

· 论 著 ·

高危神经母细胞瘤差异表达基因的生信分析

郭文才 李 锐 李龙祥 吴 勇

【摘要】目的 应用生信分析方法筛选高危神经母细胞瘤的差异表达基因(DEG)。**方法** 从 GEO 数据库下载 2 个高危神经母细胞瘤数据集(GSE49710、GSE73517),筛选 DEG,应用 GO 和 KEGG 进行富集分析,构建 PPI 网络筛选中枢基因。**结果** GSE49710 包括 34 个上调 DEG 和 284 个下调 DEG, GSE73517 包括 62 个上调 DEG 和 309 个下调 DEG。GO 分析显示,生物过程主要集中在细胞粘附、GTP 酶活性的正调节、转录的负调节、DNA 模板化、凋亡过程的负调节、中枢神经系统发育,调节的细胞成分富集在膜的组成部分、细胞外区域、质膜的组成部分、细胞外空间,调节的分子功能富集于钙离子结合、受体结合。KEGG 分析显示在可卡因成瘾、造血细胞谱系、NOD 样受体信号通路、糖胺聚糖生物合成-硫酸乙酰肝素/肝素中显著富集。PPI 网络确定 5 个中枢基因(ADRB2、MC4R、CD69、RBFOX1 和 IL7R)。**结论** ADRB2、MC4R、CD69、RBFOX1 和 IL7R 可能与高危神经母细胞瘤的发生、发展相关,也可为高危神经母细胞瘤提供潜在治疗靶点。

【关键词】 神经母细胞瘤;生物信息学;生物标志物;差异表达基因

【文章编号】 1009-153X(2021)08-0612-04 **【文献标志码】** A **【中国图书资料分类号】** R 739.41; Q 786

Bioinformatic analysis of differential expression genes in high-risk neuroblastoma

GUO Wen-cai, LI Rui, LI Long-xiang, WU Yong. Department of Neurosurgery, Bazhong Central Hospital, Sichuan 636000, China

【Abstract】 Objective To analyze the differential expression genes (DEG) of high-risk neuroblastoma using biosynthesis analysis methods. **Methods** Two high-risk neuroblastoma datasets (GSE49710, GSE73517) were downloaded from GEO database, and DEG was screened. GO and KEGG analyses were used for enrichment analysis, and a PPI network was constructed to screen central hub genes. **Results** GSE49710 included 34 up-regulated DEG and 284 down-regulated DEG, GSE73517 included 62 up-regulated DEG and 309 down-regulated DEG. GO analysis showed that biological processes were mainly involved in cell adhesion, positive regulation of GTPase activity, negative regulation of transcription, DNA templating, negative regulation of apoptotic processes, development of the central nervous system, cell components were mainly involved in membranes component, extracellular area, component of plasma membrane and extracellular space, molecular functions were mainly enriched in calcium ion binding and receptor binding. KEGG analysis showed significant enrichment in cocaine addiction, hematopoietic cell lineage, NOD-like receptor signaling pathway, glycosaminoglycan biosynthesis-heparan sulfate/heparin. The PPI network identified five central hub genes, including ADRB2, MC4R, CD69, RBFOX1 and IL7R. **Conclusions** ADRB2, MC4R, CD69, RBFOX1 and IL7R may be related to the occurrence and development of high-risk neuroblastoma, and they may also provide potential therapeutic targets for high-risk neuroblastoma.

