

## · 论著 ·

# 基于生物信息学分析方法和机器学习算法探究铁死亡参与糖尿病心脏病的机制

扫描二维码  
查看更多杨启帆<sup>1</sup>, 方柔柔<sup>1</sup>, 邬东东<sup>1</sup>, 骆家引<sup>2</sup>, 徐守竹<sup>1</sup>, 赵晶<sup>3</sup>

**【摘要】** 目的 基于生物信息学分析方法和机器学习算法探究铁死亡参与糖尿病心脏病(DHD)的机制。方法 从美国国家生物技术信息中心(NCBI)基因表达综合(GEO)数据库下载数据集GSE4745(大鼠心肌细胞测序数据)和数据集GSE26887(人心肌细胞测序数据),通过铁死亡数据库获得铁死亡基因数据集。应用R软件(4.2.1版本)中的“Limma”包分析数据集GSE4745中DHD的差异表达基因;应用“WGCNA”包构建加权基因共表达网络以筛选DHD相关基因模块;应用“Venn”包取DHD差异表达基因、DHD相关基因模块及铁死亡基因数据集交集以获取DHD相关铁死亡基因。然后针对DHD相关铁死亡基因进行LASSO回归分析以筛选DHD相关铁死亡核心基因。最后通过数据集GSE26887、蛋白质组学原始数据及既往研究验证DHD相关铁死亡核心基因。结果 从数据集GSE4745中获得491个DHD差异表达基因,其中上调基因252个、下调基因239个。以软阈值为7将DHD差异表达基因构建邻接矩阵,最终选择品红色和蓝色模块为DHD相关基因模块,共1 092个基因。进一步分析获得28个DHD相关铁死亡基因。LASSO回归分析结果显示,从28个DHD相关铁死亡基因中获得6个DHD相关铁死亡核心基因,分别为H19、Nr4a1、Decr1、Gstm1、Slc3a2、Por。根据数据集GSE26887、蛋白质组学原始数据及既往研究结果最终确定H19、Nr4a1、Decr1、Por为DHD相关铁死亡的核心基因。结论 铁死亡参与DHD的机制可能与H19、Decr1、Por表达上调及Nr4a1表达下调有关,这为铁死亡参与DHD提供了新的证据。

**【关键词】** 糖尿病血管病变;糖尿病心脏病;铁死亡;生物信息学;LASSO回归

**【中图分类号】** R 587.23 **【文献标识码】** A **DOI:** 10.12114/j.issn.1008-5971.2023.00.271

## Mechanism of Ferroptosis Participating in Diabetic Heart Disease Based on Bioinformatics Analysis Methods and Machine Learning Algorithms YANG Qifan<sup>1</sup>, FANG Rourou<sup>1</sup>, WU Dongdong<sup>1</sup>, LUO Jiayin<sup>2</sup>, XU Shouzhuzhu<sup>1</sup>, ZHAO Jing<sup>3</sup>

1.School of Public Health of Shaanxi University of Chinese Medicine, Xianyang 712046, China

2.The Second Clinical Medical College of Shaanxi University of Chinese Medicine, Xianyang 712046, China

3.Master Institute of Traditional Chinese Medicine of Shaanxi University of Chinese Medicine, Xianyang 712046, China

Corresponding author: XU Shouzhuzhu, E-mail: xsz@sntcm.edu.cn; ZHAO Jing, E-mail: zhaojing\_1207@126.com

**【Abstract】 Objective** To explore the mechanism of ferroptosis participating in diabetic heart disease (DHD) based on bioinformatics analysis methods and machine learning algorithms. **Methods** The dataset GSE4745 (rat cardiomyocyte sequencing data) and dataset GSE26887 (human cardiomyocyte sequencing data) were downloaded from the Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI). The ferroptosis gene dataset were selected from ferroptosis database. The "limma" package in R software (version 4.2.1) was used to analyze the differentially expressed genes of DHD in the dataset GSE4745; the "WGCNA" package was used to construct a weighted gene coexpression network to screen DHD related gene modules; the "Venn" package was used to package the intersection of differentially expressed genes, DHD-related gene modules and ferroptosis gene datasets, that DHD-related ferroptosis gene. Then LASSO regression analysis was performed on DHD-related ferroptosis genes to screen DHD-related ferroptosis core genes. Finally, the DHD related ferroptosis core genes was verified by dataset GSE26887, proteomics raw data and previous studies. **Results** A total of 491 DHD differentially expressed genes were obtained from the dataset GSE4745, including 252 up-regulated genes and 239 down regulated genes. DHD differentially expressed genes were constructed into an adjacency matrix with a soft threshold of 7. Finally, magenta and blue modules were selected as DHD related gene modules, with a total of 1 092 genes. Further analysis obtained DHD related ferroptosis genes. LASSO

