· 基础研究 ·

基因表达数据库对冠心病关键调控基因的生物信息学分析

艾伦娜[1,2],刘昱圻[1,3],陈韵岱[3*]

(中国人民解放军总医院:[1]第一医学中心心血管内科,[2]研究生院,北京 100853;[3] 中国人民总医院第六医学中心心血管病医学部,北京 100142)

【摘 要】 **目的** 应用基因表达(GEO)数据库对冠心病患者相关基因进行研究分析,探索新的冠心病潜在调控机制。**方法** 检索 GEO 数据库,选择冠心病患者和对照者的相关基因表达谱数据。导入处理后的冠心病患者的 GSE71226 数据集,利用 GEO2R 分析鉴定冠心病和健康人的差异表达基因(DEGs),通过分析确定冠心病患者和对照者的差异基因。采用 R 语言 ggplot2 包进行可视化相关分析。**结果** 共筛选出 575 个 DEGs。进一步构建了一个蛋白质-蛋白质相互作用(PPI)网络,以可视化 DEGs 和识别中心基因。基因本体论通路富集分析显示,上调和下调 DEGs 的各排名前 15 的信号通路参与了胰岛素分泌、心脏传导系统、生长激素分泌等相关的生物过程;京都基因与基因组百科通路富集分析显示,上调和下调 DEGs 主要与序列特异性 DNA 复制、环状腺苷单磷酸酯(cAMP)、脂肪消化吸收酶等相关的信号通路等有关。进一步构建了 PPI 网络互作图,显示五个中心基因(*SST*、*BBS*10、*CCK*、*POMC*、*HSPA*9),其中 *SST* 与冠心病的预后密切相关。**结论** 冠心病的关键调控基因包括 *SST*、*BBS*10、*CCK*、*POMC*、*HSPA*9,可能通过双链 DNA 结合、cAMP、脂肪消化吸收酶等相关的信号通路参与冠心病发病。

【关键词】 冠心病;差异表达基因;关键调控基因;生物信息学分析;通路

【中图分类号】 R541.4 【文献标志码】 A 【DOI】 10.11915/j.issn.1671-5403.2023.06.093

Bioinformatics analysis of hub regulatory genes of coronary heart disease using Gene Expression Omnibus database

Ai Lunna[1,2], Liu Yuqi[1,3], Chen Yundai[3*]

([1]Department of Cardiology, First Medical Center, [2]Medical School, Chinses PLA General Hospital, Beijing 100853, China; [3]Department of Cardiology, Sixth Medical Center, Chinese PLA General Hospital, Beijing 100142, China)

【**Abstract**】 **Objective** To analyze genes related to coronary heart disease (CHD) patients and to explore potential new mechanisms for CHD using Gene Expression Omnibus (GEO) database. **Methods** The gene expression profile dataset of CHD patients and healthy controls was searched and retrieved from the GEO database. Differentially expressed genes (DEGs) between CHD patients and controls were identified using GEO2R analysis of the processed GSE71226 dataset. Visual correlation analysis was performed using the R language ggplot2 package. **Results** A total of 575 DEGs were screened. A protein-protein interaction (PPI) network was further constructed to visualize the DEGs and identify the genes. Gene ontology pathway enrichment analysis showed that the top 15 signaling pathways upregulating and downregulating DEGs were involved in biological processes, including insulin secretion, cardiac conduction system, growth hormone secretion, *etc*. Kyoto encyclopedia of genes and genomes pathway enrichment analysis showed that upregulating and downregulating DEGs were mainly related to sequence-specific DNA replication, cyclic adenosine monophosphate (cAMP), fat digestion and absorption, and other related signaling pathways. PPI network interaction diagram showed five central genes (*SST*, *BBS*10, *CCK*, *POMC*, *HSPA*9), among which *SST* was closely related to the pathogenesis of CHD. **Conclusion** The hub regulatory genes of CHD include *SST*, *BBS*10, *CCK*, *POMC*, *HSPA*9, probably participating in the pathogenesis of CHD through DNA replication, cAMP, fat digestion and absorption, and other related signal pathways.

