

基于生物信息学和机器学习识别2型糖尿病线粒体特征基因



任 婧¹, 李宇婷¹, 刘晓琴², 孟祥龙¹

1. 山西中医药大学中药与食品工程学院 (山西晋中 030619)

2. 山东现代学院药学院 (济南 250104)

【摘要】目的 基于生物信息学与机器学习技术筛选 2 型糖尿病 (type 2 diabetes mellitus, T2DM) 骨骼肌线粒体功能相关特征基因并构建诊断模型, 为 T2DM 的临床诊断提供新思路。**方法** 整合基因表达综合数据库 (Gene Expression Omnibus, GEO) 中 T2DM 数据集, 采用 R 4.3.2 软件进行数据标准化和批次效应校正后, 筛选差异表达基因, 并与 MitoCarta 3.0 线粒体基因数据库取交集。通过最小绝对收缩和选择算子 (least absolute shrinkage and selection operator, LASSO) 回归、随机森林 (random forest, RF) 和支持向量机 (support vector machine, SVM) 算法筛选关键基因, 采用高斯混合模型 (Gaussian mixture model, GMM) 评估特征基因的可靠性, 构建逻辑回归诊断模型。利用受试者工作特征曲线 (receiver operating characteristic curve, ROC) 评估模型效能, 并在独立验证集中进行外部验证。**结果** 获得 23 个 T2DM 相关线粒体差异基因, 经机器学习算法筛选确定 *MRPS10*、*SLC25A5* 和 *TRNT1* 为核心特征基因, 其显著富集于氧化磷酸化及脂肪酸代谢通路。所构建诊断模型在训练集曲线下面积 (area under curve, AUC) 达 0.958, 在 GSE29221 (AUC=1.000) 和 GSE25724 (AUC=0.847) 验证集中均表现优异。**结论** 本研究鉴定的线粒体功能基因 *MRPS10*、*SLC25A5* 和 *TRNT1* 具有 T2DM 诊断潜力, 不仅为 T2DM 诊断提供了新型候选生物标志物, 更进一步解析线粒体功能障碍在 T2DM 发病中的分子机制奠定了基础。

【关键词】 2 型糖尿病; 线粒体基因; 生物信息学; 机器学习

【中图分类号】 R 587.1 **【文献标识码】** A

Identification of mitochondrial feature genes for type 2 diabetes mellitus based on bioinformatics and machine learning

REN Jing¹, LI Yuting¹, LIU Xiaoqin², MENG Xianglong¹

1. College of Traditional Chinese Medicine and Food Engineering, Shanxi University of Chinese Medicine, Jinzhong 030619, Shanxi Province, China

2. College of Pharmacy, Shandong Xiandai University, Jinan 250104, China

Corresponding author: MENG Xianglong, Email: sszywzh@126.com

【Abstract】Objective To screen mitochondrial function-related feature genes in

DOI: 10.12173/j.issn.1004-4337.202411053

基金项目: 山西省留学人员科技活动择优资助项目 (20230034); 山西省回国留学人员科研资助项目 (2023-156); 山西省中医药管理局科研课题 (2023ZYA2012); 山西省应用基础研究计划项目 (20210302124694); 山西中医药大学 2021 年度山西教育厅项目 (2021L364)

通信作者: 孟祥龙, 高级实验师, Email: sszywzh@126.com

<https://slyyx.whuzhmedj.com/>

the skeletal muscle of type 2 diabetes mellitus (T2DM) patients and construct a diagnostic model based on bioinformatics and machine learning techniques, in order to provide new insights for the clinical diagnosis of T2DM. **Methods** T2DM datasets from the Gene Expression Omnibus (GEO) database were integrated and processed using R 4.3.2 software for data normalization and batch effect correction. Differentially expressed genes were screened and intersected with mitochondrial genes from the MitoCarta 3.0 database. Key genes were identified through a combination of three machine learning algorithms, including least absolute shrinkage and selection operator (LASSO) regression, random forest (RF), and support vector machine (SVM). The reliability of feature genes was evaluated using Gaussian mixture models (GMM), and a logistic regression diagnostic model was constructed. The model performance was assessed using the receiver operating characteristic curve (ROC) and externally validated in independent datasets. **Results** A total of 23 T2DM-related mitochondrial differential genes were obtained. *MRPS10*, *SLC25A5*, and *TRNT1* were identified by machine learning algorithm as the core feature genes, which were significantly enriched in oxidative phosphorylation and fatty acid metabolism pathways. The diagnostic model achieved an area under curve (AUC) of 0.958 in the training set, with excellent performance in validation sets GSE29221 (AUC=1.000) and GSE25724 (AUC=0.847). **Conclusion** The mitochondrial function genes *MRPS10*, *SLC25A5*, and *TRNT1* identified in this study demonstrated potential as diagnostic biomarkers for T2DM. These findings not only provide new candidate biomarkers for diagnosing T2DM, but also lay a foundation for in-depth analysis of the molecular mechanisms of mitochondrial dysfunction in the pathogenesis of T2DM.

【Keywords】 Type 2 diabetes mellitus; Mitochondrial gene; Bioinformatics; Machine learning

2 型糖尿病 (type 2 diabetes mellitus, T2DM) 是一种全球范围内广泛流行的代谢性疾病, 其并发症严重影响患者的生存质量, 并导致致残率和死亡率显著上升^[1-2]。根据国际糖尿病联合会 (International Diabetes Federation, IDF) 的预测, 到 2045 年, 全球 T2DM 患者的数量将飙升至 7 亿^[3-6]。T2DM 的主要病因为胰岛素抵抗 (insulin resistance, IR), 长期 IR 和血糖升高对胰岛 β 细胞造成损害, 使得胰岛素的分泌能力逐渐下降, 最终导致多个器官和细胞功能紊乱^[7]。

骨骼肌作为胰岛素作用的关键靶器官, 承担着餐后葡萄糖储存的主要功能, 也是能量平衡、消耗和储存的重要场所, 已作为糖尿病 (diabetes mellitus, DM) 药物开发的新靶点。线粒体是细胞进行有氧呼吸的主要场所, 为细胞的基本生物功能提供能量, 参与维持能量代谢稳态、诱导细胞生存和凋亡、活性氧 (reactive oxygen species, ROS) 的产生, 以及钙的合成和保持钙稳态等人体重要生理功能。研究发现, 线粒体基因突变与 DM 密切相关, 在 DM 早期, IR 已存在, 且线粒体氧化功能缺陷, 表明线粒体可能是 IR 及胰岛 β 细胞分泌功能障碍的影响靶标之一^[8-9]。然而, T2DM 与线粒体功能之间的生物学机制尚未明确,

故需深入探讨线粒体在 T2DM 潜在的生物标志物, 为疾病治疗提供依据。

伴随各种组学技术的高速发展和公共数据库的广泛应用, 生物信息学和机器学习借助计算机及数学方法解析生物学数据, 在流行病学筛选、疾病易感性研究、病因探索、疾病诊断和预后标志物方面展现出巨大潜力^[10]。本研究基于生物信息学中差异分析、富集分析、诊断列线图等和机器学习算法, 筛选 T2DM 线粒体相关特征基因, 探究其与 T2DM 之间的内在关系, 为 T2DM 的临床诊断和治疗提供参考。

1 资料与方法

1.1 数据获取与预处理

本研究数据来源于基因表达综合数据库 (Gene Expression Omnibus, GEO) (<https://www.ncbi.nlm.nih.gov/geo/>)、线粒体基因数据库 MitoCarta 3.0 (<https://www.broadinstitute.org/mitocarta/mitocarta30-inventory-mammalian-mitochondrial-proteins-and-pathways>)、及基因和蛋白相互作用和功能数据库 GeneMANIA (<https://genemania.org>), 采用 R 4.3.2 软件对数据进行整理及分析, 研究技术路线见图 1。