基金项目: 国家自然科学基金资助项目(82100488, 82105016); 陕西省科技厅重点研发计划项目(2021SF-071, 2022SF-318); 陕西省教育厅重点科研计划项目(21JS012); 国家级大学生创新创业训练计划项目(202110716027)

作者单位: 1.712046陕西省咸阳市, 陕西中医药大学公共卫生学院 2.712046陕西省咸阳市, 陕西中医药大学第二临床医学院 3.712046陕西省咸阳市, 陕西中医药大学陕西省针药结合重点实验室

通信作者: 徐守竹, E-mail: xsz@sntcm.edu.cn 赵晶, E-mail: zhaojing\_1207@126.com

regression analysis results showed that six DHD related ferroptosis core genes were obtained from 28 DHD related ferroptosis genes, which were H19, Nr4a1, Decr1, Gstm1, Slc3a2 and Por, respectively. H19, Nr4a1, Decr1 and Por were finally identified as the DHD related ferroptosis core genes according to the dataset GSE26887, the proteomics raw data and previous research results.

**Conclusion** The mechanism of ferroptosis participating in DHD may be related to the up regulation of H19, DECR1 and Por expression and the down regulation of NR4A1 expression, which provides new evidence for ferroptosis participating in DHD.

**【 Key words 】** Diabetic angiopathies; Diabetic heart disease; Ferroptosis; Bioinformatics; LASSO regression

近年来我国糖尿病的发病率急剧升高，患者人数高达1.16亿，已成为威胁居民健康的主要疾病之一<sup>[1]</sup>。糖尿病可导致多种并发症，如糖尿病心血管疾病、糖尿病足、神经病变等<sup>[2]</sup>，其中糖尿病心脏病（diabetic heart disease, DHD）是引起糖尿病患者死亡的主要原因之一<sup>[3]</sup>。目前，DHD的发病机制尚未完全阐明，仍缺乏有效的治疗策略和药物。研究发现，调控铁死亡可以延缓DHD的发生<sup>[4]</sup>。但对于铁死亡与DHD的关系尚缺乏深入研究。近年来随着生物信息技术的成熟和机器学习理论体系的发展，生物信息学分析方法和机器学习算法被广泛用于揭示疾病的发生、发展<sup>[5]</sup>。本研究基于生物信息学分析方法和机器学习算法探究铁死亡参与DHD的机制，以期寻找DHD新的治疗靶点提供理论依据。

### 1 资料与方法

**1.1 数据来源** 从美国国家生物技术信息中心（National Center for Biotechnology Information, NCBI）基因表达综合（Gene Expression Omnibus, GEO）数据库（<https://www.ncbi.nlm.nih.gov/geo/>）下载数据集GSE4745（大鼠心肌细胞测序数据）和数据集GSE26887（人心肌细胞测序数据），并根据平台文件对探针进行注释。利用R软件对数据集进行矫正、标准化处理、缺失值补充等，并对基因表达矩阵数据进行对数转换。其中数据集GSE4745包含正常组样本4个和DHD组样本4个，数据集GSE26887包含正常组样本5个和DHD组样本7个。通过铁死亡数据库（<http://www.zhounan.org/ferrdb/current/>）获取铁死亡基因数据集。

