

. 论 著 .

基于生信分析的脑胶质瘤预后模型的构建与验证

夏 天 金志勇 于如同

【摘要】目的 探讨脑胶质瘤预后模型的构建方法及预测效果。**方法** 从 CGGA 数据库获得脑胶质细胞瘤的样本数据,进行差异表达基因(DEGs)分析及加权基因共表达网络分析(WGCNA),获得关键基因,采用 Lasso-Cox 方法分析有效关键基因并计算风险评分,生存曲线分析风险评分与脑胶质瘤生存预后的关系,多因素 Cox 回归和列线图分析构建预后模型,并使用 TCGA 数据库验证。**结果** 共发现 5 463 个 DEGs,其中 4 697 个显著上调,766 个显著下调。WGCNA 分析筛选出 261 个关键基因,Lasso-Cox 分析得到 11 个有效关键基因,并计算风险评分。根据风险评分中位数将脑胶质瘤分为高、低风险组,生存曲线分析显示,高风险组脑胶质瘤病人中位生存期较低风险组明显缩短($P<0.001$)。多因素 Cox 回归分析显示风险评分是脑胶质瘤预后危险因素。利用 WHO 级别、复发状态、IDH 情况、年龄、化疗及风险评分构建预后模型预测 1、3、5 年生存预后 ROC 曲线下面积分别为 0.76、0.82、0.84,预后模型的 C-Index 值为 0.778,风险评分的 C-Index 值为 0.74。**结论** 采用 CGGA 数据库中脑胶质瘤样本数据成功构建了预后模型,为临床判断脑胶质瘤预后具有一定的价值。

【关键词】 脑胶质瘤;生信分析;差异表达基因;WGCNA;预后模型

【文章编号】 1009-153X(2023)08-0504-04

【文献标志码】 A

【中国图书资料分类号】 R 742.8; R 651.1[†]

Construction and verification of brain glioma prognostic model based on bioinformatics analysis

XIA Tian^{1,2}, JIN Zhi-yong¹, YU Ru-tong¹. 1. Department of Neurosurgery, Affiliated Hospital of Xuzhou Medical University, Xuzhou 221000, China; 2. Department of Neurosurgery, Pizhou Hospital Affiliated to Xuzhou Medical University, Pizhou 221300, China

【Abstract】 Objective To investigate the method of constructing prognostic model of brain glioma and its prediction effect. **Methods** The data of glioma patients were obtained from CGGA database, and the key genes were obtained by differentially expressed genes (DEGs) analysis and weighted gene co-expression network analysis (WGCNA). The effective key genes were analyzed by Lasso-Cox method and the risk score was calculated. The relationship between risk score and survival prognosis of glioma patients was analyzed by survival curve. The prognostic model was constructed by multivariate Cox regression and nomogram analysis, and verified by TCGA database. **Results** A total of 5 463 DEGs were found, of which 4 697 were significantly up-regulated and 766 were significantly down-regulated. A total of 261 key genes was screened out by WGCNA analysis, and 11 effective key genes were screened out by Lasso-Cox analysis and the risk score was calculated. According to the median risk score, brain glioma patients were divided into high-risk and low-risk groups. Survival curve analysis showed that the median survival time of brain gliomas patients in the high-risk group was significantly shorter than that in the low-risk group ($P<0.001$). Multivariate Cox regression analysis showed that risk score was a prognostic risk factor for glioma patients. WHO grade, relapse status, IDH status, age, chemotherapy and risk score were used to construct a prognostic model to predict the survival prognosis and the area under ROC curve of 1, 3 and 5 years survival was 0.76, 0.82 and 0.84, respectively. The C-Index value of the prognostic model was 0.778 and the C-Index value of risk score was 0.74. **Conclusions** A prognostic model is successfully constructed using glioma data from CGGA database, which has certain value for clinical judgment of glioma prognosis.

