

帕金森病小脑枢纽基因及诊断模型

包成政¹,王功俊¹,罗雪莲¹,王洁²,蓝学群³,王喻¹,陈金梅¹,李雪斌^{2,3,4}

摘要 目的:探索帕金森病(PD)小脑的差异表达枢纽基因及构建诊断模型,探讨PD的发病机制,寻求潜在干预靶点。**方法:**通过GEO数据库获取PD小脑总RNA芯片GSE20314、GSE28894的数据,利用limma和WGCNA筛选出对照组人群和PD患者的差异表达基因;再通过VennDiagram寻找枢纽(Hub)基因。获取Hub基因后再使用逻辑回归分析构建诊断模型,并通过血液样本数据集GSE18838和GSE6613进行验证,计算受试者工作特征曲线(ROC)的曲线下面积(AUC)评估其诊断性能。**结果:**①获得CDKL2、DGKI、ERO1B、SBNO1、SDK2、SYNRG、VPS13C共7个Hub基因。②逻辑回归分析获取SBNO1+VPS13C构建诊断模型,ROC曲线显示在GSE20314+GSE2889、GSE18838、GSE6613中的AUC分别为0.9386、0.8663、0.6564。**结论:**Hub基因SBNO1、VPS13C可能参与PD的发生、发展过程,未来或可作为PD的潜在干预靶点,且基于SBNO1+VPS13C基因构建的诊断模型具有较好的诊断性能。

关键词 帕金森病;小脑;基因

中图分类号 R741;R741.02;R742 文献标识码 A DOI 10.16780/j.cnki.sjssgncj.20220988

Cerebellar Hub Genes and Diagnostic Model of Parkinson's disease BAO Cheng-zheng¹,WANG Gong-jun¹, LUO Xue-lian¹, WANG Jie², LAN Xue-qun³, WANG Yu¹, CHEN Jin-mei¹, LI Xue-bin^{2,3,4}. 1. Graduate School of Youjiang Medical University for Nationalities, Guangxi Baise 533000, China; 2. Key Laboratory of High Disease Prevention and Control Research in the Western Guangxi Region of Guangxi University, Affiliated Hospital of Youjiang Medical University for Nationalities, Guangxi Baise 533000, China; 3. Department of Neurology, Affiliated Hospital of Youjiang Medical University for Nationalities, Guangxi Baise 533000, China; 4. Modern Industrial College of Biomedicine and Great Health, Youjiang Medical University for Nationalities, Guangxi Baise 533000, China

Abstract Objective: To explore the differentially expressed Hub genes in the cerebellum of Parkinson's disease (PD) and to construct a diagnostic model, explore the pathogenesis of PD and seek the potential causes pretarget. **Methods:** The data of GSE20314 and GSE28894, PD cerebellar total RNA, were obtained from the GEO database. Limma and WGCNA were used to screen the differentially expressed genes between control population and PD patients. Then VennDiagram was used to search for Hub genes. Logistic regression was used to construct the diagnostic model after obtaining the hub genes, and the AUC of ROC curve was used to evaluate its diagnostic performance using data of GSE18838 and GSE6613, two data of PD blood samples. **Results:** (1) Seven Hub genes including CDKL2, DGKI, ERO1B, SBNO1, SDK2, SYNRG and VPS13C were obtained. (2) The diagnostic model was constructed based on SBNO1+VPS13C by logistic regression analysis. ROC curve showed that the AUC of GSE20314+GSE2889, GSE18838 and GSE6613 were 0.9386, 0.8663 and 0.6564, respectively. **Conclusion:** Hub genes such as SBNO1 and VPS13C may be involved in the occurrence and development of PD, which may be used as potential intervention targets for PD in the future. The diagnostic model based on SBNO1+VPS13C gene has good diagnostic performance.

