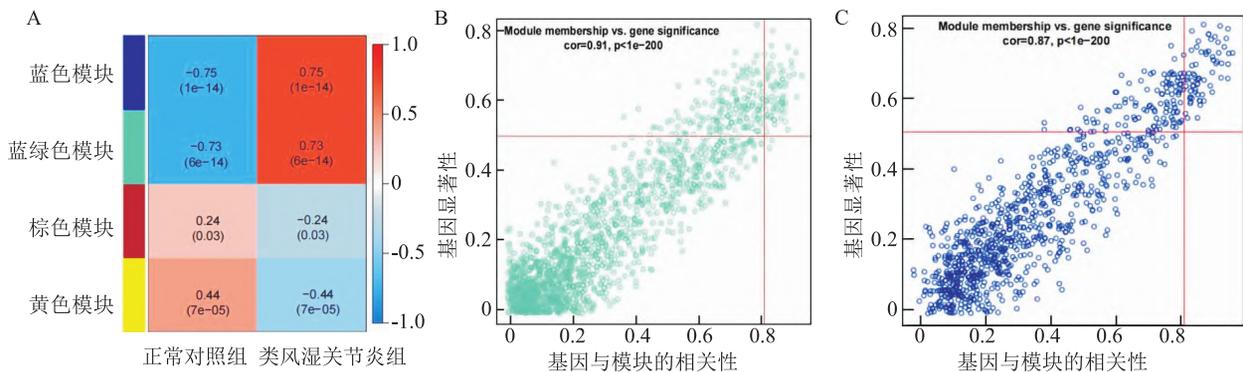


图2 基因模块与类风湿关节炎的相关性分析图

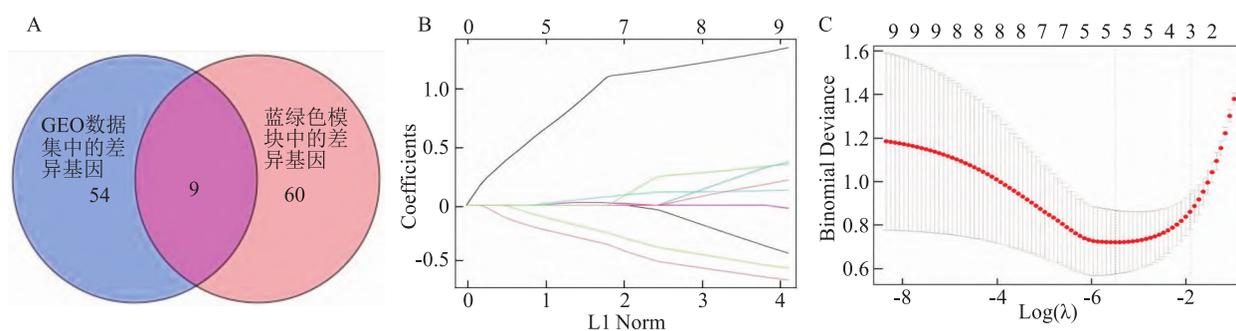


图3 类风湿关节炎特征生物标志物的鉴定和 LASSO 回归图

A:交集基因韦恩图; B:LASSO 的交叉验证图; C:最佳调谐参数选择图

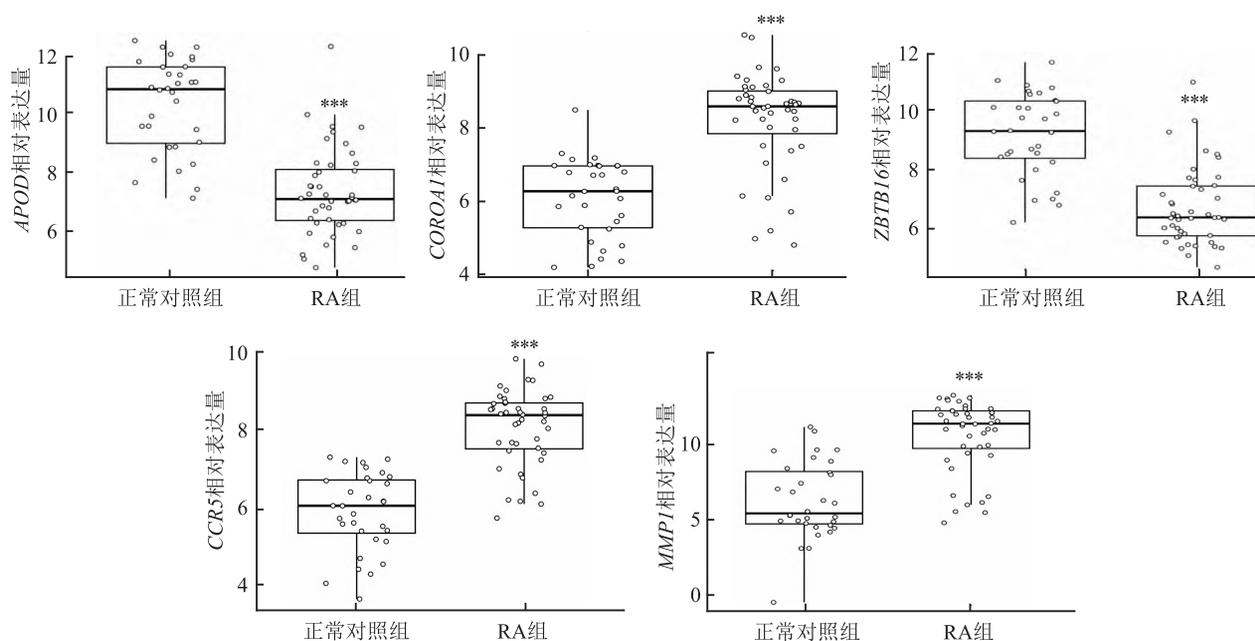


图4 正常对照组和 RA 组特征基因表达的比较

与正常对照组比较: *** $P < 0.001$

2.4 特征基因的表达情况 基因 *CCR5*、*APOD*、*ZBTB16*、*MMP1* 和 *CORO1A* 在数据集中均存在差异表达(图4)。与正常对照组相比,RA组 *CCR5*、*MMP1* 和 *CORO1A* 表达增高 ($P < 0.001$)。 *APOD* 和 *ZBTB16* 表达降低 ($P < 0.001$)。

2.5 特征基因的 ROC 曲线分析结果 从标准化表达矩阵中获取 *CCR5*、*APOD*、*ZBTB16*、*MMP1* 和 *CORO1A* 基因的表达情况,结果显示 5 个特征基因的 AUC 均大于 0.85(图5)。

