

线粒体自噬视域下的类风湿关节炎：基于多机器学习算法的互作分析

李加根, 陈跃平, 黄柯琪, 陈尚桐, 黄川洪

<https://doi.org/10.12307/2025.745>

投稿日期: 2024-08-01

采用日期: 2024-09-15

修回日期: 2024-11-22

在线日期: 2024-12-03

中图分类号:

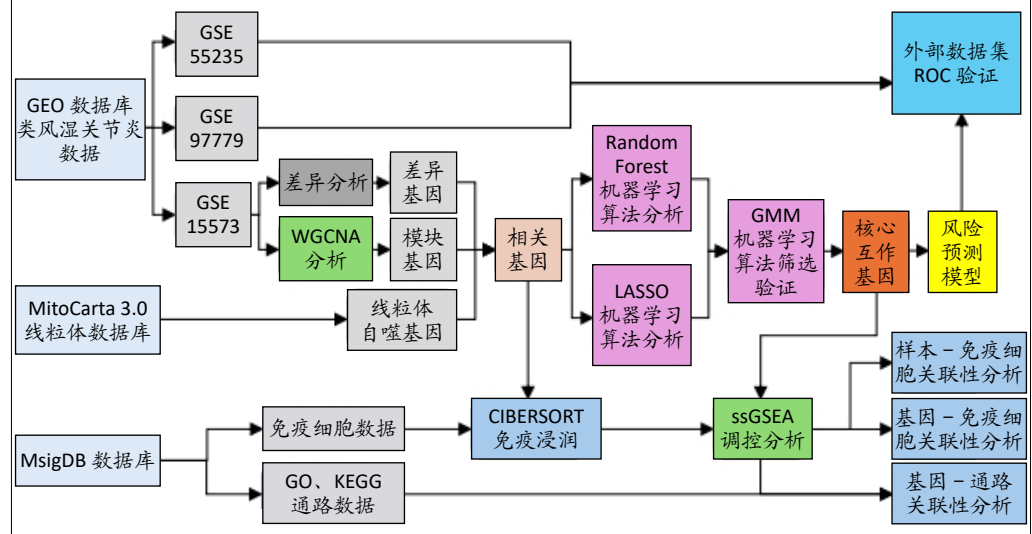
R459.9; R318; R593.22

文章编号:

2095-4344(2025)26-05595-13

文献标识码: B

文章快速阅读: 类风湿关节炎中线粒体自噬相关预测模型构建及免疫调控分析



文题释义:

类风湿关节炎: 是一种慢性、系统性的自身免疫性疾病, 主要特征为多发对称性小关节炎、疼痛和肿胀, 疾病后期往往导致严重的关节畸形甚至功能丧失。

线粒体自噬: 是一种选择性自噬, 主要过程为异常的线粒体被包裹在自噬体内随之运输到溶酶体, 从而被靶向降解和清除, 以控制线粒体质量和功能的稳定, 这一过程对维持细胞内能量平衡、减少氧化应激和维护细胞内稳态具有重要意义。

摘要

背景: 类风湿关节炎发病机制至今尚未完全清晰, 近来有研究表明线粒体自噬与类风湿关节炎存在关联性, 但关键机制有待深入探究。

目的: 利用多机器学习算法鉴定和验证类风湿关节炎中线粒体自噬过程的核心互作基因并解析其免疫调控过程。

方法: 从GEO数据库整理类风湿关节炎转录组表达数据集GSE15573作为独立训练集, GSE97779和GSE55235合集作为独立验证集。利用训练集筛选类风湿关节炎差异表达基因, 同时进行“WGCNA”分析。然后从“MitoCarta3.0”数据库下载线粒体自噬相关基因, 将其与类风湿关节炎差异基因和“WGCNA”分析模块基因取交集获得类风湿关节炎-线粒体自噬相关基因, 将相关基因进行功能富集分析以明确细胞通路。随后利用“Random Forest”和“LASSO”2种机器学习算法分别筛选特征基因, 利用“GMM”机器学习算法对前2种机器学习算法筛选的交集基因拟合验证, 以获得类风湿关节炎-线粒体自噬核心互作基因。进一步建立预测模型, 并利用外部验证集验证。最后, 采用“CIBERSORT”进行免疫浸润分析此过程中免疫细胞亚群占比及亚群之间关联性, 采用“ssGSEA”分析类风湿关节炎-线粒体自噬核心互作基因与免疫细胞亚群间关联性, 同时分析核心互作基因的相关“GO”“KEGG”生物学通路。

结果与结论: ①差异分析获得807个类风湿关节炎差异表达基因, “WGCNA”分析筛选出2个特征模块含1 208个基因, 线粒体基因数据库整理出1 136个基因, 三部分基因取交集获得53个基因类风湿关节炎-线粒体自噬相关基因; ②相关基因的功能富集分析结果显示细胞过程主要与线粒体自噬、过氧化物酶体代谢、细胞衰老、坏死性凋亡相关; ③3种机器学习算法鉴定出4个类风湿关节炎-线粒体自噬核心互作基因(DNAJA3、C12orf65、AKR7A2、PDHB); ④预测模型的风险评分受试者工作特征曲线下面积为0.989, 外部患者样本的工作特征曲线验证类风湿关节炎-线粒体自噬核心互作基因的曲线下面积均大于0.7; ⑤免疫调控分析显示类风湿关节炎中线粒体自噬过程与记忆B细胞、M0型巨噬细胞、活化的记忆性CD4 T细胞、静息态记忆性CD4 T细胞密切相关; ⑥生物学通路分析结果显示核心互作基因与821条“GO”通路强相关(|cor| > 0.8, $P < 0.001$), 与48条“KEGG”通路强相关(|cor| > 0.8, $P < 0.001$), 其中关键生物学过程与线粒体DNA代谢、线粒体DNA修复、线粒体DNA复制、线粒体基因组维护、线粒体去极化的正向调控和参与凋亡信号通路的线粒体外膜通透性的正向调控有关; ⑦上述结果证实, DNAJA3、C12orf65、AKR7A2、PDHB是类风湿关节炎中线粒体自噬过程的核心互作基因, 通过参与特定免疫过程在疾病进展中发挥关键作用, 对类风湿关节炎的诊断具有精准预测效果。

关键词: 线粒体自噬; 类风湿关节炎; 机器学习算法; 加权基因共表达网络分析; 预测模型; 外部验证; 免疫细胞; 生物学功能

广西中医药大学附属瑞康医院, 广西壮族自治区南宁市 530011

第一作者: 李加根, 男, 1998年生, 内蒙古自治区鄂尔多斯市人, 汉族, 广西中医药大学在读硕士, 主要从事脊柱、骨关节创伤性疾病的防治研究。

通讯作者: 陈跃平, 博士, 主任医师, 博士生导师, 广西中医药大学附属瑞康医院, 广西壮族自治区南宁市 530011

<https://orcid.org/0009-0006-1240-5170> (李加根); <https://orcid.org/0000-0003-3860-1568> (陈跃平)

基金资助: 国家自然科学基金项目(82360937), 项目负责人: 陈跃平; 广西自然科学基金项目(2023JJA140318), 项目负责人:

陈跃平; 广西中西医结合骨与关节退行性疾病多学科交叉创新团队(GZKJ2310), 项目参与人: 陈跃平; 广西中医药大学研究生

教育创新计划项目(YCSY2023040), 项目负责人: 李加根

引用本文: 李加根, 陈跃平, 黄柯琪, 陈尚桐, 黄川洪. 线粒体自噬视域下的类风湿关节炎: 基于多机器学习算法的互作分析 [J].

中国组织工程研究, 2025, 29(26):5595-5607.



Rheumatoid arthritis from the perspective of mitophagy: interaction analysis based on multiple machine learning algorithms

Li Jiagen, Chen Yueping, Huang Keqi, Chen Shangtong, Huang Chuanhong

Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, Guangxi Zhuang Autonomous Region, China

Li Jiagen, Master candidate, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, Guangxi Zhuang Autonomous Region, China

Corresponding author: Chen Yueping, PhD, Chief physician, Doctoral supervisor, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning 530011, Guangxi Zhuang Autonomous Region, China

Abstract

BACKGROUND: The pathogenesis of rheumatoid arthritis has not yet been fully clarified, and recent studies have shown that mitophagy is associated with rheumatoid arthritis, but the key mechanisms need to be explored in depth.

OBJECTIVE: To identify and validate the core interaction genes of mitophagy in rheumatoid arthritis using multiple machine learning algorithms and to analyze its immunoregulatory process.

METHODS: The rheumatoid arthritis transcriptome expression dataset GSE15573 was retrieved from the GEO database as an independent training set, with the GSE97779 and GSE55235 collections used as independent validation sets. The differentially expressed genes of rheumatoid arthritis were screened using the training set, and “WGCNA” analysis was also performed. Then we downloaded the mitophagy-related genes from the “MitoCarta3.0” database, and intersected them with the differentially expressed genes of rheumatoid arthritis and the genes in the “WGCNA” analysis module to obtain the rheumatoid arthritis-mitophagy-related genes, and then analyzed the related genes for functional enrichment to clarify the cellular pathways. Feature genes were initially identified using the “Random Forest” and “Lasso” algorithms. The overlapping genes from these two methods were further refined using the “GMM” algorithm to identify the core interaction genes between rheumatoid arthritis and mitophagy. A predictive model was then developed and validated using an external dataset. Finally, “CIBERSORT” was employed to analyze the proportions and interactions of immune cell subsets during immune infiltration, while “ssGSEA” was used to examine the associations between the rheumatoid arthritis-mitophagy core interaction genes and immune cell subsets. “ssGSEA” was also utilized to analyze the “GO” and “KEGG” biological pathways of the core interaction genes.

RESULTS AND CONCLUSION: (1) Totally 807 differentially expressed genes in rheumatoid arthritis were obtained by differential analysis, 1 208 genes were selected from two feature modules by “WGCNA” analysis, 1 136 genes were sorted out from the MitoCarta 3.0 database, and 53 HUB genes were obtained from the intersection of the three genes as rheumatoid arthritis-mitophagy related genes. (2) The results of functional enrichment analysis of related genes showed that the cellular processes were mainly related to mitophagy, peroxisome metabolism, cellular senescence, and necroptosis. (3) The three machine learning algorithms identified four rheumatoid arthritis-mitophagy core interaction genes (*DNAJA3*, *C12orf65*, *AKR7A2*, and *PDHB*). The area under the curve of nomoscore was 0.989, and the area under the curve values of rheumatoid arthritis-mitophagy core interaction genes verified by the receiver operating characteristic curve of external patient samples were all greater than 0.7. (5) Immunoregulatory analysis showed that the mitophagy process in rheumatoid arthritis was closely associated with memory B cells, M0 macrophages, activated memory CD4 T cells, and resting memory CD4 T cells. (6) The biological pathway analysis revealed that the core interaction genes were strongly associated with 821 “GO” pathways ($|\text{cor}| > 0.8, P < 0.001$) and 48 “KEGG” pathways ($|\text{cor}| > 0.8, P < 0.001$). The key biological processes identified were related to mitochondrial DNA metabolic process, mitochondrial DNA repair, mitochondrial DNA replication, mitochondrial genome maintenance, positive regulation of mitochondrial depolarization, and positive regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway. To conclude, *DNAJA3*, *C12orf65*, *AKR7A2*, and *PDHB* are the core interaction genes of the mitophagy process in rheumatoid arthritis, which play key roles in disease progression by participating in specific immune processes and have precise and predictive effects on the diagnosis of rheumatoid arthritis.

