

基于 GEO 数据库筛选稳定性心绞痛患者外周血 关键差异基因及诊断模型构建

侯芳霞, 刘琳, 张维, 方凤, 齐婷, 马美娟, 刘富强, 唐治国

(陕西省人民医院心血管内一科, 西安 710068)

摘要:目的 通过生物信息学方法对稳定性心绞痛患者外周血基因表达谱芯片进行分析, 获取其外周血表达谱特征并筛选关键差异表达基因作为潜在的分子标记物并构建 Nomogram 诊断模型。方法 从 NCBI 中的基因表达综合 (Gene Expression Omnibus, GEO) 数据库中下载稳定性心绞痛患者和对照组的外周血基因表达谱芯片数据集 GSE98583, 使用 R 软件 limma 包筛选出具有显著意义的差异基因 (differential expression genes, DEGs); 利用 clusterProfiler 包进行基因本体 (gene ontology, GO) 与 KEGG (kyoto encyclopedia of genes and genomes) 通路富集分析; 使用 STRING 在线分析工具构建蛋白交互网络和 Cytoscape 软件 Cytohubba 和 Mcode 插件筛选出关键基因; 以关键基因为变量构建稳定性心绞痛 Nomogram 分子诊断预测模型。结果 通过比较稳定性心绞痛患者和正常受试者外周血基因表达谱, 共筛选出 303 个差异表达基因, 其中上调基因 160 条, 下调基因 43 条; GO 和 KEGG 分析表明, 这些差异表达基因主要参与神经递质配体受体相互作用、脂肪吸收消化、钙调节信号通路、PI3K-Akt 通路、NF-kappaB 通路及氧化磷酸化等有关, 使用 Cytohubba 进一步分析, 筛选出 10 个关键基因 BDNF, GFAP, SYN1, NES, PLG, HPGDS, KCNC1, APOA4, AMBP 和 TJP1, 并建立了 Nomogram 诊断模型。结论 使用生物信息学方法揭示稳定性心绞痛外周血差异基因潜在特征, 为稳定性心绞痛的早期诊断提供新的思路。

关键词: 稳定性心绞痛; 生物信息学; 基因表达综合数据库; 差异表达基因; 富集分析

中图分类号: R541.4; Q343.1 文献标识码: A 文章编号: 1671-7414 (2022) 06-019-06

doi:10.3969/j.issn.1671-7414.2022.06.004

Identification of Hub Genes and Differential Expression Genes for Peripheral Blood Samples of Stable Angina Pectoris Based on GEO Databases

HOU Fang-xia, LIU Lin, ZHANG Wei, FANG Feng, QI Ting, MA Mei-juan, LIU Fu-qiang, TANG Zhi-guo

(Department of Cardiology, Shaanxi Provincial People's Hospital, Xi'an 710068, China)

Abstract: Objective To analyze the peripheral blood gene expression profile of patients with stable angina pectoris using bioinformatics methods, obtain the characteristics of the peripheral blood gene expression profile, and screen hub genes as potential molecular markers and construct nomogram model. **Methods** The GSE98583 dataset was downloaded from the GEO database in NCBI, which contain the peripheral blood gene expression profile of stable angina patients and the control group. Differential expression genes (DEGs) were screen using the R software limma package. Further Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the clusterProfiler package. The STRING database and Cytohubba plug-in of Cytoscape software were used to establish and visualize the protein-protein interaction (PPI) network and identify the hub genes. **Results** A total of 303 differentially expressed genes were screened between stable angina pectoris and normal subjects, including 160 up-regulated genes and 43 down-regulated genes. GO and KEGG pathway enrichment analyses found that these genes mainly take part in the process of neuroactive ligand-receptor interaction, fat digestion and absorption, calcium signaling pathway, PI3K-Akt signaling pathway, NF-kappa B signaling pathway, oxidative phosphorylation and so on. Ten key genes were identified by Cytohubba plug-in, including BDNF, GFAP, SYN1, NES, PLG, HPGDS, KCNC1, APOA4, AMBP, TJP1 and nomogram model was constructed based on these hub genes. **Conclusion** The bioinformatics method was used to reveal the potential characteristics of differential genes in peripheral blood of stable angina pectoris, providing a new idea for the early dragnosis of stable angina pectoris.

