

• 论 著 •

基于 GEO 数据库筛选分析肌肉萎缩的相关靶点及通路*

薛 艳¹, 邱凤喜¹, 郑晓栋², 陈 斌¹, 王 楠¹, 王 雅¹, 张 楠¹, 仲荣洲¹, 褚立希^{1△}

1. 上海市养志康复医院/上海市阳光康复中心临床研究中心, 上海 201600;

2. 上海市普陀区真如镇街道社区卫生服务中心, 上海 200333

摘要:目的 通过生物信息学方法分析肌肉萎缩前后的差异基因和相关信号通路, 探讨肌肉萎缩的发病机制及潜在治疗靶点。方法 从公共基因表达数据库(GEO)下载芯片数据集 GSE148152 和其注释平台 GPL17586。先运用 R 软件中的 limma 包筛选差异基因, 使用 ggplot2 和 pheatmap 程序包对差异表达基因(DEGs)进行可视化分析。然后通过 STRING 数据库进一步筛选核心靶点, 最后采用 clusterProfiler GO 和 clusterProfiler KEGG 软件包对核心靶点进行 GO 和 KEGG 富集分析。结果 通过对芯片数据集 GSE148152 中肌肉萎缩前后股外侧肌的肌肉样本进行分析性, 结果共获得 100 个 DEGs, 其中上调基因为 51 个, 下调基因为 49 个。在 String 数据库中, 运用网络拓扑特征分析, 再筛选出包括 TNNT1、FOXO3、MSTN 等在内的 31 个 DEGs 作为关键靶点。GO 富集分析表明这些 DEGs 主要涉及细胞代谢过程; KEGG 信号通路富集分析这些基因与代谢途径、过氧化物酶体增殖物激活受体、腺苷酸活化蛋白激酶等信号通路有关。结论 该次研究发现了肌肉萎缩发展中的潜在关键基因和通路, 为今后的治疗方向提供了理论依据。

关键词:肌肉萎缩; 生物信息学; 差异基因; GEO 数据库
DOI:10.3969/j.issn.1673-4130.2022.14.013 **中图法分类号:**R592
文章编号:1673-4130(2022)14-1726-04 **文献标志码:**A

Screening and analysis of related targets and pathways of muscle atrophy based on GEO database*

XUE Yan¹, QIU Fengxi¹, ZHENG Xiaodong², CHEN Bin¹, WANG Nan¹,
WANG Ya¹, ZHANG Nan¹, ZHONG Rongzhou¹, CHU Lixi^{1△}

1. Clinical Research Center, Shanghai YangZhi Rehabilitation Hospital/Shanghai Sunshine Rehabilitation Center, Shanghai 201600, China; 2. Shanghai Putuo District Zhenru Town Street Community Health Service Center, Shanghai 200333, China

Abstract: **Objective** To analyze the differential genes and related signal pathways before and after muscle atrophy by bioinformatics, and to explore the pathogenesis and potential therapeutic targets of muscle atrophy. **Methods** The microarray dataset GSE148152 and its annotation platform GPL17586 were downloaded from the public gene expression database(GEO). Firstly, limma package in R software was used to screen differential genes, and ggplot2 and pheatmap packages were used to visually analyze DEGs. Then the core targets were further screened by STRING database. Finally, GO and KEGG enrichment analysis of core targets was carried out by using clusterProfiler GO and clusterProfiler KEGG software packages. **Results** Through the analysis of the muscle samples of the lateral vastus lateralis muscle before and after muscle atrophy in the chip data set GSE148152, a total of 100 DEGs, were obtained, including 51 up-regulated genes and 49 down-regulated genes. In the String database, using the analysis of network topology characteristics, 31 DEGs including TNNT1, FOXO3, MSTN and so on are selected as core targets. GO enrichment analysis showed that these DEGs were mainly involved in cell metabolism, and KEGG signal pathway enrichment analysis of these genes were related to metabolic pathway, PPAR, AMPK and other signal pathways. **Conclusion** This study found the potential key genes and pathways in the development of muscle atrophy, which provides a theoretical basis for the direction of treatment in the future.

