论著·智能医学与预测模型

基于焦亡相关基因的骨髓炎诊断模型构建与线粒体-炎症互作机制解析:整合多组学与机器学习的生物信息学研究

张金烨[1],郑皓南[1],杨乾坤[2],余 斌[1] [1]南方医科大学南方医院创伤骨科,广州;[2]陆军军医大学(第三军医大学)第一附属医院骨科,重庆

[摘要] **目的** 金黄色葡萄球菌(*Staphylococcus aureus*,SA)诱导的骨髓炎(osteomyelitis,OM)作为骨科常见难治性感染,其早期诊断与免疫微环境解析面临严峻挑战。焦亡相关基因(pyroptosis-related genes,PRGs)的表达谱与SA-OM密切相关,这些基因可能通过线粒体相关通路影响免疫微环境,从而参与疾病进程,且特定基因组合可用于构建高精度诊断模型。本研究旨在整合多组学与机器学习,筛选SA-OM中关键的焦亡相关诊断生物标志物,构建高精度诊断模型,并解析其通过"线粒体-炎症"互作影响免疫微环境的分子机制。**方法** 基于GEO数据库收录的3个SA-OM数据集(GSE6269/GSE16129/GSE30119),共纳入143例SA-OM患者与79例健康对照样本,通过数据预处理(sva包校正批次效应)、差异表达分析(limma包筛选DE-PRGs,adj. P<0.05且|$\log_2$FC|>0.263)、共表达网络构建(WGCNA算法鉴定关键模块基因,softThreshold=5)、多组学交叉验证(Pearson相关性分析筛选MR-PRGs)、机器学习建模(SVM-RFE/LASSO/随机森林交叉验证筛选特征基因,n=9)及诊断模型构建(逻辑回归列线图模型,通过AUC、校准曲线slope与DCA评估效能),结合免疫微环境解析(CIBERSORT/ssGSEA定量分析22种免疫细胞浸润水平)。**结果** 发现23个DE-PRGs中8个关键基因构成的诊断模型在训练集(AUC=0.89,95%*CI*:0.83~0.95)与验证集(AUC=0.83,95%*CI*:0.76~0.90)均展现出优异性能,RT-qPCR实验进一步验证,在SA-OM组中,焦亡通路关键基因Caspase-1与IL-18的mRNA表达水平较对照组显著上调(P<0.05),与生物信息学结论相互印证。且*METTL3-MRPL39*轴显著富集于代谢通路与线粒体基因表达生物学过程,进一步揭示疾病组Th1/Th17细胞浸润水平较对照组升高3.2倍(P<0.001),*METTL3*表达与效应T细胞浸润呈正相关(r=0.65,P=0.008)。**结论** 研究系统阐明SA-OM中焦亡相关基因的调控网络,构建的诊断模型为早期筛查提供新工具,同时发现的线粒体-炎症互作机制及特异性免疫微环境特征为靶向治疗策略研发奠定理论基础。

[关键词] 骨髓炎;细胞焦亡;分子亚型;免疫浸润;生物标志物
[中图法分类号] R318.04;R515.9;R551.3 [文献标志码] A

Construction of a pyroptosis-related gene-based diagnostic model for osteomyelitis and analysis of the mitochondria-inflammation interaction mechanism: a bioinformatics study integrating multi-omics and machine learning

ZHANG Jinye[1], ZHENG Haonan[1], YANG Qiankun[2], YU Bin[1] [1]Department of Orthopedics and Traumatology, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong; [2]Department of Orthopedics, First Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] Objective *Staphylococcus aureus* (SA) -induced osteomyelitis (OM) is a common

[基金项目] 国家自然科学基金面上项目(82272517)
[通信作者] 余 斌,E-mail: yubin@smu.edu.cn

2097-0927/Copyright © 2026, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议 (https://creativecommons.org/licenses/by/4.0/)
引用格式:张金烨,郑皓南,杨乾坤,等. 基于焦亡相关基因的骨髓炎诊断模型构建与线粒体-炎症互作机制解析:整合多组学与机器学习的生物信息学研究[J]. 陆军军医大学学报,2026,48(07):914-927. DOI:10.16016/j.2097-0927.202602071.

Citation: ZHANG J Y, ZHENG H N, YANG Q K, *et al*. Construction of a pyroptosis-related gene-based diagnostic model for osteomyelitis and analysis of the mitochondria-inflammation interaction mechanism: a bioinformatics study integrating multi-omics and machine learning[J]. J Army Med Univ, 2026, 48(07): 914-927. DOI:10.16016/j.2097-0927.202602071.

refractory orthopedic infection, presenting substantial challenges in early diagnosis and immune microenvironment characterization. Expression profiles of pyroptosis-related genes (PRGs) are closely associated with SA-OM; these genes may influence the immune microenvironment through mitochondrial-related pathways, thereby participating in disease progression, and specific gene combinations can be utilized to construct high-precision diagnostic models. This study aims to integrate multi-omics and machine learning to screen key pyroptosis-related diagnostic biomarkers in SA-OM, construct a high-precision diagnostic model, and elucidate its molecular mechanism influencing the immune microenvironment through "mitochondria-inflammation" crosstalk. **Methods** Based on 3 SA-OM datasets (GSE6269/GSE16129/GSE30119) retrieved from the GEO database, a total of 143 SA-OM patients and 79 healthy control samples were enrolled. Data preprocessing (batch effect correction using the sva package), differential expression analysis (DE-PRGs screened *via* the limma package, adj. $P<0.05$ & $|\log_2FC|>0.263$), co-expression network construction (key module genes identified through WGCNA algorithm, softThreshold=5), multi-omics cross-validation (Pearson correlation analysis for MR-PRGs screening), machine learning modeling (feature genes selected via SVM-RFE/LASSO/random forest cross-validation, $n=9$), and diagnostic model construction (logistic regression nomogram model, efficacy evaluated through AUC, calibration curve slope, and DCA) were performed, combined with immune microenvironment analysis (CIBERSORT/ssGSEA quantitative analysis of 22 immune cell infiltration levels). **Results** Among 23 DE-PRGs, a diagnostic model comprising 8 key genes demonstrated excellent performance in both the training set (AUC=0.89, 95%*CI*: 0.83 to 0.95) and validation set (AUC=0.83, 95%*CI*: 0.76 to 0.90). RT-qPCR experiments further validated that the mRNA expression levels of the key pyroptosis pathway genes Caspase-1 and IL-18 in the SA-OM group were significantly upregulated compared with the control group ($P<0.05$), corroborating the bioinformatics findings. The *METTL3-MRPL39* axis was significantly enriched in "metabolic pathways" and "mitochondrial gene expression" biological processes. Furthermore, Th1/Th17 cell infiltration levels in the disease group were 3.2-fold higher than those in the control group ($P<0.001$), and *METTL3* expression exhibited positive correlation with effector T cell infiltration ($r=0.65$, $P=0.008$). **Conclusion** This study systematically elucidates the regulatory network of pyroptosis-related genes in SA-OM. The constructed diagnostic model provides a novel tool for early screening, while the identified mitochondrial-inflammation interplay mechanisms and specific immune microenvironment characteristics establish a theoretical foundation for the development of targeted therapeutic strategies.

[**Key words**] osteomyelitis; pyroptosis; molecular subtype; immune infiltration; biomarker

Supported by the General Program of National Natural Science Foundation of China (82272517). Corresponding author: YU Bin, E-mail: yubin@smu.edu.cn

骨髓炎(osteomyelitis, OM)是一种由微生物感染引起的骨组织炎症性疾病,在临床上构成了重大挑战[1]。其首要致病菌是金黄色葡萄球菌(*Staphylococcus aureus*,SA),该病菌在儿童及免疫功能低下个体中尤为常见[2]。骨髓炎的感染途径主要包括血源性播散、邻近感染灶的直接蔓延,以及创伤或外科手术导致的直接接种[3]。若诊断延迟或治疗不当,可能导致包括脓毒症、严重炎症反应、骨破坏、生长停滞,甚至多器官功能衰竭或死亡在内的严重并发症[4]。目前,临床管理骨髓炎的主要手段是抗生素治疗。然而,SA具备多种免疫逃逸机制,常导致疗程延长[5-7]。因此,通过筛选与验证关键生物标志物来实现金黄色葡萄球菌性骨髓炎的早期检测,对于显著改善患者预后至关重要。

细胞死亡是多细胞生物发育过程中的基本进程,其中坏死与凋亡是主要类型[8]。近期研究表明,特定基因调控着一些非凋亡性细胞死亡形式,例如细胞焦亡,其在形态学与生物化学特征上区别于凋亡和坏死[9-10],细胞焦亡的特征是细胞肿胀、质膜破裂以及促炎性细胞内含物的释放,继而引发强烈的免疫反应[11]。该过程主要由半胱氨酸天冬氨酸蛋白酶-1(cysteine aspartate-specific protease-1,Caspase-1)的激活所介导,该酶是称为"炎症小体"的多蛋白复合物的组分。此炎症小体可被多种刺激激活,包括病原体相关分子模式(pathogen-associated molecular patterns)和损伤相关分子模式(damage-associated molecular patterns)。激活后,Caspase-1切割消皮素D(gasdermin D,GSDMD)蛋白,导致细胞膜形成孔道,从而促进IL-1β和IL-18等促炎细胞因子的释放。这些细胞因子在募集并激活免疫细胞至感染或损伤部位的过程中起着关键作用[12]。同时,细菌感染可能诱发骨细胞线粒体功能障碍,导致氧化应激及后续的细胞死亡[13]尽管骨髓炎的病理生理过程涵盖免疫与炎症反应,但金黄色葡萄球菌诱导的骨髓炎中,与线粒体功能障碍相关的细胞焦亡机制仍未完全阐明。

