·论著·

铁死亡相关基因作为新型标志物预测结核潜伏感染活化风险及风险模型构建

姜吉亮[1],王文涛[1],李乐然[1],尹绍卿[1],付玉荣[2],伊正君[1]
(山东第二医科大学 1.医学检验学院分子诊断学教研室;2.基础医学院病原生物学教研室,山东 潍坊 261053)

摘要 目的 利用生物信息学和多种机器学习算法,筛选出预测结核潜伏感染(LTBI)活化风险的新型标志物并建立风险模型。**方法** 从基因表达综合数据库获取GSE112104和GSE193777数据集,通过差异基因分析和加权基因共表达网络分析筛选与LTBI活化相关的铁死亡相关差异基因(FRG-DEGs)。进一步通过LASSO、SVM-RFE和RF 3种机器学习算法筛选铁死亡相关关键基因(FRG-hubs),并通过验证集和逆转录PCR验证其可靠性。最后,利用R语言建立风险模型。**结果** 在GSE112104数据集中,与LTBI相比,活动性结核病中296个基因显著上调,1 569个基因显著下调,LTBI进展者中506个基因显著上调,1 132个基因显著下调。WGCNA共得到5个基因共表达模块,其中蓝色模块与LTBI活化相关性最强(cor=0.62,*P*=0.000 04),包含1 340个基因。将三者连同728个铁死亡相关基因(FRG)取交集,共得到8个表达趋势一致的FRG-DEGs。3种机器学习算法共筛选出4个FRG-hubs,包括*PLA2G6*、*GLS2*、*JUN*和*AMN*。逆转录PCR结果显示,随着休眠结核分枝杆菌的活化,FRG-hubs的表达逐渐降低。最后,基于FRG-hubs构建了LTBI活化风险模型,曲线下面积在0.98~1.00之间。**结论** 本研究成功筛选出用于预测LTBI活化风险的新型标志物,并构建了具有良好预测效能的LTBI活化风险模型。

关键词 结核病;结核潜伏感染;铁死亡;风险模型;机器学习
中图分类号 R521 **文献标志码** A **文章编号** 0258-4646(2025)04-0333-07
网络出版地址 https://link.cnki.net/urlid/21.1227.R.20250410.0903.008
DOI:10.12007/j.issn.0258-4646.2025.04.009

Ferroptosis-related genes as novel biomarkers for predicting the risk of latent tuberculosis infection activation and establishment of a risk model

JIANG Jiliang[1], WANG Wentao[1], LI Leran[1], YIN Shaoqing[1], FU Yurong[2], YI Zhengjun[1]
(1. Department of Molecular Diagnostics, School of Medical Laboratory Science, Shandong Second Medical University, Weifang 261053, China; 2. Department of Pathogenic Biology, School of Basic Medical Sciences, Shandong Second Medical University, Weifang 261053, China)

Abstract Objective To identify novel biomarkers for predicting the risk of latent tuberculosis infection (LTBI) activation using bioinformatics and machine-learning algorithms and to establish a risk model. **Methods** The GSE112104 and GSE193777 datasets were obtained from the Gene Expression Omnibus. Differential gene expression and weighted gene co-expression network analyses were performed to identify ferroptosis-related differentially expressed genes (FRG-DEGs) associated with LTBI activation. Three machine-learning algorithms, least absolute shrinkage and selection operator, support vector machine-recursive feature elimination, and random forest, were used to identify ferroptosis-related hub genes (FRG-hubs). The reliability of these genes was validated using independent validation datasets and reverse transcription polymerase chain reaction (PCR). A risk model was established using R software. **Results** In the GSE112104 dataset, 296 genes were upregulated and 1 569 genes were downregulated in active tuberculosis compared to those in LTBI. Among the LTBI progressors, 506 genes were upregulated and 1 132 genes were downregulated. Weighted correlation network analysis identified five gene modules, with the blue module showing the strongest correlation with LTBI activation (cor=0.62, *P*=0.000 04), containing 1 340 genes. Intersections with 728 ferroptosis-related genes resulted in eight FRG-DEGs. The machine-learning algorithms identified four FRG-hubs: *PLA2G6*, *GLS2*, *JUN*, and *AMN*, whose expression decreased with LTBI activation. Reverse transcription PCR confirmed this trend. A risk model based on these genes yielded an area under the curve of 0.98 to 1.00. **Conclusion** This study successfully identified novel biomarkers for predicting the risk of LTBI activation and developed an accurate predictive risk model.