### 1.2 研究方法

#### 1.2.1 DHD差异表达基因及DHD相关基因模块筛选

应用R软件（4.2.1版本）中的“Limma”包分析数据集GSE4745中DHD差异表达基因，设置 $P < 0.05$ 、 $|\log_2FC| \geq 0.5$ ；应用“ggplot2”包对DHD差异表达基因进行可视化；应用“WGCNA”包构建加权基因共表达网络，并通过pick Soft Threshold函数计算相关系数和软阈值，然后根据软阈值计算拓扑重叠矩阵（topological overlap matrix, TOM）以筛选DHD相关基因模块。

**1.2.2 DHD相关铁死亡基因筛选** 应用R软件（4.2.1版本）中的“Venn”包分析铁死亡基因数据集与DHD差异表达基因、DHD相关基因模块的交集，即DHD相关铁死亡基因，然后将其导入String数据库（[\[stringdb.org/\]\(https://stringdb.org/\)）进行蛋白-蛋白相互作用（protein-protein interaction, PPI）分析，其相互作用系数设置为0.15。](https://</a></p></div>
<div data-bbox=)

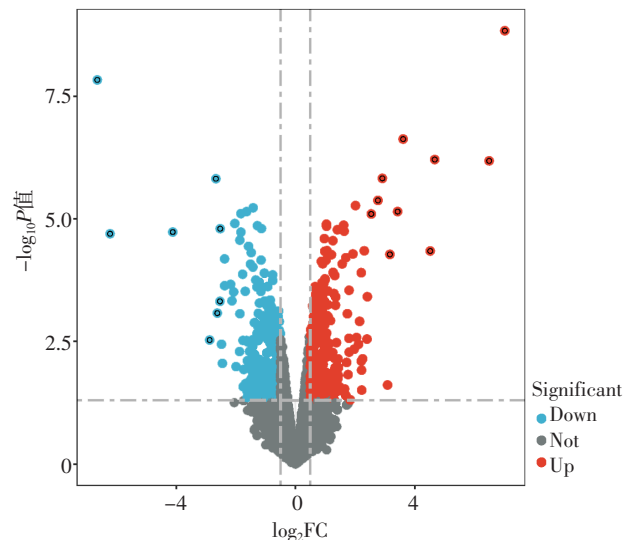
**1.2.3 铁死亡核心基因筛选** 针对DHD相关铁死亡基因进行LASSO回归分析<sup>[6]</sup>以筛选DHD相关铁死亡核心基因。

**1.2.4 铁死亡核心基因验证** 应用R软件（4.2.1版本）中的“ggplot2”包分别对数据集GSE4745和数据集GSE26887中的DHD相关铁死亡核心基因进行可视化分析。此外，检索既往研究，并分析DHD相关铁死亡核心基因表达趋势是否与既往研究结果一致。