Key words: muscular dystrophy; bioinformatics; differential genes; GEO database

* 基金项目: 上海市青年科技英才扬帆计划(21YF1443400); 上海市进一步加快中医药事业发展三年行动计划(ZY[2018-2020]-FWTX-8016); 上海市养志康复医院/上海市阳光康复中心科研培育项目(KYPY202002)。

作者简介: 薛艳, 女, 助理研究员, 主要从事中西医结合康复治疗研究。 △ 通信作者, E-mail: chulixi2@126.com。

骨骼肌是人体功能活动的基础,能够进行运动和呼吸^[1]。随着年龄的增长,血液供应减少,或肌肉固定,长时间失重,营养不良等会导致横纹肌营养障碍,肌肉纤维变细甚至消失等,使得肌肉体积缩小并且力量减退,最终引发肌肉萎缩^[2-3]。肌肉萎缩如果未进行及时治疗,很可能引发其他并发症,因此肌肉萎缩不仅危害健康,还会降低生活质量^[4]。但目前尚未有统一的治疗方案,很多药物的疗效也不太理想,因此研究者应当要重视其预防和治疗。在本研究中,通过对比 GEO 数据库中卧床前和卧床后受试者肌肉组织中的差异表达基因(DEGs),可以进一步明确肌肉萎缩的相关靶点及信号通路,由此可以为治疗肌肉萎缩提供新的思路 and 依据。

1 资料与方法

1.1 一般资料 美国国立生物技术信息中心(NCBI)的 GEO 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)是一个储存高通量测序数据的数据库。在本次研究中,以("Muscular Atrophy"[Mesh])作为关键词进行芯片数据检索,芯片类型为"Expression profiling by array",种属为"Homo sapiens",检索时间截止于 2020 年 9 月。检索获得 GSE148152 数据集,其注释平台为 GPL17586。

1.2 数据分析方法 下载基因芯片数据 GSE148152 和 GPL 17586-45144 的芯片基因注释文件,运用 KNN 算法补全缺失值,构建基因表达矩阵,对数据进行归一化处理后,运用 R 软件中的 limma 包筛选符合 $|\log\text{FoldChange}| > 1$ 且 $P < 0.05$ 的差异基因。筛选出差异基因后使用 R 软件中的 ggplot2 和 pheatmap 程序包对 DEGs 进行可视化分析,绘制火山图和热图。

1.3 蛋白质-蛋白质相互作用(PPI)网络构建 通过 STRING(<https://string.db.org/>),一个已知的 PPI 网络数据库来进一步筛选核心靶点,并且绘制 PPI 的网络图。

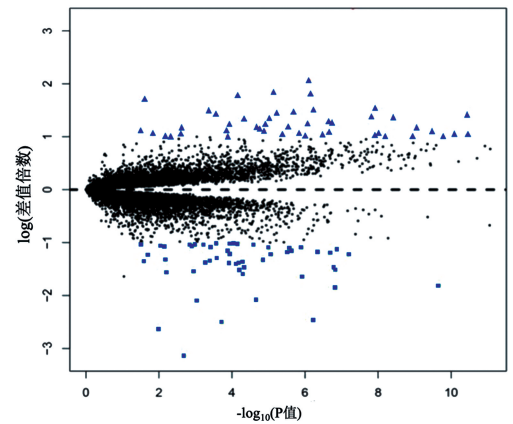
1.4 基因本体论(GO)富集分析和京都基因和基因组百科全书(KEGG)富集分析 使用 R 软件中的 clusterProfiler GO 和 clusterProfiler KEGG 软件包对 DEGs 进行 GO 和 KEGG 富集分析。GO 分析主要分析差异基因的功能。KEGG 富集可以获得差异基因的潜在生物学功能和所涉及的生物途径。

2 结果

2.1 原始数据的基本信息 选取 GSE148152 数据集中 11 名健康男性,进行 84 d 头朝下倾斜卧床休息(BR),即受试者始终保持 6°头低位(即休息、淋浴、运输、运动训练、如厕等),但在用餐时可以用肘部休息,这样可以造成肌肉萎缩。所有受试者在卧床前(PRE)和卧床后(POST),即发生肌肉萎缩的前后均取股外侧肌的肌肉进行活检,其中 PRE 组的样本编号分别为 GSM 4455104、GSM 4455106、GSM 4455108、GSM