以 T2DM 为关键词在 GEO 数据库中进行

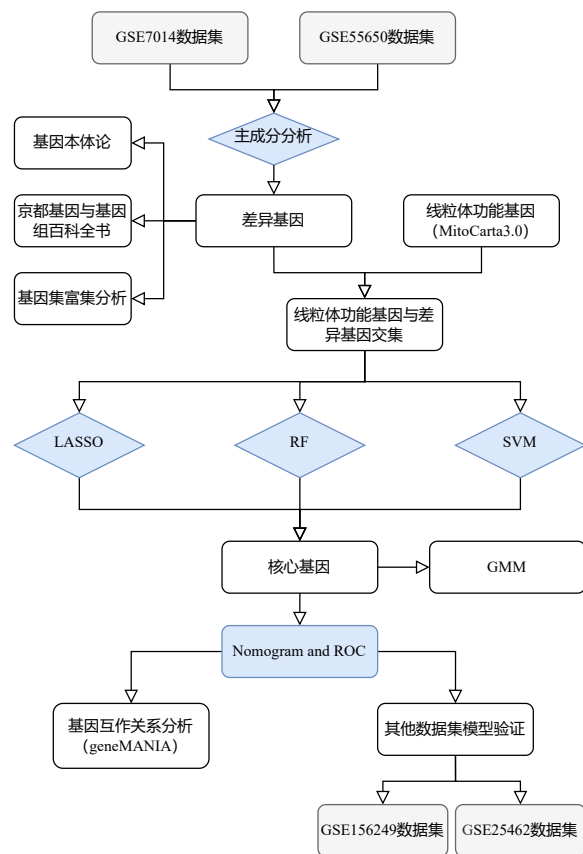


图1 技术路线图

Figure 1. Technology roadmap

注：LASSO, least absolute shrinkage and selection operator, 最小绝对值收敛和选择算子；RF, random forest, 随机森林；SVM, support vector machine, 支持向量机；GMM, Gaussian mixture model, 高斯混合模型；Nomogram, 列线图；ROC, receiver operating characteristic curve, 受试者工作曲线；geneMANIA, Gene Multiple Association Network Integration Algorithm, 基因多重关联网络整合算法数据库。

检索，筛查按照数组进行表达式分析的数据集，并根据以下标准进行筛选：①确保各数据集包含健康对照组与疾病组；②T2DM样本来源为人类骨骼肌组织；③数据集需提供可供进一步分析的原始数据。对 GEO 数据库中下载的 T2DM 数据集 GSE7014 和 GSE55650，运用 R 4.3.2 软件中的 SVA 包进行背景校正、数据整合和归一化分析后，进行主成分分析（principal component analysis, PCA），获得一个综合的 T2DM 表达数据集。

1.2 差异基因筛选

对所获的综合 T2DM 表达数据集进行差异表达基因（differentially expressed genes, DEGs）分析，即利用 R 4.3.2 软件的 limma 包，以 $\log_2FC > 0.5$ 和校正 $P < 0.05$ 为筛选标准得到差异基因，绘制差异基因的火山图。将从 MitoCarta 3.0 数据库得

到的线粒体基因与 T2DM 的差异基因取交集，得到线粒体 T2DM 差异基因（Mito-DEGs），并利用韦恩图进行可视化。

1.3 富集分析

对交集基因利用 R 4.3.2 软件的 clusterProfiler 包从基因的分子功能（molecular function, MF）、细胞组分（cellular component, CC）、生物学过程（biological process, BP）三个角度进行基因本体论（Gene Ontology, GO）、京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes, KEGG）分析和基因集富集分析（Gene Set Enrichment Analysis, GSEA），内参文件为“c5.all.v7.0.symbols.gmt”“c5.cc.v7.0.symbols.gmt”及“c2.cp.reactome.v7.0.symbols.gmt”，且以校正后归一化的富集评分（NES） > 1 ， $P < 0.05$ 为筛选条件，确定与生物标志物相关的信号通路。同时，进一步对交集的上调基因和下调基因进行 KEGG 富集分析，确定其相关的信号通路。

1.4 机器学习筛选核心基因

利用最小绝对值收敛和选择算子（least absolute shrinkage and selection operator, LASSO）回归、随机森林（random forest, RF）和支持向量机（support vector machine, SVM）算法对筛选出来的线粒体和 T2DM 交集基因进一步分析。LASSO 算法使用 R 4.3.2 软件的 glmnet 包，将得到的 Mito-DEGs 模拟总结归纳自变量 T2DM 组 - 正常组和因变量基因之间的关系，从而对样本进行取值，预测 T2DM 患者的结局，选择最佳基因。RF 算法是通过多个决策树结果取最多分类结果作为最终结果，利用 R 4.3.2 软件的 Ran-Forest 包进行随机森林计算，进行 10 折交叉验证，取精确度大于 0.8 的最少基因数量。SVM 算法是通过在特征空间中找到一个最优的超平面来进行分类，其利用 R 4.3.2 软件的 e1071 包进行计算，使用递归特征消除方式，通过减少 SVM 生成的特征向量来优化预测模型，选择 5 倍交叉验证的方式进行核算。

将得到的基因运用韦恩图进行交集基因的选择，以进一步减少算法偏差、提高特征选择的鲁棒性，同时有效避免过拟合，进而提高模型的稳定性和泛化能力，最终得到 T2DM 与线粒体相关的基因（hub Mito-DEGs, hMDs）作后续分析。将得到的基因进行高斯混合模型（Gaussian mixture model, GMM）评估，通过使用高斯分布作为参数

模型的聚类算法，验证最佳基因诊断组合。

1.5 核心基因富集分析

上述 1.4 中筛选得到核心基因，在 GSE7014 和 GSE55650 的合并数据集中的 T2DM 疾病组进行中位数差异分析，并对差异分析结果进行 GSEA 富集分析，内参文件为“h.all.v7.0.symbols.gmt”，并以校正 $P < 0.05$ 为筛选条件，分析核心关键基因与 T2DM 线粒体功能相关的信号通路。

1.6 枢纽基因及其相互作用网络的功能分析

通过 GeneMANIA 数据库可以在线构建基因-蛋白相互作用网络，并预测基因之间的相互关系，分析枢纽基因 hMDs 之间的共表达、共定位及相互作用的结果^[11]。根据得到的 hMDs 在 GeneMANIA 数据库中构建基因-蛋白相互作用网络，并对其进行功能分析。

1.7 诊断模型的建立、评价及验证

对得到的交集基因利用受试者工作特征曲线（receiver operating characteristic curve, ROC）下面积（area under curve, AUC）评价核心枢纽基因的准确性。利用 R 4.3.2 软件的 pROC 包对单基因进行分析，利用 R 包 rms 建立基于特征基因预测 T2DM 风险的列线图（Nomogram），并利用校准曲线估算列线图的预测效率，同时对得到的 hMDs 进行整体 ROC 预测分析。

GMM 可以根据概率分布将数据分类为不同类别并进行建模，用以评估所建立 ROC 模型的效果。利用 R 4.3.2 软件的 GMM 包对所选择的 hMDs 进行分析，得到最优 ROC 曲线值的分类情况。同时，选取 GSE156249 和 GSE25462 两个数据集的样本对枢纽基因进行验证，分别验证枢纽基因作为诊断模型的 AUC 值，以验证诊断模型的有效性。