Keywords tuberculosis; latent tuberculosis infection; ferroptosis; risk model; machine learning

基金项目:山东省自然科学基金(ZR2022MH024);山东第二医科大学研究生科研创新基金(2023YJSCX007)
作者简介:姜吉亮(2000-),男,硕士研究生.
通信作者:伊正君,E-mail:fuyizhengjun@163.com
收稿日期:2024-05-11
网络出版时间:2025-04-10 13:26:56

结核病是由结核分枝杆菌（*Mycobacterium tuberculosis*，Mtb）感染引起的一种慢性传染性疾病，是严重危害人类生命健康的传染病之一。根据疾病状态，可分为活动性结核病（active tuberculosis，ATB）和结核潜伏感染（latent tuberculosis infection，LTBI）。全球约1/4的人群感染Mtb，其中5%~10%可能发展成ATB[1]。研究[2]显示，对LTBI人群进行预防性治疗可以显著降低结核病的发病率。然而，对所有LTBI人群进行预防性治疗并不可行。因此，识别出最有可能在未来几年内发展成ATB的高风险人群，将使预防性治疗更加精准，从而使最有可能受益的人群获益。然而，目前的诊断方法无法早期识别高危人群，并且漏诊率很高。因此，迫切需要寻找更准确、更方便的诊断方法。

近年来，基于宿主转录组的生物标志物在结核病的诊断和监测中获得了广泛的认可[3]。Mtb感染宿主后会引发一系列病理生理变化，包括细胞死亡、炎症反应、干扰素反应和免疫激活等。这些变化会导致分子层面上的异常表达模式[4]，为结核病的诊断提供了新的分子标志物。铁是宿主防御细菌感染的关键因素，在宿主和病原体间的营养斗争中发挥关键作用[5]。研究[6]指出，血清或血浆中的铁水平与结核病的发病率有关。铁过载会引起铁死亡，铁死亡是促进Mtb在宿主中存活和传播的关键机制之一[7]。鉴于宿主严格调控大多数形式的细胞死亡，本研究利用生物信息学和多种机器学习算法，筛选铁死亡相关基因（ferroptosis-related gene，FRG）作为预测LTBI活化风险的新型标志物并建立风险模型。

1　材料与方法

1.1　数据来源

从基因表达综合数据库（Gene Expression Omnibus，GEO）获取GSE112104和GSE19377数据集。GSE112104数据集包括36例LTBI个体和14例ATB患者。经过5年的随访，36例LTBI个体中有16例进展为ATB，称为LTBI进展者（LTBI progressor，LTBIPro），用作训练集。GSE193777数据集包括11例ATB患者和35例LTBI个体，用作验证集。728个FRG来源于FerrDb数据库（http://www.zhounan.org/ferrdb/current/）。

1.2　加权基因共表达网络分析（weighted correlation network analysis，WGCNA）

采用WGCNA探索LTBI活化的差异基因模块及其枢纽基因。基于WGCNA包的pickSoftThreshold函数筛选最佳软阈值，使基因间的调控关系符合无尺度网络分布。然后，构建基因分层聚类树，利用动态剪枝法识别基因共表达模块。最后，计算各基因共表达模块与临床组别间的相关性。

1.3　差异表达基因（differentially expressed genes，DEGs）分析

使用R语言limma包筛选DEGs。将DEGs和WGCNA中与LTBI活化相关性最强的模块内基因以及FRG取交集，得到LTBI活化相关铁死亡相关差异表达基因（ferroptosis-related differentially expressed genes，FRG-DEGs）。

1.4　铁死亡相关关键基因（ferroptosis-related hub genes，FRG-hubs）的筛选

采用最小绝对收缩与选择算法（least absolute shrinkage and selection operator，LASSO）、支持向量机递归特征消除（support vector machine-recursive feature elimination，SVM-RFE）和随机森林（random forest，RF）3种机器学习算法，识别与LTBI活化相关的FRG-hubs。

1.5　骨髓源性巨噬细胞（bone marrow-derived macrophage，BMDM）的培养和感染

按照改良的Wayne模型[8]，诱导休眠结核分枝杆菌（dormant *Mycobacterium tuberculosis*，D-Mtb）。从8周龄C57BL/6小鼠的股骨和胫骨中分离BMDM，悬浮于含10% 胎牛血清（美国Gibco公司）和30 ng/mL巨噬细胞集落刺激因子的RPMI 1640培养基（美国Gibco公司）中7 d。用D-Mtb以感染复数=5感染BMDM，感染后6 h更换培养基，记为感染0 h。本研究获得山东第二医科大学实验动物伦理委员会批准（编号：2021SDL528）。

1.6　RNA的提取和逆转录PCR

使用TRIzol试剂（美国Invitrogen公司）从细胞中分离总mRNA，并使用 PrimeScript™ Ⅱ第一链cDNA合成试剂盒（日本TaKaRa公司）制备cDNA，通过逆转录PCR评估mRNA水平。

