

文章编号: 1004-0374(2004)04-0259-02

· 新思维 ·

复杂疾病 / 性状的基因定位

徐海明, 朱 军*

(浙江大学生物信息学研究所, 杭州 310029)

影响人类健康的主要是一些多发性的复杂疾病, 如肥胖、哮喘、高血压等, 这类复杂疾病相关性状的表型没有明显的孟德尔遗传模式, 多表现为连续的数量性状变异, 遗传机理较为复杂, 受多基因与环境的协同调控, 在医学上较难进行明确的诊断。数量性状基因座(quantitative trait loci, 简称 QTL)是染色体上影响性状表型变异的特定区段。随着 DNA 分子标记技术的发展和分子标记连锁图谱的建立, 已经能够把疾病 / 性状的表型变异剖解为归因于单个 QTL 的作用。这些分析可采用基于分子标记的数量性状基因定位的各种方法。通过对疾病相关基因的定位, 可以揭示控制疾病发展的各个基因位点, 了解致病的遗传学基础, 对于复杂疾病则还可明确不同位点间、各位点与环境间的相互作用, 阐明复杂疾病的发病机理, 为今后进一步定位、克隆创造条件。

连锁分析是疾病基因定位的主要分析方法, 其基本原理是通过分析两个遗传位点在家系中的共分离性来确定控制疾病表型变异的基因位点, 主要有参数和非参数两类方法。参数方法需要假定性状的遗传模式以及有关参数的值, 如外显值、重组率等, 然而对大多数疾病, 很难估计其孟德尔遗传的有关参数。非参数方法不需要假设性状遗传的模式, 但完全依赖于已知标记的遗传模式。连锁分析的功效又与个体疾病表型的正确诊断有关。因此, 特别适用于受单基因控制和表型存在明显差异的简单孟德尔疾病, 如高胆固醇血症(hypercholesterolemia)、亨廷顿氏症(huntington's disorder)、囊肿性纤维化(cystic fibrosis)。根据人类突变数据库信息, 至今已有 1 600 多个孟德尔疾病基因被成功定位, 这仅

仅是估计的人类基因的 3%, 也远少于需要定位的孟德尔疾病基因数。现有的连锁分析方法, 大都假定疾病性状只受单个基因的控制, 分析一个或多个标记位点与单个疾病基因的遗传关系。由于疾病基因的不完全外显性和病因的异质性以及各种复杂的因素, 因此都不适合于复杂疾病的基因定位与研究。

连锁分析需要利用配子重组的信息, 因此, 具有较低的解析率, 定位的位点一般在 1~10cM 之间。另一类基因定位方法是连锁不平衡(linkage disequilibrium, LD)分析方法, 这类方法不需要估计突变位点与标记位点间的重组率, 而是利用了人类在长期进化过程中累积的重组信息, 因此, 具有较高的解析率, 多用于基因的精细定位。尽管位点间的关联可能起因于 DNA 上某一个位点等位基因的突变、等位基因频率不同的多个群体的混合、奠基者群体的影响、随机的遗传漂移或直接的生物学进化, 但经过若干世代后, 相距较远(>1cM)的位点间的连锁不平衡很快就会消失, 因此, 显著的连锁不平衡体现了位点间的紧密连锁。连锁不平衡方法需要稠密的标记图谱, 一般用于基因的精细定位。对于连锁方法所得的疾病位点, 可以用 LD 方法进行进一步的精细分析。分子标记的突变会降低 LD 分析的功耗, SNP 标记(single nucleotide polymorphism)具有极低的突变频率, 而且人类染色体具有丰富的 SNPs, 是进行 LD 分析的最佳遗传标记。随着人类基因组研究的进展, SNP 图谱得到不断的加密, 为 LD 方法提供了广阔的应用前景。LD 方法的结果一般存在较高的假阳性, 不宜在全基因组进行疾病基因的检测。通过构建 SNP 单体型(haplotype), 分析单体型与性状表型的关联可以显著地提高基因定位

收稿日期: 2004-07-28

作者简介: 徐海明(1969 -), 男, 博士生, 讲师; 朱 军(1949 -), 男, 教授, 博导。有关详细的信息可检索网站 <http://www.cab.zju.edu.cn/ics/faculty/zhujun.htm>, *通讯作者。

的功效。无论是LD方法或基于单体型的关联作图方法,只适用于分析性状表型具有离散特征的人类疾病,而且每次都只能检测单个疾病基因位点,不能分析控制复杂疾病性状变异的基因间的互作,以及基因与环境间的互作。