【**Key words**】 coronary heart disease; differentially expressed genes; hub regulatory genes; bioinformatics analysis; pathways

Corresponding author: *Chen Yundai*, *E-mail*: *cyundai@vip.163.com*

冠心病是世界范围内最常见的心血管疾病[1],其中急性心肌梗死是导致死亡的主要原因之一[2]。基因调控对冠心病的发生和发展具有重要作用[3],在不同种族、性别患者中均具有一定影响[4,5]。最近,涉及数十万人的全基因组关联研究发现了许多新的冠心病相关基因位点,如影响血脂代谢和血压等的相关基因。目前,面临的主要挑战之一是进一步探索这些基因中的致病基因,并了解其导致冠心

收稿日期:2023-01-16;**接受日期**:2023-03-28

通信作者:陈韵岱,E-mail:cyundai@vip.163.com

病[6]的可能的分子机制。因此,寻找潜在的分子标志物为冠心病的早期诊断和临床治疗提供了可能。微阵列分析,别名基因表达谱,被发现对揭示分子机制和预测疾病预后有关键影响[7]。大多数微阵列数据存储在公共数据库中,如基因表达(gene expression omnibus, GEO)数据库和 Oncomine 数据库,其中 GEO 数据库是由美国国立生物技术信息中心(National Center for Biotechnology Information, NCBD)于2000年创建并维护至今的高通量基因表达数据库,包括了多种疾病相关的高通量基因表达数据。近年来,随着高通量测序和表达谱芯片手段不断发展进步,产生了海量的测序数据,生物信息学作为一门新的学科领域也由此得到了快速发展。通过生物信息学手段,可以从基因层面分析生物学功能,从而进一步了解基因表达的调控机制,根据生物分子在基因调控中的作用,进而发现人类疾病潜在的诊断及药物的分子靶点或疾病相关分子机制,心血管领域也已有相关研究。Chen 等[8]通过生物信息分析来自微阵列的基因表达谱数据,确定了单核细胞相关的3个关键基因(*CUX*1、*CTSD* 和 *ADD*3)作为早期识别急性心肌梗死(acute myocardial infarction,AMI)患者发生心力衰竭风险的潜在生物标志物;Zhu 等[9]在 GEO 数据库中获得6个数据集,并通过生物信息手段对它们进行了识别免疫相关基因的分析及诊断模型的构建,最后发现 AMI 组 C-C 趋化因子受体(C-C chemokine receptor, CCR)、巨噬细胞、中性粒细胞、Treg 细胞的比例明显高于正常组,免疫相关基因和免疫细胞与 AMI 密切相关,除此外仍有大量与 AMI 相关的生物信息学数据有待探索。本研究对 GSE71226 中冠心病患者和健康人的差异表达基因(differentially expressed genes, DEGs)进行了检测,并对这些基因进行了基因本体论(gene ontology, GO)和京都基因与基因组百科(Kyoto encyclopedia of genes and genomes, KEGG)分析。也建立了蛋白质-蛋白质相互作用(protein-protein interaction, PPI)网络,以可视化 DEGs 和识别枢纽基因。本研究旨在对 GEO 数据库中的基因表达谱进行生物信息学分析,进一步研究探讨冠心病相关的新的分子标志物,为冠心病机制探索提供新思路。

1 材料与方法

1.1 数据采集

GEO (http://www.ncbi.nlm.nih.gov/geo)是用于存储微阵列数据[10]的开放基因组数据库。孟等[11]提交的基因表达数据集 GSE71226 有6个样本,包括3例冠心病患者样本和3名健康中国汉族样本。

1.2 鉴定差异基因

应用 GEO2R(https://www.ncbi.nlm.nih.gov/geo/geo2r/)进行冠心病患者外周血总 RNA 与健康人的 DEGs 鉴定[12]。GEO2R 是基于 Web 的程序、用于检测 GEO 系列中的 DEGs、且基于 R 语言的 limma 包[13]。矫正后的 $P<0.05$ 和 $|logFC|\geq 1$ 作为上调和下调的选择条件。