Keywords: stable angina pectoris; bioinformatic analysis; GEO databases; differential expression genes; enrichment analysis

基金项目: 陕西省重点研发项目 (2021SF-329, 2022SF-476); 西安市科技计划项目 (21YXYJ0095)。

作者简介: 侯芳霞 (1978-), 女, 本科, 主管护师, 主要从事心血管疾病预防与早期识别, E-mail: zhuzhu19821112@yeah.net。

通讯作者: 唐治国 (1978-), 男, 本科, 主治医师, 主要从事冠心病介入治疗及早期防治, E-mail: tang13484600363@163.com。

冠心病是严重危害人类健康的疾病,据《中国心血管健康与疾病报告2019》报道^[1],我国估计现有冠心病患者1139万,给社会和家庭带来了沉重的负担。慢性稳定性心绞痛是冠状动脉固定性严重狭窄基础上,由于心肌负荷的增加引起的心肌急剧、短暂的缺血缺氧临床综合征,即慢性冠脉综合征,是冠心病最常见的类型^[2]。然而由于慢性心绞痛与急性冠脉综合征相比,症状较轻、心肌酶等心肌特异性损伤标记生化指标不高,很容易被患者和医生忽视,失去早期治疗的机会,进而导致病情恶化形成急性心肌梗死。因此早期识别稳定性心绞痛对冠心病的预后有着重要的意义。经典诊断稳定性心绞痛主要依赖于有关冠状动脉造影术及冠状动脉CT,这些检查因有创、费用高等不易被患者接收,基于外周血细胞能对稳定性冠心病进行早期诊断有着广阔的前景。本研究拟基于基因表达综合(Gene Expression Omnibus, GEO)数据库采用生物信息学方法对稳定性心绞痛患者外周血基因表达谱芯片进行分析,获取稳定性心绞痛外周血表达谱特征,筛选关键差异表达基因作为潜在的分子标记物,并以关键基因为变量构建Nomogram诊断预测模型,为稳定性心绞痛的早期诊断提供新思路。

1 材料与方法

1.1 资料来源 从美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)的GEO数据库中(<https://www.ncbi.nlm.nih.gov/geo/>)下载GSE98583数据集。该数据集包括12例稳定性心绞痛患者(排除糖尿病)和6例正常受试者外周全血表达谱数据,所有患者均行冠脉造影证实冠脉情况。所采用表达谱芯片为Affymetrix Human Genome U133A 2.0 Array,平台为GPL571。所有数据均可在网上免费获得。本项目是对在线数据库再分析,不涉及伦理问题。

1.2 方法

1.2.1 分析差异表达基因:采用R studio4.1.0软件对下载的基因表达谱数据进行标准化,再采用limma包分析差异基因(differential expression genes, DEGs)进行筛选^[3],筛选标准为同时满足:① $P < 0.05$;②组间差异表达倍数(fold change) ≥ 2.0 。使用ggplot2和pheatmap包对差异基因进行火山图和热图展示。

1.2.2 差异基因功能富集分析:使用R studio4.1.0软件clusterProfiler包对获取的差异基因进行基因本体(Gene ontology, GO)富集分析和KEGG通路分析,基于Benjamini-Hochberg理论对 P 值进行排序和多重校正,以FDR(false discovery rate, 伪发现率) < 0.05 为显著性的阈值^[4]。选择GO注释条目

下的生物学过程(biological process, BP)细胞成分(cellular component, CC)以及分子功能(molecular function, MF)对DEGs进行注释。

1.2.3 关键基因筛选:利用STRING v10在线分析平台(<https://string-db.org/>)构建蛋白-蛋白相互作用(PPI)网络关系,导出PPI网络关系tsv数据^[5]。通过Cytoscape 3.8.0软件实现蛋白网络的可视化,同时使用其cytoHubba插件,计算PPI网络的拓扑特征,用于计算每个蛋白节点的程度,根据中心性分值的高低,前10位基因认为是中心基因^[6]。

