

类风湿关节炎与代谢综合征共病基因的表达特征分析及诊断价值评估

丁芸发^{1,2}, 邓安霞³, 祁腾飞², 张宏斌⁴, 余浩⁵, 吴良平^{2*}

¹广东医科大学附属第二医院肝胆外科, 广东湛江 524002; ²广州中医药大学金沙洲医院甲乳代谢外科, 广东广州 510168; ³省部共建中亚高发病成因与防治国家重点实验室/新疆医科大学第一附属医院心内科, 新疆乌鲁木齐 830054; ⁴南部战区总医院基础医学实验部, 广东广州 510010; ⁵南方医科大学珠江医院甲乳外科, 广东广州 510260

[中图分类号] R589.2 [文献标志码] A [DOI] 10.11855/j.issn.0577-7402.1314.2025.1022

[声明] 本文所有作者声明无利益冲突

[引用本文] 丁芸发, 邓安霞, 祁腾飞, 等. 类风湿关节炎与代谢综合征共病基因的表达特征分析及诊断价值评估[J]. 解放军医学杂志, 2026, 51(2): 219-231.

[收稿日期] 2024-09-04 [录用日期] 2024-12-09 [上线日期] 2025-10-22

[摘要] **目的** 采用生物信息学方法分析类风湿关节炎(RA)与代谢综合征(MS)共病基因的表达特征, 挖掘潜在的生物标志物并评估其诊断价值。**方法** 利用基因表达综合数据库(GEO)的微阵列数据集探究RA及MS的基因表达模式。采用差异表达基因(DEGs)分析和加权基因共表达网络分析(WGCNA)识别RA与MS共有的基因, 并通过基因本体(GO)和京都基因与基因组百科全书(KEGG)分析以及蛋白质-蛋白质相互作用网络探究这些基因的功能。采用随机森林和最小绝对值收缩和选择算子(LASSO)算法识别关键的共享关键基因, 并利用孟德尔随机化验证关键基因与RA之间的因果关系, 同时应用XGBoost机器学习技术建立诊断预测模型。通过CIBERSORT和基因集变异分析(GSVA)评估关键基因与免疫细胞浸润及代谢通路的相关性, 并利用单细胞转录组测序数据以及临床血液样本对这些关键基因的作用及表达进行验证。**结果** 使用limma R包对GSE93777和GSE98895数据集进行分析, 分别鉴定出259个和280个DEGs, 联合WGCNA技术, 共识别了88个在RA和MS中共表达的特征基因, 这些基因显著富集于免疫反应和代谢调控的生物过程。通过机器学习进一步筛选出24个关键的共享关键基因, 它们被有效地应用于开发预后预测模型。CIBERSORT和GSVA评估结果显示, 这些关键基因与外周血中的免疫功能和代谢活动紧密相关, 同时孟德尔随机化分析提示信号转导与转录激活因子3(STAT3)与RA风险存在潜在因果关系。单细胞RNA测序结果验证了颗粒酶A(GZMA)和STAT3基因具有较显著的诊断效能。**结论** 成功鉴定了RA和MS之间共有的关键调控基因; GZMA和STAT3基因与能量代谢过程呈正相关, 其参与的代谢途径可能与细胞活动密切相关。

[关键词] 类风湿关节炎; 生物信息学; 机器学习; 单细胞; 代谢综合征

Integrated multi-omics analysis of shared gene expression signatures and their diagnostic value in rheumatoid arthritis and metabolic syndrome comorbidity

Ding Yun-Fa^{1,2}, Deng An-Xia³, Qi Teng-Fei², Zhang Hong-Bin⁴, Yu Hao⁵, Wu Liang-Ping^{2*}

¹Department of Hepatobiliary Surgery, the Second Affiliated Hospital of Guangdong Medical University, Zhanjiang, Guangdong 524002, China

²Jinshazhou Hospital, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong 510168, China

³State Key Laboratory of Causes and Prevention of High Morbidity in Central Asia/Department of Cardiology, the First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang 830054, China

⁴Department of Basic Medical Experimentation, General Hospital of Southern Theater of Operations of PLA, Guangzhou, Guangdong 510010, China

[基金项目] 广东省科技计划项目(202002020069)

[作者简介] 丁芸发, 博士研究生, 主要从事减重手术改善代谢性疾病相关机制方面的研究

[通信作者] 吴良平, E-mail: drwulp@163.com

⁵Department of Thyroid and Breast Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong 510260, China

*Corresponding author, E-mail: drwulp@163.com

This work was supported by the Guangdong Provincial Science and Technology Program (202002020069)

[Abstract] Objective To investigate the shared gene expression signatures of comorbid genes between rheumatoid arthritis (RA) and metabolic syndrome (MS) using bioinformatics approaches, identify potential biomarkers, and evaluate their diagnostic utility. **Methods** This study harnessed microarray datasets from the Gene Expression Omnibus (GEO) to explore gene expression patterns in RA and MS. Differentially expressed genes (DEGs) were identified, and weighted gene co-expression network analysis (WGCNA) was applied to uncover genes common to both conditions. Functional enrichment analyses, including Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG), alongside protein-protein interaction (PPI) network analysis, were employed to elucidate the biological roles of these shared genes. Key hub genes were subsequently screened using random forest and least absolute shrinkage and selection operator (LASSO) algorithms. Furthermore, Mendelian randomization (MR) analysis was utilized to verify causal relationships between these key genes and RA. To translate these findings into clinical application, a diagnostic prediction model was developed using the XGBoost machine learning framework. The CIBERSORT algorithm and gene set variation analysis (GSVA) were used to explore the correlations between hub genes and immune cell infiltration as well as metabolic pathway activities. Finally, the expression and potential roles of these hub genes were rigorously validated through single-cell RNA sequencing (scRNA-seq) data and clinical blood samples. **Results** Analysis of the GSE93777 and GSE98895 datasets using limma R package identified 259 and 280 DEGs, respectively. Integration with WGCNA revealed 88 genes co-expressed in both RA and MS. Functional enrichment analysis revealed that these genes were significantly enriched in biological processes related to immune response and metabolic regulation. Subsequent refinement using machine learning algorithms (LASSO and random forest) pinpointed 24 key hub genes, which were then used to construct a prognostic prediction model. These hub genes demonstrated significant associations with immune functions and metabolic activities in peripheral blood. Additionally, Mendelian randomization (MR) analysis suggested a potential causal relationship between signal transducer and activator of transcription 3 (STAT3) and RA risk. Analysis of scRNA-seq data and clinical blood samples confirmed the diagnostic significance of two prominent hub genes: granzyme A (GZMA) and STAT3. **Conclusions** Key regulatory genes shared between RA and MS have been successfully identified. The GZMA and STAT3 genes are positively correlated with energy metabolism processes, suggesting that the metabolic pathways in which they participate may be closely associated with cellular activities.

[Key words] rheumatoid arthritis; bioinformatics; machine learning; single-cell; metabolic syndrome

类风湿关节炎(rheumatoid arthritis, RA)是一种慢性、全身性的自身免疫性疾病,主要影响关节及其周围的软组织^[1-2]。RA患者因全身性炎症而导致心血管疾病的风险增加,这可能使其预期寿命缩短5~10年^[3]。代谢综合征(metabolic syndrome, MS)临床较为常见,与心血管疾病风险增加相关,且具有空腹血糖升高、甘油三酯水平升高、高密度脂蛋白水平降低、腰围增加和高血压等一系列特征^[4]。越来越多的证据表明,RA与MS的多个组分存在关联,包括体重的异常变化、血脂异常、特定的脂肪因子模式及胰岛素抵抗,这些组分均可能增高心血管病患者的病死率^[5]。研究发现,在RA患者中存在较多MS,且这类患者发展为中重度RA的风险更高^[6]。研究还发现,MS相关RA(MS-RA)患者的疾病活动度高于非MS相关RA患者,且疾病活动度与MS的发生发展呈正相关^[7],提示MS可能为RA的发展提供了促炎微环境。近年来,代谢系统与免疫系统的交叉作用受到广泛关注^[8],这些相互作用的变化可能导致自身免疫性疾病的发展^[9]。脂肪因子(如瘦素、脂联素及抵抗素等)在代谢过程中扮演着重要角色,不仅与MS的发生有关,还参与RA的炎症过程及免疫

调节^[10]。目前,综合应用生物信息学手段来诊断疾病及识别致病基因已日益普及^[11-13],但对于RA与MS共有的诊断基因以及它们之间的关联基因,目前的认识仍有限。因此,本研究通过整合现有公共基因表达数据集,综合运用生物信息学方法及机器学习技术,挖掘RA与MS共病中的潜在关键基因,采用孟德尔随机化分析关键基因与疾病之间的因果关系,借助单细胞转录组学验证这些基因在疾病发生发展中的作用,并通过实时定量聚合酶链反应(qRT-PCR)实验验证关键基因是否可作为共病的筛查标志物;旨在鉴定出在RA及MS中起关键作用的关键基因,从而为MS-RA的防治提供新的思路及方法。

1 资料与方法

1.1 研究设计与分析流程 本研究采用多组学整合分析策略,以分析RA与MS共病的潜在分子机制。首先,从美国基因表达综合数据库(gene expression omnibus, GEO)获取RA和MS的基因表达数据及临床样本信息,用于差异表达基因(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)。随后,通过功能富集分析、蛋白质互作网络构建、机器学习筛选关键基因,并构建诊断预测模型。最后进一步利用免疫浸润分析、代谢通路活性评估及单细胞转录组数据验证关键基因的表达与功能,并通过孟德尔随机化分析初步探索关键基因与RA的潜在因果关系。