Keywords Parkinson's disease; cerebellum; gene

帕金森病(Parkinson's disease, PD)是第二大退行性病变,临床主要表现为静止性震颤、运动迟缓、肌强直、姿势步态异常等症状^[1]。随着疾病进展,患者可因治疗效果差、并发症多等导致生活质量严重下降,甚至死亡。深入研究PD的发病机制,寻找更多早期诊断标志物及潜在干预靶点使PD得以早诊断、早治疗是现阶段的热门研究方向。目前PD的研究主要集中在基底节区。有学者提出小脑有可能参与PD震颤的发生或发展过程^[2],也有学者认为小脑-丘脑-皮质回路

有可能参与PD震颤^[3]。小脑的功能异常有可能是PD冻结步态的基础^[4],使用神经调节技术可改善PD患者运动症状^[5],刺激小脑可以降低PD患者左旋多巴治疗诱发的运动障碍的严重程度,小脑可能参与PD的疾病进程^[6]。小脑在PD的发病机制中具有重要地位,但目前相关研究较少,同时探索PD的诊断方式也是现行研究的热点。

本研究基于生物信息学方法,从GEO数据库获取PD小脑总RNA芯片GSE20314、GSE28894的数据,通过对该数据进行

作者单位

- 1. 右江民族医学院
研究生院
广西 百色 533000
- 2. 右江民族医学院
附属医院广西高校
桂西地区高发病防
治研究重点实验室
广西 百色 533000
- 3. 右江民族医学院
附属医院神经内科
广西 百色 533000
- 4. 右江民族医学院
生物医药与大健康
现代产业学院
广西 百色 533000

基金项目

国家自然科学基金
(No. 81860226)
右江民族医学院横
向课题(No.YYHX
2021001)

收稿日期

2022-11-26

通讯作者

李雪斌
13507766338@
139.com

limma 差异分析和加权基因共表达网络分析(weighted gene co-expression network analysis, WGCNA), 并通过 VennDiagram 寻找枢纽(Hub)基因。再使用逻辑回归分析构建诊断模型, 并通过 2 个血液样本数据集进行验证, 计算受试者工作特征曲线(receiver operator characteristic curve, ROC)的曲线下面积(area under the curve, AUC)评估其诊断性能。探索小脑在 PD 中的作用及构建具有应用价值的诊断模型。

1 资料与方法

1.1 数据采集

数据集来自 NCBI 基因表达综合公共数据库(GEO)。GSE20314 和 GSE28894 是在小脑中表达的 RNA 数据集, GPL96 注释的 GSE20314 包括 4 个对照组人群和 4 个 PD 患者的小脑 RNA 注释样本; GPL6104 注释的 GSE28894 包括 14 个对照组人群和 15 个 PD 患者的小脑 RNA 注释样本。在分析前将 2 样本去批次及归一化, 见图 1。GSE18838 和 GSE6613 是外周血 RNA 表达的 2 个数据集, 其中 GPL5175 注释的 GSE18838 包括 11 个正常人样本和 17 个 PD 患者样本, GPL96 注释的 GSE6613 包括 22 个正常人样本和 50 个 PD 患者样本。

1.2 方法

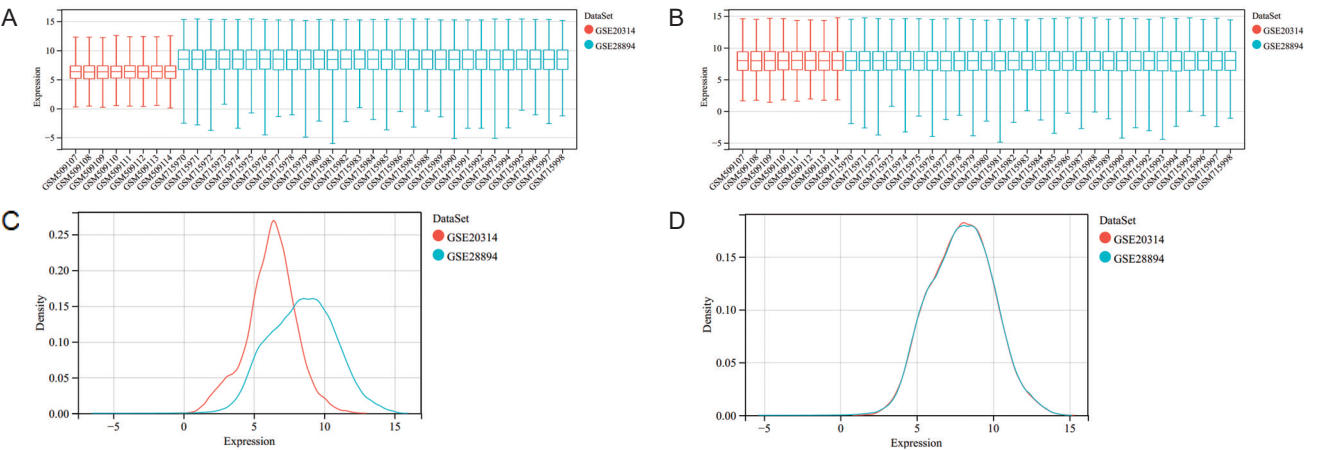
1.2.1 差异表达分析 Limma 是一种基于广义线性模型的差异表达筛选方法^[7]。在分析前将 GSE20314 和 GSE28894 利用“inSilicoMerging”包进行合并, 再使用经典贝叶斯方法进行去除批次效应^[8]。我们使用 R 软件 limma 包(version 3.40.6)进行差异分析。P<0.05 和差异倍数 1.1 倍的基因被视为差异基因(DEGs)。DEGs 的筛选及热图和火山图绘制使用 Sangerbox 在线分析工具(<http://vip.sangerbox.com/home.html>)。