与人类复杂疾病相关的、用于疾病诊断的生理生化性状多为数量性状,表现为连续的变异,性状的变异往往受众多基因与环境的共同调控,相互间又存在一定程度的互作。由于受伦理学的限制,人类不能像动植物群体那样进行有控制的交配试验,只能针对不同人群进行疾病相关性状的科学调查,因此样本的不同个体间会存在复杂的环境差异,这种个体间的环境差异以及基因与环境间的互作会极大地干扰QTL的分析,降低基因定位的功效。控制复杂疾病/性状的基因遗传效应又可分解为加性、显性、上位性效应,基因与环境的互作效应也可作相应的分解,这样的分解是研究复杂疾病/性状遗传机理的关键。复杂疾病性状的基因定位需要发展新的统计分析方法。

混合线性模型可包括多项固定效应和随机效应,相应的统计分析方法特别适合于复杂遗传现象的分析和非平衡数据的处理。浙江大学生物信息学研究所于1998年提出了适用于植物群体的基于混合线性模型的QTL定位分析方法^[1],在随后的几年中,针对植物QTL定位的各种交配群体(DH群体、回交群体、F₂群体、RIL群体),推导并给出了混合模型QTL定位的相关参数^[2],并研制了相应的计算机分析软件(QTLMapper),可以从Internet地址<http://www.cab.zju.edu.cn/ics/faculty/zhujun.htm>免费下载该软件,软件同时附有详细的使用手册(QTLMapperManual.pdf)。该软件能够分析QTL遗传主效应(加性、显性、上位性)以及QTL与环境的

互作效应(加性×环境、显性×环境、上位性×环境),提出的方法和软件已被国内外学者采用。Li等^[3]运用基于混合线性模型的基因定位方法分析了水稻单株粒重和单株生物量的自交衰退和杂种优势,得出一个重要结论:大部分与自交衰退以及杂种优势有关的QTL都与其他QTL存在上位性互作。Chen等^[4]运用同样的方法定位了水稻和大麦的疾病抗性位点,并讨论了这些疾病抗性位点在水稻和大麦之间的共线性关系。浙江大学研究复杂疾病或性状的基因定位,将发展人类复杂疾病的多基因遗传模型和相应的统计分析方法,能够无偏分析各个基因的遗传主效应、基因间的上位性效应,以及这些效应与环境的互作效应,并研制开发相应的计算机分析软件。在此基础上,进一步分析我国人群中哮喘和大肠癌复杂疾病/性状的表现型变异和分子标记遗传信息,研究中国人群哮喘病、大肠癌等复杂疾病的相关基因的相互作用,确定新的相关基因位点,估算复杂疾病或性状的上位性遗传效应和QTL与环境的互作效应。

[参 考 文 献]

- [1] 朱 军. 复杂数量性状基因定位的混合线性模型方法[A]. 王连铮, 戴晋瑞. 全国作物育种学术讨论会论文集[C]. 北京: 中国农业科技出版社, 1998, 11-20.
- [2] Wang D L, Zhu J, Li Z K L, et al. Mapping QTLs with epistatic effects and QTL-environment interactions by mixed model approaches. *Theor Appl Genet*, 1999, 99: 1255-1264
- [3] Li Z K, Luo L J, Mei H W, et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. *i. biomass and grain yield*. *Genetics*, 2001, 158 (4): 1737-1753
- [4] Chen H L, Wang S P, Xing Y Z, et al. Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to *Pyricularia* in rice and barley. *Proc Natl Acad Sci USA*, 2003, 100 (5): 2544-2549

· 情报服务 ·

中国科学院上海生命科学信息中心生命科学图书馆咨询研究部业务工作简介

生命科学图书馆咨询研究部主要提供以下服务: 科技管理决策咨询服务、科研进展咨询服务、科技动态跟踪服务、定题服务、专题数据库建设服务、生命科学学科信息门户网站资源导航服务。其中, 定题服务是根据科研人员的特定需求, 为其提供最新的相关文献信息, 以便及时掌握和了解某专题的国际、国内最新研究进展情况。对于长期用户, 还可为其一直追踪某方面的最新研究, 定期提供相关信息或文献调研报告。我们开展集团和个人的咨询业务。目前主要跟踪的专题有: 干细胞、艾滋病、SARS、空间生命科学与技术等。如果您需要以上方面的信息服务, 请与我们联系:

联系人: 高柳滨 电话 021-64334724 E-mail: gaolb@sibs.ac.cn 邮编 200031

联系地址: 上海市岳阳路319号 中科院上海生命科学信息中心生命科学图书馆咨询研究部