

网络出版时间: 2019-6-10 17:48 网络出版地址: <http://kns.cnki.net/kcms/detail/34.1065.R.20190606.1642.002.html>

利用基因表达芯片探索多发性硬化的关键基因和通路

黄帆 唐玉兰 廖韦静 邝慧敏 蓝岚

摘要 目的 以 GEO 数据库中 GSE21942 系列芯片数据为分析对象, 分析多发性硬化(MS) 中的高通量基因表达来获得差异表达的基因, 筛选出与多发性硬化相关的关键基因和通路。方法 利用 R 语言与 DAVID、STRING 等方法对差异表达的基因进行 GO 和 KEGG 富集分析, 构建和分析蛋白质-蛋白质相互作用网络。结果 88 个基因被鉴定为差异表达的基因, 其中 76 个基因(86.4%) 上调, 12 个基因(13.6%) 下调。Pathway 通路分析显示, 差异表达基因主要与造血细胞谱系、B 细胞受体信号通路、破骨细胞分化、氮代谢等有关, 进一步分析筛选出 ALAS2、CA1、SNCA、HBB、IL8 等核心基因, 这些基因在血红素生物合成、突触传递、神经退行性疾病、血脑屏障破坏等生物功能中起重要作用。结论 与健康人相比, 多发性硬化样本中相关靶基因表达有较大差异, 这些数据可能为多发性硬化相关的发病机制和临床治疗提供新思路。

关键词 多发性硬化; 基因芯片; 差异表达基因; 生物标志物
中图分类号 R 741

文献标志码 A 文章编号 1000-1492(2019)07-1009-07
doi: 10.19405/j.cnki.issn1000-1492.2019.07.002

多发性硬化(multiple sclerosis, MS) 是一种中枢神经系统的炎症性和神经退行性疾病, 其发生是由遗传和环境因素共同作用的结果^[1]。在西方国家 MS 是造成青壮年神经性残疾中仅次于外伤的第二大因素, 据 2013 年统计, 估计全球约有 230 万人患有 MS, 患病率约为(50~300) 人/10 万, 且由于印度和中国等大型人群相对缺乏数据, 数据可能被低估^[2]。早期 MS 的特征通常是神经功能缺损急性发作, 其依赖于中枢神经系统急性炎症性脱髓鞘病变的区域及炎症反应的程度, 而导致病变髓鞘再生的形成机制尚未完全了解。位于主要染色体 6p21 中组织相容性复合体(major histocompatibility complex,

MHC) 区域内的人类白细胞抗原(human leukocyte antigen, HLA) 等位基因, 已被确定为该疾病的主要遗传决定因素。另外, 已经描述了超过 100 种非 MHC MS 的易感性变体, 携带已知的易感性变体相关基因参与调节免疫细胞分化或信号传导^[3]。不恰当的诊断标准用于不典型症状的脱髓鞘患者是导致其误诊的主要原因, 遗传研究的目标是实现更精确的表示疾病发病机制中的相关基因、通路和网络, 并利用这些信息发现预防、治疗和修复的新靶点。

该研究拟基于 GEO 基因表达数据, 通过生物信息学分析多发性硬化疾病状态, 构建基因网络并筛选潜在的关键分子靶点, 为寻找 MS 发病机制提供新途径, 或许可以用于早期诊断多发性硬化并为临床治疗提供新思路。

1 材料与方法

1.1 数据来源 Gene Expression Omnibus(GEO) 数据库隶属于美国国立卫生研究院的美国国立生物技术信息中心, 是高通量基因表达数据和杂交阵列、芯片、微阵列的数据库。以“Multiple Sclerosis”为搜索词进入数据库获取基因表达谱 GSE21942 种属为人类 芯片平台 GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array。该芯片数据包括 14 例多发性硬化患者及 15 例健康人的基因表达阵列数据。

1.2 差异表达基因(differentially-expressed genes, DEGs) 处理 使用 R 语言和 Bioconductor(<http://www.bioconductor.org/>) 进行基因芯片数据分析。利用 GPL570 平台对应的 hgu133plus2.db 注释 R 包进行基因探针注释, 取每个探针的最大值作为该探针的表达值, 除去无法注释的探针。通过稳健多元阵列平均(robust multivariate array average, RMA) 对原始表达数据进行对数转换, 背景校正和归一化处理。按照 $|\log_2 \text{FC}| > 1$ 和 $P < 0.05$ 作为筛选标准, 使用 limma 包筛选 14 名正常人样本和 15 名多发性硬化患者样本基因得出 88 个 DEGs, 绘制 DEGs 热图和火山图。

1.3 样本的主成分分析 主成分分析可将高维数

2019-03-14 接收

基金项目: 国家自然科学基金(编号: 81260188、81460194); 广西自然科学基金(编号: 2012GXNSFAA053082、2015GXNSFAA139212)

作者单位: 广西医科大学第一附属医院老年神经内科, 广西 530021
作者简介: 黄帆, 女, 硕士研究生;

唐玉兰, 女, 主任医师, 硕士生导师, 责任作者, E-mail: tan-yulan7@163.com

据处理成低维数据,利用 R 语言 factoextra 包对 DEGs 数据集进行主成分分析。

1.4 样本的聚类分析 聚类分析是一种无监督机器学习分析方法,可以把相似的对象分成不同的组别或更多的子集。利用 R 语言 cluster 包进行聚类分析,并利用 factoextra 包将聚类分析结果可视化处理。

