中枢性 PNET 患者基因表达谱变化及生物信息学分析

汪　颖　李　朋　王　博　刘丕楠

摘　要　目的　依据基因表达芯片数据筛选原始神经外胚层肿瘤(primitive neurotodermal tumor,PNET)患者肿瘤组织的差异表达基因及关键蛋白,为其临床治疗提供新的理论依据。**方法**　从基因芯片公共数据库(Gene Expression Omnibus,GEO)获取 GSE74195 基因芯片数据集,筛选 PNET 肿瘤组织与正常小脑组织的差异表达基因,同时进行功能注释(GO)、通路分析(KEGG)及蛋白互作网络分析。**结果**　筛选出 340 个表达显著上调的差异基因(FC > 3, $P < 0.01$),经功能注释和通路分析,发现这些基因富集于细胞周期、有丝分裂及细胞增殖等过程,KEGG 通路富集显示主要集中于细胞周期和 ECM 受体相互作用等通路。经蛋白质相互作用网络构建,筛选出 5 个关键蛋白分子,即 UBA52、UBC、CDK1、CENPK 及 NDE1。**结论**　筛选出 PNET 肿瘤中高表达基因及关键蛋白,为 PNET 发病机制及潜在的治疗靶点研究提供新思路。

关键词　原始神经外胚层肿瘤　差异基因　生物信息学

中图分类号　R73　　**文献标识码**　A　　**DOI**　10.11969/j.issn.1673-548X.2020.03.008

Changes in Gene Expression Profiles and Bioinformatics in Central Primitive Neurotodermal Tumor Patients.　*Wang Ying, Li Peng, Wang Bo, et al. Beijing Neurosurgical Institute, Capital Medical University, Beijing 100050, China*

Abstract　Objective　To screen the differential expression genes and key proteins of gene expression microarray data in primitive neurotodermal tumor patients, to provide a new theoretical basis for clinical treatment. **Methods**　Microarray data set GSE74195 was obtained from GEO to screen differentially expressed genes between PNET tumors and normal cerebellum tissues. Functional annotation (GO), pathway analysis (KEGG) and protein interaction network analysis were also performed. **Results**　A total of 340 differentially expressed genes (FC > 3, $P < 0.01$) were screened. Functional annotation and pathway analysis showed that these genes were mainly concentrated in cell cycle, mitosis and cell proliferation. KEGG pathway enrichment showed that they were mainly concentrated in cell cycle and ECM receptor interaction. Five key protein molecules, namely UBA52, UBC, CDK1, CENPK and NDE1, were screened through the construction of protein interaction network. **Conclusion**　Screening of genes and key protein molecules with high expression in PNET tumors provides a new way to study the pathogenesis and potential therapeutic targets of PNET.

Key words　Primitive neurotodermal tumor; Differential expression genes; Bioinformatics

原始神经外胚层肿瘤(primitive neurotodermal tumor,PNET)为一种罕见的高度恶性的神经系统肿瘤,由 Hart 等[1]于 1973 年首次报道,分为中枢性和外周性两大类[2]。其中中枢性 PNET 多发于儿童,肿瘤呈浸润性增长,且侵袭性强、恶性程度高、预后差[3]。现阶段中枢性 PNET 治疗以手术切除加辅助放化疗为主,总体疗效较差,且发病原因不明[4]。

生物学中的基因芯片技术可用于肿瘤分子相关途径的分类、鉴定和预测,并可识别与临床相关的肿瘤分子标志物及潜在治疗靶点,从而可对疾病的临床诊断及治疗起提示作用[5,6]。

本研究通过对 GEO 数据库中下载的中枢性 PNET 肿瘤组织基因芯片数据集 GSE74195 进行综合生物信息学分析,探讨基因表达改变后,中枢性 PNET 分子生物学功能的改变及影响该过程的关键蛋白,以期进一步探讨中枢 PNET 疾病发生、发展相关分子机制,为中枢 PNET 的临床治疗提供新思路。

材料与方法

1. 材料:登录美国基因表达汇编(Gene Expression Omnibus, GEO)数据库,检索 PNET 研究相关的基因表达芯片数据,获取芯片分析平台为 GPL570 (HG - U133 - Plus - 2) Affymetrix 表达谱芯片数据集 GSE74195。GSE74195 包含 5 例中枢 PNET 肿瘤标本和 5 例正常小脑组织对照标本。该研究中所获得的肿瘤标本为术后立即至于液氮中,并一直在 -80℃保存直至使用,正常组织标本获得于没有脑肿瘤疾病史的患者。