本研究利用GEO数据库,并应用机器学习算法,以鉴定具有诊断潜力的最佳差异表达焦亡相关基因。采用受试者工作特征曲线评估了这些基因的诊断准确性,并使用独立数据集进行了验证。此外,本研究构建了列线图模型以预测金黄色葡萄球菌性骨髓炎的发病率。基于这些关键的差异表达焦亡相关基因。进一步深入解析了SA-OM样本相较于对照组的免疫微环境特征,探究其免疫细胞浸润模式的差异。

1 材料与方法

1.1 GEO数据库数据获取与预处理

从NCBI GEO[14](Gene Expression Omnibus,GEO,http://www.ncbi.nlm.nih.gov/geo/)数据库中下载数据编号为GSE6269的数据集,143个样本,其中患病信息包含OM的样本有21个,对照样本6个。测序平台有3个,GPL96［HG-U133A］Affymetrix Human Genome U133A Array,GPL570［HG-U133_Plus_2］Affymetrix Human Genome U133 Plus 2.0 Array和GPL2507 Sentrix Human-6 Expression BeadChip。直接下载经过处理后的探针表达矩阵,同时下载相应的平台注释文件对探针进行基因名转化,对于不同探针对应同一基因名,取平均值作为该基因的表达值。同时下载GSE16129数据集,126个样本,其中患病信息包含OM的样本有70个,对照样本29个。测序平台有3个,GPL96［HG-U133A］Affymetrix Human Genome U133A Array,GPL97［HG-U133B］Affymetrix Human Genome U133B Array和GPL6106 Sentrix Human-6 v2 Expression BeadChip,使用同样的方法进行表达矩阵处理,对2个数据集的表达矩阵使用sva包Version 3.52.2[15]（https://bioconductor.org/packages/release/bioc/html/sva.html)进行去批次合并处理,最终包含91例患病样本,35例正常对照样本,该数据集作为训练集。另外下载GSE30119数据集,共143例样本,包含38例OM患病样本,44例正常对照样本。测序平台为GPL6947 Illumina HumanHT-12 V3.0 expression beadchip,与训练集处理类似,直接下载经过处理后的表达矩阵,下载注释文件进行gene symbol转化。

1.2 差异表达焦亡相关基因筛选

基于上述处理合并训练集、数据集,并进行去批次处理,对去批次前和去批次后分别进行聚类可视化展示,以观察去批次的变化。之后在合并去批次的训练集中将样本按照性质分成患病和正常样本,利用R语言limma包Version 3.34.7[16](https://bioconductor.org/packages/release/bioc/html/limma.html)筛选患病与正常组间的显著差异基因,计算各基因的显著性P值和差异倍数,选取调整差异显著性adj. P<0.05且$|\log_2FC|$>0.263作为筛选阈值[17-18],此阈值对应于表达量变化约20%,是微阵列数据差异分析中常用的温和阈值,旨在筛选出具有生物学意义的差异表达基因,绘制热图和火山图展示。同时绘制PCA聚类图以展示患病与正常分组的聚类情况。

1.3 加权基因共表达网络构建

基于训练集的表达矩阵,利用R包WGCNA[19](Version 1.61,https://cran.r-project.org/web/packages/WGCNA/)对训练集表达矩阵基因进行分析,从而鉴定高度协同变化的基因集模块。

在WGCNA算法中,定义的基因共表达矩阵中的元素为基因的相关系数的加权值,权重的选择标准为使得每个基因网中包含的基因之间的连接服从无尺度网络分布(scale-free networks,SFN)。这里的加权值即为软阈值,首先通过设定一系列的幂值,分别计算各个幂值下连接度k与p(k)的相关系数的平方值及平均连接度,然后选取合适的幂值取值,使得网络中的基因之间的连接服从SFN特性,选择幂值=5的依据是:此时网络拓扑结构拟合度首次达到0.9以上,且平均连接度处于合理水平,符合SFN构建要求;其次基于聚类和动态剪枝法,设置参数(minModuleSize=30:每个模块至少包含30个基因;mergeCutHeight=0.25:合并模块时的相似系数为0.25)将高相关性基因聚集为模块;最后通过计算模块与表型相关性,这里的表型指的是OM患病与正常对照分组,选取P<0.05且与表型相关性最高的模块作为有意义的关键模块。

1.4 线粒体-焦亡基因交叉验证

对差异分析得到的DEGs和WGCNA分析得到的模块基因分别与数据库线粒体基因和焦亡基因进行交集识别,得到DEGs-WGCNA-MRG和DEGs-WGCNA-PRG,然后对DEGs-WGCNA-MRG和DEGs-WGCNA-PRG进行相关性分析,相关性算法为Pearson算法,以相关系数|R|>0.3且P<0.001为阈值[20-21]。识别OM中线粒体相关焦亡基因,记为共表达基因,为了降低报告的高错误发现率,使用OM和对照样本之间的Wilcoxon秩和检验进一步筛选这些共表达基因,观察其显著性是否为P<0.05。

1.5 富集分析

利用常用的R语言程序包clusterProfiler[22](Version 4.10.0,https://www.bioconductor.org/packages/3.10/bioc/html/clusterProfiler.html)对上述共表达基因分别进行GO[23](包括Biological Process、Cell Component、Molecular Function)以及KEGG pathway[24]富集分析,选择P<0.05的条目被认为是显著富集结果,之后选择部分显著条目进行展示。

1.6 PPI网络构建

结合STRING[25](Version:10.0,http://www.string-db.org/)数据库预测分析共表达基因编码蛋白之间是

否存在互作关系。输入基因集为上述得到的共表达基因,物种为homo。参数PPI score设为0.15。使用Cytoscape[26](Version:3.9.1)中的cytoHubba[27]插件中拓扑分析算法得到MCC、MNC、Degree和EPC等关键性指标,对每种指标选取排名前15的基因,4种指标得到的交集基因为候选基因。

1.7 3种机器学习算法筛选特征基因

在数据集中基于患病和对照分组采用3种机器学习算法进行分析,分别为R语言e1071包(版本:1.7-14,https://cran.r-project.org/web/packages/e1071/index.html)支持向量机递归特征消除机器学习算法(support vector machine recursive feature elimination,SVM-RFE)[28],R语言glmnet包(版本:4.1.3,https://cran.r-project.org/web/packages/glmnet/index.html)[29]的LASSO回归模型,R语言randomForest包(版本:4.7-1.1,https://cran.r-project.org/web/packages/randomForest/index.html)的随机森林算法,在使用SVM-RFE、LASSO和随机森林算法时,均采用了交叉验证策略来评估模型稳定性并防止过拟合。对这3种方法筛选得到的关键基因取交集得到交集基因用于训练集中进一步特征基因的筛选。

1.8 特征基因的表达与ROC验证

从训练集和验证集中提取候选基因的表达数据,并使用Wilcoxon秩和检验来计算特征基因在OM患病和对照样本间的差异显著性,利用R语言pROC[30](版本:1.12.1,https://cran.r-project.org/web/packages/pROC/index.html)中的ROC曲线方法分别在训练集和验证集中对各候选基因进行诊断评估,观察各基因的AUC值。选取训练集和验证集中差异都显著且上下调一致的基因作为特征基因。之后在训练集和验证集中对特征基因进行相关性分析。

1.9 免疫细胞与特征基因的相关性

考虑到免疫细胞在OM发展中的重要作用,通过免疫浸润算法评估免疫微环境。通过CIBERSORT[31-32]算法和ssGSEA算法[33]分别计算各样本微环境中各免疫细胞的值,CIBERSORT得到免疫细胞的浸润占比值,ssGSEA得到各免疫细胞的富集评分。比较患病和正常分组间免疫浸润值的差异,绘制热图与箱线图展示,以观察各免疫细胞的差异情况。使用Pearson算法对免疫细胞进行相关性分析,同时计算特征基因和各免疫细胞浸润值的相关性。

1.10 列线图

基于训练集样本验证成功的特征基因的表达量,利用R语言中R包rms Version 6.7-0(https://mran.microsoft.com/web/packages/rms/index.html)[34],基于各特征基因构建列线图,并进行校准曲线分析和决策曲线DCA以评估列线图诊断性能。基于列线图构建的诊断模型,分别计算训练集和验证集的得分,利用列线图得分绘制箱线图与ROC曲线以评估列线图诊断模型在训练集和验证集中的诊断性能。

1.11 GSEA分析

利用clusterProfiler包基于KEGG基因集进行GSEA富集分析,对于每种特征基因分别进行分析,在OM患病样本中按基因表达量将样本分为高低表达分组,设置阈值为$P<0.05$,展示最显著富集的10条通路以研究特征基因在OM患病样本中可能参与的功能。

1.12 金黄色葡萄球菌的复苏、定量和冻存

本研究使用的金黄色葡萄球菌菌株的复苏、培养、定量与保存均遵循标准微生物学操作规范。具体流程如下:从-80 ℃冰箱取出冻存菌种,37 ℃水浴快速复苏后,接种于胰蛋白酶大豆肉汤,于37 ℃、200×*g*震荡培养16~18 h至对数生长期;细菌经PBS洗涤后,用分光光度计调整菌悬液浓度至[*OD*(600)]=0.5(约1×10^8 CFU/mL)以用于后续实验;长期保存时,将菌液与无菌二甲基亚砜(终浓度15%)混合,置于-80 ℃冻存。所有操作均在生物安全柜内按生物安全二级(BSL-2)规范进行。