1.5 基因本体论和 pathway 分析 DEGs 将获得的 DEGs 通过在线软件数据库(DAVID, <https://david.ncifcrf.gov/>) 进行注释、可视化分析,以确定过度表现的 GO 类别和 Pathway 生物通路。GO 分析可以确定大量基因的生物学意义,并对基因产物功能进行分类,包括生物过程(biological processes, BP)、分子功能(molecular functions, MF) 和细胞组分(cellular components, CC)。通过 Pathway 生物通路富集分析,筛选的基因可能与两个或更多信号传导途径有关,由于不同途径中的基因相同,因此途径之间的重叠是不可避免的。基于 KEGG 数据库对 DEGs 进行基因信号通路富集分析,选择 DEGs 富集最显著的 10 个功能进行排序并分析。

1.6 蛋白-蛋白相互作用信号网络分析 DEGs 用于检索交互基因/蛋白质的搜索工具(STRING)数据库(<http://string.embl.de/>) 提供了 PPI 的关键评估和整合,用于评估直接(物理)和间接(功能)关联的 DEGs。本研究通过 STRING 数据库绘制靶基因编码蛋白互作网络图,以便了解差异基因之间的相互调控关系。

2 结果

2.1 DEGs 数据 对数化标准处理前后的多发性硬化组和健康对照组基因芯片数据结果见图 1、2,可知本研究标本经对数化处理后数据均一性较好,两组之间具有可比性。

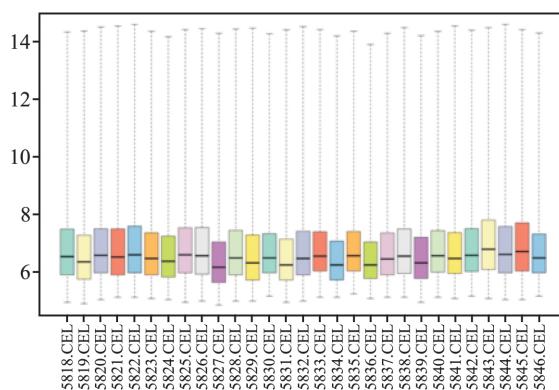


图 1 样本芯片数据对数化标准处理前

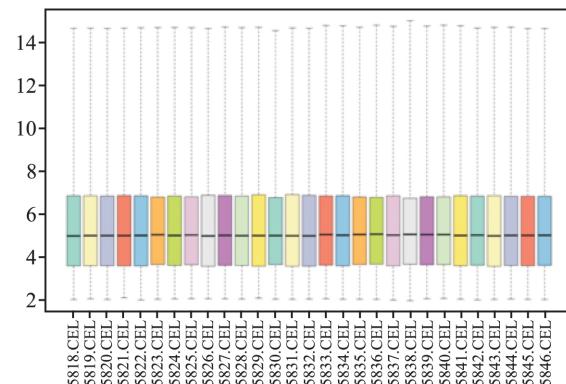


图 2 样本芯片数据对数化标准处理后

2.2 DEGs 筛选结果 通过对两组数据的 DEGs 进行筛选,构建火山图并显示 88 个 DEGs(图 3A),与对照组比较,多发性硬化组上调基因 76 个、下调基因 12 个,图 3C 显示了上调基因和下调基因中有显著差异的 10 个;88 个基因构建的热图(图 4),其中横坐标表示 15 个 MS 组和 14 个健康组样本;红色代表上调的基因,绿色代表下调的基因,差异越大,颜色越深。

2.3 样本聚类分析 通过 R 语言 factoextra 包的增强聚类分析函数对差异基因数据集进行层次聚类,自动计算最佳聚类簇两簇,并且对结果进行可视化处理(图 5)。此聚类结果表明,筛选的差异基因能较好的区分出健康组和 MS 组。

2.4 DEGs 数据集主成分分析 使用 R 语言 FactoMineR 包对差异基因数据集进行主成分分析,实现高维数据降维成为简化的数据,用 factoextra 包可视化主成分分析结果(图 6A、B),主要有贡献率最大的两个主成分,主成分分析结果能较好地区分健康组和 MS 组(图 6C)。

2.5 DEGs GO 分析结果 采用线软件数据库(DAVID, <https://david.ncifcrf.gov/>) 对 DEGs 进行 GO 功能富集分析,生物学过程中,富含 MS 中 DEGs 的 GO 术语包括固有免疫反应、血液凝固、体液免疫反应和氧转运。在细胞组分类别中,富含 GO 的术语主要是血红蛋白复合物、质膜和细胞外间隙。在分子功能分类中,包括蛋白结合、血红素结合和跨膜信号受体活性(图 7);从红色到蓝色,颜色越蓝,表示负相关程度越大;颜色越红,表示正相关程度越大(图 8)。

