

缺血性心肌病不同阶段基因表达的改变模式研究

刘志华^{1,2} 马晨光²

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(北京易咨众科科技有限公司 北京 100084)

摘 要 文章选取缺血性心肌病患者和对照组收集的外周血, 对其采用微阵列数据集进行转录组分析, 分为两个部分进行。在第一部分, 从 3 组缺血性心肌病患者样本与健康对照的数据集分析中, 确定了 3 个重要基因——成纤维细胞生长因子结合蛋白 2 (FGFBP2)、葡萄糖-果糖氧化还原酶结构域 (GFOD1) 和伴皮层下囊肿的巨脑性白质脑病 (MLC1), 可作为潜在的 mRNA 生物标志物。在第二部分, 从 3 个时间点 (发作当天、恢复 4~6 天及恢复 6 个月) 收集了 2 组基因表达数据集。基因代谢途径的差异性分析表明, 与对照组对比, 发作组和恢复组涉及到炎症和免疫途径; 与恢复 6 个月组对比, 发作组和恢复组 (4~6 天) 涉及到代谢途径或神经分泌。这显示了缺血性心肌病的发作和恢复期间的生物学过程变化。实验结果表明, 3 种潜在的 mRNA 生物标志物——FGFBP2、GFOD1 和 MLC1, 涉及到缺血性心肌病不同的代谢通路, 具有连续的生物学过程变化。

关键词 缺血性心肌病; 生物信息学; 生物标志物; 微阵列; 转录组

中图分类号 R 394.8 **文献标志码** A

Altered Gene Expression Through Different Stages of Ischemic Cardiomyopathy

LIU Zhihua¹ MA Chenguang²

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Beijing EasyTech Co. Ltd, Beijing 100084, China)

Abstract In this study, transcriptome analysis was conducted on the microarray data sets of peripheral blood collected from patients with ischemic cardiomyopathy and controls. The analysis was carried out in two phases. In phase 1, by comparing three sets of ischemic cardiomyopathy samples versus healthy controls, we identified three key genes—fibroblast growth factor binding protein 2 (FGFBP2), glucose-fructose oxidoreductase domain containing 1 (GFOD1), and megalencephalic leukoencephalopathy with subcortical cysts 1 (MLC1). These were considered as the potential mRNA biomarkers candidates. In phase 2, two gene expression data sets were collected on three points in time (the day of attack, 4-6 days of recovery, and 6 months of recovery). Differential analysis of gene pathways revealed that the seizure and recovery group were involved in the inflammatory and immune pathways compared with the control group. The seizure and recovery group (4-6 days) were involved

收稿日期: 2017-08-01 修回日期: 2017-08-14

基金项目: 深圳市基础研究项目 (JCYJ20150401145529007); 北京市科技新星计划资助 (Z141102001814075); 深圳市技术开发项目 (CXZZ20151015163619907)

作者简介: 刘志华 (通讯作者), 博士, 教授, 博士生导师, 研究方向为生物信息学与健康医疗大数据, E-mail: zh.liu@siat.ac.cn; 马晨光, 研究员, 研究方向为大数据。

in the metabolic pathways or nerve secretion compared with the 6 months of recovery group. The experimental results show that the three potential mRNA biomarkers (FGFBP2, GFOD1 and MLC1) can be involved different pathways of ischemic cardiomyopathy, exhibiting continuous changes in the biological process.

Keywords ischemic cardiomyopathy; bioinformatics; biomarker; microarray; transcriptome

1 引 言

目前的研究已表明^[1], 心血管疾病的死亡率远远高于癌症和艾滋病, 成为危害人类健康的第一杀手。虽然近几年在血管重建技术、药物治疗、心脏康复和器官移植等医疗技术方面获得了重大进展, 但心血管疾病, 包括缺血性心肌病仍然是发达国家人口死亡的首要原因。虽然吸烟、高血清胆固醇、高血压和糖尿病等经典的疾病危险因素, 可在一定程度上预测疾病发生和预后情况, 但还不足以提供完全准确的疾病诊断^[2]。这是因为当前这些风险预测模型的预测价值的准确率相当有限。而这种“检测差距”是因为缺血性心肌病患者常常缺乏常规危险因素, 高达 20% 的患者没有传统的危险因素, 40% 的患者往往只有一个^[3,4]。目前疾病风险的预测更加重视非传统和遗传学风险因素的影响。因此, 找到缺血性心肌病相关的预测风险因子, 降低死亡率和改善预后是现阶段公共卫生的重点。