1.13 RT-qPCR分析

首先,使用TRIzol法从样本中提取总RNA,经质量检测合格后,将其逆转录为cDNA。针对目标基因设计特异性引物,在QuantStudio 5荧光定量PCR仪上进行扩增反应,反应体系包含SYBR Green Master Mix、引物及cDNA模板。通过构建质粒标准曲线评估扩增效率,并利用熔解曲线分析验证扩增特异性。以*GAPDH*作为内参基因,采用$2^{-\Delta\Delta Ct}$法计算目标基因的相对表达量,所有实验均设置技术重复及无模板对照以确保结果可靠性。

2 结果

2.1 焦亡相关基因显著差异表达及关键基因筛选

根据方法部分,对2个数据集的不同测序平台的结果进行合并去批次处理,去批次前,各平台测序结果的聚类分布图如图1A所示,为分别聚类,去批次后聚类分布图(图1B),各平台的结果都聚在一起,说明去批次结果较好。之后在患病和对照样本中差异表达分析图(1D、E),得到739个差异基因,其中198个上调基因,541个下调基因,对差异基因进行火山图展示(图1C)。绘制热图展示各基因在样本中的表达情况(图1B)所示。对患病样本和正常样本进行PCA聚类分析,2组样本都有各自明显的聚类特征。随后,对训练集表达矩阵进行WGCNA分析,选取软阈值为5(图2A、B);接着基于聚类和动态剪枝法,将高相关性基因聚集为模块,再对这些模块进行聚类,将相关系数>0.75,即相异系数<0.25的模块合并掉,最终共整合成

15个模块(图2C)所示。进一步,通过计算每个模块特征向量基因(特征向量基因为某一特定模块第一主成分基因,代表了该模块内基因表达的整体水平)与表型(这里指样本的OM患病与正常对照分组)之间的相关性,对各模块与表型信息作热图显示(图2D),可以看到,蓝色模块(213个基因;相关系数r=0.58且P<0.05)与患病样本呈最显著的正相关性,黄色模块(170个基因;相关系数r=−0.54且P<0.05)以及红色模块(139个基因;相关系数r=−0.31且P<0.05)与患病样本呈显著的负相关性,因此将这3个模块基因作为关键模块基因。

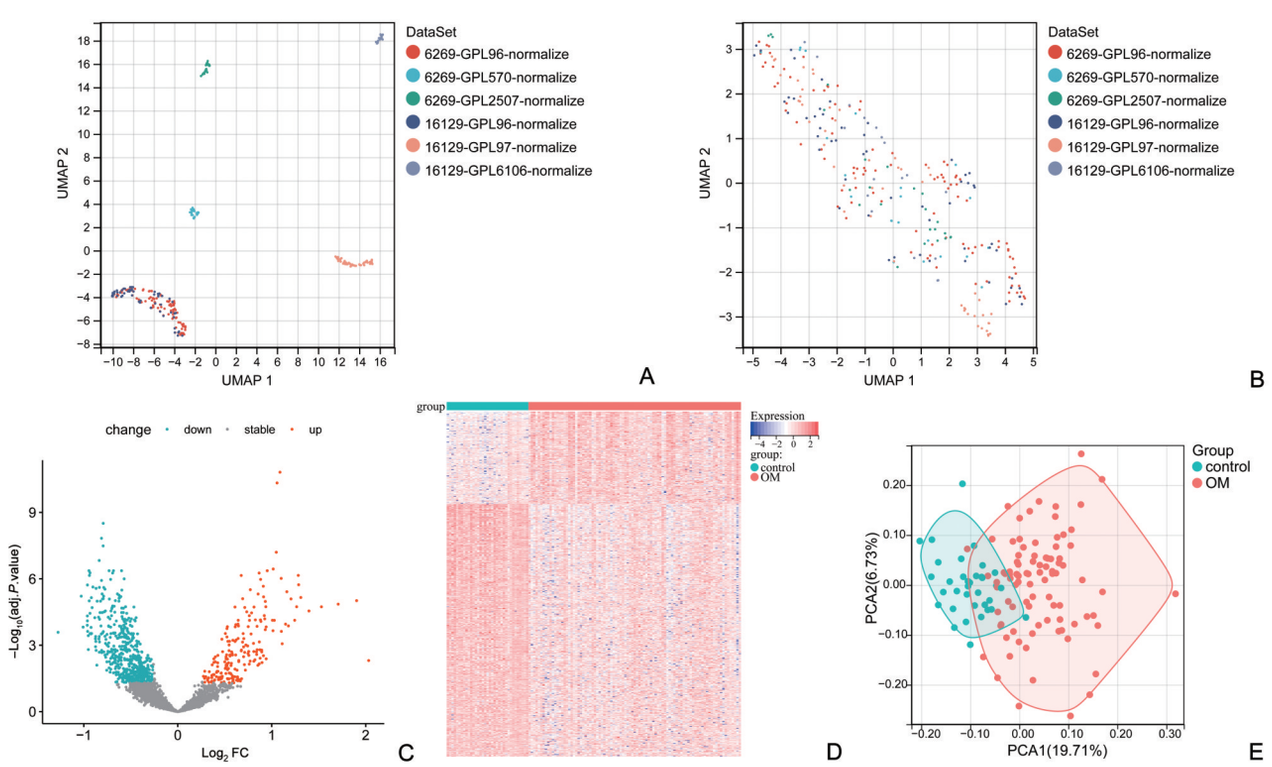
此外,通过RT-qPCR对焦亡通路中的2个关键执行分子进行验证,结果显示,与对照组相比,SA-OM组中炎症小体组分Caspase-1及下游效应细胞因子IL-18的mRNA表达水平均显著上调(P<0.05,图3)。该实验结果与前期生物信息学分析中发现的"焦亡相关基因(PRGs)在SA-OM中显著差异表达"这一结论相互印证,为后续构建以焦亡相关基因为基础的诊断模型提供了直接的实验证据支持。

2.2 核心调控模块鉴定与功能富集

按方法所述,对差异显著基因DEG,WGCNA关键模块基因和线粒体相关基因MRG取交集(图2E),得到23个交集基因,同时对DEG,WGCNA基因和焦亡相关基因MRG取交集(图2F)得到3个交集基因。对上述两组交集基因进行相关性分析,分析结果详见附表8,按照方法所述阈值进行筛选,得到显著相关基因21个,记为共表达基因,对共表达基因进行差异检验,所有共表达基因在SA-OM组与对照组间的表达差异均具有统计学意义(P<0.05,图2G)。对共表达基因进行KEGG信号通路和GO功能富集分析,共筛选得到了显著相关的4条KEGG信号通路和306个GO项,根据P值选取合适的显著富集项作图展示,选择全部显著的KEGG展示,GO选择BP,CC和MF各四条最显著的展示(图2H、I),KEGG通路富集分析显示,差异基因主要富集于代谢通和膦酸酯与次膦酸酯代谢。GO功能富集分析表明,差异基因在生物学过程中显著富集于线粒体基因表达在细胞组分中显著富集于线粒体,在分子功能中显著富集于RNA结合。

2.3 线粒体-焦亡共表达基因网络构建

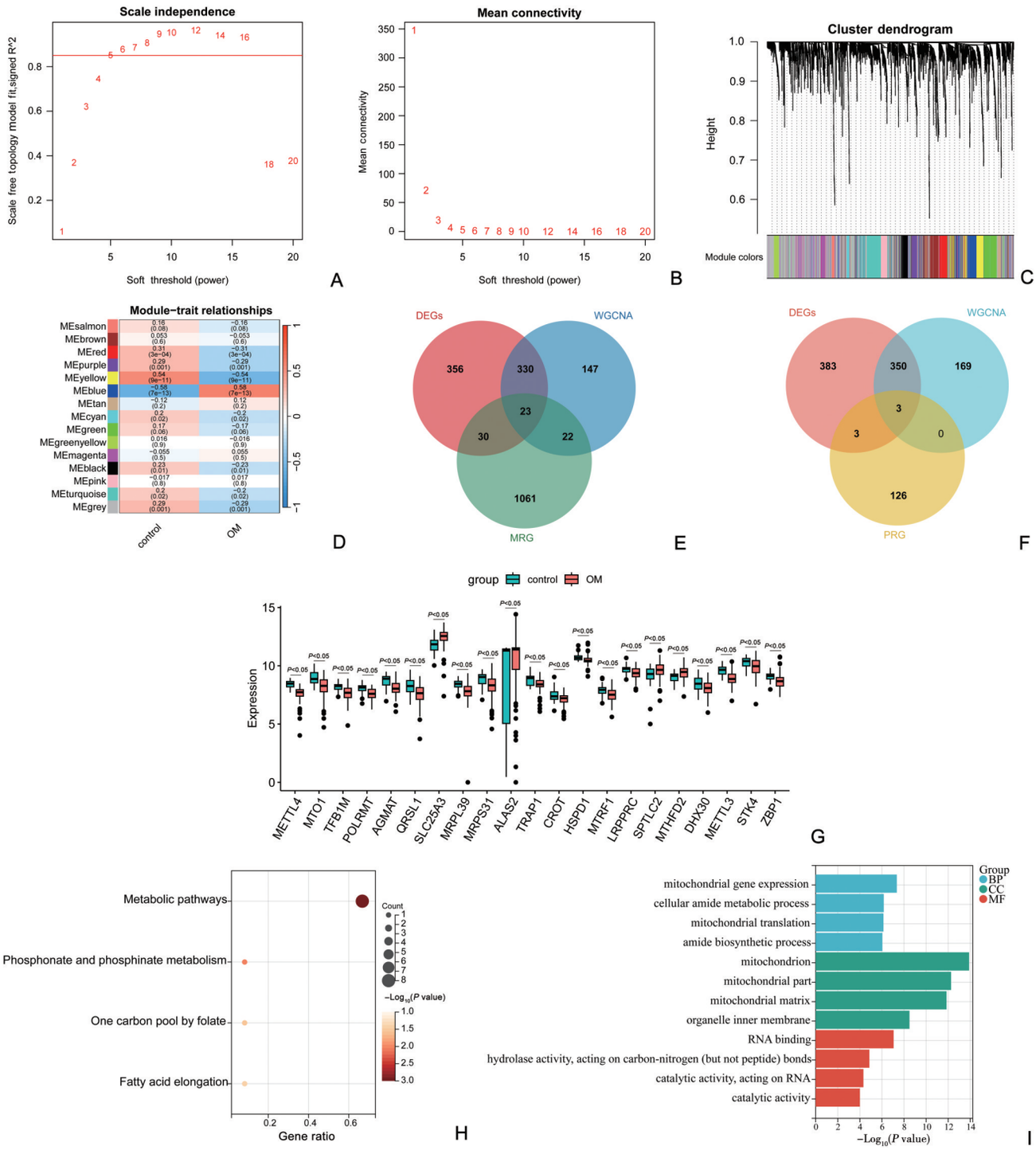
基于模型基因使用STRING数据库构建蛋白互作网络图,对互作网络各节点蛋白进行关键性分析(图



