

## 颈动脉斑块与斑块旁组织基因差异表达和信号通路分析

夏红利<sup>1</sup>, 刘国利<sup>1</sup>, 吴岸森<sup>1</sup>, 赵蓉<sup>1</sup>, 李敏才<sup>1,2</sup>

(1. 咸宁市中心医院血管外科 湖北科技学院附属第一医院, 湖北省咸宁市 437100;

2. 湖北科技学院基础医学院, 湖北省咸宁市 437100)

[关键词] 动脉硬化斑块; 基因表达数据库; 基因芯片; 核心基因; 信号传导通路

[摘要] **目的** 基于颈动脉粥样硬化斑块中基因芯片数据分析差异性基因表达、信号传导通路及蛋白质网络分析。**方法** 从基因表达数据库(GEO)下载 GSE43292 基因芯片数据,数据预处理后,使用在线 R 软件分析颈动脉斑块和斑块旁组织中差异表达的基因,运用生物医学信息学 GSEA 方法进行富集 GO 分析和 Pathway 分析,并对差异性基因行蛋白质网络分析。**结果** 芯片数据分析显示,颈动脉斑块和斑块旁组织中表达水平显著上调的有 87 个基因,表达水平显著下调的有 60 个基因( $P$  均 $<0.01$ ),并绘制成热图。GSEA 的 GO 功能富集分析发现,差异表达多与抗原结合、丝氨酸水解酶活性、趋化因子受体结合相关。GSEA 的 Pathway 富集分析显示,差异表达上调基因与造血干细胞系、溶酶体、细胞因子受体相互作用等通路中富集。STRING 分析筛选出 IL-8、CXCL-10、SELE、MMP-9、IL-18 共 5 个核心基因。**结论** 通过生物信息学方法对 GEO 基因芯片数据分析,发现了动脉粥样硬化斑块中差异性基因的相关信号传导通路及蛋白质的核心基因,为动脉粥样硬化研究提供基础资料。

[中图分类号] R6

[文献标识码] A

### The differential gene expression and signaling pathway analysis between the atherosclerotic plaque and the distant tissue in carotid artery

XIA Hongli<sup>1</sup>, LIU Guoli<sup>1</sup>, WU Anseng<sup>1</sup>, ZHAO Rong<sup>1</sup>, LI Mincai<sup>1,2</sup>

(1. Department of Vascular Surgery, Xianning Central Hospital & the First Affiliated Hospital of Hubei University of Science and Technology, Xianning, Hubei 437100, China; 2. School of Basic Sciences, Hubei University of Science and Technology, Xianning, Hubei 437100, China)

[KEY WORDS] atherosclerosis; gene expression omnibus database; gene chip data; core genes; signaling pathways

[ABSTRACT] **Aim** The differential gene expression, signaling pathways and protein network were analyzed from the gene chip data for carotid atherosclerotic plaques. **Methods** GSE43292 was downloaded from the gene expression omnibus (GEO) database and the data was adjusted. The significant differential genes were analyzed by R software in GSE43292. The gene ontology (GO) analysis and pathway analysis were performed by bioinformatics GSEA, and the significant differential gene was analyzed by Cytoscape. **Results** After GSE43292 was analyzed, the 87 up-regulated genes and the 60 down-regulated genes were found as the significant different expression gene ( $P<0.01$ ) between the atherosclerotic plaques and the distant tissues in carotid artery. The heatmap showed the different expression genes. The enriched GO analysis of GSEA revealed that the main differential expressions were associated with the antigen binding, the serine hydrolase activity and the chemokine receptor binding. The pathway analysis of GSEA revealed that the main differential expression were associated with hematopoietic stem cells, lysosomal and cytokine receptor interactions. Cytoscape analysis performed the five core genes, including IL-8, CXCL-10, SELE, MMP-9 and IL-18. **Conclusions** The bioinformatics analysis of GSE43292 data revealed that the differential genes enriched the related signaling pathways and the core genes were found between the atherosclerotic plaque and the distant tissue in carotid atherosclerotic plaques. These research provided the basic data for atherosclerosis research and treatment.