基金项目:国家自然科学基金青年科学基金资助项目(81502453)

作者单位:100070　北京市神经外科研究所(汪颖、刘丕楠);100070　首都医科大学附属北京天坛医院(李朋、王博、刘丕楠)

通讯作者:刘丕楠,电子信箱:liupinan@yeah.net

2. 差异表达基因筛选：使用在线分析工具 GEO2R（https://www.ncbi.nlm.nih.gov/geo/geo2r/）对数据集 GSE74195 进行处理，以条件 FC > 3 且 $P < 0.01$ 进行差异基因筛选，获得 PNET 肿瘤组织中显著高表达的差异基因列表。

3. GO 富集分析和 KEGG 通路分析：将上述所获得的显著差异表达基因列表导入在线生物信息数据库 WebGestalt，进行 GO 功能分析及 KEGG 通路分析，以 FDR < 0.05 为筛选条件。

4. 可能作用药物筛选：将上述所获得的显著差异表达基因列表导入使用在线生物信息数据库 WebGestalt 进行作用药物筛选，以 FDR < 0.05 为筛选条件。

5. 蛋白质的相互作用网络建设：使用 String 在线分析网站（https://string-db.org/，version 11.0）对上述获得的显著差异基因列表进行分析，设置可信度 0.7（confidence：0.7）为差异有统计学意义，选择实验或数据库支持进行筛选条件限制，分析得到相关蛋白质的相互作用网络图，根据差异基因与其相互作用蛋白的连线数量进一步确定蛋白质相互作用网络中的关键蛋白。

结　果

1. 显著高表达差异基因的获得：依据 FC > 3 且 $P < 0.01$ 的筛选条件对所获得的差异表达基因进行筛选，结果获得 340 个相对于小脑正常组织，中枢性 PNET 肿瘤组织中显著高表达的差异基因。差异基因表达谱宏观表象见图 1。

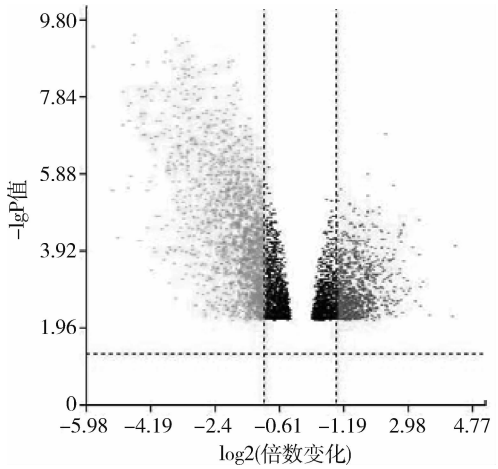

图 1　差异基因火山图

2. 差异基因功能通路富集分析：将上述筛选获得的 340 个显著上调的差异基因映射到 GO 数据库。GO 生物过程分析发现差异基因主要富集在细胞有丝分裂、细胞周期及细胞增殖等生物学过程；分子功能包括二磷酸多糖蛋白糖转移酶活性、低聚糖基转移酶活性、rRNA 结合、细胞外基质结构成分、核糖体的结构成分等的功能；细胞组成分析显示这些差异基因集中在低聚糖基转移酶复合物、浓缩染色体着丝粒区及染色体区。GO 分析结果见表 1。KEGG 通路分析显示，差异基因富集的信号通路有细胞周期和 ECM 受体相互作用等，详见图 2。

3. 可能作用药物筛选：将显著高表达的差异基因输入 WebGestalt 在线数据库，对可能有效的药物进行筛选，以 FDR < 0.05 为筛选条件。数据库选择 GLAD4U 时，按富集程度差异，获得前 10 个对 PNET 可能有效的药物候选，分别为紫杉烷类（taxanes）、紫杉醇（paclitaxel）、胃蛋白酶（pepsin）、奈替米星（netilmicin）、顺铂（cisplatin）、铂化合物（platinum compounds）、放线菌素类（actinomycines）、达克汀霉素（dactinomycin）、抗肿瘤药（antineoplastic agents）及蛋白激酶抑制剂（protein kinase inhibitor），结果如图 3 所示。数据库选择 DrugBank 时，青蒿醇（artenimol）为唯一候选药物（富集比例 5.5749）。