A: clustering distribution of the integrated dataset before batch effect removal; B: clustering distribution of the integrated dataset after batch effect removal; C: volcano plot of differentially expressed genes between disease and control samples (Red: upregulated genes, Blue: downregulated genes); D: heatmap of significantly differentially expressed genes; E: PCA clustering plot of disease and control sample

图1 训练数据集的批次效应校正、差异表达基因筛选及主成分分析

Figure 1 Batch effect correction, differential gene screening, and principal component analysis of the training dataset



A: WGCNA power value selection result; B: WGCNA power value selection result; C: module clustering dendrogram showing gene hierarchical clustering and dynamically identified modules, with colors representing distinct co-expression modules; D: correlation heatmap between WGCNA modules and the disease/control grouping; E: Venn diagram of intersections among differentially expressed genes (DEGs), key WGCNA module genes, and mitochondrial-related genes (MRGs); F: Venn diagram of intersections among differentially expressed genes (DEGs), key WGCNA module genes, and pyroptosis-related genes (PRGs); G: differential expression box plot of the shared genes; H:KEGG pathway enrichment analysis of the co-expressed genes;I: GO functional enrichment analysis of the co-expressed genes

图2 加权基因共表达网络分析、关键模块鉴定及交集基因功能富集

Figure 2 Weighted gene co-expression network analysis, key module identification, and functional enrichment of intersection genes

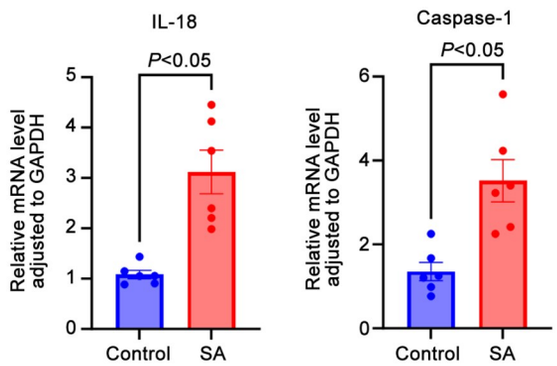


图3 IL-18与Caspase-1在对照与SA处理组的相对mRNA水平比较

Figure 3 Relative mRNA levels of IL-18 and Caspase-1 between the control and *Staphylococcus aureus* (SA) groups

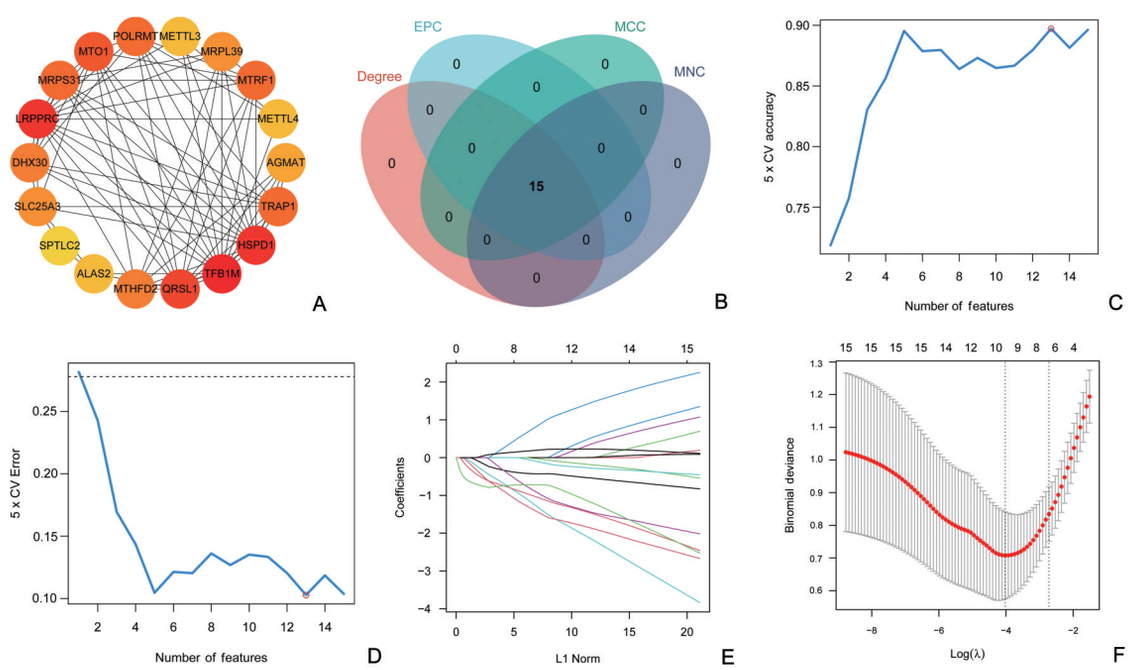
4A),颜色越红表示节点连线数越多,即在网络中越关键,上一步筛选的21个共表达基因中有18个基因在网络图中。按方法所述,对网络节点蛋白进行关键性拓扑分析,然后对MCC、MNC、Degree和EPC指标前15的节点进行取交集(图4B),15个基因全部为交集基因。

2.4 3种机器学习算法筛选特征基因

在数据集中基于患病和对照分组采用3种机器学习算法进行分析,分别为SVM-RFE、LASSO回归和随机森林算法。SVM-RFE分析结果显示,前13个基因被选取为最优组合(图4C、D)。LASSO回归10个基因被选取为最优组合(图4E、F)。随机森林算法结果显示,全部基因的重要性值都>1,故结果都被保留(图5A),随机森林中树数与错误率之间的关系(图5B)。对这3种方法筛选得到的关键基因取交集(图5C),依据方法得到8个交集基因,即*MRPL39*,*MTO1*,*POLRMT*,*AGMAT*,*METTL3*,*TFB1M*,*TRAP1*和*DHX30*。

2.5 诊断模型诊断效能评估

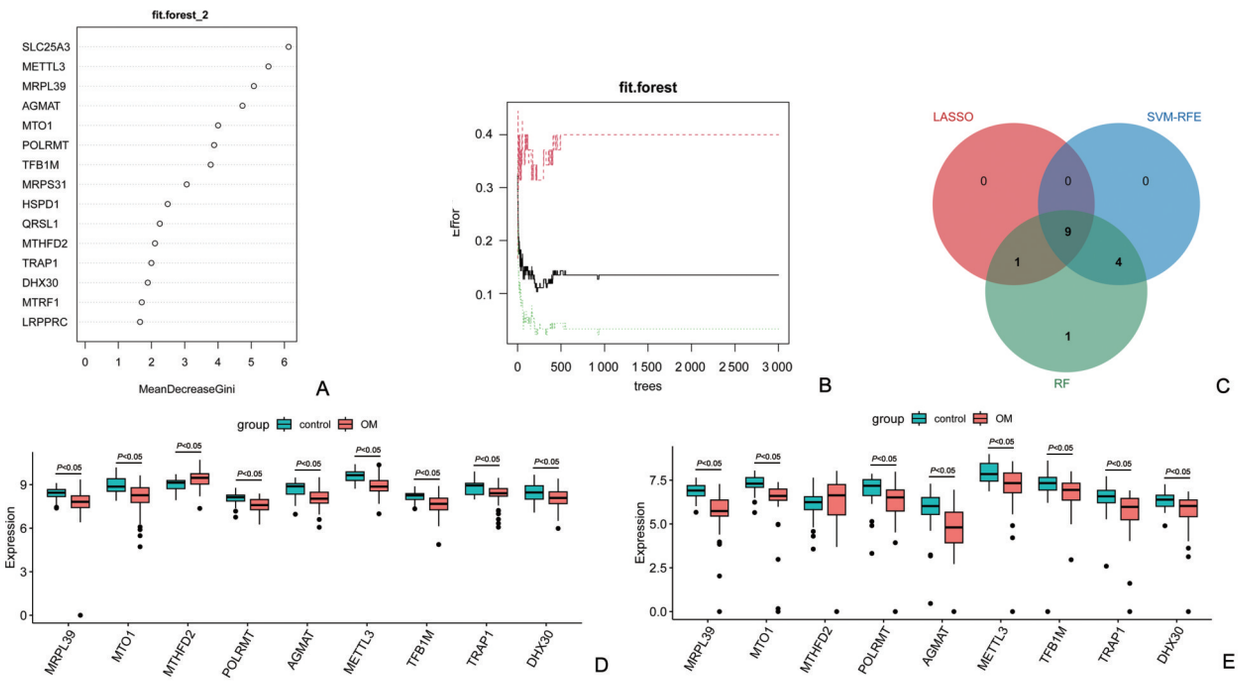
对经过3种机器学习筛选得到的交集基因进行验证,按方法所述,在训练集和验证集中提取交集基因表达量,在患病和对照组间进行差异分析,训练集中的差异结果(图5D),验证集中差异结果(图5E),可以看到,除*MTHFD2*外,其余基因在训练集和验证集中都是显著差异的,且上下调一致,都是下调,因此被确定为8个显著基因。除*MTHFD2*外的其余基因在训练集和验证集的ROC曲线的AUC值基本都>0.7(图6A、B),因此被确定为特征基因。在训练集和验证集中展示特征基因的相关性热图——训练集结果(图6C)和验证集结果(图6D),2个数据集中对应基因的相关性正负方向都是一致的。