Key words: mitophagy; rheumatoid arthritis; machine learning algorithm; weighted gene co-expression network analysis; predictive model; external validation; immune cells; biological function

Funding: National Natural Science Foundation of China, No. 82360937 (to CYP); Natural Science Foundation of Guangxi Zhuang Autonomous Region, No. 2023JJA140318 (to CYP); Guangxi Integrated Bone and Joint Degenerative Disease Multidisciplinary Innovation Team Project, No. GZKJ2310 (to CYP [project participant]); Innovation Project of Guangxi Graduate Education of Guangxi University of Chinese Medicine, No. YCSY2023040 (to LJG)

How to cite this article: LI JG, CHEN YP, HUANG KQ, CHEN ST, HUANG CH. Rheumatoid arthritis from the perspective of mitophagy: Interaction analysis based on multiple machine learning algorithms. *Zhongguo Zuzhi Gongcheng Yanjiu*. 2025;29(26):5595-5607.

0 引言 Introduction

类风湿关节炎是一种系统性自身免疫性疾病，患者临床表现为对称性的多发小关节炎、疼痛和肿胀^[1]，病理特征包括滑膜增生、软骨退化、成纤维样滑膜细胞浸润软骨、骨面及软骨下骨侵蚀，疾病后期往往导致严重的功能障碍，甚至关节功能丧失或残疾^[2]。类风湿关节炎全球患病率据报道约为 0.46%，范围从 0.39%–0.54% 不等^[3]。虽然类风湿关节炎可以在任何年龄段发病，但 40 岁以上患病风险显著增加，女性受影响的概率是男性的两到三倍^[4]，而且患病率呈逐渐上升趋势。目前有研究表明类风湿关节炎的发病机制与包括 *HLA-DRB1* 在内多个等位基因的异常表达、表观遗传因素如 DNA 甲基化和组蛋白乙酰化等生物过程有关^[5-6]。由于类风湿关节炎具有多因素病因和十分复杂的发病机制，至今仍是一项医学难题^[7-8]，因此探讨类风湿关节炎发病过程的重要机制，确定关键

基因及相关分子作用过程具有重要意义。

线粒体是一种分布非常广泛的双层膜细胞器，几乎存在于机体所有细胞中^[9]，主要功能是产生三磷酸腺苷为细胞提供能量，同时还具有调节钙稳态、细胞增殖和代谢等重要功能^[10]。通过自噬选择性地清除线粒体称为线粒体自噬，是线粒体质量控制的重要机制^[11]，其过程主要体现为：在泛素化依赖或非泛素化依赖的受体介导途径下，受损的线粒体被包裹在自噬体内，转运到溶酶体进行降解和再循环，这对于线粒体的基础周转和生理条件的维持至关重要^[12]。但目前关于线粒体自噬在疾病中的作用仍存在争议，一些研究表明，线粒体自噬通过发挥细胞保护作用可以延缓疾病进展，如在骨关节炎中，线粒体自噬可以调节活性氧水平，进一步增加机体软骨细胞的存活率^[13]。还有研究表明，线粒体损伤可能与类风湿关节炎的发病有关^[9]。

机器学习方法是一类新兴的涵盖多种算法在内的数据驱动研究方法,已广泛应用于生物医学领域^[14]。尽管已有研究利用机器学习算法探索类风湿关节炎的关键基因,但尚无从线粒体自噬角度研究其发病机制的方法学研究。此研究首次应用多种机器学习算法包括随机森林(random forest, RF)、LASSO(least absolute shrinkage and selection operator)和高斯混合模型(gaussian mixture model, GMM)算法,深入分析类风湿关节炎中的线粒体自噬相关机制。“RF”算法通过构建多个决策树,评估相关基因的重要性,具有处理多元共线性和缺失数据的优势;“LASSO”算法则通过引入L1正则化项,在高维数据中有效筛选出关键特征,适用于小样本研究^[15]。为了确保“RF”和“LASSO”算法筛选结果的可靠性,使用“GMM”算法进行概率建模,识别潜在分布模式,从而进一步验证并筛选核心互作基因。

此研究旨在通过多种机器学习算法稳健并准确地鉴定类风湿关节炎疾病进展中线粒体自噬核心互作基因,进一步构建风险预测模型评估类风湿关节炎的患病风险,为类风湿关节炎的早期临床决策提供新的思路,同时为类风湿关节炎中线粒体自噬机制研究提供新的视角。

1 资料和方法 Data and methods

1.1 设计 “WGCNA”分析、“GO”与“KEGG”功能富集分析、“RF”机器学习算法、“LASSO”机器学习算法、“GMM”机器学习算法、类风湿关节炎预测模型构建及验证、外部样本验证、“CIBERSORT”免疫浸润、“ssGSEA”免疫调控分析、“ssGSEA”生物学功能分析。

1.2 时间及地点 实验于2024年5-7月在广西中医药大学附属瑞康医院完成。

1.3 资料 从“GEO DataSets”公共数据库(<https://www.ncbi.nlm.nih.gov/>)筛选类风湿关节炎相关微阵列数据集,搜索关键词为“rheumatoid arthritis”,具体检索语句如下: (“arthritis, rheumatoid” [MeSH Terms] OR rheumatoid arthritis[All Fields]) AND “Homo sapiens” [porgn] AND (“gse” [Filter] AND “Expression profiling by array” [Filter]),共获得198个相关数据集,按样本量从高到底排序,过滤掉与类风湿关节炎诊断二分类变量不相关的数据集,筛选获得训练集GSE15573芯片,获得验证集GSE97779芯片与GSE55235芯片(表1)。

1.4 方法

1.4.1 类风湿关节炎差异表达基因筛选 使用R软件中“limma”包与“affy”包,对GSE15573、GSE97779和GSE55235基因数据集进行数据清洗与矩阵整理,对基因表达矩阵数据去除重复值、去除缺失值、表达分布归一化和log₂标准化处理,制作分组矩阵、比较矩阵并进一步利

表1 | “GEO DataSets”数据集信息

Table 1 | “GEO DataSets” dataset information

序 号	数据集	所属 平台	样本数量		组织来源	数据类型	数据 分类
			正常	类风湿关节炎			
1	GSE15573	GPL6102	15	18	外周血液	全基因组转录组	训练集
2	GSE97779	GPL570	5	9	滑膜	全基因组转录组	验证集
3	GSE55235	GPL96	10	10	滑膜	全基因组转录组	验证集

用“贝叶斯”函数处理校正。对于独立训练集GSE15573进行差异基因筛选。采用平均绝对折叠变化法,计算所有差异基因log FC绝对值的平均值与二倍绝对值的标准差之和,从而确定符合数据分布情况的log FC Cut Off阈值,当log FC绝对值大于此阈值且当 $-\log_{10}(P\text{Value}) < 0.05$ 时,所筛选的基因作为类风湿关节炎差异表达基因,继而使用“ggplot2”函数包和“pheatmap”函数包绘制类风湿关节炎差异表达基因的火山图和差异化热图。

1.4.2 “WGCNA”分析 加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA)是一种用于分析基因相关性的系统生物学模式方法,主要应用于不同样品间的基因表达数据。该方法能够识别在表达水平上表现出高度一致变化的基因组,并通过分析这些基因组之间的关系及其与特定表型的关联,来识别潜在的生物标志物或治疗靶标。为了分析数据清洗后的基因表达矩阵,使用R中的“WGCNA”和“limma”包,首先去除非典型样本,然后对各样本进行聚类,继而构建拓扑重叠矩阵和无标度网络,随之进行动态基因模块的融合,并识别与临床表型最一致的模块,最后将所挑出模块中的基因进行组织汇总,获得“WGCNA”特征基因。

1.4.3 类风湿关节炎-线粒体自噬相关基因筛选及“GO”和“KEGG”分析 从“MitoCarta3.0”数据库(<https://www.broadinstitute.org/mitocarta/>)中下载人类线粒体相关表达基因^[16]。使用类风湿关节炎差异表达基因、“WGCNA”特征基因和“MitoCarta 3.0”数据库基因取交集得到类风湿关节炎-线粒体自噬相关基因,继而将交集过程利用“VennDiagram”函数包绘制成Venn图。然后使用R函数包“enrichplot”和“clusterProfiler”,将整理好的类风湿关节炎-线粒体自噬相关基因进行“GO”和“KEGG”功能富集分析并可视化。

1.4.4 “RF”机器学习算法筛选核心基因 运用R软件的“randomForest”和“ggplot2”包,对类风湿关节炎-线粒体自噬相关基因表达数据进行“RF”算法分析,以筛选出具有显著分类能力的“RF Genes”特征基因。

1.4.5 “LASSO”机器学习算法筛选核心基因 利用R软件的“broom”和“glmnet”包,对类风湿关节炎-线粒体自噬相关基因表达数据进行“LASSO”机器学习算法筛选,获得显著特征基因“LASSO Genes”。

1.4.6 “GMM”机器学习算法同级验证核心基因 使用“GMM”算法相关 R 包“mclust”和“SimDesign”对“RF Genes”特征基因和“LASSO Genes”特征基因同时分别进行独立聚类验证,使筛选出的特征基因得到进一步确认,确保其在分类任务中具有稳定的表现和高准确性。进一步将验证后的“RF Genes”与“LASSO Genes”作交集,以获得类风湿关节炎-线粒体自噬(RA-Mitophagy)核心互作基因,并进行 Venn 图可视化。

1.4.7 预测模型构建及验证 基于类风湿关节炎-线粒体自噬核心互作基因,通过引用 R 包“rms”“nomogramFormula”“Hmisc”“rmda”构建二项逻辑回归模型,继而预测个体是否存在患类风湿关节炎的风险。通过绘制 Nomogram 图将预测模型可视化,并生成预测分数“nomoscore”。同时基于 Bootstrap 方法对模型进行校准,绘制校准曲线后观察所构建预测模型的拟合效果,从而进行内部验证。最后进行临床决策曲线分析,以评估预测模型的临床效用。

1.4.8 外部患者样本的工作特征曲线验证 综合“sva”“tinyarray”“FactoMineR”“factoextra”R 包对 GSE97779 与 GSE55235 数据集进行去批次合并,并在合并前后进行主成分分析以验证去批次效果,从而制作外部验证所使用的独立验证集。在独立外部验证集中,对预测模型对应的类风湿关节炎-线粒体自噬核心互作基因进行诊断效能外部验证。通过“pROC”R 函数包绘制受试者工作特征曲线,观察各核心互作基因所对应曲线下面积,当曲线下面积 > 0.6 时,代表其诊断效能良好,以确保筛选出的特征基因在实际应用中的稳定性和准确性,以进一步支持核心互作基因在分类任务中的有效性,从而为后续研究提供可靠的依据。

1.4.9 “CIBERSORT”免疫浸润分析 结合“MsigDB”数据库(<https://www.gsea-msigdb.org/gsea/msigdb/>)22 种免疫细胞“gmt”文件数据^[17],并基于“CIBERSORT”“ggpubr”“reshape2”“tidyverse”“preprocessCore”“doParallel”R 包函数,深入分析类风湿关节炎样本中的免疫细胞浸润情况,识别出与类风湿关节炎关联性较高的关键免疫细胞类型并分析免疫细胞亚群之间的相关性,继而通过两种可视化手段展示分析结果。

1.4.10 “ssGSEA”单样本-免疫细胞亚群间关联性分析 采用“ssGSEA”(Single-sample Gene Set Enrichment Analysis)方法对类风湿关节炎样本进行免疫浸润分析。“ssGSEA”是一种扩展的基因集合富集分析方法,用于评估每个样本中预定义基因集的富集程度,与传统“GSEA”不同,“ssGSEA”可以在单个样本上进行分析,而不是对样本组进行比较。通过“ssGSEA”分析,深入了解每个样本中免疫相关基因集的活性,为类风湿关节炎的免疫机制研究