【Key words】 Glioma; Bioinformatics analysis; Differentially expressed genes; WGCNA; Prognostic model

脑胶质细胞瘤是颅内原发性恶性肿瘤之一,起源于脑神经细胞,占颅内原发性肿瘤的 50%~60%^[1]。根据 2016 年 WHO 中枢神经系统肿瘤分类,胶质细

胞瘤分为 I~IV 级,其中 I~II 级为低级别胶质细胞瘤,III~IV 级为高级别胶质细胞瘤。胶质母细胞瘤(glioblastoma multiforme, GBM)属于 WHO 分级 IV 级,恶性程度最高^[2,3]。当前胶质瘤的标准治疗措施(包括手术联合术后放疗等)取得一定的进展,但胶质瘤的预后仍不理想,GBM 中位生存期在 15 个月以下^[4],复发者的中位生存期在 3~6 个月^[5]。因此,GBM 复发是临床治疗的棘手问题。深入研究 GBM 复发的发生、发展机制,寻找其复发的分子标志物,对其治疗

doi:10.13798/j.issn.1009-153X.2023.08.008

作者单位:221000 江苏徐州,徐州医科大学附属医院神经外科/徐州医学大学神经系统疾病研究所(夏 天、金志勇、于如同);221300 江苏邳州,徐州医科大学附属医院邳州医院神经外科(夏 天)

通讯作者:于如同, E-mail: yu.rutong@163.com

有重要意义。

近年来,随着生物信息学的发展,测序及基因芯片等手段被大量应用。目前,已有 GBM 预后相关基因如 MEOX2、HMGB1 等的研究报道,然而单个基因芯片结果使用的技术平台、临床样本不同,导致现有分析结果常存在一定差异^[6]。通过生物信息学方法分析的差异表达基因(differentially expressed genes, DEGs)与加权基因共表达网络分析(weighted gene correlation network analysis, WGCNA)得到的关键基因进行整合,可以获得与 GBM 临床特征密切相关的基因,对 GBM 复发的分子机制及预后预测有重要意义。本文对中国胶质瘤基因组图谱(CGGA)数据库中已知数据进行分析,得到 DEGs,利用 WGCNA 筛选出关键基因,并与病人临床特征综合分析,得到原发与复发的特征基因,进而建立临床预后预测模型并验证。

1 资料与方法

1.1 胶质细胞瘤相关基因数据来源 基因数据来自 CGGA 数据库(<http://cgga.org.cn/index.jsp>)的 mRNA_693 数据,标准化预处理后,得到 662 个胶质细胞瘤样本,其中 405 个原发样本,257 个复发样本。

1.2 DEGs 的筛选 使用 R 软件包 limma (version 3.40.6)进行分析,获得复发与原发脑胶质细胞瘤的 DEGs。筛选标准: $P<0.01$,差异倍数 >2.0 。然后,对差异基因进行整合分析,使用 R 软件包绘制差异基因火山图。

1.3 WGCNA 分析关键基因 利用 WGCNA 软件对筛选出来的 DEGs 进行相关矩阵和平均连锁法分析,然后,使用幂函数构建加权邻接矩阵。为使网络可视化,邻接矩阵被转换为拓扑重叠矩阵(topological overlap matrix, TOM),测量基因的网络连通性,灰色模块为无法被分配给任何模块的基因集合。通过计算模块特征向量与基因的表达相关性以获得 MM。根据截止标准($|MM|>0.8$, $|GS|>0.3$),临床显著模块中具有高连接性的基因被确定为关键基因。

1.4 功能富集分析 使用 R 软件包 clusterProfiler(version 3.14.3)程序进行 GO 富集分析和 KEGG 信号通路分析,以获得基因集富集的结果。

1.5 Lasso-Cox 生存分析 将提取的关键基因使用 R 软件包 glmnet,整合生存时间、生存状态和基因表达数据,利用 Lasso-Cox 方法进行回归分析并算出相关基因的风险系数。

1.6 预后模型和列线图的评估和验证 依据风险评

分中位值将 CGGA 数据集病人分为高、低风险两组, Kaplan-Meier 曲线比较生存差异,用 R 软件包 pROC (version 1.17.0.1)绘制 ROC 曲线并计算曲线下面积(area under curve, AUC)。使用 R 软件 rms 包构建风险评分和临床特征的列线图模型,bootstrap 法重复抽样计算一致性指数(C-Index)并绘制校准曲线。使用 TCGA 数据库验证列线图模型预测的有效性。