A: protein-protein interaction (PPI) network of the co-expressed genes, where a redder node color indicates a higher number of connections, i. e., a more central node; B: Venn diagram showing the intersection of the top genes identified by four centrality measures, MCC, MNC, Degree, and EPC, in the PPI network; C: the SVM-RFE algorithm used for feature selection; D: relationship between the number of genes and the error rate in SVM-RFE; E: distribution of LASSO coefficients; F: the binomial deviance versus log (lambda) for the LASSO model, where the 2 vertical dashed lines indicate the positions of lambda. min, left black line, and lambda. 1se, right black line

图4 交集基因的蛋白质互作网络分析与核心枢纽基因筛选

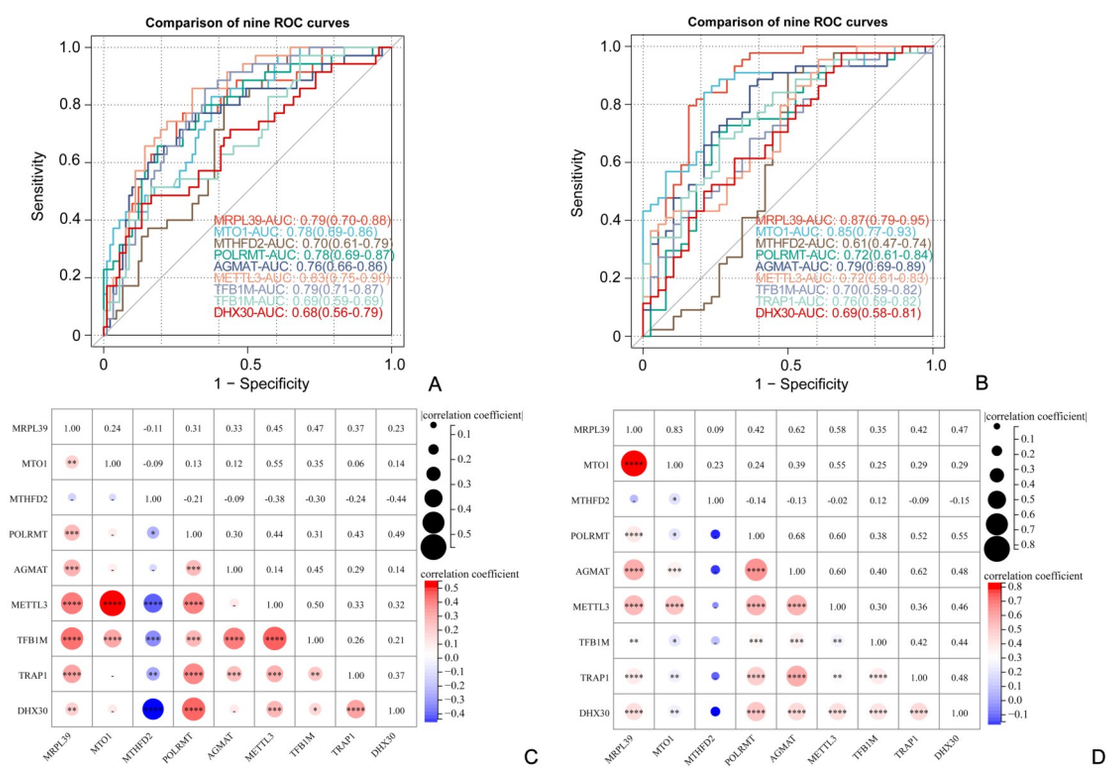
Figure 4 Protein-protein interaction network analysis of the intersection genes and core hub gene identification



A: relative importance of genes in the random forest; B: relationship between the number of trees and the error rate in the random forest; C: Venn diagram of the intersection of genes screened by the 3 machine learning algorithms; D: differential expression box plot of the intersection genes in the training set; E: differential expression box plot of the intersection genes in the validation set

图5 基于3种机器学习算法的特征基因筛选与表达验证

Figure 5 Feature gene screening based on 3 machine learning algorithms and expression validation



A: ROC curves of the intersection genes in the training set; B: ROC curves of the intersection genes in the validation set; C: correlation heatmap of the feature genes in the training set; D: correlation heatmap of the feature genes in the validation set

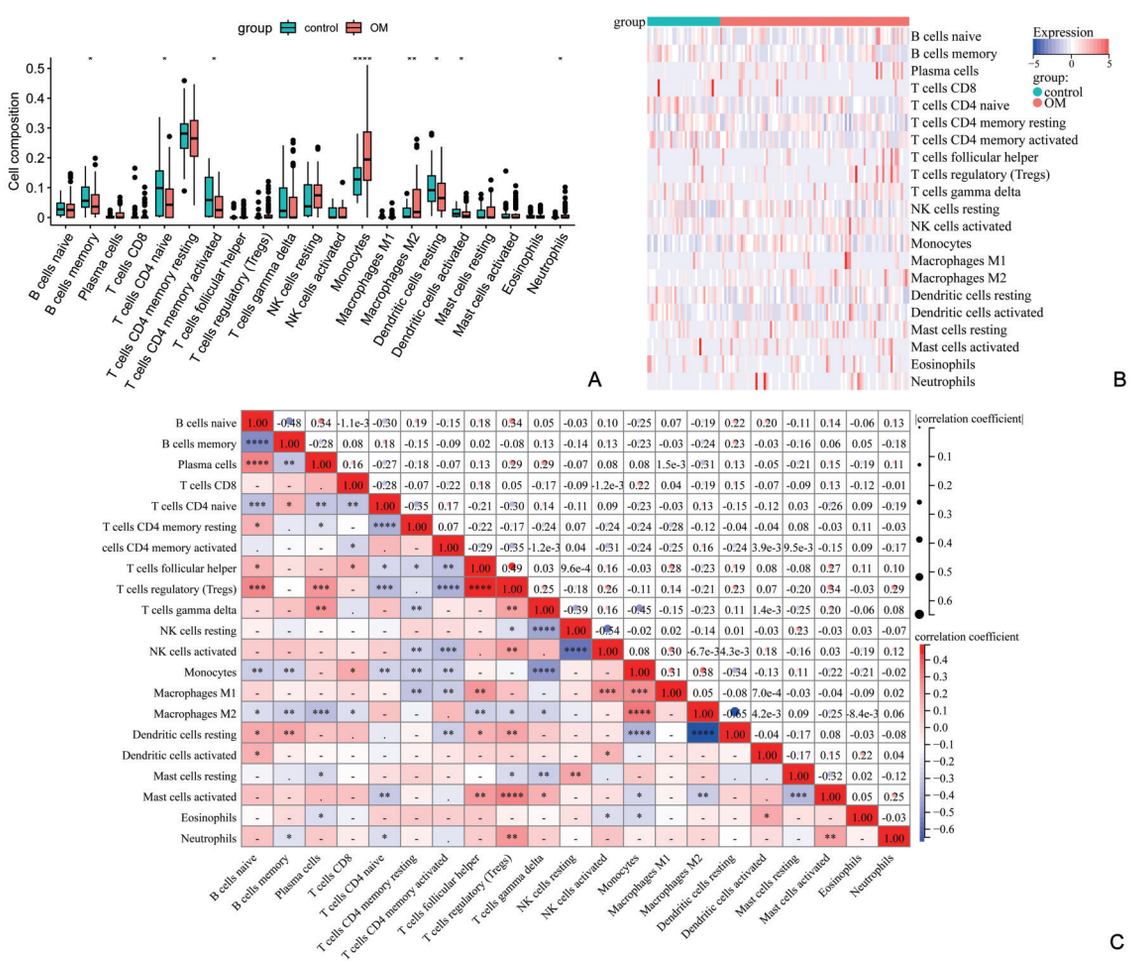
图6 特征基因的诊断效能评估(ROC曲线)及其表达相关性

Figure 6 Diagnostic efficacy evaluation (ROC curves) and expression correlation of the feature genes