2.6 特征基因与免疫细胞浸润之间的关系 通过 R 语言绘制 5 个特征基因与免疫细胞的相关性热图,纵坐标代表免疫细胞,横坐标代表 5 个特征基因,模块中颜色代表 5 个特征基因与免疫细胞间的相关性,模块中颜色深浅代表该免疫特征基因与免疫细胞间的相关性的强弱,蓝色代表负相关,红色代

表正相关(图6)。发现 *CCR5*、*MMP1* 和 *CORO1A* 与多种免疫细胞呈正相关,其中 *CCR5*、*MMP1* 和 *CORO1A* 均与伽马三角洲 T 细胞($\gamma\delta$ T 细胞)呈正相关(均 $P < 0.05$)。 *CCR5* 和 *CORO1A* 与巨噬细胞、调节性 T 细胞、单核细胞、骨髓源性抑制细胞呈正相关(均 $P < 0.05$)。而 *APOD* 基因与多种免疫细胞呈负相关($P < 0.05$)。

2.7 土茯苓黄酮类化合物的基本信息 检测数据经峰匹配与校准后共得到 133 个化合物离子用于后续分析。通过与标准品、保留时间、精确分子量和二级质谱比对共结构鉴定出 20 个化合物(表2)。

2.8 土茯苓黄酮类化合物与特征基因的结合情况 通过分子对接发现 14 个化合物与 *MMP1* 有较好的结合($T_Score > 5$);16 个化合物与 *CCR5* 有较好的结合($T_Score > 5$);17 个化合物与 *APOD* 有较好