4455110、GSM 4455112、GSM 4455114、GSM 4455116、GSM 4455117、GSM 4455119、GSM 4455120 和 GSM 4455121;POST 组的样本编号分别为 GSM 4455105、GSM 4455107、GSM 4455109、GSM 4455111、GSM 4455113、GSM 4455115、GSM 4455127、GSM 4455128、GSM 4455129、GSM 4455130 和 GSM 4455131。

2.2 DEGs 的筛选结果 运用 R 语言分析受试者卧床前后股外侧肌的肌肉组织中的基因表达,结果共获得 100 个 DEGs,其中卧床后上调基因为 51 个,同时下调基因为 49 个。见图 1。



注:图中三角形表示上调基因,方形表示下调基因,圆形表示未有明显改变的基因。

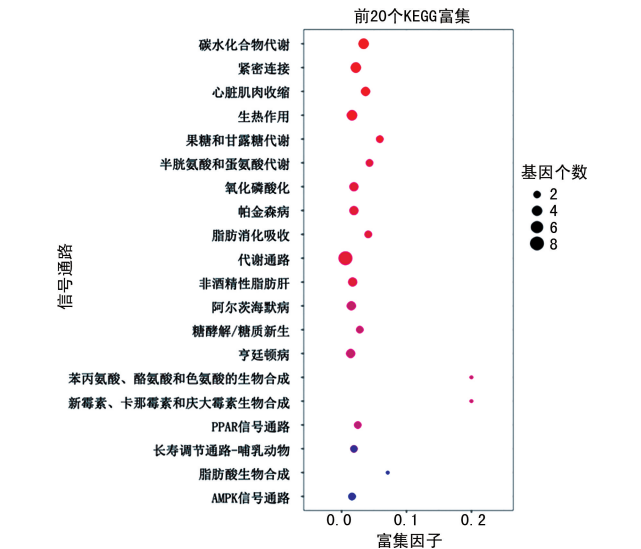
图 1 DEGs 的火山图

2.3 DEGs 的 PPI 网络构建 在 String 数据库中,通过构建 100 个 DEGs 的 PPI 网络,然后运用网络拓扑特征分析,以“自由度”、“接近中心性”和“介数”3 个主要参数为基础,通过筛选各项参数的数值均 > 0 的 31 个 DEGs 作为关键靶点,其中发生肌肉萎缩后肌肉组织中表达显著上调的基因为 MYOZ2、MYH7B、TECRL、CASQ2、CKMT2、TNNT1、LDHB、MYL6B、GYS1、MYL3、SMPX、FABP3、MDH1、TNNI1、UQCRI0、NDUFAB1、GOT2、ACSL1、FBP2、ATP5G1、HK2 和 COX5A,同时表达显著下调的基因为 SESN1、ABCA1、FOXO3、MSTN、ADH1B、ACTN3、MYH1、PDK4 和 MYH8。

2.4 DEGs 的 GO 功能富集分析 GO 数据库总共有三大类,分别是生物学过程(BP)、细胞定位(CC)和分子功能(MF),各自描述了基因产物可能行使的分子功能,所处的细胞环境,以及参与的生物学过程。通过 R 分析,将两组中的 DEGs 进行 GO 功能富集分析,总共获得 43 个具有显著意义的 GO 条目:在 BP 中,有 21 条显著富集条目(48.8%),主要涉及细胞过程、代谢过程和多细胞生物的过程等;在 CC 中,有 12 条显著富集条目(27.9%),主要涉及细胞、细胞部分和细胞器等;在 MF 中,有 10 条显著富集条目(23.3%),主要涉及结合、催化活性和结构分子活动等。

2.5 DEGs 的 KEGG 通路富集分析 首先通过对 31

个核心 DEGs 进行 KEGG 通路的聚类分析,发现肌肉萎缩前后的核心 DEGs 所属的信号通路主要集中于以下三大类:代谢、生物体系统和人类疾病,并且参与调控了能量代谢(包含 3 个基因)、内分泌系统(包含 6 个基因)、循环系统(包含 4 个基因)、环境适应(包含 4 个基因)、代谢疾病(4 个基因)等相关生理病理机制,进一步通过富集分析发现 DEGs 具体涉及的通路主要是碳代谢、果糖和甘露糖代谢等代谢途径及过氧化物酶体增殖物激活受体(PPAR)、腺苷酸活化蛋白激酶(AMPK)等信号通路(图 2)。