1.7 统计学方法 使用 R 软件 4.2.1 版本进行分析, R 语言相关程序包(limma、clusterProfiler、glmnet、pROC、rms、survival 等)进行处理。Kaplan-Meier 法分析生存差异,用风险比(hazard ratio, HR)及其 95% 置信区间(confidence interval, CI)表示;C-index >0.70 具有较好的准确度; $P<0.05$ 为差异有统计学意义。

2 结果

2.1 DEGs 的筛选结果 发现 5 463 个 DEGs,其中 4 697 个显著上调,766 个显著下调。

2.2 关键基因的分析结果 将 5 463 个 DEGs 进行 WGCNA 分析,黑色和褐色模块与疾病相关性最高,其中黑色模块含 685 个基因,褐色模块含 2 009 个基因。以 $|MM|>0.8$ 和 $|GS|>0.2$ 为筛选条件,共筛选出 261 个关键基因。

2.3 功能富集分析 对 261 个关键基因 GO 和 KEGG 分析,发现这些基因主要与细胞蛋白质代谢、内质网蛋白质合成、鞘磷脂类信号通路、AMPK 信号通路等有关。

2.4 风险评分的设置 将 261 个关键基因进行 Lasso 分析,得到 20 个候选基因,然后纳入多因素 Cox 回归分析得到 11 个有效基因,包括 FAM120AOS、COMMD2、TRMT6、PRUNE、PROSC、VAMP7、CAP1、DSCR3、PGM2、NDUFS4、MRPS30。用这 11 个有效基因构建风险预后模型:风险评分 $=0.057 \times \text{FAM120AOS} + 0.042 \times \text{COMMD2} + 0.101 \times \text{TRMT6} + 0.081 \times \text{PRUNE} - 0.020 \times \text{PROSC} - 0.086 \times \text{VAMP7} + 0.008 \times \text{CAP1} - 0.031 \times \text{DSCR3} + 0.061 \times \text{PGM2} - 0.005 \times \text{NDUFS4} - 0.05 \times \text{MRPS30}$ 。

2.5 风险评分与脑胶质瘤病人预后的关系 随胶质瘤病理级别增高,其风险评分明显增高,WHO 分级 IV 级脑胶质瘤风险评分平明显高于 WHO 分级 II、III 级脑胶质瘤($P<0.001$;图 1A)。根据风险评分中位数将脑胶质瘤分为高、低风险组,生存曲线分析显示,高风险组脑胶质瘤病人中位生存期较低风险组明显缩短($P<0.001$;图 1B)。