1.3 统计学分析 差异表达基因分析采用非配对Wilcoxon检验,差异表达基因分析及富集分析中的 P 值经过BH(Benjamin and Hochberg)法校正得到FDR值。根据确定核心基因,通过R软件中的RMS包进行Logistic回归,并建立冠心病稳定型心绞痛的Nomogram模型。所有统计分析均在R软件(<https://www.r-project.org/>)中实现,除差异表达分析及富集分析外,双侧 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 稳定性心绞痛外周血差异表达基因筛选 火山图(图1A)所示,与正常受试者外周血基因表达相比,稳定性心绞痛患者有303个差异表达基因,其中上调基因160条,下调基因43条,使用前40条差异基因热图(图1B)显示,差异基因可明显将稳定性心绞痛患者和正常受试者区分开,说明稳定性心绞痛患者和正常受试者外周血表达谱存在明显差异。

2.2 差异基因富集分析 见图2。将303条差异基因使用R软件clusterProfiler包进行GO富集分析发现,在生物学过程中神经递质转运、突触传递、神经肽类信号通路等明显富集(图2A);在细胞成分中,差异基因集中在质膜、细胞外基质、突触膜等位置(图2B);在分子功能方面,主要富集在细胞外配体门控离子通道活性、突触后神经递质受体活性、神经递质结合等方面(图2C)。KEGG通路分析显示差异表达基因主要参与神经递质配体相互作用、脂肪吸收消化、钙调节信号通路、PI3K-Akt通路、NF-kappa B通路及氧化磷酸化等通路(如图2D)。

2.3 关键基因筛选 见图3。我们对差异基因进行了蛋白互作PPI网络构建,该网络包含152个点,258条边,我们采用Cytohubba插件的NCC算法进一步分析,筛选出10个关键基因BDNF, GFAP, SYN1, NES, PLG, HPGDS, KCNC1, APOA4, AMBP和TJP1。颜色由红至黄逐渐改变,表示得分逐渐降低。