2.6 DEGs pathway 分析结果 表 1 显示了 4 个差异最显著的富集通路,其主要涉及造血细胞谱系、B

细胞受体信号通路、破骨细胞分化、氮代谢等。

2.7 DEGs 信号网络分析结果 利用 STRING 在线数据库构建差异基因所编码的蛋白互作网络, 可

得出 27 个蛋白存在相互作用(图 9)。图 9A 是差异基因表达蛋白互作网络, 图 9B 所示的是互作网络中筛选出的核心基因节点数, 其中节点数

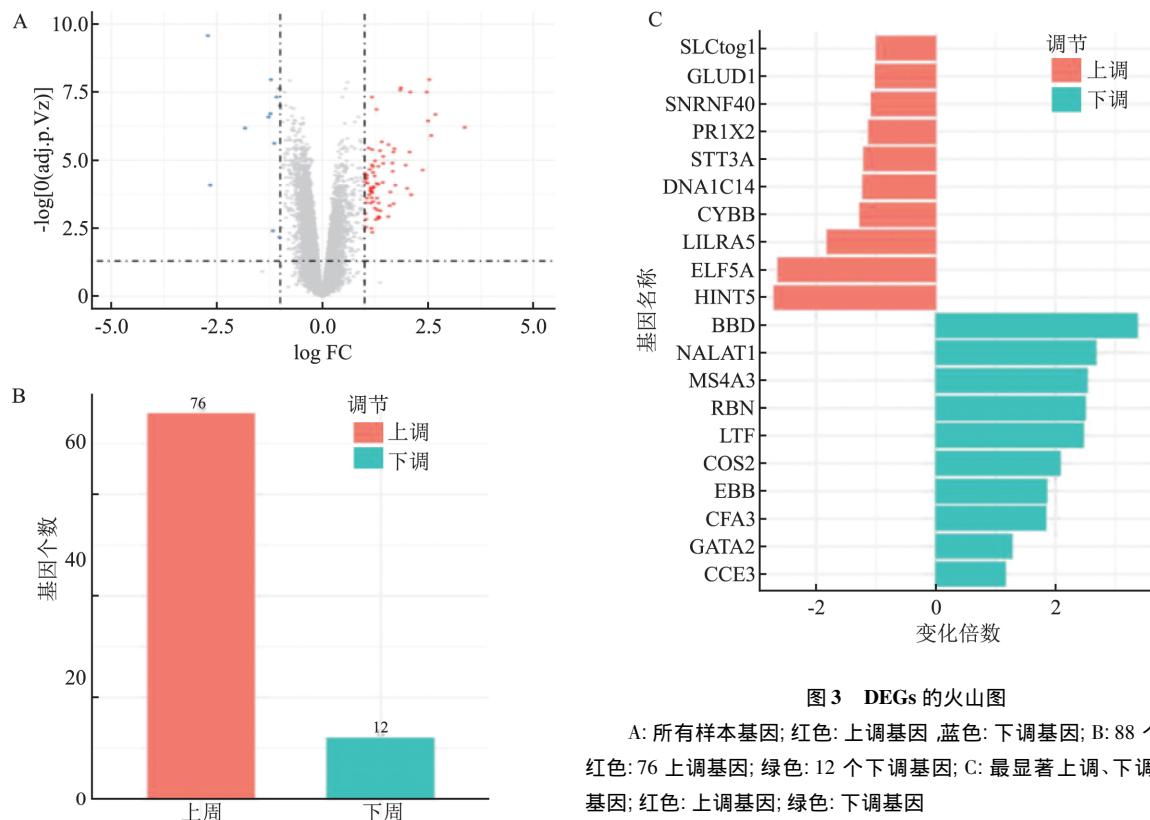


图 3 DEGs 的火山图

A: 所有样本基因; 红色: 上调基因, 蓝色: 下调基因; B: 88 个 DEGs; 红色: 76 上调基因; 绿色: 12 个下调基因; C: 最显著上调、下调的前 10 基因; 红色: 上调基因; 绿色: 下调基因