[收稿日期] 2018-11-12

[修回日期] 2019-01-05

[基金项目] 湖北省自然科学基金(2018CFB737);湖北科技学院科学基金(2018-19XZY02,2019-20XZ01)

[作者简介] 夏红利,博士,副主任医师,研究方向为动脉闭塞症及周围动脉病变,E-mail 为 mobei\_hu@126.com。通信作者 李敏才,博士,教授,研究方向为动脉硬化及成体干细胞,E-mail 为 mincai@163.com。

在当今社会中动脉粥样硬化(atherosclerosis, As)是众多心血管疾病共同的病理基础,As病理特征为不稳定斑块形成、动脉管壁硬化、管腔狭窄,其狭窄程度与急性心血管事件的发生相关<sup>[1]</sup>。As普遍被认为是慢性进展性炎症性疾病<sup>[2]</sup>,与炎症反应、细胞大量凋亡、新生血管形成等相关<sup>[3-4]</sup>,诱导这些病理生理过程的基因表达、分子机制、蛋白质网路相互作用并不完全清楚。研究As基因表达谱、探讨基因之间的调节关系对揭示As发病机制、防治心脑血管疾病发生发展有重要意义。

生物学研究的新领域——大数据基因表达芯片技术的出现,通过比较基因表达谱变化,筛选出与疾病相关的基因<sup>[5]</sup>,还可以对差异性基因参与的生物学过程、通路分析研究<sup>[6]</sup>。为研究比较As斑块与周围组织的基因表达谱的差异性,以GSE43292芯片数据为样本<sup>[7]</sup>,运用基因富集分析(gene set enrichment analysis, GSEA)软件,采用生物信息学方法对基因本体(gene ontology, GO)、京都基因和基因组百科全书(kyoto encyclopedia of gene and genomes, KEGG)通路(pathway)和蛋白质相互网络(String)分析可能存在的分子机制,探索相关生物学通路和风险预测,为As治疗靶点奠定理论基础。

## 1 材料和方法

### 1.1 样本数据来源

在PUBMED的GEO数据库中输入“atherosclerosis”,选择GSE43292的芯片数据作为研究对象。该芯片数据样本来自32例颈动脉As患者,每例患者有As斑块中心部位组织、As斑块旁组织,均为经颈动脉内膜剥脱术的病变血管。

### 1.2 数据分析

下载GSE43292的CEL文件,对探针水平的信号值进行数据归一化和基因注释处理,最终转化为基因表达水平的中位值<sup>[6,8]</sup>。

### 1.3 差异性基因分析

对32例颈动脉患者As斑块组织、斑块旁组织基因表达值,通过在线工具GEO2R软件分析差异性基因,继而对差异性基因进行火山图表示。通过倍比法(fold-change, FC)和假发现率(false-discovery rate, FDR)筛选显著差异性基因,筛选条件:PDR值 $<0.05$ ,  $P$ 值 $\leq 0.01$ 且 $|\log FC| > 1$ 。对显著性差异化基因进行聚类热图分析<sup>[9]</sup>。

### 1.4 GSEA富集分析

对处理后的GSE43292数据表达值,采用GSEA

软件进行GO功能富集的生物学过程和KEGG的Pathway富集分析<sup>[10]</sup>,分析次数设置为1 000,并按分析的富集分数(ES)、ES进行标准化(NES)值排序。

### 1.5 蛋白质网路分析及核心基因分析

对GSEA分析得到的显著性差异基因(阈值 $>0.4$ )进行STRING分析(<https://string-db.org>),将分析得到的数据导入Cytoscape得到蛋白质表达网路图,并进行核心基因表达分析<sup>[11]</sup>。

### 1.6 统计学方法

结果均采用相关软件检验标准, $P < 0.05$ 为差异有统计学意义。在线R软件GEO2R分析差异表达基因,采用倍数变化和显著性分析,设置 $P$ 值 $\leq 0.01$ 且 $|\log FC| > 1$ 。GSEA软件GO功能富集分析和KEGG通路富集分析采用软件推荐 $P$ 值 $<0.01$ 。PPI软件和Cytoscape软件分析采用推荐 $P$ 值 $<0.05$ 。