注:圆点的大小表示该条通路下 DEGs 的个数,圆点越大表示基因数越多,位置越靠上表示富集越明显。

图 2 DEGs 的 KEGG 通路富集分析

3 讨 论

肌肉萎缩主要与年龄增长、运动损伤、疾病因素、营养不良及神经源性等原因相关^[5-7]。其中老年性肌肉萎缩不容忽视,根据一项美国的横断面调查显示,60 岁以上老年人肌肉萎缩的男女总发病率分别为 75.5% 和 35.4%^[8]。而随着我国人口老龄化日趋加重,肌肉萎缩的发病率也逐年攀高^[9]。肌肉萎缩给家庭和个人带来的不良影响是巨大的,所以随着社会对健康要求的不断提高,已然是一个备受关注的问题^[10]。因此为了进一步明确肌肉萎缩的病理机制,研究者借助 GEO 数据库中的基因芯片数据进行深入挖掘分析,共获得了 31 个核心 DEGs,其中共同表达下调的 DEGs 有 22 个,共同表达上调的 DEGs 有 9 个,包括慢速骨骼肌型肌钙蛋白(TNNI1)、叉头蛋白转录因子 3(FOXO3)、肌肉生长抑制素(MSTN)等关键基因。

TNN1 是肌钙蛋白 I(TnI)的同源基因编码的亚型之一,TnI 是横纹肌细丝中的肌钙蛋白复合物的抑制亚单位,在钙调节收缩和舒张中发挥核心作用^[11-12]。其中 TNNI1 可以通过抑制钙介导的肌动蛋白和肌球蛋白复合物的构象变化来抑制肌肉收缩,其表达的高低能反映骨骼肌内慢肌纤维的水平^[13]。

FOXO3 属于叉头基因转录因子家族(FOXO)的亚型,其与细胞凋亡和肌肉再生有关^[14-15]。有研究报道,FOXO3 可以通过同时调控泛素-蛋白酶体和自噬-溶酶体系统的蛋白质(萎缩基因)来触发骨骼肌肉萎缩^[16]。AMPK 是调节骨骼肌线粒体功能和氧化应激的关键因子,可通过影响骨骼肌的线粒体质量控制调节线粒体功能^[17]。有研究发现,AMPK/FOXO3 信号通路在骨骼肌肉萎缩中会被激活,而在阻断 AMPK/FOXO3 信号通路后,则可以有效逆转小鼠的肌肉萎缩^[18],说明 FOXO3 和 AMPK 二者在肌肉萎缩中起着重要的调节作用。

MSTN 是肌肉生长的负调控因子,可以抑制成肌细胞的增殖和分化^[19]。通过抑制 MSTN 基因的表达,能够抑制肌肉萎缩,增强肌力^[20]。MSTN 进入细胞核的主要途径是 Smad 激活途径和 Akt 抑制途径,有研究显示 MSTN 通过介导 Akt/mTOR 信号通路可引发肌肉萎缩^[21-22]。另外还有一些研究发现 PPAR 参与全身炎症反应并且调节骨骼肌氧化代谢,而 MSTN 则可能通过调节 PPAR 来影响骨骼肌的脂肪生成^[17,23]。

综上所述,本次研究主要利用生物信息学技术分析,获得了与肌肉萎缩病理机制相关的一些重要靶基因及信号通路,其中包含 TNNI、FOXO3 和 MSTN 等基因,这些靶点又可能通过调控 AMPK、PPAR 等信号通路引发肌肉萎缩,因此今后可开展针对这些特定的靶点及信号通路的相关疗效及病理机制研究。