【Key words】 Neuroblastoma; Bioinformatics; Biomarkers; Differential expression gene

神经母细胞瘤占 15 岁以下儿童肿瘤的 7%,可分为低危、中危和高危^[1]。目前,中、低危神经母细胞瘤的长期生存率超过 90%^[2,3],但高危神经母细胞瘤的长期生存率仍低于 50%^[4]。研究表明,许多分子标志物与肿瘤的发生发展密切相关,可用于早期筛查^[5]。GEO 是一个国际公共数据库,为探讨肿瘤基因表达提供了基础^[6]。我们从 GEO 数据库下载 2 个神经母细胞瘤数据集(GSE49710、GSE73517),筛选高危神经母细胞瘤差异表达基因(differential

expression gene, DEG),为高危神经母细胞瘤的诊治提供参考。

1 资料与方法

1.1 基因表达和临床数据的收集 两个基因表达数据集(GSE49710、GSE73517)来自 NCBI 的 GeneExpressionSynthesis (GEO) 数据库,可从网站(<http://www.ncbi.nlm.nih.gov/geo/>)获取。GSE49710 阵列数据由 Zhang 等提交,包括 498 例神经母细胞瘤。GSE73517 数据集由 105 个神经母细胞瘤组成,由 Henrich 等提交。GSE49710 和 GSE73517 数据集基于 GPL16876 平台(Agilent-020382 Human Custom Microarray44k;发布于 2013 年 3 月 28 日)。

1.2 DEG 的鉴定 应用 Affy 软件包([doi:10.13798/j.issn.1009-153X.2021.08.012](http://cran.r-proj-</p></div><div data-bbox=)

ect.org/)实现稳健的多阵列平均算法,将原始阵列数据转换为表达式值,并进行背景校正以归一化和汇总探针。基于R语言的“limma”软件包的配对 t 检验用于分析高危和非高危神经母细胞瘤之间的DEG。调整后的 P 值 <0.05 和 $|\log_2FC|>1.5$ 被认为是DEG筛选的临界值。“Venny”是一个网络分析工具(<https://bioinfo.gp.cnb.csic.es/tools/venny/index.html>),用于分析交集EDG。

1.3 GO 和 KEGG 对 DEG 进行富集分析 GO (<http://www.geneontology.org>)为基因数据提供功能分类,包括生物过程、细胞成分和分子功能。因此,GO分析是一种广泛使用的基因和基因产物注释工具。KEGG (<http://www.genome.ad.jp/kegg/>)是一个网络网站,分析、解释和可视化基因功能。DAVID(<http://david.abcc.ncifcrf.gov/>)是一个带注释的、可视化的、全面的发现数据库,也是一种基因功能分类的在线工具,可用于评估基因的生物学功能。利用DAVID网站进行GO富集分析和KEGG通路分析,探讨DEG的功能。 $P<0.05$ 认为具有统计学意义。

1.4 PPI 网络构建与中枢基因鉴定 STRING(版本:11.0,<https://string-db.org>)用于识别相互作用的基因和蛋白质,将DEG导入其中构建PPI网络,显示物理和功能相互作用。选择总分 >0.4 的蛋白质对进行PPI网络构建。此外,Cytoscape软件(V.3.8.0)用于可视化PPI网络。使用cytoHubba中的12种方法(Betweenness、BottleNeck、Closeness、ClusteringCoefficient、Degree、DMNC、EcCentricity、EPC、MCC、MNC、Radality 和 Stress)对 hub 基因进行排序和评估,最终生成 hub 基因网络;DAVID用于GO的 hub 基因富集分析,KEGG进一步说明结果的可靠性。

2 结果

2.1 高危神经母细胞瘤的 DEG 识别 GSE49710 包括 34 个上调 DEG 和 284 个下调 DEG, GSE73517 包括 62 个上调 DEG 和 309 个下调 DEG。见图 1。

2.2 高危神经母细胞瘤 DEG 的 GO 和 KEGG 富集分析 Venny 图显示补充表, 255 个 DEG 中, 31 个上调(图 2a)和 224 个下调(图 2b)。GO 分析显示, 高危神经母细胞瘤 DEG 调节的生物过程主要集中在细胞粘附、GTP 酶活性的正调节、转录的负调节、DNA 模板化、凋亡过程的负调节、中枢神经系统发育, 调节的细胞成分富集在膜的组成部分、细胞外区域、质膜的组成部分、细胞外空间, 调节的分子功能富集于钙离子结合、受体结合(图 2c)。KEGG 分析显示高危