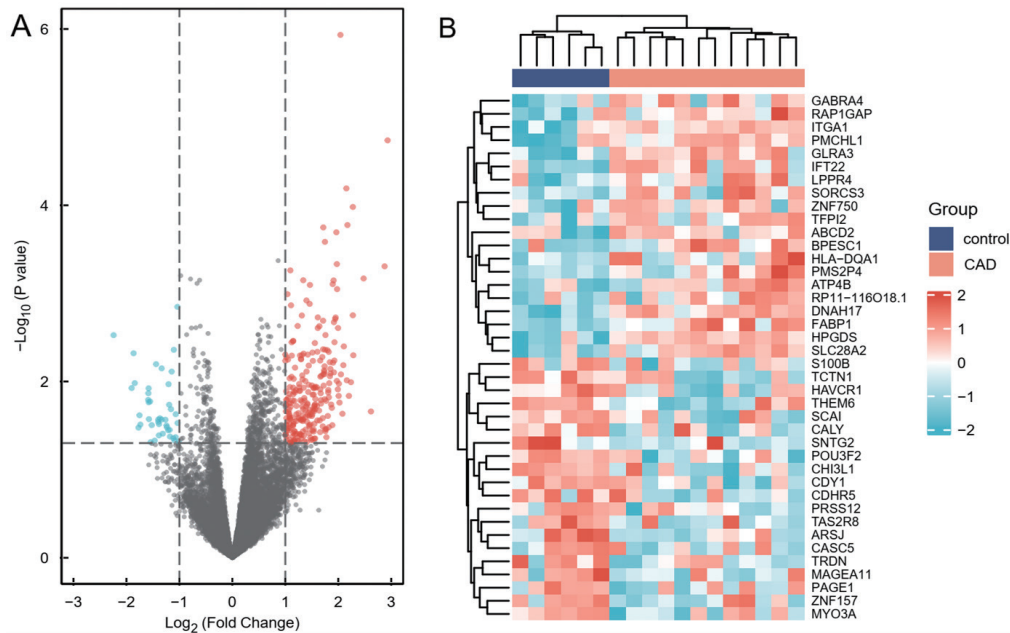
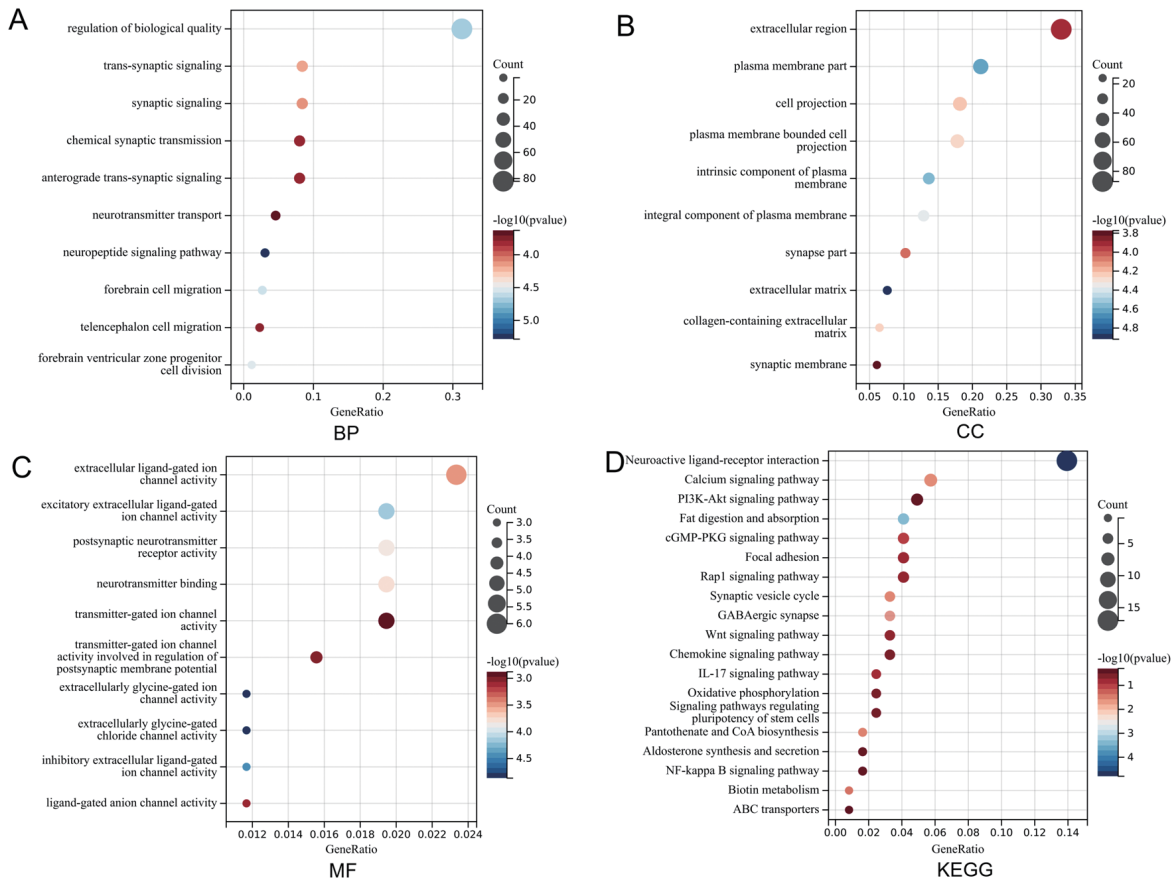


图1 稳定性心绞痛患者与正常受试者差异基因的火山图 (A) 和热图 (B)

