

· 综述 ·

脑卒中病因流行病学的家系研究

王雪茵 胡永华

【关键词】 脑卒中; 家系研究; 遗传流行病学

Family-based studies on the etiological epidemiology of stroke WANG Xue-yin, HU Yong-hua. Department of Epidemiology and Bio-statistics, School of Public Health, Peking University, Beijing 100191, China

Corresponding author: HU Yong-hua, Email: yhhu@bjmu.edu.cn

This work was supported by grants from the National Natural Science Foundation of China (No. 81172744 and No. 81230066).

【Key words】 Stroke; Family-based study; Genetic epidemiology

脑卒中是一种由多个遗传和环境因素影响的病因异质性疾病^[1],大量研究表明遗传因素在脑卒中的发生、发展、结局和预后等方面发挥着不容忽视的作用。家系研究通过发现与共享基因和(或)共享环境有关的危险因素来揭示脑卒中等复杂性疾病的潜在病因,是探索遗传和环境相对作用、确定疾病易感基因的最佳策略。近 10 余年来,不断涌现的各种组学技术、现代分子生物学检测、计算机信息化技术以及高效能统计分析方法为家系资料在遗传流行病学研究中的应用提供了可行性。本文对脑卒中尤其是缺血性脑卒中遗传易感性的家系研究进行综述。

一、聚集性研究

1. 提示遗传易感性的证据:2012 年美国心脏协会(AHA)指出高血压、吸烟、高脂血症、糖尿病是脑卒中的传统危险因素,肥胖、心房纤颤、缺乏体育活动、无症状性颈动脉狭窄、冠心病、终末期肾病和慢性肾病、口服避孕药等也会增加脑卒中的发病风险^[2]。除此之外,家族聚集性研究、家系研究及双生子研究均指出遗传因素在脑卒中发病过程中发挥着极为重要的作用。2011 年荷兰学者 Knottnerus 等^[3]的一项病例对照研究表明,脑卒中家族史是腔隙性脑卒中的独立危险因素,多因素分析发现其与双亲脑卒中家族史($OR=6.46$, 95% CI : 1.96 ~ 21.33)、父亲脑卒中家族史($OR=4.00$, 95% CI : 1.18 ~ 13.56)和母亲脑卒中家族史($OR=5.40$, 95% CI : 1.14 ~ 25.61)均有较强关联。此外,阳性家族史还与子代脑卒中死亡率有关。2012 年 JACCS 对日本 22 763 名男性和 30 928 名女性平均 15.9 年的随访发现,若双亲有脑卒中病史,子代的脑卒中死亡风险将升高,多因素调整的相对风险(HR)和人群归因危险度百分比(PAF)分别为 1.25(95% CI :

1.12 ~ 1.40)和 4.8%^[4]。

遗传因素对脑卒中的影响因其类型和发病年龄而变化。2012 年伦敦大学和脑卒中与痴呆症研究组的一项联合研究发现,缺血性脑卒中中存在明显的遗传性,但不同亚型之间具有显著差异:大动脉粥样硬化性和心源性栓塞性脑卒中的遗传度分别为 40.3% 和 32.6%,而小动脉闭塞性缺血性脑卒中的遗传度仅为 16.1%^[5]。此外,遗传因素对脑卒中的影响在不同种族、不同地域的人群中亦发挥着不同的作用。Ariyaratnam 等^[6]的 Meta 分析发现, APOE 基因($OR=2.18$, 95% CI : 1.52 ~ 3.13)和 ACE(I/D)多态性($OR=1.90$, 95% CI : 1.23 ~ 2.93)与非欧洲血统人群缺血性脑卒中有关;而 FV Leiden Arg506Gln($OR=1.33$, 95% CI : 1.12 ~ 1.58)、MTHFR C677T($OR=1.33$, 95% CI : 1.12 ~ 1.58)、PT G20210A($OR=1.44$, 95% CI : 1.11 ~ 1.86)、ACE(I/D)($OR=1.21$, 95% CI : 1.08 ~ 1.35)与欧洲血统人群缺血性脑卒中相关联^[7]。另外还有报道其他与脑卒中相关基因,如 PDE4D、IL1RN、NOS1、PITX2 和 ZFHX3 基因等^[8-11]。

