

· 综 述 ·

遗传流行病学

张卫华

产生和发展

群体遗传学在 1920 年被确认为一门学科。1960 年它又分为两个分枝,一个与历史过程有关,将遗传学与人类学以及在其他物种的古生物学和分类学结合。另一分枝与同时期人群的遗传有关,将遗传学与流行病学、医学、心理学以及在其他物种的农学相结合^[1]。后一分枝在本世纪 80 年代中期产生了遗传流行病学,并被人们普遍接受。它是遗传学与流行病学相互渗透产生的一门完整的独立的边缘学科,合并了流行病、统计学方法与分子遗传学技术,把许多背景不同的研究者联系到了一起。它是群体遗传学中最年轻和发展最快的分枝。Morton^[1]认为群体遗传学的每个方面,只要不是主要涉及进化的,就属于遗传流行病学的内容。

遗传学与流行病学的结合是近代对复杂疾病病因认识的需要。1954 年 Neel 和 Schull 就提出新的学科术语“流行病遗传学”(epidemiological genetics),即研究遗传和疾病的关系。后来 Morton 和 Chung 等一致认为流行病和遗传学方法的结合是必需的、不可避免的,而决定家族聚集性的因素有可能是纯环境的,“流行病遗传学”未考虑环境假设,容易产生偏颇,故于 1978 年正式使用“遗传流行病学”术语^[2,3],即遗传流行病学是“研究亲属中疾病和病因、分布和控制以及研究在人群中疾病遗传的原因的一门学科”。这一定义为人们所普遍接受。

研究方法

包括群体和家系研究。家系研究一般包括三个连续的步骤:有无家族聚集性、遗传和环境因素的作用如何、遗传机制。

一、群体研究:流行病学的主要内容包括研究环

境病因的作用,所以遗传流行病学也包括群体的病因研究,即研究环境与遗传因素及其交互作用在疾病发生中的作用^[4,5]。众所周知,几乎所有的疾病都涉及遗传和非遗传因素及其相互作用。肿瘤发生的“二次打击”学说就是早期著名的将流行病学分析用于医学遗传学的成功例子^[4]。研究发病率在时间、地点、胎次等的变化,以及比较疾病易感基因在病例与对照分布的关联研究都属于群体研究的内容。

二、家族聚集性研究:家族聚集现象间接反映了遗传因素的存在。研究方法包括病例对照和定群研究。判断标准如下:①患者亲属的患病率或发病率高于对照亲属或普通人群;②有家族史患者亲属的发病率高于随机人群患者亲属;③与患者血缘关系越近发病率越高;④数量性状亲属对之间的相关大于非亲属对^[6]。

三、双生子和养子研究:用于区分遗传和非遗传病因。双生子研究假设单卵双生子(MZ)和异卵双生子(DZ)所处的环境相似,如果 MZ 的疾病一致率高于 DZ,表明家族因素中有遗传因子的存在。养子研究是通过比较患病养子的血缘亲属和非血缘亲属的发病危险性,或比较患病和非患病养子的血缘亲属的发病危险性来证实遗传因素的存在。存在的问题是样本较难获得,对合子确认方法不明确会造成偏倚,对于大多数复杂疾病,双生子中仅很少一部分是家族遗传的,使平均一致率降低,难以取得强的遗传因素的证据。因此,考虑其他变量以识别可能的疾病异质性显得很重要^[7]。

四、遗传方式的研究(分离分析):确定了遗传因素的存在,就要了解与该病相关的位点数和这些位点的基因的遗传方式。由于实际的遗传方式可能很复杂,一般考虑简单化模型,即证实单位点(或最多 2~3 个位点)的存在。该位点的等位基因对患病易感性有主效应,将剩余的易感性变异分为家族和非家族的因素。通常估计家系资料在不同的遗传模型

下的概率,研究哪种模型最适合此批数据。最大似然方法用于比较模型和获得参数的估计值,如基因频率和年龄特异的外显率。还可以检验基因型和发病年龄的关系以及环境因素的效应。