1.2 数据选择 以“RA”和“MS”为关键词,在GEO数据库中检索相关数据,最终选取两个与RA相关的基因表达数据集:(1)GSE17755,包含112例RA患者及45例健康对照样本;(2)GSE93777,包含232例RA患者和34例健康对照样本。在MS的研究中选取了GSE98895数据集(包含20例MS患者和20例健康对照样本),并从GSE235508数据集中获取了RA患者的单细胞测序数据。4个数据集的详细情况见附表1。为确保研究结果的可靠性,应用多种统计方法控制潜在的混杂因素,如性别、年龄等。研究工作流程详见附图1。

1.3 临床样本收集 回顾性收集2024年10月—2025年7月广州中医药大学金沙洲医院收治的MS患者、RA患者、MS合并RA(MS+RA)患者各6例。RA的诊断符合美国风湿病学会(ACR)的标准^[14]。MS诊断标准:超重或肥胖(体重指数 $>25.0\text{ kg/m}^2$)、高血糖(空腹血糖 $>6.1\text{ mmol/L}$ 或餐后2h血糖 $>7.8\text{ mmol/L}$,或已确诊并接受治疗的糖尿病)、高血压(收缩压/舒张压 $>140/90\text{ mmHg}$ 或正在接受治疗的高血压)、血脂异常(空腹甘油三酯 $>1.7\text{ mmol/L}$,或男性高密度脂蛋白胆固醇 $<0.9\text{ mmol/L}$ 、女性高密度脂蛋白胆固醇 $<1.0\text{ mmol/L}$),以上标准至少满足3项。同时从同一医院体检中心中招募健康对照组6例,均不符合Mets诊断标准且无RA家族史。收集所有参与者的临床资料,包括详细的病史、体格检查、生化血液检测(血糖、血脂、血压)及影像学检查(腹部超声、CT)结果。由于采集滑膜组织具有侵入性,难以获取,因此缺乏健康对照组的滑膜组织,而血液样本能够反映全身的炎症和代谢状态,故本研究选择采集Mets、RA患者的血液样本进行比较。采血后2h分离血浆、血清及白细胞,储存于 $-80\text{ }^\circ\text{C}$ 备用。在样本采集和处理过程中严格遵循无菌操作规程,以确保数据的质量及可靠性。收集上述外周血样本用于后续分子生物学实验分析(qPCR),对生物信息学分析筛选出的关键枢纽基因进行验证,并探究其表达水平与临床表型之间的关联。所有参与者的个人信息均已匿名化处理以保护其隐私。本研究已获广州中医药大学金沙洲医院伦理委员会审批(B2020-005-02),并严格遵循《赫尔辛基宣言》开展。

1.4 DEGs的鉴定和可视化 对GSE17755、GSE93777数据集进行WGCNA分析,以识别与RA、Mets相关

的关键基因模块。使用WGCNA软件包中的pickSoftThreshold函数确定合适的软阈值,并构建基因网络。在进行WGCNA分析前,从每个数据集的基因表达矩阵中选择中位绝对偏差值最高的前5000个基因,以确保所选基因具有足够的变异性并与疾病表型相关。在构建网络时,去除含有缺失值或低方差的基因,以确保数据的完整性及可靠性。根据无标度拓扑的相关性标准,设定无标度 R^2 值为0.9,构建邻接矩阵,并转换为拓扑重叠矩阵,以衡量基因间的相似性,为模块划分提供基础。使用基于平均链的层次聚类方法,将表达模式相似的基因聚类为模块。模块划分时,设定最小模块大小为30个基因,确保模块的生物学意义,并将模块切割高度设定为0.25,以区分不同模块。未能形成独立模块的基因归类为灰色模块。最后,通过计算模块与RA、Mets性状之间的Pearson相关系数,识别与疾病表型密切相关的关键基因模块。

1.5 通路及功能富集分析 通过综合分析DEGs及WGCNA的结果,筛选具有特征性的生物标志物。利用Metascape平台进行深入分析,包括基因本体(Gene Ontology, GO)功能富集分析及京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)通路分析,以进一步揭示特征基因在RA、MS中显著富集的生物过程及分子通路。筛选时设定严格条件,最小基因重叠数为3,最小富集倍数为1.5,以确保筛选的GO项及KEGG通路具有较高可信度。将富集阈值设定为校正后 $P<0.05$ 。

1.6 基于机器学习算法的关键基因筛选与验证 使用随机森林及最小绝对值收缩和选择算子(LASSO)等机器学习技术筛选关键枢纽基因。这些基因筛选自DEGs与WGCNA分析结果的交集。为确保分析结果的准确性,首先使用sva R包对GSE93777、GSE17755(验证集)及GSE98895数据集进行批次效应校正,以减少不同实验批次间的潜在差异,并增强数据集间的可比性。然后,通过主成分分析(principal component analysis, PCA)确认数据集间的一致性。在随机森林分析阶段,通过构建分类模型来评估特征的重要性,以识别对疾病分类有明显贡献的基因。随后使用glmnet R包中的LASSO回归技术对筛选后的数据进行降维,提炼出最具预测价值的基因。最后通过结合这两种算法筛选在疾病进程中发挥核心作用的关键基因。

1.7 基于关键基因的预后模型构建 使用R语言中的XGBoost包构建基于极端梯度提升(XGBoost)的分类模型,输入特征为关键基因的表达水平。选择GSE93777数据集(与RA相关)及GSE98895数据集(与MS相关)作为训练集,以训练XGBoost模型。由于

缺乏专门针对MS的验证数据集,使用与RA相关的GSE17755数据集作为验证集,以评估模型的泛化能力。模型性能通过受试者操作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)及精确率-召回率曲线进行评估。AUC衡量模型的整体区分能力,而精确率-召回率曲线则衡量模型的精确率及召回率,为模型预测效能提供全面评估。

1.8 关键基因表达与免疫浸润和代谢通路的相关性分析 CIBERSORT是一种基于基因表达数据的反卷积算法,可用于分析外周血样本以识别及量化免疫细胞亚群的组成。重点分析GSE93777及GSE98895数据集中的22种免疫细胞类型,以 $P<0.05$ 作为统计阈值,通过计算Pearson相关系数来评估基因表达水平与免疫细胞亚群浸润水平之间的相关性,并使用Vioplot及pheatmap R包进行数据可视化。Vioplot用于展示基因表达与免疫细胞浸润之间的相关性分布,pheatmap用于创建热图。通过基因集变异分析(gene set variation analysis, GSVA)评估基因集中表达的动态变化。利用GSVA R包量化50条关键代谢通路的活性,以探索RA及MS中活性水平显著变化的通路,显著性标准设为 $P<0.05$ 。进一步计算Pearson相关系数,以深入分析关键基因表达与代谢通路活性水平之间的相关性。采用pheatmap R包进行数据可视化。

1.9 单细胞数据处理 从GEO数据库下载并提取GSE235508数据集中的单细胞RNA测序(scRNA-seq)原始数据,采用Seurat 4.1.0软件在RStudio中进行数据处理。通过设定线粒体基因表达占比阈值为20%进行数据质量控制,以确保分析结果的准确性。筛选每个样本中变异性最高的3000个基因,并通过方差稳定化变换(VST)对数据进行归一化处理。使用Seurat中的ScaleData函数对高变异基因进行标准化处理,并通过RunPCA函数执行PCA,以降低数据维度并识别主要变异因素。基于前20个主成分,使用FindNeighbors及FindClusters函数进行细胞聚类,识别出17个不同的细胞群体。然后使用RunUMAP函数在低维空间中进行非线性降维,以可视化细胞群体。通过Single R包自动化标注细胞类型,并使用VlnPlot函数分析不同细胞中潜在生物标志物及代谢途径的表达模式。采用Wilcoxon秩和检验评估代谢途径评分的差异是否具有统计学意义。最后,使用FeaturePlot函数直观展示潜在生物标志物及代谢途径的分布,呈现它们在不同细胞类型中的位置。

1.10 外周血单核细胞的采集与处理 将含有抗凝剂的血液样本从采集管转移到已加入同等体积PBS的15 ml离心管中,轻轻吹打混匀,并在新的15 ml离心管中加入等体积的外周血单个核细胞(PBMC)分

离液,随后将稀释后的血液样本缓慢加入分离液上层,使血液样本浮于PBMC分离液上方。样本在20~25 °C下500×g离心35 min,随后用巴氏吸管轻轻吸取白细胞层并转移至新的15 ml离心管中,加入PBS直至总体积达10 ml,洗涤细胞样本后,在4 °C下500×g离心10 min去除上清液。向离心管中加入1 ml新鲜配制的1×红细胞裂解液,轻轻吸打混合后静置5 min,以裂解PBMC中残留的红细胞。裂解完成后向离心管中加入PBS至10 ml终止裂解,在4 °C下500×g离心10 min去除上清液。最后,向管中加入1 ml三唑溶液,轻轻混匀后,将混合物转移至1.5 ml EP管中。

1.11 采用qRT-PCR检测颗粒酶A(GZMA)及信号转导与转录激活因子3(STAT3)的mRNA表达水平 使用TRIpure总RNA提取试剂盒(武汉科鹿生物科技有限责任公司)从收集的临床样本中提取总RNA。随后使用QuantStudio 6 Flex System PCR仪器和EnTurbo™ SYBR Green PCR SuperMix试剂盒将分离的总RNA反转录为互补DNA(cDNA)。每个样本在3个独立孔中进行处理。引物序列:GZMA基因,正向AGGTTGATTGATGTGGGACAGCAG,反向GCAGGAGAGAAACGACAACCTGAGAG; STAT3基因,正向GAGAAGGACATCAGCGCAAGAC,反向GATAGACCAGTGGAGACACCAGGAT;以GAPDH作为内参基因。采用 $2^{-\Delta\Delta Ct}$ 法计算mRNA的相对表达量。