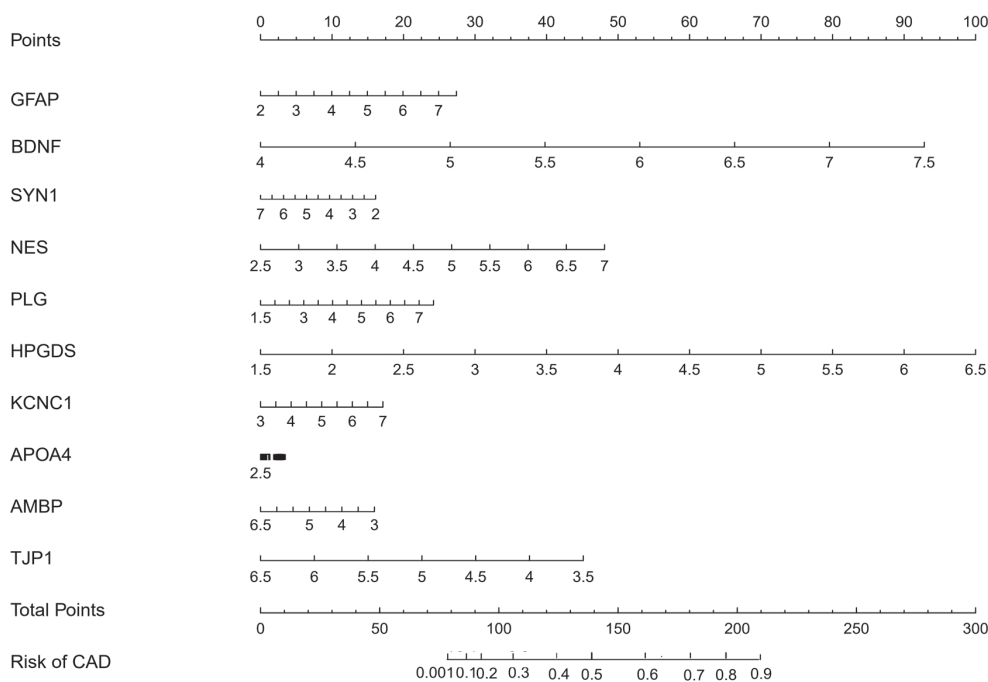
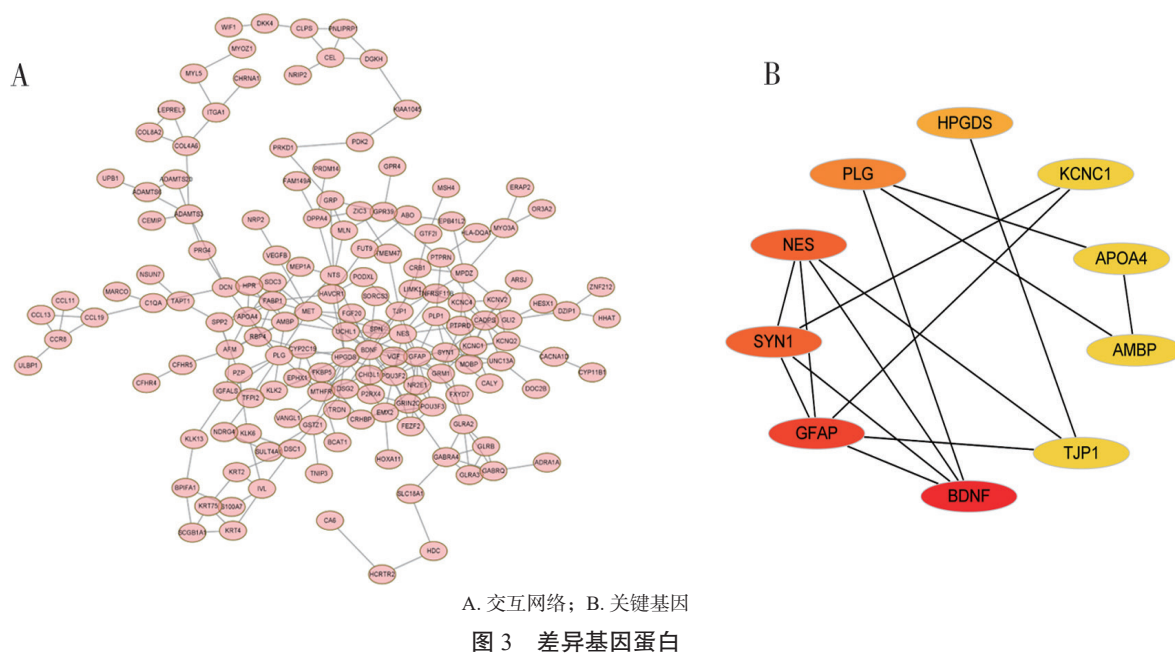
2.4 基因 Nomogram 预测诊断模型的建立 见图4。我们使用 10 个关键基因 BDNF, GFAP, SYN1, NES, PLG, HPGDS, KCNC1, APOA4, AMBP 和 TJPI 变量构建 Nomogram 模型, 根据某患者每个变量实际情况找到其对应的刻度, 向上投射到顶

部的标尺 (Points) 即可得出每个变量的分值, 分值相加即为总分值 (total points), 根据总得分值向下投射即可得到该患者诊断概率。从图 4 我们可以看出在 10 个关键基因中, BDNF, HPGDS 表达水平在模型中的权重明显较其他基因高。