2.6 免疫细胞与特征基因的相关性

按方法所述,利用CIBERSORT和ssGSEA算法分别对训练数据集样本进行免疫细胞浸润分析,然后利用R语言中的秩和检验比较患病和对照样本中不同免疫细胞分布的差异性(图7A),共筛选得到了8个分布显著差异的免疫细胞,其中最显著的是单核细胞。同时绘制热图以展示各免疫细胞在不同样本中的分布情况(图7B)。对免疫细胞浸润值进行相关性分析(图7C,图8A),然后对差异显著的免疫细胞和特征基因进行相关性分析,其中记忆B细胞和所有特征基因都成正相关关系,除MRPL39,POLRMT和TRAP1外都为显著相关。Monocyte和所有特征基因都成负相关关系,除MTO1,POLRMT和DHX30外都为显著相关。

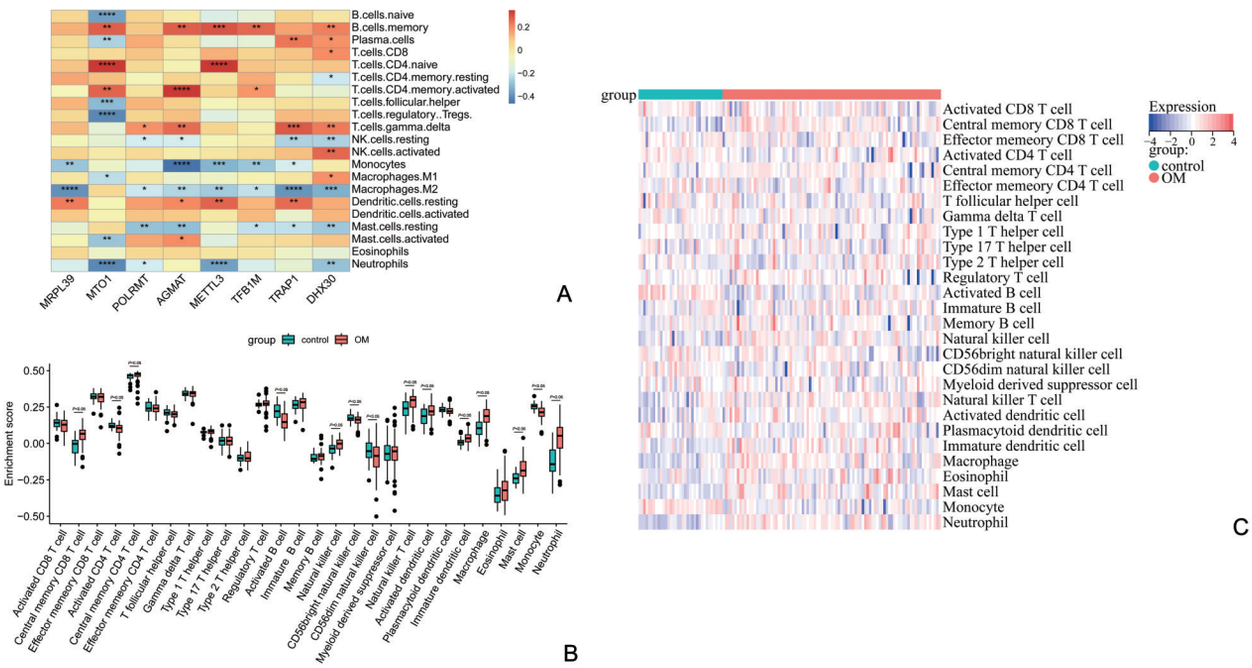
利用R语言中的秩和检验比较患病和对照样本中不同免疫细胞富集分数的差异性,绘制箱线图(图8B),共筛选得到了14个分布显著差异的免疫细胞。同时绘制热图以展示各免疫细胞在不同样本中的富集分数分布情况(图8C)。对免疫细胞富集分数进行相关性分析,结果如图8所示,然后对差异显著的免疫细胞和特征基因进行相关性分析,其中活化B细胞和所有特征基因都成显著正相关关系。自然杀伤细胞、自然杀伤T细胞、巨细胞及中性粒细胞和所有特征基因都成显著负相关关系。基于8个特征基因在训练集中构建列线图(图9A),并绘制校准曲线和决策曲线(图9B、C),校准曲线基本沿着对角线分布,决策曲线基本都分布在all线和none线之上的区域,说明模型具有良好的诊断性能。利用诊断列线图模型计算训练集和验证集各样本的列线图得分,绘制差异箱线图,并绘制ROC曲线评估性能(图9D、E)。验证集的结果(图10A、B),在训练集和验证集中列线图得分的箱线图差异都是显著的,ROC曲线的AUC值均>0.7的,说明特征基因的列线图得分能较好地区分患病和对照组样本。



A: differential box plot of immune cells between disease and control samples in the CIBERSORT algorithm; B: differential heatmap of immune cells between disease and control samples in the CIBERSORT algorithm; C: correlation heatmap of immune cells in the CIBERSORT algorithm

图7 基于CIBERSORT算法的免疫细胞浸润景观分析

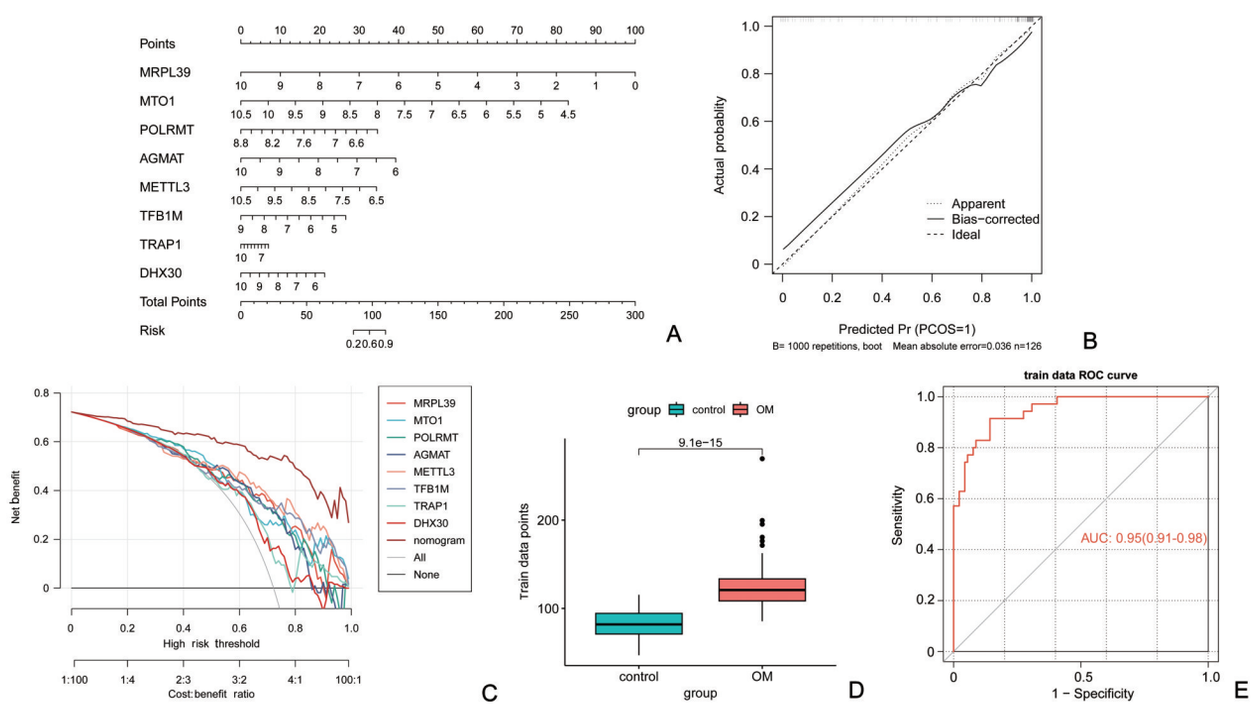
Figure 7 Immune cell infiltration landscape analysis based on the CIBERSORT algorithm



A: correlation heatmap between feature genes and immune cells in the CIBERSORT algorithm; B: differential box plot of immune cells between disease and control samples in the ssGSEA algorithm; C: differential heatmap of immune cells between disease and control samples in the ssGSEA algorithm

图8 特征基因与免疫细胞的相关性及基于ssGSEA的免疫富集分析

Figure 8 Correlation between feature genes and immune cells, and immune enrichment analysis based on ssGSEA



A: diagnostic nomogram constructed with feature genes in the training set; B: calibration curve of the nomogram; C: decision curve of the nomogram; D: differential box plot of the nomogram scores in the training set; E: ROC curve of the nomogram scores in the training set

图9 基于特征基因构建的列线图诊断模型及其性能评估

Figure 9 Nomogram diagnostic model constructed based on feature genes and its performance evaluation

