

doi: 10.3978/j.issn.2095-6959.2019.01.005  
View this article at: <http://dx.doi.org/10.3978/j.issn.2095-6959.2019.01.005>

## 基于GEO数据库芯片的心肌梗死标志物的筛选与生物信息学分析

魏文<sup>1</sup>, 汪逵<sup>2</sup>

(恩施土家族苗族自治州中心医院 1. 内科心血管中心; 2. 神经外科, 湖北 恩施 445000)

**[摘要]** 目的: 对首次急性心肌梗死(acute myocardial infarction, AMI)患者基因芯片数据进行生物信息学分析, 寻找表达特征基因谱。方法: 利用GEO数据库中高通量基因芯片数据筛选出AMI信息的芯片。采用GO基因功能注释并进行蛋白质相互作用网络可视化分析。结果: 经过数据分析, 这些差异性表达的基因被富集到不同的生物学过程或分子功能的子集中。结论: 通过对GEO数据库中AMI的表达数据分析研究而选出的22个差异表达基因, 可为该疾病的早期诊断治疗和靶向药物的开发提供重要理论依据。

**[关键词]** 心肌梗死; 生物信息学; GEO数据库

## Screening and bioinformatics analysis of myocardial infarction markers Based on GEO database chip

WEI Wen<sup>1</sup>, WANG Kui<sup>2</sup>

(1. Department of Cardiology; 2. Department of Neurosurgery, Enshi Tujia and Miao Autonomous Prefecture Central Hospital, Enshi Hubei 445000, China)

**Abstract** **Objective:** To perform bioinformatics analysis on the gene chip data of patients with first acute myocardial infarction (AMI) to find the gene expression profile. **Methods:** We used the high-throughput gene chip data in the GEO database to screen out the AMI information chip. The GO gene function was used for annotation and visual analysis of the protein interaction network. **Results:** After data analysis, these differentially expressed genes were enriched in subsets of different biological processes or molecular functions. **Conclusion:** Twenty-two differentially expressed genes were selected by analyzing the expression data of AMI in the GEO database, which can provide an important theoretical basis for the early diagnosis and treatment of this disease and the development of targeted drugs.

**Keywords** myocardial infarction; bioinformatics; GEO database

冠状动脉疾病(coronary artery disease, CAD)是一种常见的复杂疾病, 是由动脉粥样硬化、炎症和代谢疾病引起的。急性心肌梗死(acute myocardial infarction, AMI)是CAD的一种特殊类型。已知的

冠心病危险因素包括高血压、吸烟、高脂血症、糖尿病和炎症。近年来研究<sup>[1]</sup>表明: 脂质代谢和炎症在动脉粥样硬化的发生、发展中起关键作用, 尤其是斑块破裂和AMI。尽管在药物治疗、血运重建以

及器官移植等方面取得了重大进展, 但CAD仍然是成人死亡的主要原因<sup>[2-3]</sup>。传统的心血管危险因素(包括高血压, 糖尿病和吸烟)的评估在疾病预防和预后方面具有重要作用, 但不足以完全预测复发事件的风险<sup>[4-5]</sup>。分子标志如脑钠肽(brain natriuretic peptide, BNP)、C型反应性蛋白(C-reactive protein, CRP)和其他血清炎症标志物在这方面获得了越来越多的关注, 但仅提供了适度增加的预测能力, 寻找更敏感的疾病标志物以揭示疾病的初始失调和潜在机制是当前研究急需解决的问题<sup>[6]</sup>。

不同群体的全基因组关联研究<sup>[7-8]</sup>发现了大量的CAD基因位点, 但群体遗传位点占CADs的10%。迄今为止, CAD(包括AMI)的遗传原因和潜在分子机制仍不清楚。因此, 低频率和罕见的遗传变异大的影响可能是人类常见疾病的主要原因<sup>[8]</sup>。这些生物标志物可以阐明新的分子缺陷, 并为疾病的发生发展的预测提供更精确的证据。