2 结果

2.1 数据集获取及分析

从 GEO 数据库中获得 T2DM 骨骼肌的数据集

GSE7014、GSE55650、GSE156249 和 GSE25462，选择具有 T2DM 疾病及正常健康人群的数据集进行分析，其中 GSE7014 和 GSE55650 作为主要研究数据，GSE156249 和 GSE25462 作为验证序列，见表 1。在 MitoCarta 3.0 数据库中查找并下载人类线粒体相关基因 1 136 个^[12]。

2.2 2型糖尿病与线粒体相关基因结果

将数据集 GSE7014 与 GSE55650 进行整合，并对整合后的数据进行 PCA 分析。将整合的 T2DM 数据集归一化，其中 T2DM 组包含 32 例 T2DM 患者，对照组 17 例。在去除批次影响后，两个数据集之间的差异显著降低（图 2-A、图 2-B）。对整合的 T2DM 数据集进行 DEGs 分析，鉴定出上调差异基因 399 个、下调差异基因 210 个（图 2-C）。1 136 个线粒体相关基因同 T2DM 差异基因取交集，得到 23 个相关基因（图 2-D），其中 T2DM 上调 13 个、下调 10 个（图 2-E）。

2.3 富集分析结果

GO 分析结果表明，主要涉及有机磷酸酯转运（organophosphate ester transport）、线粒体转运（mitochondrial transport）、核苷酸转运（nucleotide transport）、线粒体跨膜转运（mitochondrial transmembrane transport）、线粒体膜电位的调节（negative regulation of mitochondrion organization）、线粒体腺苷二磷酸跨膜转运（mitochondrial ADP transmembrane transport）、线粒体腺苷三磷酸跨膜转运（mitochondrial ATP transmembrane transport）等（图 3-A）。KEGG 分析结果表明，主要涉及细胞衰老（cellular senescence）、坏死（necroptosis）、cGMP-PKG 信号通路（cGMP-PKG signaling pathway）、核糖体（ribosome）、糖尿病性心肌病（diabetic cardiomyopathy）、脂肪酸代谢（lipoic acid metabolism）等通路（图 3-B）。GSEA 富集分析发现，在线粒体部分（mitochondrial part）、钙粘着蛋白绑定（cadherin binding）、细胞对含氧化合物反应（cellular response to oxygen containing

表1 GEO数据库中2型糖尿病数据集统计

Table 1. Summary of type 2 diabetes mellitus datasets in GEO database

序号	序列号	芯片平台	样本来源	样本情况	物种	分类
1	GSE 7014	GPL 570	骨骼肌	20例T2DM、6例正常对照	人类	主序列
2	GSE 55650	GPL 570	骨骼肌	12例T2DM、11例正常对照	人类	主序列
3	GSE 156249	GPL 11532	骨骼肌	7例T2DM、18例正常对照	人类	验证序列
4	GSE 25462	GPL 570	骨骼肌	10例T2DM、15例正常对照	人类	验证序列

注：T2DM，type 2 diabetes mellitus，2型糖尿病。

compound)、前体代谢物和能量的生成 (generation of precursor metabolites and energy)、先天性免疫系统反应 (reactome innate immune system)、免疫系统反应性细胞因子信号传导 (reactome cytokine signaling in immune system) 及白细胞介素的信号传导 (reactome signaling by interleukins) 等方面存在较明显的富集 (图 3-C—图 3-E)。上调基因的 KEGG 富集分析表明, 主要在糖酵解/糖异生 (Glycolysis / Gluconeogenesis)、柠檬酸循环/三羧酸循环 (Citrate cycle /TCA cycle) 及尼克酸和尼克酰胺代谢 (Nicotinate and nicotinamide metabolism) 等多个信号通路与线粒体功能相关 (图 3-F)。下调基因的 KEGG 富集分析表明, 主要在环磷酸鸟苷-蛋白激酶 G 信号通路 (cGMP-PKG signaling pathway)、糖尿病心肌病 (diabetic cardiomyopathy) 及戊糖和葡萄糖醛酸相互转化 (pentose and glucuronate interconversions) 等与 T2DM 及线粒体相关的通路 (图 3-G)。

2.4 机器学习筛选核心基因结果

将 T2DM 和线粒体取交集得到的 23 个基因进行 LASSO 分析, 得到 *OCIAD1*、*MRPS10*、*TCAIM*、*NIPSNAP3B*、*TSTD1*、*PPM1K*、

SLC25A36、*SLC25A5*、*SLC25A6*、*TRNT1*、*DNLZ*、*GLRX2* 12 个特征基因 (图 4-A、图 4-B)。将 23 个交集基因进行 SVM 算法分析, 得到 *MRPS10*、*TRNT1*、*DNLZ*、*NIPSNAP3B*、*SLC25A5*、*TCAIM*、*OCIAD1*、*PPM1K*、*PRELID1*、*GLRX2*、*SLC25A36*、*COQ10A*、*PAM16*、*NNT*、*MRPL24*、*ALDH5A1*、*DCXR*、*PDPI*、*SLC25A6*、*SYNJ2BP* 20 个特征基因。在正确率验证图中, 20 个特征基因的准确率为 0.871, 表明结果较为可信 (图 4-C); 20 个特征基因的错误率为 0.129, 表明错误率低 (图 4-D)。对其进行 RF 分析, 将第 1 次出现最高精确度时得到的基因作为结果, 分别为 *MRPL24*、*SLC25A5*、*TRNT1*、*MRPS10* (图 4-E)。

将三种算法取交集, 得到 *MRPS10*、*SLC25A5* 和 *TRNT1* 三个 T2DM 和线粒体特征基因 (图 4-F), GMM 验证得到 7 种模型, 评估最优 ROC 诊断值, 当基因数为 3 时, 即 *MRPS10*、*SLC25A5* 和 *TRNT1*, 其 AUC 值最高 (AUC=0.957 7) (图 4-G)。