五、易感基因的定位(连锁研究):当研究提示一或多个主基因与疾病发生有关,就必须识别这(些)基因的分子特征。连锁分析通过检查几个位点上的等位基因以何种组合从一代传递到下一代,来推断两个或以上位点(以及疾病位点)的相对位置。该方法依赖于可获得的多态性标记。减数分裂时来自父母的不同染色体间进行遗传物质的交叉和交换。染色体上两位点之间交叉的概率为重组分数(θ)。当 θ 为 0 时,表示两位点未出现交叉,二者相距极近, θ 为 0.5 时,表示两位点的等位基因在减数分裂时随机组合,故两位点不在同一染色体上,或在同一染色体上相距很远。重组分数越小,两位点相距越近。

lod score 方法假设在已知的遗传模型和某一 θ 值(0~0.5)下,计算观察家系的似然度(likelihood)与 $\theta=0.5$ 时似然度的比值的对数(lod 值),最大 lod 值下的 θ 值是最可能的重组分数。若 lod 值 ≥ 3 ,一般认为有连锁, ≤ 2 则认为无连锁。虽然最终的分子分析将物理定位这一基因,但连锁分析极其重要^[8]。lod score 方法需要知道疾病位点的参数,如遗传方式、基因频率和外显率,参数估计错误会降低检验效能。故近年越来越多地应用非参数方法,如患病同胞对、患病亲属对、传递不平衡检验(transmission/disequilibrium test, TDT)等。TDT 法是分析异质性的父母某些标记等位基因是否比其它更可能被传递到患病子女^[9]。

随着 DNA 技术的发展以及多态分子标记的发现,近年来开始用全基因组扫描的方法来进行疾病基因定位,这种方法需大量的分子标记,通常多于 300 个位点,但该方法大大提高了发现疾病基因的机会。

六、遗传与环境的相互作用:基因控制机体对环境的敏感性,这一现象在生物界很常见。遗传与环境相互联系,基因可能影响个体对高危环境的选择、修饰机体对环境因子的易感性。例如,基因可控制人体对饮食钠的高血压反应、对环境毒物的降解效应、长期吸烟对肺功能的影响、对油脂和胆固醇消费改变的血浆脂质反应^[10]等。一些功能性多态基因的不同表达可能需要环境的影响,或因多态性影响代谢途径,而后者又依赖于环境所提供的底物的可

获得性,或因转录环境信号给基因的转录因子与多态调节顺序的相互作用不同^[11]。如高密度脂蛋白(HDL)与冠心病有显著负相关。胆固醇酯转移蛋白(CETP)可促进 HDL 的转化。因此 CETP 活性增加使冠心病的危险性增加。而酒精消费对 CETP 活性可能是一重要调节因素。饮酒者 CETP/TaqIB 多态等位基因(B2)与血浆 HDL 胆固醇呈强正相关。在 B2B2 纯合子的个体,酒精对心肌梗塞有显著的保护作用^[11]。

遗传流行病学在肿瘤^[7]、精神疾病^[8]、心血管病^[11]、糖尿病等复杂疾病领域得到了广泛应用。以下列举了遗传流行病学方法在肿瘤和精神疾病研究中的应用。

在肿瘤和精神疾病中的应用

一、肿瘤遗传流行病学:肿瘤是由于自发的或环境暴露引起的生殖细胞(可遗传)或体细胞内基因突变,导致 DNA 水平的损伤引起的^[12]。肿瘤遗传流行病学的主要目标是解决遗传突变在何种程度上与肿瘤的发生和发展有关。临床经常可见患者的家系成员患相同或不同部位的肿瘤,患者一级亲属的肿瘤发病机率一般增加 2~3 倍。大多数双生子研究发现 MZ 比 DZ 的疾病一致率高。

家系成员除了共享遗传物质,许多行为和环境特征也可能相似,因此一些研究用 logistic 回归方法分析遗传和环境变量的贡献。多数发现一些家系的肿瘤主要与遗传因素有关。如当控制了每个家系成员的年龄、性别、职业和酒精、吸烟消费后仍然发现肺癌患者的亲属肺癌发病率升高^[13]。多数肿瘤在控制了年龄后,家族史是最重要的危险因素。危险性大小与患病亲属数和发病年龄呈函数关系。家族性患者的发病年龄更早。

对结肠癌、乳腺癌、前列腺癌、肺癌、卵巢癌以及软组织肉瘤的分离分析结果大多提示存在一或多个稀有的常染色体显性遗传基因的作用。基因频率多在 0.002~0.006 之间,即归于这些基因的病例占所有病例的 5%~10%,剩余为散发,终身外显率高于 60%。归于这(些)基因的病例部分随发病年龄的增加而降低。连锁分析成功的例子是乳腺癌和卵巢癌基因的发现。1990 年报道乳腺癌与 17q21 连锁^[14],继而发现了位于 17 号染色体的 BRCA1^[15],证实了分离分析的结果。遗传与环境相互作用研究发现,停经前有 ERBB2 过表达的乳腺癌患者多在年轻时