目前的心脏生物标志物基于已知炎症和机体代谢途径而开发。因此, 这些生物标志物提供了与已知或正在测量的信息相关的信息。相比之下, 新兴技术开始允许系统、全面地揭示与疾病状况和疾病结果相关的基因变异<sup>[9]</sup>。基因表达谱的微阵列提供了一个全面的工具来探索疾病的基本机制, 并确定疾病相关基因和失调的途径<sup>[10]</sup>。既往研究<sup>[11-12]</sup>中也证实多种心肌疾病应用基于组织的微阵列基因表达谱有助于鉴别缺血性与非缺血性病因。然而, 基于血液的全基因组基因表达谱来探索首次AMI后患者的疾病发病机制和疾病进展的研究开展较少。

本研究检索GEO数据库中首次AMI后患者相关基因芯片, 鉴定差异基因, 并利用GEO等在线数据库系统地分析差异基因的相互关系及其调控网络, 为深入研究AMI遗传疾病的分子机制及早期诊断提供思路。

## 1 资料与方法

### 1.1 数据集的获取和资料分析

以“Acute myocardial infarction”作为搜索词, 进入美国国立生物中心的GEO数据库([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo))中搜索已公布的AMI基因芯片数据集, 获取一个数据集GSE48060, GSE48060来源于梅奥诊所医学系内科及移植中心(Division of General Internal Medicine and Transplant Center, Department of Medicine, Mayo Clinic)。采用Affymetrix Human Genome U133 Plus 2.0 Array平台, 以外周血为研究组织, 纳入对象为首次AMI患

者, 病程不超过48 h, 最终纳入31例患者, 以22例正常人作对照。

### 1.2 差异表达基因的分析

基因差异表达分析采R语limma包进行, limma包是基于bioconductor专门用于处理表达谱芯片数据的基于R语言的软件包。差异基因的筛选条件为差异倍数log fold change>1, adjust<0.05。

### 1.3 功能富集分析和蛋白的相互作用

获得差异表达以后采用DAVID(<http://david.abcc.ncifcrf.gov>)数据库。并对以上基因采用string-db数据库进行蛋白相互作用网络分析。P<0.05为筛选条件。

## 2 结果

### 2.1 AMI 患者基因差异表达分析

进行差异分析之前, 需要对数据进行预处理。如存在缺失值或基因对应多个探针等特殊情况。随后进行背景校正、标准化以及汇总, 根据图1可以看出, 处理后的芯片数据值分布效果较好。进一步进行差异分析, 发现与健康对照组比较, AMI患者外周血样本中22个基因存在差异表达, 其中20个表达下调以及2个表达上调(图2, 表1)。