2.5 核心基因的基因集富集分析

对差异基因进行 GSEA 富集分析, 结果显示, *MRPS10*、*SLC25A5* 和 *TRNT1* 基因在氧化磷酸化

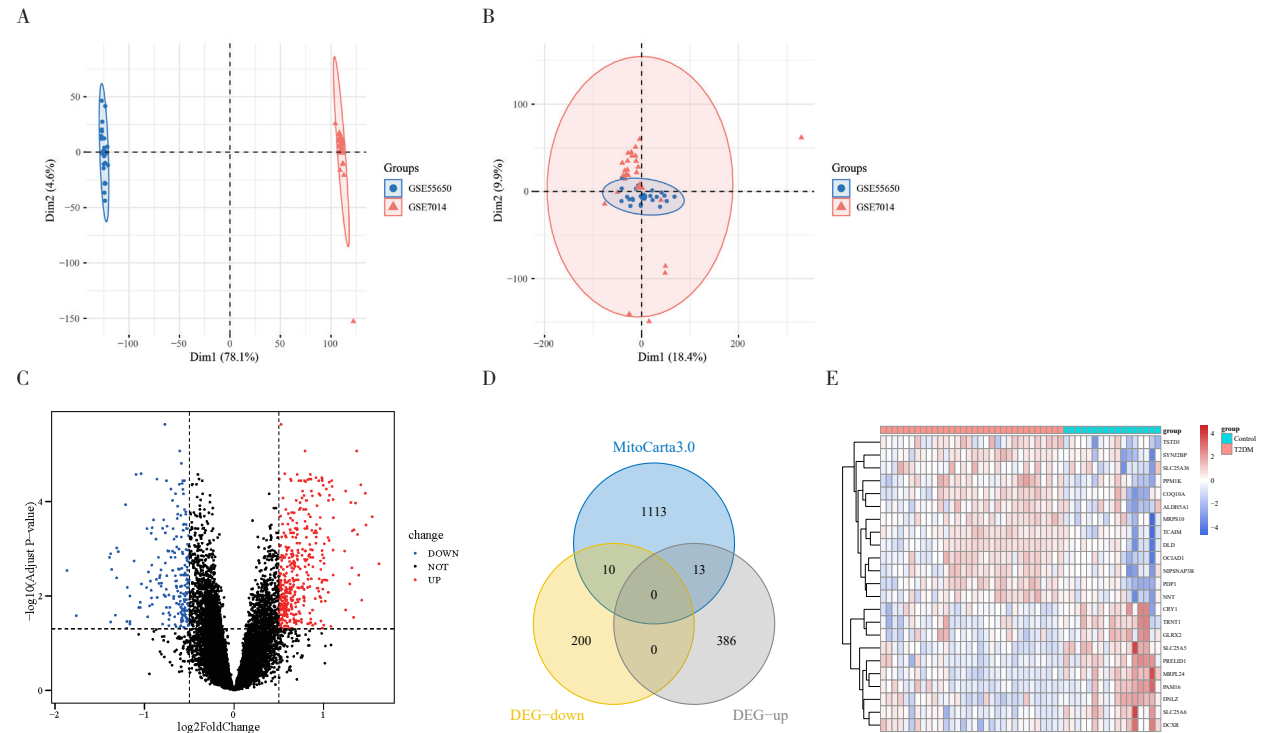


图2 2型糖尿病与线粒体基因筛选图

Figure 2. Screening map of type 2 diabetes mellitus and mitochondrial genes

注: A. 批次合并前主成分分析图; B. 批次合并后主成分分析图; C. 2型糖尿病差异基因分析火山图 (红色标记代表上调基因, 蓝色代表下调基因); D. 线粒体基因与2型糖尿病差异基因分析韦恩图; E. 23个交集基因热图。

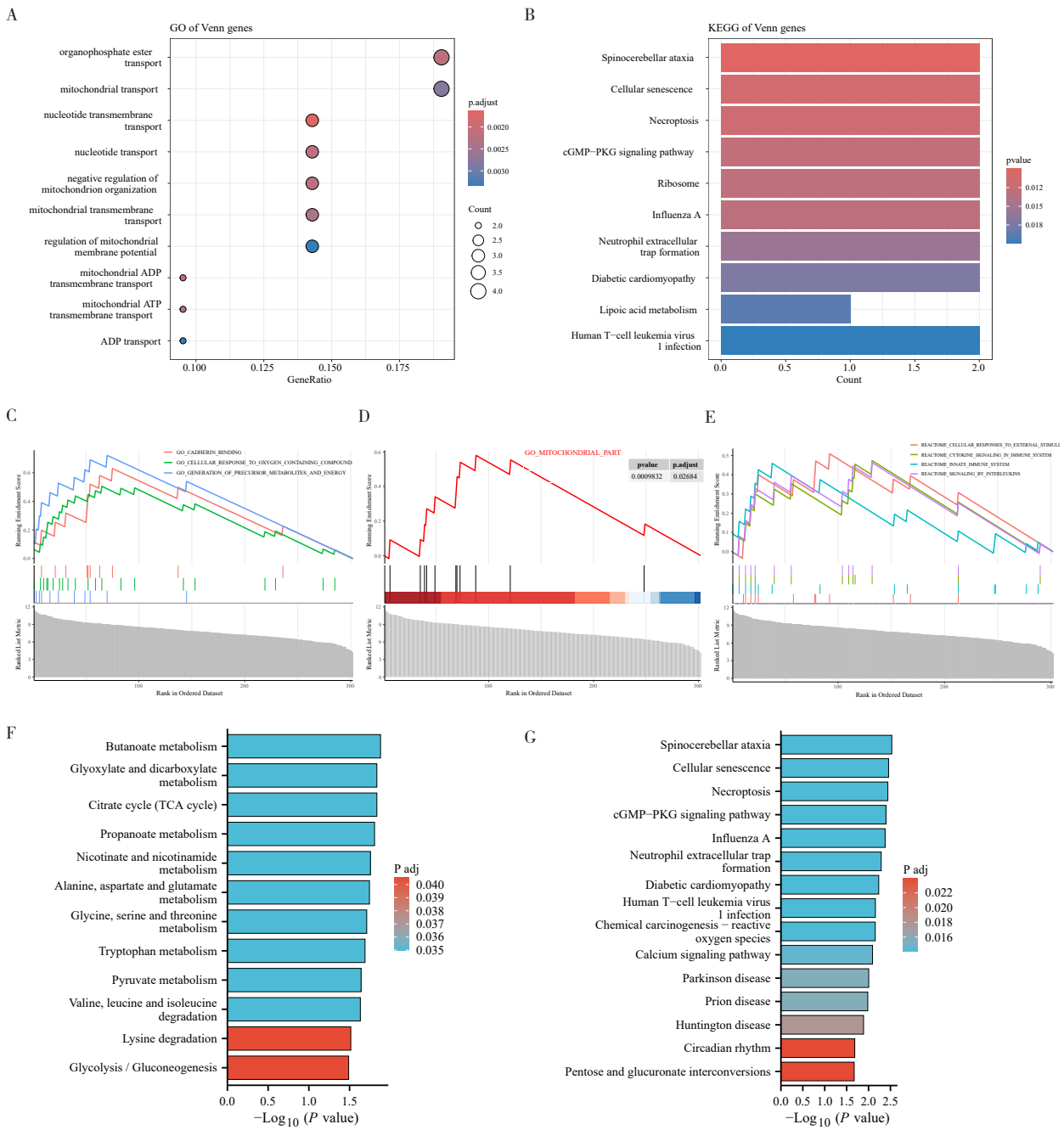


图3 富集分析

Figure 3. Enrichment analysis

注：A. 基因本体论（Gene Ontology, GO）富集分析；B. 京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes, KEGG）富集分析；C. 基因集富集分析（Gene Set Enrichment Analysis, GSEA）1；D. 基因集富集分析2；E. 基因集富集分析3；F. 13个上调基因的KEGG富集分析；G. 10个下调基因的KEGG富集分析。

（oxidative phosphorylation）、P53信号通路 [tumor protein 53 (TP53) pathway]、凋亡（apoptosis）、PI3K/AKT/MTOR信号通路 [phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT)/mechanistic target of rapamycin (mTOR) pathway]、活性氧代谢（reactive oxygen species pathway）及肿瘤坏死因子 α 通过核因子 κ B信号传导（tumor necrosis factor alpha

signaling via nuclear factor kappa-light-chain-enhancer of activated B cells）等众多通路中有显著意义（图5）。

2.6 枢纽基因相互作用分析结果

在 GeneMANIA 数据库对 *MRPS10*、*SLC25A5* 和 *TRNT1* 这三个枢纽基因及其相互作用的基因进行功能分析，结果显示，三者的主要功能体现在线粒体核糖体、器质核糖体、线粒体基因表达、