## 2 结果

### 2.1 基因芯片数据差异性分析

通过对探针水平的信号值进行质控、变异稳定性分析、背景矫正、数据归一化处理,对基因表达谱中位值进行差异基因分析,发现As斑块中心组织与斑块旁组织的差异表达基因9 840个( $P < 0.05$ ),9 840个差异性基因表达绘制成火山图(图1)。进一步筛选出极其显著性差异表达基因,筛选条件为 $P < 0.01$ 且 $|\log FC| > 1$ ,共筛选出147个,其中表达上调基因(红色)87个,表达下调基因(绿色)60个(图1)。为显示147个极其显著性差异基因之间的关系,按Person系数进行聚类,绘制成热图,其表达结果如图2。

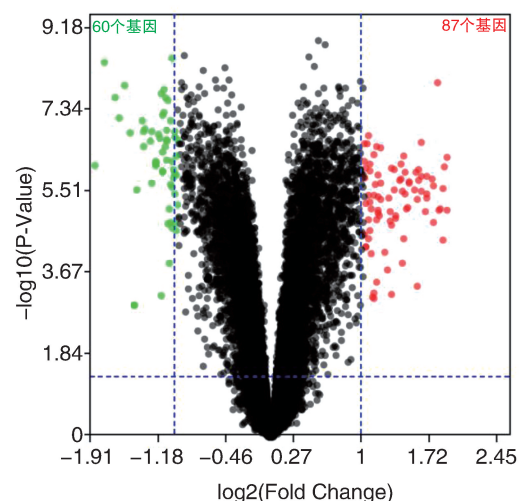


图1. 火山图 共有9 840个差异性基因,显著性差异性基因147个,上调基因87个(红色),下调基因60个(绿色)。

Figure 1. Volcano plot

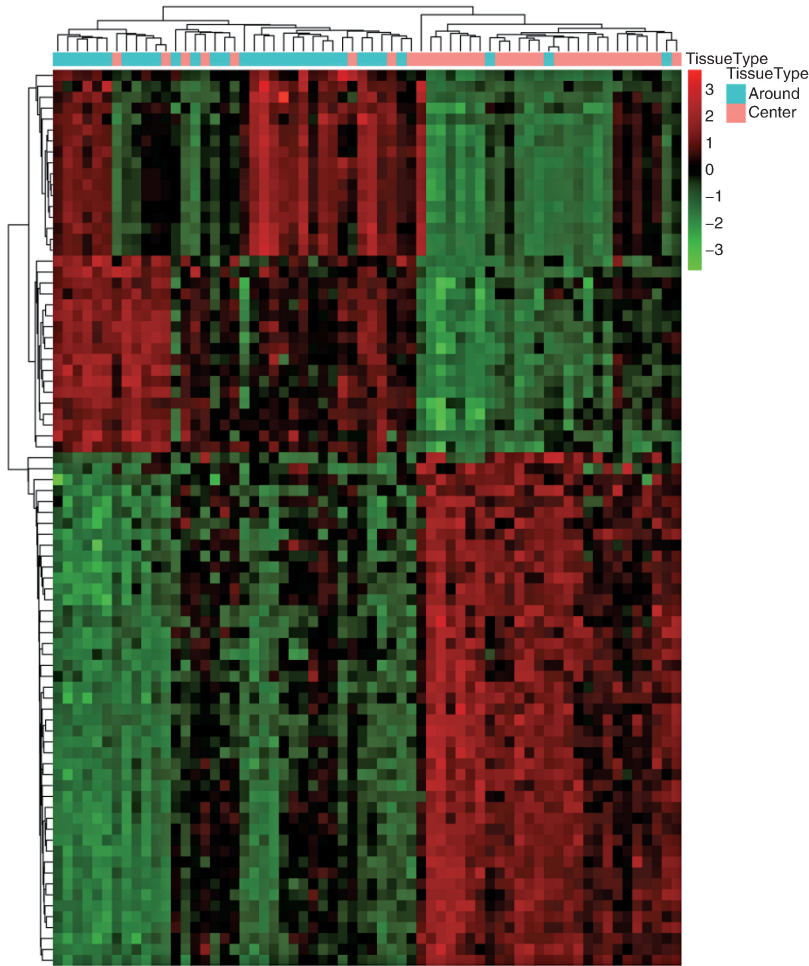


图 2. 147 个显著性差异表达基因的聚类热图

Figure 2. Clustering heat map of 147 significantly different expression genes

2.2 GO 功能富集分析