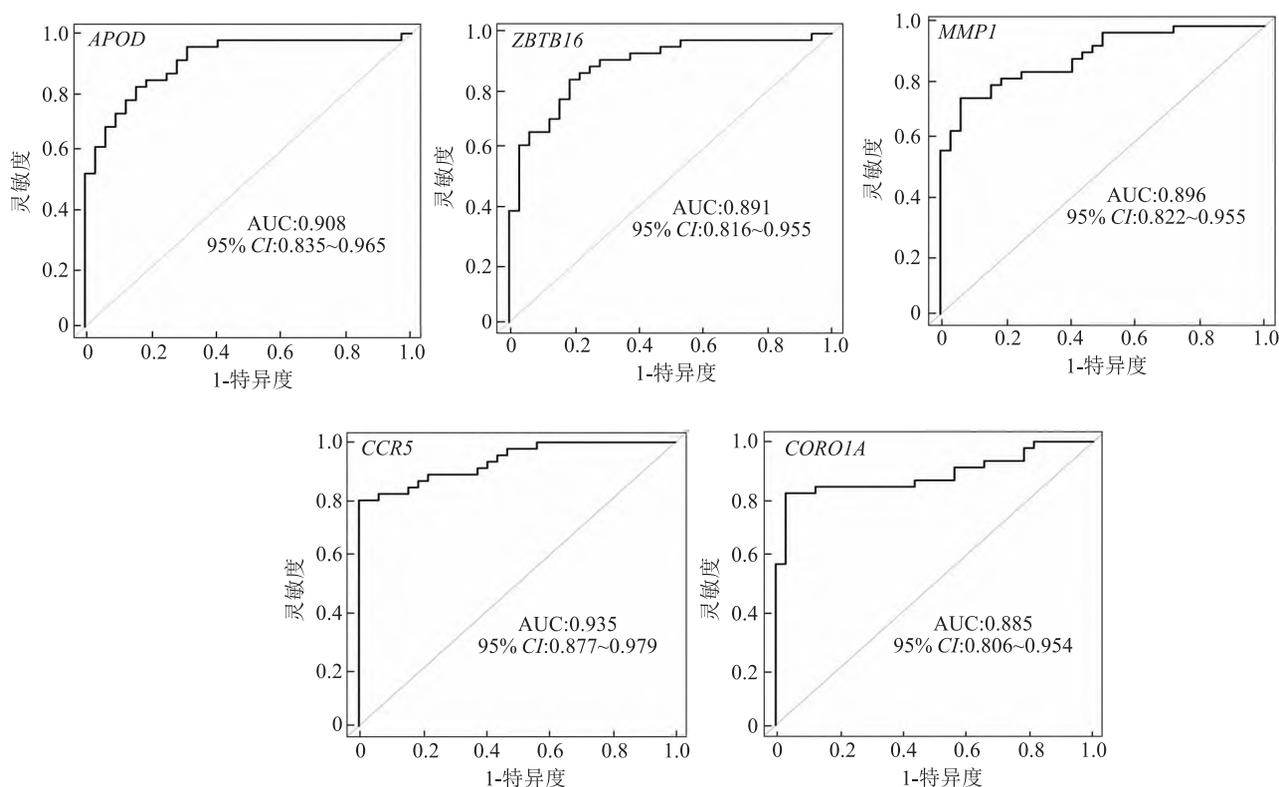


图5 特征基因的 ROC 曲线

表2 鉴定出的土茯苓黄酮类化合物

编号	出峰时间	化合物名称	分子量	离子形式	质荷比
1	0.79	Trilobatin	C ₂₁ H ₂₄ O ₁₀	[M-H] ⁻	435.13
2	2.36	(+)-Catechin hydrate	C ₁₅ H ₁₆ O ₇	[M-H] ⁻	289.07
3	3.27	kaempferol 7-neohesperidoside	C ₂₇ H ₃₀ O ₁₅	[M-H] ⁻	593.15
4	4.02	Mulberrin	C ₂₅ H ₂₆ O ₆	[M-H] ⁻	421.16
5	4.05	Phloridzin	C ₂₁ H ₂₄ O ₁₀	[M-H] ⁻	435.13
6	4.79	Hematoxylin	C ₁₆ H ₁₄ O ₆	[M-H] ⁻	301.07
7	5.09	Phloretin	C ₁₅ H ₁₄ O ₅	[M-H] ⁻	273.08
8	5.12	Kaempferitrin	C ₂₇ H ₃₀ O ₁₄	[M-H] ⁻	755.20
9	5.74	Taxifolin	C ₁₅ H ₁₂ O ₇	[M-H] ⁻	303.05
10	5.93	Luteolin 7-glucuronide	C ₂₁ H ₁₈ O ₁₂	[M-H] ⁻	461.07
11	6.61	Kaempferol	C ₁₅ H ₁₀ O ₆	[M-H] ⁻	285.04
12	6.65	hesperidin	C ₂₈ H ₃₄ O ₁₅	[M-H] ⁻	609.18
13	7.06	Apigenin 7-O-glucuronide	C ₂₁ H ₁₈ O ₁₁	[M-H] ⁻	445.08
14	8.27	Isosakuranin	C ₂₂ H ₂₄ O ₁₀	[M-H] ⁻	447.13
15	8.56	Scutellarein	C ₁₅ H ₁₀ O ₆	[M-H] ⁻	285.04
16	9.42	Formononetin	C ₁₆ H ₁₂ O ₄	[M-H] ⁻	267.07
17	9.60	Tectorigenin	C ₁₆ H ₁₂ O ₆	[M-H] ⁻	299.06
18	12.14	Neobavaisoflavone	C ₂₀ H ₁₈ O ₄	[M-H] ⁻	321.11
19	13.59	Licochalcone-C	C ₂₁ H ₂₂ O ₄	[M-H] ⁻	337.14
20	16.47	Isoxanthohumol	C ₂₁ H ₂₂ O ₅	[M-H] ⁻	353.14