1.3 差异基因的 KEGG 和 GO 富集

注释、可视化和综合发现数据库 DAVID (https://david.ncifcrf.gov/)是执行 GO 和 KEGG[8]的有效方法。GO 提供了转录组数据的分析平台以解释基因和基因产物在分子功能(molecular function, MF)、生物过程(biological progress, BP)和细胞成分(cellular component, CC)[9]三个关键领域的特征。KEGG 是研究基因和分子系统基因功能的系统数据库[14]。它是一种从分子水平理解生物功能的数据源,特别是海量分子数据集的评估。$P<0.05$ 为差异有统计学意义。采用 R 语言 ggplot2 包进行可视化相关分析。

1.4 PPI 网络构建

检索相互作用基因的搜索工具(STRING, http://www.string-db.org/)研究 DEGs 的 PPI 网络,Cytoscape(San Diego, California)进行绘制[15]。STRING 数据库收集、排序和组合所有公开可访问的 PPI 数据资源,并通过计算预测进行匹配。CytoHubba 根据连接程度选择枢纽基因。CytoHubba 为 Cytoscape 应用程序,它使用一些拓扑计算预测和研究每个系统[16]中的重要枢纽和子网络。

2 结 果

2.1 差异基因鉴定

GSE71226 数据集包括3例冠心病和3名健康样本。应用 GEO2R 技术探索2组差异表达的基因。最终发现了225个上调的 DEGs 和350个下调的 DEGs(图1)。

2.2 GO 和 KEGG 鉴定

为了更系统地理解这些筛选的 DEGs,本研究通过 DAVID 进行了功能分析。上调的前15名 DEGs 主要富含 BP,涉及调节胰岛素分泌、心脏传导系统发育和心脏心房形态发生。在 CC 中,二纤维细胞主要与细胞-细胞连接、基底外侧血浆膜和高尔基腔有关。此外,MF 分析发现,DEGs 主要富含序列特异性的双链 DNA 结合、双链 DNA 结合和丝氨酸型内肽酶活性(图2A)。下调的前15名 DEGs 在 BP 的非运动纤毛组装、线粒体内膜组织和中心体蛋白定位中富集。它们在 CC 中富含 S-腺苷蛋氨酸

(S-adenosylmethionine, SAM)复合物、线粒体膜间空间桥接(mitochondrial inter-membrane-space bridging, MIB)复合物和线粒体外膜转位酶复合物。它们在 MF 中富含 snoRNA 结合、核糖核酸内切酶活性、产生 5′-磷酸单核酸酯和 M7g(5′)pppN 二磷酸酶活性(图 2B)。

KEGG 通路的富集图谱详见图 3。这些上调的 DEGs 富含神经活性配体-受体相互作用、环状腺苷单磷酸酯(cyclic adenosine monophosphate, cAMP)信号通路、GnRH 分泌和脂肪消化吸收,而下调的 DEGs 富含 DNA 复制、过氧化物酶体、RNA 转运和维生素消化。

2.3 PPI 网络建设

构建 PPI 网络,研究 DEGs 在冠心病外周血总 RNA 中的表达产物。上调的 PPI 网络有 713 个节点和 1 687 条边,下调的 PPI 网络有 410 个节点和 882 条边。基于图 4 中的连通性显示了 PPI 网络。前 5 个枢纽基因分别为 *SST*、*CCK*、*POMC*、*HSPA*9 和 *BBS*10。