提供重要的参考依据。此研究主要通过 R 软件“GSVA”及“GSEABase”函数包对样本及相关免疫细胞数据集进行处理分析,获取各样本与免疫细胞亚群间相关性数据,进行样本-免疫细胞相关性热图可视化及分组箱式表达差异可视化。

1.4.11 类风湿关节炎-线粒体自噬核心互作基因-免疫细胞亚群间关联性分析 采用“ggcorrplot”函数包进行分析,获取核心互作基因与免疫细胞亚群间关联性数据集,然后绘制各个核心互作基因与免疫细胞亚群间相关性“棒棒糖”图,最后利用“linkET”函数包,构建核心互作基因-免疫细胞亚群间关联网络矩阵图。

1.4.12 基于“ssGSEA”的核心互作基因相关生物学功能分析 采用“ssGSEA”方法,结合“MsigDB”数据库中“GO”与“KEGG”通路相关“gmt”文件数据,分析核心互作基因与相关生物学功能之间的关联程度,并筛选出其中与线粒体自噬过程密切相关的基因-功能映射,以从线粒体自噬角度阐明核心互作基因的相关生物学功能,为探索类风湿关节炎的发病机制提供新的视角和理论依据。

1.5 主要观察指标 ①类风湿关节炎-线粒体自噬相关基因的筛选及关键细胞通路可视化结果;②多机器学习算法下类风湿关节炎-线粒体自噬核心互作基因的筛选及验证;③预测模型的构建与外部验证;④核心互作基因与免疫细胞亚群间的关联性 & 免疫调控过程;⑤核心互作基因与线粒体相关的生物学功能分析结果。

1.6 统计学分析 采用 R (版本: 4.4.1) 软件进行数据分析。采用 GraphPad Prism (版本: 10.3.0) 和 R (版本: 4.4.1) 软件进行绘图。实验均重复 3 次,以非配对 t 检验进行组间差异比较,以 Spearman 相关性分析研究基因、免疫细胞、生物学功能或通路间关联程度, $P < 0.05$ 为差异有显著性意义。此研究统计学方法由广西中医药大学生物医学专家指导与审核。

2 结果 Results

2.1 类风湿关节炎差异表达基因筛选结果 通过 R 函数包“limma”和“affy”分析类风湿关节炎组和正常组的基因转录组数据,并根据平均绝对折叠变化法获得的 $\log_2 \text{FC Cut Off}$ 及 $-\log_{10}(P\text{-Value}) < 0.05$, 筛选差异基因总计 807 个,其中上调基因数量为 439 个,下调基因数量为 368 个。将差异基因进行火山图可视化(图 1A),并对上调及下调基因中表达量最高的各 30 位基因进行热图可视化(图 1B)。

2.2 “WGCNA”分析结果 “WGCNA”可作为一种初步筛选重要性基因的强大方法,通过对标准化后的矩阵数据进一步处理,发现数据集中无不良样本及基因被删除,然后对所有样本聚类,发现 GSM389724 与 GSM389732 两

例样本离群, 当“cutHeight”=45 时, 切除离群样本。通过计算不同软阈值的“scale free topology model fit”, 继而构建无标度网络(图 2A)。随后, 集合不同表达水平基因的基因模块被分离出来, 通过聚类分析和动态剪树法识别模块, 并合并相似的模块(图 2B)。计算模块特征基因与临床表型(正常对照和类风湿关节炎)之间的相关性, 生成相关性矩阵和 P 值矩阵, 继而绘制模块与临床表型的相关性热图(图 2C), 结果显示绿松石色模块($r=\pm 0.66$, $P=6\times 10^{-5}$)、粉色模块($r=\pm 0.54$, $P=0.002$) 意义最显著。提取上述模块基因, 其中绿松石色模块含基因 970 个, 粉色模块含基因 238 个, 共有 1 208 个模块特征基因与临床表型最相关。

2.3 类风湿关节炎-线粒体自噬相关基因筛选及“GO”和“KEGG”分析结果 从“MitoCarta 3.0”数据库中下载人类线粒体特征表达基因, 数据集包含 1 136 个基因, 以 807 个类风湿关节炎差异表达基因、1 208 个“WGCNA”特征基因和人类线粒体 1 136 个特征表达基因取交集, 获得类风湿关节炎-线粒体自噬相关基因总计 53 个(图 3A)。对上述交集基因进行“GO”和“KEGG”功能富集分析, “GO”富集分析排行前 5 位的可视化结果显示生物过程主要与有氧呼吸、细胞呼吸、电子传递链、有氧电子传递链、ATP 合成偶联电子传递有关, 细胞组分主要与含蛋白线粒体复合物、线粒体内膜、线粒体基质、呼吸链复合物、线粒体呼吸体相关, 分子功能主要与电子传递活性、NADH 脱氢酶(泛醌)活性、NADH 脱氢酶(醌)活性、NADH 脱氢酶活性、NAD(P)H 脱氢酶(醌)活性有关(图 3B)。“KEGG”功能富集分析结果包含 4 条细胞过程通路, 富集程度由高到低依次为线粒体自噬、过氧化物酶体代谢、细胞衰老、坏死性凋亡(图 3C)。

2.4 “RF”机器学习算法筛选核心诊断基因 采用 R 函数包“randomForest”的“RF”算法对 53 个类风湿关节炎-线粒体自噬相关基因进行进一步筛选并对随机森林模型误差率曲线图(图 4A)和基因重要性排名图(图 4B)可视化, 获得 30 个相关性较高的基因, 当重要性评分 > 0.6 时, 获得“RF”算法相关性最佳的 11 个基因, 分别为 SLIRP、NDUFB3、PUS1、DNAJA3、PDHA1、PDHB、PTRH2、C12orf65、AKR7A2、UQCRCQ、ECHS1。

2.5 “LASSO”机器学习算法筛选核心诊断基因 通过 R 函数包“tidyverse”“broom”和“glmnet”对 53 个类风湿关节炎-线粒体自噬相关基因进行“LASSO”回归分析, 绘制出“LASSO”回归模型交叉验证误差曲线图(图 5A)和“LASSO”回归系数路径图(图 5B)。根据交叉验证误差曲线图可知, “LASSO”筛选的特征基因数量在 6-9 之间, 输出基因筛选结果, 获得 9 个相关性最高的基因, 包括 DNAJA3、C12orf65、AKR7A2、HADH、PDHB、BLOC1S1、

COX7C、SLC25A5、NDUFS5。

2.6 “GMM”机器学习算法同级验证核心互作基因 将“RF”算法与“LASSO”算法筛选结果取交集, 进行 Venn 图可视化(图 6A), 取得 4 个交集基因分别为 DNAJA3、C12orf65、AKR7A2、PDHB。通过 R 函数包“mclust”和“SimDesign”对“RF”和“LASSO”回归分析筛选出的特征基因分别进行“GMM”独立聚类验证, “RF”和“LASSO”的基因筛选结果分别被归聚为 9 个类别, 验证“RF”结果下创建了 2 047 个训练模型(图 6B); 验证“LASSO”结果下创建了 511 个训练模型(图 6C)。输出两侧最佳模型结果, 包括 DNAJA3、PUS1、AKR7A2、PTRH2、ECHS1、PDHB、C12orf65、SLC25A5、COX7C 共 9 个基因, 其中 DNAJA3、C12orf65、AKR7A2、PDHB 均位于“GMM”算法下最优训练模型基因结果内, 因此被筛选作为 4 个类风湿关节炎-线粒体自噬核心互作基因。

2.7 预测模型构建及验证结果 在二分类逻辑回归分析模式下, 对 4 个类风湿关节炎-线粒体自噬核心互作基因进行诊断列线图模型构建, 列出各个基因的分值区间, 以用来评估基因表达水平在不同分值下的风险贡献, 并通过相加每个基因的分值计算出总分, 通过总分评估相对应的风险概率。总分越高, 患类风湿关节炎的风险越大(图 7A)。同时基于 Bootstrap 方法, 重复 1 000 次, 进行模型校正曲线绘制(图 7B), 以进行内部验证, 结果发现校正曲线与理想曲线贴合度高(均方误差 = 0.005 15), 证明模型的构建精度良好。进一步构建临床决策曲线(图 7C), 发现预测模型曲线与所有样本患病曲线之间离散度较高, 表明模型的临床效用较高。

2.8 外部患者样本的工作特征曲线验证 将 GSE97779 与 GSE55235 数据集去批次合并获得外部患者样本的独立验证集, 通过对比去批次前后的主成分分析(图 8A, B), 发现去批次效果良好。在独立外部验证集基因表达数据中进行核心互作基因的受试者工作特征曲线分析(表 2)并进行可视化(图 8C), 结果发现所有核心互作基因的曲线下面积均大于 0.7 ($P < 0.01$), 表明诊断效能良好, 证明 4 个类风湿关节炎-线粒体自噬核心互作基因在实际应用中对类风湿关节炎的预测具有良好的稳定性与准确性。

表 2 | 核心互作基因在外部验证集中的受试者工作特征曲线分析
Table 2 | Receiver operating characteristic curve analysis for core interaction genes in the external validation set

基因	曲线下面积	95% 置信区间		P 值	截断值	约登指数	灵敏度	特异性
		下限	上限					
DNAJA3	0.779	0.625	0.933	0.006	9.181	0.498	0.632	0.867
C12orf65	0.888	0.777	0.998	< 0.001	8.540	0.667	1.000	0.667
AKR7A2	0.775	0.615	0.935	0.007	10.363	0.533	1.000	0.533
PDHB	0.783	0.630	0.935	0.005	10.075	0.512	0.579	0.933

2.9 “CIBERSORT”免疫浸润分析结果 通过“CIBERSORT”算法,利用“MsigDB”数据库的22种免疫细胞“gmt”数据和类风湿关节炎差异基因表达矩阵数据进行免疫浸润分析,进行免疫细胞堆积比例图可视化(图9A)及免疫细胞相关性热图可视化(图9B),可以看出呈高正相关性的免疫细胞对有活化的记忆性CD4 T细胞与 $\gamma\delta$ T细胞、活化的记忆性CD4 T细胞与静息态肥大细胞、M1型巨噬细胞与嗜酸性粒细胞、 $\gamma\delta$ T细胞与静息态肥大细胞。其中, $\gamma\delta$ T细胞与静息态肥大细胞、活化的记忆性CD4 T细胞、静息态树突细胞、M0型巨噬细胞、中性粒细胞均具有较高正相关性。呈高负相关性的免疫细胞对有 $\gamma\delta$ T细胞与静息态NK细胞、活化的记忆性CD4 T细胞与静息态NK细胞、活化的肥大细胞与静息态树突细胞。其中静息态NK细胞与静息态肥大细胞、活化的记忆性CD4 T细胞、 $\gamma\delta$ T细胞、静息态树突细胞、M0型巨噬细胞、中性粒细胞均有高负相关性。结果表明,在类风湿关节炎线粒体自噬过程中, $\gamma\delta$ T细胞与活化的记忆性CD4 T细胞的免疫状态较为活跃。

2.10 “ssGSEA”单样本-免疫细胞亚群间关联性分析结果 通过R包“GSVA”相关函数初步分析各样本和免疫细胞亚群间的相关性,鉴定不同样本中相关免疫细胞的浸润丰度,继而进行免疫浸润热图可视化(图10A)及表达差异分组箱式图可视化(图10B)。免疫浸润热图显示类风湿关节炎组 $\gamma\delta$ T细胞、巨噬细胞的表达显著高于正常组,正常组调节性T细胞的表达显著高于类风湿关节炎组。表达差异分组箱式图显示类风湿关节炎组与正常组间活化的CD8 T细胞、CD56亮型自然杀伤细胞、嗜酸性粒细胞、 $\gamma\delta$ T细胞、1型辅助性T细胞、2型辅助性T细胞的表达均具有显著性差异($P < 0.001$)。