**1.2.5 铁死亡核心基因蛋白质组学验证** 在PubMed下载XI等<sup>[7]</sup>研究并整理文中蛋白质组学原始数据，根据该原始数据对DHD相关铁死亡核心基因进行验证。

## 2 结果

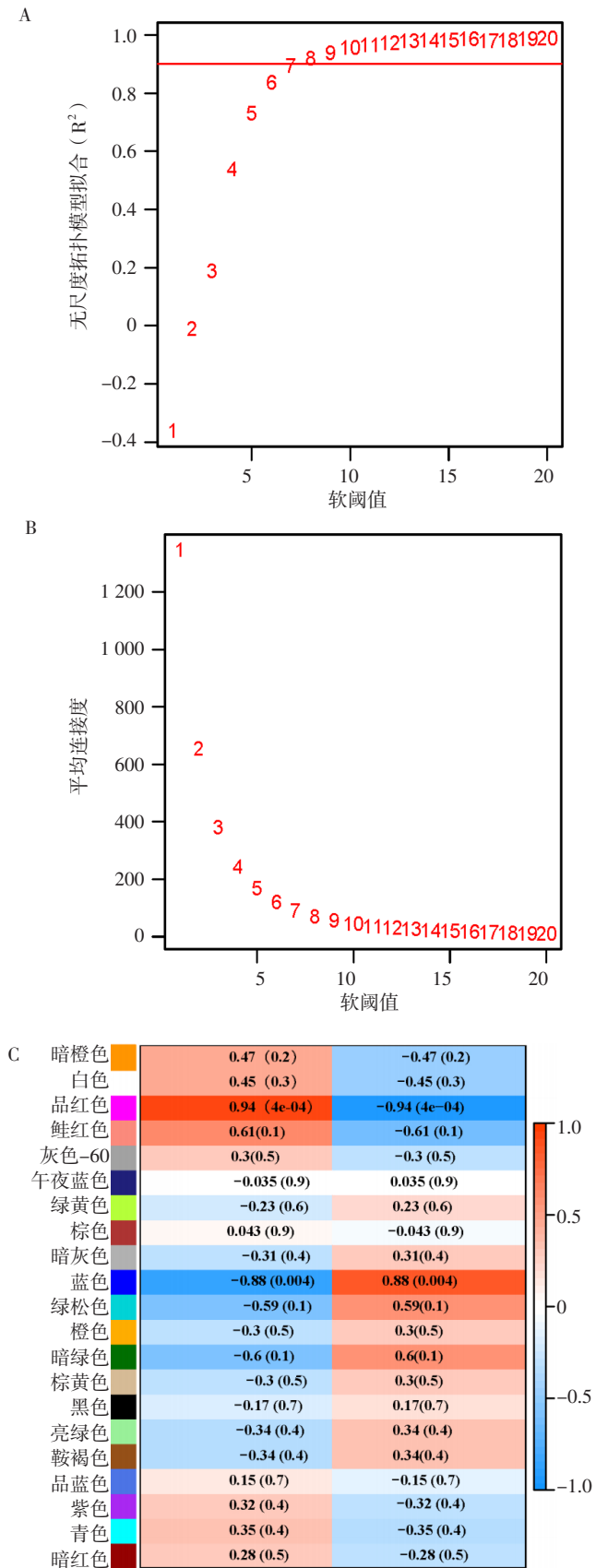
**2.1 DHD差异表达基因及相关基因模块** 将数据集GSE4745进行矫正、标准化处理、缺失值补充后获得4 000个DHD表达基因，进一步分析获得491个DHD差异表达基因，其中上调基因252个、下调基因239个，见图1。通过pick Soft Threshold函数计算相关系数为0.90时对应的软阈值为7，见图2A、2B。以软阈值为7将DHD差异表达基因构建邻接矩阵，基因模块的最小基因数为30，最终选择品红色和蓝色模块为DHD相关基因模块，共1 092个基因，见图2C。



注：红色点为上调基因，蓝色点为下调基因

图1 DHD差异表达基因火山图

Figure 1 Volcano map of DHD differentially expressed gene



注: A为不同软阈值对应的无尺度拓扑模型拟合 ( $R^2$ ), B为不同软阈值对应的平均连接度, C为基因模块

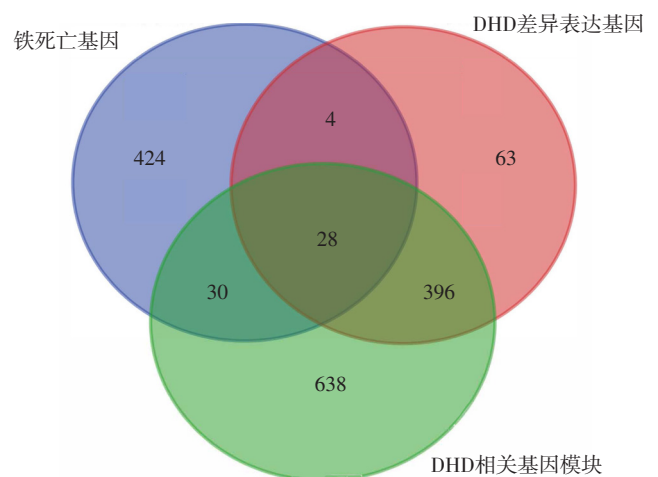
图2 加权基因共表达网络分析结果

Figure 2 Weighted gene co-expression network analysis results

2.2 DHD相关铁死亡基因 应用“Venn”包共获得28个DHD相关铁死亡基因,分别为H19、Nr4a1、Hspb1、Decr1、Acot1、Ptpn18、Gja1、Ech1、Pparg、Alox15、Mpc1、Smad7、Nr1d2、Atg13、Por、Enpp2、Arntl、Pdk4、Ppara、Pgrmc1、Acadsb、Gstm1、Bcat2、Slc3a2、Lamp2、Fabp4、Srebfl、Nr1d1,见图3。将28个DHD相关铁死亡基因导入String数据库进行PPI分析,结果显示,PPI网络共28个节点、118条边。