2.6 预后模型的构建及验证 多因素 Cox 回归分析显

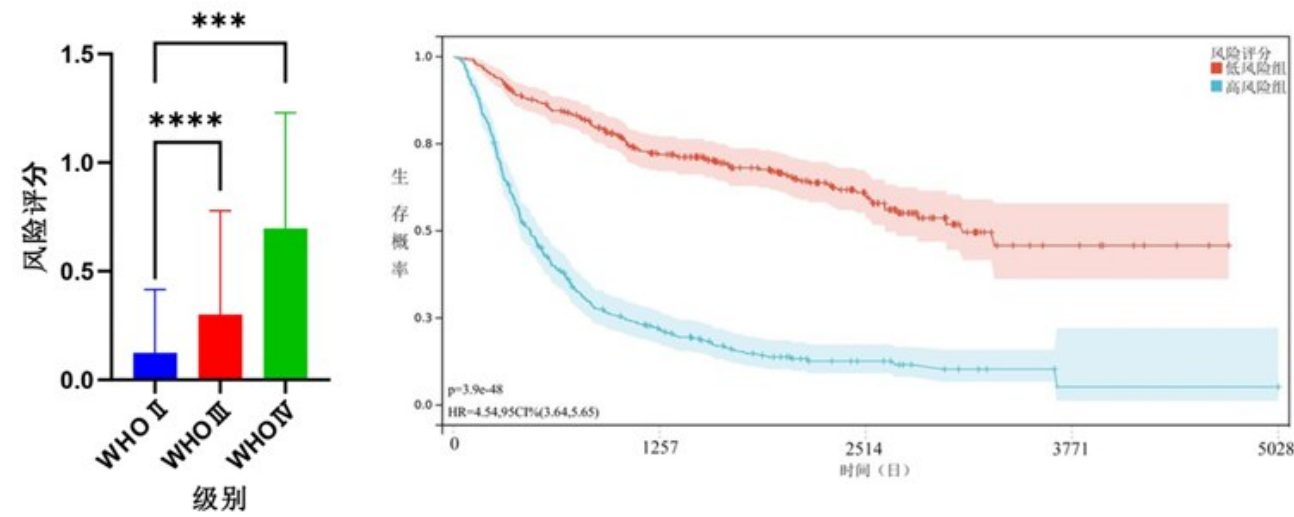


图1 风险评分与脑胶质瘤病人预后的关系
A. 风险评分与脑胶质瘤病理级别的关系,*** $P<0.001$;B. 风险评分与脑胶质瘤生存预后的关系

示病理级别、复发状态、IDH情况、年龄及风险评分是脑胶质瘤预后危险因素,而化疗是保护因素(图2A)。利用WHO级别、复发状态、IDH情况、年龄、化疗及风险评分构建列线图模型见图2B。

ROC曲线分析显示预后模型预测1、3、5年AUC分别为0.76、0.82、0.84(图3A)。用bootstrap法重复抽样1000次,预后模型的C-Index值为0.778,风险评分的C-Index值为0.74(图3B)。绘制预测1、2、3、5年校准曲线,并使用TCGA数据库验证(图3C、3D)。

3 讨论

胶质细胞瘤是最常见的颅内肿瘤之一,其治疗方法有手术、放疗、化疗、靶向治疗等,标准疗法是手术切除联合放疗和替莫唑胺化疗^[7]。但是,胶质细胞瘤的预后仍然很差,尤其是GBM,恶性程度高,治疗难度大,复发率高,生存率低。尽管已经有一些胶质细胞瘤复发预测的研究^[8],但是长期随访样本不足,能够准确预测胶质细胞瘤复发或预后的研究不多。

本文从CGGA和TCGA数据库下载胶质细胞瘤样本的基因表达数据和临床预后数据,应用R软件工具筛选出DEGs,并进一步筛选出11个有效关键基因。研究表明FAM120AOS与驱动蛋白和运动蛋白相关,参与微管运动,其介导IL13R α 2激活FAK、PI3K、AKT、mTOR信号通路,FAM120AOS在人类肿瘤细胞系和55%的人类肿瘤标本中过表达^[9]。COMMD2的表达与肿瘤干细胞、TMB、MSI、免疫细胞浸润、免疫抑制和肿瘤药物敏感性有关,在大多数肿瘤中也是强表达^[10]。TRMT6通过调节细胞周期、

PI3K/AKT、TGF- β 、MTORC1、NOTCH和MYC途径参与胶质瘤的进展^[11]。PRUNE是DHH(Asp-His-His)磷酸酯酶蛋白家族的一员,对细胞运动非常重要,并与肿瘤进展有关^[12]。PROSC的双等位基因突变与维生素B6依赖性遗传性癫痫有关^[13,14]。VAMP7是细胞毒性T淋巴细胞质膜融合机制的重要组成部分,细胞毒性T淋巴细胞通过释放细胞毒性物质来杀死肿瘤细胞,敲除VAMP7的表达显著降低T细胞的杀伤效率^[15]。CAP1也是一种肿瘤相关蛋白,非小细胞肺癌病人血清CAP1蛋白浓度显著升高,而血清CAP1蛋白浓度越高,非小细胞肺癌病人的总生存期