2.11 类风湿关节炎-线粒体自噬核心互作基因-免疫细胞亚群间关联性分析结果 首先,通过R包“ggcorrplot”分析核心互作基因与免疫细胞亚群间的相关性,根据4个核心互作基因的表达矩阵数据与“CIBERSORT”免疫浸润结果数据,利用Shapiro-Wilk检验与封装好的Spearman相关性分析函数进行相关计算,获得基因与免疫细胞亚群间的关联性数据,继而进行核心互作基因-免疫细胞关联性“棒棒糖”图可视化(图11A-D)。图中所示,DNAJA3、C12orf65、PDHB与记忆B细胞均具有强正相关性,DNAJA3、AKR7A2、PDHB与M0型巨噬细胞均呈强负相关性,此外,C12orf65还与活化的记忆性CD4 T细胞呈强负相关性,AKR7A2与静息态记忆性CD4 T细胞呈强正相关性。

其次,通过R包“linkET”进行汇总性可视化,绘制核心互作基因-免疫细胞亚群间关联性网络矩阵图(图11E),多核心互作基因共同作用时,其中以记忆B细胞、活化的记忆性CD4 T细胞、调节性T细胞、M0型巨噬细胞、

中性粒细胞的参与最为重要。同时,在类风湿关节炎线粒体自噬过程中,活化的记忆性CD4 T细胞与 $\gamma\delta$ T细胞、静息态肥大细胞之间的正相关表达作用最强,单核细胞和中性粒细胞之间的负相关表达作用最强。

2.12 基于“ssGSEA”的核心互作基因相关生物学功能分析结果 “ssGSEA”算法下共获得821种与类风湿关节炎-线粒体自噬核心互作基因相关的“GO”生物学功能($|\text{cor}| > 0.8, P < 0.001$),其中生物过程数量为600种,细胞组分数量为72种,分子功能数量为149种,并筛选出与每个核心互作基因正相关性和负相关性最强的前3位生物学功能进行展示(表3)。同时获得48条与类风湿关节炎-线粒体自噬核心互作基因相关的“KEGG”信号通路($|\text{cor}| > 0.8, P < 0.001$),同样筛选出与每个核心互作基因正相关性和负相关性最强的前3位信号通路进行展示(表4)。最后,逐条筛选上述全部获得的821种“GO”生物学功能和48条“KEGG”信号通路,发现有6种生物学功能与线粒体自噬密切相关(表5)。

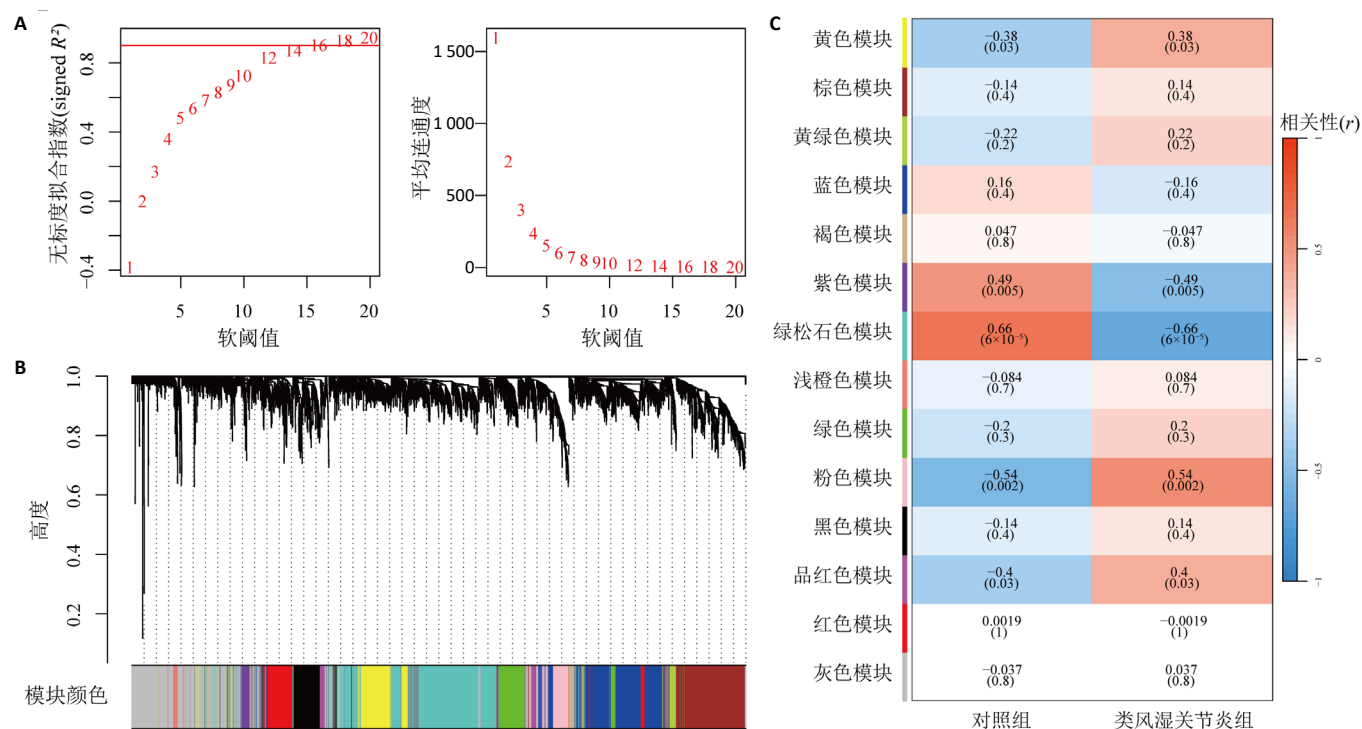
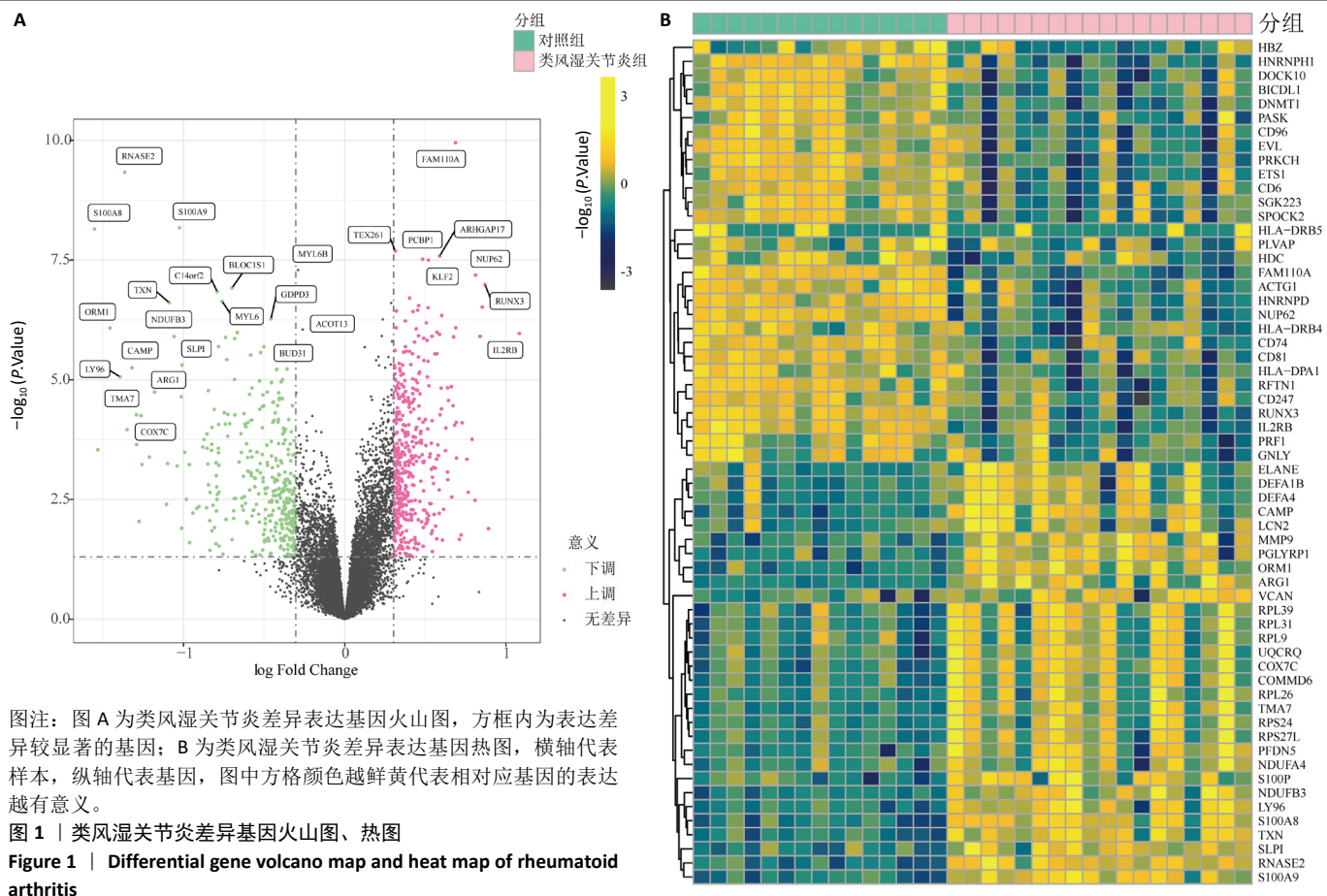
3 讨论 Discussion

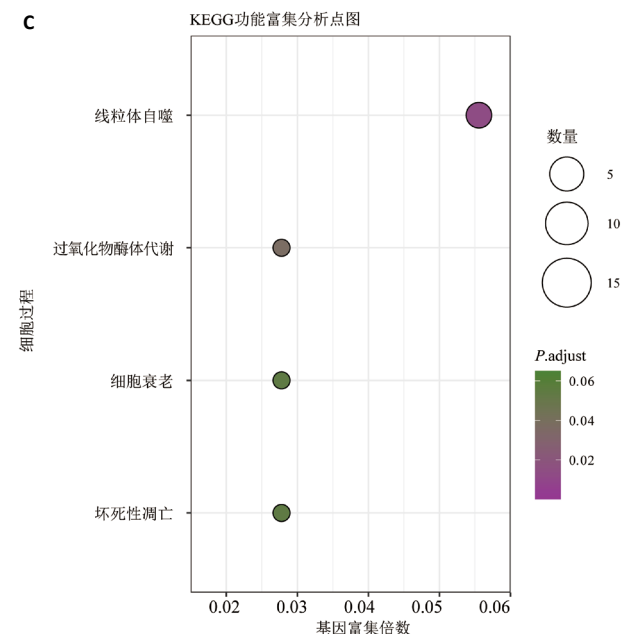
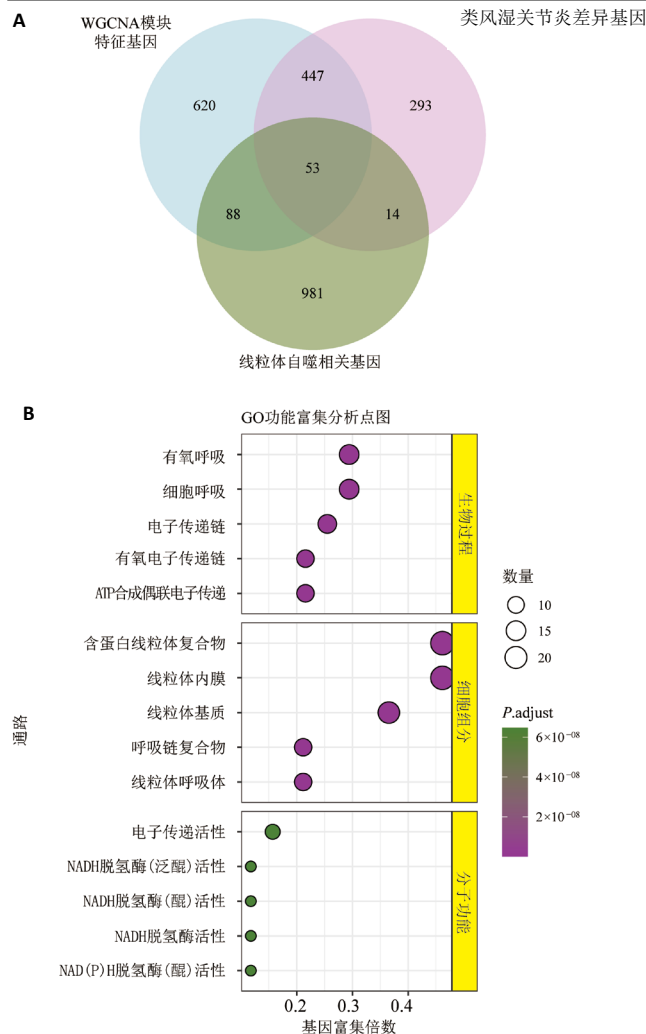
类风湿关节炎是最常见的风湿性疾病之一,其病程隐匿且渐进,常导致关节肿胀和疼痛,最终引发残疾,并带来高昂的社会经济成本^[18-21]。尽管传统和靶向抗风湿药物(csDMARD和b/tsDMARD)能够缓解病情,但约40%的患者对这些疗法反应不佳,导致类风湿关节炎被认为是“难治性”疾病^[22]。因此,早期诊断和治疗尤为重要,但目前尚无明确的诊断标准^[1],深入研究类风湿关节炎的发病机制并构建预测模型对临床决策具有重大意义。