服过口服避孕药; CYP2 基因与吸烟相关的组织学类型的肺癌有关; 卵巢癌家族史与口服避孕药相关。

二、精神疾病遗传流行病学: 某些精神疾病有家族聚集现象。双生子研究发现焦虑和抑郁症存在共同的遗传成分, 且可解释 30% 的表型方差。用多变量分析法对抑郁、焦虑、恐惧、恐慌、贪食、酒精中毒等六种精神疾病的研究发现, 一组遗传成分主要与抑郁和焦虑症有关, 另一组与恐惧、恐慌、贪食有关。酒精中毒的遗传成分虽然明显, 但很大程度上是疾病特异的。精神分裂症 MZ 和 DZ 的疾病一致率分别为 53% 和 15%。与养子的非血缘亲属或对照养子的血缘亲属和非血缘亲属相比, 精神分裂症患者的血缘亲属发生该病的风险升高。血缘父亲嗜酒者比非嗜酒者的血缘养子发生酒精中毒的风险显著升高。

精神分裂症和躁狂抑郁症是否有单主基因效应还有争议, 一些证据更支持多基因遗传而不是单主基因遗传, 且至少 3 个位点与精神分裂症有关。很可能有多个微效基因, 稀有主基因存在于某些高家族性病例中。连锁研究成功的将早发性家族性早老性痴呆 (Alzheimer 病) 的主基因定位于 21 号和 14 号染色体, 这依赖于该病的多代连续遗传、显性基因和高外显率。躁狂抑郁症的候选区域在 Xq26~28 和 18 号染色体的着丝粒周区, 精神分裂症在 6 号染色体短臂。

群体相关分析发现早老性痴呆与阿朴脂蛋白 E (ApoE) 的 ε4 等位基因相关, 酒精中毒与东方人群醛脱氢酶 2 (ALDH2) 低活性形式有负相关。十年前, 多数研究只是估计精神疾病的遗传方式, 现在开始研究基因与环境的相互作用、家族性环境的性质、家族因素与社会和其他环境因素在发病中的相互关系^[10]。考察遗传因素与社会事件对抑郁症的相互作用, 发现一年内的不利生活事件是家族性的, 遗传和环境因素各占 20%。抑郁症患者的亲属生活事件发生率高, 可能由于共享家族因素。遗传因素加强了不幸生活事件引起抑郁的作用。

结 语

遗传流行病学同时研究遗传和环境危险因素, 尤其是致病过程中可能的基因-环境间的交互作用, 今后将继续致力于揭示疾病的遗传特征、发现复

杂疾病的遗传基因、研究遗传与环境的相互作用, 以识别高发病风险的个体, 通过控制危险因素制定基因型特异的预防和治疗策略。

参 考 文 献

- 1 Morton NE. Genetic epidemiology. *Ann Hum Genet.* 1997, 61:1.
- 2 Morton NE. *Outline of Genetic Epidemiology.* New York: Karger, 1982:1-5.
- 3 Morton NE, Chung CS. *Genetic Epidemiology.* New York: Academic Press, 1978:ix-x.
- 4 Friedman JM. Genetics and epidemiology, congenital anomalies and cancer. *Am J Hum Genet.* 1997, 60:469.
- 5 沈福民, 孟炜. 群体遗传学与遗传流行病学. 见: 沈福民主编, 遗传医学. 上海: 上海医科大学出版社, 1994: 127-143.
- 6 胡应, 沈福民. 遗传流行病学概述. 见: 耿贯一主编, 流行病学. 第三卷. 第 2 版. 北京: 人民卫生出版社, 1996:669-683.
- 7 Claus EB. The genetic epidemiology of cancer. *Cancer Surveys.* 1996, 25:13.
- 8 Sham P. Genetic epidemiology. *Br Med Bul.* 1996, 52: 408.
- 9 Spielman RS, McGinnis RE and Ewens WJ. Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet.* 1993, 52:506.
- 10 Kendler KS. Genetic epidemiology in psychiatry. *Arch Gen Psychiatry.* 1995, 52:895.
- 11 Cambien F. Insight into the genetic epidemiology of coronary heart disease. *Ann Med.* 1996, 28:465.
- 12 Stazi MA, Sampogna F. Genetic epidemiology: a powerful approach for the comprehension of cancer etiology. *Ann Ist Super Sanita.* 1996, 32:167.
- 13 Sellars TA, Elston RC, Stewart C, et al. Familial risk of cancer among randomly selected cancer probands. *Genet Epidemiol.* 1988, 5:381.
- 14 Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990, 250:1684.
- 15 Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science.* 1994, 266:66.

(收稿: 1998-03-25 修回: 1998-04-15)