1.12 统计学处理 采用SPSS 20.0及R4.3.3软件进行统计分析和生物信息学分析。计量资料符合正态分布者以 $\bar{x}\pm s$ 表示,两组间比较采用 t 检验,多组间比较采用单因素方差分析,进一步的两两比较采用Tukey's HSD法;不符合正态分布者以 $M(Q_1, Q_3)$ 表示,组间比较采用Kruskal-Wallis H 检验,进一步两两比较采用Dunn检验并经Bonferroni校正。计数资料以例(%)表示,组间比较采用 χ^2 检验或Fisher精确检验。生物信息学分析中,差异基因的筛选采用limma包,筛选标准为错误发现率(FDR) <0.05 ,关键基因鉴定采用随机森林与LASSO回归分析,模型效能及诊断价值采用ROC曲线及AUC进行评估,相关性分析采用Pearson检验。 $P<0.05$ 为差异有统计学意义。

2 结 果

2.1 RA与MS中的DEGs鉴定 使用limma R包对RA相关的GSE93777数据集进行分析,鉴定出259个DEGs,其中220个基因表达上调,39个基因表达下调(图1A)。MS相关的GSE98895数据集中共鉴定出280个DEGs,其中179个基因表达上调,101个基因表达下调(图1B)。GSE93777及GSE98895数据集

DEGs的热图表达模式见附图2、附图3,在两个数

据集的DEGs中发现21个共同的基因(图1C)。

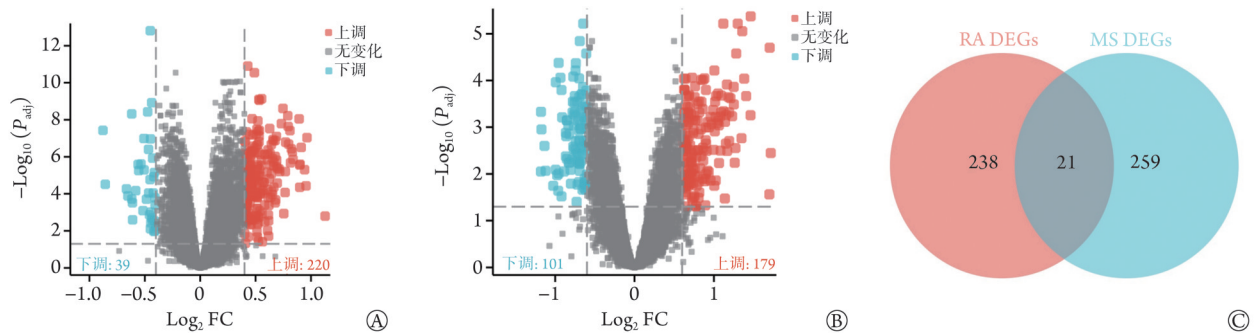


图1 类风湿关节炎(RA)与代谢综合征(MS)中差异表达基因(DEGs)的火山图

Fig.1 Volcano plot of differentially expressed genes (DEGs) between rheumatoid arthritis (RA) and metabolic syndrome (MS)

A. 数据集 GSE93777 差异基因表达火山图(红色点表示基因表达上调, 蓝色点表示基因表达下调); B. 数据集 GSE98895 差异基因表达火山图(红色点表示基因表达上调, 蓝色点表示基因表达下调); C. 差异基因交集韦恩图(交集处为共有基因)。FC. 差异倍数

2.2 RA和MS的加权基因共表达网络分析 在WGCNA分析中, 样本聚类树状图结果显示, 所有样本均表现出良好的聚类特性, 未发现明显的离群样本(图2A、B)。软阈值分析确定功率值5为构建网络的最佳参数, 此时平均连通数分布平稳(图2C、D)。基于此参数, 共划分出9个不同的基因共表达模块(图2E)。这些基因模块与临床特征之间的相关性分析结果显示, RA与MS明显相关, 尤其是在棕色模块中(图2F)。对棕色模块中的基因与RA、MS进行相关性分析, 结果显示, 棕色模块的特征基因与MS和RA均呈正相关($r=0.435$, $P=0.0006$, 图2G; $r=0.314$, $P=0.0003$, 图2H)。

2.3 富集分析及蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络构建 WGCNA分析识别出的棕色模块包含340个基因, 并通过比较筛选出24个候选基因, 富集分析结果显示, 这些基因在免疫反应正向调控、白细胞增殖、细胞因子信号转导、炎症反应调节和T细胞激活等过程中显著富集, 且在细胞代谢和丝裂原活化蛋白激酶(mitogen-activated protein kinase, MAPK)信号通路中也有显著富集(附图4A)。此外, 这些基因不仅富集于免疫相关通路, 且在代谢相关通路中也显著富集(附图4B)。在STRING数据库构建的PPI网络中, 白细胞介素(interleukin, IL)-1 β 、MAPK14及甘油醛-3-磷酸脱氢酶(glyceraldehyde-3-phosphate dehydrogenase, GAPDH)是排名前3位的关键基因(附图4C)。

2.4 随机森林及LASSO鉴定并验证潜在的共享关键基因 采用随机森林算法对340个候选基因进行重要性评估, 筛选出84个关键基因(图3A)。采用LASSO回归分析从DEGs集合中进一步筛选出30个基因(图3B), 重要性排名前30位的基因见图3C, LASSO回归系数路径变化见图3D。通过比较两种方法筛选结果的交集, 最终确定24个共有的关键基

因, 这些基因显示出极高的诊断潜力(图3E)。采用ROC曲线验证这些共享关键基因的诊断效能, 结果显示, *GZMA*和*STAT3*具有最高的AUC值(分别为0.87、0.88)(图3F, 附图5)。

2.5 基于XGBoost的预后模型性能评估 预后模型在训练集(GSE93777)中表现出优异的判别性能, 其AUC值与精确率-召回率曲线下面积分别达到0.983和0.997(附图6A)。验证集的AUC值分别达到了0.739和0.868, 提示预后模型在预测RA方面准确性较高(附图6B)。此外, 为检验模型是否适用于识别MS患者, 使用同一组中心基因对MS数据集(GSE98895)进行训练, 也取得了较好的预测效果, AUC值均为1(附图6C)。

2.6 免疫细胞浸润与共享关键基因的相关性 免疫浸润分析结果显示, 与健康对照相比, RA患者的初始B细胞和初始CD4⁺T细胞减少, M2型巨噬细胞、静息肥大细胞及中性粒细胞增多(附图7A)。MS患者表现为活化记忆CD4⁺T细胞及中性粒细胞减少, 静息NK细胞增多(附图7B)。同时, 本研究探究24个关键基因与免疫细胞浸润之间的关联, 结果显示, MS患者的*STAT3*基因与活化的NK细胞、活化的肥大细胞、活化记忆CD4⁺T细胞呈正相关, *GZMA*基因与M0型巨噬细胞、 γ 8T细胞及单核细胞呈正相关, 而*AAK1*、*FBX021*、*JARID2*及*FLNB*基因则与静息肥大细胞、单核细胞呈负相关(附图7C)。在RA数据集中也发现关键基因与免疫细胞的相关性(附图7D)。此外, Pearson相关性分析显示关键基因与代谢通路的相关性, 50个标志性代谢通路中有9个参与代谢过程(附图7E)。

2.7 共定位分析和孟德尔随机化分析 基于前述差异表达与WGCNA分析所确定的关键基因, 将其作为暴露因素进行孟德尔随机化分析, 利用TwoSampleMR包识别与之相关的单核苷酸多态性