自噬是指细胞通过自我吞噬降解其成分的过程,分为非选择性和选择性自噬。1967年,DETER等^[23]首次提出自噬的概念。线粒体自噬是一种选择性自噬,由LEMASTERS于2005年首次提出,主要通过受体介导的机制靶向降解受损的线粒体,是维持细胞健康的重要过程^[24-26]。

在类风湿关节炎的进展中,自噬与成纤维细胞样滑膜细胞的存活密切相关,且抑制自噬可以减少破骨细胞的数量,从而减轻骨侵蚀^[27]。此外,与成骨细胞和破骨细胞增殖、分化相关的多种信号通路,包括PINK1/Parkin,均与线粒体自噬的调控过程有关^[28]。以上均表明线粒体自噬在骨稳态中发挥重要作用。

然而,线粒体自噬对类风湿关节炎疾病进展的具体影响方式及其中关键机制目前尚未清晰,此研究综合“GEO DataSets”数据库、“MitoCarta3.0”数据库、“MsigDB”数据库数据,利用多种机器学习算法成功筛选出4个类风湿关节炎-线粒体自噬核心互作基因,同时考虑核心互作基因可能与类风湿关节炎疾病进展中的关键发病机

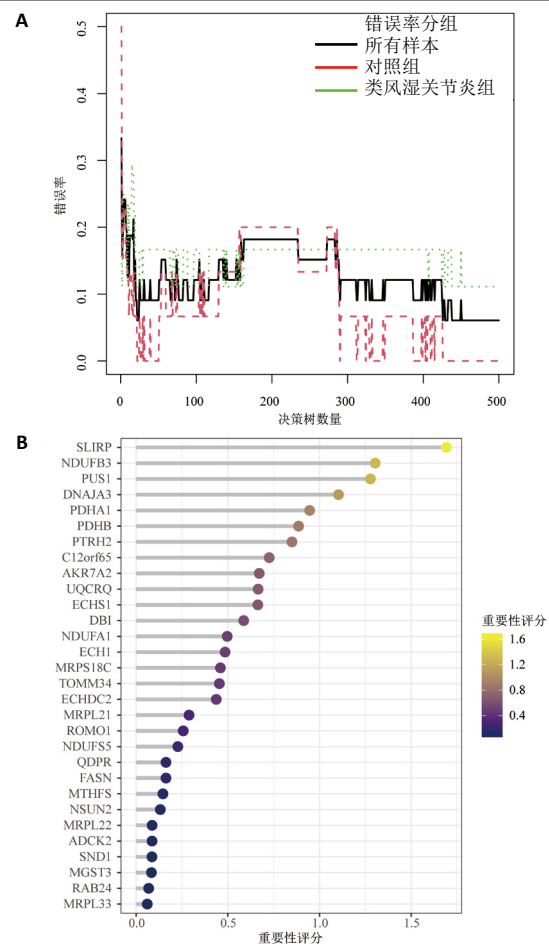




图注: 图A为类风湿关节炎差异表达基因、线粒体自噬基因和“WGCNA”模块特征基因交集 Venn 图; B为“GO”功能富集分析, 结果分别为生物过程、细胞组分与分子功能; C为“KEGG”富集功能分析结果, 结果显示线粒体自噬基因富集倍数最高 ($P < 0.05$)。

图3 | 类风湿关节炎-线粒体自噬相关基因筛选及“GO”“KEGG”富集分析结果

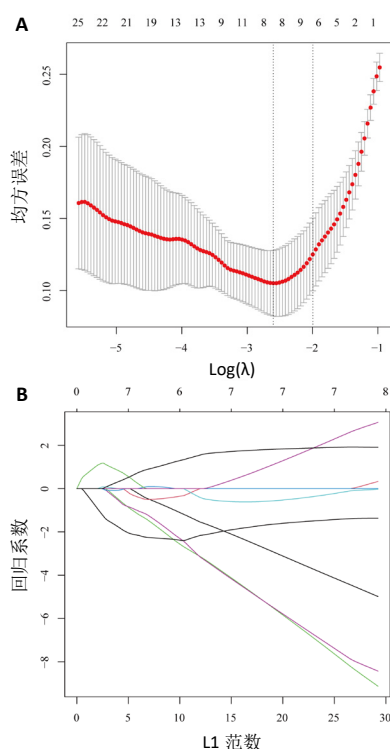
Figure 3 | Screening of rheumatoid arthritis-mitophagy related genes and enrichment analysis results in “GO” and “KEGG”



图注: 图A为类风湿关节炎-线粒体自噬相关基因随机森林(RF)模型误差率曲线图, 红线表示对照组误差, 绿线表示类风湿关节炎组误差, 黑色线条表示所有样品误差; B为类风湿关节炎-线粒体自噬相关基因评分排序结果, 颜色越鲜黄表示重要性评分越高。

图4 | “Random Forest”机器学习算法筛选“RF”特征基因

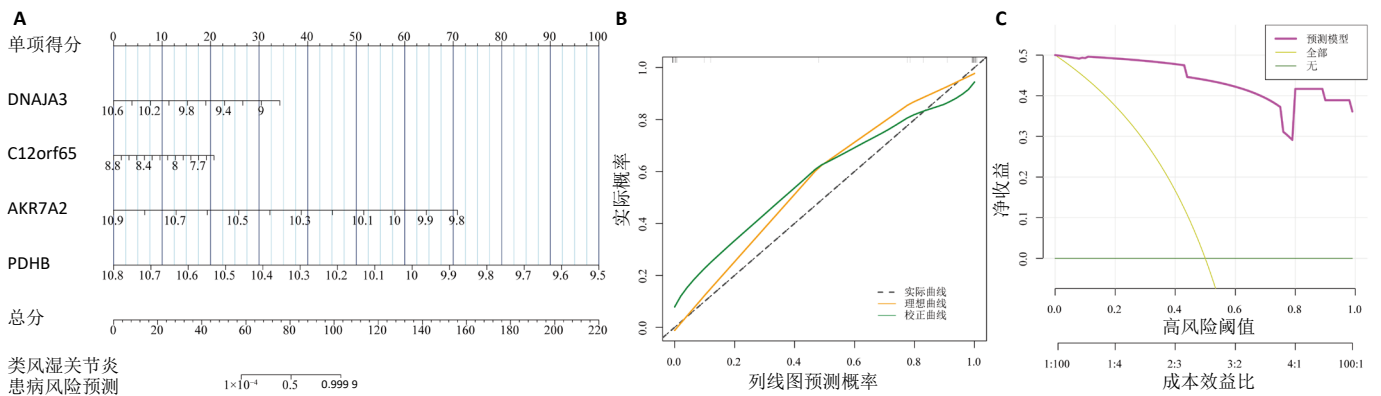
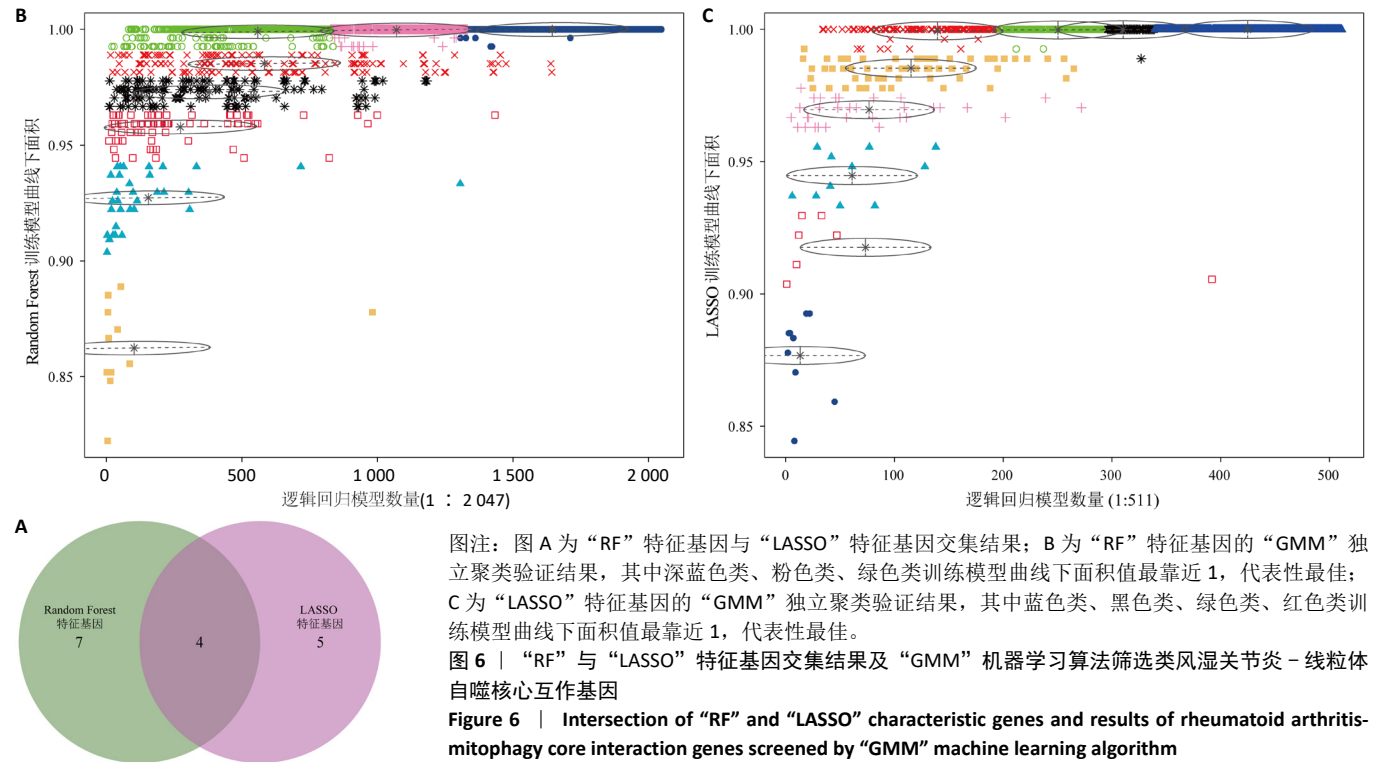
Figure 4 | Selection of “RF” characteristic genes using the “Random Forest” machine learning algorithm