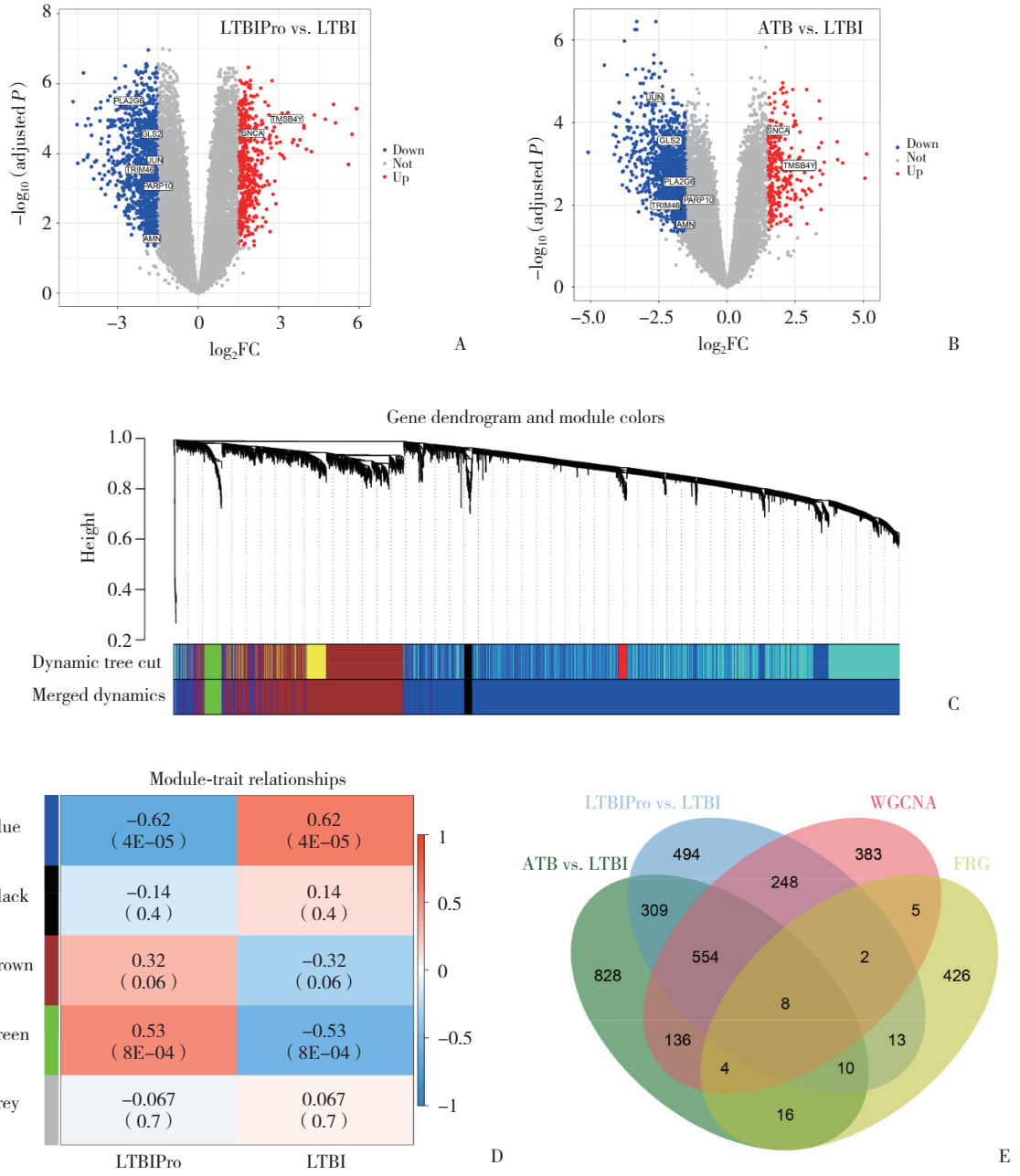
1.7　基于FRG-hubs建立风险模型

使用R语言rms包创建列线图，预测LTBI活化风险。使用校准曲线、决策曲线、受试者操作特征（receiver operating characteristics，ROC）曲线评估模型效能。

2　结果

2.1　筛选参与LTBI活化的FRG

在训练集GSE112104中，设定$P<0.05$和$|\log_2FC|>1.5$为过滤条件，分别对ATB与LTBI、LTBIPro与LTBI样本表达谱进行差异分析。结果发现，与LTBI相比，LTBIPro中506个基因显著上调，1 132个基因显著下调（图1A）；ATB中296个基因显著上调，1 569个基因显著下调（图1B）。WGCNA结果显示，当软阈值为5时，基因间的调控关系符合无尺度网络分布（图1C），共得到5个基因共表达模块（图1D），其中蓝色模块与LTBI活化相关性最强（cor=0.62，$P=0.000\ 04$），包含1 340个枢纽基因。将ATB与LTBI、LTBIPro与LTBI的差异基因与WGCNA蓝色模块内的基因以及728个FRG取交集，共得到8个趋势一致的FRG-DEGs（图1E）。