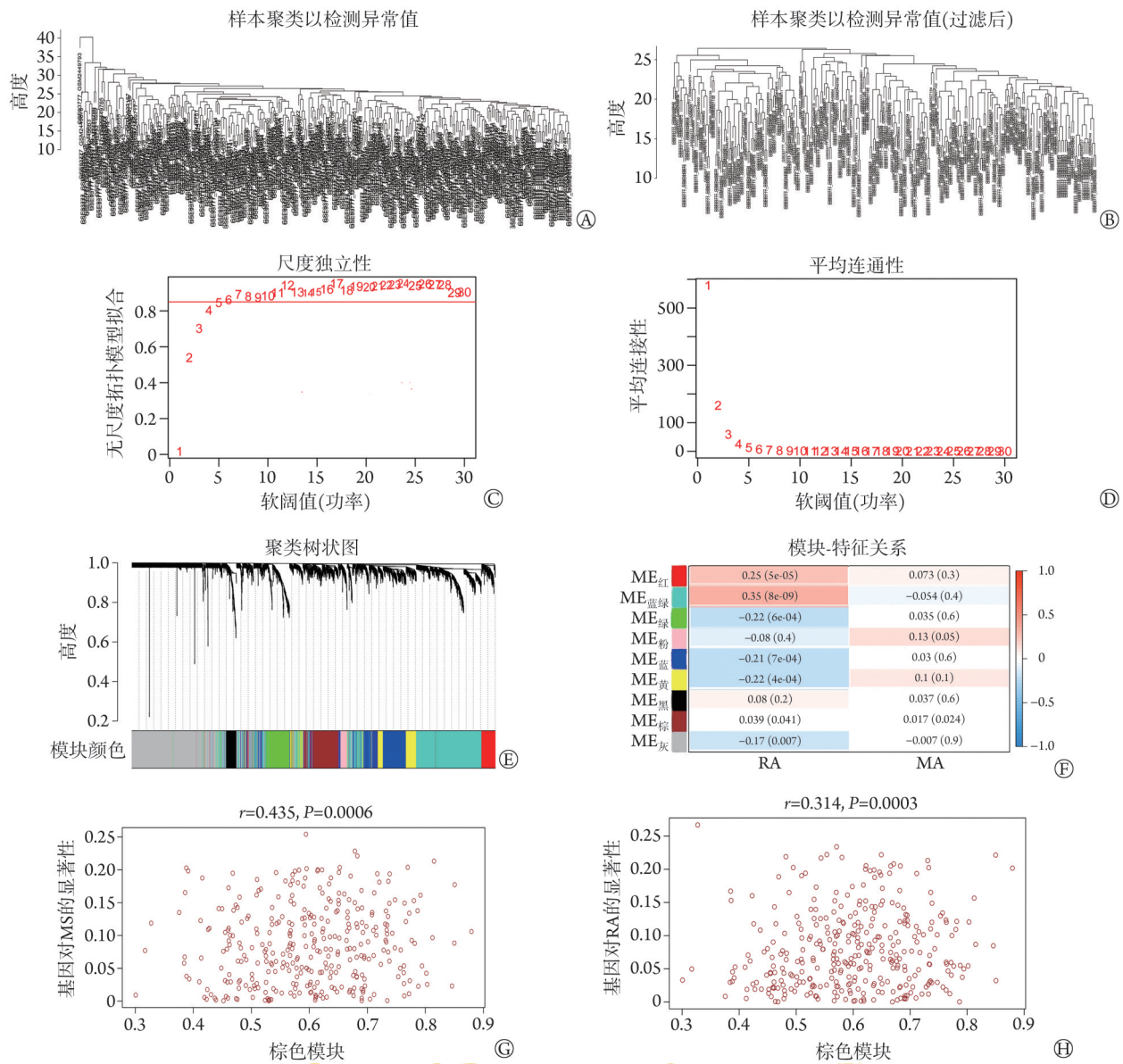


图2 RA与MS的加权基因共表达网络分析(WGCNA)

Fig.2 Weighted gene co-expression network analysis (WGCNA) of RA and MS

A. 基于全部基因表达数据的样本聚类树状图,若样本的分支与其他样本分支的连接高度明显较高则说明有离群值[红色线代表健康组被错误地诊断为RA的假阳性率,黑色线代表所有样本(健康+RA)的平均错分概率,绿色线代表将RA患者错分为健康的误差率]; B. 基于去批次后样本表达水平的聚类,每个分支代表数据集中的一个样本,若样本的分支与其他样本分支的连接高度相近,则没有离群值; C. 软阈值筛选分析图,选择拟合指数 $R^2 \geq 0.9$ 的最小软阈值(图中红色横线处, $R^2=0.9$ 时,软阈值为5); D. 最佳软阈值下的平均连通度分布图,软阈值为5时,后续基因间的平均连接数分布平稳,符合无标度网络构建要求; E. 聚类树下显示的起源模块及合并模块,不同颜色对应不同基因模块; F. 模块特征基因与RA和MS的相关性热图; G. 模块特征基因与MS发病率之间的相关图; H. 模块特征基因与RA发病率之间的相关图。RA. 类风湿关节炎; MS. 代谢综合征; ME. 基因模块

(SNPs)。设定 $P=0.00000005$ 进行筛选,并通过连锁不平衡的校正,同时排除了 F 值 <10 的SNPs,最终筛选出59个SNPs。这些SNPs被纳入后续研究以分析其与RA的关联。孟德尔随机化中逆方差加权法(IVW)分析显示,基因*STAT3*遗传特征与RA的发病风险增加存在正向因果关系($OR=1.09$, $95\%CI$ 1.02~1.18, $P=0.0149$)(图4A)。此外,采用共定位分析探究遗传位点对RA的联合效应,通过TwoSampleMR

包筛选与关键基因相关的SNPs,并应用严格的统计阈值排除潜在的连锁不平衡及低效的遗传标记,结果揭示了若干SNPs与RA风险的显著关联(图4B)。

2.8 单细胞数据集验证关键基因表达 对GSE235508数据集的分析共纳入11 178个细胞,经降维与聚类后,识别出16个不同的细胞簇(图5A—C)。细胞类型注释将其归类为T细胞、单核细胞、NK细胞、B细胞及血小板5大群体(图5D)。与健康对照组相

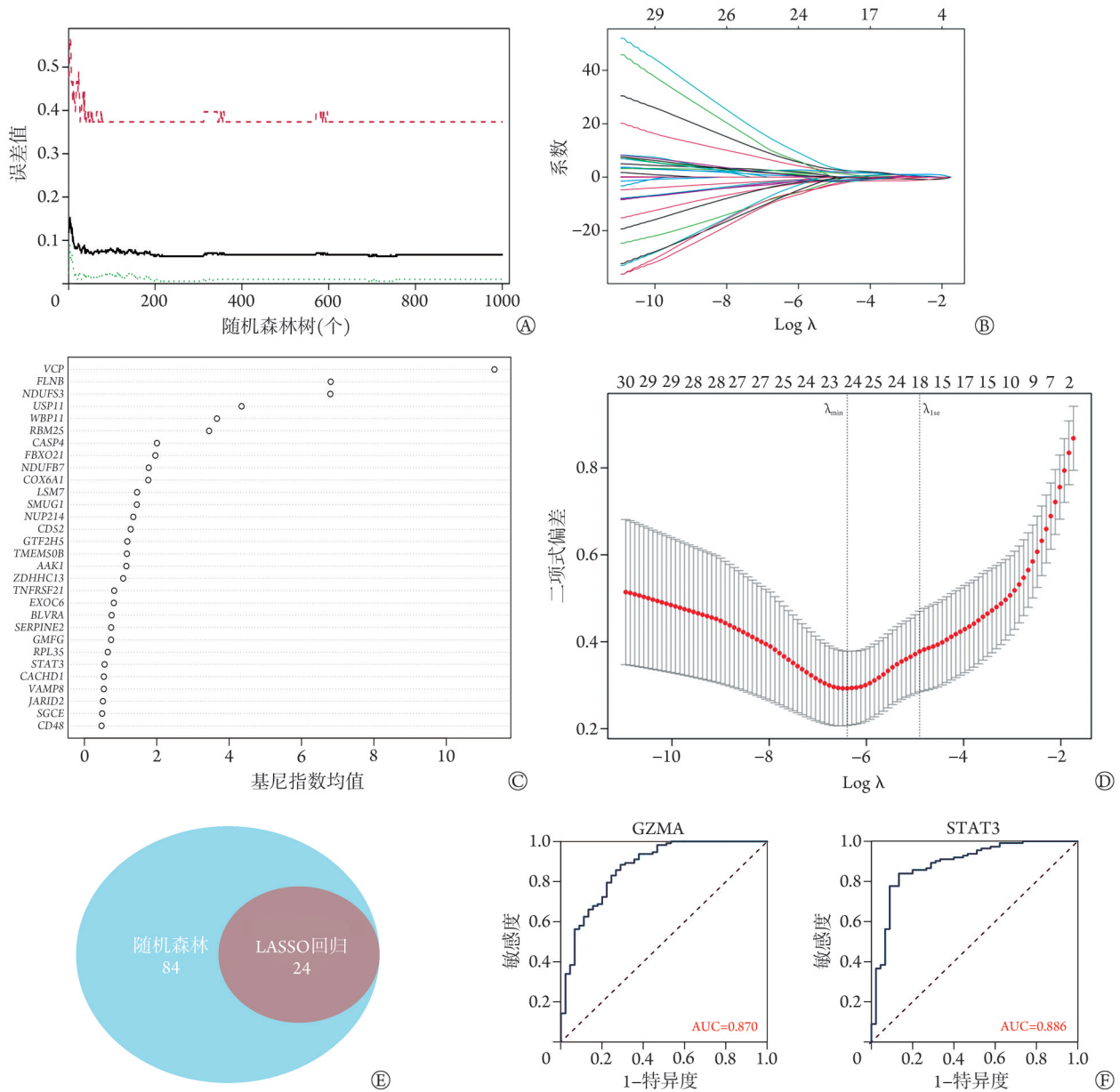


图3 随机森林及LASSO筛选RA与MS的关键基因
 Fig.3 Random forest and LASSO screening for hub genes of RA and MS

A. 随机森林误差率变化图，曲线趋于平稳时对应的树为模型需要的最少树的数量[红色线代表健康组的假阳性率，黑色线代表所有样本(健康+RA)的平均错分概率，绿色线代表将RA患者错分为健康的误差率]；B. 在选择LASSO时对调整参数进行十次交叉验证的结果，上图中的每条系数曲线代表一个基因；C. 从随机森林中识别出的前30个重要基因，MeanDecreaseGini(基尼指数均值)显示了基因的相对重要性排名；D. LASSO分析二项式偏差图，垂直实线代表部分似然偏差SE，左侧竖线(λ_{min})对应曲线最低点，表示平均交叉验证误差最小处 λ 值，右侧竖线(λ_{1se})表示最小误差+1个标准误差 λ 值，二者分别提示最优模型(保留基因数 $n=20$)与模型最简且误差可接受的 λ 值；E. 差异基因交集韦恩图；F. 关键基因GZMA、STAT3诊断RA的ROC曲线。RA. 类风湿性关节炎；MS. 代谢综合征；LASSO. 最小绝对收缩和选择算子算法

比，RA患者中源自T细胞、单核细胞和NK细胞的多个亚群比例发生明显变化($P < 0.05$)。对关键基因STAT3和GZMA的分析发现，二者在GSE17755和GSE93777数据集中的表达增加(图5E、F)，同时，基因表达热图提示STAT3和GZMA在NK细胞、T细胞中表达增加(图5G)。

2.9 非线性降维算法(uniform manifold approximation

and projection, UMAP)技术分析 UMAP技术可展示关键基因在细胞群体中的表达模式，结果显示，在T细胞和NK细胞中，STAT3和GZMA基因的表达水平升高。其中，图6A显示在T细胞中STAT3基因的表达水平明显上调(RA组青色点部分)，B细胞与T细胞中，GZMA基因表达明显上调(RA组黄色点部分)；图6B显示，在氧化应激条件下，RA患者中