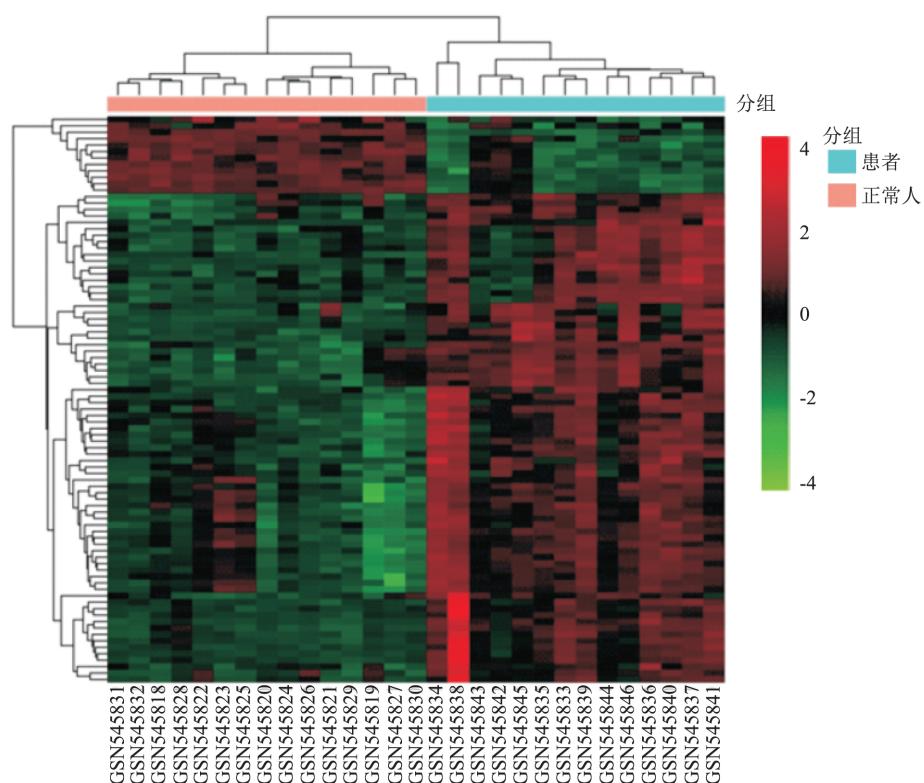


图 4 差异基因表达值参差聚类热图

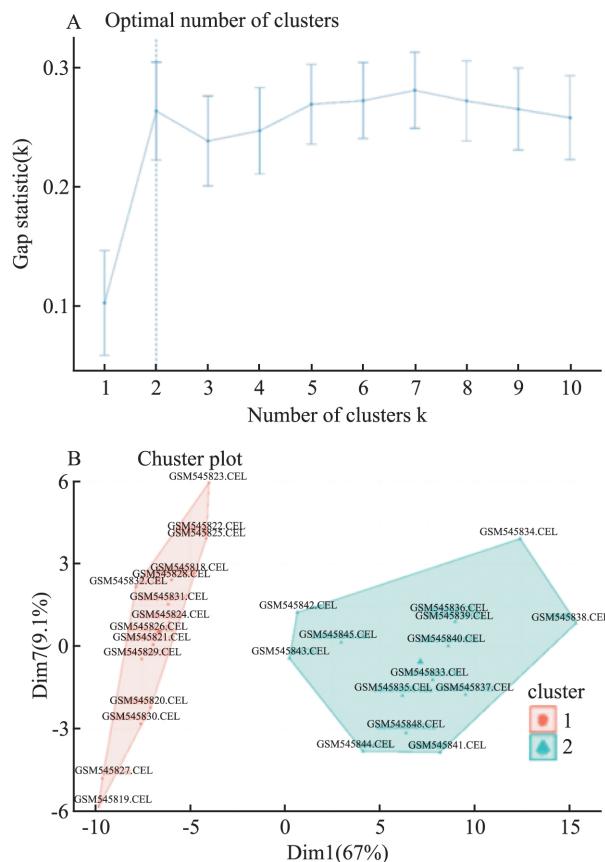


图5 聚类分析图
A: 最佳聚类簇为2; B: 绿色是正常人

表1 DEGs 的 KEGG 通路分析		
通路注释	基因数	基因名称
造血细胞谱系	6	CD3G, CR2, IL7, MS4A1, CD24, ITGB3
B 细胞受体信号通路	4	CR2, RASGRP3, CD72, BLNK
破骨细胞分化	4	CYBB, LILRA5, ITGB3, BLNK
氮代谢	2	GLUD1, CA1

≥4 的 DEGs 如 HBD、IL-8、SNCA、ALAS2 等。

3 讨论

MS 的特征在于脱髓鞘和进行性神经功能障碍，既往研究报道了线粒体参与 MS 中的神经变性和残疾，包括核编码的电子传递链亚基基因的表达降低和呼吸复合物的抑制^[4]。本研究中通过使用微阵列的方法从基因水平揭示 MS 发病机制可能涉及的关键基因：ALAS2、CA1、SNCA、HBB、IL8 等。其中血红蛋白亚基 β（亦称 β-珠蛋白，hemoglobin-β，HBB）与 α 珠蛋白（HBA）一起构成成人中最常见的血红蛋白形式 HbA，血红蛋白是一种可以诱导局部氧化应激、炎症和组织损伤的反应性分子^[5]，变异的血红蛋白表达可能与各种神经退行性疾病有