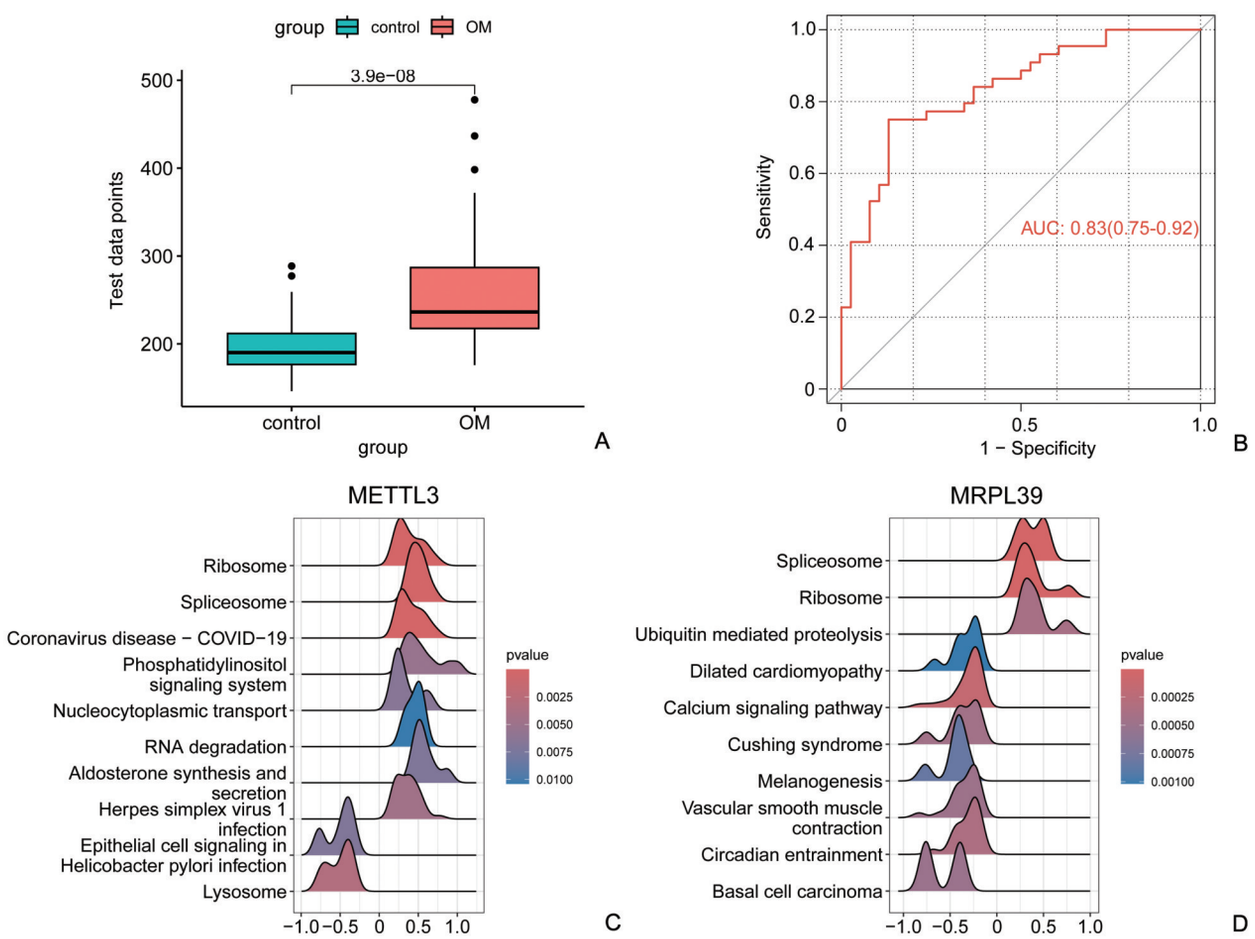
分别基于8种特征基因在OM患病样本中的高低表达分组进行GSEA富集分析。*METTL3*的结果显示,27个显著富集的通路,16个上调,9个下调,选择最显著的10条通路可视化展示(图10C)。*MRPL39*的结果显示,69个显著富集的通路,13个上调,56个下调,选择最显著的10条通路可视化展示(图10D)。

3 讨论

3.1 *METTL3-MRPL39*轴构成焦亡调控网络的核心枢纽

本研究验证了焦亡相关基因通过调控线粒体功能与免疫微环境参与SA-OM的发病机制。我们通过多组学联合分析筛选出的8个关键基因(*METTL3*、*MRPL39*、*POLRMT*等),其功能显著富集于"线粒体基因表达"与"代谢通路",这提示它们并非经典的焦亡执行分子(如GSDMD、Caspase-1),而是可能通过调控线粒体这一细胞器功能,进而影响炎症细胞命运和免疫应答。

其中,*METTL3-MRPL39*轴的发现为上述机制提供了具体线索。*METTL3*编码的m6A甲基转移酶,可通过修饰线粒体相关转录本,广泛影响免疫细胞的代谢重编程和炎症因子(如IL-1β)的产生。本研究观察到其表达与促炎因子水平及效应T细胞浸润呈正相关,支持了其作为炎症"放大器"的角色。另一方面,*MRPL39*作为线粒体核糖体蛋白,直接参与氧化磷酸化复合物的组装。其在SA-OM样本中的表达下调,可能损害线粒体电子传递链功能,导致活性氧(reactive oxygen species,ROS)过量产生和ATP合成不足。这种线粒体功能障碍状态,一方面可直接激活NLRP3炎症小体等焦亡关键传感器,另一方面可通过释放线粒体



A: differential box plot of the nomogram scores in the validation set; B: ROC curve of the nomogram scores in the validation set; C: GSEA pathway enrichment analysis for METTL3; D: GSEA pathway enrichment analysis for MRPL39

图10 诊断列线图在独立验证集中的性能及关键基因的GSEA通路富集

Figure 10 Performance of the diagnostic nomogram in the independent validation set and GSEA pathway enrichment of key genes

DNA等损伤相关分子模式,进一步扩增免疫反应。

因此,本课题组提出一个整合性的致病模型:在SA感染背景下,以*METTL3*和*MRPL39*为代表的基因表达失衡,共同导致了骨髓微环境中的线粒体功能障碍(表现为代谢紊乱与ROS蓄积)。这种功能障碍不仅是细胞能量危机和氧化应激的标志,更充当了触发免疫细胞(如巨噬细胞、T细胞)发生炎症性焦亡、并释放大量促炎因子的关键开关,最终塑造了本研究观察到的Th1/Th17偏倚的免疫微环境。这一定位使得*METTL3-MRPL39*轴超越了单纯的生物标志物范畴,成为连接上游感染信号与下游炎症失控的核心枢纽。

3.2 诊断模型的临床转化价值

本研究所构建的高效能诊断模型,其8个特征基因均指向上述线粒体-炎症调控网络,这从临床转化角度印证了该通路在SA-OM中的核心地位。模型优于传统炎症标志物的性能,可能正是因为其捕捉了更上游的、由线粒体功能障碍驱动的特异性分子紊乱。

基于机器学习初筛并经验证后确定的8个特征基因构建的诊断模型展现出优异的临床转化潜力。训练集ROC曲线下面积达0.89(95%*CI*:0.83~0.95),验证集AUC仍保持0.83(95%*CI*:0.76~0.90),显著优于传统炎症标志物CRP(AUC=0.72)和ESR(AUC=0.68)。列线图模型的校准曲线斜率为0.91,DCA曲线显示净获益率超过50%,提示该模型具有较好的临床决策价值。值得注意的是,MTHFD2虽在差异表达分析中被剔除,但其参与的叶酸代谢通路可能在疾病进展中发挥潜在作用,提示多组学整合分析的重要性。

3.3 免疫微环境特征的治疗启示

通过CIBERSORT和ssGSEA双平台验证发现,疾病组呈现典型的Th1/Th17极化特征,其中$CD8^+$T细胞浸润水平较对照组升高3.2倍($P<0.001$)。有趣的是,关键基因*METTL3*与效应T细胞浸润呈显著正相关($r=0.65$),其机制可能与*METTL3*通过m6A修饰调控T细胞线粒体代谢,从而影响其增殖与功能有关;而*MRPL39*则与调节性T细胞比例呈负相关($r=-0.58$)。这种免疫微环境的分型特征为精准治疗提供新靶点:高*METTL3*表达亚型可能对JAK抑制剂更敏感,而低*MRPL39*亚型可能更适合免疫检查点阻断治疗。此外,本研究发现自然杀伤细胞活性受损与疾病严重程度呈正相关,这为间充质干细胞联合NK细胞输注的治疗策略提供理论依据。

本研究发现的关键基因,特别是*METTL3*与*MRPL39*,不仅与特定的免疫细胞浸润水平显著相关,其已知的生物学功能也为解释这种相关性提供了合理的机制线索。本研究假设,它们可能通过调控免疫细胞的代谢与功能状态,共同塑造了SA-OM中观察到的促炎性免疫微环境。

对于METTL3,其作为关键的m6A RNA甲基转移酶,可能通过转录后修饰广泛调控免疫细胞的基因表达程序。本研究中其表达与活化的$CD8^+$T细胞及Th1/Th17细胞特征正相关,提示其可能通过m6A修饰促进T细胞受体信号下游分子(如IFNG、TBX21)或糖酵解相关基因(如HK2、PDK1)的转录本稳定性和翻译效率,从而驱动效应T细胞的代谢重编程、增殖与炎症因子(如IFN-γ, IL-17)产生,加剧炎症反应[34-35]。

对于MRPL39,其作为线粒体核糖体蛋白,其下调导致的线粒体功能障碍可能是连接基因表达与免疫表型的关键环节。MRPL39缺陷引起的氧化磷酸化受损和ROS累积,可产生双重免疫调节效应:一方面,ROS可作为第二信使直接激活巨噬细胞和中性粒细胞中的NLRP3炎症小体,促进IL-1β等焦亡相关细胞因子的释放,并驱动单核/巨噬细胞向促炎的M1表型极化;另一方面,线粒体功能紊乱导致的代谢压力(如ATP不足)可能破坏调节性T细胞的稳定性和抑制功能,从而解除对效应T细胞的抑制,间接助长了Th1/Th17反应的优势[36]。

3.4 研究局限性及改进方向

本研究存在以下局限性:①样本量较小(训练集91例),需扩大队列验证;②仅纳入GEO公共数据,缺乏独立临床样本验证;③本研究已通过qPCR对关键基因的表达进行了初步验证,但机制研究仍不够深入,未来需通过动物实验或进一步的细胞功能实验(如检测焦亡执行蛋白、线粒体功能指标等)验证*METTL3-MRPL39*轴的核心调控作用;④构建多中心前瞻性队列进行模型优化;⑤探索焦亡相关circRNA在骨髓炎中的作用机制;⑥开展基于CRISPR-Cas9的基因编辑研究,验证关键节点基因的功能。

本研究系统揭示SA诱导骨髓炎中焦亡相关基因的调控网络,构建的诊断模型具有较高的临床转化价值。*METTL3-MRPL39*轴介导的线粒体-炎症互作机制为理解骨髓炎发病机理提供新视角,基于免疫微环境特征的个体化治疗策略有望改善患者预后。

利益冲突 本文所有作者声明不存在利益冲突

参考文献:

[1] LEW D P, WALDVOGEL F A. Osteomyelitis[J]. Lancet, 2004, 364(9431): 369-379. DOI:10.1016/S0140-6736(04)16727-5.