4. 蛋白质－蛋白质相互作用网络构建：蛋白质与蛋白质之间相互作用网络的构建，可进一步明确各差异表达基因或蛋白之间的相互作用关系。基于此，本研究将上述获得的 340 个显著上调的差异表达基因上传到 String 在线工具，经分析获得差异蛋白互作网络图，其中包括 305 个圆点（代表差异基因或蛋白），938 条线（代表相互作用关系）。根据差异基因与其相互作用蛋白的连线数量，确定了 5 个关键蛋白分子。

讨　论

临床研究表明肿瘤的遗传学特征可影响其生物学行为、治疗及预后，这些研究从分子层面揭示了肿瘤发生、发展的可能原因，从而为临床诊断及治疗奠定基础[7]。而现阶段中枢性 PNET 病因不明且发病率低，临床治疗难度大[8]。为此，本研究通过生物信息学技术挖掘相关数据，以期获得中枢性 PNET 肿瘤潜在治疗靶点，为进一步实验研究提供理论支持。

本研究共分析获得 340 个显著高表达基因，通过对这些差异基因进行 GO 分析及 KEGG 通路筛选，发现上调基因主要参与细胞周期过程，利用数据库筛选可能有效的药物，发现除作用于细胞周期的紫杉烷类、铂化合物及常见抗肿瘤药外，分析结果提示青蒿醇也可能对中枢性 PNET 肿瘤有一定效果，而此前未见有报道，可能对今后中枢性 PNET 肿瘤的治疗有一定提示作用。

表 1　PNET 高表达差异基因 GO 富集分析结果

基因功能	GO 编号	GO 注释	基因数量	FDR 值	富集比例
生物学过程	GO:0140014	有丝核分裂	278	4.601×10^{-11}	5.628
生物学过程	GO:0044772	有丝分裂细胞周期相变	486	0	4.774
生物学过程	GO:0044770	细胞周期相变	524	4.02×10^{-13}	4.428
生物学过程	GO:0051301	细胞分裂	576	1.508×10^{-12}	4.122
生物学过程	GO:1903047	有丝分裂细胞周期过程	780	0	4.012
生物学过程	GO:0007346	有丝分裂细胞周期的调控	572	2.11×10^{-11}	3.962
生物学过程	GO:0000278	有丝分裂细胞周期	916	0	3.829
生物学过程	GO:0022402	细胞周期过程	1255	1.723×10^{-12}	2.881
生物学过程	GO:0007049	细胞周期	1710	0	2.619
生物学过程	GO:0008283	细胞增殖	1953	1.003×10^{-10}	2.293
细胞组分	GO:0008250	低聚糖基转移酶复合物	13	4.61×10^{-8}	33.639
细胞组分	GO:0000779	浓缩染色体,着丝粒区	117	9.40×10^{-9}	8.543
细胞组分	GO:0098687	染色体区	329	1.55×10^{-8}	4.747
细胞组分	GO:0030055	细胞基质连接	409	2.51×10^{-10}	4.735
细胞组分	GO:0005925	黏着斑	402	5.47×10^{-10}	4.662
细胞组分	GO:0005924	细胞基质黏附结	405	5.47×10^{-10}	4.628
细胞组分	GO:0044427	染色体部分	869	5.47×10^{-10}	3.235
细胞组分	GO:0005694	染色体	997	2.26×10^{-10}	3.133
细胞组分	GO:1990904	核糖核蛋白复合物	825	1.34×10^{-8}	3.105
细胞组分	GO:0070062	胞外外泌体	2000	7.81×10^{-13}	2.530
分子功能	GO:0004579	二磷酸多糖蛋白糖转移酶活性	9	0.0001	33.627
分子功能	GO:0004576	低聚糖基转移酶活性	10	0.0001	30.264
分子功能	GO:0019843	rRNA 连接	64	0.0002	8.512
分子功能	GO:0005201	细胞外基质结构成分	157	0	5.783
分子功能	GO:0003735	核糖体的结构成分	154	0.0001	5.503
分子功能	GO:0008022	protein C 末端连接	187	0.0005	4.532
分子功能	GO:0019900	激酶连接	701	0.0001	2.763
分子功能	GO:0005198	结构分子活性	757	0.0001	2.719
分子功能	GO:0003723	RNA 连接	1581	1.17×10^{-7}	2.374
分子功能	GO:0044877	含蛋白复合物连接	1053	0.0003	2.242