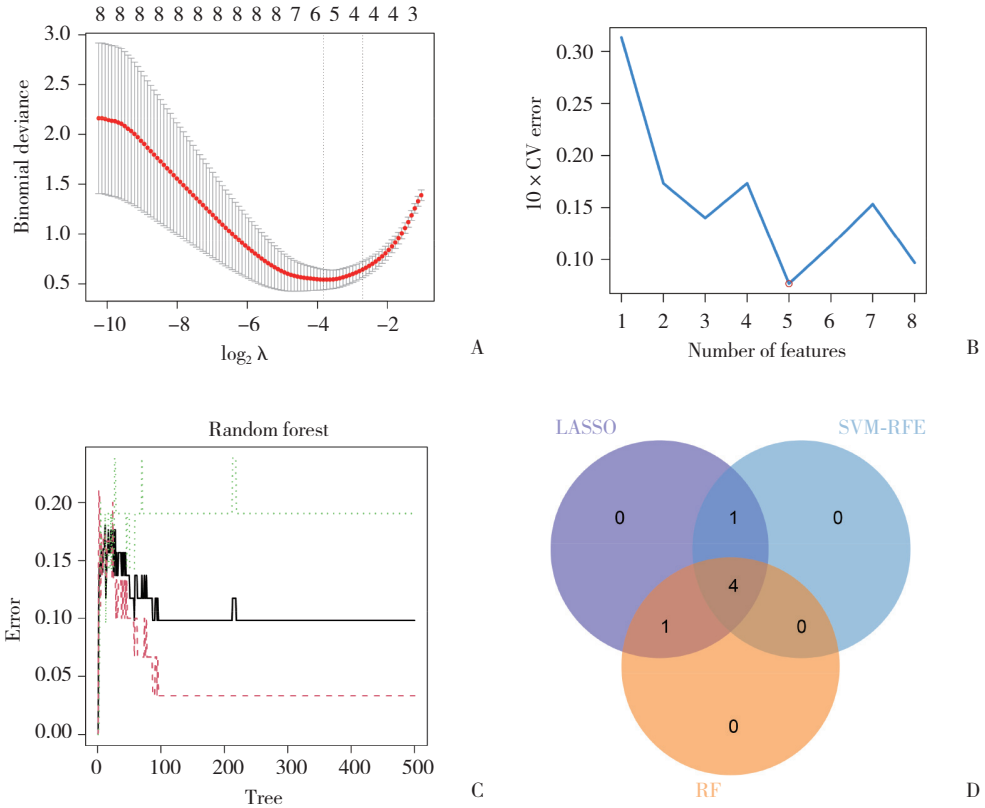
A，volcano plot of genes differentially expressed between LTBI and LTBIPro；B，volcano plot of genes differentially expressed between ATB and LTBI；C，heatmap of module-trait relationships showing clustering of genes associated with LTBI activation；D，heatmap depicting the correlation between gene co-expression modules and LTBI activation；E，Venn diagram illustrating the selection of FRG-DEGs.

图1　参与LTBI活化的FRG的筛选

Fig.1　Screening of FRG involved in LTBI activation

2.2　机器学习鉴定FRG-hubs

运用LASSO、SVM-RFE、RF 3种独立机器学习算法，从8个FRG-DEGs中筛选FRG-hubs。在LASSO回归中，通过10倍交叉验证调整最优惩罚参数λ，鉴定出6个基因（图2A），分别为*PLA2G6*、*TMSB4Y*、*SNCA*、*GLS2*、*JUN*、*AMN*。在SVM-RFE中，$n = 5$时分类误差最小（0.076 7），确定了5个基因（图2B），分别为*JUN*、*TMSB4Y*、*GLS2*、*PLA2G6*、*AMN*。RF共筛选出5个基因（图2C），分别为*SNCA*、*TMSB4Y*、*PPARD*、*KMT2D*、*PARP15*、*GLS2*。最终将3种机器学习共有的4个特征基因（图2D），包括*PLA2G6*、*GLS2*、*JUN*和*AMN*，作为FRG-hubs，用于后续分析。分别在训练集GSE112104和验证集GSE193777中验证其表达，结果发现，与LTBI相比，LTBIPro和ATB中FRG-hubs均显著下调（图3A）。ROC曲线（图3B）显示出良好的诊断效能，曲线下面积在0.76~1.00之间。



A，least absolute shrinkage and selection operator（LASSO）algorithm；B，support vector machine-recursive feature elimination（SVM-RFE）algorithm；C，random forest（RF）algorithm；D，Venn diagram illustrating the FRG-hubs common to the LASSO，SVM-RFE，and RF algorithms.