目前, 分子标记物如脑钠肽(Brain Natriuretic Peptide, BNP)、C 反应蛋白(C Reactive Protein, CRP)和其他血清炎症指标已经在缺血性心肌病预测方面得到证实。然而, 这些生物标志物在甲状腺疾病、肾衰竭和充血性心力衰竭患者中也具有较高的表达水平^[5]。因此, 亟需寻找新的更敏感的风险标记及改善个人预防的疾病筛选策略。这样的生物标志物必须具备: 识别危险个体的能力、能检测结果的重复稳定性、早期干预的良好治疗效果以及能基于细胞的靶点治疗来控制缺血性心肌病的病理过程等特点。而使用微

阵列的全基因组表达分析是检测新的生物标志物和鉴定新药物靶点的方法^[6,7]。因此, 本文主要围绕以下两个方面展开研究: (1) 研究缺血性心肌病患者外周血细胞模型中的基因表达模式, 找出病理发生的关键基因; (2) 识别与缺血性心肌病相关的生物学过程的动态变化和发展。

2 材料与amp;方法

2.1 实验材料

本实验材料分为两个部分, 在第一部分, 我们选择了 3 个微阵列数据集(GSE66360、GSE48060 和 GSE60993), 包含缺血性心肌病患者和具有正常超声心动图的健康对照。其中, 数据集 GSE60993 使用 Illumina HumanWG-6 v3.0 表达珠芯片, 从含有 17 例患者和 7 个对照样品的外周血中收集。另外 2 个数据集 GSE48060 和 GSE66360 均使用 Affymetrix 人类基因组 U133 Plus 2.0 阵列平台, 分别包括 49 例患者与 50 例来自循环内皮细胞的对照、31 例患者与 21 例外周血对照组比较。

在第二部分, 为了找到与缺血性心肌病发作和病情恢复的基因表达变化情况, 我们选择了 2 个微阵列表达数据集(GSE59867 和 GSE62646)做进一步分析。其中, 数据集 GSE59867 收集 ST 段抬高型心肌梗死患者样品($n=111$)。为了排除相关基因的影响, 对照组为病情稳定且无疾病史的患者($n=46$), 而不是健康对照组。数据集 GSE62646 由 28 名患者和 14 名对照组患者组成, 具有相同的纳入标准。在 3 个时间点: 急性

发作的第 1 天(简称“发作组”)、治疗恢复 4~6 天(简称“恢复组”)、治疗恢复 6 个月后(简称“6 个月组”)收集血液样品。采用 Affymetrix HuGene 1.0 ST 阵列分析研究组和对照组分离的外周血液单核细胞中的 mRNA 水平。

2.2 分析方法

2.2.1 加权基因共表达网络分析

加权基因共表达网络分析(Weighted Gene co-Expression Network Analysis, WGCNA)是一种共表达网络的分析方法^[8]。它在 R WGCNA 软件包中实现,是基于基因表达数据来构建基因共表达的网络。本文使用 WGCNA 方法对数据集的差异表达基因进行独立分析。首先,对于每组基因,通过计算相关矩阵构建加权相关网络。其次,通过将相关矩阵提高到 8、10 或 12 的适当幂来构建邻接矩阵。之后基于相邻矩阵,计算每对基因的拓扑重叠度。最后,对拓扑重叠度进行了平均层次聚类,并使用动态混合树切割算法来剪切聚类树。为了获得大而不同的模块,我们将最小模块的大小设置为 30 个基因,并将模块的最小高度设置为 0.2。最终对每个基因模块和疾病状态的 Pearson 相关系数值、模块显著性值和 Pearson 相关系数检验值进行评估,为进一步分析提供补充评估。

2.2.2 基因功能富集分析

如果基因具有多个探针位点,则根据该探针位点的平均值来过滤数据。预处理后,我们对这些数据集进行患者组和对照组的差异分析,差异表达基因的统计学显著性阈值为 0.05。

对于模块的每个基因,使用超几何分布测试来评估功能富集分析。计算公式如下^[9]:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-1}}{\binom{N}{n}}$$

其中, N 是所有具有注释的基因数量; n 是 N 中

模块基因的数量; M 是注释到特定条目的所有基因的数量; m 是 M 中模块基因的数量。计算的 P 值通过 Bonferroni 校正,将校正的 P 值 0.05 作为阈值。