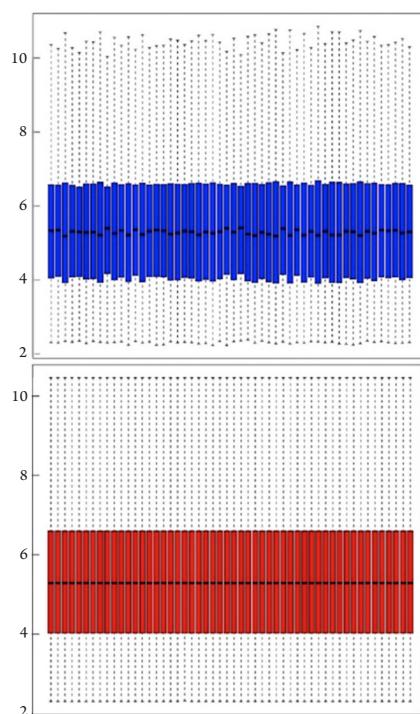


图1 各样本芯片数据进行预处理的箱体图

Figure 1 Box diagram of pre-processing of each sample chip data

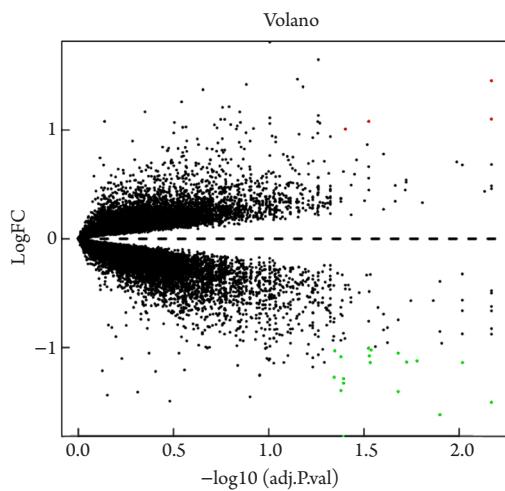


图2 差异基因表达谱的火山图

Figure 2 Volcanic map of differential gene expression profiles

## 2.2 GO 富集分析

GO分析发现差异基因功能主要集中在以下几个方面：GO:0016021~integral component of membrane(膜的整体组分)；GO:0050776~regulation of immune response(免疫应答调节)；GO:0004888~transmembrane signaling receptor activity(跨膜信号受体活性)；GO:0016020~membrane(膜)以及GO:0005886~plasma membrane(质膜)。对于符合要求的以上的GO，采用R语言ggplot2包进行可视化呈现(表2，图3)。

## 2.3 蛋白相互作用网络可视化分析

对于分析获得的差异基因，进一步采用string数据库进行蛋白相互作用网络分析呈现，发现有5个基因之间存在直接或者间接联系(图4)。

表1 AMI患者的外周血基因表达差异分析

Table 1 Difference analysis of gene expression in peripheral blood of patients with AMI

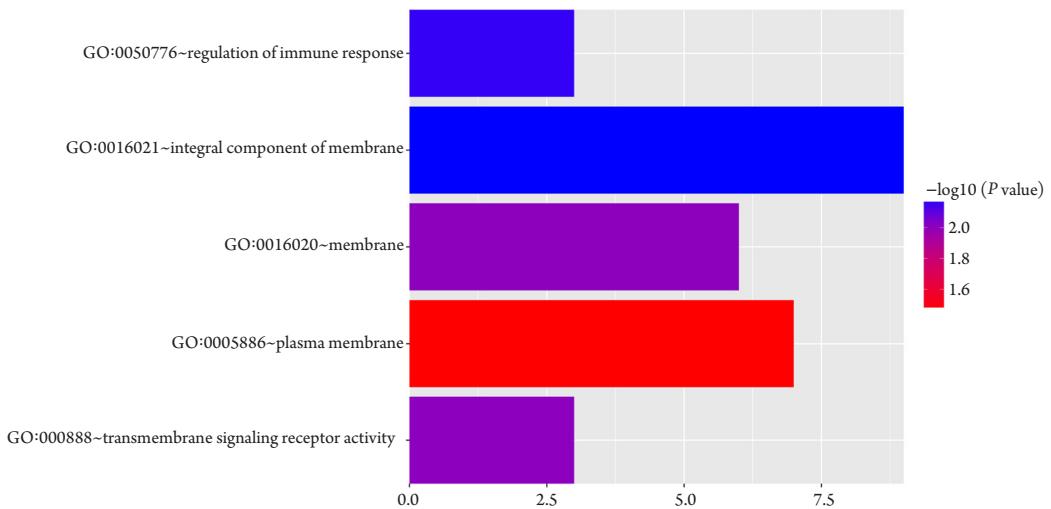
基因	log FC	AveExpr	t	P	adj.P	B
RPS11	1.726831	9.715312	5.533154	9.67E-07	0.006762	5.328382
LOC400499	1.550077	7.947541	5.285285	2.36E-06	0.006762	4.538805
KLRB1	-1.75237	10.3	-5.23384	2.84E-06	0.006762	4.375969
AK021987	-1.56845	6.325456	-5.01593	6.17E-06	0.009604	3.690981
KLRF1	-1.80848	9.778662	-4.87303	1.02E-05	0.012595	3.246469
BF509781	-1.56088	6.161799	-4.7487	1.57E-05	0.016637	2.86309
MMD	-1.56725	9.43519	-4.69137	1.92E-05	0.018895	2.687467
SMAD7	-1.52513	6.296098	-4.62393	2.42E-05	0.020875	2.481861
GZMB	-1.70175	9.710675	-4.61576	2.49E-05	0.020875	2.457038
AI683805	-1.51157	9.272419	-4.43246	4.66E-05	0.028911	1.904416
T90348	-1.56872	4.043076	-4.41779	4.90E-05	0.029228	1.860591
PRF1	-1.5375	9.184815	-4.40114	5.18E-05	0.029678	1.810908
ASGR2	-0.46077	7.993025	4.393029	5.32E-05	0.029798	1.786721
BF111108	-1.50313	4.427554	-4.3752	5.65E-05	0.029839	1.733637
IL4R	-0.49624	8.362564	4.218765	9.55E-05	0.039487	1.272023
P2RY12	-1.66385	6.67997	-4.18516	0.000107	0.040493	1.173847
AK024584	-1.90778	6.384099	-4.18014	0.000108	0.040493	1.159197
BG430958	-1.64299	5.006146	-4.17809	0.000109	0.040493	1.153244
R71414	-1.54283	4.143904	-4.14366	0.000122	0.041779	1.053107
AKR1C3	-1.6975	6.924522	-4.12817	0.000129	0.041779	1.008171
ENPP4	-1.51519	6.578008	-4.09909	0.000142	0.044948	0.924053
KLRD1	-1.63677	8.235487	-4.09016	0.000146	0.045279	0.898275