的结合(T_Score > 5);8 个化合物与 CORO1A 有较好的结合(T_Score > 5);3 个化合物与 ZBTB16 有较好的结合(T_Score > 5);同时发现, Mulberrin 和 Neobavaisoflavone 化合物能同时与 5 个特征基因结合(表 3),有进一步研究的价值。

3 讨论

RA 是一种常见的自身免疫性疾病,能反复和持续激活先天性和获得性免疫系统,导致免疫耐受失败、大量炎性细胞因子产生、致残性软骨和骨骼损

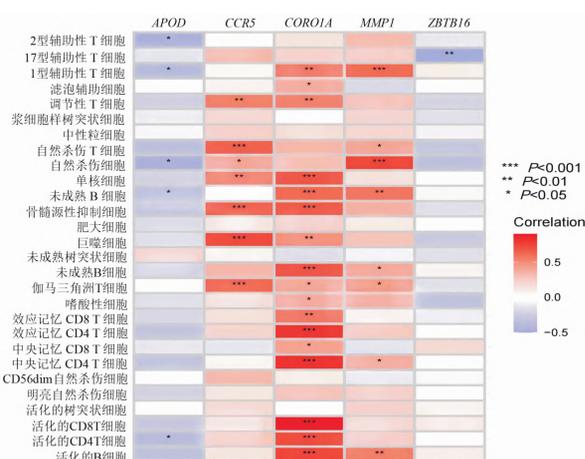


图6 特征基因与免疫细胞的相关性热图

表3 土茯苓黄酮类化合物与特征基因结合总分

化合物名称	CCR5	APOD	ZBTB16	MMP1	CORO1A
Trilobatin	6.6	6.5	4.3	8.1	5.7
(+)-Catechin hydrate	5.8	5.0	3.3	3.6	5.1
kaempferol 7-neohesperidoside	7.3	7.7	3.3	6.0	4.6
Mulberrin	6.3	8.4	5.4	7.6	5.1
Phloridzin	5.6	7.3	4.4	7.3	4.5
Hematoxylin	4.6	3.6	2.6	4.3	4.0
Phloretin	6.2	6.4	4.0	5.7	5.5
Kaempferitrin	5.4	5.8	2.4	5.7	5.0
Taxifolin	5.8	6.0	3.1	5.6	4.4
Luteolin 7-glucuronide	6.1	8.0	4.3	6.2	2.1
Kaempferol	4.0	4.3	2.0	4.0	4.2
hesperidin	7.1	9.9	3.9	7.5	3.6
Apigenin 7-O-glucuronide	7.7	5.8	3.8	8.2	6.5
Isosakuranin	5.8	6.8	3.3	6.3	4.2
Scutellarein	3.9	5.0	1.8	4.1	4.4
Formononetin	4.2	7.3	2.0	3.9	3.4
Tectorigenin	5.5	4.2	2.3	5.2	4.1
Neobavaisoflavone	5.4	6.7	5.1	6.5	5.4
Licochalcone-C	6.1	7.3	3.5	6.0	4.1
Isoxanthohumol	5.5	6.5	5.2	4.3	5.0

伤,也会出现肺部和心脏的全身性炎症反应。本研究通过生物信息学方法,对 GEO 数据库中 RA 相关数据深度挖掘,发现 5 个 RA 特征基因(*CCR5*, *APOD*, *ZBTB16*, *MMP1*, *CORO1A*),且 5 个 RA 特征基因与多种免疫细胞相关。RA 的发病过程中趋化因子通过调节免疫和炎症反应发挥介质作用。*CCR5* 是一种 G 蛋白偶联受体。已有研究^[9]发现 *CCR5* 参与了包括 RA、骨质疏松症在内的多种疾病的发病。滑膜 T 细胞中的 *CCR5* 水平升高,并与 RA 的发生发展有关,*CCR5* 在外周血单核细胞和 T 细胞中的表达水平低,但在体内外经炎症刺激后表达明显增强;*APOD* 是多功能蛋白质,涉及炎症、抗氧化等。在多种组学中 *APOD* 表达影响糖尿病的发生和发展^[10]。但其在 RA 的表达情况尚未明确,值得进一