图2　FRG-hubs的识别

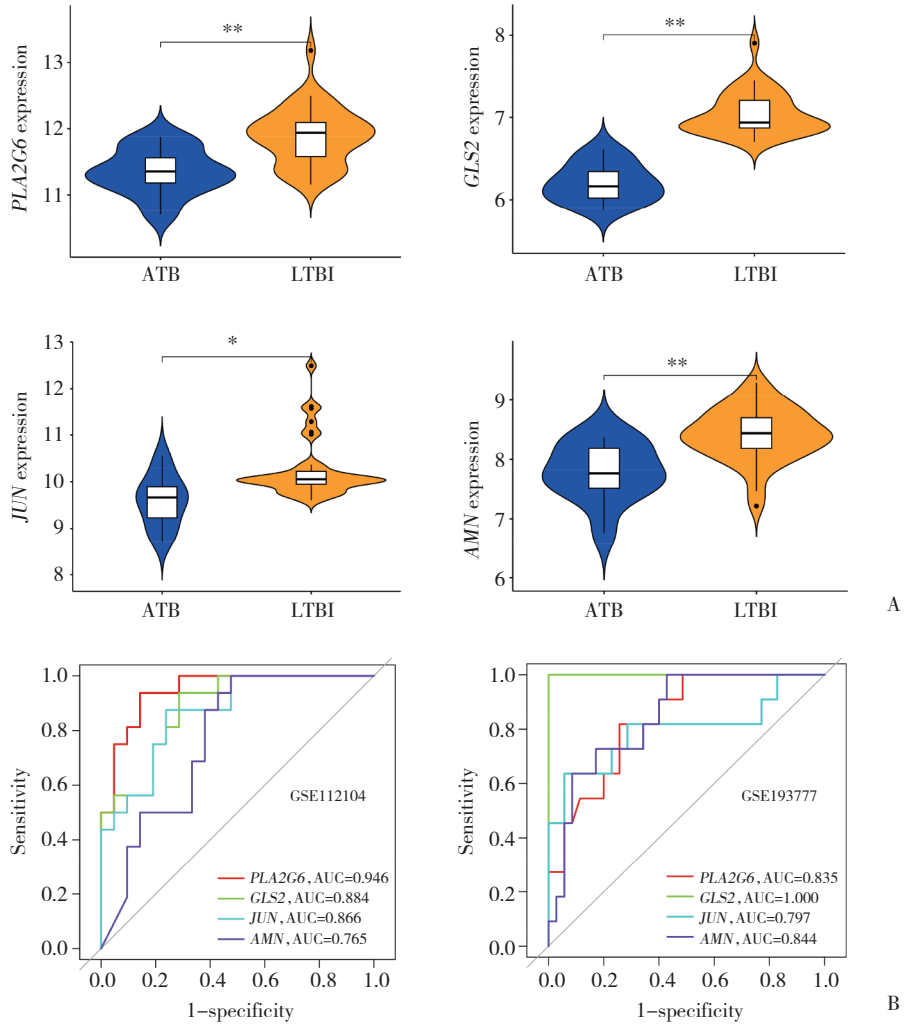
Fig.2　Identification of FRG-hubs

2.3　D-Mtb活化后FRG-hubs的表达逐渐减少

为了进一步研究LTBI活化过程中FRG-hubs的表达，建立LTBI细胞感染模型。首先诱导D-Mtb，透射电子显微镜下观察，发现D-Mtb细胞壁显著增厚（图4A），证明D-Mtb被成功诱导。D-Mtb感染BMDM后，金胺O和尼罗红双重染色显示，随着感染时间的延长，绿色荧光逐渐增强，而红色荧光逐渐减弱，表明细胞内D-Mtb逐渐活化（图4B）。逆转录PCR结果显示，随着感染时间的延长，FRG-hubs的表达水平逐渐降低（图4C）。

2.4　基于FRG-hubs构建LTBI活化风险模型

基于FRG-hubs开发LTBI活化风险模型，计算活化风险，风险评分=（$-10 \times$ *PLA2G6*$-3.9279 \times$ *GLS2*$-5.1599 \times$ *JUN*$+1.2553 \times$ *AMN*）（图5A）。校准曲线与理想曲线相近，说明模型的预测能力较好（图5B）。决策曲线变量在很大一个概率阈值区间内高于参考线，说明该变量模型较好（图5C）。应用ROC曲线评估模型的诊断效能，结果发现，4个FRG-hubs组合在训练集GSE112104（图5D）和验证集GSE193777（图5E）的曲线下面积分别为0.982和1.000。



A, expression of FRG-hubs in the GSE193777 dataset; B, diagnostic performance of FRG-hubs in distinguishing ATB and LTBI in the GSE112104 and GSE193777 datasets. *$P < 0.01$, **$P < 0.001$.