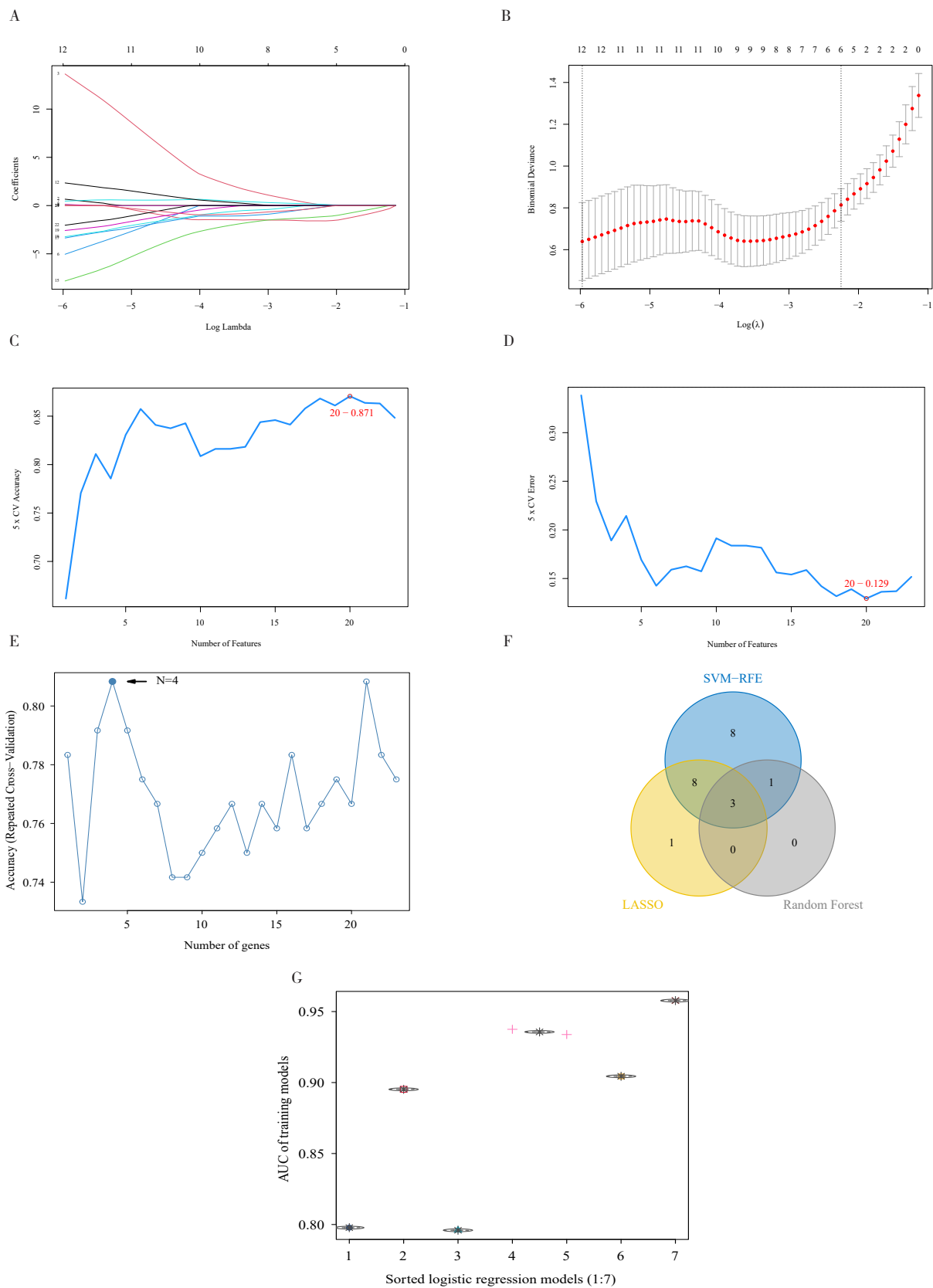


图4 利用机器学习算法识别特征基因

Figure 4. Identification of feature genes using machine learning algorithms

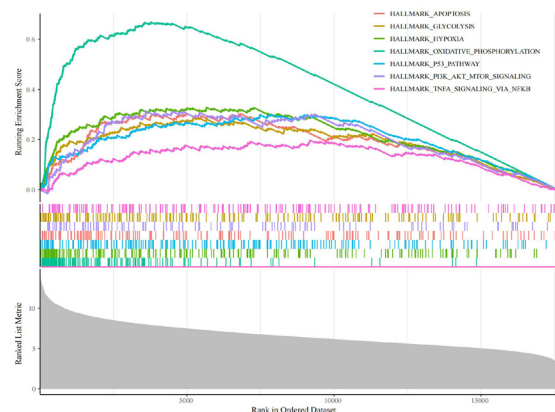
注：A. 最小绝对值收敛和选择算子（least absolute shrinkage and selection operator, LASSO）模型中最优参数的选择；B. 最优值确定12个特征基因分布；C. 支持向量机（support vector machine, SVM）模型准确性；D. 支持向量机（support vector machine, SVM）模型误差值；E. 随机森林（random forest, RF）分析基因数目；F. Venn共有基因选择；G. 高斯混合模型（Gaussian mixture model, GMM）分析。

线粒体翻译、线粒体蛋白复合体、核糖体亚基、胰岛素分泌等上,表明三个枢纽基因与 T2DM 和线粒体功能密切相关(图 6)。

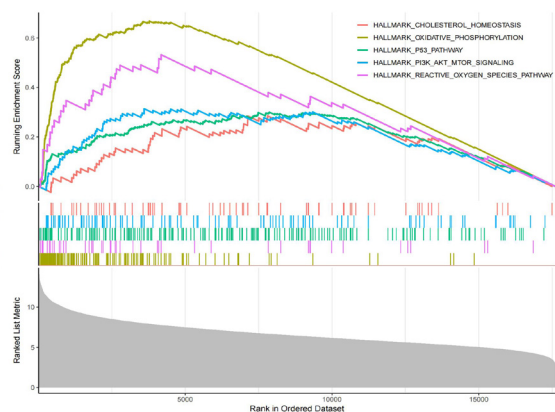
2.7 诊断模型构建及评价

在 T2DM 数据集中,三个枢纽基因 *MRPS10*、*SLC25A5* 和 *TRNT1* 作为诊断模型进行 ROC 评价,

A



B



C



图5 基因集富集分析结果

Figure 5. Results of Gene Set Enrichment Analysis

注: A. *MRPS10* 在疾病组中差异基因基因集富集分析结果; B. *SLC25A5* 在疾病组中差异基因基因集富集分析结果; C. *TRNT1* 在疾病组中差异基因基因集富集分析结果。

MRPS10 (AUC=0.798)、*SLC25A5* (AUC=0.895) 和 *TRNT1* 基因 (AUC=0.796) 均表现出良好的诊断价值(图 7-A—图 7-C)。同时,构建了基于三个枢纽基因的列线图模型以预测 T2DM 发病风险(图 7-D),每个基因均对应一个评分标准,为临床筛查和预测提供一定依据,且列线图的校准曲线验证了模型的预测性能较好(图 7-E)。最后,分析得到三个枢纽基因为诊断模型的 ROC 曲线,以评价列线图的预测能力,该模型总体 AUC 值为 0.958,说明枢纽基因具有较高的诊断价值(图 7-F)。

此外,应用逻辑回归计算另外两个独立的数据集 GSE156249 和 GSE25462 以验证模型的准确性,得到三个基因作为诊断模型的评判效果, GSE156249 数据集的 AUC 值为 1.000(图 7-G),而 GSE25462 数据集的 AUC 值为 0.847(图 7-H),均有较好的诊断性,说明该模型对 T2DM 的诊断较为稳定。