2. 研究策略及统计学方法:家族聚集性研究可通过家族史法和家族研究法,询问并收集先证者及其亲属的疾病发生情况、实验室检查及遗传学标记等资料,计算先证者亲属疾病频率的测量指标(如患病率、发病率、累积发病率),估算 OR 、 RR 值或家族史得分等,从而估计亲属的疾病风险并确定其危险因素^[12]。定量资料则可计算亲属间的相关系数、估计遗传度并构建家庭内疾病模型。

二、家系研究

1. 近期研究概况:2002 年冰岛 deCODE 研究组应用家系资料进行全基因组连锁分析,发现染色体 5q12 与脑卒中高度连锁($LOD=4.4$),在剔除出血性脑卒中家系后,发现其连锁程度增加($LOD=4.86$)^[13]。随后又对该区域进行精确定位,发现 PDE4D 基因与心源性栓塞性和大动脉粥样硬化性脑卒中相关联。2009 年该研究组又报道染色体 16q22 中 ZFHX3 基因 rs7193343 与缺血性脑卒中存在关联($OR=1.11$, 95% CI : 1.04 ~ 1.17)^[14]。

基于瑞典 MONICA 方案,通过脑卒中登记系统对 56 个家系进行分析,验证 deCODE 报道的 PDE4D 基因与脑卒中的连锁关系^[15],随后又发现了 9 个与脑卒中连锁的染色体区域^[16]。2009 年在对 7 个有共同祖先的核心家系的全基因组连锁研究中,发现染色体 9q31 ~ q33 与脑卒中高度连锁($LOD=4.81$)^[17]。

SWISS 研究组在美国和加拿大的 70 个临床中心募集 223 名脑卒中先证者、248 名患病一致同胞和 84 名患病不一致同胞,发现同胞间发病年龄的一致率为 0.83,缺血性脑卒

DOI: 10.3760/cma.j.issn.0254-6450.2013.06.026

基金项目:国家自然科学基金(81172744, 81230066)

作者单位:100191 北京大学公共卫生学院流行病与卫生统计学系

通信作者:胡永华, Email: yhhu@bjmu.edu.cn

中类型的一致率为 33.8%, 全基因组扫描未发现具有统计学意义的缺血性脑卒中位点, 但指出染色体 3p(NOS1) 和 6p 很可能与其相关^[18,19]。

北曼哈顿家系研究(NOMAFS)选择易于测量的脑卒中中间表型, 探索加勒比拉美裔人群脑卒中的基因、环境危险因素及其之间的交互作用^[20]。从 2005 年起, NOMAFS 先后报道颈动脉扩张功能的遗传度约为 20%^[21]; 颈动脉内膜中层厚度的遗传度为 0.65^[22]; 染色体 12p11 (MLOD=3.11) 上 D12S1042 附近的一个数量性状位点与左心室质量相连锁^[23]; 2 号 (MLOD=1.77) 和 17 号 (MLOD=2.66) 染色体上存在与同型半胱氨酸水平 (Hcy) 连锁较显著的数量性状位点^[24]。但该研究仅针对加勒比拉美裔人群, 其结果仍需在其他人中验证。

2. 研究策略及统计学方法: 家系研究不仅可避免因人群分层而导致 I 类错误增加的问题, 而且能够分别进行连锁和关联分析, 使二者互为补充, 相得益彰^[25]。家系资料通常包括核心家系 (parent-offspring trios)、患病一致同胞对 (concordant sib-pairs)、患病不一致同胞对 (discordant sib-pairs)、包含多个患病个体的扩展家系 (extended pedigrees with multiple affected individuals) 以及血缘家系 (consanguineous families) 等^[26]。家系研究包括家系风险的决定因素、遗传模型的决定因素、测定基因型的估计风险度 (外显率)、遗传相关性研究、按家族史对风险度进行分层、遗传易感个体风险度的影响因素研究、危险因素的前瞻性研究^[27]。