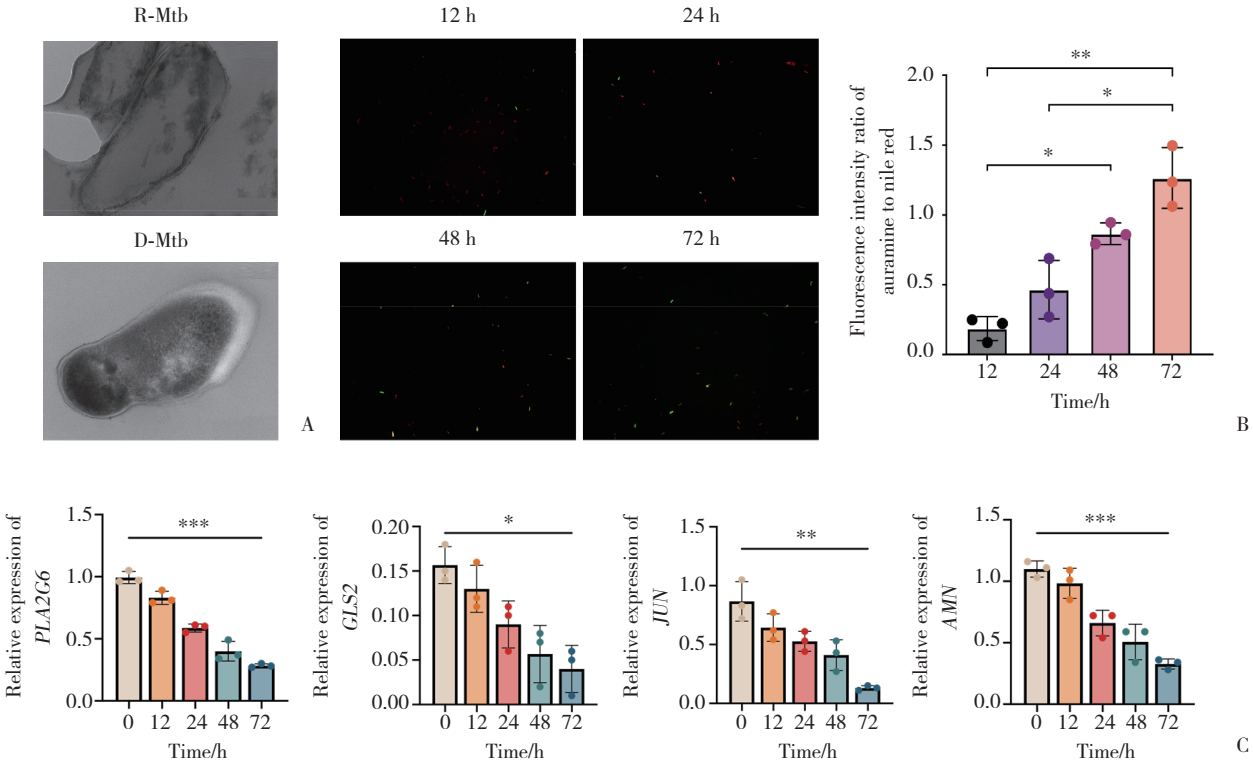
图3　FRG-hubs表达和诊断效能验证

Fig.3　Validation of FRG-hubs expressions and diagnostic efficacy

3　讨论

抑制宿主的铁死亡可以阻碍Mtb的传播[9]。在LTBI活化过程中，铁蛋白水平升高，导致细胞内铁浓度升高，诱导氧化应激和铁死亡，从而加剧LTBI活化[10]。本研究通过综合生物信息学分析以及多种机器学习算法筛选FRG，作为预测LTBI活化风险以及区分ATB和LTBI的新型标志物。首先，从LTBI人群和LTBIPro人群中得到LTBI活化的潜在风险基因，在LTBI人群和ATB患者中得到ATB相关基因。然后，通过WGCNA选择与LTBI活化最相关的模块，将三者与FRG取交集，得到8个表达趋势一致的FRG-DEGs。为了得到最佳的基因模型，通过LASSO、SVM-RFE、RF 3种独立机器学习算法，从8个FRG-DEGs中鉴定出4个FRG-hubs，分别为*PLA2G6*、*GLS2*、*JUN*和*AMN*。在训练集和验证集中验证其表达，结果发现，与LTBI相比，LTBIPro和ATB中FRG-hubs均显著下调。