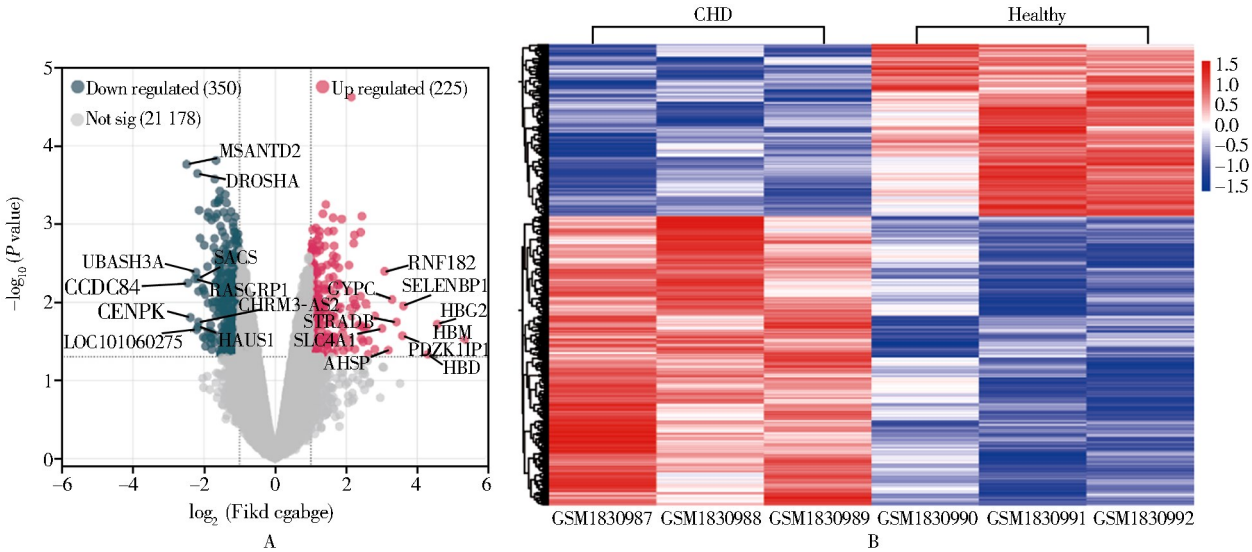


图 1 GSE71226 的差异表达分析

Figure 1 Differential expression analysis of GSE71226

A: volcano plot; B: clustering heat map of DEGs, red indicates up-regulated DEGs and blue indicates down-regulated DEGs. DEGs: differentially expressed genes; CHD: coronary heart disease.

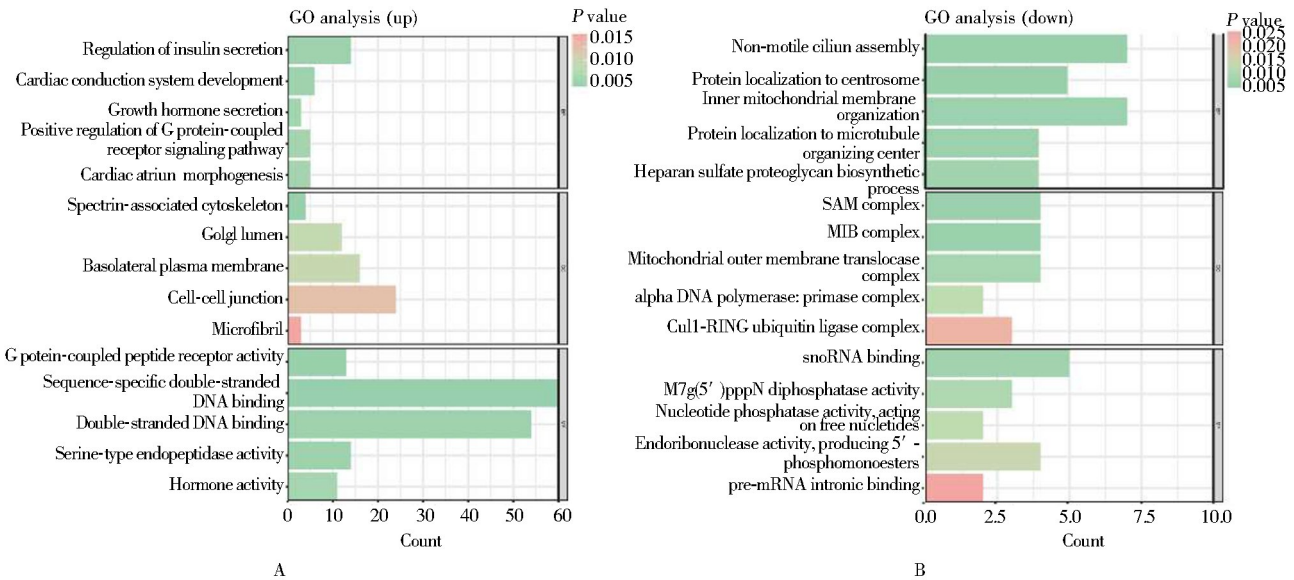


图 2 GO 富集分析

Figure 2 GO enrichment analysis

A: GO enrichment analysis of up-regulated genes in BP, CC and MF; B: GO enrichment analysis of down-regulated genes in BP, CC and MF. GO: gene ontology; SAM: S-adenosylmethionine; MIB: mitochondrial intermembrane spacebridging; BP: biological progress; CC: cellular component; MF: molecular function.