1.2.2 加权基因共表达网络分析 基因之间共表达及

基因与表型之间的关系探索使用在线分析工具 Sangerbox (<http://vip.sangerbox.com/home.html>)。首先分别计算每个基因的 MAD(median absolute deviation), 剔除 MAD 最小的前 50% 的基因。利用 R 软件包 WGCNA 的 goodSamplesGenes 方法去除离群的基因和样本, 进一步构建 scale-free co-expression network。首先对所有配对基因分别采用 Pearson 相关矩阵和平均连锁法, 然后用幂函数 $A_{mn}=|C_{mn}|^{\beta}$ (C_{mn} =基因 m 和基因 n 之间的 Pearson 相关关系, A_{mn} =基因 m 和基因 n 之间的邻接关系) 构造加权邻接矩阵(β 是一个能强调基因之间强相关性和惩罚弱相关性的软阈值参数)。选择适当的幂后, 将邻接关系转化为一个拓扑重叠矩阵(TOM), 该矩阵可以度量一个基因的网络连通性, 定义为该基因与所有其他基因的邻接关系之和作为网络基因比, 并计算相应的不相似度(1-TOM)。为了将具有相似表达谱的基因划分为基因模块, 根据基于 tom 的不相似度度量进行平均连锁层次聚类。为进一步计算模块特征基因的不同相似度, 选择一条切线作为模块树状图, 合并一些模块, 并对每个模块赋予不同的颜色描绘。模块隶属度(MM)是基因表达值与模块 ME 的相关系数, 表示该基因与该模块的相关性。基因显着性(GS)是基因表达值与表型之间的相关系数, 代表基因与表型之间的关联。

1.2.3 枢纽(Hub)基因鉴定 使用“VennDiagram”包获取 DEGs 和 WGCNA 获得的关键模块基因的交集, 进而获得与 PD 较为密切相关的 Hub 基因, Hub 基因在对照组人群和 PD 患者小脑中的表达使用箱线图表示。

1.2.4 逻辑回归模型 逻辑回归是一种广义线性回归分析模型, 可用于疾病的自动诊断。本研究采用 2 个响应变量的逻辑回归, 响应变量为 1 时表示 PD 样本, 响应变量为 0 时表示正常样本。逐步回归分析最初用



注: (A、C)去批次前; (B、D)去批次后

图1 GSE20314、GSE28894箱线图和Density图

于消除对响应变量不显著的因素,仅保留显著的因素以简化模型。逐步回归从模型中迭代地添加或删除变量,直到保留最显著因素。之后使用逻辑回归来拟合这些显著因素与响应变量之间的关系。最后使用受试者工作特征曲线(ROC)和ROC曲线下面积(AUC)评估模型的诊断效果。

1.3 统计分析

所有分析在 Sangerbox、SPSS 23.0、Graphpad Prism 9.4.0 中进行,根据是否符合正态分布选择 *t* 检验或秩和检验; $P<0.05$ 为差异有统计学意义。