“adj.P”是调整后的P值；“P”是贝叶斯检验后得到的P值；“B”是贝叶斯检验后得到的标准差的对数化值；“log FC”是两组表达值间以2为底对数化的变化倍数。

“adj.P” is the adjusted P value; “P” is the P value obtained after Bayesian test; “B” is the logarithmic value of standard deviation obtained after Bayesian test; “log FC” is the multiple of the logarithmic change between two groups of expression values.

**表2 GO富集分析汇总****Table 2 Summary of GO enrichment analysis**

Term	Count	P	Genes	FDR
GO:0019835~cytolysis	3	9.757472686138433E-5	PRF1, MMD, GZMB	0.11
GO:0030246~carbohydrate binding	4	3.139088893666228E-4	KLRB1, ASGR2, KLRF1, KLRD1	0.31
GO:0007166~cell surface receptor signaling pathway	4	8.476492581859565E-4	KLRB1, ASGR2, KLRF1, KLRD1	0.99
GO:0016021~integral component of membrane	9	0.006477301722014285	KLRB1, P2RY12, PRF1, ASGR2, IL4R, ENPP4, MMD, KLRF1, KLRD1	5.67
GO:0050776~regulation of immune response	3	0.006877289545375131	KLRB1, KLRF1, KLRD1	7.81
GO:0004888~transmembrane signaling receptor activity	3	0.009709938071131144	KLRB1, KLRF1, KLRD1	9.19
GO:0016020~membrane	6	0.00971738274428522	PRF1, ENPP4, MMD, GZMB, RPS11, KLRF1	8.39
GO:0005886~plasma membrane	7	0.034519518764643445	KLRB1, P2RY12, PRF1, SMAD7, ENPP4, KLRF1, KLRD1	27.05
GO:0005887~integral component of plasma membrane	4	0.06052303093041587	P2RY12, IL4R, MMD, KLRF1	42.91
GO:0043235~receptor complex	2	0.08051740370530616	IL4R, KLRD1	52.94