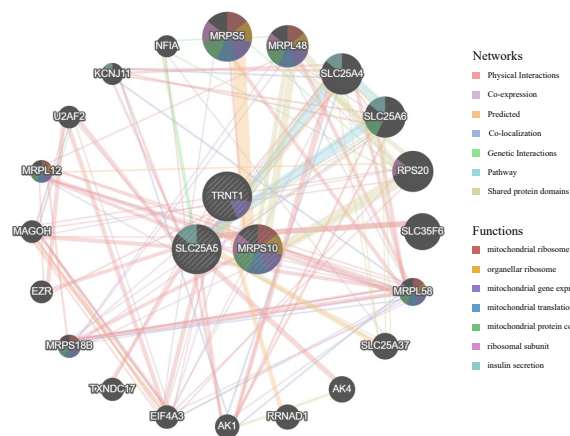


图6 基因相互作用图

Figure 6. Gene interaction map

3 讨论

骨骼肌是 IR 重要的靶组织之一^[13],调节全身血糖功能的发挥,与 T2DM 及其并发症的发生发展有密切的关系,具体的生物学机制涉及脂质异位沉积^[14-15]、氧化应激及内质网应激^[16]、晚期糖基化终末产物累积^[17]、神经病变^[18]及肾病^[19]等。其中以骨骼肌的线粒体功能最为突出,主要通过碳水化合物和脂质的酶氧化获得主要能量,而线粒体 ATP 的产生对骨骼肌细胞的功能和结构完整性至关重要^[20]。研究表明,通过不同的机制抑制胰岛素信号传导,会导致骨骼

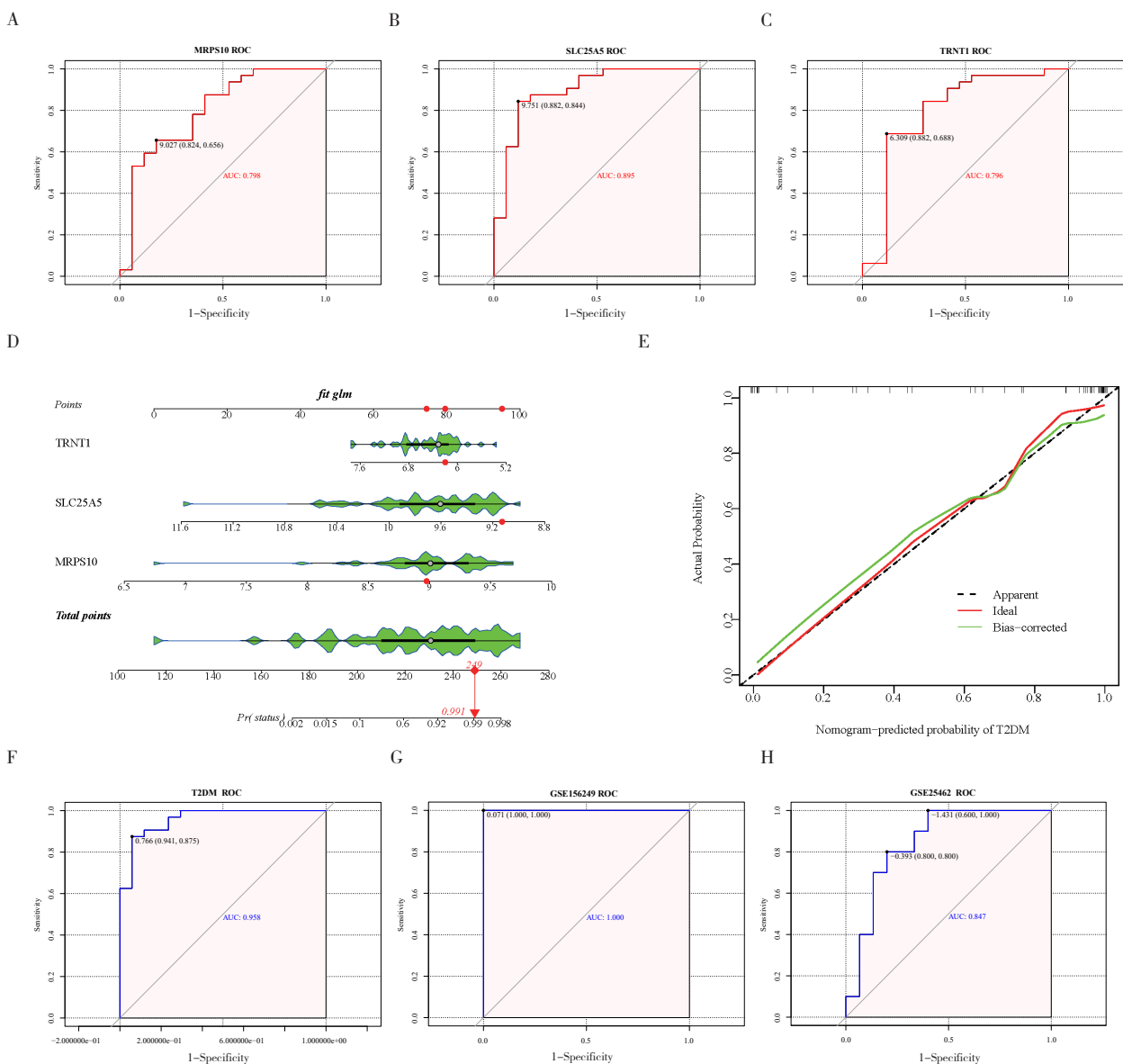


图7 列线图及受试者工作特征曲线评价

Figure 7. Nomogram and receiver operating characteristic curve evaluation

注: A. *MRPS10* 的受试者工作特征曲线诊断图; B. *SLC25A5* 的受试者工作特征曲线诊断图; C. *TRNT1* 的受试者工作特征曲线诊断图; D. 三个基因的Nomogram模型图; E. Nomogram校准曲线; F. T2DM的受试者工作特征曲线诊断图; G. GSE156249的受试者工作特征曲线验证图; H. GSE25462的受试者工作特征曲线验证图。

肌的氧化能力降低和线粒体功能受损^[21-22]。同时,葡萄糖和脂肪酸代谢紊乱会导致脂肪酸和糖酵解中间体在骨骼肌细胞中积累,进一步激活蛋白激酶C、核因子- κ B等相关通路,从而干扰胰岛素信号传导^[23-25]。由于氧化磷酸化降低而导致的ATP消耗减少,会进一步损害参与胰岛素信号转导的许多酶的催化活性,从而加剧IR^[26]。除产生能量外,线粒体还是ROS的主要来源,骨骼肌线粒体功能障碍会导致细胞内ROS负担增加,从而导致骨骼肌IR^[27]。此外,

IR和高血糖也可能导致线粒体功能障碍,形成恶性循环,进而损害葡萄糖稳态^[28]。

本研究表明,*MRPS10*、*SLC25A5*和*TRNT1*在T2DM中显著影响细胞凋亡、PI3K/AKT/MTOR信号通路、氧化磷酸化和脂肪酸代谢等通路,这些通路与线粒体功能有关^[29],进而可能导致IR,从而加剧T2DM及其并发症的发生^[30]。此外,构成膜转运蛋白的超级蛋白家族中的*SLC25A5*在骨骼肌中高度广泛表达,能够将ADP导入线粒体基质进行ATP合成,同时将