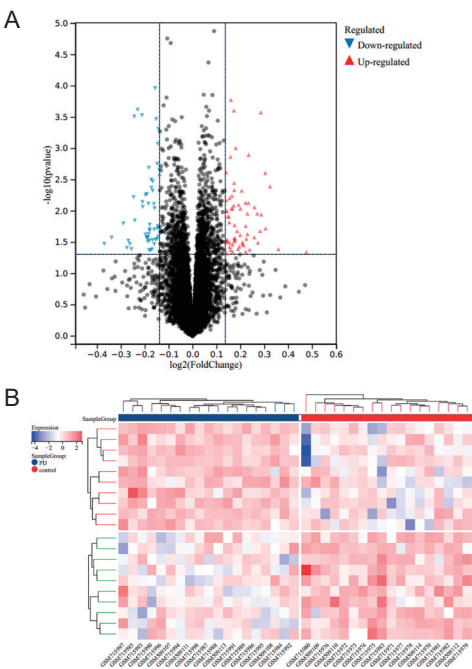
2 结果

2.1 Limma 差异基因

DEGs 以 $P<0.05$ 、差异倍数 1.1 倍的基因作为筛选条件,共得到 116 个 DEGs,使用火山图及热图(显示 *t* 值前 10 位及最后 10 位基因)表示,见图 2。

2.2 WGCNA 结果

从 2 个数据集集合选取基因并用于构建加权基因共表达网络。当软阈值功率设置为 16 时,尺度独立性达到 0.89,平均连接值为 0.29。当最小模块大小设置为 30 时,合并距离 <0.25 的模块,最终获得 13 个共表达模块;值得注意的是 grey 模块被认为无法被分配给任何模块的基因集合。然后对每个模块与临床特征进行相关性分析。purple 模块与 PD 的正相关性最高($r=0.35, P=0.03$)。计算与基因的表达相关性获得 GS,同时计算模块特征向量与基因的表达相关性获得 MM,基于截止标准($|MM|>0.6$ 及 $|GS|>0.1$),purple 模块中具有高连通性的 188 个重要基因被提取出来用于进一