图注: 图A为“LASSO”交叉验证误差曲线图, 虚线对应区间代表合适 λ 取值下的所筛选基因数量范围; B为“LASSO”回归系数路径图, X轴表示系数的L1范数, L1范数越小, λ 越大, 较大 λ 下系数为0的模型将被抛弃。

图5 | “LASSO”机器学习算法筛选“LASSO”特征基因

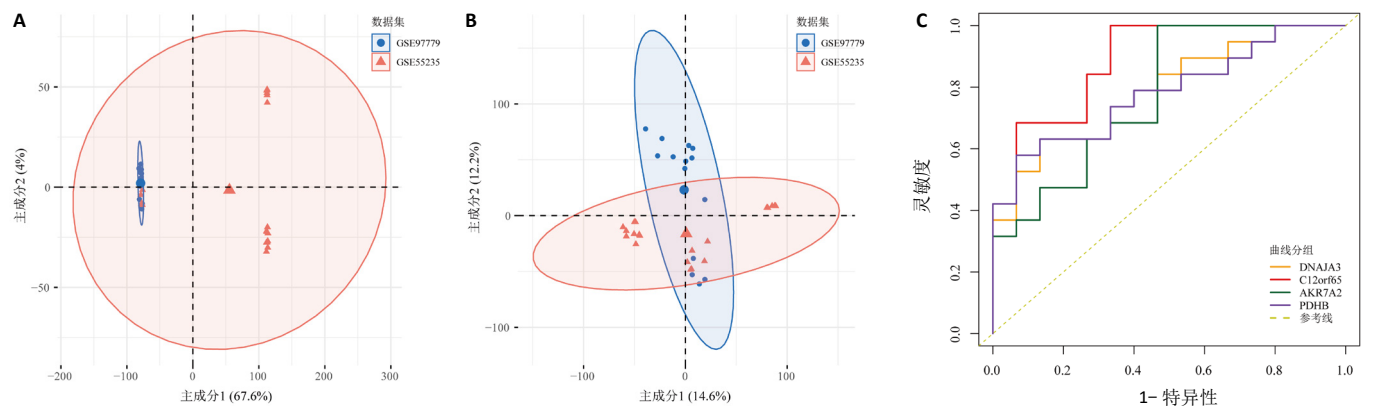
Figure 5 | Selection of “LASSO” characteristic genes using the “LASSO” machine learning algorithm



图注：图 A 为去批次前的主成分分析，显示 2 个数据集样本的基因表达离散程度较为明显；B 为去批次后的主成分分析，显示 2 个数据集样本的基因表达离散程度显著降低；C 为 4 个核心互作基因的受试者工作特征曲线图。

图 7 | 类风湿关节炎-线粒体自噬核心互作基因预测类风湿关节炎患病风险模型的构建及评估

Figure 7 | Construction and evaluation of a rheumatoid arthritis risk prediction model based on rheumatoid arthritis-mitophagy core interaction genes



图注：图 A 为去批次前的主成分分析，显示 2 个数据集样本的基因表达离散程度较为明显；B 为去批次后的主成分分析，显示 2 个数据集样本的基因表达离散程度显著降低；C 为 4 个核心互作基因的受试者工作特征曲线图。

图 8 | 类风湿关节炎-线粒体自噬核心互作基因在独立外部验证集中的受试者工作特征曲线分析

Figure 8 | Receiver operating characteristic curve analysis of rheumatoid arthritis-mitophagy core interaction genes in an independent external validation gene set

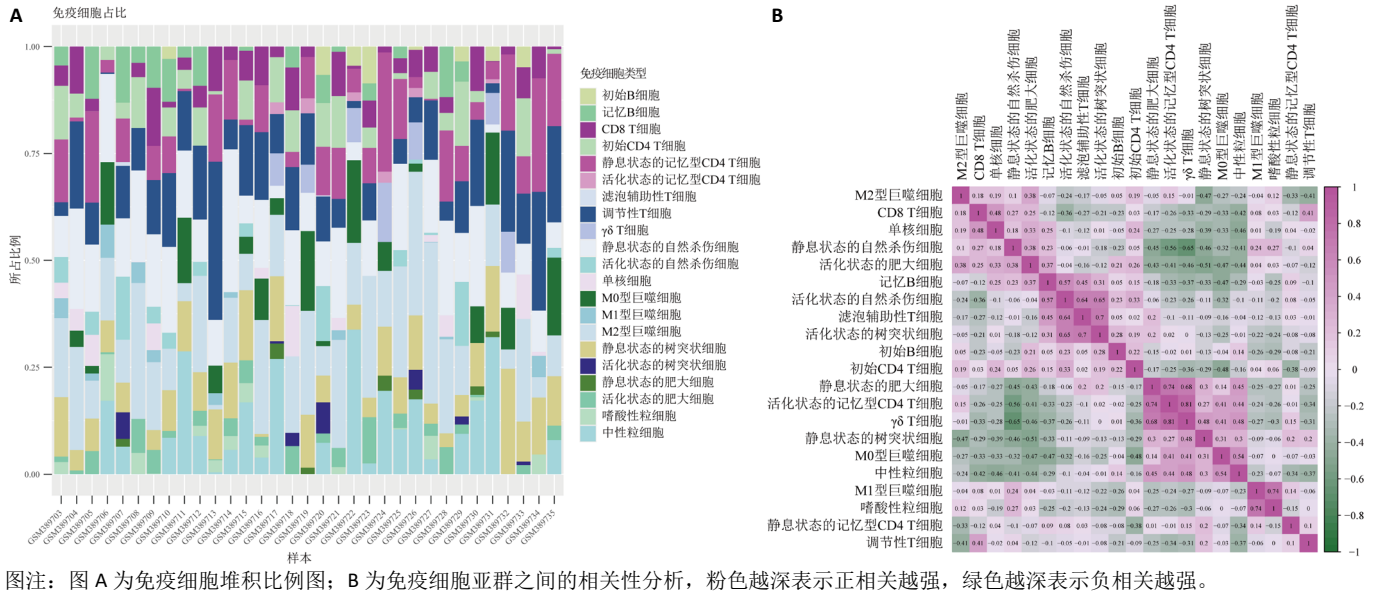


图 9 | 类风湿关节炎样本中免疫细胞浸润程度的分析

Figure 9 | Analysis of immune cell infiltration levels in rheumatoid arthritis samples

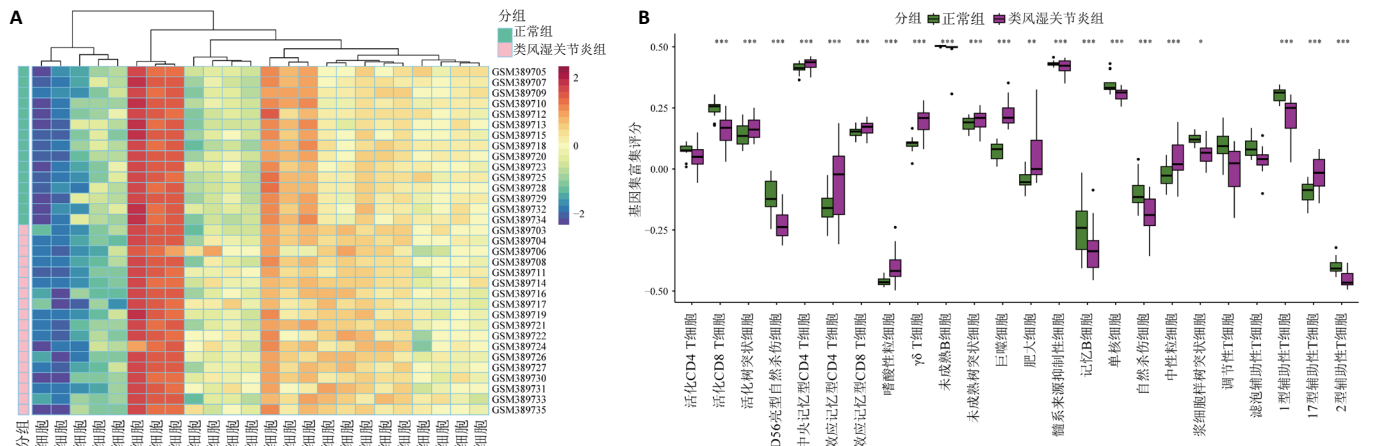


图 10 | “ssGSEA” 样本 - 免疫细胞亚群间相关性分析结果

Figure 10 | The correlation analysis between samples and immune cells using “ssGSEA”

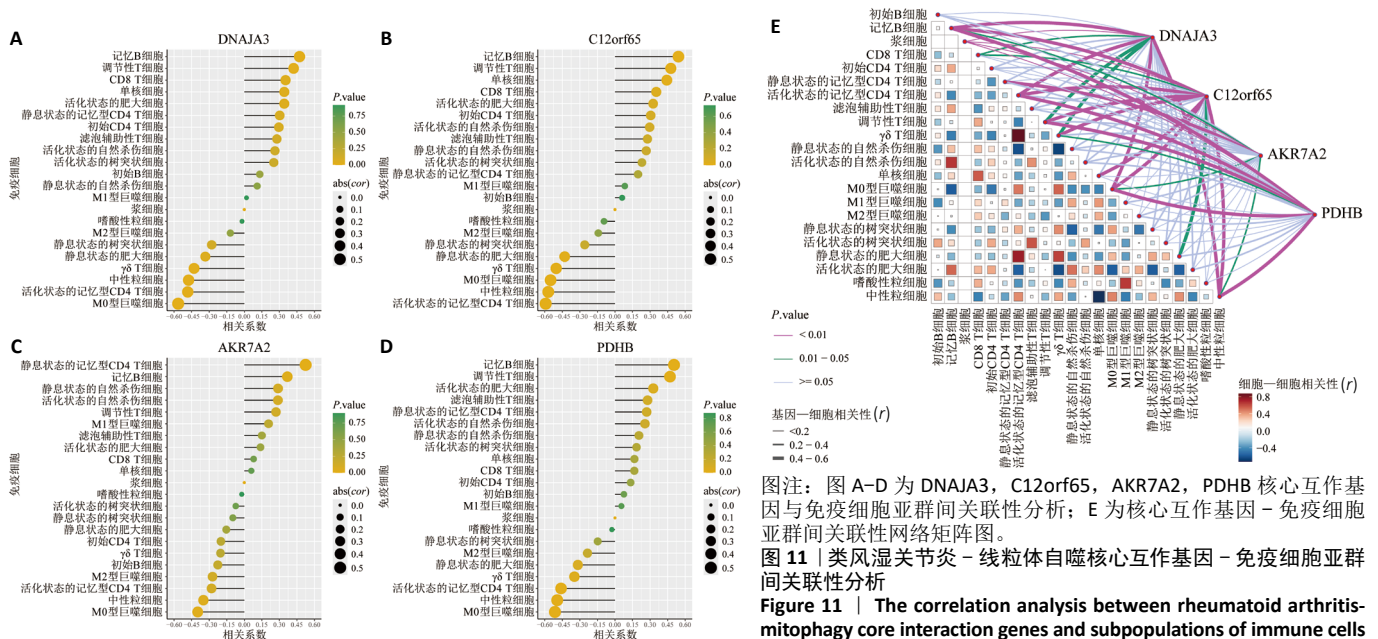


图 11 | 类风湿关节炎 - 线粒体自噬核心交互基因 - 免疫细胞亚群间关联性分析

Figure 11 | The correlation analysis between rheumatoid arthritis-mitophagy core interaction genes and subpopulations of immune cells