神经母细胞瘤 DEG 在可卡因成瘾、造血细胞谱系、NOD 样受体信号通路、糖胺聚糖生物合成-硫酸乙酰肝素/肝素中显著富集(图 2d)。

2.3 PPI 网络构建、模块分析和枢纽基因确定 STRING 分析 255 个 DEG 生成 PPI 网络见图 3。在 Cytoscape 中使用 MCODE 插件识别中心模块的基因, 对得分最高的模块(5.852)进行 GO 富集分析, 共鉴定出 5 个重要的中心模块, 得分最高的模块的 GO 分析表明其与细胞表面受体信号通路、质膜和早期内体有关(图 4a、4b)。同时, 使用 Cytoscape 中 cytohubba 插件的“Degree”算法来计算 EDG 的重要性排名, 绘制 top25DEGs 的 PPI 网络图并进行 GO 分析, 结果表明与腺苷酸环化酶调节 G 蛋白偶联受体信号通

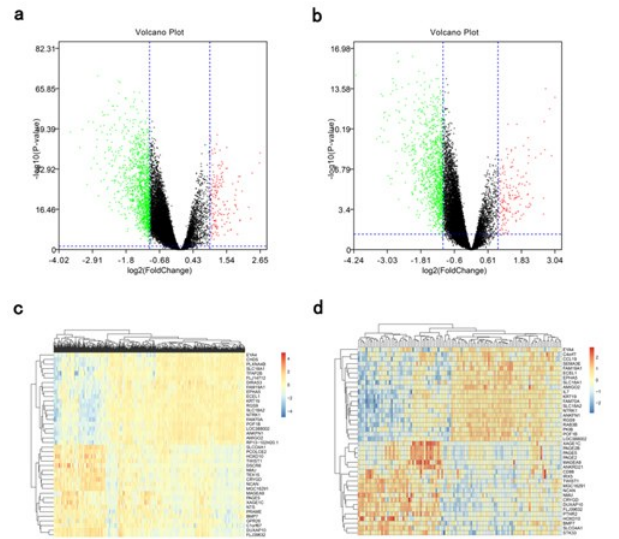


图1 高危神经母细胞瘤差异表达基因分析
a. GSE49710 火山图; b. GSE73517 火山图; c. GSE49710 基因表达热图; d. GSE73517 基因表达热图

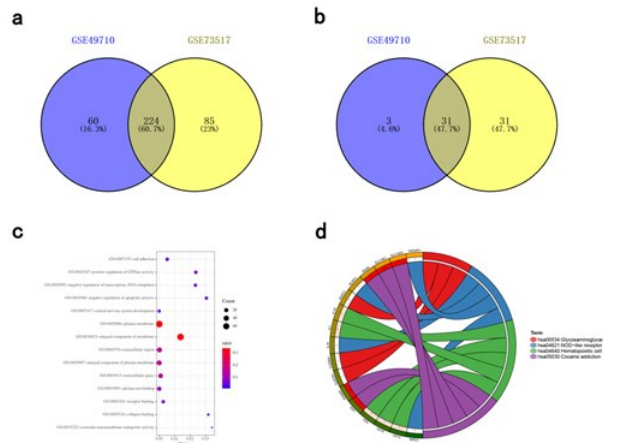


图2 高危神经母细胞瘤差异表达基因的富集分析
a. GSE49710 和 GSE73517 下调的差异表达基因; b. GSE49710 和 GSE73517 上调的差异表达基因; c. 信号通路富集 255 个差异表达基因; d. GO 富集分析 255 个差异表达基因