(1) 连锁研究: 连锁分析的主要目标是确定与疾病性状有关的基因组区域, 分为参数连锁分析和非参数连锁分析^[25]。参数连锁分析中, 首先要求已知或假定疾病模型 (如隐性或显性遗传等), 然后使用似然比方法检验遗传标记和疾病的非独立分配情况, 使用剥皮算法 (peeling algorithm)、隐藏 Markov 模型 (hidden Markov model) 和基于计算机模拟的 MCMC 运算 (Markov Chain Monte Carlo algorithms) 可扩展似然比检验的应用范围。非参数连锁分析通过比较患病亲属基因组特定区域的血缘一致性 (IBD) 来确定其与孟德尔期望是否存在显著性差异, 包括 Abel 等提出的最大似然二项式法 (maximum-likelihood binomial method) 和纯合子作图法 (Homozygosity mapping) 等。

(2) 关联研究: 1993 年 Spielman 等提出了适用于核心家系的传递不平衡检验 (transmission-disequilibrium test, TDT), 此方法不需要基因分型和等位基因频率信息, 但不适合脑卒中等晚发性疾病的研究。此后, 一些研究者对 TDT 方法进行扩展, 提出了基于家系的患病对照 (affect-family-based controls, AFBAC) 方法、以家系为基础的关联检验 (family-based association test, FBAT), 以及专门应用于扩展家系连锁不平衡 (linkage disequilibrium, LD) 分析的家系不平衡检验 (pedigree disequilibrium test, PDT)。2003 年 Martin 等提出了 APL (association in the presence of linkage) 检验, 此方法通过估计 IBD 参数可以正确推断出连锁区域中缺失的父母基因型, 且检验效能高于 PDT^[26]。2004 年 Gordon 等^[28]

提出了 TDTac 方法, 可在基因分型中允许孟德尔错误的存在, 并提供了不同类型的错误模型。2007 年 Zhao 等提出基于熵的 TDT (the entropy-based TDT) 方法, 其检验效能高于 TDT, 但 I 类错误发生率增加^[29]。

(3) 全基因组关联研究 (GWAS): 该研究能够同时检测大量疾病易感位点, 但多重假设检验将导致假阳性率升高, 统计分析时需对其进行校正^[30]。另外, 当不存在人群分层问题时, 家系研究在 GWAS 中发挥的效能低于以人群为基础的研究设计。van Steen 等^[25]的研究显示, 当先证者人数不少于 1500、SNPs 最小等位基因频率不小于 0.01、遗传度为中度 (> 0.03) 时, GWAS 可以发挥最好的研究效能。家系资料的 GWAS 统计学方法有两大类。第一类是传统的基于家系设计的关联分析, 如 TDT、PDT、FBAT 等; 第二类是基于人群的适用于家系数据的分析方法。前者不受人群分层和错误假定表型模型的影响, 并能应用于多种类型家系设计的分析; 后者通过基于人群分析的原理能够达到最大的统计学效能。2011 年 Won 等^[31]提出了一种基于家系资料 GWAS 的新方法, 即使用 Liptak 统计方法将家系内与家系间的关联信息结合为一个综合统计量, 计算公式: $Z_i = w_{FBAT} Z_{pFBAT} + w_T Z_{pTIO}$ 。此方法既能获得最优的效能水平, 同时也不受人群分层的影响, 结果稳健可靠。

三、双生子研究

1. 近期重要发现: 双生子研究被视为一种特殊类型的家系设计, 能够有效地区分共享基因和共享环境的影响。经典的双生子研究方法通过观察同卵双生子 (monozygotic twin) 和异卵双生子 (dizygotic twin) 间的差异来推测遗传因素在疾病进程中的作用^[32]。Flossmann 等^[33]的系统综述中阐述了一项针对美国国家科学院全国研究理事会 (NAS-NRC) 的双生子登记系统中 9475 对男性双生子的邮件问卷调查, 发现同卵双生子的先证者一致率 (17.7%) 是异卵双生子先证者一致率 (3.6%) 的 4.3 倍。之后, 又对此登记系统中的 4345 对同卵双生子、5240 对异卵双生子和 1261 对未知卵型的双生子随访达 10 年, 发现同卵双生子的脑卒中一致率 (12.8%) 显著高于异卵双生子 (8.0%, $\chi^2 = 4.085$, $P < 0.05$)。Bak 等^[34]以丹麦双生子出生登记系统、死因登记系统和出院登记系统为基础, 建立了两个不同年龄的脑卒中双生子队列, 分别包含 990 对和 869 对同性的双生子, 均发现同卵双生子患脑卒中的风险高于异卵双生子, 多因素调整 RR 值分别为 2.1 (95% CI: 1.3 ~ 3.3) 和 1.5 (95% CI: 0.9 ~ 2.4), 遗传度为 0.32, 再次证明了遗传因素在脑卒中发病中的作用。