A. GO: 生物学过程; B. GO: 细胞成分; C. GO: 分子功能; D. KEGG pathway 分析

图2 差异基因的富集分析



3 讨论

近年来随着基因芯片技术的成熟以及生物信息学的迅猛发展,对复杂疾病样本进行高通量分析,已成为当前揭示疾病发病机制及发现新的分子标记物常用的手段。GEO 数据库隶属于美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI), 免费提供微阵列及其他形式的高通量数据,还提供可以完全注释原始数据和经处理数据的基本信息,支持对原始数据、处理后的数据和原数据的存档,是目前全球最大、最全面的

公共基因表达资源^[3]。数据共享平台能提供给研究人员详实、可信度高的第一手资料,我们可以通过科学的算法和严谨的统计学分析,借助合理的生物信息学分析,进而帮助我们从小样本数据中筛选出关键基因,从而揭示基因在疾病发生和发展过程中的重要作用,并鉴定生物标志物来精确地预测早期诊断、预后判断并为疾病提供潜在的治疗靶点^[4-5]。本研究基于 GEO 数据库分析比较稳定性心绞痛患者和正常受试者外周血基因表达谱,发现两组外周血表达谱存在明显差异,筛选出 303 个明显差异表

达基因,并筛选出10个关键基因,这为无创外周血生物标记物早期识别稳定型心绞痛提供思路。

脑源性神经营养因子 (brain-derived neurotrophic factor, BDNF) 是1982年BARDE等首先在猪脑中发现的一种具有神经营养作用的蛋白质。BDNF由小分子二聚体蛋白质构成,通过与原肌球蛋白激酶受体B以及p75神经生长素受体结合后维持神经元分化和存活,促进轴突生长,并参与调节突触可塑性。越来越多研究显示,BDNF也存在于心脏、肺、脾脏、骨骼肌等器官,同时巨噬细胞、淋巴细胞、内皮细胞和平滑肌细胞中也有表达,且BDNF通路对心血管系统的发育及诱导其对缺血损伤的反应至关重要。ZIEROLD等^[6]研究显示巨噬细胞、平滑肌细胞和成纤维细胞是粥样硬化斑块中BDNF的主要来源,BDNF可能通过激活冠脉NADPH酶系来诱导氧化应激导致斑块不稳定。人群研究发现血浆BDNF浓度与三酰甘油、低密度脂蛋白-胆固醇和纤维蛋白原水平、糖尿病状态、年龄呈负相关,与心绞痛患者血浆高密度脂蛋白-胆固醇水平和血小板计数呈正相关^[7]。与此同时,JIN等^[8]研究发现循环中BDNF水平对稳定性心绞痛识别有重要的诊断价值,ESMAELI等^[9]报道BDNF水平与冠脉病变严重程度呈正相关。同时BDNF与微血管性心绞痛和冠脉钙化也有着密切关系^[10-11],更有人群队列研究发现BDBF水平可作为冠心病发生和全因死亡率的四年独立预测因子,低血浆BDNF水平可能与未来的冠状动脉事件和这些患者的死亡有关^[12]。与既往研究结果一致,本研究也显示稳定性心绞痛外周血BDNF表达与正常受试者存在明显差异,有望成为稳定性心绞痛潜在的诊断分子标记物。

造血前列腺素D合成酶 (hematopoietic prostaglandin D synthase, HPGDS) 是一种sigma类谷胱甘肽-S-转移酶家族成员,该酶催化前列腺素H₂ (prostaglandin H₂, PGH₂) 向前列腺素D₂ (prostaglandin D₂, PGD₂) 的转化,并在免疫系统和肥大细胞的前列腺素的生产中起作用。PGD₂是一种来源于花生四烯酸代谢的关键脂质介质,通过其受体促进炎症的消退和调节T细胞功能,进而参与了心血管疾病包括动脉粥样硬化、高血压、肺部高压、主动脉瘤以及其他代谢性疾病等的形成^[13-14]。最近有学者发现PGD₂可以通过其受体DP₁/DP₂促进血管平滑肌细胞表型转化进而导致动脉粥样硬化的形成^[15]。作为PGD₂的降解酶HPGDS在心血管疾病中的角色尚未被报道,本研究通过生物信息学方法发现HPGDS在稳定性心绞痛患者中高表达,同时在构建的Nomogram预测诊断模型中权重较高,提示其在稳定性心绞痛诊断过程中可能有潜在