2.3 DHD相关铁死亡核心基因 LASSO回归分析结果显示,从28个DHD相关铁死亡基因中共获得6个DHD相关铁死亡核心基因,分别为H19、Nr4a1、Decr1、Gstm1、Por、Slc3a2,见图4~5。

2.4 DHD相关铁死亡核心基因的验证结果 在数据集GSE26887中,H19、Nr4a1、Decr1、Gstm1的表达趋势与数据集GSE4745一致,而Por和Slc3a2的表达趋势与数据集GSE4745不一致。在蛋白质组学原始数据中,可以提取到Decr1、Gstm1、Por、Slc3a2的蛋白质数据,其中Decr1、Por、Slc3a2的蛋白质表达趋势与数据集



注: DHD=糖尿病心脏病

图3 DHD相关铁死亡基因韦恩图

Figure 3 Venn diagram of DHD-related ferroptosis gene

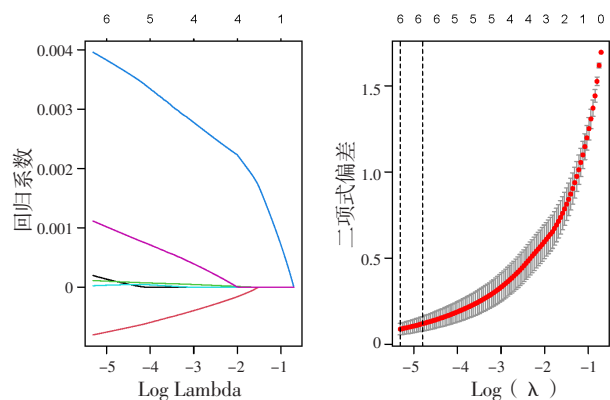
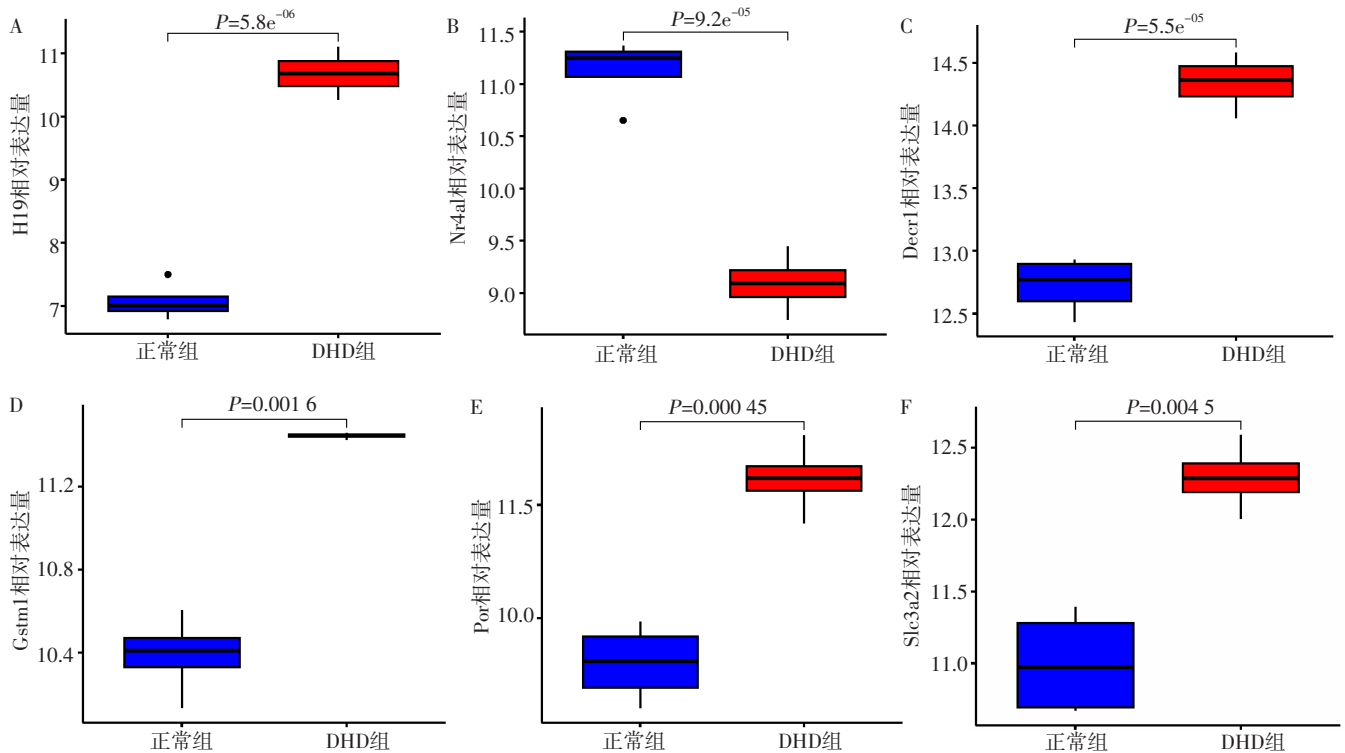