步深入研究;在基质金属蛋白酶(matrix metalloproteinases, MMPs)家族中,*MMP1* 属于胶原酶和间质溶解素类,并且有研究^[11]证明 *MMP1* 的水平与 RA 疾病活动度相关,并且可以预测 RA 的功能和影像学结果,是预测 RA 早期骨破坏的指标。*CORO1A* 基因参与自身免疫病的发生发展,其是特异性调节因子,参与维持外周初始 T 细胞稳态^[12]。以上提示这些特征基因在 RA 的发生和发展中发挥重要作用。

RA 发病机制复杂,多种免疫细胞参与 RA 的发病及进展过程,通过探索 RA 特征基因与免疫细胞之间的关系,可以更好地了解 RA 的免疫学机制。本研究筛选特征基因与免疫细胞的相关性发现 *CCR5*、*MMP1* 和 *CORO1A* 与多种免疫细胞呈正相关,其中 *CCR5*、*MMP1* 和 *CORO1A* 均与 $\gamma\delta$ T 细胞呈正相关。*CCR5* 和 *CORO1A* 与巨噬细胞、调节性 T 细胞、单核细胞呈正相关。单核细胞和巨噬细胞在 RA 的发病机制中起着核心作用,其能通过产生不同种类的趋化因子、细胞因子和生长因子来调节免疫和炎症反应,导致骨和软骨的破坏,诱导 RA 的发生^[13];调节性 T 细胞(Treg)在多种疾病中发挥作用。是辅助性 T 细胞的特殊亚群之一,Treg 能有效防止自身免疫性疾病的发展。RA 患者的 Treg 细胞以功能受损和表型改变为特征^[14]; $\gamma\delta$ T 细胞是免疫系统的一类特殊细胞,已有研究^[15]显示 $\gamma\delta$ T 细胞在 RA 发病中具有启动和保护双重作用。RA 作为自身免疫性疾病,其自身反应性 CD4⁺ T 细胞、巨噬细胞、炎症细胞因子、趋化因子等水平异常升高。而本研究发现特征基因与免疫细胞的相关性,对未来相关的科学研究提供理论依据。

综上所述,基于差异表达数据利用 LASSO 回归筛选出 5 个 RA 特征基因,在 ROC 曲线分析中, *CCR5*, *APOD*, *ZBTB16*, *MMP1*, *CORO1A* 的 AUC 值均大于 0.85,这 5 个 RA 特征基因作为 RA 诊断标志物具有潜在价值。此外,本研究选发现土茯苓黄酮类化合物与 RA 诊断标志物均有结合作用,其治疗机制可能与免疫相关,由于本研究尚未进行体内实验研究,具体机制有待进一步体内实验验证。

参考文献

[1] Jin S, Li M, Fang Y, et al. Chinese registry of rheumatoid arthritis (CREDIT): II. prevalence and risk factors of major comorbidities in Chinese patients with rheumatoid arthritis[J]. Arthritis Res Ther, 2017, 19(1):251.
 [2] 陈小鹏. 土茯苓落新妇苷基于 NF- κ B 信号通路抑制大鼠类风湿性关节炎的研究[D]. 合肥:安徽中医药大学, 2022.