的诊断价值。

总之,本研究基于GEO数据库进行生物信息学分析,发现稳定性心绞痛患者和正常受试者外周血基因表达谱存在明显差异,同时发现BDNF和HPGDS为代表的基因可能成为稳定性心绞痛的潜在的诊断分子标记物。但由于样本人群少且为非亚裔人群,有待于今后进一步扩张人群样本进行验证。

参考文献:

- [1] 中国心血管健康与疾病报告编写组. 中国心血管健康与疾病报告2019概要[J]. 中国循环杂志, 2020, 35(9): 833-854.
The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases in China 2019: an updated summary [J]. Chinese Circulation Journal, 2020, 35(9): 833-854.
- [2] KNUUTI J, WIJNS W, SARASTE A, et al. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes: the task force for the diagnosis and management of chronic coronary syndromes of the European society of cardiology (ESC) [J]. European Heart Journal, 2020, 41(3): 407-477.
- [3] 闫小妮, 田国祥, 郭晓娟, 等. GEO数据库架构、申请及数据提取方法与流程[J]. 中国循证心血管医学杂志, 2019, 11(2): 134-137.
YAN Xiaoni, TIAN Guoxiang, GUO Xiaojuan, et al. GEO database architecture, application and data extraction methods and processes [J]. Chinese Journal of Evidence-Based Cardiovascular Medicine, 2019, 11(2): 134-137.
- [4] 陈龙梅, 杨振华. 基于GEO数据库对类风湿性关节炎相关基因筛选及生物信息学分析[J]. 现代检验医学杂志, 2021, 36(2): 49-52, 78.
CHEN Longmei, YANG Zhenhua. Gene screening and bioinformatics analysis of rheumatoid arthritis based on GEO database [J]. Journal of Modern Laboratory Medicine, 2021, 36(2): 49-52, 78.
- [5] 吴良银, 李文丽, 刘俊. 基于GEO数据的病毒相关性肝癌潜在生物标志物的筛选及生物信息学分析[J]. 现代检验医学杂志, 2021, 36(6): 106-110.
WU Liangyin, LI Wenli, LIU Jun. Screening and bioinformatics analysis of potential biomarkers for virus-associated hepatocellular carcinoma based on GEO data [J]. Journal of Modern Laboratory Medicine, 2021, 36(6): 106-110.
- [6] ZIEROLD S, BUSCHMANN K, GACHKAR S, et al. Brain-derived neurotrophic factor expression and signaling in different perivascular adipose tissue depots of patients with coronary artery disease [J]. Journal of the American Heart Association, 2021, 10(6): e018322.

(下转第69页)