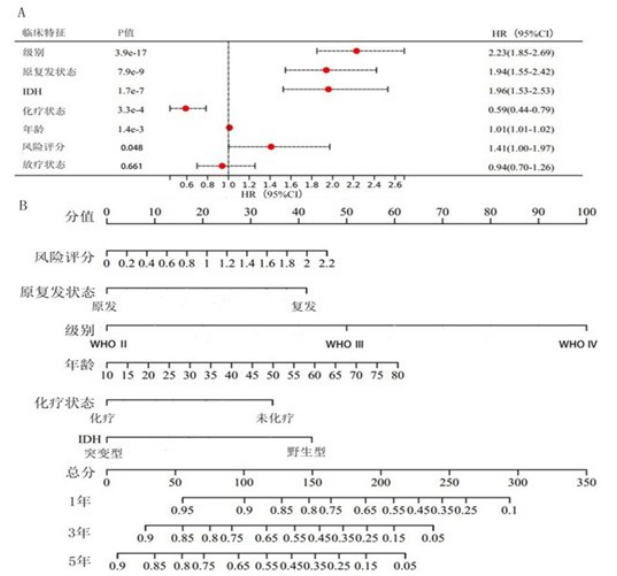


图2 多因素Cox回归和列线图模型分析风险评分与脑胶质瘤预后的关系
A. 多因素Cox回归分析风险评分与临床预后的关系;B. 列线图分析风险评分预测脑胶质细胞瘤1、3、5年生存率

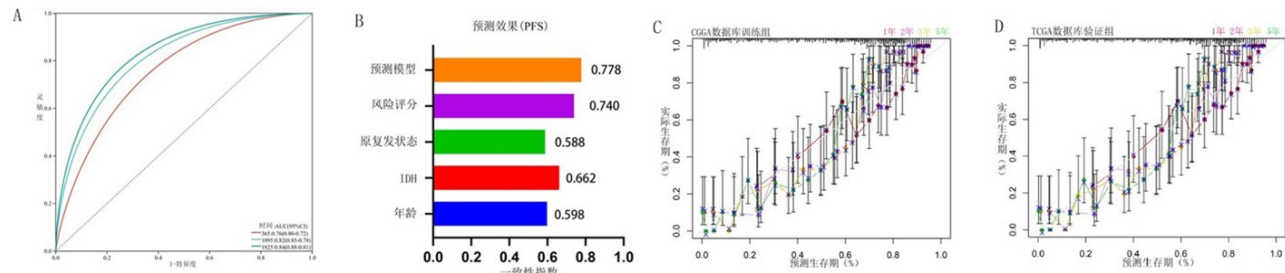


图3 ROC曲线分析预后模型预测脑胶质瘤病人预后的效果及验证

A. ROC曲线;B. 预后模型的C-Index值比较;C. CGGA数据库训练组校准图;D. TCGA数据库验证组校准图

越短,CAP1的磷酸化可以促进肿瘤细胞的增殖、迁移和侵袭^[16]。DSCR3能够增强GBM对谷氨酰胺的摄取,在复发性GBM中高表达。研究表明,沉默DSCR3的表达可增加GBM对替莫唑胺的敏感性^[17]。PGM2的过表达与肺腺癌的进展和不良预后相关,并且PGM2与低效的免疫细胞浸润相关^[18]。NDUFS4基因的失活会导致严重的神经系统疾病,是导致Leigh综合征的基因,通过ERK磷酸化调节突触素表达参与神经可塑性,其缺失导致线粒体功能障碍,突触素表达减少^[19,20]。MRPS30高表达与乳腺癌病人不良预后呈正相关,敲低MRPS30表达能明显抑制癌细胞增殖和侵袭,诱导乳腺癌细胞凋亡^[21]。

本文筛选出的11个有效关键基因中,FAM120AOS、COMMD2、TRMT6、PRUNE、VAMP7、CAP1、PGM2等6个关键基因为高危基因,在肿瘤细胞中高表达,且表达越高,总生存越短;而PROSC、VAMP7、DSCR3、NDUFS4、MRPS30等5个有效关键基因是胶质瘤复发的保护因素。本文利用这11个有效关键基因构建风险预后模型,再根据风险评分将脑胶质瘤分为高、低风险组,高风险组脑胶质瘤病人中位生存期较低风险组明显缩短($P<0.001$);多因素Cox回归分析显示风险评分是脑胶质瘤病人生存预后的独立危险因素,包含风险评分的预后模型对脑胶质瘤病人1、3、5年生存预后具有良好的预测价值。但是这些关键基因在脑胶质瘤复发中的分子机制仍需要进一步探究。

【参考文献】

[1] 盛博,陈炜. 脑胶质瘤相关性癫痫治疗的研究进展[J]. 医学综述,2021,27(8):1528-1533.