**图3 可视化GO富集分析****Figure 3 Visualization GO enrichment analysis**

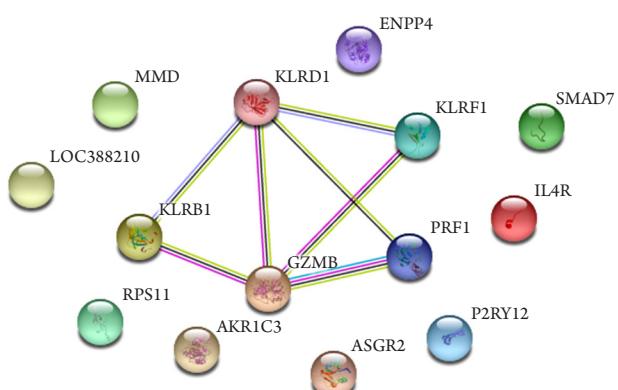


图4 蛋白相互作用网络

Figure 4 Protein interaction network

### 3 讨论

传统的AMI危险因素如吸烟、糖尿病等虽然对于疾病的预防和预后的判断具有一定的参考价值，但是这种危险因素的特异性存在一定局限性<sup>[13]</sup>。当前阶段识别并鉴定出新的生物标志物，有利于AMI的早期筛查和诊断。

本研究采用生物信息学方法，通过筛选GEO芯片数据，分析获得AMI患者外周血标本中的基因差异性表达，发现与健康对照组比较，AMI患者外周血样本中22个基因存在差异表达，其中20个表达下调以及2个表达上调。进一步对差异基因进行GO功能富集分析和蛋白相互作用网络构建，初步对其差异性表达进行了探索。

差异表达的基因中有2个表达量上调较为明显，分别为RPS11和LOC400499。说明这2个基因在AMI患者中属于危险因素。后者在既往研究<sup>[14]</sup>中被认为与心肌生长相关，这可能涉及间充质干细胞以及胚胎干细胞衍生心肌细胞的机制<sup>[15]</sup>。此外，LOC400499被发现参与了心肌损伤后的炎症以及修复过程<sup>[16]</sup>。但是RPS11与AMI的相关性目前尚未由文献报道，这可能成为一个潜在的生物标志物。

本研究进一步对所有差异基因进行富集分析，将功能上一致的基因进行合并后，发现主要集中以下功能：GO:0016021~integral component of membrane(膜的整体组分)；GO:0050776~regulation of immune response(免疫应答调节)；GO:0004888~transmembrane signaling receptor activity(跨膜信号受体活性)；GO:0016020~membrane(膜)以及GO:0005886~plasma membrane(质膜)。全部为AMI患者血液低表达的基因。这些基因的功能主要涉及细胞膜和免疫应答调节。在早期研究<sup>[17]</sup>中就已经发现内皮细胞膜上

的调节性蛋白对于血栓栓塞性疾病的发生具有重要作用，如血栓调节蛋白作为凝血酶介导的蛋白C活化的辅助因子，血栓调节蛋白辅因子功能受损也可能导致血栓栓塞性疾病的促血栓形成异常。AMI发病中也常伴有辅助性T细胞和炎性因子的失衡，同时内皮细胞膜的组成成分与这种免疫应答调节存在一定的相关性，如在AMI患者中能发现外周血内皮细胞膜微粒与IL-2与INF-γ呈正相关<sup>[18]</sup>。结合蛋白相互作用网络后，本研究挑选出了存在相互作用的基因：KLRD1，KLRF1，KLRB1，GZMB和PRF1。其中KLRB1基因与Yan等<sup>[19]</sup>的研究结果一致，该基因表达量的减少与AMI患者自然杀伤细胞数量减少和功能障碍有密切关系。Zhang等<sup>[20]</sup>和Chen等<sup>[21]</sup>的研究中均提示了GZMB和PRF1基因之间的表达存在一定的相关性，且二者均被认识与AMI的发病相关。近些年也认为这2个基因均参与泛素化和T细胞相关通路<sup>[22]</sup>。目前尚未有足够的研究证实KLRD1和KLRF1基因参与AMI的发生，但是从本研究结果可以发现，这2个基因事实上与其他几个基因之间一起参与了多项生物学功能。因此，本研究结果可为AMI生物标志物的寻找提供一个新的依据和线索。