采用 GSEA 软件研究差异表达基因的富集分析,经过 GSEA 做 GO 功能富集分析,发现许多生物学过程参与 As 病变过程。GO 功能分析显示斑块中上调基因参与 36 个生物学过程,排在前三的生物学过程( $P$  值 $\approx 0$ ,  $FDR \approx 0$ )为抗原结合、丝氨酸水解

酶活性、趋化因子受体结合等,分析结果 NES、ES 值见图 3。GO 功能分析显示斑块中下调基因参与 35 个生物学过程,排在前三的生物学过程( $P$  值 $\approx 0$ ,  $FDR \approx 0$ )为肌性结构组成、细胞外谷氨酸门控离子通道活性、WW 区域结合等,分析结果 NES、ES 值见图 4。

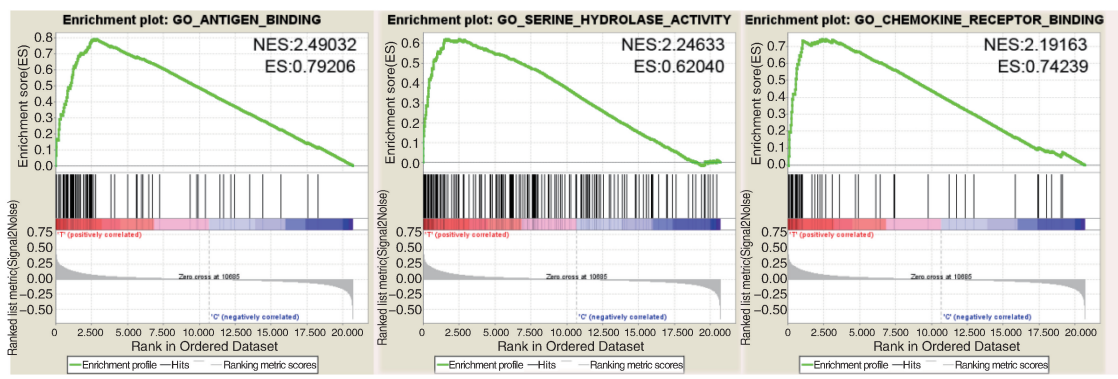


图 3. GSE43292 显著性上调基因的 GO 分析

Figure 3. GO analysis of up-regulated genes in GSE43292



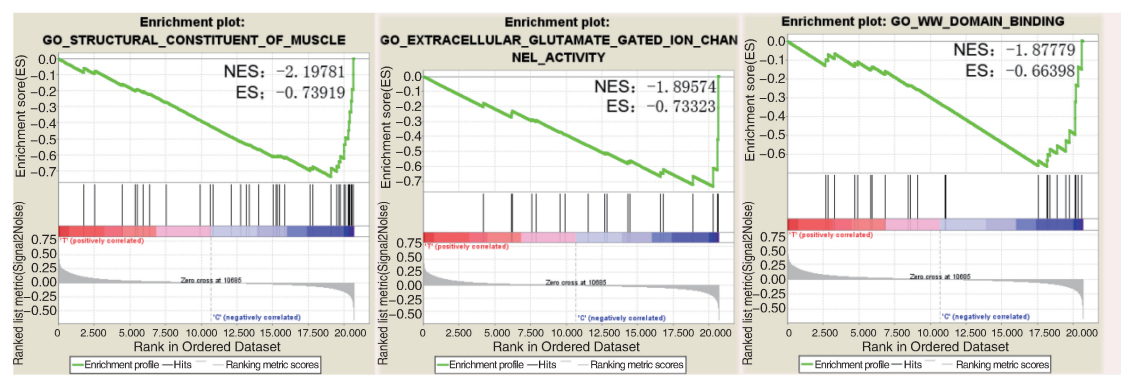


图 4. GSE43292 显著性下调基因的 GO 分析  
Figure 4. GO analysis of down-regulated genes in GSE43292