参考文献

- [1] 程宙明. 抗阻和有氧训练对大学生最大摄氧量的影响及其运动适应机制的研究[D]. 上海:华东师范大学,2017.
- [2] 王哲培. 被动牵伸运动结合电针对废用性肌肉萎缩小鼠骨骼肌的影响[D]. 北京:北京体育大学,2019.
- [3] DUTT V, GUPTA S, DABUR R, et al. Skeletal muscle atrophy: Potential therapeutic agents and their mechanisms of action[J]. Pharmacol Res, 2015, 99(1): 86-100.
- [4] 张瑾, 闫坚强, 史仍飞. 维生素 D 在骨骼肌功能中的作用[J]. 营养学报, 2021, 43(5): 514-519.
- [5] 张汇敏. 健身气功干预老年人衰老性肌肉萎缩功效研究[D]. 武汉:武汉体育学院,2016.
- [6] CAO R Y, LI J, DAI Q Y, et al. Muscle atrophy: present and future[J]. Adv Exp Med Biol, 2018, 1088(1): 605-624.
- [7] DING S G, DAI Q Y, HUANG H T, et al. An overview of muscle atrophy[J]. Adv Exp Med Biol, 2018, 1088(1): 3-19.
- [8] 金惠林. 老年人肌少症与不同运动的相关性指标分析[D]. 杭州:浙江大学,2018.
- [9] 郑艺. 运动干预支持在老年肌肉减少症病人中应用的研究进展[J]. 全科护理, 2020, 18(30): 4100-4103.
- [10] 余康. 爬梯运动联合低强度脉冲超声对去势大鼠股四头肌肉萎缩的改善效果及分子机制研究[D]. 西安:陕西师范大学,2019.

- [4] MOGHIMI M, KORDAFSHARI G, KENARI H M. Hashimoto's thyroiditis[J]. J Infect Dis, 2019, 23(6): 435-440.
- [5] 詹玲, 陈创, 孙圣荣. 免疫微环境在桥本甲状腺炎与甲状腺癌中作用研究进展[J]. 中华实用诊断与治疗杂志, 2020, 34(6): 640-642.
- [6] BHAT T, TELI S, RIJAL J, et al. Neutrophil to lymphocyte ratio and cardiovascular diseases: a review[J]. Expert Rev Cardiovasc Ther, 2013, 11(1): 55-59.
- [7] 王迪, 梁伟娟, 许雷鸣, 等. 桥本甲状腺炎西药治疗进展[J]. 医学综述, 2021, 27(22): 4501-4505.
- [8] DUNTAS L H. Environmental factors and autoimmune thyroiditis[J]. Nat Clin Pract Endocrinol Metab, 2008, 4(8): 454-460.
- [9] JAILLON S, GALDIERO M R, DEL PRETE D, et al. Neutrophils in innate and adaptive immunity[J]. Semin Immunopathol, 2013, 35(4): 377-394.
- [10] ITALIANO J E, MAIRUHU A T, FLAUMENHAFT R. Clinical relevance of microparticles from platelets and megakaryocytes[J]. Current Opinion Hematol, 2010, 17(6): 578-584.
- [11] BAYIR O, KARAGOZ T, OCAL B, et al. Predictive role of neutrophil-lymphocyte and platelet-lymphocyte ratios in thyroid nodules with cytological diagnosis of "undetermined significance" and "suspicious for malignancy"[J]. ENT Updates, 2017, 7(1): 28-32.
- [12] KIM S M, KIM E H, KIM B H, et al. Association of the preoperative neutrophil-to-lymphocyte count ratio and platelet-to-lymphocyte count ratio with clinicopathological characteristics in patients with papillary thyroid cancer[J]. Endocrinol Metab, 2015, 30(4): 494-501.
- [13] HAN S W, KANG S Y, KIM S K, et al. Clinical significance of blood neutrophil-to-lymphocyte ratio in patients with papillary thyroid carcinoma[J]. Endocr Surg, 2014, 14(1): 184-189.
- [14] USLU A U, KUCUK A, SAHIN A, et al. Two new inflammatory markers associated with disease activity score-28 in patients with rheumatoid arthritis: neutrophil-lymphocyte ratio and platelet-lymphocyte ratio[J]. Int J Rheum Dis, 2015, 18(1): 731-735.
- [15] KESKIN H, KAYA Y, CADIRCI K, et al. Elevated neutrophil-lymphocyte ratio in patients with euthyroid chronic autoimmune thyreoiditis[J]. Endocr Regul, 2016, 50(3): 148-153.
- [16] 陈玲敏, 蔡林辉, 郑雨薇, 等. NLR 与 PLR 对鉴别诊断自身免疫性甲状腺炎的临床意义[J]. 健康研究, 2019, 39(5): 566-569.
- [17] AJJAN R A, WEETMAN A P. The pathogenesis of hashimoto's thyroiditis: further developments in our understanding[J]. Horm Metab Res, 2015, 47(10): 702-710.
- [18] CHEN C R, HAMIDI S, BRALEY-MULLEN H, et al. Antibodies to thyroid peroxidase arise spontaneously with Age in NOD. H-2h4 mice and appear after thyroglobulin antibodies[J]. Endocrinology, 2010, 151(9): 4583-4593.
- [19] 黄园园, 刘亚楠, 于世鹏. CD4⁺ T 淋巴细胞相关细胞因子与桥本氏甲状腺炎发病机制的研究进展[J/CD]. 中华诊断学电子杂志, 2019, 7(3): 214-216.
- [20] 顾雪疆, 吴森敏, 徐静, 等. 中性粒细胞/淋巴细胞比值和血小板/淋巴细胞比值在甲状腺毒症患者鉴别诊断中的意义[J]. 中华内分泌代谢杂志, 2017, 33(6): 491-496.
- [21] 吕珏, 陈瑶, 曹茜茜, 等. 联合应用 NLR, PLR 及 RDW 对桥本甲状腺炎的诊断价值[J]. 慢性病杂志, 2020, 21(12): 61-63.