[2] 程诗琦. CCL8在脑胶质瘤中表达的临床意义及相关机制研究[D]. 郑州大学,2020.

[3] 广东中科清紫医疗科技有限公司. 免疫相关基因在预测

弥漫性胶质瘤预后的试剂盒和系统中的应用:CN202110681350.9[P]. 2021-10-8.

[4] 郭兆刚,栾涛,邓超,等. 胶质瘤中TRP基因家族预后模型的开发和验证[J]. 武警医学,2022,33(8):675-680.

[5] 林艺,王策,康勋,等. 胶质母细胞瘤复发相关基因的筛选、表达和临床预后分析[J]. 中国肿瘤杂志,2022,32(1):23-33.

[6] 黄晓珊,曾慧,周碧兰,等. 整合生物信息学方法与加权基因共表达网络分析联合分析脑胶质瘤特征基因[J]. 中国现代应用药学,2020,37(19):2311-2316.

[7] 连露露,范小璇,赵晓平,等. 复发性脑胶质瘤的临床治疗进展[J]. 国际神经病学神经外科学杂志,2022,49(2):84-88.

[8] LI GZ, WU F, ZENG F, *et al.* A novel DNA repair-related nomogram predicts survival in low-grade gliomas [J]. CNS Neurosci Ther, 2020, 27: 186-195.

[9] BARTOLOMÉ RA, GARCÍA-PALMERO I, TORRES S, *et al.* IL13 receptor $\alpha 2$ signaling requires a scaffold protein, fam120a, to activate the FAK and PI3K pathways in colon cancer metastasis [J]. Cancer Res, 2015, 75(12): 2434-2444.

[10] TAI PP, WANG ZW, CHEN XY, *et al.* Multi-omics analysis of the oncogenic value of copper metabolism-related protein COMMD2 in human cancers [J]. Cancer Med, 2022, 12(10): 11941-11959.

[11] WANG BB, NIU LH, WANG ZY, *et al.* RNA m1A methyltransferase TRMT6 predicts poorer prognosis and promotes malignant behavior in glioma [J]. Front Mol Biosci, 2021, 8: 692130.

[12] ZOLLO M, AHMED M, FERRUCCI V, *et al.* PRUNE is crucial for normal brain development and mutated in microcephaly with neurodevelopmental impairment [J]. Brain, 2017, 140(4): 940-952.

(下转第512页)

neurological function through downregulating NOX2-ROS in severe traumatic brain injury [J]. *Drug Des Devel Ther*, 2018, 12: 3847-3854.

[13] 曾子桓, 张 灏, 陈伟强, 等. 颅脑损伤后继发性脑损伤发病机制的研究进展 [J]. *中国临床神经外科杂志*, 2019, 24 (12): 777-779.

[14] LIN X, XU Z, WANG P, *et al*. Role of PiCCO monitoring for the integrated management of neurogenic pulmonary edema following traumatic brain injury: a case report and literature review [J]. *Exp Ther Med*, 2016, 12(4): 2341-2347.

[15] LIU Y, LU J, WANG X, *et al*. Erythropoietin-derived peptide protects against acute lung injury after rat traumatic brain injury [J]. *Cell Physiol Biochem*, 2017, 41(5): 2037-2044.

[16] EASTMAN CL, D'AABROSIO R, GANESH T. Modulating neuroinflammation and oxidative stress to prevent epilepsy and improve outcomes after traumatic brain injury [J]. *Neuropharmacology*, 2020, 172: 107907.

[17] ISMAIL H, SHAKKOUR Z, TABET M, *et al*. Traumatic brain injury: oxidative stress and novel anti-oxidants such as mitoquinone and edaravone [J]. *Antioxidants (Basel)*, 2020, 9(10): 943.