3 实验结果

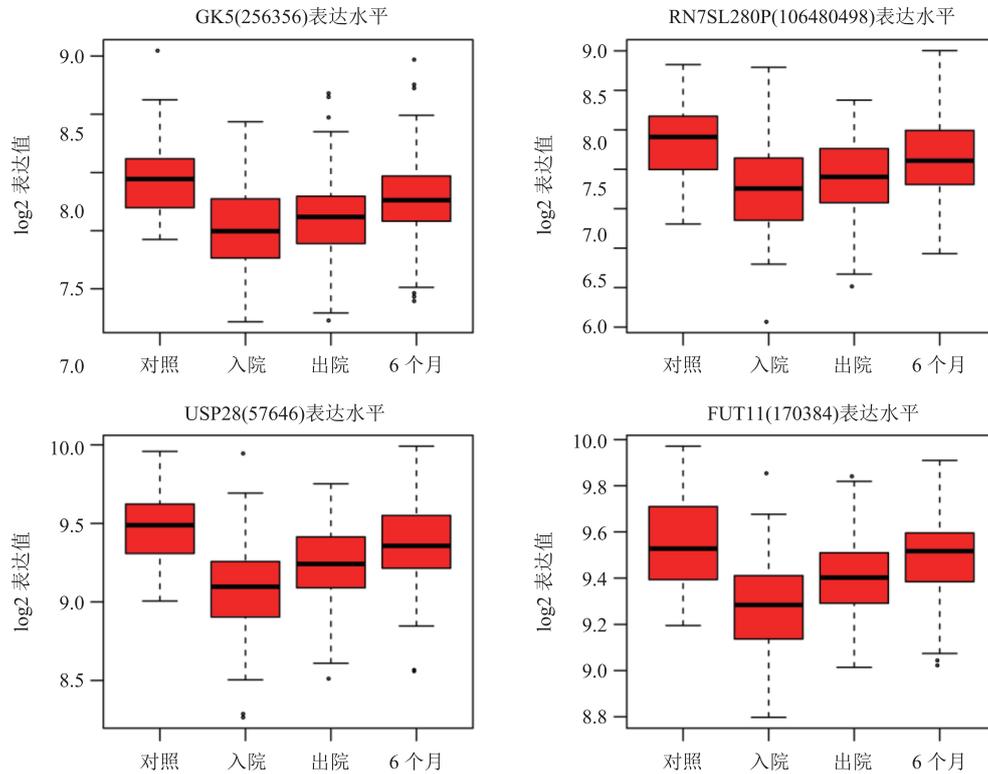
3.1 疾病和健康对照组基因微阵列差异分析

分析结果显示,3 组病例中差异表达基因之间存在非常显著的重叠,支持数据的鲁棒性,并指示这 3 种情况下缺血性心肌病特异性表达变化是一致的。我们在 GSE66360、GSE48060 和 GSE60993 中分别鉴定出 4672、3185、3660 个基因,显示患者和对照组之间的显著表达变化。

接下来,应用 WGCNA 对基因表达差异进行分析,并计算与疾病状态相关的模块值,提供模块的定量评估。我们对每个模块参与的基因进行功能注释和途径富集分析发现,选定的模块中差异表达基因之间仅有 3 个基因重叠:成纤维细胞生长因子结合蛋白 2(FGFBP2)、葡萄糖-果糖氧化还原酶结构域(GFOD1)和伴皮层下囊肿的巨脑性白质脑病(MLC1),涉及到血管重建、糖代谢、伴皮层下囊肿的巨脑性脑白质病。

3.2 疾病的发作、恢复及后期的基因表达分析

为了研究从疾病发生到后期恢复不同时间点的基因表达和涉及到的通路的变化,对发作和恢复这两个时间点与恢复后 6 个月的时间点及对照组分别作了基因表达差异分析(图 1),分别为发作组对比 6 个月组、恢复组对比 6 个月组、发作组对比对照组、恢复组对比对照组。此外,应用 WGCNA 来确定共表达基因的模块。我们对 4 组数据选中的模块做重叠基因分析,分别获得 427、179、172、157 个重叠基因。对这 4 组重叠基因做通路富集分析后发现,与对照组相比,发作组和恢复组涉及相似的代谢通路,如自然杀伤细胞介导的细胞毒性、移植物抗宿主病、抗



x 轴表示不同的分析组；y 轴表示 log₂ 归一化的基因表达水平；入院对应的是发作时间；出院对应的是恢复 4~6 天；6 个月对应的是恢复 6 个月后

图 1 不同时间点所选基因表达曲线图

Fig. 1 Expression data from microarray experiments for chosen genes

原加工和呈递反应；与恢复后 6 个月组相比，发作组和恢复组涉及相似的代谢通路，如氧化磷酸化、亨廷顿氏病和帕金森病等。

此外，我们对上述 4 组重叠基因进行逐个病例分析，并在所有患者中鉴定了 19 个常见基因，前 10 个重要基因被选中做进一步分析。虽然在患者之间基因相对表达水平不同，但其显著变化趋势是相同的。