2.3 KEGG 通路富集分析

经过 GSEA 做 KEGG 通路富集分析,发现许多通路调节参与 As 病变。KEGG 通路分析显示斑块中上调基因参与 5 个通路调节,排在前三的通路调节为造血干细胞系、溶酶体、细胞因子受体相互作用

用等,分析结果 NES、ES 值见图 5。KEGG 通路分析显示斑块中下调基因参与 6 个通路调节,排在前三的通路调节为酪氨酸代谢、苯丙氨酸代谢、扩张型心肌病等,分析结果 NES、ES 值见图 6。

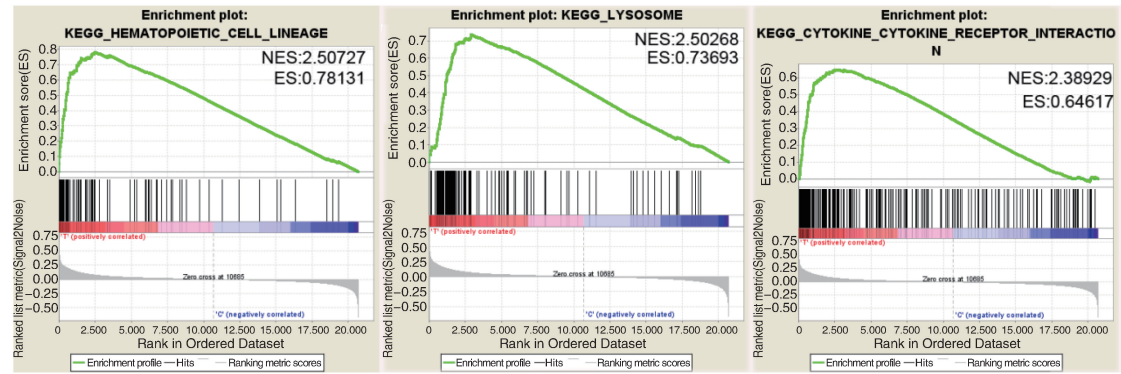


图 5. GSE43292 显著性上调基因的 Pathway 分析  
Figure 5. KEGG pathway analysis of up-regulated genes in GSE43292

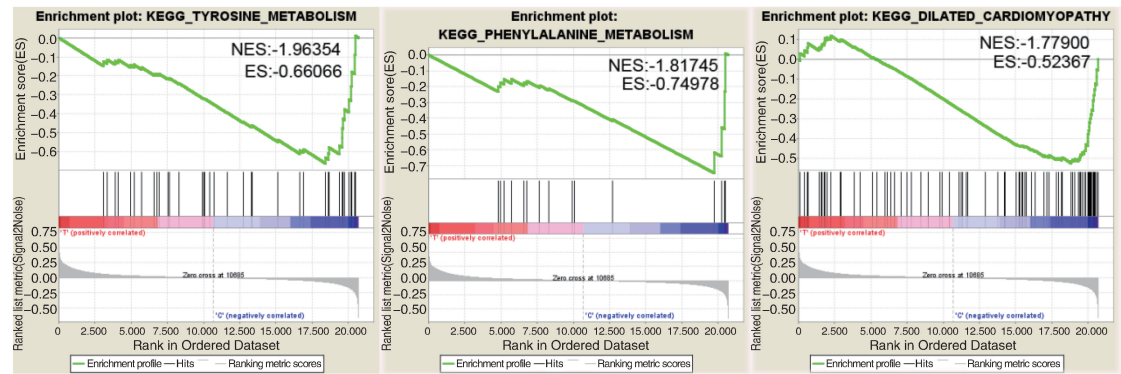


图 6. GSE43292 显著性下调基因的 Pathway 分析  
Figure 6. KEGG pathway analysis of down-regulated genes in GSE43292