图4 LASSO回归分析结果

Figure 4 LASSO regression analysis results



注：A为H19，B为Nr4a1，C为Decr1，D为Gstm1，E为Por，F为Slc3a2

图5 DHD相关铁死亡核心基因表达箱式图

Figure 5 Box plot of the expression of DHD-related ferroptosis core gene

GSE4745中相关基因表达趋势一致，而Gstm1的蛋白质表达趋势与数据集GSE4745中相关基因表达趋势不一致；未提取到H19、Nr4a1的蛋白质数据，其中H19为非编码RNA，见表1。鉴于测序数据和蛋白质组学数据中样本量有限，将数据集GSE26887、蛋白质组学原始数据及既往研究结果<sup>[8-13]</sup>中至少符合2项者认为验证成功，故最终确定H19、Nr4a1、Decr1、Por为DHD相关铁死亡核心基因。

### 3 讨论

DHD是一种致死率较高的疾病，属于糖尿病患者慢性且严重的并发症。铁死亡是细胞依赖铁与氧化应激相关的死亡方式，与经典细胞死亡方式不同，其具有独特的形态学特征，主要包括线粒体嵴溶解和线粒体轮廓变小等<sup>[14]</sup>。此外，脂质过氧化物在细胞内积聚是铁死亡

发生的直接原因，而细胞内铁离子和活性氧介导的芬顿反应可进一步促进脂质过氧化物在细胞内积聚。研究发现，调控铁死亡可以延缓DHD的发生<sup>[4]</sup>。本研究基于生物信息学分析方法和机器学习算法探究了铁死亡参与DHD的可能机制。

本研究结果显示，H19、Nr4a1、Decr1、Gstm1、Por、Slc3a2是DHD相关铁死亡的核心基因，根据数据集GSE26887、蛋白质组学原始数据及既往研究最终确定H19、Nr4a1、Decr1、Por为DHD相关铁死亡的核心基因。H19是一个在DHD发展过程中发挥重要调控作用的长链非编码RNA，其与心脏肥厚、心肌纤维化的发生密切相关<sup>[15]</sup>。ZHUANG等<sup>[8]</sup>研究证实，H19在衰老型心肌细胞和衰老小鼠心脏组织中呈高表达，其可以通过miR-19a/SOCS1/p53/p21途径促进心肌细胞衰老。此外，在扩张型心肌病大鼠中抑制H19可减少心肌细胞凋亡并改善大鼠左心室结构和功能<sup>[16]</sup>。另有研究表明，H19表达下调后可通过调控miR-106b-5p/ACSL4轴而发挥促进细胞增殖、提高细胞活力、抑制脑微血管内皮细胞铁死亡的作用，进而防止脑出血事件的发生<sup>[17]</sup>。故推测下调H19可以抑制铁死亡，进而保护心肌细胞。Nr4a1又称为Nur77，是一种核激素受体，属于Nr4a亚家族成员<sup>[18]</sup>，其被证实可以通过多种机制抑制巨噬细胞的炎症反应，如抑制NF-κB信号通路或线粒体代谢重编程<sup>[19]</sup>，因而上调Nr4a1表达可减轻巨噬细胞的炎症