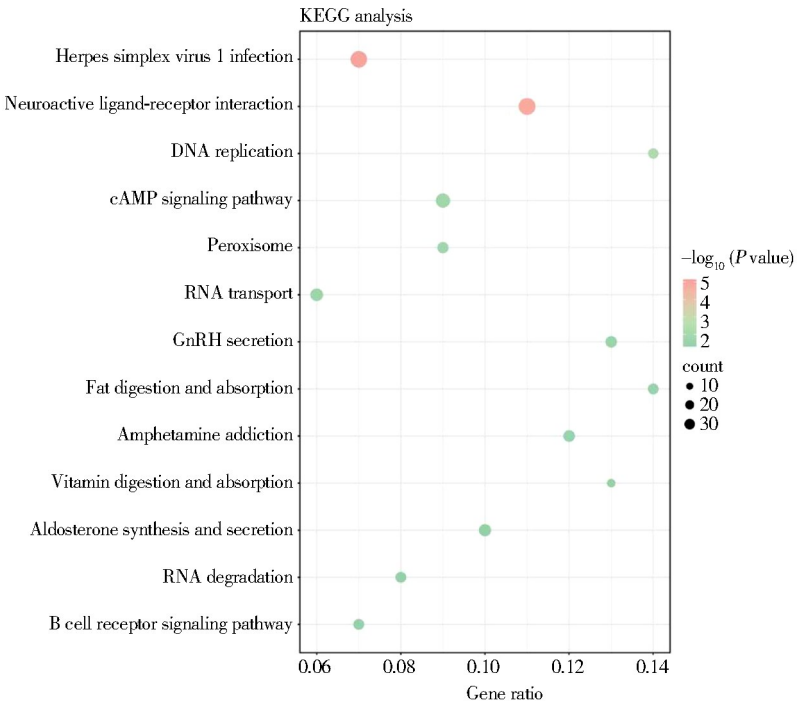


图3 DEGs的KEGG分析

Figure 3 KEGG analysis of DEGs

Red is the KEGG pathway upregulated by DEGs, blue is the KEGG pathway downregulated by DEGs. The higher the count, the higher the correlation between enriched pathways and DEGs. DEGs: differentially expressed genes; KEGG: Kyoto encyclopedia of genes and genomes.