## 2.4 蛋白质相互作用和核心基因分析

将 147 个极其显著性差异表达基因输入在线 STRING 软件,分析结果显示节点值 99,边缘值 81, PPI 富集  $P$  值  $<1.0 \times 10^{-16}$ ,主要有单-多细胞器官过程、生物定量调节、血液循环、刺激反应等生物学过

程,有酪氨酸代谢、长期潜在性通路调节。将分析得到的蛋白质之间的相互作用得分值输入 Cytoscape,得到蛋白质相互作用网路图。在评分  $>300$  条件下进行核心基因筛选,得到 IL-8、CXCL-10、SELE、MMP-9、IL-18 共 5 个核心基因(图 7)。

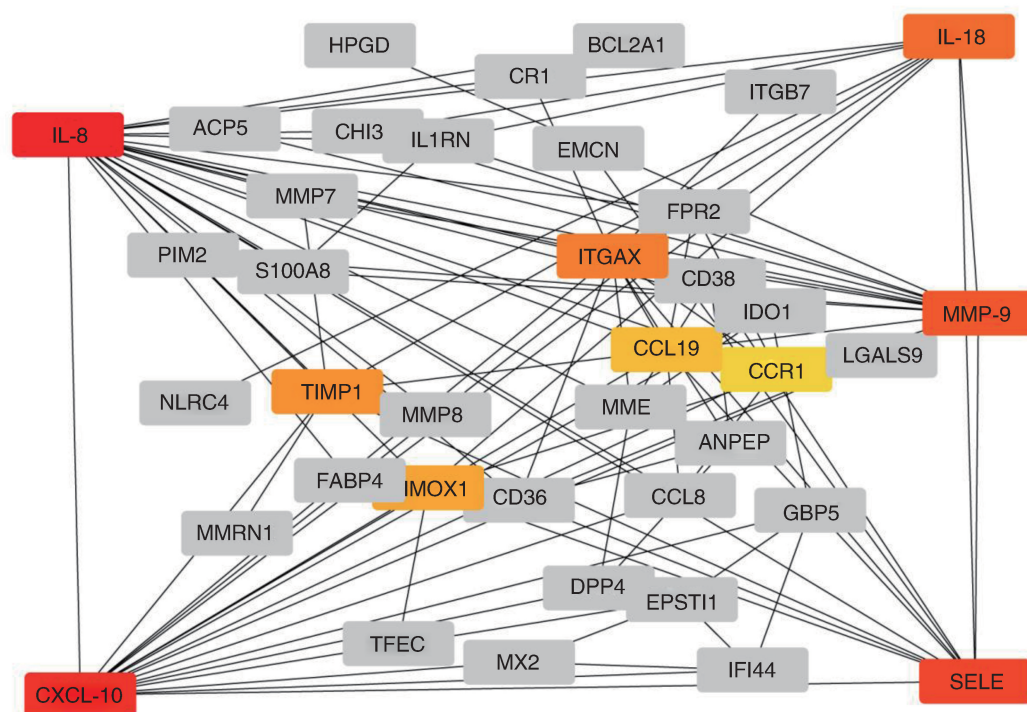


图 7. 显著性差异基因的蛋白质网络图 彩色标记的核心基因位于枢纽位置。

Figure 7. Protein network of differential expression genes

## 3 讨论

颈动脉粥样硬化性血管疾病往往导致颅内或颅外动脉的闭塞或严重狭窄,易造成缺血性脑卒中。血管狭窄程度越严重,斑块越不稳定,发生脑卒中的风险越高。发现 As 发展的分子机制,对发现新药物靶点有重大的积极意义。