结局	暴露	方法	SNP数量	β	P	OR(95%CI)
RA	JARID2	IVW	7	0.009851283	0.71527621	1.01(0.96~1.06)
RA	BLVRA	IVW	5	0.005393358	0.81592277	1.01(0.96~1.05)
RA	VAMP8	IVW	3	0.012325129	0.74606589	1.01(0.94~1.09)
RA	SMUG1	IVW	2	-0.052166001	0.04197986	0.95(0.90~1.00)
RA	NUP214	IVW	2	-0.085362475	0.22209339	0.92(0.80~1.05)
RA	STAT3	IVW	2	0.090092967	0.01499567	1.09(1.02~1.18)
RA	LSM7	IVW	3	0.039317288	0.12954225	1.04(0.99~1.09)
RA	SERPINE2	IVW	6	0.025375152	0.30707824	1.03(0.98~1.08)
RA	GZMA	IVW	3	0.027624899	0.04848555	1.03(0.95~1.11)
RA	FLNB	IVW	5	-0.085513049	0.13221250	0.92(0.82~1.03)
RA	TNFRSF21	IVW	7	0.083685654	0.09421383	1.09(0.99~1.20)
RA	SGCE	IVW	4	-0.112057212	0.03563062	0.89(0.81~0.99)
RA	CD52	IVW	4	-0.007457241	0.88032639	0.99(0.90~1.09)
RA	ZDHHC13	IVW	2	0.018109119	0.65152022	1.02(0.94~1.10)
RA	CASP4	IVW	6	0.064871922	0.27557544	1.07(0.95~1.20)

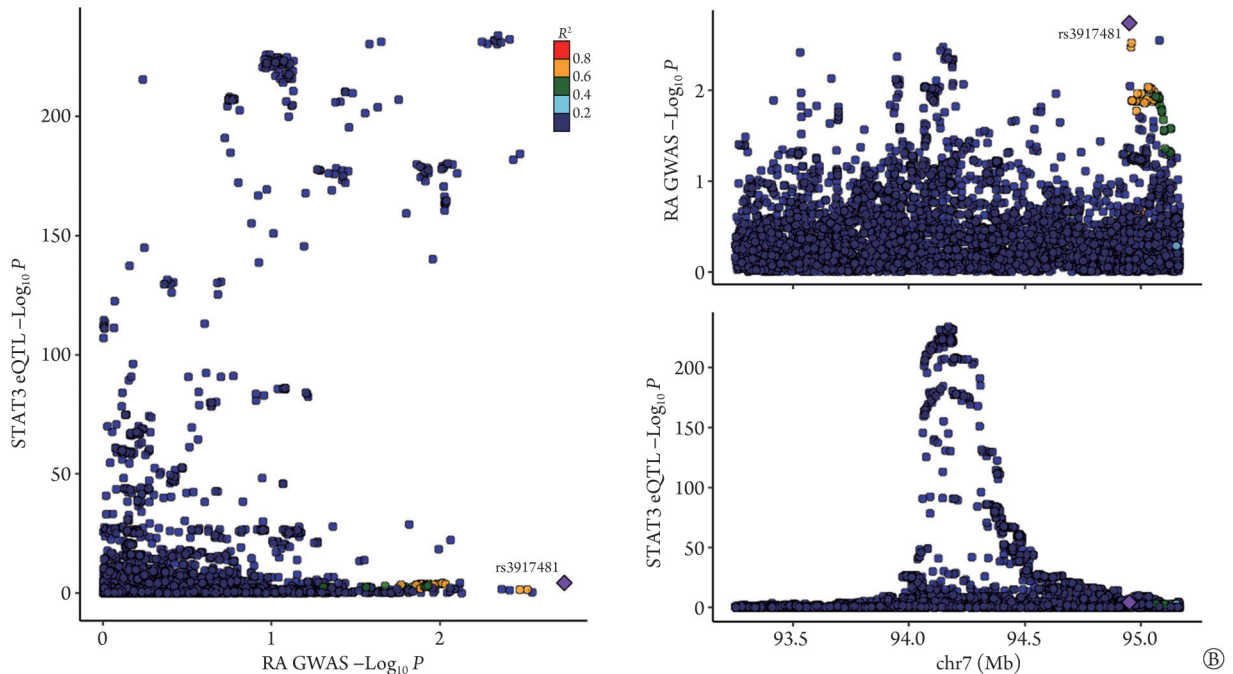


图4 类风湿性关节炎(RA)与RA关键基因的孟德尔随机化和共定位分析

Fig.4 Mendelian randomization and co-localization analysis of RA and key RA genes

A. 孟德尔随机化中差异基因的森林图(使用具有有效工具变量的基因进行分析,其余的被排除); B. 差异基因的共定位关联图分析(若点集中在对角线附近,表明两个SNPs与RA间高度共定位)。SNP. 单核苷酸多态性; IVW. 逆方差加权法; STAT3. 信号转导与转录激活因子3; eQTL. 表达数量性状基因座; GWAS. 全基因组关联研究; chr. 染色体

STAT3基因的表达呈上调趋势。

2.10 临床样本中关键基因的表达 各组的体重指数(BMI)、空腹血糖及炎症指标[类风湿因子(RF)、C-反应蛋白(CRP)、红细胞沉降率(ESR)]的总体差异均有统计学意义($P<0.001$)。进一步两两比较发现, Mets组及RA+MS组的BMI及空腹血糖均明显高于健康对照组及RA组($P<0.001$); RA组及RA+MS组的RF、CRP和ESR均明显高于健康对照组及MS组($P<0.01$); RA+MS组的上述炎症指标亦明显高于RA组($P<0.01$)。尽管RA+MS组与RA组在病程上差异无统计学意义($P=0.664$),但RA+MS组的疾病活动度

(DAS28评分)明显高于RA组($P=0.011$)(表1)。RT-qPCR检测结果显示,与健康对照组相比,MS组、RA组及RA+MS组的STAT3和GZMA基因表达水平均明显上调($P<0.05$)(附图8)。

3 讨论

近年来,测序技术的应用领域日益广泛,特别是生物信息学分析和机器学习的结合使研究者可以更深入地研究疾病的关键基因、诊断和预后指标、潜在的信号通路及治疗靶点,这些都为全面理解疾病提供了丰富的数据支持^[15-16]。RA与MS之间的联