(收稿日期: 2021-06-20 修回日期: 2022-07-02)

(上接第 1728 页)

- [11] 路倩. 心肌细胞核内的肌钙蛋白 I 作为非经典转录因子调 ATP2a2/SERCA2a 表达的机制研究[D]. 重庆: 重庆医科大学, 2019.
- [12] MANAS-GARCÍA L, BARGALLÓ N, GEA J, et al. Muscle phenotype, proteolysis, and atrophy signaling during reloading in mice: effects of curcumin on the gastrocnemius[J]. Nutrients, 2020, 12(2): 388.
- [13] 李景龙. 去负荷肌肉萎缩小鼠运动耐力的变化及机制研究[D]. 北京: 北京体育大学, 2012.
- [14] 孟繁博. 运动联合枸杞多糖延缓衰老小鼠骨骼肌肉萎缩的机制研究[D]. 武汉: 武汉体育学院, 2020.
- [15] CHIARA N, LAURA S, DARLING D S, et al. Regulation of muscle atrophy-related genes by the opposing transcriptional activities of ZEB1/CtBP and FOXO3[J]. Nucleic Acids Res, 2018, 46(20): 10697-10708.
- [16] MAMMUCARI C, MILAN G, ROMANELLO V, et al. FoxO3 controls autophagy in skeletal muscle in vivo[J]. Cell Metab, 2007, 6(6): 458-471.
- [17] 钟卫权. 不同负荷游泳训练对去卵巢大鼠骨骼肌氧化应激和自噬的影响研究[D]. 上海: 上海体育学院, 2020.
- [18] LIU J, PENG Y, WANG X, et al. Mitochondrial dysfunction launches dexamethasone-induced skeletal muscle atrophy via AMPK/FOXO3 Signaling[J]. Mol Pharm, 2016, 13(1): 73-84.
- [19] 罗琳, 杨金鹏, 王松涛, 等. 肥胖性肌肉萎缩的分子机制研究进展[J]. 中国康复理论与实践, 2017, 23(5): 553-557.
- [20] 马淑梅, 陈瑶, 刘莉. 肌肉萎缩症治疗的研究进展[J]. 世界临床药物, 2017, 38(11): 790-792.
- [21] 伏智亮, 沈如玉, 王绮绮, 等. 肌肉生长抑制素调控肌肉和脂肪组织代谢的研究进展[J]. 畜牧与兽医, 2019, 51(7): 121-126.
- [22] 范蕊, 吴宗辉, 陈晓琳, 等. 电针对创伤性脊髓损伤大鼠下肢骨骼肌肉萎缩相关蛋白表达的影响[J]. 中国康复理论与实践, 2019, 25(10): 1133-1139.
- [23] 黄榆杰. 白藜芦醇对肥胖性肌肉衰减症的作用及机制研究[D]. 重庆: 中国人民解放军陆军军医大学, 2019.

(收稿日期: 2021-06-08 修回日期: 2022-07-02)