近来出现的芯片技术促进了生物信息学的迅猛发展,基因芯片的微阵列研究具有很大的潜力,可以为复杂疾病的发病机制提供新的见解。基因芯片的大数据测试能发现基因表达改变或基因表达谱改变,为揭示疾病的分子机制提供了帮助。本研究利用 PUBMED 已发表的 As 斑块基因芯片 GSE42392 数据,按照 T 检验分析方法,共获取了 147 个极其显著性差异表达基因,其中上调基因 87 个,下调基因 60 个,它们共同构成了斑块形成的显著性差异表达基因谱,与斑块的发生、形成密切相关。

GSEA 是 GO 功能富集分析和 KEGG 通路富集

分析的常用分析软件。GO 功能富集分析显示 As 斑块中心与斑块周边的差异在于:抗原结合<sup>[12-13]</sup>、丝氨酸水解酶活性、趋化因子受体结合<sup>[14]</sup>等。丝氨酸水解酶活性、趋化因子受体结合在以往研究被证实<sup>[4]</sup>。抗原结合是一个新的发现:我们一直认为 As 病变是炎症性反应,很少考虑有抗原-抗体反应过程。As 是否为免疫相关的病变过程,这应该是一个新的方向和思路<sup>[2,13,15]</sup>,需要在后续研究中验证;如能验证 As 与免疫反应有关,可以为治疗提供新的策略。在 KEGG 通路富集研究中,发现 As 斑块中有造血干细胞系、溶酶体、细胞因子受体相互作用等,与我们以往研究巨噬细胞参与 As 形成的结论相一致。我们曾经研究过造血干细胞参与 As 的形成,特别与平滑肌前体细胞的调控相关<sup>[16]</sup>,骨髓干细胞参与 As 平滑肌前体细胞调控过程<sup>[17]</sup>。

在 As 基因分析中,与斑块旁组织相比,我们发现一些下调的生物学过程和通路调节:如 GO 功能富集的肌性结构组成<sup>[18]</sup>、细胞外谷氨酸门控离子

通道活性、WW 区域结合生物过程下调,符合 As 中平滑肌表型转化现象,与相关的文献报道研究相一致<sup>[19]</sup>;KEGG 通路调节的酪氨酸代谢<sup>[20]</sup>、苯丙氨酸代谢、扩张型心肌病等通路下调,提示酪氨酸、苯丙氨酸代谢途径被抑制,激活酪氨酸苯丙氨酸的代谢途径、抑制扩张型心脏病对 As 预防提供了新思路。

我们需要了解 As 斑块中蛋白的相互作用,因而在 As 斑块差异表达基因谱中,通过 STRING 分析结果值导入 Cytoscape 分析:主要有 5 个重要的核心基因 IL-8、CXCL-10、SELE、MMP-9、IL-18。核心基因提示 As 中心部位的炎症因子<sup>[21]</sup>、趋化因子、基质金属蛋白酶等关键性核心基因聚集<sup>[22]</sup>,与相关文献报道 As 中炎症、免疫应答、溶酶体通路、细胞因子信号通路等相一致<sup>[22]</sup>。

我们从生物信息学角度分析了颈动脉斑块与斑块旁组织之间基因表达差异,分析了主要差异表达基因谱的生物学过程和通路调节,列出相关强度值排在前 3 位 NES、ES 值,为 As 的病变过程和分子通路调节提供基础数据支持。我们根据差异表达基因谱,分析了相关的蛋白质关系,列出相关强度值最高的 5 个关键核心基因,可以作为临床 As 预防提供参考意见。在基因数据分析中,采用的分析方法差异、阈值的不同可能会得出不同的分析结果<sup>[23]</sup>,这些分析结果为深层机制提供线索,进一步通过实际的病例、样本予以验证,本研究需要在今后的实验中收集样本验证上述分析结果。