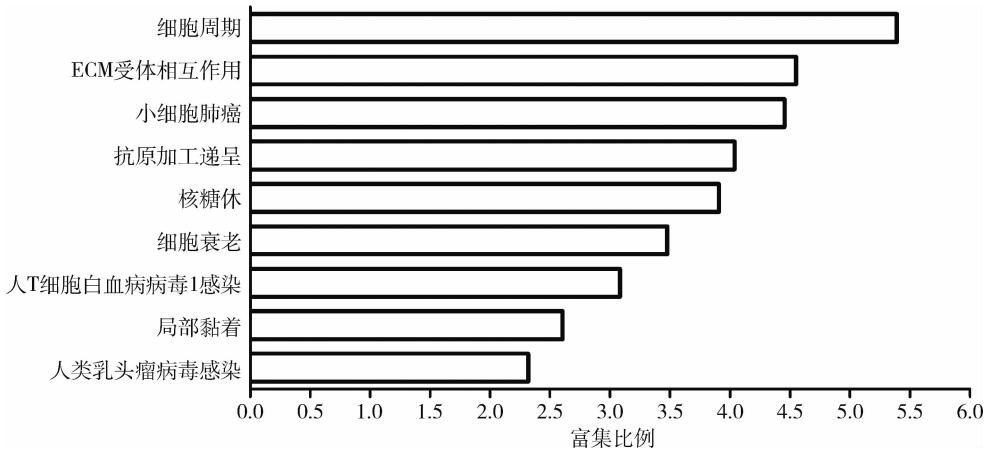

图 2　差异表达基因通路分析

进一步构建的蛋白互作网络图中,本研究获取了 5 个关键蛋白分子,分别是泛素核糖体融合蛋白 52(UBA52,连线 43)、泛素 C(UBC,连线 35)、周期蛋白依赖性激酶 1(CDK1,连线 31)、着丝粒蛋白 K(CENPK,连线 29)、神经发育蛋白 1(NDE1,连线 19)。这些蛋白可能在中枢性 PNET 肿瘤的发生、发展中发挥重要作用。蛋白互作网络图见图 4。