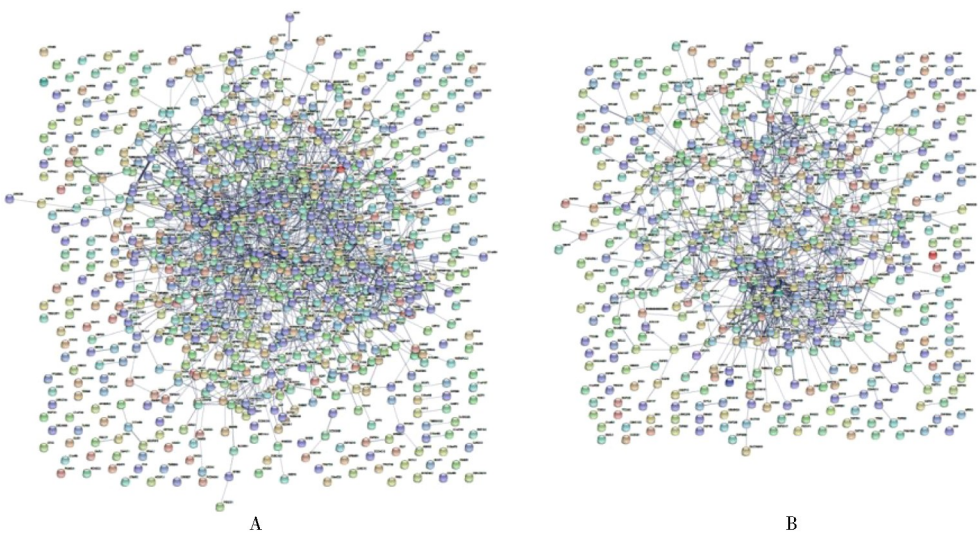


图4 PPI网络

Figure 4 PPI network

A: PPI network diagram of up-regulated DEGs; B: PPI network diagram of down-regulated. Circles represent genes and the linkage between two genes represents protein interactions. DEGs: differentially expressed genes; PPI: protein-protein interaction.

3 讨 论

冠心病,即冠状动脉动脉粥样硬化性心脏病,属常见心脏病[17]。经皮冠状动脉腔内成形术已被应用于许多冠心病患者的初期阶段[18]。然而,冠心病症状在早期往往不典型。因此,发现相关临床生物信息学对冠心病防治变得越来越重要。

本研究发现多种关键基因相关功能与冠心病的进展相关,或可能参与冠心病的发生发展,如序列特异性双链DNA结合,已被证明与心脏病相关。端粒是一种富含g的串联DNA序列,呈双链状,随着细胞分裂而缩短,它们可以通过保存和延长端粒减缓

这一过程。一些相关研究表明,心肌细胞和内皮细胞中端粒酶的表达和活性显著升高,表明端粒酶调节心脏疾病中的组织修复[19]。这表明,双链 DNA 结合是治疗冠心病的一个重要靶点。

同时,本研究也发现了与冠心病进展相关的多种信号通路,如 cAMP 信号通路。环核苷酸 cAMP 和 cGMP 是心脏生理和病理的重要调节因子,也是心脏细胞中的主要第二信使。在环核苷酸水解磷酸二酯酶的水平上,有越来越多的证据表明 cAMP 和 cGMP 信号通路磷酸二酯酶之间存在功能上显著的串扰[20]。在多种心脏疾病中,cAMP/cGMP 微域和亚细胞浓度分布被强烈重构,导致磷酸二酯酶进一步参与 cAMP/cGMP 信号通路调节[20]。

本研究构建了 PPI 网络,并在冠心病 *SST*、*CCK*、*POMC*、*HSPA*9 和 *BBS*10 中鉴定了几个枢纽基因。根据文献报道,这些枢纽基因与心脏病有关[21-23]。例如,载脂蛋白 C3(apolipoprotein C-Ⅲ,APOC3)是一种脂蛋白,而脂蛋白是动脉粥样硬化的重要因素。因此,参与脂蛋白代谢的基因可能是冠心病的候选基因[24]。*APOC*3 基因具有许多多态性,包括 *C*-482*T*、*T*-455*C*、*SST*1 和 *C*1100*T*[25]。其中,*SST* 是我们从 PPI 网络中提取的关键基因之一,其多态性与冠心病风险之间的关系仍需要进一步研究。

*BBS*10 是 Bardet-Biedl 综合征(Bardet-Biedl syndrome, BBS)基因家族[26]的成员。BBS 是一种常染色体退性功能障碍,以进行性视网膜变性、多指和认知障碍为特征。许多 BBS 蛋白存在于细胞的基底体、纤毛轴突和中心周区域,它们通过微管相关运输[27]在细胞内发挥作用。它们可能影响其他纤毛或基底体蛋白作为分子伴侣的折叠或稳定性,具体表现为其表达阻断前脂肪细胞的纤毛发生,从而增加心肌纤维化的可能性。一般来说,心肌纤维化与心脏代谢异常、心肌缺血、冠状动脉粥样硬化[28]等因素有关。因此,寻找新的肌纤维治疗靶点对冠心病[29]及 CCK、POMC、HSPA9 的研究也很有帮助,目前还没有相关研究表明这些枢纽基因与冠心病有很强的相关性。这些差异表达基因可能与疾病相关,但也可能只是伴随现象,未来仍然需要进一步验证分析。

综上,对冠心病患者和健康样本的芯片数据进行分析,揭示了冠心病的部分机制。本研究分析的枢纽基因和信号通路可能是冠心病诊断的有前途的生物标志物和潜在靶点。

【参考文献】

[1] Vilahur G, Badimon JJ, Bugiardini R, *et al*. Perspectives: the burden of cardiovascular risk factors and coronary heart disease in Europe and worldwide[J]. Eur Heart J Suppl, 2014, 16(Suppl_A): A7-A11. DOI: 10.1093/eurheartj/sut003.