#### [参考文献]

- [1] Tarkin JM, Joshi FR, Rudd JH. PET imaging of inflammation in atherosclerosis[J]. *Nat Rev Cardiol*, 2014, 11(8): 443-457.
- [2] Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis[J]. *Cell*, 2011, 145(3): 341-355.
- [3] Pfeiler S, Gerdes N. Atherosclerosis: cell biology and lipoproteins focus on anti-inflammatory therapies[J]. *Curr Opin Lipidol*, 2018, 29(1): 53-55.
- [4] Back M, Hansson GK. Anti-inflammatory therapies for atherosclerosis[J]. *Nat Rev Cardiol*, 2015, 12(4): 199-211.
- [5] Sackmann EK, Fulton AL, Beebe DJ. The present and future role of microfluidics in biomedical research[J]. *Nature*, 2014, 507(7491): 181-189.
- [6] Hou Y, Zhang X, Hou X, et al. Rapid pathogen identification using a novel microarray-based assay with purulent meningitis in cerebrospinal fluid[J]. *Sci Rep*, 2018, 8(1): 15965.
- [7] Ayari H, Bricca G. Identification of two genes potentially associated in iron-heme homeostasis in human carotid plaque using microarray analysis[J]. *J Biosci*, 2013, 38(2): 311-315.
- [8] Giacomello S, Lundeberg J. Preparation of plant tissue to enable spatial transcriptomics profiling using barcoded microarrays[J]. *Nat Protoc*, 2018, 13(11): 2425-2446.
- [9] Zhang Z, Murtagh F, Van Poucke S, et al. Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R[J]. *Ann Transl Med*, 2017, 5(4): 75.
- [10] Croken MM, Qiu W, White MW, et al. Gene Set Enrichment Analysis (GSEA) of *Toxoplasma gondii* expression data sets links cell cycle progression and the bradyzoite developmental program[J]. *BMC Genomics*, 2014, 15(1): 515.
- [11] Carlin DE, Demchak B, Pratt D, et al. Network propagation in the cytoscape cyberinfrastructure[J]. *PLoS Comput Biol*, 2017, 13(10): e1005598.
- [12] Nour-Eldine W, Joffre J, Zibara K, et al. Genetic depletion or hyperresponsiveness of natural killer cells do not affect atherosclerosis development[J]. *Circ Res*, 2018, 122(1): 47-57.
- [13] Getz GS, Reardon CA. Natural killer T cells in atherosclerosis[J]. *Nat Rev Cardiol*, 2017, 14(5): 304-314.
- [14] van den Borne P, Quax PH, Hoefer IE, et al. The multifaceted functions of CXCL10 in cardiovascular disease[J]. *Biomed Res Int*, 2014, 2014: 893106.
- [15] Wang Y, Subramanian M, Yurdagul A, et al. Mitochondrial fission promotes the continued clearance of apoptotic cells by macrophages[J]. *Cell*, 2017, 171(2): 331-345.
- [16] Yu J, Li Y, Li M, et al. Oxidized low density lipoprotein-induced transdifferentiation of bone marrow-derived smooth muscle-like cells into foam-like cells in vitro[J]. *Int J Exp Pathol*, 2010, 91(1): 24-33.
- [17] Li M, Yu J, Li Y, et al. CXCR4 positive bone mesenchymal stem cells migrate to human endothelial cell stimulated by ox-LDL via SDF-1alpha/CXCR4 signaling axis[J]. *Exp Mol Pathol*, 2010, 88(2): 250-255.
- [18] Durham AL, Speer MY, Scatena M, et al. Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness[J]. *Cardiovasc Res*, 2018, 114(4): 590-600.
- [19] Rosenson RS, Brewer HB, Ansell BJ, et al. Dysfunctional HDL and atherosclerotic cardiovascular disease[J]. *Nat Rev Cardiol*, 2016, 13(1): 48-60.
- [20] Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis[J]. *Nat Med*, 2013, 19(5): 576-585.
- [21] Kojima Y, Volkmer JP, McKenna K, et al. CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis[J]. *Nature*, 2016, 536(7614): 86-90.
- [22] Wang J, Sun C, Gerdes N, et al. Interleukin 18 function in atherosclerosis is mediated by the interleukin 18 receptor and the Na-Cl co-transporter[J]. *Nat Med*, 2015, 21(7): 820-826.
- [23] Nai W, Threapleton D, Lu J, et al. Identification of novel genes and pathways in carotid atheroma using integrated bioinformatic methods[J]. *Sci Rep*, 2016, 6: 18764.

(此文编辑 许雪梅)