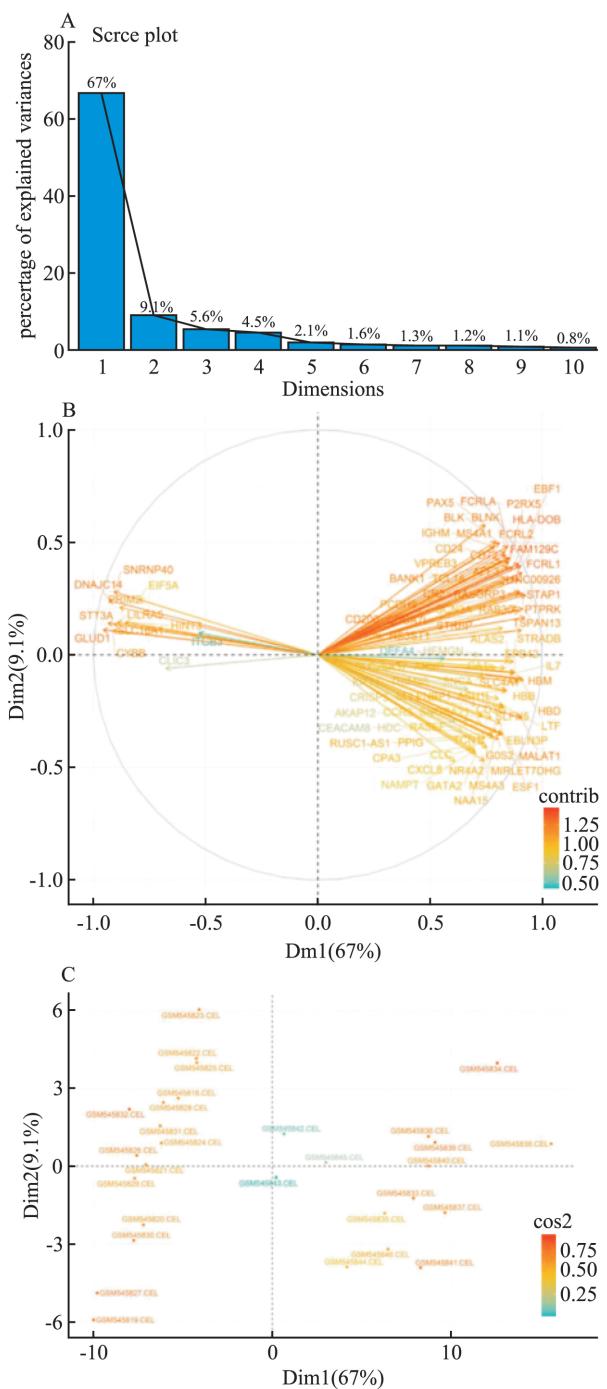


图6 样本和DEGs数据集主成分分析
A: DEGs 数据集主成分分析; B: 样本数据集主成分分析; C: 样本数据集主成分分析(左侧为健康组 右侧为 MS 组)

关^[6]；HBB 被发现在 MS 大脑病变皮层分离的线粒体部分中水平增加，被认为可能是将神经元能量学与细胞核中组蛋白的表观遗传变化联系起来的机制的一部分，并且可以通过支持神经元代谢在 MS 中提供神经保护^[7]。碳酸酐酶（carbonic anhydrase 1，CA1）通过激肽释放酶原激活和丝氨酸蛋白酶因子 XIIa 生成来介导出血性视网膜和脑血管通透性，这