4 讨论

4.1 患者与健康对照组基因表达差异分析

在我们挑选的数据集 GSE66360、GSE48060 和 GSE60993 中，GSE66360 并没有相关文献发表，GSE48060 主要针对疾病组和对照组的差异

表达基因做了相关通路分析，发现与机体的免疫应答相关，并寻找与临床复发相关的关键基因。而 GSE60993 也是从疾病组和对照组寻找出差异基因，并在相关组织中进行了酶联免疫吸附测定 (Enzyme Linked Immunosorbent Assay, ELISA) 蛋白验证。

我们通过使用 WGCNA 方法降低非重要基因来减少数据集的复杂性，并筛选最相关的模块进行分析。生物学过程主要涉及炎症和免疫应答途径，其中包括自然杀伤细胞介导的细胞毒性、移植植物抗宿主病、同种异体移植排斥、抗原加工和表达以及细胞因子-细胞因子受体相互作用与缺血性心脏损伤后病理变化的相关机制。这也表明免疫应答、炎症和凋亡途径相关的基因表达水平升高，可能和心肌病理过程与循环细胞炎症转录

变化相关。实际上, 通过直接多方面测量白细胞的炎症 mRNA 谱的研究显示, 涉及白细胞的炎症反应与缺血性心肌病的发病机理密切相关^[10]。另一方面, 在白细胞转录的调节机制研究中, 分析血液白细胞基因表达模式的变化进一步证明了心肌缺血可以导致外周血相关转录基因的表达变化^[11,12]。

值得一提的是, 这些差异基因的功能分析揭示其参与了 I 型糖尿病和自身免疫性甲状腺疾病的代谢通路。与目前国内外已有的研究^[13-15]相比, 有了进一步的进展。另有相关研究证明, 亚临床型甲状腺功能减退症(伴随着心脏收缩功能减退、动脉粥样硬化脂质分布、高血压和炎症反应)和甲状腺机能亢进(伴随着冠状动脉存储功能下降、心律不齐、血液高凝固风险)等临床症状对心血管系统具有负面影响^[16,17]。因此, 甲状腺激素失调通常导致心脏功能和基因表达变化, 即使生理范围内的甲状腺激素的小变化也可能导致心血管疾病的不良发展。

在关键基因的获取上, 我们挑选了 GSE48060 数据集、GSE66360 数据集及 GSE60993 数据集中相关系数最高的 3 个模块取交集, 得到 3 个关键基因——FGFBP2、GFOD1 和 MLC1, 可作为缺血性心肌病早期诊断的潜在生物标志物。这比从 3 个数据集中分别单独挑选的关键基因更具可靠性。这一发现意味着缺血性心肌病的致病机制与涉及血管内病变、免疫应答和脑源性因子调节系统的基因表达相关。其中, FGFBP2 编码成纤维细胞生长因子结合蛋白家族的成员, 通过调节蛋白激酶调控网络使心血管免受缺血损伤。FGFBP2 是对葡萄糖代谢、脂质代谢和胰岛素敏感性具有多效性的代谢激素^[18]。GFOD1 在抗细胞衰老中发挥了重要的作用, 是细胞能量的主要来源和细胞能量代谢的基础, 在衰老进程中处于中心地位^[19]。MLC1 编码膜蛋白, 在脑星形胶质细胞和循环血细胞, 特别是单

核细胞中高度表达。虽然已有研究提出 MLC1 具有血脑屏障和脑-脑脊液屏障的作用^[20], 但其功能仍然未知。Lanciotti 等^[21]观察到 MLC1 调节星形胶质细胞中的表皮生长因子受体信号, 提供了关于 MLC1 发病机理的新信息。这些研究表明, 血液转录组大规模基因表达分析可用于缺血性心肌病的诊断和鉴定疾病的相关机制。