研究表明,肿瘤是一种细胞周期性疾病,恶性肿

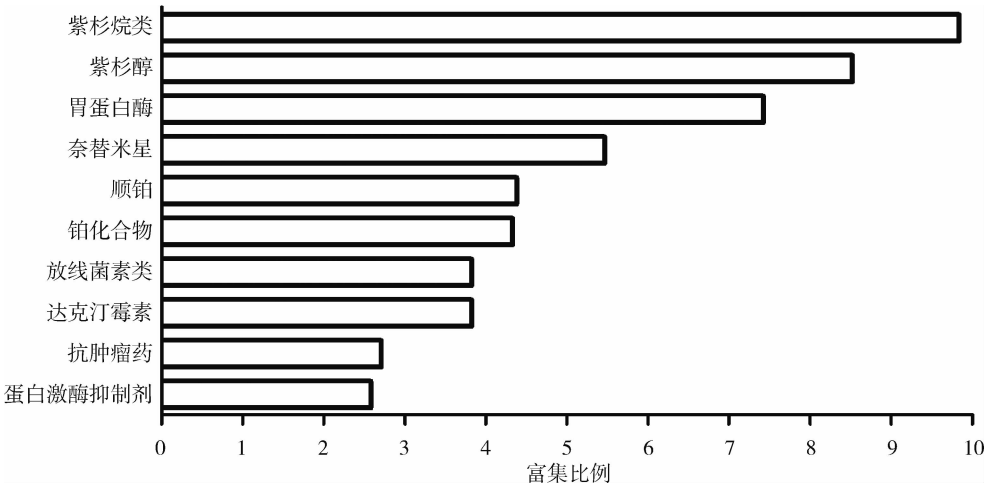

图 3　差异基因可能有效药物筛选

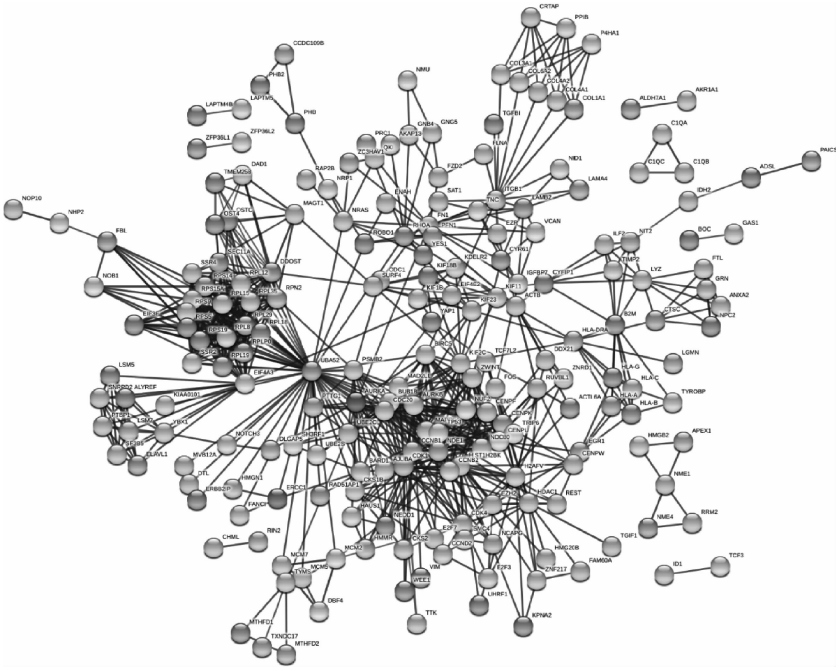

图 4　差异表达基因/蛋白互作网络图

瘤的发生、发展与细胞周期调节机制的紊乱有密切联系[9]。而笔者获取的 5 个关键蛋白，都参与了细胞周期的调控过程，其中 CENPK 蛋白由 269 个氨基酸组成，编码基因位于染色体 5q12.3 上，是着丝粒蛋白家族成员之一。而报道显示，着丝粒调节失调或功能紊乱可促进肿瘤的发生[10]。CENPK 主要定位于着丝粒的内层，可对 CENPA 与其他着丝粒组分的有效组装进行调控，且在卵巢癌、肝癌等高表达，并与预后相关，可能扮演着致癌基因的角色[11~13]。提示 CENPK 在调控着丝粒的组装和功能方面起重要作用，进而影响肿瘤的发生、发展。但目前关于 CENPK 的研究较少，调控机制尚不清楚。靶向调控 PNET 中 CENPK 的表达可能为 PNET 治疗提供新的策略，但需进一步扩大样本量进一步验证 CENPK 在中枢性 PNET 组织中的表达，在蛋白水平上探讨其与中枢性 PNET 临床病理特征及预后的关系。

此外，UBA52 是由泛素和核糖体蛋白 L40 组成的融合蛋白，在体内可以被泛素羧基端水解酶快速精准的剪切为泛素分子和核糖体蛋白 L40。泛素可形成多聚泛素化链对底物修饰，进一步通过泛素蛋白酶体途径特异性降解细胞周期相关蛋白，从而阻滞细胞周期。

细胞周期蛋白依赖性激酶 1（cyclin - dependent kinases 1，CDK1）为由 CDC2 基因编码，相对分子质量为 3.4×10^4 的蛋白质，是丝氨酸/苏氨酸蛋白激酶家族的成员之一。CDK1 可与 cyclinB 组成复合物，可促使细胞周期从 G_2 期进入 M 期，从而影响细胞有丝分裂过程[14]。NDE1 蛋白位于中心体，作为调节动力蛋白功能的多蛋白复合物的一部分与其他中心体成分相互作用[15]。文献报道显示该蛋白对大脑皮质发育至关重要，在脑增殖室区皮质神经元祖细胞分裂过程中，可通过控制有丝分裂纺锤体的方向来调节神经元的产生。由此可见，这几个差异蛋白主要集中在细胞周期及有丝分裂过程，与中枢性 PNET 发生、发展有密切关系，而这种机制是否参与了 PNET 的发病尚待进一步的研究。

本研究采用生物信息学研究方法筛选出中枢性 PNET 肿瘤组织相关的差异表达基因，并对其进行了功能预测，为寻找与中枢性 PNET 发生、发展及治疗相关靶点的研究提供了新思路。

参考文献

1 Hart MN, Earle KM. Primitive neuroectodermal tumors of the brain in children[J]. Cancer, 1973, 32(4):890 - 897

2 Vermeulen JF, van Hecke W, Spliet WG, *et al*. Pediatric primitive neuroectodermal tumors of the central nervous system differentially express granzyme inhibitors[J]. PLoS One, 2016, 11(3):e0151465

3 沈捷，袁燕，胡克，等. 原始神经外胚层肿瘤预后因素分析[J]. 中华肿瘤杂志，2018，40(8)：614 - 618

4 冯梅，谭志博，付彬玉，等. 中枢型和外周型原始神经外胚层肿瘤的临床特征及预后因素分析[J]. 中国肿瘤临床与康复，2014，21(9):1097 - 1099