路、胞外区和肽激素结合密切相关(图4c、4d)。最终得到5个中枢基因为 ADRB2、MC4R、CD69、RBFOX1 和 IL7R。

3 讨论

早诊断和早治疗是延长高危神经母细胞瘤生存时间的关键。随着生物信息学的发展,DNA 微阵列越来越多地用于研究肿瘤的早期诊断、治疗和预后

评估^[7]。与单队列研究相比,多队列研究往往具有较低的假阳性和假阴性率^[8]。然而,由于批次效应和生物学差异等原因,来自不同平台的多个微阵列可能会掩盖和混淆真实情况^[9]。为了提高 DEG 鉴定的可靠性,我们在同一平台上选择了两个微阵列数据集,共鉴定出 689 个 DEG,其中 255 个 DEG 在两个数据集中有显著差异;进一步 GO 富集显示 DEG 的表达主要与生物过程有关,如细胞粘附、GTP 酶活性的正调控、转录的负调控等;分子功能主要涉及钙离子结合和受体结合;KEGG 富集分析显示 DEG 在可卡因成瘾、造血细胞谱系和 NOD 样受体信号通路中显著富集。细胞粘附的减少、细胞能量供应的增加以及某些基因表达的失衡是肿瘤发生、发展的关键步骤^[10,11]。Tajbakhsh 等^[12]指出钙离子在肿瘤的发生和发展中起着至关重要的作用。NOD 样受体在肿瘤转移和炎症性疾病中也起着至关重要的作用^[13,14]。本文筛选出的 DEG 与这些文献报道的分子功能密切相关。

为识别出的 DEG 构建 PPI 网络,并根据度级别定义关键基因,最终确定 5 个中枢基因 (DRB2、MC4R、CD69、RBFOX1 和 IL7R)。HUB 基因功能富集分析表明神经母细胞瘤的发生与环化酶调节 G 蛋白偶联受体信号、胞外区、肽激素结合有关。对 5 个 HUB 基因进行目标数据集的生存分析,筛选出具有预后意义的 ADRB2、MC4R 和 RBFOX1。ADRB2 属于由肾上腺素或去甲肾上腺素激活的跨膜 G 蛋白偶