- [3] Liu K, Chen S, Lu R. Identification of important genes related to ferroptosis and hypoxia in acute myocardial infarction based on WGCNA[J]. *Bioengineered*, 2021, 12(1):7950-63.
- [4] 张贵斌, 张闫斌, 许 涵, 等. 基于 GEO 数据库筛选多发性硬化症关键基因 *Kcnc1*、*Kcnc2* 的分析研究[J]. *中国免疫学杂志*, 2023, 39(8):1600-4.
- [5] 李开晃. WGCNA 筛选炎症相关基因 *PTTG1* 在骨性关节炎中作用与机制研究[D]. 南昌:南昌大学, 2023.
- [6] 杜紫璇. 基于 spike-and-slab lasso 和 lasso 方法构建低级别胶质瘤放射敏感性预测模型[D]. 苏州:苏州大学, 2022.
- [7] 宋书祎, 兰先明, 徐 静, 等. 附子二萜类生物碱 UHPLC-Q-Ex-active Orbitrap MS 的系统分析鉴定[J]. *药物分析杂志*, 2023, 43(6):918-29.
- [8] 袁桥玉, 刘新桥. 基于分子对接技术的红毛七抗炎化学成分虚拟筛选[J]. *化工设计通讯*, 2022, 48(10):218-21.
- [9] 王友强, 兰由玉, 李世勇, 等. *CCR5* 基因沉默对 RA 大鼠滑膜细胞炎症反应的影响[J]. *西部医学*, 2021, 33(9):1300-4, 1310.
- [10] 蒋圣娟, 丛 艺, 张士瑾, 等. 载脂蛋白 ApoD: 系统发生、基因结构与表达调控[J]. *中国海洋大学学报(自然科学版)*, 2008(3):436-46.
- [11] 代静杨. 艾灸对类风湿关节炎患者血清中 MMP1、MMP3、VEGF 含量的影响[D]. 成都:成都中医药大学, 2019.
- [12] 丁 文. Coro1a 在高钠环境下内皮细胞中表达的改变及其对内皮细胞功能的调节作用[D]. 合肥:安徽医科大学, 2022.
- [13] 王晨宇, 邱 波. 免疫细胞在类风湿关节炎发病机制中作用的研究进展[J]. *现代医学*, 2023, 51(1):121-5.
- [14] 汤小雨, 王 琛, 韩 乐, 等. CP-25 通过调控 T 细胞亚群治疗大鼠胶原性关节炎[J]. *安徽医科大学学报*, 2019, 54(7):1001-8.
- [15] 韩燕英, 王友莲. $\gamma\delta$ T 细胞在自身免疫病发病中的作用[J]. *江西医药*, 2019, 54(8):988-92.

Study on immune characteristic genes of rheumatoid arthritis and the relationship with flavonoids of *Smilax glabra* Roxb.

Yang Xin¹, Huang Cong¹, Yao Xueming²

¹*School of Basic Medicine, Guizhou University of Traditional Chinese Medicine, Guiyang 550025;*

²*Dept of Rheumatology and Immunology, The Second Affiliated Hospital of Guizhou University of Traditional Chinese Medicine, Guiyang 550025)*

Abstract Objective To use the GEO dataset and bioinformatics techniques, such as LASSO logistic regression, ssGSEA, and WGCNA, to screen for RA diagnostic markers and investigate the impact of earthly flavonoids in *Smilax glabra* Roxb. on specific immune cell infiltration, to screen for rheumatoid arthritis (RA) diagnostic markers on specific immune cell infiltration and to analyze the combination of flavonoids in *Smilax glabra* Roxb. and diagnostic markers. **Methods** The normal control group and RA gene chip were obtained from the Gene Expression Omnibus database. The R 4.3.0 WGCNA software package was used to integrate and analyze the dataset, identify co-expression modules and associated trait information, and screen key modules closely related to RA. LASSO regression analysis was performed using the glmnet package in R to identify characteristic genes for RA. The area under the receiver operating characteristic (ROC) curve was used to evaluate the diagnostic value of the characteristic genes in RA. The gene expression data of the normal control group and RA group were subjected to quantitative immune cell infiltration analysis using the GSVA, limma, and GSEABase packages in R. The chemical components of earthworm flavonoids in *Smilax glabra* Roxb. were analyzed based on UHPLC-Q-Exactive Orbitrap MS. The correlation between flavonoids and characteristic genes was assessed through molecular docking. **Results** The LASSO regression algorithm selected 5 characteristic genes (apolipoprotein D, zinc finger and BTB domain containing 16, C-C chemokine receptor type 5, matrix metalloproteinase 1, coronin-1A). The area under ROC curve of all 5 characteristic genes was greater than 0.85, which exhibited positive correlations with various immune cells. Twenty earthworm flavonoids of *Smilax glabra* Roxb. were identified using UHPLC-Q-Exactive/MS, and Mulberrin and Neobavaisoflavone were well combined with 5 immune characteristic genes. **Conclusion** Flavonoids compounds of *Smilax glabra* Roxb. have good combination with RA immune characteristic genes, providing a scientific basis for RA immunomodulation therapy and early diagnosis.

Key words rheumatoid arthritis; *Smilax glabra* Roxb.; immune cell infiltration analysis; characteristic gene; molecular docking; flavonoids compounds