[2] Anderson JL, Morrow DA. Acute myocardial infarction[J]. N Engl J Med, 2017, 376(21): 2053-2064. DOI: 10.1056/NEJMra1606915.

[3] Ma X, Li SF, Qin ZS, *et al*. Propofol up-regulates expression of ABCA1, ABCG1, and SR-B1 through the PPARγ/LXRα signaling pathway in THP-1 macrophage-derived foam cells[J]. Cardiovasc Pathol, 2015, 24(4): 230-235. DOI: 10.1016/j.carpath.2014.12.004.

[4] Ren M, Reilly RT, Sacchi N. Sasa health exerts a protective effect on Her2/NeuN mammary tumorigenesis[J]. Anticancer Res, 2004, 24(5A): 2879-2884.

[5] Xie X, Pan X, Zhang W, *et al*. A context hierarchical integrated network for medical image segmentation[J]. Comput Electr Eng, 2022, 101: 108029. DOI: 10.1016/j.compeleceng.2022.108029.

[6] Xie X, Pan X, Shao F, *et al*. Mci-net: multi-scale context integrated network for liver CT image segmentation[J]. Comput Electr Eng, 2022, 101: 108085. DOI: 10.1016/j.compeleceng.2022.108085.

[7] Zhou M, Zhao H, Wang Z, *et al*. Identification and validation of potential prognostic lncRNA biomarkers for predicting survival in patients with multiple myeloma[J]. J Exp Clin Cancer Res, 2015, 34(1): 102. DOI: 10.1186/s13046-015-0219-5.

[8] Chen Q, Yin Q, Song J, *et al*. Identification of monocyte-associated genes as predictive biomarkers of heart failure after acute myocardial infarction[J]. BMC Med Genomics, 2021, 14(1): 1-13. DOI: 10.1186/s12920-021-00890-6.

[9] Zhu X, Yin T, Zhang T, *et al*. Identification of immune-related genes in patients with acute myocardial infarction using machine learning methods[J]. J Inflamm Res, 2022, 15: 3305-3321. DOI: 10.2147/JIR.S360498.

[10] Gao LJ, Zhao X, Li JG, *et al*. Bioinformatics analysis of genes related to pathogenesis of major depression disorder[J]. Acta Physiol Sin, 2018, 70(4): 361-368. DOI: 10.13294/j.aps.2018.0053.

[11] Meng F, Fan L. Expression data from coronary heart disease in Chinese Han people[EB/OL]. [2019-03-25]. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71226.

[12] Li GM, Zhang CL, Rui RP, *et al*. Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy[J]. Eur Rev Med Pharmacol Sci, 2018, 22(11): 3553-3569. DOI: 10.26355/eurrev_201806_15182.

[13] Hu G, Su Y, Kang B H, *et al*. High-throughput phenotypic screen and transcriptional analysis identify new compounds and targets for macrophage reprogramming[J]. Nat Commun, 2021,

12(1): 773. DOI: 10.1038/s41467-021-21066-x.

[14] Kanehisa M, Sato Y, Kawashima M, *et al*. KEGG as a reference resource for gene and protein annotation[J]. Nucleic Acids Res, 2016, 44(D1): D457-D462. DOI: 10.1093/nar/gkv1070.

[15] Su G, Morris J H, Demchak B, *et al*. Biological network exploration with Cytoscape 3[J]. Curr Protocols Bioinform, 2014, 47(1): 8-13. DOI: 10.1002/0471250953.bi0813s47.

[16] Zhao CH, Qu L, Zhang H, *et al*. Identification of breast cancer-related circRNAs by analysis of microarray and RNA-sequencing data: an observational study[J]. Medicine, 2019, 98(46): e18042. DOI: 10.1097/MD.0000000000018042.