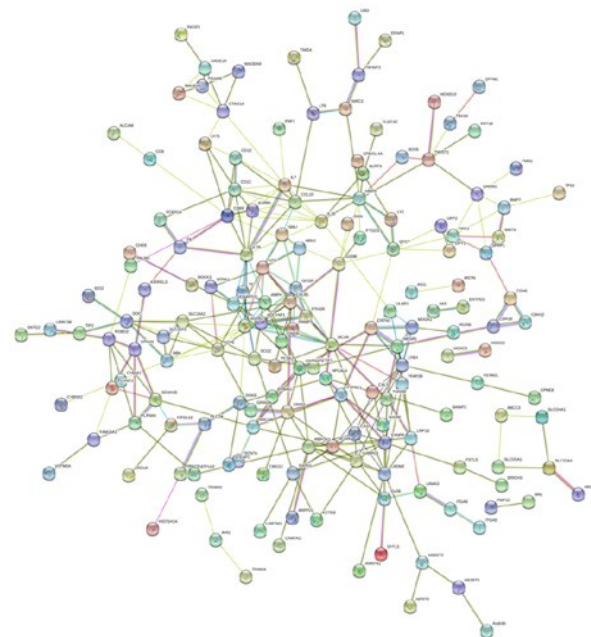


图3 高危神经母细胞瘤 255 个差异表达基因的蛋白质相互作用网络图

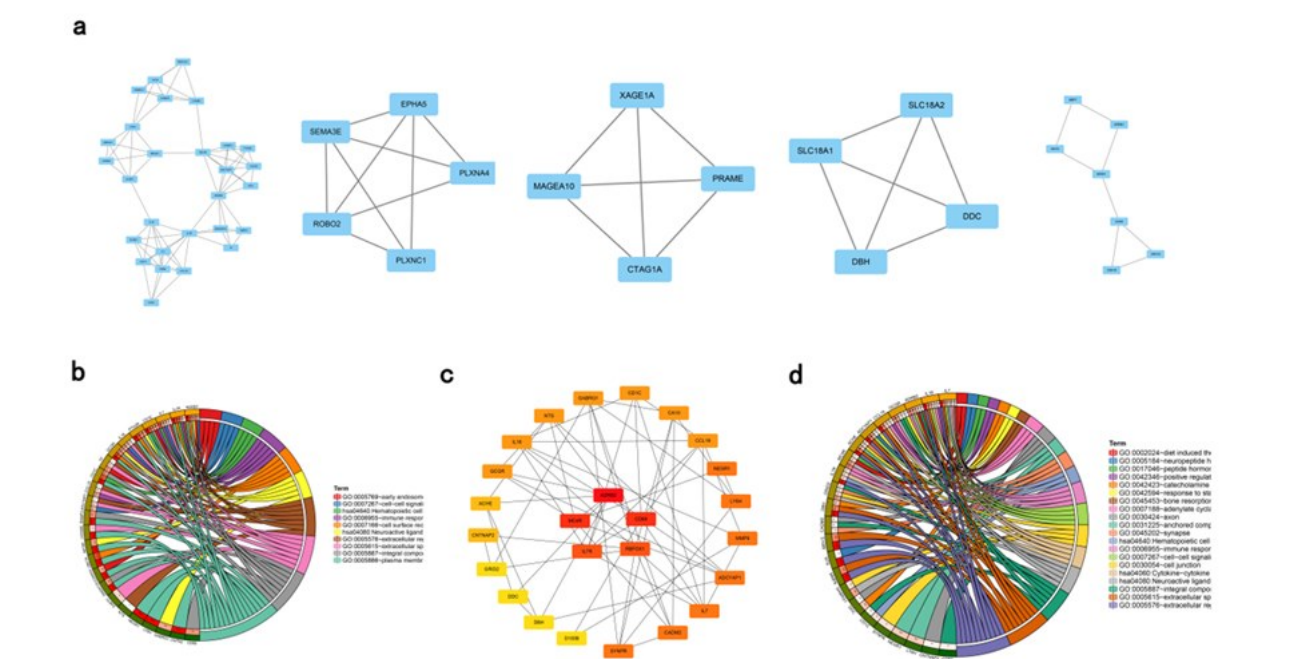


图4 高危神经母细胞瘤中枢基因的鉴定

a. MCODE 基因的拓扑图;b. 最高分模块的富集基因本体通路;c. 前 25 个基因的网络拓扑图;d. 前 25 个基因的基因本体富集通路

联受体的超家族 A^[15],与 cAMP 的产生有关。cAMP 与肿瘤的发生发展相关^[16,17]。研究表明,ADRB2 与许多肿瘤的发生、发展、进展或转移有关^[18],以及肿瘤耐药性^[19],甚至有作为分子靶点进行靶向治疗的潜力^[20]。这提示 ADRB2 基因可能是高危神经母细胞瘤潜在靶分子和早期检测和预后的潜在标志物。

总之,ADRB2、MC4R、CD69、RBFOX1 和 IL7R 为高危神经母细胞瘤的中枢基因。这些基因可能成为提高高危神经母细胞瘤诊断、优化化疗和预测预后的潜在靶点,也可能为高危神经母细胞瘤的治疗提供潜在的治疗靶点。

【参考文献】

[1] Koneru B, LopezG, Farooqi A, *et al.* Telomere maintenance mechanisms define clinical outcome in high-risk neuroblastoma [J]. *Cancer Res*, 2020, 80(12): 2663-75.

[2] Strother DR, London WB, Schmidt ML, *et al.* Outcome after surgery alone or with restricted use of chemotherapy for patients with low-risk neuroblastoma: results of Children's Oncology Group study P9641 [J]. *J Clin Oncol*, 2012, 30(15): 1842-8.

[3] Baker DL, Schmidt ML, Cohn SL, *et al.* Outcome after reduced chemotherapy for intermediate-risk neuroblastoma [J]. *N Engl J Med*, 2010, 363(14): 1313-23.