2. 研究策略及统计学方法: 2012 年 van Dongen 等^[35]指出当今组学时代 (omics era) 下的双生子研究, 在识别和理解复杂性状的分子通路及机制中发挥着重要作用, 包括①遗传和环境因素的定量分析: 广泛用于不同性状由相同或不同遗传和环境因素影响程度的估计, 并通过检验特定性状的遗传度是否在不同环境暴露水平下发生改变以推断多基因与环境交互作用的模型; ②表型不一致双生子 (discordant twins) 的应用价值: 双生子对照法 (co-twin control method) 应用表

型不一致双生子设计可区分反映因果关系的关联和基因或环境的混杂效应导致的关联;③分子表型和数量性状变异的原因:目前先进的分子检测技术能够在产生表型相似性的分子过程水平上估计双生子间的相似程度,有利于发现新的疾病病因通路。

综上所述,脑卒中的遗传易感性研究日益受到关注,家系资料通过各种不同类型的研究设计和高效能统计学方法,在探索脑卒中遗传易感性方面显示出独特的优势,其中选择具有相同遗传基础的亲属作为对照,可有效避免人群分层的影响,并考虑了等位基因传递的信息,可用于多种家系类型、多个性状的复杂表型分析;各种高级统计学方法极大地提高了家系研究的检验效能;提高研究效率,能同时进行连锁分析和关联分析,结果准确可靠。虽然研究对象募集困难、基因分型工作量较大等问题对家系研究的实施提出了挑战,但计算机管理系统和高通量检测技术的不断发展可以解决这些困扰。