表1 DHD相关铁死亡核心基因筛选结果

Table 1 Screening results of DHD-related ferroptosis core genes

基因	数据集 GSE4745	数据集 GSE26887	蛋白质组学 原始数据	是否与既往研究 结果一致
H19	↑	↑	-	是 <sup>[8]</sup>
Nr4a1	↓	↓	-	是 <sup>[9]</sup>
Decr1	↑	↑	↑	是 <sup>[10]</sup>
Gstm1	↑	↑	↓	否 <sup>[11]</sup>
Por	↑	↓	↑	是 <sup>[12]</sup>
Slc3a2	↑	↓	↑	否 <sup>[13]</sup>

注：↑表示上调，↓表示下调，-表示无相关数据

反应并改善心功能<sup>[9]</sup>。由此推测,上调Nr4a1表达可抑制铁死亡,减轻炎症反应,进而改善心功能。NASSAR等<sup>[10]</sup>研究证实,Decr1可导致多不饱和脂肪酸在细胞内积累,进而增强线粒体氧化应激和脂质过氧化,并诱导铁死亡,推测Decr1可能是DHD的潜在治疗靶点。Por是一种位于内质网的氧化还原酶,其可以参与能量调节、调控糖异生和氧化还原稳态<sup>[20]</sup>。研究发现,Por缺乏可以抑制铁死亡诱导剂引起的过氧化磷脂水平升高,而Por升高可以促进细胞色素P450血红素成分发生芬顿反应,从而促进脂质过氧化<sup>[12]</sup>,推测Por可能通过促进脂质过氧化而促进铁死亡,进而调控DHD的发展。

综上所述,铁死亡参与DHD的机制可能与H19、Decr1、Por表达上调及Nr4a1表达下调有关,这为铁死亡参与DHD提供了新的证据,也为DHD的防治提供了新策略,但该结论仍需进一步研究证实。

作者贡献:徐守竹、赵晶进行文章的构思与设计,负责文章的质量控制及审校,并对文章整体负责、监督管理;杨启帆、方柔柔进行研究的实施与可行性分析;杨启帆、邬东东进行数据收集、整理、分析;杨启帆、骆家引进行结果分析与解释;杨启帆负责撰写、修订论文。