ATP 输出到细胞质中^[31]。*MRPS10* 属于线粒体核糖体蛋白 (MRPs)，是线粒体复合体结构和功能完整性的重要组成部分，其异常表达会对线粒体核糖产生应激反应，导致线粒体功能障碍，影响骨骼肌质量和功能，从而导致 IR 的加剧^[32]。*TRNT1* 最常见临床表型是常染色体隐性铁粒幼细胞性贫血、周期性发热和发育迟缓等^[33]，且与骨骼肌病变有关，其部分功能丧失或异常表达会对线粒体功能造成影响，引起线粒体肌病，推断可能会导致骨骼肌异常^[34]。对 *MRPS10*、*SLC25A5* 和 *TRNT1* 这三个线粒体功能基因进行 GSEA 单基因富集分析发现，三个基因的异常可能损害氧化磷酸化过程，导致 ATP 合成减少和活性氧积累^[34-36]。本研究通过对 T2DM 异常线粒体基因进行分析，发现 *MRPS10*、*SLC25A5* 和 *TRNT1* 可能作为 T2DM 诊断的标志物，具有良好的 ROC 诊断评价效果；通过基因交互分析发现，三者在线粒体核蛋白、胰岛素表达、细胞器核糖体、线粒体基因表达等方面存在交互作用，可能对骨骼肌中线粒体功能产生一定影响，从而导致 IR，进而延长 T2DM 病程。

本研究存在一定局限性：首先，本研究基于公共 GEO 数据库的样本分析，未能涵盖临床样本的异质性；其次，研究主要采用生物信息学和机器学习方法，缺乏实验来证实研究中识别出的 *MRPS10*、*SLC25A5* 和 *TRNT1* 基因在 T2DM 中的作用；最后，本研究聚焦于有限的线粒体基因，可能忽视其他涉及 T2DM 复杂病理生理学的关键因素和途径。未来的研究应将实验验证与计算分析相结合，以增强结果的稳健性，更全面地阐明线粒体基因在 T2DM 病理病程发展过程中的分子机制。

综上所述，本研究构建了基于 *MRPS10*、*SLC25A5* 和 *TRNT1* 三个基因的诊断模型，探讨了 T2DM 线粒体共同核心基因 *MRPS10*、*SLC25A5* 和 *TRNT1* 的潜在功能机制，揭示了其可能在骨骼肌中影响线粒体功能并影响血糖稳定，为进一步深入研究 T2DM 线粒体基因表达及诊断提供了依据。

参考文献

1 Luo JQ, Shu Y, Zhang W. Editorial: pharmacogenomics and pharmacomicrobiomics in type 2 diabetes mellitus (T2DM)[J].

- Front Endocrinol (Lausanne), 2023, 14: 1287807. DOI: [10.3389/fendo.2023.1287807](https://doi.org/10.3389/fendo.2023.1287807).
- 2 王富军, 王文琦. 《中国 2 型糖尿病防治指南 (2020 年版)》解读[J]. 河北医科大学学报, 2021, 24(12): 1365-1371. [Wang FJ, Wang WQ. Interpretation of the "China guidelines for the prevention and treatment of type 2 diabetes (2020 edition)"[J]. Journal of Hebei Medical University, 2021, 24(12): 1365-1371.] DOI: [10.3969/j.issn.1007-3205.2022.12.001](https://doi.org/10.3969/j.issn.1007-3205.2022.12.001).
- 3 Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition[J]. Diabetes Res Clin Pract, 2019, 157: 107843. DOI: [10.1016/j.diabres.2019.107843](https://doi.org/10.1016/j.diabres.2019.107843).
- 4 Wang L, Peng W, Zhao Z, et al. Prevalence and treatment of diabetes in China, 2013-2018[J]. JAMA, 2021, 326(24): 2498-2506. DOI: [10.1001/jama.2021.22208](https://doi.org/10.1001/jama.2021.22208).
- 5 Dascalu AM, Serban D, Tanasescu D, et al. The value of white cell inflammatory biomarkers as potential predictors for diabetic retinopathy in type 2 diabetes mellitus (T2DM)[J]. Biomedicines, 2023, 11(8): 2106. DOI: [10.3390/biomedicines11082106](https://doi.org/10.3390/biomedicines11082106).
- 6 Cui K, Li Z. Identification and analysis of type 2 diabetes-mellitus-associated autophagy-related genes[J]. Front Endocrinol (Lausanne), 2023, 14: 1164112. DOI: [10.3389/fendo.2023.1164112](https://doi.org/10.3389/fendo.2023.1164112).
- 7 Ma L, Li Y. Cognitive function and insulin resistance in elderly patients with type 2 diabetes[J]. Neurol Res, 2017, 39(3): 259-263. DOI: [10.1080/01616412.2017.1281199](https://doi.org/10.1080/01616412.2017.1281199).
- 8 官伟康, 吕静, 杨朝鲜. 线粒体质量控制新途径: 线粒体衍生囊泡[J]. 解剖学报, 2021, 52(1): 152-156. [Guan WK, Lyu J, Yang CX. A new pathway of mitochondrial quality control: mitochondrial-derived vesicles[J]. Acta Anatomica Sinica, 2021, 52(1): 152-156.] DOI: [10.16098/j.issn.0529-1356.2021.01.025](https://doi.org/10.16098/j.issn.0529-1356.2021.01.025).
- 9 杨迎, 夏维, 庞志宇, 等. 2 型糖尿病患者线粒体呼吸链复合体 I 活性的临床观察[J]. 中国糖尿病杂志, 2016, 24(8): 721-725. [Yang Y, Xia W, Pang ZY, et al. Observational study of mitochondrial complex I kinetic activity in type 2 diabetes[J]. Chinese Journal of Diabetes, 2016, 24(8): 721-725.] DOI: [10.3969/j.issn.1006-6187.2016.08.013](https://doi.org/10.3969/j.issn.1006-6187.2016.08.013).
- 10 张鼎, 赵亚双. 生物信息学在分子流行病学中的应用[J]. 中华疾病控制杂志, 2021, 25(1): 20-24. [Zhang D, Zhao YS. Applications of bioinformatics in molecular epidemiology[J]. Chinese Journal of Disease Control and Prevention, 2021, 25(1): 20-24.] DOI: [10.16462/j.cnki.zhjbkz.2021.01.005](https://doi.org/10.16462/j.cnki.zhjbkz.2021.01.005).
- 11 Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function[J]. Nucleic Acids Res, 2010, 38(Web Server issue): W214-W220. DOI: [10.1093/nar/gkq537](https://doi.org/10.1093/nar/gkq537).
- 12 Rath S, Sharma R, Gupta R, et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations[J]. Nucleic Acids Res, 2021, 49(D1): D1541-D1547. DOI: [10.1093/nar/gkaa1011](https://doi.org/10.1093/nar/gkaa1011).
- 13 Xu K, Feng X, Xu Z, et al. Association of sarcopenia with