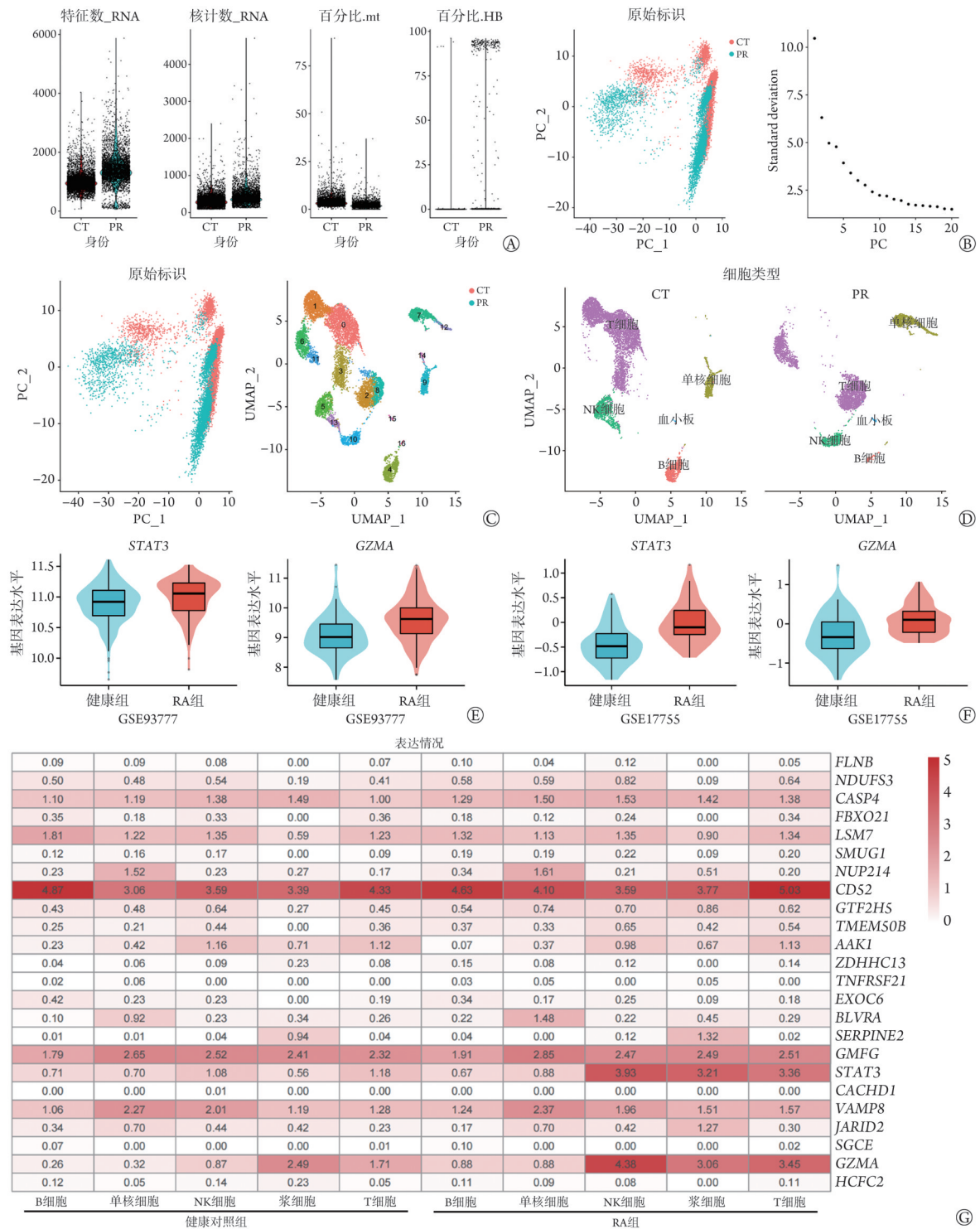


图5 单细胞数据集验证类风湿关节炎(RA)与代谢综合征(MS)共享关键基因的表达

Fig.5 Single-cell dataset validates expression of key genes of RA and MS

A. 单细胞数据质量控制(特征数_RNA显示了每个细胞中检测到的基因数量,显示是否有过高或过低的异常值;核计数_RNA显示了每个细胞中测序到的RNA分子总数,显示是否有过高或过低的异常值;百分比_mt显示了每个细胞中线粒体基因表达占总RNA表达的百分比,线粒体基因表达百分比过高提示细胞受压或死亡;百分比_HB显示了每个细胞中血红蛋白基因表达占总RNA表达的百分比,高血红蛋白基因表达百分比提示细胞质量较差); B. 单细胞数据降维聚类 and 变异性解释[主成分图(左)中,观察点的分布可了解两组细胞在空间中的聚类情况,以及它们之间的相似性或差异性;方差解释图(右)中,方差解释量的下降速度开始变缓的点之前的主成分通常被认为是重要的]; C. 非线性降维技术分析(UMAP_1和UMAP_2代表UMAP算法生成的两个降维坐标;不同颜色代表不同的细胞团); D. 非线性降维技术分析结果(不同颜色代表不同的细胞类型,图中标注了注释的细胞类型); E. GSE93777中STAT3和GZMA表达水平的小提琴图; F. GSE17755中STAT3和GZMA表达水平的小提琴图; G. 差异基因在免疫细胞中的表达水平(颜色越深,表达水平越高)。CT. 健康对照组单细胞数据集; PR. RA组单细胞数据集

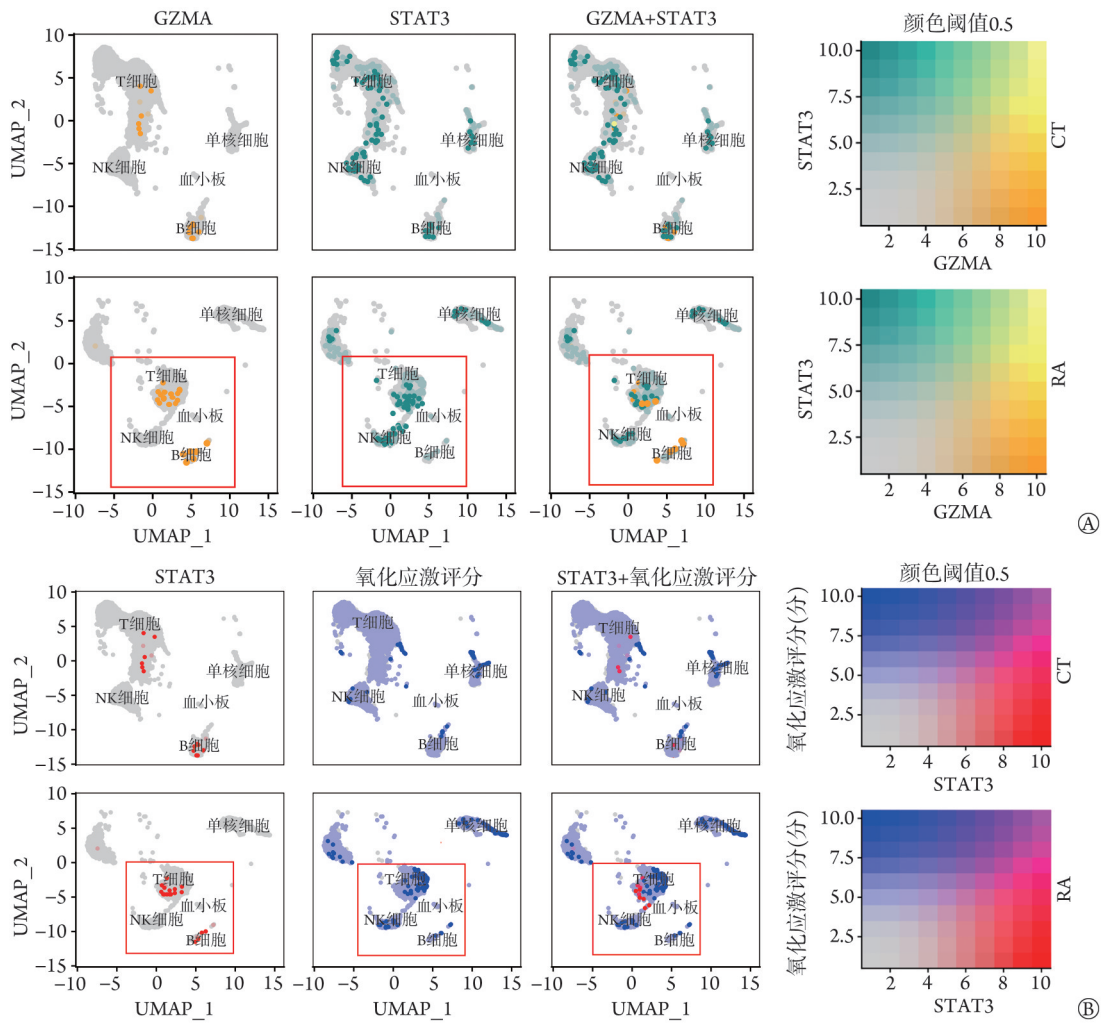


图6 类风湿关节炎(RA)与代谢综合征(MS)共享的关键基因表达的可视化

Fig.6 Visualization of expression of key gene from RA and MS

A. 对照(CT)组与RA组患者中GZMA、STAT3基因表达的UMAP可视化(黄色点示GZMA表达,青色点示STAT3表达);B. 对照组与RA患者中STAT3基因及氧化应激代谢通路评分的UMAP可视化(红色点示STAT3基因表达,蓝点示氧化应激)。GZMA. 颗粒酶A; STAT3. 信号转导与转录激活因子3; UMAP. 均匀流形近似与投影

表1 各组临床样本的基线资料比较

Tab.1 Baseline information on clinical samples

指标	健康对照组(n=6)	MS组(n=6)	RA组(n=6)	RA+MS组(n=6)	$\chi^2/F/Z/t$	P
性别[例(%)]					2.15	0.131
男	3(50.0)	3(50.0)	2(33.3)	3(50.0)		
女	3(50.0)	3(50.0)	4(66.6)	3(50.0)		
年龄(岁, $\bar{x}\pm s$)	55.8 \pm 5.2	58.1 \pm 8.5	54.5 \pm 5.3	60.5 \pm 3.5	1.18	0.340
体重指数(kg/m ² , $\bar{x}\pm s$)	23.5 \pm 1.3	34.9 \pm 1.9 ⁽²⁾	22.2 \pm 3.5 ⁽³⁾	33.0 \pm 0.9 ⁽²⁾⁽⁵⁾	52.05	<0.001
空腹血糖[mmol/L, M(Q ₁ , Q ₃)]	5.36(5.18, 5.47)	8.81(8.26, 11.32) ⁽²⁾	6.46(5.82, 6.71) ⁽¹⁾⁽³⁾	11.41(8.99, 12.67) ⁽²⁾⁽⁵⁾	18.46	<0.001
病程(年, $\bar{x}\pm s$) [*]	NA	NA	6.48 \pm 3.79	7.36 \pm 2.98	-0.45	0.664
DAS28($\bar{x}\pm s$) [*]	NA	NA	4.96 \pm 0.91	6.21 \pm 0.59	-3.12	0.011
RF(U/ml, $\bar{x}\pm s$)	9.00 \pm 3.74	10.66 \pm 3.26	172.00 \pm 103.85 ⁽²⁾⁽⁴⁾	293.33 \pm 110.57 ⁽²⁾⁽⁴⁾⁽⁶⁾	16.06	<0.001
CRP(mg/L, $\bar{x}\pm s$)	1.65 \pm 0.42	9.93 \pm 2.65 ⁽²⁾	30.50 \pm 10.03 ⁽²⁾⁽⁴⁾	55.16 \pm 8.95 ⁽²⁾⁽⁴⁾⁽⁶⁾	91.54	<0.001
ESR(mm/h, $\bar{x}\pm s$)	12.66 \pm 1.21	28 \pm 4.77 ⁽²⁾	52.33 \pm 10.85 ⁽²⁾⁽⁴⁾	74.66 \pm 8.04 ⁽²⁾⁽⁴⁾⁽⁶⁾	135.42	<0.001

与健康对照组比较, (1) P <0.05, (2) P <0.01; 与MS组比较, (3) P <0.05, (4) P <0.01; 与RA组比较: (5) P <0.05, (6) P <0.01; *病程与DAS28的统计比较采用独立样本 t 检验, 仅在RA组与RA+MS组间进行。RA. 类风湿关节炎; MS. 代谢综合征; DAS28. 关节疾病活动度评分; RF. 类风湿因子; CRP. C反应蛋白; ESR. 红细胞沉降率; NA. 无数据