注:(A)Limma 差异分析火山图;(B)Limma 差异分析热图
图 2 Limma 差异分析火山图及热图

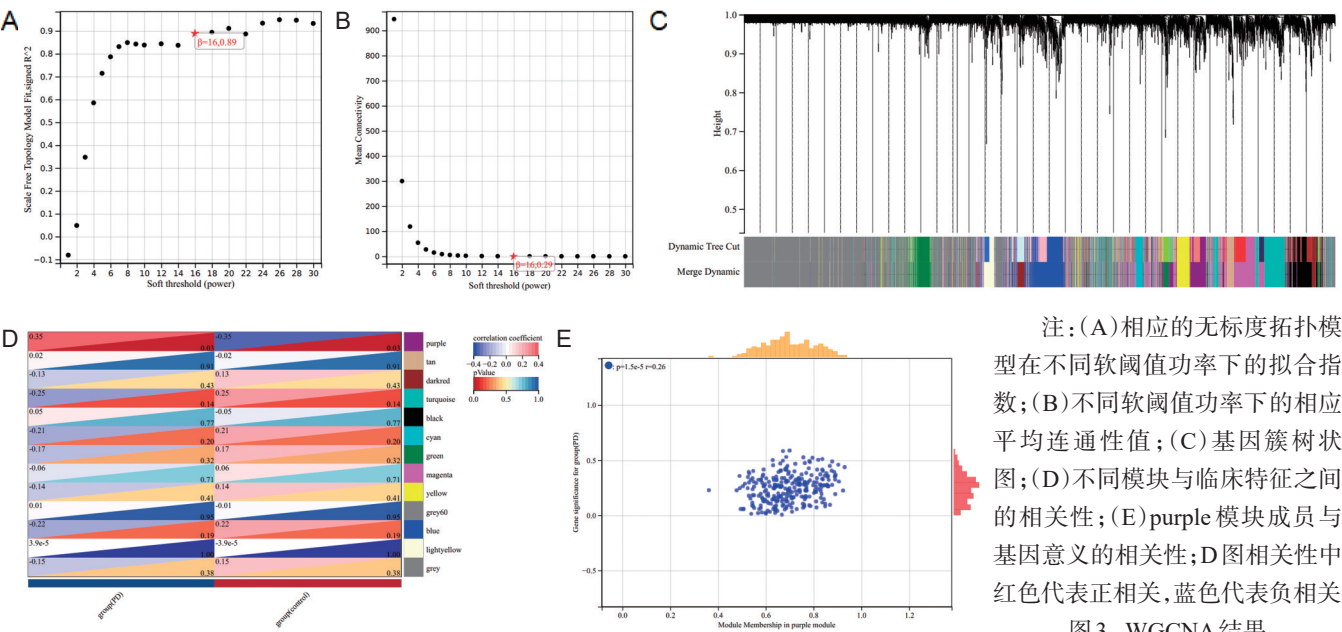
步分析,见图 3。

2.3 Hub 基因

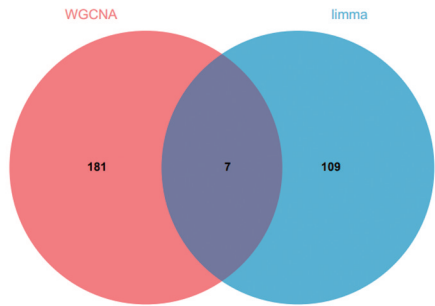
使用 VennDiagram 获得到 7 个 Hub 基因,使用 Veen 图表示,见图 4;使用箱线图表示各基因在各组的表达,见图 5。

2.4 诊断模型的构建和验证

使用逻辑回归算法构建基因预测模型,逐步回归分析提示 SBNO1、VPS13C 对 PD 的诊断影响最显著。在小脑样本数据集 GSE20314、GSE28894 中,SBNO1+VPS13C ROC 曲线 AUC 值为 0.9386,见图 6A。在血液样本数据集 GSE18838、GSE6613 中进一步验证该模型



注:(A)相应的无标度拓扑模型在不同软阈值功率下的拟合指数;(B)不同软阈值功率下的相应平均连通性值;(C)基因簇树状图;(D)不同模块与临床特征之间的相关性;(E)purple 模块成员与基因意义的相关性;D 图相关性中红色代表正相关,蓝色代表负相关
图 3 WGCNA 结果



注:Limma 的 116 个 DEGs 与 WGCNA 的 purple 模块 188 个重要基因的交集

图4 Veen图

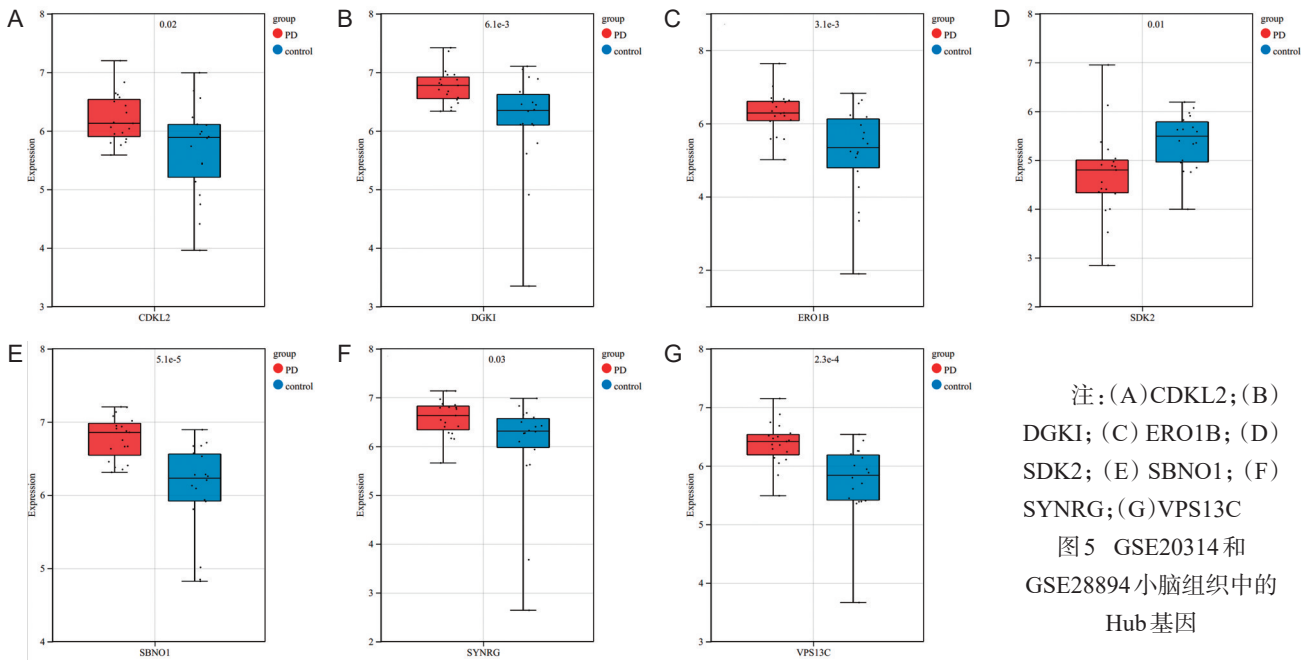
提示具有良好的诊断性能,AUC 值分别为 0.8663、0.6564,见图 6B、C。