表 3 | 基于“ssGSEA”的核心互作基因相关“GO”生物学功能分析结果
Table 3 | GO biological function analysis of core interaction genes based on ssGSEA

互作基因	生物学功能		相关性 (cor)	P 值
	类型	名称		
AKR7A2	分子功能	环氧化物水解酶活性	0.922	< 0.000 1
AKR7A2	分子功能	作用于醚键的水解酶活性	0.922	< 0.000 1
AKR7A2	生物过程	小分子生物合成过程	0.868	< 0.000 1
AKR7A2	生物过程	雄性配子生成	-0.866	< 0.000 1
AKR7A2	生物过程	蛋白质重折叠	-0.853	< 0.000 1
AKR7A2	生物过程	成纤维细胞凋亡过程的正向调控	-0.853	< 0.000 1
C12orf65	分子功能	二酰基甘油依赖性丝氨酸 / 苏氨酸激酶活性	0.920	< 0.000 1
C12orf65	分子功能	小核糖体亚基 rRNA 结合	0.917	< 0.000 1
C12orf65	生物过程	Fcε 受体信号通路	0.903	< 0.000 1
C12orf65	分子功能	血管紧张素受体活性	-0.924	< 0.000 1
C12orf65	生物过程	脂蛋白定位	-0.907	< 0.000 1
C12orf65	生物过程	Toll 样受体 9 信号通路正向调控	-0.907	< 0.000 1
DNAJA3	生物过程	CD4 ⁺ αβ T 细胞活化的正向调控	0.950	< 0.000 1
DNAJA3	生物过程	Fcε 受体信号通路	0.946	< 0.000 1
DNAJA3	生物过程	Th17 型免疫反应的正向调控	0.946	< 0.000 1
DNAJA3	生物过程	一氧化氮合酶的生物合成过程	-0.938	< 0.000 1
DNAJA3	生物过程	细胞对拓扑错误蛋白的应答	-0.930	< 0.000 1
DNAJA3	生物过程	树突状细胞细胞因子的生成	-0.920	< 0.000 1
PDHB	细胞组分	α-酮酸脱氢酶复合体	0.915	< 0.000 1
PDHB	细胞组分	丙酮酸脱氢酶复合体	0.915	< 0.000 1
PDHB	分子功能	作用于醛或酮基供体、以二硫化物为受体的氧化还原酶活性	0.915	< 0.000 1
PDHB	生物过程	未成熟 B 细胞分化	-0.868	< 0.000 1
PDHB	生物过程	参与痛觉感知的机械刺激检测	-0.849	< 0.000 1
PDHB	生物过程	神经元投射分枝	-0.845	< 0.000 1

表 4 | 基于“ssGSEA”的核心互作基因相关“KEGG”信号通路分析结果
Table 4 | KEGG signaling pathway analysis of core interaction genes based on ssGSEA

互作基因	生物学功能	相关性 (cor)	P 值
AKR7A2	糖原降解	0.850	< 0.000 1
AKR7A2	生长因子受体酪氨酸激酶 -RAS-ERK 信号通路	0.800	< 0.000 1
AKR7A2	HIV Vpr 介导的 CDC25 细胞周期 G ₂ /M 调控	0.797	< 0.000 1
AKR7A2	自噬小体成核、延长、成熟及 Sequestosome 1 样受体	-0.845	< 0.000 1
AKR7A2	TBK1 介导的自噬小体形成	-0.845	< 0.000 1
AKR7A2	PRNP-PI3K-NOX2 信号通路	-0.808	< 0.000 1
C12orf65	朊病毒 PrP ^{Sc} 构象介导的 mGluR5-Ca ²⁺ 凋亡通路	0.831	< 0.000 1
C12orf65	朊病毒 PrP ^{Sc} 构象介导的钙运输	0.723	0.001 0
C12orf65	HBV HBx 介导的 ERK 信号通路	-0.724	0.000 7
C12orf65	CD80/CD86/CD28-PI3K 信号通路	-0.839	< 0.000 1
C12orf65	突变导致的 PSEN1 异常前向轴突运输	-0.839	< 0.000 1
C12orf65	乙酰肝素硫酸的生物合成	-0.833	< 0.000 1
DNAJA3	RTK-PLCG-ITPR 信号通路	0.884	< 0.000 1
DNAJA3	BCR-PLCG-ITPR 信号通路	0.880	< 0.000 1
DNAJA3	Ca ²⁺ -PLCD-ITPR 信号通路	0.880	< 0.000 1
DNAJA3	TLR4-IRF3/7 信号通路	-0.920	< 0.000 1
DNAJA3	锌介导的前向轴突运输	-0.872	< 0.000 1
DNAJA3	CD80/CD86/CD28-PI3K 信号通路	-0.853	< 0.000 1
PDHB	双功能糖基化酶介导的碱基切除和链切割	0.830	< 0.000 1
PDHB	NEIL 糖基化酶介导的碱基切除和链切割	0.830	< 0.000 1
PDHB	β-氧化	0.820	< 0.000 1
PDHB	KSHV vGPCR 介导的 Gβγ-PI3K-JNK 信号通路	-0.864	< 0.000 1
PDHB	CCR/CXCR 介导的 Gβγ-PI3K-RAC 信号通路	-0.864	< 0.000 1
PDHB	TLR1/2/4-NFκB 信号通路	-0.825	< 0.000 1

制密切相关，建立诊断风险模型，并利用外部类风湿关节炎患者及正常人的巨噬细胞转录组数据进行代入验证，进一步通过免疫浸润分析类风湿关节炎-线粒体自噬核心互作基因与免疫细胞亚群间的互作关系，以解析探讨

表 5 | 核心互作基因与线粒体自噬相关生物学功能分析结果
Table 5 | Biological function analysis of core interaction genes associated with mitochondrial autophagy

互作基因	通路	相关性		P 值
		相关方向	cor	
C12orf65	参与凋亡信号通路的线粒体外膜通透性的正向调控	负相关	-0.839	< 0.000 1
DNAJA3	线粒体 DNA 复制	正相关	0.924	< 0.000 1
DNAJA3	线粒体 DNA 代谢	正相关	0.876	< 0.000 1
DNAJA3	线粒体基因组维护	正相关	0.876	< 0.000 1
DNAJA3	参与凋亡信号通路的线粒体外膜通透性的正向调控	负相关	-0.853	< 0.000 1
PDHB	线粒体 DNA 修复	正相关	0.830	< 0.000 1
PDHB	线粒体去极化的正向调控	正相关	0.830	< 0.000 1

线粒体自噬在类风湿关节炎疾病中的作用机制。

类风湿关节炎-线粒体自噬相关基因的“GO”“KEGG”富集分析结果十分明确，其中以线粒体自噬、电子传递链、含蛋白线粒体复合物、线粒体内膜、线粒体基质、线粒体呼吸体、坏死性凋亡相关细胞组分与过程最为重要。电子传递链包括复合物 I - IV，这些复合物在线粒体内膜上排列，负责电子的传递和 ATP 的生成，线粒体自噬通过清除受损的线粒体复合物，防止活性氧的过度积累^[29]。受损的线粒体通常会产生过量的活性氧，这些活性氧可以通过氧化线粒体 DNA、蛋白质和脂质，进一步加剧细胞损伤和炎症反应^[30]。此外，线粒体自噬与坏死性凋亡的关系也已被证实，线粒体自噬通过清除产生过量活性氧的线粒体，可以抑制坏死性凋亡的发生^[31]，这在类风湿关节炎中尤为重要，因为滑膜细胞中的线粒体功能障碍和活性氧积累可以触发坏死性凋亡，导致滑膜组织的进一步破坏和炎症加剧。

目前有关类风湿关节炎的研究十分庞杂，包括铁死亡、铜死亡等热点^[32-33]。但对于线粒体自噬在类风湿关节炎发病过程中相关基因层面的研究较少，此研究通过“WGCNA”分析及“RF”“LASSO”“GMM”多机器学习算法构建可靠的诊断模型，模型特征基因包括 DNAJA3、C12orf65、AKR7A2、PDHB，该模型在训练集以及外部样本验证数据中均表现出良好的预测性能，同时也为类风湿关节炎中线粒体自噬的发生、发展机制过程提供新的见解。

关于核心互作基因 DNAJA3，在免疫调控方面发现其与记忆 B 细胞、调节性 T 细胞正相关性较强，与 M0 型巨噬细胞、活化的记忆性 CD4 T 细胞负相关性最强。已有研究证实，DNAJA3 在增强免疫记忆和维持免疫耐受中具有重要作用，其缺失可能导致免疫功能紊乱，尤其影响 B 细胞的发育和功能，同时会导致线粒体质量和膜电位的改变^[34]。研究发现，线粒体功能障碍可能导致 DNAJA3 的异常积累，同时伴有多种干扰素刺激基因 (如 Ifi44 和 Gbp3) 以及模式识别受体和胞质核酸传感器 (如 Ddx58 和 Ifih1)

的显著表达,继而激活先天免疫反应,这表明 DNAJA3 在调节线粒体应激和免疫反应中扮演着重要角色^[35]。此外,还有研究表明 DNAJA3 能够影响线粒体蛋白质折叠过程,从而影响线粒体的正常功能^[36]。但目前尚无有关 DNAJA3 通过影响线粒体自噬而影响类风湿关节炎发病过程的研究,此研究的筛选结果可作为类风湿关节炎临床研究的一个新的切入方向。

对于核心互作基因 PDHB,在免疫调控过程中发现其同样与记忆 B 细胞、调节性 T 细胞正相关性最强,但相较 DNAJA3 新增了与中性粒细胞的高负相关性。有研究发现,PDHB 基因在丙酮酸脱氢酶复合体中起关键作用,该复合体对于丙酮酸转化为乙酰辅酶 A 的过程至关重要,而类风湿关节炎患者 PDHB 基因的表达下调,表明它可能在类风湿关节炎的病理机制中起相关作用^[37]。另有研究表明,在类风湿关节炎疾病进展中,PDHB 可能通过影响细胞能量代谢和氧化应激反应参与细胞死亡过程,尤其与铜离子的应激反应密切相关^[38]。然而,对于 PDHB、线粒体自噬和类风湿关节炎的共同作用方式目前仍无相关研究,此研究结果可为其提供新的参考。

此研究发现 C12orf65 是另外一个重要的类风湿关节炎-线粒体自噬核心互作基因,目前相关研究仅发现其通过影响线粒体核糖体相关的质量控制机制而影响线粒体的正常功能^[39]。此外,目前尚无 AKR7A2 基因、线粒体自噬与类风湿关节炎共同作用的研究。C12orf65 与 AKR7A2 对于类风湿关节炎诊断预测的临床价值有待进一步更深入的研究证实。

此研究通过免疫浸润研究及构建核心互作基因-免疫细胞亚群关联性网络矩阵图发现,在类风湿关节炎线粒体自噬过程中,以记忆 B 细胞、活化的记忆性 CD4 T 细胞、调节性 T 细胞、M0 型巨噬细胞、中性粒细胞的参与最为频繁,可能考虑作为免疫调控过程中的核心免疫细胞。恰有相关研究表明,线粒体代谢与 T 细胞、巨噬细胞和中性粒细胞等免疫细胞有重要相关作用^[40]。另有学者表明,先天免疫激活是线粒体功能障碍直接或间接影响类风湿关节炎发病的一大重要方式,在这个过程中 T 细胞受到异常氧化磷酸化途径的影响而加剧炎症反应,巨噬细胞代谢表型发生改变而降低骨稳态^[41]。通过系列现有研究,可以发现免疫细胞的调控与类风湿关节炎的发病及线粒体自噬息息相关。