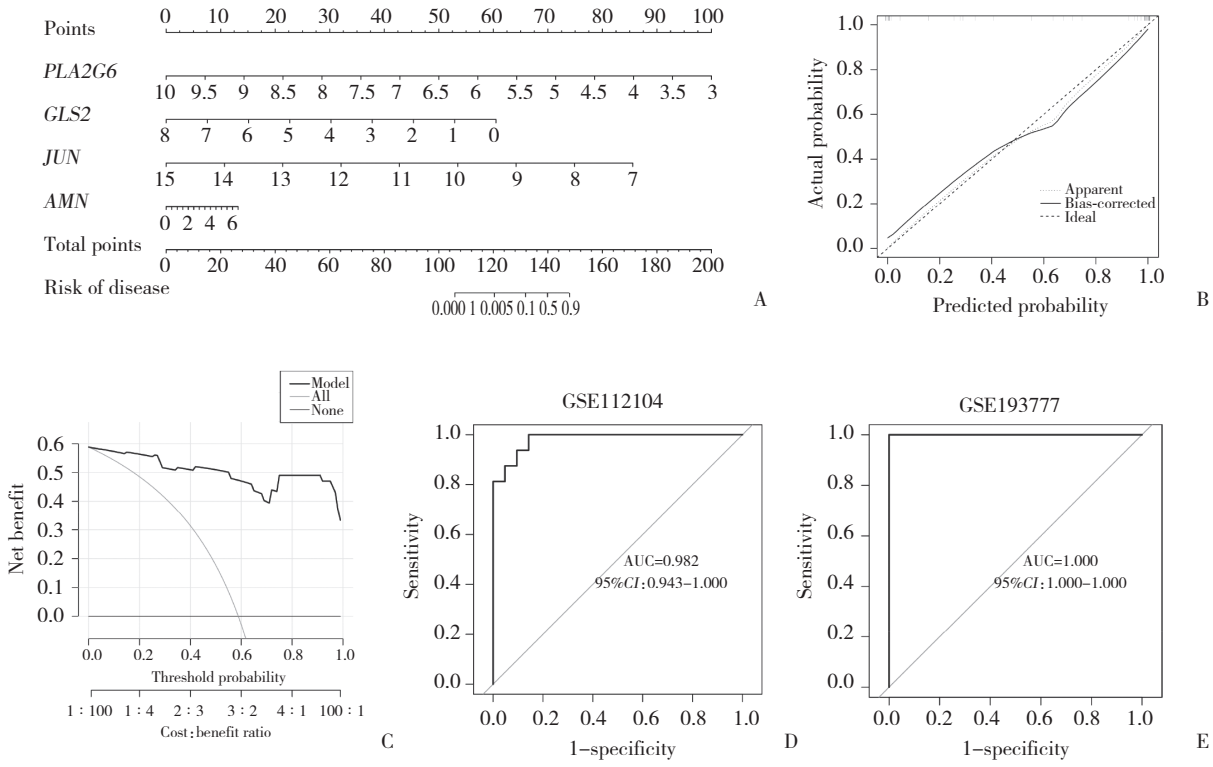
为了进一步验证研究结果，本研究构建D-Mtb感染巨噬细胞活化模型，使用D-Mtb感染BMDM，分别在12、24、48和72 h提取细胞RNA，逆转录PCR检测FRG-hubs表达量。结果发现，随着D-Mtb的活化，其表达量不断降低。有研究[11]表明，*PLA2G6*与免疫逃逸有关。*GLS2*是一种谷胱酰胺代谢酶，影响T细胞的分化和炎性细胞因子的产生[12]。*JUN*可以和γ干扰素近端启动子结合，从而促进其表达[13]。在LTBI中，Mtb处于潜伏状态，受宿主免疫系统的控制。*JUN*可能通过调节免疫细胞的活化、增殖和细胞因子的产生，影响LTBI的发展和活化过程。



A, R-Mtb and D-Mtb were observed under a transmission electron microscope (×25); B, double staining with auramine O and Nile red at 12, 24, 48, and 72 h after D-Mtb infection of macrophages (×1 000); C, expression trend of FRG-hubs after activation of D-Mtb. *P < 0.01, **P < 0.001, ***P < 0.000 1.

图4　D-Mtb的活化抑制FRG-hubs的表达

Fig.4　Activation of D-Mtb suppresses expression of FRG-hubs



A, column plot based on FRG-hubs; B, calibration curve; C, decision curve; D, diagnostic performance of the risk model in distinguishing ATB and LTBI in the GSE112104 dataset; E, diagnostic performance of the risk model in distinguishing ATB and LTBI in the GSE193777 dataset.