些现象诱发增殖性糖尿病性视网膜病变和糖尿病性黄斑水肿疾病进展,是视力丧失的主要原因^[8],这可能与多发性硬化临床表现中的视力受损发生机制

相关。 α -突触核蛋白(α -synuclein, SNCA)的过量产生可能是帕金森病的致病因素,其主要存在于神经细胞(神经元)的突触前末端,通过聚集突触小

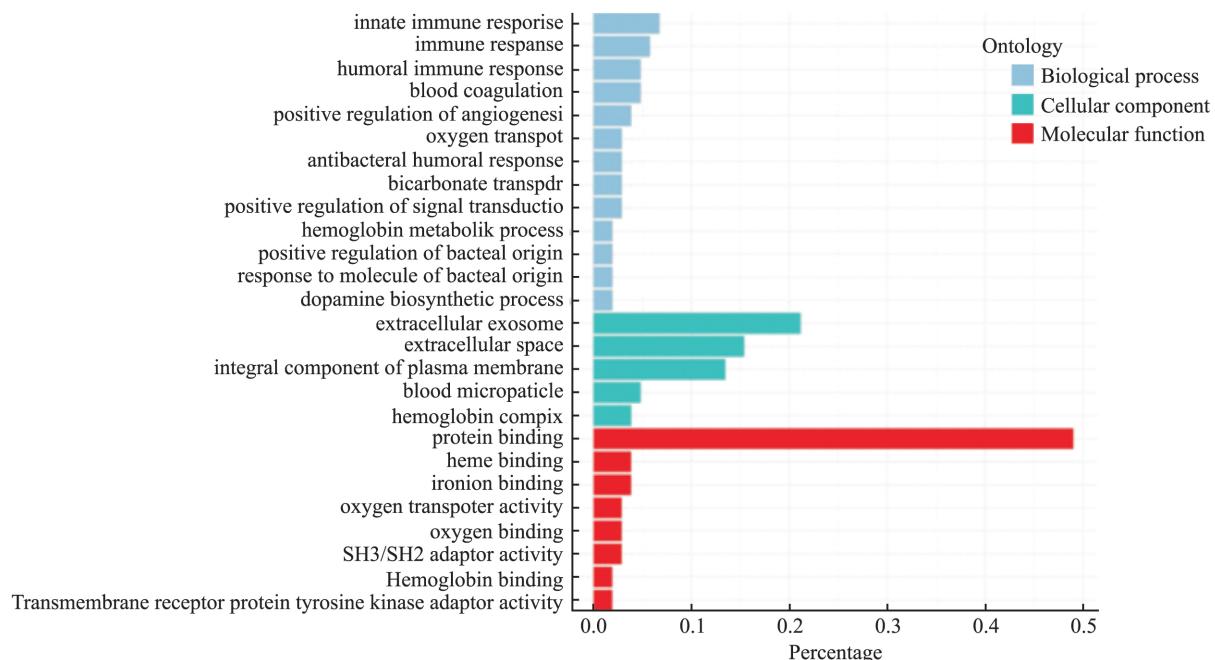


图7 GO功能富集分析差异基因

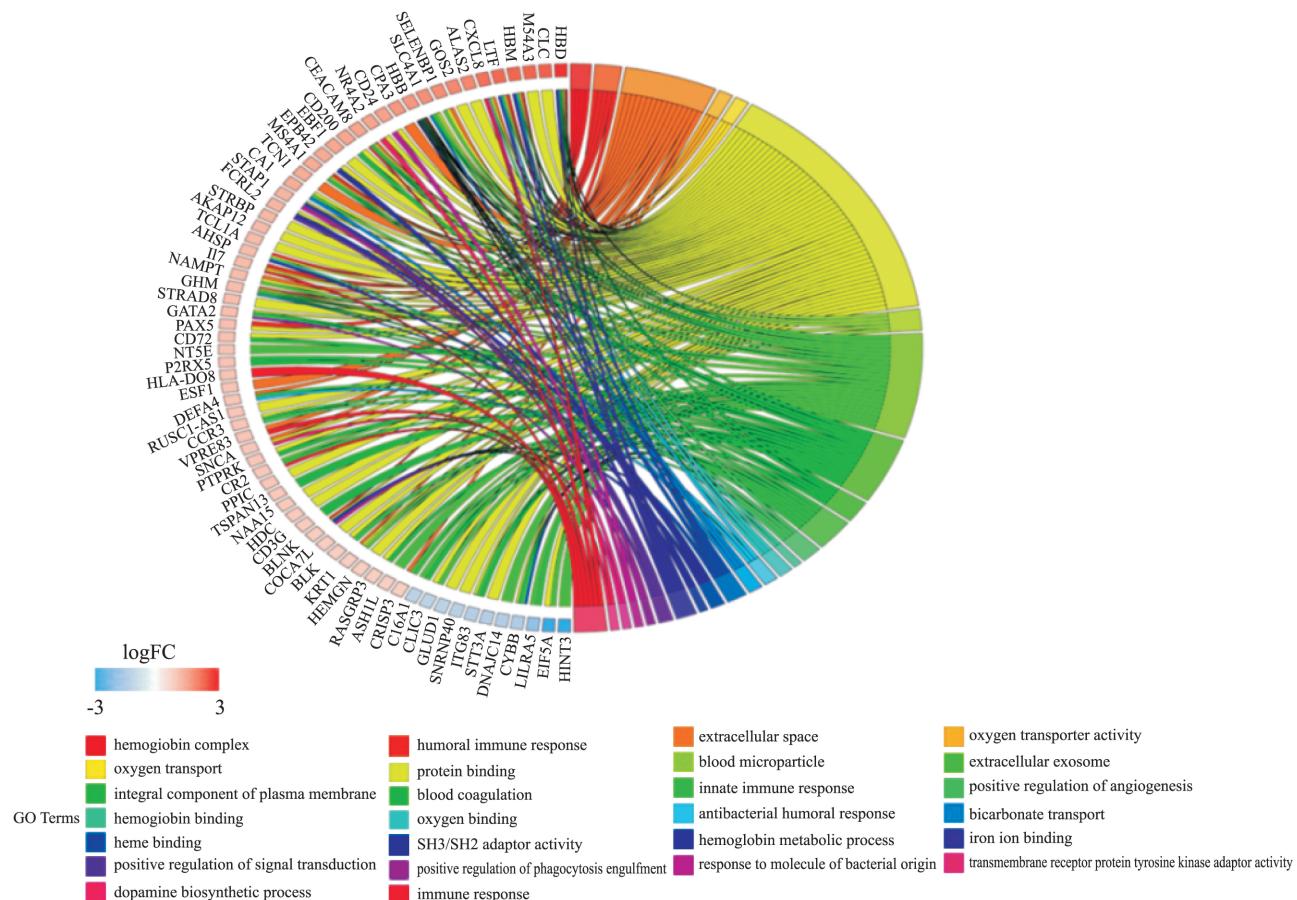


图8 差异基因GO功能富集分析

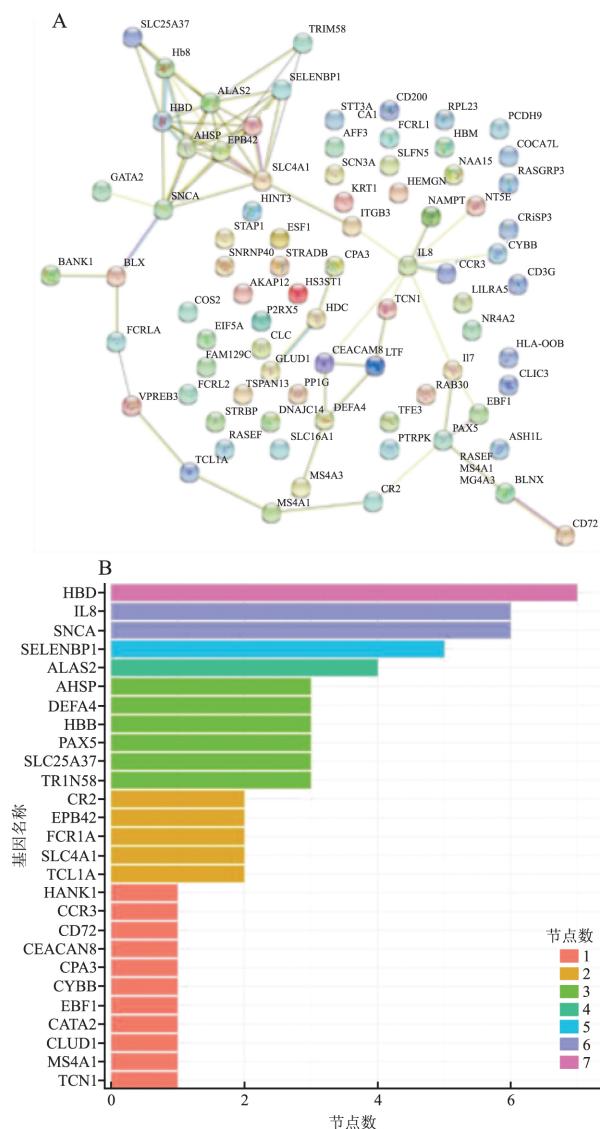


图 9 差异基因表达蛋白互相作用网络及核心基因节点数

A: DEGs 蛋白互相作用网络分析; B: DEGs 共表达显著差异的节点数

泡在维持突触前终末突触小泡的供应方面发挥作用^[9],它也可能有助于调节多巴胺的释放。白细胞介素-8(interleukin-8, IL-8)是先天免疫系统反应中免疫反应的重要介质,其分泌增加了氧化应激,从而引起炎症细胞的募集和诱导氧化应激介质的进一步增加,使其成为局部炎症的关键因素^[10]; IL-8 也被称为嗜中性粒细胞趋化因子,能诱导靶细胞趋化,主要是嗜中性粒细胞以及其他粒细胞,导致它们向感染部位迁移,IL-8 在抵达后也会诱导吞噬作用;研究发现,MS 患者脑脊液中 IL-8 的水平显著高于对照组,血清 IL-8 水平显著低于对照组,这些差异可能与血脑屏障的损伤有关^[11]。由此,本文筛选的 DEGs 可能在 MS 发病相关的炎症反应或神经退行

性过程中扮演重要角色。

通过对 DEGs 进行 GO 分析显示,二者差异最显著的功能主要涉及体液免疫反应、固有免疫反应、蛋白结合、血红素结合和跨膜信号受体活性等方面。通过对二者进行 pathway 分析发现,其差异最显著富集通路主要涉及造血细胞谱系、B 细胞受体信号通路、破骨细胞分化、氮代谢等方面。