系已引起广泛关注,有研究发现MS各组分与RA的临床表现密切相关^[9,17]。在分子层面,MS相关的脂肪因子及其引起的细胞代谢重编程对RA病程有显著影响^[5]。尽管如此,利用生物信息学方法分析RA与MS关联性的研究仍然相对有限。本研究通过多种生物信息学分析工具和机器学习技术,挖掘了MS与RA的致病基因,探究两者之间的联系,并揭示MS和RA的免疫细胞浸润特性,旨在为MS-RA的治疗提供新的思路和方法。

本研究的富集分析结果表明,MS与RA的病理机制主要涉及炎症反应的调控和免疫细胞的激活。研究发现,与加权基因共表达网络分析识别的关键模块相重叠的基因不仅参与中性粒细胞的脱颗粒过程,还在炎症反应调节及巨噬细胞活性氧和一氧化氮生成的控制中表现出显著富集。这些基因也在MS与RA共有的DEGs富集路径中占据重要地位。中性粒细胞作为慢性炎症性疾病进展的关键因子,其活化过程与MS及其相关疾病如2型糖尿病、非酒精性脂肪肝病和肥胖密切相关,这些疾病通常伴随脂肪组织异常累积,从而引发慢性低度炎症状态和免疫系统失衡,包括中性粒细胞的过度活化^[18]。研究发现,肥胖和2型糖尿病患者的中性粒细胞活化程度明显高于健康人群,检测外周血中性粒细胞弹性蛋白酶和髓过氧化物酶水平可评估该活化的程度^[19]。此外,通过测量NE的血清浓度^[20]或MPO的血浆浓度^[21]也可评估中性粒细胞的活化状态。进一步研究发现,肥胖患者的 α_1 -抗胰蛋白酶(A1AT)水平下降,且这种下降与体重指数呈负相关^[22]。在RA的病理过程中,中性粒细胞在血液和关节滑膜液中产生的活性氧水平明显升高^[23-24]。活性氧水平升高可在血液中引起免疫球蛋白G(IgG)的结构变化,与免疫原性的增强和类风湿因子免疫复合物的形成密切相关^[25]。本研究揭示了MS中代谢基因的紊乱现象,特别是与能量代谢相关的途径。这种代谢异常同样在RA中有所体现,暗示了两者之间可能共享的代谢机制;此外,免疫细胞的异常分化过程可能加剧了RA的病理进展,为探讨MS和RA的共同病理基础提供了潜在线索。

本研究发现,RA和MS患者STAT3和GZMA的表达水平明显升高,可能为未来针对这两种疾病的治疗提供新的方向。STAT3最初于20世纪90年代被发现,是急性期反应因子(APRF)复合体的关键成分之一,该复合体由白细胞介素-6(IL-6)激活^[26-28]。STAT3的激活是一个多因素过程,涉及多种细胞因子和生长因子,这些因子通过与IL-6信号转导共受体-糖蛋白130(gp130)结合(如IL-6、抑瘤素M、IL-11)或与同型二聚体细胞因子受体[如粒细胞集落

刺激因子(G-CSF)]结合发挥作用^[29]。此外,STAT3还参与由细胞内信号蛋白如活化的小G蛋白(Ras)或酪氨酸激酶癌蛋白(如Src家族成员)^[30-32]引发的信号级联反应,这些反应在细胞生理和病理过程中发挥着关键作用。小鼠模型中的T细胞特异性STAT3基因敲除实验显示,STAT3在IL-6介导的T细胞存活过程中起关键作用,这一作用独立于B细胞淋巴瘤-2蛋白(Bcl-2)的调控机制^[33-34]。在T细胞受体(TCR)激活后,IL-6、转化生长因子- β (TGF- β)及其分泌的IL-21信号级联可促进辅助性T细胞17(Th17)的生成^[35-37]。STAT3对IL-6和IL-21信号的响应对Th17细胞的形成发挥重要作用。在分子层面,STAT3能够促进Th17细胞系特化转录因子-视黄酸受体相关孤儿受体 γ t和 α (ROR γ t和ROR α)的表达,在Th17细胞的发育中发挥决定性作用^[38]。此外,STAT3还参与IL-23受体(IL-23R)和IL-17的表达上调^[39]。在IL-23的协同作用下,IL-23R信号可进一步增强Th17细胞的发育,而Th17细胞分泌的IL-17主要发挥其效应功能^[40-42]。IL-17可通过促进抗微生物肽、趋化因子和粒细胞生成细胞因子等免疫效应物的产生,介导Th17细胞的免疫作用^[43]。因此,STAT3不仅通过直接影响粒细胞的生成,还通过调节Th17系在中性粒细胞驱动的炎症反应中发挥作用。GZM蛋白酶是一类由细胞毒性T淋巴细胞(CTL)和自然杀伤(NK)细胞分泌的酶^[44]。GZMA作为一种促炎介质,在RA^[45-46]、银屑病(psoriasis)^[47]和骨关节炎(osteoarthritis)^[48]等多种炎症性疾病中发挥重要作用。近期有研究指出,在RA患者的关节滑液和滑膜组织中,GZM蛋白酶水平升高,可能与炎症反应和关节损伤有关^[49]。GZMA可通过多种促炎机制发挥作用,其中包括独立穿透目标细胞并激活白细胞介素-1 β (IL-1 β)的裂解,作为炎症介质来发挥作用^[50]。Wensink等^[51]发现,将GZMA与Toll样受体2(TLR2)和TLR4的激动剂共同作用于单核细胞时,可明显促进肿瘤坏死因子- α (TNF- α)、IL-6和IL-8等促炎因子的释放。此外,GZMA在细胞外通过裂解尿激酶及蛋白酶激活受体-1(PAR-1)和PAR-2,进一步促进炎症反应^[52]。本研究也发现,在RA和MS患者中,GZMA的表达水平呈上升趋势,提示GZMA可能在这两种疾病的发病过程中起重要作用,因此进一步探索GZMA在RA和MS中的具体机制可能有助于发现新的治疗靶点。

GSE17755和GSE93777数据集主要包含RA患者的样本,但未明确排除MS患者的可能性。因此,本研究不能排除患者疾病表型混杂对结果的影响,这可能对相关基因作为诊断工具的特异性和可靠性造成一定的干扰。鉴于此,建议未来的研究应更加细

致地审查患者的纳入标准,并可能需要对RA和MS患者进行区分研究,区分后将有助于更准确地评估发现的基因表达模式在RA与MS诊断中的作用。尽管本研究结果可能受到混杂因素的影响,但基因表达模式在RA和MS患者中仍具有重要的生物学意义。此外,本研究构建了一个基于XGBoost的预后模型,旨在预测患者在未来某一特定时间内发生特定健康事件的风险。预后模型的构建对于临床决策至关重要,可以帮助医师预测患者的长期健康结果,并据此制订个性化的治疗计划。与诊断模型相比,预后模型能够帮助医疗专业人员评估治疗干预的长期效果,并为患者提供更精确的预后信息。尽管本研究的预后模型显示出良好的预测性能,但仍存在一定的局限性:(1)本研究的模型是基于单一中心的数据构建的,可能限制其在不同人群和环境中的泛化能力;(2)样本量有限,可能影响模型的稳定性和可靠性。未来需要在更大的多中心样本中验证本研究的预后模型,并探索其在不同种族人群和不同区域的适用性。此外,笔者计划进一步研究预后模型在不同临床场景下的应用,包括在患者教育、疾病管理和健康政策制定中的作用。通过不断的优化和验证,基于XGBoost的预后模型将成为临床实践中一个有价值的工具,可帮助改善患者的预后及生活质量。

尽管生物信息学在疾病诊断标志物的识别中扮演着重要角色,样本间存在的个体差异性却可能引起诊断过程中的误判,表现为假阳性或假阴性的结果,进而影响生物标志物的敏感度和特异度。因此,未来的研究需要制定更为严格的标准来降低样本间的异质性,从而提高诊断标志物的准确性和可信度。需要注意的是,关键基因的临床转化仍面临诸多挑战,包括不同人群的遗传异质性、基因表达的动态变化及环境因素的影响等,因此,未来的研究应着重于大规模多中心验证研究,并探索其与现有治疗手段的协同作用。