[4] Holmes K, PotschgerU, Pearson ADJ, *et al.* Influence of surgical excision on the survival of patients with stage 4 high-risk neuroblastoma: a report from the HR-NBL1/SIOPEN study [J]. *J Clin Oncol*, 2020, 38(25): 2902-2915.

[5] Pstrag N, Ziemnicka K, Bluyssen H, *et al.* Thyroid cancers of follicular origin in a genomic light: in-depth overview of common and unique molecular marker candidates [J]. *Mol Cancer*, 2018, 17(1): 116.

[6] Toro-Domínguez D, Martorell-Marugán J, López-Domínguez R, *et al.* ImaGEO: integrative gene expression meta-analysis from GEO database [J]. *Bioinformatics*, 2019, 35(5): 880-882.

[7] Xia XQ, Jia Z, Porwollik S, *et al.* Evaluating oligonucleotide properties for DNA microarray probe design [J]. *Nucleic Acids Res*, 2010, 38(11): e121.

[8] Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of

p-values [J]. *Bioinformatics*, 2003, 19(10): 1236-42.

[9] Kupfer P, Guthke R, Pohlers D, *et al.* Batch correction of microarray data substantially improves the identification of genes differentially expressed in rheumatoid arthritis and osteoarthritis [J]. *BMC Med Genomics*, 2012, 5: 23.

[10] Karstens KF, BellonE, Polonski A, *et al.* Expression and serum levels of the neural cell adhesion molecule L1-like protein (CHL1) in gastrointestinal stroma tumors (GIST) and its prognostic power [J]. *Oncotarget*, 2020, 11(13): 1131-1140.

[11] Svensmark JH, Brakebusch C. Rho GTPases in cancer: friend or foe [J]. *Oncogene*, 2019, 38(50): 7447-7456.

[12] Tajbakhsh A, Pasdar A, Rezaee M, *et al.* The current status and perspectives regarding the clinical implication of intracellular calcium in breast cancer [J]. *J Cell Physiol*, 2018, 233(8): 5623-5641.

[13] Ohashi K, Wang Z, Yang YM, *et al.* NOD-like receptor C4 inflammasome regulates the growth of colon cancer liver metastasis in NAFLD [J]. *Hepatology*, 2019, 70(5): 1582-1599.

[14] Coll RC, Robertson AA, Chae JJ, *et al.* A small-molecule inhibitor of the NLRP3 inflammasome for the treatment of inflammatory diseases [J]. *Nat Med*, 2015, 21(3): 248-255.

[15] Philipp M, Hein L. Adrenergic receptor knockout mice: distinct functions of 9 receptor subtypes [J]. *Pharmacol Ther*, 2004, 101(1): 65-74.

[16] Sapio L, Gallo M, Illiano M, *et al.* The natural camp elevating compound forskolin in cancer therapy: is it time [J]? *J Cell Physiol*, 2017, 232(5): 922-927.

[17] Berdantin LB. Diabetes and cancer: Debating the link through Ca⁽²⁺⁾/cAMP signalling [J]. *Cancer Lett*, 2019, 448: 128-131.

[18] Zhang X, Zhang Y, He Z, *et al.* Chronic stress promotes gastric cancer progression and metastasis: an essential role for ADRB2 [J]. *Cell Death Dis*, 2019, 10(11): 788.

[19] Xie WY, He RH, Zhang J, *et al.* betablockers inhibit the viability of breast cancer cells by regulating the ERK/COX2 signaling pathway and the drug response is affected by ADRB2 singlenucleotide polymorphisms [J]. *Oncol Rep*, 2019, 40(1): 341-350.

[20] Kulik G. ADRB2-targeting therapies for prostate cancer [J]. *Cancers (Basel)*, 2019, 11(3): E358.

(2021-06-18 收稿, 2021-06-26 修回)