3 讨论

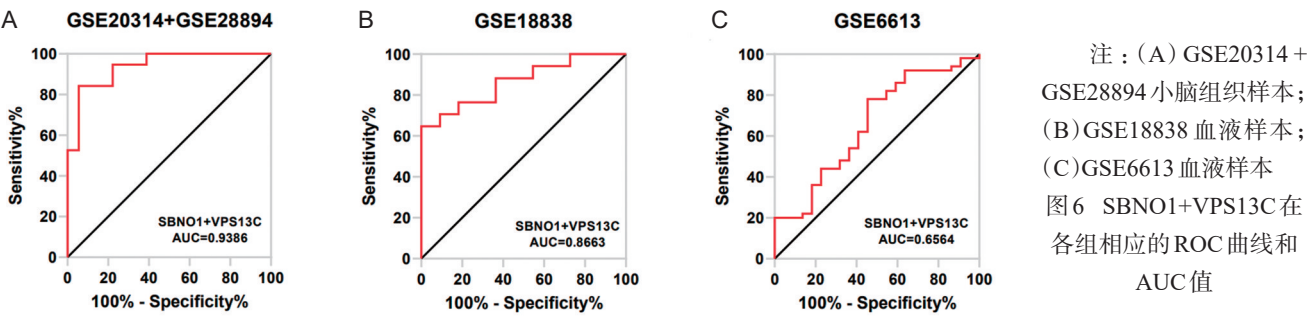
PD 主要为中脑黑质多巴胺能神经元缺失和纹状体多巴胺神经递质减少,导致出现运动迟缓、肌强直等症状,治疗上常以补充多巴胺为主要治疗手段^[9]。现有研究表明 PD 震颤产生与小脑有较为明显的关系,PD 患者腹侧被盖区小脑连接性高于黑质小脑连接性,在正常人群中这种现象则会相反^[10]。Kawabata 等^[11]发现小脑和基底神经节间的功能连接不仅与运动相关,与认知表现也有联系。小脑的功能或形态调节不仅与运动症状相关,与一些非运动症状也有不可忽视的联系,在 PD 中小脑发挥着代偿的作用^[12]。早期 PD 患者中,小脑-丘脑-皮质回路的募集与运动特征的进展呈正相关,同样也参与 PD 的补偿机制^[13]。甚至有学者提出小脑可能参与 PD 的猝死机制^[14]。Rusholt 等^[15]的体视学研究,揭示了 PD 的白质改变情况。基于现有小脑参与