5 孙娜，范颂，郭亚收，等. HIV 相关神经认知障碍患者脑组织基因表达谱的变化及生物信息学分析[J]. 中华疾病控制杂志，2017，(12)：1277 - 1281

6 郭亚收，徐晓峰，李楠，等. 高危前列腺癌患者基因表达谱的变化及生物信息学分析[J]. 现代肿瘤医学，2019，27(8):1281 - 1284

7 李文一，周俊林. 原始神经外胚层肿瘤的研究进展[J]. 国际医学放射学杂志，2014，37(3):221 - 224

8 李培培，申淑景，王晔，等. 原始神经外胚层肿瘤 52 例的临床特征及预后[J]. 实用医学杂志，2017，33(15):2554 - 2557

9 张百红. 肿瘤的细胞周期治疗[J]. 中国肿瘤临床，2018，45(6)：316 - 319

10 Scully R. The spindle - assembly checkpoint, aneuploidy, and gastrointestinal cancer[J]. N Engl J Med, 2010; 363: 2665 - 2666

11 Okada M, Cheeseman IM, Hori T, *et al*. The CENP - H - I complex is required for the efficient incorporation of newly synthesized CENP - A into centromeres[J]. Nat Cell Biol, 2006, 8: 446 - 457

12 Lee YC, Huang CC, Lin DY, *et al*. Overexpression of centromere protein K (CENPK) in ovarian cancer is correlated with poor patient survival and associated with predictive and prognostic relevance [J]. Peer J, 2015; 3: e1386

13 Wang H, Liu W, Liu L, *et al*. Overexpression of centromere protein K (CENP - K) gene in hepatocellular carcinoma promote cell proliferation by activating AKT/TP53 signal pathway[J]. Oncotarget, 2017, 8: 73994 - 74005

14 石静，马洪，段晓峰. CDK1 在口腔鳞状细胞癌中的表达及意义[J]. 贵州医科大学学报，2019，44(1):64 - 67

15 Shimizu S, Ishino Y, Tohyama M, *et al*. NDE1 positively regulates oligodendrocyte morphological differentiation[J]. Sci Rep, 2018, 8(1):7644

（收稿日期：2019 - 07 - 08）
（修回日期：2019 - 07 - 10）

（接第 114 页）

18 Uvuz F, Kilic S, Yilmaz N, *et al*. Relationship between preterm labor and thrombophilic gene polymorphism: A prospective sequential cohort study [J]. Gynecol Obstet Invest, 2009, 68(4): 234 - 238

19 Hiltunen LM, Laivuori H, Rautanen A, *et al*. Factor V Leiden as a risk factor for preterm birth - a population - based nested case - control study [J]. J Thromb Haemost, 2011, 9(1): 71 - 78

20 Lykke JA, Bare LA, Olsen J, *et al*. Thrombophilias and adverse pregnancy outcomes: Results from the Danish National Birth Cohort [J]. J Thromb Haemost, 2012, 10(7): 1320 - 1325

21 Tiwari D, Bose PD, Das S, *et al*. MTHFR (C677T) polymorphism and PR (PROGINS) mutation as genetic factors for preterm delivery, fetal death and low birth weight: a Northeast Indian population based study [J]. Meta Gene, 2015, 31(3): 31 - 42

22 Marseglia LM, Nicotera A, Salpietro V, *et al*. Hyperhomocysteinemia and MTHFR polymorphisms as antenatal risk factors of white matter abnormalities in two cohorts of late preterm and full term newborns [J]. Oxid Med Cell Longev, 2015, 2015: 1 - 8

23 Yalda FZ, Omid K, Dormohammad KT, *et al*. Investigation of the association between C677T polymorphism of the MTHFR gene and plasma homocysteine level in recurrent fetal miscarriage [J]. Obstet Gynaecol Res, 2019, 45(8):1442 - 1447

24 骆婵珍. 母亲高同型半胱氨酸血症对胎儿生长发育的影响 [D]. 广州：暨南大学，2012

25 Chen J, Chen L, Zhu LH, *et al*. Association of methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism with preterm delivery and placental abruption: a systematic review and meta - analysis [J]. Acta Obstet Gynecol Scand, 2016, 95(2): 157 - 165

26 Walid AA, Abdulsamad W, Samer A, *et al*. Association of methylenetetrahydrofolate reductase C677T and A1298C gene polymorphisms with recurrent pregnancy loss in syrian women [J]. Reprod Sci, 2017, 24(9): 1275 - 1279

（收稿日期：2019 - 07 - 04）
（修回日期：2019 - 09 - 07）