4.2 患者不同时间点的基因转录组差异分析

许多研究发现, 外周血单核细胞在与血管重塑和局部炎症反应中发挥着重要作用^[22,23]。因此, 外周血单核细胞的激活表达反映炎症程度, 可能与缺血性心肌病患者的病情进展相关。急性冠状动脉综合征(ACS)患者的预后很大程度上取决于急性期心肌损伤的程度。在这部分, 我们选择了两个外周血单核细胞基因表达谱(GSE59867 和 GSE62646), 用于鉴定缺血性心肌病随不同时间变化的生物学相关过程的变化情况。其中 GSE59867 和 GSE62646 数据集本身也进行了 4 组差异基因比较: 发作组对比对照组、恢复组对比对照组、发作组对比恢复 6 个月组、恢复组对比恢复 6 个月组。GSE59867 更侧重于发展为或没有发展为心力衰竭的基因表达差别。而 GSE62646 更侧重于这 4 组比较差异基因涉及的代谢通路变化。而我们对这两组数据集进行组合差异分析, 并进行 WGCNA 模块筛选, 对这 4 组对照都涉及的差异基因挑选前 10 个关键基因进行逐个病例分析。

结果显示, 与对照组相比, 发作组和恢复组涉及的生物学过程大致相同, 与机体的炎症和免疫反应相关, 与第一部分的结果一致。与恢复 6 个月组相比, 发作组和恢复组涉及的生物学过程大致相同, 与机体代谢途径和神经系统相关。而这种在不同时间涉及机体相似的代谢通路和神经系统通路, 在目前国内外已有的研究分析中并没有发现这种规律。一种可能是与患者病情发作后药物干预和心肌功能恢复相关, 另一种可能是由

于冠状动脉的部分或完全闭塞,神经分泌涉及对心功能不全发展的保护性反应。已有研究证实,酪氨酸激酶受体 B (TrkB) 是脑源性神经营养因子的高亲和力受体,可在动脉粥样硬化形成过程中保护内皮完整性,并在缺血性心肌病发展中起到未知的保护作用^[24]。此外,自主神经系统的长期激活可能增加动脉粥样硬化、心肌梗死或猝死的风险^[25,26]。

5 结 论

精准诊断缺血性心肌病的病理特征的最大障碍是心脏组织样本的限制性^[27,28]。然而转录组技术为发现疾病特异性机制提供了新的机会,并提供诊断和预后价值。在本研究中,我们筛选出 3 个重要的基因——FGFBP2、GFOD1 和 MLC1,并且发现缺血性心肌病在 3 个不同时间段(发作、恢复 4~6 天和恢复后 6 个月)内涉及到机体的不同代谢通路。这为确定潜在的心肌缺血相关生物标志物,并揭示与缺血性心肌病发展相关的调节网络和代谢途径提供了一定的理论依据。

参 考 文 献

- [1] Members WG, Mozaffarian D, Benjamin EJ, et al. Heart disease and stroke statistics-2016 update: a report from the American Heart Association [J]. *Circulation*, 2016, 133(4): 38-60.
- [2] Law MR, Watt HC, Wald NJ. The underlying risk of death after myocardial infarction in the absence of treatment [J]. *Archives of Internal Medicine*, 2002, 162(21): 2405-2410.
- [3] Ajani UA, Ford ES. Has the risk for coronary heart disease changed among U.S. adults? [J]. *Journal of the American College of Cardiology*, 2006, 48(6): 1177-1182.
- [4] Hozawa A, Folsom AR, Sharrett AR, et al. Absolute and attributable risks of cardiovascular disease incidence in relation to optimal and borderline risk factors: comparison of African American with white subjects-atherosclerosis risk in communities study [J]. *Archives of Internal Medicine*, 2007, 167(6): 573-579.
- [5] Ge Y, Wang TJ. Identifying novel biomarkers for cardiovascular disease risk prediction [J]. *Journal of Internal Medicine*, 2012, 272(5): 430-439.
- [6] Hall PA, Reis-Filho JS, Tomlinson IP, et al. An introduction to genes, genomes and disease [J]. *Journal of Pathology*, 2010, 220(2): 109-113.
- [7] Heidecker B, Hare JM. The use of transcriptomic biomarkers for personalized medicine [J]. *Heart Failure Reviews*, 2007, 12(1): 1-11.
- [8] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [J]. *BMC Bioinformatics*, 2008, 9(1): 559.
- [9] Wen F, Batliwalla F, Gregersen PK, et al. Overlapping probabilities of top ranking gene lists, hypergeometric distribution, and stringency of gene selection criterion [C] // International Conference of the IEEE Engineering in Medicine & Biology Society, 2006: 5531-5534.
- [10] Suresh R, Li X, Chiriac A, et al. Transcriptome from circulating cells suggests dysregulated pathways associated with long-term recurrent events following first-time myocardial infarction [J]. *Journal of Molecular and Cellular Cardiology*, 2014, 74(3): 13-21.
- [11] Calvano SE, Xiao W, Richards DR, et al. Corrigendum: a network-based analysis of systemic inflammation in humans [J]. *Nature*, 2005, 437(7061): 1032-1037.
- [12] 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.
- [13] Maciejak A, Kiliszek M, Michalak M, et al. Gene expression profiling reveals potential prognostic biomarkers associated with the progression of heart failure [J]. *Genome Medicine*, 2015, 7(1): 26.
- [14] Kiliszek M, Burzynska B, Michalak M, et al. Altered gene expression pattern in peripheral blood