PD 发病机制的学说,相关药物的开发也逐步开展,如姜黄素膳食补充剂可能对 PD 的小脑相关症状具有保护作用^[16]。越来越多证据表明小脑在 PD 中的重要性。

人类 VPS13 基因的突变是造成神经发育和神经退行性疾病的原因,VPS13 可能在细胞器之间的脂质交换中起作用。作为 VPS13 之一的液泡蛋白 13C(VPS13C)与 PD 有着较为密切的联系,而 VPS13A 与舞蹈症棘皮细胞增多症存在联系^[17]。通过全外显子组测序已成功识别导致家族性 PD 的基因,发现 VPS13C 位列其中^[18]。已有研究提示 VPS13C 变异与 PD 相关^[19]。PD 和路易体痴呆均为路易体病,VPS13C 与两者都有联系。VPS13C 中罕见的错义突变与路易体病相关,隐性复合杂合错义突变可能对 VPS13C 的表达和功能产生不同的影响。罕见的隐性复合杂合错义突变的组合可降低 VPS13C 表达并导致路易体病风险增加^[20]。由此可见 VPS13C 与 PD 间具有密切关系。在 PD 细胞模型中,可见 VPS13C 部分定位于线粒体的外膜。VPS13C 沉默与较低的线粒体膜电位、线粒体片段化、呼吸频率增加、PINK1/Parkin 依赖的线粒体恶化以及响应线粒体损伤的 PARK2 转录上调有关。VPS13C 功能丧失会增加线粒体对压力的脆弱性,从而激活 PINK1/Parkin 依赖性线粒体质量控制途径^[21],或许由此参与 PD 的发病过程。不同 PD 患者群之间 VPS13C 存在着些许区别。VPS13C 基因的单核苷酸多态性 rs2414739 与等位基因模型的 PD 易感性存在显著联系,但这不包括中国人群^[22,23]。但在另外一项研究中筛选中国早发性 PD 患者队列中的 VPS13C 突变,结果发现 PD 患者中仍有携带 VPS13C 的复合杂合突变^[24]。



注:(A)CDKL2;(B) DGKI; (C) ERO1B; (D) SDK2; (E) SBNO1; (F) SYNRG; (G)VPS13C
图5 GSE20314和 GSE28894小脑组织中的 Hub 基因



VPS13C 也与东亚人群有着显著联系^[25]。在伊朗人中，VPS13C rs2414739 在 PD 与正常人群中也存在显著差异^[26]。VPS13C 在不同人群中是否有着区别，需进一步扩大样本去验证。综上研究可见 VPS13C 与 PD 间密切的关系，这与我们的研究结果相符。我们的研究结果进一步发现 VPS13C 与 SBNO1 基因共同构建基于小脑的 PD 诊断模型具有良好的诊断性能。在外周血液中，其诊断性能较脑组织有所下降，这或许与脑组织样本往往比血液样本更能代替 PD 的病理改变有关。综合本研究结果及现有研究报道，我们推测该模型对 PD 患者的诊断或许存在一定程度上的指导意义。

鉴于 PD 与小脑间存在着相应联系，我们认为 SBNO1、VPS13C 等 Hub 基因有可能参与 PD 的发生、发展过程，未来或可作为 PD 的潜在干预靶点，且 SBNO1+VPS13C 基因构建的诊断模型具有较好的诊断性能。但其中的相关机制及应用价值仍需进一步探讨及研究。

参考文献

[1] Xu S, He XW, Zhao R, et al. Cerebellar functional abnormalities in early stage drug-naïve and medicated Parkinson’s disease[J]. J Neurol, 2019, 266: 1578-1587.
[2] Lefaivre SC, Brown MJ, Almeida QJ. Cerebellar involvement in Parkinson’s disease resting tremor[J]. Cerebellum Ataxias, 2016, 3: 13.
[3] Lewis MM, Du G, Sen S, et al. Differential involvement of striato- and cerebello-thalamo-cortical pathways in tremor- and akinetic/rigid-predominant Parkinson’s disease[J]. Neuroscience, 2011, 177: 230-239.
[4] harti K, Suppa A, Pietracupa S, et al. Abnormal Cerebellar Connectivity Patterns in Patients with Parkinson’s Disease and Freezing of Gait[J]. Cerebellum, 2019, 18: 298-308.
[5] França C, de Andrade DC, Teixeira MJ, et al. Effects of cerebellar neuromodulation in movement disorders: A systematic review[J]. Brain Stimul, 2018, 11: 249-260.
[6] Coutant B, Frontera JL, Perrin E, et al. Cerebellar stimulation prevents Levodopa-induced dyskinesia in mice and normalizes activity in a motor network[J]. Nat Commun, 2022, 13: 3211.
[7] Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic Acids Res, 2015, 43: e47.
[8] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods[J]. Biostatistics, 2007, 8: 118-127.
[9] Geibl FF, Henrich MT, Oertel WH. Mesencephalic and extramesencephalic dopaminergic systems in Parkinson’s disease[J]. J