参 考 文 献

- [1] Krafft PR, Bailey EL, Lekic T, et al. Etiology of stroke and choice of models. *Int J Stroke*, 2012, 7(5):398-406.
- [2] Roger VL, Go AS, Lloyd-Jones DM, et al. Heart disease and stroke statistics-2012 update: a report from the American Heart Association. *Circulation*, 2012, 125(1):e2-220.
- [3] Knotnerus ILH, Gielen M, Lodder J, et al. Estimating the magnitude of genetic factors by calculating the genetic relative risk of stroke in first-ever lacunar stroke patients. *PLoS One*, 2011, 6(6):e21439.
- [4] Eguchi E, Iso H, Wada Y, et al. Parental history and lifestyle behaviors in relation to mortality from stroke among Japanese men and women: the Japan Collaborative Cohort Study. *J Epidemiol*, 2012, 22(4):331-339.
- [5] Bevan S, Traylor M, Adib-Samii P, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke*, 2012, 43(12):3161-3167.
- [6] Ariyaratnam R, Casas JP, Whittaker J, et al. Genetics of ischaemic stroke among persons of non-European descent: a meta-analysis of eight genes involving approximately 32 500 individuals. *PLoS Med*, 2007, 4(4):e131.
- [7] Jerrard-Dunne P, Cloud G, Hassan A, et al. Evaluating the genetic component of ischemic stroke subtypes: a family history study. *Stroke*, 2003, 34(6):1364-1369.
- [8] Worrall BB, Brodt TG, Brown RD Jr, et al. IL1RN VNTR polymorphism in ischemic stroke: analysis in 3 populations. *Stroke*, 2007, 38(4):1189-1196.
- [9] Milton AG, Aykanat VM, Hamilton-Bruce MA, et al. Association of the phosphodiesterase 4D (PDE4D) gene and cardioembolic stroke in an Australian cohort. *Int J Stroke*, 2011, 6(6):480-486.
- [10] Manso H, Krug T, Sobral J, et al. Variants within the nitric oxide synthase 1 gene are associated with stroke susceptibility. *Atherosclerosis*, 2012, 220(2):443-448.
- [11] Traylor M, Farrall M, Holliday EG, et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*, 2012, 11(11):951-962.
- [12] Hu YH. Genetic epidemiology. Beijing: Peking University Medical Press, 2008: 109. (in Chinese)
胡永华. 遗传流行病学. 北京: 北京大学医学出版社, 2008: 109.
- [13] Nakayama T, Asai S, Sato N, et al. PDE4D gene in the STRK1 region on 5q12: susceptibility gene for ischemic stroke. *Curr Med Chem*, 2007, 14(30):3171-3178.
- [14] Gudbjartsson DF, Holm H, Gretarsdottir S, et al. A sequence variant in ZFX3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet*, 2009, 41(8):876-878.
- [15] Nilsson-Ardnor S, Wiklund PG, Lindgren P, et al. Linkage of ischemic stroke to the PDE4D region on 5q in a Swedish population. *Stroke*, 2005, 36(8):1666-1671.
- [16] Nilsson-Ardnor S, Janunger T, Wiklund PG, et al. Genome-wide linkage scan of common stroke in families from northern Sweden. *Stroke*, 2007, 38(1):34-40.
- [17] Janunger T, Nilsson-Ardnor S, Wiklund PG, et al. A novel stroke locus identified in a northern Sweden pedigree: linkage to chromosome 9q31-33. *Neurology*, 2009, 73(21):1767-1773.
- [18] Meschia JF, Kissela BM, Brodt TG, et al. The Siblings With Ischemic Stroke Study (SWISS): a progress report. *Clin Med Res*, 2006, 4(1):12-21.
- [19] Meschia JF, Nalls M, Matarin M, et al. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. *Stroke*, 2011, 42(10):2726-2732.
- [20] Sacco RL, Sabala EA, Rundek T, et al. Design of a family study among high-risk Caribbean Hispanics: the Northern Manhattan Family Study. *Ethn Dis*, 2007, 17(2):351-357.
- [21] Juo SH, Rundek T, Lin HF, et al. Heritability of carotid artery distensibility in Hispanics: the Northern Manhattan Family Study. *Stroke*, 2005, 36(11):2357-2361.
- [22] Sacco RL, Blanton SH, Slifer S, et al. Heritability and linkage analysis for carotid intima-media thickness: the family study of stroke risk and carotid atherosclerosis. *Stroke*, 2009, 40(7):2307-2312.
- [23] Wang L, Beecham A, Di Tullio MR, et al. Novel quantitative trait locus is mapped to chromosome 12p11 for left ventricular mass in Dominican families: the Family Study of Stroke Risk and Carotid Atherosclerosis. *BMC Med Genet*, 2009, 10:74.
- [24] Della-Morte D, Beecham A, Rundek T, et al. Genetic linkage of serum homocysteine in Dominican families: the Family Study of Stroke Risk and Carotid Atherosclerosis. *Stroke*, 2010, 41(7):1356-1362.
- [25] van Steen K, McQueen MB, Herbert A, et al. Genomic screening and replication using the same data set in family-based association testing. *Nat Genet*, 2005, 37(7):683-691.
- [26] Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. *Nat Rev Genet*, 2011, 12(7):465-474.
- [27] Hopper JL, Bishop DT, Easton DF. Population-based family studies in genetic epidemiology. *Lancet*, 2005, 366(9494):1397-1406.
- [28] Gordon D, Haynes C, Johnnidis C, et al. A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. *Eur J Hum Genet*, 2004, 12(9):752-761.
- [29] Ewens W, Li M. Comments on the entropy-based transmission/disequilibrium test. *Hum Genet*, 2008, 123(1):97-100.
- [30] Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet*, 2006, 7(5):385-394.
- [31] Won S, Wilk JB, Mathias RA, et al. On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies. *PLoS Genet*, 2009, 5(11):e1000741.
- [32] Jermendy G, Horvath T, Littvay L, et al. Effect of genetic and environmental influences on cardiometabolic risk factors: a twin study. *Cardiovasc Diabetol*, 2011, 10:96.
- [33] Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*, 2004, 35(1):212-227.
- [34] Bak S, Gaist D, Sindrup SH. Genetic liability in stroke: a long-term follow-up study of danish twins. *Stroke*, 2002, 33(3):769-774.
- [35] van Dongen J, Slagboom PE, Draisma HH, et al. The continuing value of twin studies in the omics era. *Nat Rev Genet*, 2012, 13(9):640-653.

(收稿日期:2013-01-03)

(本文编辑:张林东)