【附加材料】

附表1及附图1—8见<https://dx.doi.org/10.11855/j.issn.0577-7402.1314.2025.1022F>。

请扫描右方二维码获取。



【参考文献】

- Gravallese EM, Firestein GS, Koscal N, *et al.* What is rheumatoid arthritis? [J]. *N Engl J Med*, 2024, 390(13): e32.
- 尤君怡,梁国强,宋秀道. TRIM31与慢性肝病的孟德尔随机化及生物信息学分析[J]. *医学新知*, 2025, 35(3): 303-311.
- Gravallese EM, Firestein GS. Rheumatoid arthritis-common origins, divergent mechanisms[J]. *N Engl J Med*, 2023, 388(6): 529-542.
- Fahed G, Aoun L, Bou Zerdan M, *et al.* Metabolic syndrome: updates on pathophysiology and management in 2021[J]. *Int J Mol Sci*, 2022, 23(2): 786.
- Kerekes G, Nurmohamed MT, González-Gay MA, *et al.* Rheumatoid arthritis and metabolic syndrome[J]. *Nat Rev Rheumatol*, 2014, 10(11): 691-696.
- Grzechnik K, Targońska-Stepniak B. Metabolic syndrome and rheumatoid arthritis activity: an analysis of clinical, laboratory, and ultrasound parameters[J]. *Nutrients*, 2023, 15(22): 4756.
- Karvounaris SA, Sidiropoulos PI, Papadakis JA, *et al.* Metabolic syndrome is common among middle-to-older aged Mediterranean patients with rheumatoid arthritis and correlates with disease activity: a retrospective, cross-sectional, controlled, study[J]. *Ann Rheum Dis*, 2007, 66(1): 28-33.
- Zmora N, Bashardes S, Levy M, *et al.* The role of the immune system in metabolic health and disease[J]. *Cell Metab*, 2017, 25(3): 506-521.
- Medina G, Vera-Lastra O, Peralta-Amaro AL, *et al.* Metabolic syndrome, autoimmunity and rheumatic diseases[J]. *Pharmacol Res*, 2018, 133: 277-288.
- Francisco V, Ruiz-Fernández C, Pino J, *et al.* Adipokines: linking metabolic syndrome, the immune system, and arthritic diseases[J]. *Biochem Pharmacol*, 2019, 165: 196-206.
- Rhodes DR, Chinnaiyan AM. Integrative analysis of the cancer transcriptome[J]. *Nat Genet*, 2005, 37(S6): S31-S37.
- Sajda P. Machine learning for detection and diagnosis of disease[J]. *Annu Rev Biomed Eng*, 2006, 8: 537-565.
- 徐天波,刘德国,顾增辉,等. 衰老基因在骨肉瘤治疗、预后及肿瘤微环境中的作用研究[J]. *解放军医学杂志*, 2024, 49(5): 557-569.
- Aletaha D, Neogi T, Silman AJ, *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative[J]. *Ann Rheum Dis*, 2010, 69(9): 1580-1588.
- Zhu EY, Shu XR, Xu Z, *et al.* Screening of immune-related secretory proteins linking chronic kidney disease with calcific aortic valve disease based on comprehensive bioinformatics analysis and machine learning[J]. *J Transl Med*, 2023, 21(1): 359.
- Li JC, Wang GH, Xv X, *et al.* Identification of immune-associated genes in diagnosing osteoarthritis with metabolic syndrome by integrated bioinformatics analysis and machine learning[J]. *Front Immunol*, 2023, 14: 1134412.
- Luo PY, Xu WL, Ye D, *et al.* Metabolic syndrome is associated with an increased risk of rheumatoid arthritis: a prospective cohort study including 369, 065 participants[J]. *J Rheumatol*, 2024, 51(4): 360-367.
- Herrero-Cervera A, Soehnlein O, Kenne E. Neutrophils in chronic inflammatory diseases[J]. *Cell Mol Immunol*, 2022, 19(2): 177-191.
- Ali M, Jasmin S, Fariduddin M, *et al.* Neutrophil elastase and myeloperoxidase mRNA expression in overweight and obese subjects[J]. *Mol Biol Rep*, 2018, 45(5): 1245-1252.
- Nijhuis J, Rensen SS, Slaats Y, *et al.* Neutrophil activation in morbid obesity, chronic activation of acute inflammation[J]. *Obesity (Silver Spring)*, 2009, 17(11): 2014-2018.
- Rayman MP, Stranges S. Epidemiology of selenium and type 2 diabetes: can we make sense of it [J]. *Free Radic Biol Med*, 2013, 65: 1557-1564.

- [22] Mansuy-Aubert V, Zhou QL, Xie XY, *et al.* Imbalance between neutrophil elastase and its inhibitor α 1-antitrypsin in obesity alters insulin sensitivity, inflammation, and energy expenditure[J]. *Cell Metab*, 2013, 17(4): 534-548.
- [23] Kundu S, Ghosh P, Datta S, *et al.* Oxidative stress as a potential biomarker for determining disease activity in patients with rheumatoid arthritis[J]. *Free Radic Res*, 2012, 46(12): 1482-1489.
- [24] El Benna J, Hayem G, Dang PM, *et al.* NADPH oxidase priming and p47phox phosphorylation in neutrophils from synovial fluid of patients with rheumatoid arthritis and spondylarthropathy[J]. *Inflammation*, 2002, 26(6): 273-278.
- [25] Rasheed Z. Hydroxyl radical damaged immunoglobulin G in patients with rheumatoid arthritis: biochemical and immunological studies[J]. *Clin Biochem*, 2008, 41(9): 663-669.
- [26] Lütticken C, Wegenka UM, Yuan J, *et al.* Association of transcription factor APRF and protein kinase Jak1 with the interleukin-6 signal transducer gp130[J]. *Science*, 1994, 263(5143): 89-92.
- [27] Standke GJ, Meier VS, Groner B. Mammary gland factor activated by prolactin on mammary epithelial cells and acute-phase response factor activated by interleukin-6 in liver cells share DNA binding and transactivation potential[J]. *Mol Endocrinol*, 1994, 8(4): 469-477.
- [28] Zhong Z, Wen Z, Darnell JE Jr. Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6[J]. *Science*, 1994, 264(5155): 95-98.
- [29] Heim MH, Gamboni G, Beglinger C, *et al.* Specific activation of AP-1 but not Stat3 in regenerating liver in mice[J]. *Eur J Clin Invest*, 1997, 27(11): 948-955.
- [30] Yu CL, Meyer DJ, Campbell GS, *et al.* Enhanced DNA-binding activity of a Stat3-related protein in cells transformed by the Src oncoprotein[J]. *Science*, 1995, 269(5220): 81-83.
- [31] Ram PT, Horvath CM, Iyengar R. Stat3-mediated transformation of NIH-3T3 cells by the constitutively active Q205L Galphao protein [J]. *Science*, 2000, 287(5450): 142-144.
- [32] Zhang Y, Turkson J, Carter-Su C, *et al.* Activation of Stat3 in v-Src-transformed fibroblasts requires cooperation of Jak1 kinase activity [J]. *J Biol Chem*, 2000, 275(32): 24935-24944.
- [33] Takeda K, Kaisho T, Yoshida N, *et al.* Correction: Stat3 activation is responsible for IL-6-dependent T cell proliferation through preventing apoptosis: generation and characterization of T cell-specific Stat3-deficient mice[J]. *J Immunol*, 2015, 194(7): 3526.
- [34] Liu X, Lee YS, Yu CR, *et al.* Loss of STAT3 in CD4⁺ T cells prevents development of experimental autoimmune diseases[J]. *J Immunol*, 2008, 180(9): 6070-6076.
- [35] Laurence A, Tato CM, Davidson TS, *et al.* Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation[J]. *Immunity*, 2007, 26(3): 371-381.
- [36] Houston S. STAT3 and autoimmunity[J]. *Nat Immunol*, 2023, 24(1): 1.
- [37] Zhou L, Ivanov II, Spolski R, *et al.* IL-6 programs TH-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways[J]. *Nat Immunol*, 2007, 8(9): 967-974.
- [38] Qin Z, Wang R, Hou P, *et al.* TCR signaling induces STAT3 phosphorylation to promote TH17 cell differentiation[J]. *J Exp Med*, 2024, 221(3): e20230683.
- [39] Paradowska-Gorycka A, Wajda A, Romanowska-Próchnicka K, *et al.* Th17/treg-related transcriptional factor expression and cytokine profile in patients with rheumatoid arthritis[J]. *Front Immunol*, 2020, 11: 572858.
- [40] Damasceno LEA, Prado DS, Veras FP, *et al.* PKM2 promotes Th17 cell differentiation and autoimmune inflammation by fine-tuning STAT3 activation[J]. *J Exp Med*, 2020, 217(10): e20190613.
- [41] De Simone V, Franzè E, Ronchetti G, *et al.* Th17-type cytokines, IL-6 and TNF- α synergistically activate STAT3 and NF- κ B to promote colorectal cancer cell growth[J]. *Oncogene*, 2015, 34(27): 3493-3503.
- [42] Lee SH, Kim EK, Kwon JE, *et al.* Ssu72 attenuates autoimmune arthritis via targeting of STAT3 signaling and Th17 activation[J]. *Sci Rep*, 2017, 7(1): 5506.
- [43] Li X, Bechara R, Zhao J, *et al.* IL-17 receptor-based signaling and implications for disease[J]. *Nat Immunol*, 2019, 20(12): 1594-1602.
- [44] Chowdhury D, Lieberman J. Death by a thousand cuts: granzyme pathways of programmed cell death[J]. *Annu Rev Immunol*, 2008, 26: 389-420.
- [45] Santiago L, Mena C, Arias M, *et al.* Granzyme A contributes to inflammatory arthritis in mice through stimulation of osteoclastogenesis[J]. *Arthritis Rheumatol*, 2017, 69(2): 320-334.
- [46] Chemin K, Gerstner C, Malmström V. Effector functions of CD4⁺ T cells at the site of local autoimmune inflammation—lessons from rheumatoid arthritis[J]. *Front Immunol*, 2019, 10: 353.
- [47] Cheuk S, Martini E, Bergh K, *et al.* Granzyme A potentiates chemokine production in IL-17-stimulated keratinocytes[J]. *Exp Dermatol*, 2017, 26(9): 824-827.
- [48] Jaime P, García-Guerrero N, Estella R, *et al.* CD56⁺/CD16⁻ Natural Killer cells expressing the inflammatory protease granzyme A are enriched in synovial fluid from patients with osteoarthritis[J]. *Osteoarthr Cartil*, 2017, 25(10): 1708-1718.
- [49] Zheng Y, Zhao J, Shan Y, *et al.* Role of the granzyme family in rheumatoid arthritis: Current Insights and future perspectives[J]. *Front Immunol*, 2023, 14: 1137918.
- [50] Ma F, Wang X, Qiu Z, *et al.* NK-derived exosome miR-1249-3p inhibits Mycobacterium tuberculosis survival in macrophages by targeting SKOR1[J]. *Cytokine*, 2024, 175: 156481.
- [51] Wensink AC, Kok HM, Meeldijk J, *et al.* Granzymes A and K differentially potentiate LPS-induced cytokine response[J]. *Cell Death Discov*, 2016, 2: 16084.
- [52] Reches G, Piran R. Par2-mediated responses in inflammation and regeneration: choosing between repair and damage[J]. *Inflamm Regen*, 2024, 44(1): 26.

(责任编辑: 张小利)