实际上，AMI患者血液转录组的靶向基因表达研究<sup>[23-24]</sup>已经表明：在AMI发生之后，白细胞转录机制的调节和循环白细胞的炎性标志物的过度表达发生独立于患者危险因素或炎性标志物CRP的血浆水平，进一步表明心肌缺血导致外周血中转录基因表达的改变。鉴于这一现象，外周血微阵列基因表达谱与富集分析技术结合，似乎提供了一种可行且非侵入性的方法，用于鉴定在AMI发生时来自循环细胞的不同的生物学相关的基因表达谱。

但是本研究尚有部分局限性。其中最重要的是GSE48060不仅提供了AMI与健康对照人群的对比数据，而且还提供了复发与非复发的AMI的数据。所以对AMI患者有无复发的数据的亚组分析应该是下一步研究的方向，这可能为复发的遗传易感性研究提供一个方向。同时对已经发表的文献进行分析后也发现，本研究与既往张新春等<sup>[25]</sup>的研究结果并不完全一致，这种差异产生的主要原因是挑选了不同的原始芯片，如张新春等<sup>[25]</sup>的研究主要采用的是GSE60993，而本研究结果来自于GSE48060，这也提示了未来多芯片联合筛选差异基因可能是采用生信技术寻找差异基因的趋势。