此外,还基于“ssGSEA”算法进行了核心互作基因的相关生物学功能分析,发现机器学习算法所筛选出的 4 个类风湿关节炎-线粒体自噬核心互作基因与广泛的生物学功能和信号通路之间存在着错综复杂的关联。有关线粒体功能方面,DNAJA3 表达上调与线粒体 DNA 的复制、

代谢、基因组维护的正向调控密切相关,PDHB 表达上调则与线粒体 DNA 的修复和去极化的正向调控显著相关,但就参与凋亡信号通路的线粒体外膜通透性的正向调控这一过程而言,C12orf65 和 DNAJA3 均表现出强负相关性。作者考虑这可能与疾病状态下机体宏观生理稳态的自我调节有关,具体反映在线粒体自噬加速以维持细胞微观稳态的需求。然而,目前尚无明确的研究能够证实这一推论,这一领域仍需进一步深入探索。

与既往研究相比,此研究的优势体现在首次将高斯混合模型应用于疾病靶点基因的筛选领域。此前尚无文献报道“GMM”机器学习算法用于此类基因研究,而该研究将其与“RF”机器学习算法和“LASSO”机器学习算法相结合,确保了基因筛选的可靠性和诊断模型的稳健性,进一步拓展了机器学习在疾病靶点基因筛选中的应用,这也为探索线粒体自噬在类风湿关节炎中的作用机制提供了重要的技术基础。

综上所述,此研究综合多种机器学习算法,筛选出类风湿关节炎中与线粒体自噬相关的核心互作基因 DNAJA3、C12orf65、AKR7A2 和 PDHB,并作为疾病诊断标志物而构建了稳健的疾病预测模型。通过进一步分析免疫调控过程,发现类风湿关节炎中线粒体自噬与记忆 B 细胞、M0 型巨噬细胞、活化的记忆性 CD4 T 细胞、静息态记忆性 CD4 T 细胞显著相关。生物学功能分析表明,核心互作基因与线粒体 DNA 代谢、DNA 修复、DNA 复制、基因组维护以及线粒体去极化的正向调控密切相关,同时还参与了凋亡信号通路中线粒体外膜通透性的正向调控。这些发现不仅为类风湿关节炎的诊断提供了新的分子标志物,也为未来的治疗策略提供了潜在的靶点,对提升类风湿关节炎的精准诊断和个性化治疗具有重要意义,为类风湿关节炎的早期临床诊断和干预及机制探索提供了崭新切入点。

作者贡献: 李加根负责提出研究思路、数据分析及文章撰写;陈跃平负责研究的全局指导;黄柯琪负责收集数据;陈尚桐负责收集文献;黄川洪负责数据校对和论文审查。

利益冲突: 文章的全部作者声明,在课题研究和文章撰写过程中不存在利益冲突。

开放获取声明: 这是一篇开放获取文章,根据《知识共享许可协议》“署名-非商业性使用-相同方式共享 4.0”条款,在合理引用的情况下,允许他人以非商业性目的基于原文内容编辑、调整和扩展,同时允许任何用户阅读、下载、拷贝、传递、打印、检索、超级链接该文献,并为之建立索引,用作软件的输入数据或其它任何合法用途。

版权转让: 文章出版前全体作者与编辑部签署了文章版权转让协议。

出版规范: 该文章撰写遵守国际医学期刊编辑委员会《学术研究实验与报告和医学期刊编辑与发表的推荐规范》。文章出版前已经过专业反剽窃文献检测系统进行 3 次查重。文章经小同行外审专家双盲外审,同行评议认为文章符合期刊发稿宗旨。

4 参考文献 References

- [1] DI MATTEO A, BATHON JM, EMERY P. Rheumatoid arthritis. *Lancet*. 2023;402(10416):2019-2033.
- [2] LEE EE, SHIN A, LEE J, et al. All-cause and cause-specific mortality of patients with rheumatoid arthritis in Korea: A nation-wide population-based study. *Joint Bone Spine*. 2022;89(1):105269.
- [3] ALMUTAIRI K, NOSSENT J, PREEN D, et al. The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review. *Rheumatol Int*. 2021;41(5):863-877.
- [4] SCOTT IC, WHITTLE R, BAILEY J, et al. Rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis epidemiology in England from 2004 to 2020: An observational study using primary care electronic health record data. *Lancet Reg Health Eur*. 2022;23:100519.
- [5] WOUTERS F, MAURITS MP, VAN BOHEEMEN L, et al. Determining in which pre-arthritis stage HLA-shared epitope alleles and smoking exert their effect on the development of rheumatoid arthritis. *Ann Rheum Dis*. 2022;81(1):48-55.
- [6] NEMTSOVA MV, ZALETAEV DV, BURE IV, et al. Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis. *Front Genet*. 2019;10:570.
- [7] GRAVALLESE EM, FIRESTEIN GS. Rheumatoid Arthritis- Common Origins, Divergent Mechanisms. *N Engl J Med*. 2023;388(6):529-542.
- [8] ALIVERNINI S, FIRESTEIN GS, MCINNES IB. The pathogenesis of rheumatoid arthritis. *Immunity*. 2022;55(12):2255-2270.
- [9] CLAYTON SA, MACDONALD L, KUROWSKA-STOLARSKA M, et al. Mitochondria as Key Players in the Pathogenesis and Treatment of Rheumatoid Arthritis. *Front Immunol*. 2021;12:673916.
- [10] KAN S, DUAN M, LIU Y, et al. Role of Mitochondria in Physiology of Chondrocytes and Diseases of Osteoarthritis and Rheumatoid Arthritis. *Cartilage*. 2021;13(2_suppl):1102S-1121S.
- [11] ASHRAFI G, SCHWARZ TL. The pathways of mitophagy for quality control and clearance of mitochondria. *Cell Death Differ*. 2013;20(1):31-42.
- [12] ONISHI M, YAMANO K, SATO M, et al. Molecular mechanisms and physiological functions of mitophagy. *EMBO J*. 2021;40(3):e104705.
- [13] ANSARI MY, KHAN NM, AHMAD I, et al. Parkin clearance of dysfunctional mitochondria regulates ROS levels and increases survival of human chondrocytes. *Osteoarthritis Cartilage*. 2018;26(8):1087-1097.
- [14] GOODSWEN SJ, BARRATT JLN, KENNEDY PJ, et al. Machine learning and applications in microbiology. *FEMS Microbiol Rev*. 2021;45(5):fuab015.
- [15] 李晨曦, 石婕, 魏巍, 等. 颞下颌关节 PVNS 和 RA 共病机制及潜在治疗靶点的初步研究 [J]. *生物信息学*, 2025, 23(1):71-80.
- [16] RATH S, SHARMA R, GUPTA R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res*. 2021;49(D1):D1541-D1547.
- [17] CASTANZA AS, RECLA JM, EBY D, et al. Extending support for mouse data in the Molecular Signatures Database (MSigDB). *Nat Methods*. 2023;20(11):1619-1620.
- [18] QIAN H, DENG C, CHEN S, et al. Targeting pathogenic fibroblast-like synoviocyte subsets in rheumatoid arthritis. *Arthritis Res Ther*. 2024; 26(1):103.
- [19] MANKIA K, D'AGOSTINO MA, ROWBOTHAM E, et al. MRI inflammation of the hand interosseous tendons occurs in anti-CCP-positive at-risk individuals and may precede the development of clinical synovitis. *Ann Rheum Dis*. 2019;78(6):781-786.
- [20] ALETAHA D, SMOLEN JS. Diagnosis and Management of Rheumatoid Arthritis: A Review. *JAMA*. 2018;320(13):1360-1372.
- [21] SHI Y, ZHOU M, CHANG C, et al. Advancing precision rheumatology: applications of machine learning for rheumatoid arthritis management. *Front Immunol*. 2024;15:1409555.
- [22] NAGY G, ROODENRIJS NMT, WELSING PM, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. *Ann Rheum Dis*. 2021;80(1):31-35.
- [23] DETER RL, DE DUVE C. Influence of glucagon, an inducer of cellular autophagy, on some physical properties of rat liver lysosomes. *J Cell Biol*. 1967;33(2):437-449.
- [24] MIZUSHIMA N, KOMATSU M. Autophagy: renovation of cells and tissues. *Cell*. 2011;147(4):728-741.
- [25] LEMASTERS JJ. Selective mitochondrial autophagy, or mitophagy, as a targeted defense against oxidative stress, mitochondrial dysfunction, and aging. *Rejuvenation Res*. 2005;8(1):3-5.
- [26] WANG Y, LIU N, LU B. Mechanisms and roles of mitophagy in neurodegenerative diseases. *CNS Neurosci Ther*. 2019;25(7):859-875.
- [27] WANG S, DENG Z, MA Y, et al. The Role of Autophagy and Mitophagy in Bone Metabolic Disorders. *Int J Biol Sci*. 2020;16(14):2675-2691.
- [28] HUANG T, WANG Y, YU Z, et al. Effect of mitophagy in the formation of osteomorphs derived from osteoclasts. *iScience*. 2023;26(5):106682.
- [29] ZHAO RZ, JIANG S, ZHANG L, et al. Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *Int J Mol Med*. 2019; 44(1):3-15.
- [30] WANG S, LONG H, HOU L, et al. The mitophagy pathway and its implications in human diseases. *Signal Transduct Target Ther*. 2023;8(1):304.
- [31] YANG YD, LI ZX, HU XM, et al. Insight into Crosstalk Between Mitophagy and Apoptosis/Necroptosis: Mechanisms and Clinical Applications in Ischemic Stroke. *Curr Med Sci*. 2022;42(2):237-248.
- [32] LIU Y, LUO X, CHEN Y, et al. Heterogeneous ferroptosis susceptibility of macrophages caused by focal iron overload exacerbates rheumatoid arthritis. *Redox Biol*. 2024;69:103008.
- [33] HAN J, LUO J, WANG C, et al. Roles and mechanisms of copper homeostasis and cuproptosis in osteoarticular diseases. *Biomed Pharmacother*. 2024;174:116570.
- [34] SAYSON SL, FAN JN, KU CL, et al. DNAA3 regulates B cell development and immune function. *Biomed J*. 2024;47(2):100628.
- [35] MALETZKO A, KEY J, WITTIG I, et al. Increased presence of nuclear DNAA3 and upregulation of cytosolic STAT1 and of nucleic acid sensors trigger innate immunity in the ClpP-null mouse. *Neurogenetics*. 2021;22(4):297-312.
- [36] SHIN CS, MENG S, GARBIS SD, et al. LONP1 and mtHSP70 cooperate to promote mitochondrial protein folding. *Nat Commun*. 2021;12(1):265.
- [37] WU C, TAN S, LIU L, et al. Transcriptome-wide association study identifies susceptibility genes for rheumatoid arthritis. *Arthritis Res Ther*. 2021;23(1):38.
- [38] JIANG M, LIU K, LU S, et al. Verification of cuproptosis-related diagnostic model associated with immune infiltration in rheumatoid arthritis. *Front Endocrinol (Lausanne)*. 2023;14:1204926.
- [39] DESAI N, YANG H, CHANDRASEKARAN V, et al. Elongational stalling activates mitoribosome-associated quality control. *Science*. 2020; 370(6520):1105-1110.
- [40] BREDA CNS, DAVANZO GG, BASSO PJ, et al. Mitochondria as central hub of the immune system. *Redox Biol*. 2019;26:101255.
- [41] MA C, WANG J, HONG F, et al. Mitochondrial Dysfunction in Rheumatoid Arthritis. *Biomolecules*. 2022;12(9):1216.

(责任编辑: LCH, MZH, ZN, WL)