小胶质细胞属于成体组织中存在的髓系细胞谱系,其在器官发生过程中从不同于造血干细胞的卵黄囊红细胞-骨髓祖细胞(yolk-sac erythro-myeloid progenitors, EMPs)发育而来,研究显示:BRAF(V600E)在小鼠 EMP 中的镶嵌表达导致组织驻留巨噬细胞的克隆性扩增和严重的迟发性神经退行性疾病^[12]。部分研究^[13]已经证明,在特定条件下 Tregs 可以产生炎性细胞因子,其中自身反应性 CD4⁺ T 细胞对髓鞘自身抗原发生免疫应答参与发病过程,自身反应性 CD8⁺ T 细胞和 B 细胞等也参与免疫损伤作用。Strom et al^[14]利用基于 ApoE -/- 小鼠颈动脉周围的血管周围环的放置的新内膜形成模型来确定 B 细胞和 B 细胞亚群是否赋予针对损伤发展的保护,结果显示源自淋巴结的 B2-B 细胞或纯化的 CD21(hi) CD23(hi) CD24(hi) B 细胞向同基因小鼠的继发性转移减少了损伤大小和炎症,而不改变血清胆固醇水平,IL-10 阻断或转移 IL-10 缺陷型 B 细胞阻止了淋巴结衍生的 B 细胞介导的保护,这可能为多发性硬化中的免疫调节方法开辟道路。Niedziela et al^[15]通过评估血清一氧化氮及其反应性衍生物(NOx)作为复发缓解型多发性硬化的氮类和炎症参数之一,并比较各种类型的减少一氧化氮和炎性生物标志物的疾病缓解疗法的有效性,在一线药物治疗的受试者中证实血清 NOx 水平和 MS 持续时间之间呈负相关。随后的前瞻性研究将需要进一步确定这些核心基因在 MS 发病机制中的功能。

2017 年修订版 MS 诊断标准提出,应结合 MRI 和血清学检测及临床特征和病史,探索特发性炎症疾病的差异诊断,包括视神经脊髓炎谱系障碍以及其他可以类似 MS 的复发性疾病。视神经脊髓炎谱系疾病和 MS 同为中枢神经系统脱髓鞘疾病,二者的临床表现和影像学特征相似,但治疗策略却有所不同,如预防 MS 复发的疾病修饰治疗(如 β 干扰素、芬戈莫德、那他珠单抗)可加重视神经脊髓炎^[16]。因此,筛选识别期生物靶标,结合基因功能和参与的通路分析,有助于研究疾病的发病机制,为尽早确诊及制定治疗方案提供判别依据。