[2] BURY D C, ROGERS T S, DICKMAN M M. Osteomyelitis: diagnosis and treatment[J]. Am Fam Physician, 2021, 104(4): 395-402.

[3] URISH K L, CASSAT J E. *Staphylococcus aureus* osteomyelitis: bone, bugs, and surgery[J]. Infect Immun, 2020, 88(7): e00932-e00919. DOI:10.1128/IAI.00932-19.

[4] SCHALLER J G. Arthritis and infections of bones and joints in children[J]. Pediatr Clin North Am, 1977, 24(4): 775-790. DOI:10.1016/s0031-3955(16)33497-6.

[5] JIN M C, WU X H, HU J, *et al*. EGFR-MEK1/2 cascade negatively regulates bactericidal function of bone marrow macrophages in mice with *Staphylococcus aureus* osteomyelitis[J]. PLoS Pathog, 2024, 20(8): e1012437. DOI: 10.1371/journal.ppat.1012437.

[6] HAABER J, PENADÉS J R, INGMER H. Transfer of antibiotic resistance in *Staphylococcus aureus*[J]. Trends Microbiol, 2017, 25(11): 893-905. DOI: 10.1016/j.tim.2017.05.011.

[7] HOWDEN B P, GIULIERI S G, WONG FOK LUNG T, *et al*. *Staphylococcus aureus* host interactions and adaptation[J]. Nat Rev Microbiol, 2023, 21(6): 380-395. DOI: 10.1038/s41579-023-00852-y.

[8] PENG J J, SONG W T, YAO F, *et al*. Involvement of regulated necrosis in blinding diseases: Focus on necroptosis and ferroptosis[J]. Exp Eye Res, 2020, 191: 107922. DOI: 10.1016/j.exer.2020.107922.

[9] RAO Z P, ZHU Y T, YANG P, *et al*. Pyroptosis in inflammatory diseases and cancer[J]. Theranostics, 2022, 12(9): 4310-4329. DOI:10.7150/thno.71086.

[10] VASUDEVAN S O, BEHL B, RATHINAM V A. Pyroptosis-induced inflammation and tissue damage[J]. Semin Immunol, 2023, 69: 101781. DOI: 10.1016/j.smim.2023.101781.

[11] KOVACS S B, MIAO E A. Gasdermins: effectors of pyroptosis[J]. Trends Cell Biol, 2017, 27(9): 673-684. DOI:10.1016/j.tcb.2017.05.005.

[12] HSU S K, LI C Y, LIN I L, *et al*. Inflammation-related pyroptosis, a novel programmed cell death pathway, and its crosstalk with immune therapy in cancer treatment[J]. Theranostics, 2021, 11(18): 8813-8835. DOI: 10.7150/thno.62521.

[13] MENDELSOHN D H, NIEDERMAIR T, WALTER N, *et al*. Ultrastructural evidence of mitochondrial dysfunction in osteomyelitis patients[J]. Int J Mol Sci, 2023, 24(6): 5709. DOI:10.3390/ijms24065709.

[14] BARRETT T, WILHITE S E, LEDOUX P, *et al*. NCBI GEO: archive for functional genomics data sets: update[J]. Nucleic Acids Res, 2013, 41(Database issue): D991-D995. DOI:10.1093/nar/gks1193.

[15] LEEK J T, JOHNSON W E, PARKER H S, *et al*. The sva package for removing batch effects and other unwanted variation in high-throughput experiments[J]. Bioinformatics, 2012, 28(6): 882-883. DOI: 10.1093/bioinformatics/bts034.

[16] RITCHIE M E, PHIPSON B, WU D, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Res, 2015, 43(7): e47. DOI:10.1093/nar/gkv007.

[17] ZHANG L, CHEN S K, ZENG X X, *et al*. Revealing the pathogenic changes of PAH based on multiomics characteristics[J]. J Transl Med, 2019, 17(1): 231. DOI: 10.1186/s12967-019-1981-5.

[18] LUO A, JIA Y F, HAO R R, *et al*. Proteomic and phosphoproteomic analysis of right ventricular hypertrophy in the pulmonary hypertension rat model[J]. J Proteome Res, 2024, 23(1): 264-276. DOI:10.1021/acs.jproteome.3c00546.

[19] LANGFELDER P, HORVATH S. WGCNA: an R package for weighted correlation network analysis[J]. BMC Bioinformatics, 2008, 9: 559. DOI: 10.1186/1471-2105-9-559.

[20] COHEN J. Statistical power analysis for the behavioral science (2nd Ed)[M]. Hillsdale, NJ: Lawrence Erlbaum Associates. 1988, 83.

[21] SUN Z, GAO Z Y, XIANG M M, *et al*. Comprehensive analysis of lactate-related gene profiles and immune characteristics in lupus nephritis[J]. Front Immunol, 2024, 15: 1329009. DOI:10.3389/fimmu.2024.1329009.

[22] YU G C, WANG L G, HAN Y Y, *et al*. clusterProfiler: an R package for comparing biological themes among gene clusters[J]. OMICS, 2012, 16(5): 284-287. DOI:10.1089/omi.2011.0118.

[23] ASHBURNER M, BALL C A, BLAKE J A, *et al*. Gene Ontology: tool for the unification of biology[J]. Nat Genet, 2000, 25(1): 25-29. DOI:10.1038/75556.

[24] KANEHISA M, GOTO S. KEGG Kyoto encyclopedia of genes and genomes[J]. Nucleic Acids Res, 2000, 28(1): 27-30. DOI:10.1093/nar/28.1.27.

[25] SZKLARCZYK D, MORRIS J H, COOK H, *et al*. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible[J].

Nucleic Acids Res, 2017, 45(D1): D362-D368. DOI: 10.1093/nar/gkw937.

[26] SHANNON P, MARKIEL A, OZIER O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome Res, 2003, 13(11): 2498-2504. DOI:10.1101/gr.1239303.

[27] CHIN C H, CHEN S H, WU H H, *et al*. cytoHubba: identifying hub objects and sub-networks from complex interactome[J]. BMC Syst Biol, 2014, 8(Suppl 4): S11. DOI:10.1186/1752-0509-8-S4-S11.

[28] SANZ H, VALIM C, VEGAS E, *et al*. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels[J]. BMC Bioinformatics, 2018, 19(1): 432. DOI:10.1186/s12859-018-2451-4.

[29] GAO J B, KWAN P W, SHI D M. Sparse kernel learning with LASSO and Bayesian inference algorithm[J]. Neural Netw, 2010, 23(2): 257-264. DOI: 10.1016/j.neunet.2009.07.001.

[30] ROBIN X, TURCK N, HAINARD A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves[J]. BMC Bioinformatics, 2011, 12: 77. DOI: 10.1186/1471-2105-12-77.

[31] CHEN B B, KHODADOUST M S, LIU C L, *et al*. Profiling tumor infiltrating immune cells with CIBERSORT[J]. Methods Mol Biol, 2018, 1711: 243-259. DOI: 10.1007/978-1-4939-7493-1_12.

[32] KAWADA J I, TAKEUCHI S, IMAI H, *et al*. Immune cell infiltration landscapes in pediatric acute myocarditis analyzed by CIBERSORT[J]. J Cardiol, 2021, 77(2): 174-178. DOI:10.1016/j.jjcc.2020.08.004.

[33] HÄNZELMANN S, CASTELO R, GUINNEY J. GSVA gene set variation analysis for microarray and RNA-seq data[J]. BMC Bioinformatics, 2013, 14: 7. DOI: 10.1186/1471-2105-14-7.

[34] DE KESEL J, FIJALKOWSKI I, PIETERS T, *et al*. HNRNPC and m6A RNA methylation control oncogenic transcription and metabolism in T-cell leukemia[J]. Blood, 2025, 146(3): 275-290. DOI:10.1182/blood.2024026848.

[35] WANG A T, HUANG H Y, SHI J H, *et al*. USP47 inhibits m6A-dependent c-Myc translation to maintain regulatory T cell metabolic and functional homeostasis[J]. J Clin Invest, 2023, 133(23): e169365. DOI:10.1172/JCI169365.

[36] CHEN S Z, SAEED A F U H, LIU Q, *et al*. Macrophages in immunoregulation and therapeutics[J]. Sig Transduct Target Ther, 2023, 8: 207. DOI:10.1038/s41392-023-01452-1.

(**收稿**:2026-02-13;**修回**:2026-03-26)

(编辑 栾 嘉)