本文无利益冲突。

## 参考文献

- [1] 邵韦巧.机器学习分类算法对糖尿病数据应用研究[D].兰州:兰州大学,2021.
- [2] 郑尔昌,邹金串,薛成斌,等.糖尿病联合并发症发病风险计算与预测[J].华侨大学学报(自然科学版),2022,43(4):498-510.DOI:10.11830/ISSN.1000-5013.202110012.
- [3] RAWAL S, MANNING P, KATARE R. Cardiovascular microRNAs: as modulators and diagnostic biomarkers of diabetic heart disease [J]. *Cardiovasc Diabetol*, 2014, 13: 44. DOI: 10.1186/1475-2840-13-44.
- [4] WANG X, CHEN X X, ZHOU W Q, et al. Ferroptosis is essential for diabetic cardiomyopathy and is prevented by sulforaphane via AMPK/NRF2 pathways [J]. *Acta Pharm Sin B*, 2022, 12(2): 708-722. DOI: 10.1016/j.apsb.2021.10.005.
- [5] MAINALI S, DARSIE M E, SMETANA K S. Machine learning in action: stroke diagnosis and outcome prediction [J]. *Front Neurol*, 2021, 12: 734345. DOI: 10.3389/fneur.2021.734345.
- [6] LIU L, WANG C H, YIN L H, et al. Study on influence factors of pressure injury risk in the elderly inpatients with kidney disease based on LASSO regression [J]. *Open J Prev Med*, 2020, 10(6): 95-107. DOI: 10.4236/ojpm.2020.106007.
- [7] XI Y B, CHEN D P, DONG Z H, et al. Multi-omics insights into potential mechanism of SGLT2 inhibitors cardiovascular benefit in diabetic cardiomyopathy [J]. *Front Cardiovasc Med*, 2022, 9: 999254. DOI: 10.3389/fcvm.2022.999254.
- [8] ZHUANG Y T, LI T T, XIAO H W, et al. LncRNA-H19 drives cardiomyocyte senescence by targeting miR-19a/SOCS1/p53 axis [J]. *Front Pharmacol*, 2021, 12: 631835. DOI: 10.3389/fphar.2021.631835.
- [9] LI Q Q, LI Y T, HUANG W, et al. Loss of lipocalin 10 exacerbates diabetes-induced cardiomyopathy via disruption of Nr4a1-mediated anti-inflammatory response in macrophages [J]. *Front Immunol*, 2022, 13: 930397. DOI: 10.3389/fimmu.2022.930397.
- [10] NASSAR Z D, MAH C Y, DEHAIRS J, et al. Human DECR1 is an androgen-repressed survival factor that regulates PUFA oxidation to protect prostate tumor cells from ferroptosis [J]. *eLife*, 2020, 9: e54166. DOI: 10.7554/eLife.54166.
- [11] TIN A, SCHARPF R, ESTRELLA M M, et al. The loss of GSTM1 associates with kidney failure and heart failure [J]. *J Am Soc Nephrol*, 2017, 28(11): 3345-3352. DOI: 10.1681/ASN.2017030228.
- [12] ZOU Y L, LI H X, GRAHAM E T, et al. Cytochrome P450 oxidoreductase contributes to phospholipid peroxidation in ferroptosis [J]. *Nat Chem Biol*, 2020, 16(3): 302-309. DOI: 10.1038/s41589-020-0472-6.
- [13] WU F X, XIONG G Y, CHEN Z J, et al. SLC3A2 inhibits ferroptosis in laryngeal carcinoma via mTOR pathway [J]. *Hereditas*, 2022, 159(1): 6. DOI: 10.1186/s41065-022-00225-0.
- [14] STOCKWELL B R, FRIEDMANN ANGELI J P, BAYIR H, et al. Ferroptosis: a regulated cell death nexus linking metabolism, redox biology, and disease [J]. *Cell*, 2017, 171(2): 273-285. DOI: 10.1016/j.cell.2017.09.021.
- [15] ZHUO C J, JIANG R H, LIN X D, et al. LncRNA H19 inhibits autophagy by epigenetically silencing of DIRAS3 in diabetic cardiomyopathy [J]. *Oncotarget*, 2017, 8(1): 1429-1437. DOI: 10.18632/oncotarget.13637.
- [16] ZHANG Z, GAO W, LONG Q Q, et al. Increased plasma levels of lncRNA H19 and LIPCAR are associated with increased risk of coronary artery disease in a Chinese population [J]. *Sci Rep*, 2017, 7(1): 7491. DOI: 10.1038/s41598-017-07611-z.
- [17] CHEN B, WANG H R, LV C L, et al. Long non-coding RNA H19 protects against intracerebral hemorrhage injuries via regulating microRNA-106b-5p/acyl-CoA synthetase long chain family member 4 axis [J]. *Bioengineered*, 2021, 12(1): 4004-4015. DOI: 10.1080/21655979.2021.1951070.
- [18] ZHAO Y, BRUEMMER D. NR4A orphan nuclear receptors: transcriptional regulators of gene expression in metabolism and vascular biology [J]. *Arterioscler Thromb Vasc Biol*, 2010, 30(8): 1535-1541. DOI: 10.1161/ATVBAHA.109.191163.
- [19] LI L, LIU Y, CHEN H Z, et al. Impeding the interaction between Nur77 and p38 reduces LPS-induced inflammation [J]. *Nat Chem Biol*, 2015, 11(5): 339-346. DOI: 10.1038/nchembio.1788.
- [20] PANDEY A V, FLÜCK C E. NADPH P450 oxidoreductase: structure, function, and pathology of diseases [J]. *Pharmacol Ther*, 2013, 138(2): 229-254. DOI: 10.1016/j.pharmthera.2013.01.010.

(收稿日期: 2023-04-13; 修回日期: 2023-07-24)

(本文编辑: 谢武英)