参考文献

- [1] Michel L. Environmental factors in the development of multiple sclerosis [J]. *Retour Au Numéro* 2018, 174(6): 372–7.
- [2] Browne P, Chandraratna D, Angood C, et al. Atlas of multiple sclerosis 2013: A growing global problem with widespread inequity [J]. *Neurology*, 2014, 83(11): 1022–4.
- [3] Andlauer T F M, Buck D, Antony G, et al. Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation [J]. *Sci Adv* 2016, 2(6): e1501678.
- [4] Singhal N K, Alkhayer K, Shelestak J. Erythropoietin upregulates brain hemoglobin expression and supports neuronal mitochondrial activity [J]. *Mol Neurobiol* 2018, 55(10): 8051–8.
- [5] Bamm V V, Henein M E L, Sproul S L J, et al. Potential role of ferric hemoglobin in MS pathogenesis: effects of oxidative stress and extracellular methemoglobin or its degradation products on myelin components [J]. *Free Radic Biol Med* 2017, 112: 494–503.
- [6] Vanni S, Zattini M, Moda F, et al. Hemoglobin mRNA changes in the frontal cortex of patients with neurodegenerative diseases [J]. *Front Neurosci* 2018, 12: 8.
- [7] Brown N, Alkhayer K, Clements R, et al. Neuronal hemoglobin expression and its relevance to multiple sclerosis neuropathology [J]. *J Mol Neurosci* 2016, 59(1): 1–17.
- [8] Gao B B, Clermont A, Rook S, et al. Extracellular carbonic anhydrase mediates hemorrhagic retinal and cerebral vascular permeability through prekallikrein activation [J]. *Nat Med*, 2007, 13(2): 181.
- [9] Diao J, Jacqueline B, Sandro V, et al. Native α -synuclein induces clustering of synaptic-vesicle mimics via binding to phospho-
- lipids and synaptobrevin-2/VAMP2 [J]. *Elife*, 2013, 2(2): e00592.
- [10] Vlahopoulos S, Boldogh I, Casola A, et al. Nuclear factor-kappaB-dependent induction of interleukin-8 gene expression by tumor necrosis factor alpha: evidence for an antioxidant sensitive activating pathway distinct from nuclear translocation [J]. *Blood*, 1999, 94(6): 1878–89.
- [11] Matejcíková Z, Mareš J, Sládková V, et al. Cerebrospinal fluid and serum levels of interleukin-8 in patients with multiple sclerosis and its correlation with Q-albumin [J]. *Mult Scler Relat Disord*, 2017, 14: 12–5.
- [12] Mass E, Jacome-Galarza C E, Blank T, et al. A somatic mutation in erythro-myeloid progenitors causes neurodegenerative disease [J]. *Nature* 2017, 549(7672): 389–93.
- [13] Agahozo M C, Peferoen L, Baker D, et al. CD20 therapies in multiple sclerosis and experimental autoimmune encephalomyelitis targeting T or B cells? [J]. *Mult Scler Relat Disord* 2016, 9: 110–7.
- [14] Strom A C, Cross A J, Cole J E, et al. B regulatory cells are increased in hypercholesterolaemic mice and protect from lesion development via IL-10 [J]. *Thromb Haemost* 2015, 114(4): 835–47.
- [15] Niedziela N, Adamczyk-Sowa M, Niedziela J T, et al. Assessment of serum nitrogen species and inflammatory parameters in relapsing-remitting multiple sclerosis patients treated with different therapeutic approaches [J]. *Biomed Res Int* 2016, 2016: 4570351.
- [16] 张瑛, 管阳太. 2017 修订版 McDonald 多发性硬化诊断标准解读 [J]. *神经病学与神经康复学杂志* 2018, 14(1): 1–5.

Exploring key genes and pathways in multiple sclerosis using gene expression chips

Huang Fan, Tang Yulan, Liao Weijing, et al

(Dept of Elderly Neurology, The First Affiliated Hospital of Guangxi Medical University, Guangxi 530021)

Abstract Objective Based on the GSE21942 series of chip data in the GEO database, the high-throughput gene expression in multiple sclerosis (MS) was analyzed to obtain differentially expressed genes (DEGs), and key genes and pathways related to multiple sclerosis were screened. **Methods** Using the R language and DAVID, STRING and other methods to carry out GO and KEGG enrichment analysis of DEGs, construct and analyze the protein-protein interaction network. **Results** Eighty-eight genes were identified as differentially expressed genes, of which 76 (86.4%) were up-regulated and 12 (13.6%) were down-regulated. Pathway analysis found that differential genes are mainly related to hematopoietic cell lineage, B cell receptor signaling pathway, Osteoclast differentiation, Nitrogen metabolism, etc. Further analysis, screening for core genes such as ALAS2, CA1, SNCA, HBB, IL8, etc. It played an important role in biological functions such as heme biosynthesis, synaptic transmission, neurodegenerative diseases, and blood-brain barrier destruction. **Conclusion** Compared with healthy people, the expression of relevant target genes in MS samples is quite different. These data may provide new insights for MS function and molecular mechanism research.

Key words multiple sclerosis; gene chip; differentially-expressed genes; biomarkers