- osteoporosis in Chinese patients with type 2 diabetes[J]. BMC Musculoskelet Disord, 2024, 25(1): 226. DOI: [10.1186/s12891-024-07323-2](https://doi.org/10.1186/s12891-024-07323-2).
- 14 Dlamini M, Khathi A. Prediabetes-associated changes in skeletal muscle function and their possible links with diabetes: a literature review[J]. Int J Mol Sci, 2023, 25(1): 469. DOI: [10.3390/ijms25010469](https://doi.org/10.3390/ijms25010469).
- 15 Zhu S, Tian Z, Torigoe D, et al. Aging- and obesity-related perimyscular adipose tissue accelerates muscle atrophy[J]. PLoS One, 2019, 14(8): e0221366. DOI: [10.1371/journal.pone.0221366](https://doi.org/10.1371/journal.pone.0221366).
- 16 Arab Sadeghabadi Z, Abbasalipourkabir R, Mohseni R, et al. Investigation of oxidative stress markers and antioxidant enzymes activity in newly diagnosed type 2 diabetes patients and healthy subjects, association with IL-6 level[J]. J Diabetes Metab Disord, 2019, 18(2): 437-443. DOI: [10.1007/s40200-019-00437-8](https://doi.org/10.1007/s40200-019-00437-8).
- 17 Yamamoto M, Sugimoto T. Advanced glycation end products, diabetes, and bone strength[J]. Curr Osteoporos Rep, 2016, 14(6): 320-326. DOI: [10.1007/s11914-016-0332-1](https://doi.org/10.1007/s11914-016-0332-1).
- 18 Tesfaye S, Selvarajah D. Advances in the epidemiology, pathogenesis and management of diabetic peripheral neuropathy[J]. Diabetes Metab Res Rev, 2012, 28 Suppl 1: 8-14. DOI: [10.1002/dmrr.2239](https://doi.org/10.1002/dmrr.2239).
- 19 Fung FY, Koh YLE, Malhotra R, et al. Prevalence of and factors associated with sarcopenia among multi-ethnic ambulatory older Asians with type 2 diabetes mellitus in a primary care setting[J]. BMC Geriatrics, 2019, 19(1): 122. DOI: [10.1186/s12877-019-1137-8](https://doi.org/10.1186/s12877-019-1137-8).
- 20 Fiorentino TV, Monroy A, Kamath S, et al. Pioglitazone corrects dysregulation of skeletal muscle mitochondrial proteins involved in ATP synthesis in type 2 diabetes[J]. Metabolism, 2021, 114: 154416. DOI: [10.1016/j.metabol.2020.154416](https://doi.org/10.1016/j.metabol.2020.154416).
- 21 Koves TR, Ussher JR, Noland RC, et al. Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance[J]. Cell Metab, 2008, 7(1): 45-56. DOI: [10.1016/j.cmet.2007.10.013](https://doi.org/10.1016/j.cmet.2007.10.013).
- 22 Petersen KF, Dufour S, Befroy D, et al. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes[J]. N Engl J Med, 2004, 350(7): 664-671. DOI: [10.1056/NEJMoa031314](https://doi.org/10.1056/NEJMoa031314).
- 23 Sriwijitkamol A, Christ-Roberts C, Berria R, et al. Reduced skeletal muscle inhibitor of kappaB content is associated with insulin resistance in subjects with type 2 diabetes: reversal by exercise training[J]. Diabetes, 2006, 55(3): 760-767. DOI: [10.2337/diabetes.55.03.06.db05-0677](https://doi.org/10.2337/diabetes.55.03.06.db05-0677).
- 24 Tantiwong P, Shanmugasundaram K, Monroy A, et al. NF- κ B activity in muscle from obese and type 2 diabetic subjects under basal and exercise-stimulated conditions[J]. Am J Physiol Endocrinol Metab, 2010, 299(5): E794-E801. DOI: [10.1152/ajpendo.00776.2009](https://doi.org/10.1152/ajpendo.00776.2009).
- 25 Reyna SM, Ghosh S, Tantiwong P, et al. Elevated toll-like receptor 4 expression and signaling in muscle from insulin-resistant subjects[J]. Diabetes, 2008, 57(10): 2595-2602. DOI: [10.2337/db08-0038](https://doi.org/10.2337/db08-0038).
- 26 Kang J, Heart E, Sung CK. Effects of cellular ATP depletion on glucose transport and insulin signaling in 3T3-L1 adipocytes[J]. Am J Physiol Endocrinol Metab, 2001, 280(3): E428-E435. DOI: [10.1152/ajpendo.2001.280.3.E428](https://doi.org/10.1152/ajpendo.2001.280.3.E428).
- 27 Wei Y, Chen K, Whaley-Connell AT, et al. Skeletal muscle insulin resistance: role of inflammatory cytokines and reactive oxygen species[J]. Am J Physiol Regul Integr Comp Physiol, 2008, 294(3): R673-R680. DOI: [10.1152/ajpregu.00561.2007](https://doi.org/10.1152/ajpregu.00561.2007).
- 28 Nishikawa T, Edelstein D, Du XL, et al. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage[J]. Nature, 2000, 404(6779): 787-790. DOI: [10.1038/35008121](https://doi.org/10.1038/35008121).
- 29 Ramasubbu K, Devi Rajeswari V. Impairment of insulin signaling pathway PI3K/Akt/mTOR and insulin resistance induced AGEs on diabetes mellitus and neurodegenerative diseases: a perspective review[J]. Mol Cell Biochem, 2023, 478(6): 1307-1324. DOI: [10.1007/s11010-022-04587-x](https://doi.org/10.1007/s11010-022-04587-x).
- 30 Bonnard C, Durand A, Peyrol S, et al. Mitochondrial dysfunction results from oxidative stress in the skeletal muscle of diet-induced insulin-resistant mice[J]. J Clin Invest, 2008, 118(2): 789-800. DOI: [10.1172/JCI32601](https://doi.org/10.1172/JCI32601).
- 31 Le J, Chen Y, Yang W, et al. Metabolic basis of solute carrier transporters in treatment of type 2 diabetes mellitus[J]. Acta Pharm Sin B, 2024, 14(2): 437-454. DOI: [10.1016/j.apsb.2023.09.004](https://doi.org/10.1016/j.apsb.2023.09.004).
- 32 Garros RF, Paul R, Connolly M, et al. MicroRNA-542 promotes mitochondrial dysfunction and SMAD activity and is elevated in intensive care unit-acquired weakness[J]. Am J Respir Crit Care Med, 2017, 196(11): 1422-1433. DOI: [10.1164/rccm.201701-0101OC](https://doi.org/10.1164/rccm.201701-0101OC).
- 33 Slade A, Kattini R, Campbell C, et al. Diseases associated with defects in tRNA CCA addition[J]. Int J Mol Sci, 2020, 21(11): 3780. DOI: [10.3390/ijms21113780](https://doi.org/10.3390/ijms21113780).
- 34 Fatica T, Naas T, Liwak U, et al. TRNT-1 deficiency is associated with loss of tRNA integrity and imbalance of distinct proteins[J]. Genes (Basel), 2023, 14(5): 1043. DOI: [10.3390/genes14051043](https://doi.org/10.3390/genes14051043).
- 35 Gao C, Zhang C, Wen L, et al. Regulation of reactive oxygen species and the role of mitochondrial apoptotic-related genes in rheumatoid arthritis[J]. Sci Rep, 2025, 15(1): 2165. DOI: [10.1038/s41598-025-85460-x](https://doi.org/10.1038/s41598-025-85460-x).
- 36 Zhu S, Wang W, Zhang J, et al. Slc25a5 regulates adipogenesis by modulating ERK signaling in OP9 cells[J]. Cell Mol Biol Lett, 2022, 27(1): 11. DOI: [10.1186/s11658-022-00314-y](https://doi.org/10.1186/s11658-022-00314-y).

收稿日期: 2024 年 11 月 12 日 修回日期: 2025 年 02 月 09 日
本文编辑: 张 苗 黄 笛

引用本文: 任婧, 李宇婷, 刘晓琴, 等. 基于生物信息学和机器学习识别2型糖尿病线粒体特征基因[J]. 数理医药学杂志, 2025, 38(4): 237-247. DOI: [10.12173/j.issn.1004-4337.202411053](https://doi.org/10.12173/j.issn.1004-4337.202411053).
Ren J, Li YT, Liu XQ, et al. Identification of mitochondrial feature genes for type 2 diabetes mellitus based on bioinformatics and machine learning[J]. Journal of Mathematical Medicine, 2025, 38(4): 237-247. DOI: [10.12173/j.issn.1004-4337.202411053](https://doi.org/10.12173/j.issn.1004-4337.202411053).