- mononuclear cells in patients with acute myocardial infarction [J]. *PloS One*, 2012, 7(11): e50054.
- [15] 侯振江, 侯建章, 王凤玲. 甲状腺疾病与心血管疾病研究进展 [J]. *医学综述*, 2015, 21(23): 4296-4298.
- [16] Faiza AQ. Thyroid hormone profile in patients with acute coronary syndrome [J]. *Iranian Red Crescent Medical Journal*, 2015, 17(7): e26919.
- [17] Cokkinos DV, Chrysanthopoulos S. Thyroid hormones and cardiac remodeling [J]. *Heart Failure Reviews*, 2016, 21(4): 365-372.
- [18] Lin Z, Pan X, Wu F, et al. Fibroblast growth factor 21 prevents atherosclerosis by suppression of hepatic sterol regulatory element-binding protein-2 and induction of adiponectin in mice [J]. *Circulation*, 2015, 131(21): 1861-1871.
- [19] Manini TM. Energy expenditure and aging [J]. *Ageing Research Reviews*, 2010, 9(1): 1-11.
- [20] Boor PK, De Groot K, Waisfisz Q, et al. MLC1: a novel protein in distal astroglial processes [J]. *Journal of Neuropathology & Experimental Neurology*, 2005, 64(5): 412-419.
- [21] Lanciotti A, Brignone MS, Visentin S, et al. Megalencephalic leukoencephalopathy with subcortical cysts protein-1 regulates epidermal growth factor receptor signaling in astrocytes [J]. *Human Molecular Genetics*, 2016, 25(8): 1543-1558.
- [22] Bauters A, Ennezat PV, Tricot O, et al. Relation of admission white blood cell count to left ventricular remodeling after anterior wall acute myocardial infarction [J]. *American Journal of Cardiology*, 2007, 100(2): 182-184.
- [23] Aoki S, Nakagomi A, Asai K, et al. Elevated peripheral blood mononuclear cell count is an independent predictor of left ventricular remodeling in patients with acute myocardial infarction [J]. *Journal of Cardiology*, 2011, 57(2): 202-207.
- [24] Jiang H, Huang S, Li X, et al. Tyrosine kinase receptor B protects against coronary artery disease and promotes adult vasculature integrity by regulating Ets1-mediated VE-cadherin expression [J]. *Arteriosclerosis Thrombosis & Vascular Biology*, 2015, 35(3): 580-588.
- [25] Ozturk S, Yalvac HD, Sivri N, et al. Anxiety and depression scores in patients with coronary artery disease and coronary artery ectasia [J]. *International Journal Cardiology*, 2015, 186: 299-301.
- [26] Hausberg M, Hillebrand U, Kisters K. Addressing sympathetic overactivity in major depressive disorder [J]. *Journal of Hypertension*, 2007, 25(10): 2004-2005.
- [27] Devaux Y, Bousquenaud M, Rodius S, et al. Transforming growth factor β receptor 1 is a new candidate prognostic biomarker after acute myocardial infarction [J]. *BMC Medical Genomics*, 2011, 4(1): 83.
- [28] Kim J, Ghasemzadeh N, Eapen DJ, et al. Gene expression profiles associated with acute myocardial infarction and risk of cardiovascular death [J]. *Genome Medicine*, 2014, 6(5): 40.