[18] WANG H, ZHOU X, WU LY, *et al*. Aucubin alleviates oxidative stress and inflammation via Nrf2-mediated signaling activity in experimental traumatic brain injury [J]. *J Neuroinflammation*, 2020, 17(1): 188.

[19] BAECKER J, WARTCHOW K, SEHM T, *et al*. Treatment with the neurotrophic protein S100B increases synaptogenesis after traumatic brain injury [J]. *J Neurotrauma*, 2020, 37(8): 1097-1107.

[20] Jones C, Harmon C, Mccann M, *et al*. S100B outperforms clinical decision rules for the identification of intracranial injury on head CT scan after mild traumatic brain injury [J]. *Brain Inj*, 2020, 34(3): 407-414.

[21] 尹文国, 翁山山, 赖仕宇, 等. 联合 GCS 评分、CT 评分与血清 S100B 蛋白可评估急性颅脑创伤病人损伤程度及早期预后 [J]. *南方医科大学学报*, 2021, 41(4): 543-548.

[22] Park DW, Park SH, Hwang SK. Serial measurement of S100B and NSE in pediatric traumatic brain injury [J]. *Childs Nerv Syst*, 2019, 35(2): 343-348.

[23] Schmidt AP, Carmona M. Serum biomarkers of brain injury: S100B protein, cognitive dysfunction, and major noncardiac surgery [J]. *Braz J Anesthesiol*, 2020, 70(6): 570-572.

[25] Kiiski H, Langsjo J, Tenhunen J, *et al*. S100B, NSE and MMP-9 fail to predict neurologic outcome while elevated S100B associates with milder initial clinical presentation after aneurysmal subarachnoid hemorrhage [J]. *J Neurol Sci*, 2018, 390: 129-134.

(2022-10-29 收稿, 2023-04-04 修回)

(上接第 507 页)

[13] TREMIÑO L, FORCADA-NADAL A, RUBIO V. Insight into vitamin B6-dependent epilepsy due to PLPBP (previously PROSC) missense mutations [J]. *Hum Mut*, 2018, 39(7): 1002-1013.

[14] JENSEN KV, FRID M, STÖDBERG T, *et al*. Diagnostic pitfalls in vitamin B6-dependent epilepsy caused by mutations in the PLPBP gene [J]. *JIMD Rep*, 2019, 50(1): 1-8.

[15] CHITIRALA P, RAVICHANDRAN K, GALGANO D, *et al*. Cytotoxic granule exocytosis from human cytotoxic T lymphocytes is mediated by VAMP7 [J]. *Front Immunol*, 2019, 10(1): 1855.

[16] ZENG J, LI X, LIANG L, *et al*. Phosphorylation of CAP1 regulates lung cancer proliferation, migration, and invasion [J]. *J Cancer Res Clin Oncol*, 2021, 148(1): 137-153.

[17] LIN R, XU YM, XIE SD, *et al*. Recycling of SLC38A1 to the plasma membrane by DSCR3 promotes acquired temozolomide resistance in glioblastoma [J]. *J Neurooncol*, 2022, 157(1): 15-26.

[18] YU HD, YU Z, QIN C, *et al*. High expression of glycolysis-related PGM2 gene in relation to poor prognosis and deficient immune cells infiltration in lung adenocarcinoma: a study based on bioinformatics analysis [J]. *J Thorac Dis*, 2022, 14(9): 3488-3499.

[19] SHIL SK, KAGAWA Y, UMARU BA, *et al*. Ndufs4 ablation decreases synaptophysin expression in hippocampus [J]. *Sci Rep*, 2021, 11(1): 10969.

[20] KAHLHÖFER F, GANSEN M, ZICKERMANN V. Accessory subunits of the matrix arm of mitochondrial complex I with a focus on subunit NDUFS4 and its role in complex I function and assembly [J]. *Life (Basel)*, 2021, 11(5): 455.

[21] WU BL, PAN YB, LIU GH, *et al*. MRPS30-DT knockdown inhibits breast cancer progression by targeting Jab1/Cops5 [J]. *Front Oncol*, 2019, 9: 1170.

(2023-03-22 收稿, 2023-07-26 修回)