[17] Montone RA, Niccoli G, Russo M, *et al*. Clinical, angiographic and echocardiographic correlates of epicardial and microvascular spasm in patients with myocardial ischaemia and non-obstructive coronary arteries[J]. Clin Res Cardiol, 2020, 109(4): 435-443. DOI: 10.1007/s00392-019-01523-w.

[18] 陈丽芳. 灯盏花素注射液治疗冠心病心绞痛的临床应用[J]. 中国循证心血管医学杂志, 2016, 8(4): 500-502. DOI: 10.3969/j.issn.1674-4055.2016.04.36.
Chen LF. Clinical effect observation of breviscapine injection on the treatment of coronary heart disease angina pectoris[J]. Chin J Evid-Based Cardiovasc Med, 2016, 8(4): 500-502. DOI: 10.3969/j.issn.1674-4055.2016.04.36.

[19] Ogami M, Ikura Y, Ohsawa M, *et al*. Telomere shortening in human coronary artery diseases[J]. Arterioscler Thromb Vasc Biol, 2004, 24(3): 546-550. DOI: 10.1161/01.ATV.0000117200.46938.e7

[20] Weber S, Zeller M, Guan K, *et al*. PDE2 at the crossway between cAMP and cGMP signalling in the heart[J]. Cell Signal, 2017, 38: 76-84. DOI: 10.1016/j.cellsig.2017.06.020.

[21] Lotierzo M, Dupuy AM, Kalmanovich E, *et al*. sST2 as a value-added biomarker in heart failure[J]. Clin Chim Acta, 2020, 501: 120-130. DOI: 10.1016/j.cca.2019.10.029.

[22] Kurosawa M, Iijima S, Funakoshi A, *et al*. Cholecystokinin-8 (CCK-8) has no effect on heart rate in rats lacking CCK-A receptors[J]. Peptides, 2001, 22(8): 1279-1284. DOI: 10.1016/s0196-9781(01)00452-1.

[23] Millington WR, Rosenthal DW, Unal CB, *et al*. Localization of pro-opiomelanocortin mRNA transcripts and peptide immunoreactivity in rat heart[J]. Cardiovasc Res, 1999, 43(1): 107-116. DOI: 10.1016/s0008-6363(99)00076-0.

[24] Mounkes LC, Zhong W, De Silva HV, *et al*. Evaluation of the role of lipoprotein metabolism genes in systemic cationic liposome-mediated gene transfer *in vivo*[J]. Hum Gene Ther, 2001, 12(16): 1939-1954. DOI: 10.1089/104303401753204526.

[25] Olivieri O, Stranieri C, Bassi A, *et al*. ApoC-Ⅲ gene polymorphisms and risk of coronary artery disease[J]. J Lipid Res, 2002, 43(9): 1450-1457. DOI: 10.1194/jlr.M200145-JLR200.

[26] Wei Q, Gu YF, Zhang QJ, *et al*. Lztfl1/BBS17 controls energy homeostasis by regulating the leptin signaling in the hypothalamic neurons[J]. J Mol Cell Biol, 2018, 10(5): 402-410. DOI: 10.1093/jmcb/mjy022.

[27] Hey CAB, Larsen LJ, Tümer Z, *et al*. BBS proteins affect ciliogenesis and are essential for hedgehog signaling, but not for formation of iPSC-derived RPE-65 expressing RPE-like cells[J]. Int J Mol Sci, 2021, 22(3): 1345. DOI: 10.3390/ijms22031345.

[28] Liu S, Duan W, Zhou X. Cardioprotective effects of HO-1-loaded collagen-targeted phase change nanoparticles on cardiomyocytes following acute myocardial infarction[J]. Sci Adv Mater, 2021, 13(9): 1685-1690. DOI: 10.1166/sam.2021.4056.

[29] Guo W, Wise ML, Collins FW, *et al*. Avenanthramides, polyphenols from oats, inhibit IL-1β-induced NF-κB activation in endothelial cells[J]. Free Rad Biol Med, 2008, 44(3): 415-429. DOI: 10.1016/j.freeradbiomed.2007.10.036.

(编辑: 温玲玲)