Neural Transm (Vienna), 2019, 126: 377-396.
[10] O’Shea IM, Popal HS, Olson IR, et al. Distinct alterations in cerebellar connectivity with substantia nigra and ventral tegmental area in Parkinson’s disease[J]. Sci Rep, 2022, 12: 3289.
[11] Kawabata K, Watanabe H, Bagarinao E, et al. Cerebello-basal ganglia connectivity fingerprints related to motor/cognitive performance in Parkinson’s disease[J]. Parkinsonism Relat Disord, 2020, 80: 21-27.
[12] Wu T, Hallett M. The cerebellum in Parkinson’s disease[J]. Brain, 2013, 136: 696-709.
[13] J Blesa, Trigo-Damas I, Dileone M, et al. Compensatory mechanisms in Parkinson’s disease: Circuits adaptations and role in disease modification [J]. Exp Neurol, 2017, 298: 148-161.
[14] Scorza CA, Guimarães-Marques M, Girão ERC, et al. Sudden death in Parkinson’s disease: Cerebellum in court[J]. J Neurol Sci, 2020, 414: 116854.
[15] Rusholt EHL, Salvesen L, Brudek T, et al. Pathological changes in the cerebellum of patients with multiple system atrophy and Parkinson’s disease-a stereological study[J]. Brain Pathol, 2020, 30: 576-588.
[16] Fikry H, Saleh LA, Abdel Gawad S. Neuroprotective effects of curcumin on the cerebellum in a rotenone-induced Parkinson’s Disease Model[J]. CNS Neurosci Ther, 2022, 28: 732-748.
[17] Kumar N, Leonzino M, Hancock-Cerutti W, et al. VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites[J]. J Cell Biol, 2018, 217: 3625-3639.
[18] Jansen IE, Ye H, Heetveld S, et al. Discovery and functional prioritization of Parkinson’s disease candidate genes from large-scale whole exome sequencing[J]. Genome Biol, 2017, 18(1): 22.
[19] F Hopfner, Mueller S-H, Szymczak S, et al. Rare Variants in Specific Lysosomal Genes Are Associated With Parkinson’s Disease[J]. Mov Disord, 2020, 35: 1245-1248.
[20] Smolders S, Philtjens S, Crosiers D, et al. Contribution of rare homozygous and compound heterozygous VPS13C missense mutations to dementia with Lewy bodies and Parkinson’s disease[J]. Acta Neuropathol Commun, 2021, 9: 25.
[21] Lesage S, Drouet V, Majounie E, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy[J]. Am J Hum Genet, 2016, 98: 500-513.
[22] Bai X, Liu X, Li X, et al. Association between VPS13C rs2414739 polymorphism and Parkinson’s disease risk: A meta-analysis[J]. Neurosci Lett, 2021, 754135879.
[23] Wang L, Cheng L, Li NN, et al. Association of four new candidate genetic variants with Parkinson’s disease in a Han Chinese population[J]. Am J Med Genet B Neuropsychiatr Genet, 2016, 171B: 342-347.
[24] Gu X, Li C, Chen Y, et al. Mutation screening and burden analysis of VPS13C in Chinese patients with early-onset Parkinson’s disease[J]. Neurobiol Aging, 2020, 94311.
[25] Zou M, Li R, Wang JY, et al. Association analyses of variants of SIPA1L2, MIR4697, GCH1, VPS13C, and DDRGK1 with Parkinson’s disease in East Asians[J]. Neurobiol Aging, 2018, 68157-68159.
[26] Safaralizadeh T, Jamshidi J, Esmaili Shandiz E, et al. SIPA1L2, MIR4697, GCH1 and VPS13C loci and risk of Parkinson’s diseases in Iranian population: A case-control study[J]. J Neurol Sci, 2016, 369: 1-4.
(本文编辑:唐颖馨)