- [4] CERNOCH M, VIKLICKY O. Complement in kidney transplantation[J]. *Front Med (Lausanne)*, 2017, 4:66.
- [5] CHOUDHRY N, LI Ke, ZHANG Ting, et al. The complement factor 5a receptor 1 has a pathogenic role in chronic inflammation and renal fibrosis in a murine model of chronic pyelonephritis[J]. *Kidney Int*, 2016, 90(3):540-554.
- [6] BAO Xue, BORNE Y, MUHAMMAD I F, et al. Complement C3 and incident hospitalization due to chronic kidney disease: a population-based cohort study[J]. *BMC Nephrol*, 2019, 20(1):61.
- [7] LAURA R, ALBERTO O. Timing of eculizumab therapy for C3 glomerulonephritis[J]. *Clinical Kidney Journal*, 2015, 8(4): 449-452.
- [8] 饶斯逸, 张晶, 崔炯, 等. 补体 C3 参与小鼠缺血再灌注急性肾损伤后肾脏炎症及慢性纤维化 [J]. *中华高血压杂志*, 2022, 30(2):145-152.
- RAO Siyi, ZHANG Jing, CUI Jiong, et al. Complement C3 participates in inflammation and post-injury fibrosis in mice with ischemia-reperfusion induced acute kidney injury[J]. *Chinese Journal of Hypertension*, 2022, 30(2):145-152.
- [9] LIU Fang, GOU Rong, HUANG Jun. et al Effect of anaphylatoxin C3a, C5a on the tubular epithelial-myofibroblast transdifferentiation in vitro[J]. *Chin Med J (Engl)*, 2011, 124(23):4039-4045.
- [10] PENG Qi, WU Weiju, WU Kunyi, et al. The C5a/C5aR1 axis promotes progression of renal tubulointerstitial fibrosis in a mouse model of renal ischemia/reperfusion injury[J]. *Kidney Int*, 2019, 96(1):117-128.
- [11] CHMILEWSKY F, ABOUT I, CHUNG S H. Pulp fibroblasts control nerve regeneration through complement activation[J]. *J Dent Res*, 2016, 95(8):913-922.
- [12] 肖广智, 高旋, 丁进. 血清补体 C3, C4 水平正常的活动性狼疮性肾炎患者的临床特征与实验室相关指标分析 [J]. *现代检验医学杂志*, 2022, 37(5): 39-43.
- XIAO Guangzhi, GAO Xuan, DING Jin. Analysis of clinical and laboratory characteristics of the patients of active lupus nephritis with normal levels of serum complement C3 and C4[J]. *Journal of Modern Laboratory Medicine*, 2022, 37(5): 39-43.
- [13] HU Chao, LI Long, DING Peipei, et al. Complement inhibitor CR1g/FH ameliorates renal ischemia reperfusion injury via activation of PI3K/AKT signaling[J]. *J Immunol*, 2018, 201(12):3717-3730.
- [14] CHEN Wei, YUAN Hui, CAO Wenmin, et al. Blocking interleukin-6 trans-signaling protects against renal fibrosis by suppressing STAT3 activation[J]. *Theranostics*, 2019, 9(14):3980-3991.
- [15] LIU Ziming, ZHU Shiming, QIN Xiangjing, et al. Silencing of C5a receptor gene with siRNA for protection from Gram-negative bacterial lipopolysaccharide-induced vascular permeability[J]. *Mol Immunol*, 2010, 47(6):1325-1333.
- 收稿日期: 2022-10-04
修回日期: 2022-11-10

(上接第23页)

- [7] MONISHA K G, PRABU P, CHOKKALINGAM M, et al. Clinical utility of brain-derived neurotrophic factor as a biomarker with left ventricular echocardiographic indices for potential diagnosis of coronary artery disease[J]. *Scientific Reports*, 2020, 10(1): 16359.
- [8] JIN Hong, CHEN Yifei, WANG Bilei, et al. Association between brain-derived neurotrophic factor and von Willebrand factor levels in patients with stable coronary artery disease[J]. *BMC Cardiovascular Disorders*, 2018, 18(1): 23.
- [9] ESMAEILI F, MANSOURI E, EMAMI M A, et al. Association of serum level and DNA methylation status of brain-derived neurotrophic factor with the severity of coronary artery disease[J]. *Indian Journal of Clinical Biochemistry*, 2022, 37(2): 159-168.
- [10] ASLAN G, POLAT V, BOZCALI E, et al. Evaluation of serum platelet-derived growth factor receptor-β and brain-derived neurotrophic factor levels in microvascular angina[J]. *Anatolian Journal of Cardiology*, 2020, 24(6): 397-404.
- [11] JIN Hong, JI Jingjing, ZHU Yi, et al. Brain-derived neurotrophic factor, a new predictor of coronary artery calcification[J]. *Clinical and Applied Thrombosis/Hemostasis*, 2021, 27: 1076029621989813.
- [12] KAESS B M, PREIS S R, LIEB W, et al. Circulating brain-derived neurotrophic factor concentrations and the risk of cardiovascular disease in the community[J]. *Journal of the American Heart Association*, 2015, 4 (3): e001544.
- [13] KONG Deping, YU Ying. Prostaglandin D(2) signaling and cardiovascular homeostasis[J]. *Journal of Molecular and Cellular Cardiology*, 2022, 167: 97-105.
- [14] PAWELZIK S C, AVIGNON A, IDBORG H, et al. Urinary prostaglandin D(2) and E(2) metabolites associate with abdominal obesity, glucose metabolism, and triglycerides in obese subjects[J]. *Prostaglandins & Other Lipid Mediators*, 2019, 145: 106361.
- [15] LEE H S, YUN S J, HA J M, et al. Prostaglandin D(2) stimulates phenotypic changes in vascular smooth muscle cells[J]. *Experimental & Molecular Medicine*, 2019, 51(11): 1-10.
- 收稿日期: 2022-03-29
修回日期: 2022-07-20