综上，对循环细胞全基因组转录组分析鉴定可能为预测AMI的易感性提供一个线索。本研究初步证明这种方法可能代表一种新颖的非侵入性平

台, 用于预测潜在的治疗靶点, 以优化首次AMI患者的心血管预后。

## 参考文献

1. Shapiro MD, Fazio S. From lipids to inflammation: new approaches to reducing atherosclerotic risk[J]. Circ Res, 2016, 118(4): 732-749.
2. Go AS, Mozaffarian D, Roger VL, et al. Executive summary: heart disease and stroke statistics--2013 update: a report from the American Heart Association[J]. Circulation, 2013, 127(1): 143-152.
3. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2013 update: a report from the American Heart Association[J]. Circulation, 2013, 127(1): e6-e245.
4. Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making[J]. JAMA, 2000, 284(7): 835-842.
5. Law MR, Watt HC, Wald NJ. The underlying risk of death after myocardial infarction in the absence of treatment[J]. Arch Intern Med, 2002, 162(21): 2405-2410.
6. Ge Y, Wang TJ. Identifying novel biomarkers for cardiovascular disease risk prediction[J]. J Intern Med, 2012, 272(5): 430-439.
7. Björkegren JLM, Kovacic JC, Dudley JT, et al. Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders[J]. J Am Coll Cardiol, 2015, 65(8): 830-845.
8. Lettre G. Rare and low-frequency variants in human common diseases and other complex traits[J]. J Med Genet, 2014, 51(11): 705-714.
9. Gerszten RE, Wang TJ. The search for new cardiovascular biomarkers[J]. Nature, 2008, 451(7181): 949-952.
10. Pedrotty DM, Morley MP, Cappola TP. Transcriptomic biomarkers of cardiovascular disease[J]. Prog Cardiovasc Dis, 2012, 55(1): 64-69.
11. Beisvag V, Lehre PK, Midelfart H, et al. Aetiology-specific patterns in end-stage heart failure patients identified by functional annotation and classification of microarray data[J]. Eur J Heart Fail, 2006, 8(4): 381-389.
12. Kitchens MM, Ye SQ, Irizarry RA, et al. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy[J]. Circulation, 2004, 110(22): 3444-3451.
13. Burke GM, Genuardi M, Shappell H, et al. Temporal associations between smoking and cardiovascular disease, 1971 to 2006 (from the Framingham Heart Study)[J]. Am J Cardiol, 2017, 120(10): 1787-1791.
14. Tiburcy M, Zimmermann WH. Modeling myocardial growth and hypertrophy in engineered heart muscle[J]. Trends Cardiovasc Med, 2014, 24(1): 7-13.
15. Carvalho E, Verma P, Hourigan K, et al. Myocardial infarction: stem cell transplantation for cardiac regeneration[J]. Regen Med, 2015, 10(8): 1025-1043.
16. Liehn EA, Postea O, Curaj A, et al. Repair after myocardial infarction, between fantasy and reality: the role of chemokines[J]. J Am Coll Cardiol, 2011, 58(23): 2357-2362.
17. Ohlin AK, Morser J, Ohlin H. Soluble thrombomodulin antigen in plasma is increased in patients with acute myocardial infarction treated with thrombolytic therapy[J]. Thromb Res, 1996, 82(4): 313-322.
18. 张文, 陈文强, 朱斌, 等. 冠心病患者外周血内皮细胞膜微粒、辅助性T细胞细胞因子水平变化及临床意义[J]. 国际检验医学杂志, 2017, 38(2): 245-247.  
ZHANG Wen, CHEN Wenqiang, ZHU Bin, et al. Changes and clinical significance of membrane microparticles and helper T cell cytokines in peripheral blood endothelial cells of patients with coronary heart disease[J]. International Journal of Laboratory Medicine, 2017, 38(2): 245-247.
19. Yan W, Zhou L, Wen S, et al. Differential loss of natural killer cell activity in patients with acute myocardial infarction and stable angina pectoris[J]. Int J Clin Exp Pathol, 2015, 8(11): 14667-14675.
20. Zhang S, Liu W, Liu X, et al. Biomarkers identification for acute myocardial infarction detection via weighted gene co-expression network analysis[J]. Medicine (Baltimore), 2017, 96(47): e8375.
21. Chen J, Yu L, Zhang S, et al. Network analysis-based approach for exploring the potential diagnostic biomarkers of acute myocardial infarction[J]. Front Physiol, 2016, 7: 615.
22. Joehanes R, Ying S, Huan T, et al. Gene expression signatures of coronary heart disease[J]. Arterioscler Thromb Vasc Biol, 2013, 33(6): 1418-1426.
23. Calvano SE, Xiao W, Richards DR, et al. A network-based analysis of systemic inflammation in humans[J]. Nature, 2005, 437(7061): 1032-1037.
24. Wettinger SB, Doggen CJ, Spek CA, et al. High throughput mRNA profiling highlights associations between myocardial infarction and aberrant expression of inflammatory molecules in blood cells[J]. Blood, 2005, 105(5): 2000-2006.
25. 张新春, 郑景辉, 莫云秋, 等. 基于GEO数据库芯片的早期ST段抬高性心肌梗死标志物的筛选及生物信息学分析[J]. 当代医学, 2018, 24(12): 1-6.  
ZHANG Xinchun, ZHENG Jinghui, MO Yunqiu, et al. GEO database-based bioinformatics study on gene chip of markers of early ST segment elevation myocardial infarction[J]. Contemporary Medicine, 2018, 24(12): 1-6.

**本文引用:** 魏文, 汪逵. 基于GEO数据库芯片的心肌梗死标志物的筛选与生物信息学分析[J]. 临床与病理杂志, 2019, 39(1): 27-32.  
doi: 10.3978/j.issn.2095-6959.2019.01.005

**Cite this article as:** WEI Wen, WANG Kui. Screening and bioinformatics analysis of myocardial infarction markers Based on GEO database chip[J]. Journal of Clinical and Pathological Research, 2019, 39(1): 27-32.  
doi: 10.3978/j.issn.2095-6959.2019.01.005