图5　基于FRG-hubs的风险模型的建立和评估

Fig.5　Establishment and evaluation of risk model based on FRG-hubs

*AMN*编码amnionless蛋白，这是一种细胞膜蛋白，在维生素B12的运输和吸收中发挥关键作用[14]。研究[15]表明，维生素B12在调节免疫细胞功能和抗炎过程中发挥重要作用。因此，AMN作为维生素B12的运输蛋白，可能通过调节免疫细胞中维生素B12的水平，影响LTBI的发展和活化过程。FRG-hubs主要与免疫调节、免疫细胞激活以及炎症反应有关，提示FRG-hubs可能影响免疫细胞对Mtb的识别，并调节细胞内代谢，引起机体免疫微环境改变，从而影响LTBI的活化。

鉴于FRG-hubs在LTBI活化过程中的关键作用，本研究建立了LTBI活化风险模型。ROC曲线、校准曲线和决策曲线证明该模型具有良好的准确性和特异度，可以作为LTBI活化风险的预测工具。

本研究仍存在一定的局限性。首先，样本量较小，这可能限制了研究结果的准确性。其次，FRG-hubs影响LTBI活化的具体机制仍不清楚。因此，需要进一步进行临床样本的验证及对FRG-hubs功能的研究。

参考文献：

[1] 舒薇，刘宇红. 笃志创新 躬行致远：世界卫生组织《2023年全球结核病报告》结核病科学研究章节解读[J]. 中国防痨杂志，2024，46(6)：613-617. DOI：10.19982/j.issn.1000-6621.20240159.

[2] KOWADA A. Interferon-gamma release assay for tuberculosis screening of solid organ transplant recipients is cost-effective [J]. J Infect，2019，78(1)：58-65. DOI：10.1016/j.jinf.2018.07.003.

[3] WARSINSKE H，VASHISHT R，KHATRI P. Host-response-based gene signatures for tuberculosis diagnosis：a systematic comparison of 16 signatures [J]. PLoS Med，2019，16(4)：e1002786. DOI：10.1371/journal.pmed.1002786.

[4] MAI D，JAHN A，MURRAY T，et al. Exposure to *Mycobacterium* remodels alveolar macrophages and the early innate response to *Mycobacterium tuberculosis* infection [J]. PLoS Pathog，2024，20(1)：e1011871. DOI：10.1371/journal.ppat.1011871.

[5] 韩亚，张政，张亚恒. 全基因测序探索多系统萎缩患者外周血液中铁死亡相关核心基因及其药物靶点[J]. 中国医科大学学报，2024，53(9)：821-826. DOI：10.12007/j.issn.0258-4646.2024.09.010.

[6] BOELAERT JR，VANDECASTEELE SJ，APPELBERG R，et al. The effect of the host's iron status on tuberculosis [J]. J Infect Dis，2007，195(12)：1745-1753. DOI：10.1086/518040.

[7] MEUNIER E，NEYROLLES O. Die another way：ferroptosis drives tuberculosis pathology [J]. J Exp Med，2019，216(3)：471-473. DOI：10.1084/jem.20190038.

[8] WAYNE LG，HAYES LG. An in vitro model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of non-replicating persistence [J]. Infect Immun，1996，64(6)：2062-2069. DOI：10.1128/iai.64.6.2062-2069.1996.

[9] AMARAL EP，COSTA DL，NAMASIVAYAM S，et al. A major role for ferroptosis in *Mycobacterium tuberculosis*-induced cell death and tissue necrosis [J]. J Exp Med，2019，216(3)：556-570. DOI：10.1084/jem.20181776.

[10] DAI Y，SHAN W，YANG Q，et al. Biomarkers of iron metabolism facilitate clinical diagnosis in *Mycobacterium tuberculosis* infection [J]. Thorax，2019，74(12)：1161-1167. DOI：10.1136/thoraxjnl-2018-212557.

[11] AKHADE AS，ATIF SM，LAKSHMI BS，et al. Type 1 interferon-dependent repression of NLRC4 and iPLA2 licenses down-regulation of *Salmonella* flagellin inside macrophages [J]. Proc Natl Acad Sci U S A，2020，117(47)：29811-29822. DOI：10.1073/pnas.2002747117.

[12] HE W，HU Y，CHEN D，et al. Hepatocellular carcinoma-infiltrating γδ T cells are functionally defected and allogenic Vδ2^{+} γδ T cell can be a promising complement [J]. Clin Transl Med，2022，12(4)：e800. DOI：10.1002/ctm2.800.

[13] SAMTEN B，TOWNSEND JC，WEIS SE，et al. CREB，ATF，and AP-1 transcription factors regulate IFN-γ secretion by human T cells in response to mycobacterial antigen [J]. J Immunol，2008，181(3)：2056-2064. DOI：10.4049/jimmunol.181.3.2056.

[14] JEON YM，KWON Y，LEE S，et al. Vitamin B12 reduces TDP-43 toxicity by alleviating oxidative stress and mitochondrial dysfunction [J]. Antioxidants (Basel)，2021，11(1)：82. DOI：10.3390/antiox11010082.

[15] KATHER S，KACZA J，PFANNKUCHE H，et al. Expression of the cobalamin transporters cubam and MRP1 in the canine ileum-Upregulation in chronic inflammatory enteropathy [J]. PLoS One，2024，19(1)：e0296024. DOI：10.1371/journal.pone.0296024.

（编辑　陈　姜）