高校生物学教学研究（电子版）2012年3月，2(1)：20–23
ISSN 2095-1574　CN 11-9307/R　bioteach.hep.com.cn
DOI 10.3868/j.issn 2095-1574.2012.01.004

遗传重组率在复杂疾病基因定位中的应用及其拓展

蒋善群(✉)，查向东，尹若春

安徽大学生命科学学院，合肥，230039

摘　要：基因定位是遗传学研究的重要内容，也是遗传学教学的重点和难点。本文着重阐述定位人类复杂疾病易感基因的关键方法，包括连锁分析、关联分析、候选基因筛查、全基因组关联分析以及全基因组外显子深度测序等方面，系统地深入地探讨方法的原理及其在复杂疾病的基因定位中的应用。帮助学生深入理解复杂疾病的机理及理论研究方法，为以后从事复杂疾病的遗传学研究奠定理论基础。

关键词：复杂疾病，基因定位，连锁分析，关联分析

Application and Extension of Recombination Frequency in Gene Mapping of Complex Diseases

JIANG Shan-qun(✉), ZHA Xiang-dong, YIN Ruo-chun

School of Life Science, Anhui University, Hefei 230039, China

通过测定连锁基因之间的重组频率进行基因定位，是经典遗传学的重要内容，也是遗传学教学的重点之一。1913年C.B. 布里奇斯首先在果蝇中通过X染色体的不离开现象证实了白眼基因是在X染色体上。同年A. H. 斯特蒂文特根据两个基因之间的距离愈远则交换频率愈高这一假设，首先在果蝇中进行了基因定位工作。1945年E.B. 刘易斯在果蝇中发现与中胸发育有关的几个基因相邻接，构成一个复合座位或称为基因复合体或拟等位基因系列。1960年J. 莫诺和F. 雅各布报道大肠杆菌中与乳糖发酵有关的几个基因紧密连锁，构成一个操纵子。可见基因的位置并不是和它们的功能完全无关的，因此基因定位有助于了解基因的功能。

收稿日期：2012–01–05；**修回日期：**2012–01–25
基金项目：安徽大学211三期教学质量工程（39020013）；安徽大学人才队伍建设项目（02203104）
通讯作者：蒋善群，E-mail: shanqunjiang@gmail.com

1　复杂疾病中的基因定位：连锁分析

常见大多数疾病不符合孟德尔遗传，被称为复杂疾病[1]。这些疾病是由多成分引起的，且每一种成分对疾病整体风险贡献较小。其中一些成分可能来自遗传因素，而另一些成分可能是非遗传的环境因素，如年龄或吸烟。调节复杂疾病风险的遗传变异可能增加对疾病的易感性。

连锁分析（linkage analysis）是基于家系研究的一种方法，是单基因遗传病定位克隆方法的核心。它利用遗传标记在家系中进行分型（genotyping），再利用数学手段计算遗传标记在家系中是否与疾病产生共分离，从而计算重组率。基因定位的连锁分析是根据基因在染色体上呈直线排列，不同基因相互连锁成连锁群的原理，即应用被定位的基因与同一染色体上另一基因或遗传标记相连锁的特点进行定位。重组DNA和分子克隆技术的出现，发现了许多遗传标记——多态位点，利用某个拟定位的基因是否与某个遗传标记物存在连锁关系以及连锁的紧密程度就能将该基因定位

到染色体的一定部位，使经典连锁方法获得新的广阔用途，成为人类基因定位的重要手段。

连锁分析的目的是要找到最接近未知的致病遗传基因缺陷的标志物，从根本上构建遗传图谱。连锁分析的核心是计算（θ）值[2]。θ是重组率，用于定量描述两个基因位点的接近程度。如果θ是50%或以上，那么两个位点完全自由重组，因为它们是存在于不同的染色体或相距甚远相同的染色体上。在这种情况下，θ简单地表示为50%。重组率可用来计算比值得分的Log对数值或LOD值[2]。LOD值等于两个位点连锁的似然函数除以两个位点非连锁的似然函数。一个较大的LOD值显示两个位点紧密连锁，而负LOD值显示两个位点在减数分裂过程中非共分离。传统上，LOD值>3被认为是具有显著性水平。因此，每个标记物与疾病位点的距离可以被评估，从而含有目的致病基因DNA易感区域可以被缩小。通常，参数或非参数方法可以用于计算LOD值，这取决于某些参数的设置，包括对疾病的遗传模式、人群致病等位基因及标志物基因的发生频率[2]。

自从1980年全基因组连锁方法提出后，运用全基因组连锁分析和定位克隆的方法已鉴定了1 300多个导致人类疾病的基因，大多数为单基因遗传病。该分析依赖于大量的家系资料，适用于发现研究样本的主效基因，对独立家系的中效或微效基因的检出率较差，当λs（同胞对再发风险）< 4时会降低研究样本的连锁效果[3]。等位基因共占（allele sharing method）法[4]是观察受累同胞或家系成员间标记位点等位基因的共占情况，属于非参数连锁分析，包括受累同胞对（affected sib pair, ASP） 分析及家系成员（APM）分析。传统的连锁分析结果只能确定基因组内20~30 cM 的区域，通常还需基于连锁不平衡作进一步精细定位。

2 复杂疾病中的基因定位：关联分析

关联分析为非参数性分析，不需设定遗传方式等各种参数，并且连锁不平衡的检出力高于家系连锁分析。在多基因疾病中，不但可检出主效基因，还可检出相对风险率小于5.0%的次效基因，这正是同一位点相关分析阳性而连锁分析阴性的原因之一。

关联分析（association analysis）是确定与疾病表型变化相关的遗传变异的常用方法，在本质上属于病例—对照研究，用于比较无血缘关系的患病与非患病个体某一基因的等位基因出现频率的差异，从而确定与疾病关联的基因。关联研究推断，遗传变异在患病人群中比在正常人群更常见。在一定程度上，复杂的疾病是由一些列等位基因引起的，且可能是常见的（在人群中至少有1%的发生频率），但对疾病的易感性影响较小。在检测微效的易感基因方面，关联分析较连锁分析具有更大的把握度。但是，使用关联研究解析复杂的疾病也存在内在的挑战。检测易感基因研究的把握度取决于几个因素，主要是样本量大小，还有致病等位基因频率及其效应大小。此外还要考虑到如何使患者组与正常对照组相匹配，人群、地理和社会背景等。而在这些不同条件下，等位片段的频率往往有很大的差异，这一现象被称为群体分层（population stratification）。为克服这一问题，在研究方案的设计上必须注重病例组与对照组相匹配，对家系样本需增加患者父母未传递的等位片段作匹配比较。当某一特定等位片段在传递时出现的概率比随机的概率显著增多时，则认为存在连锁不平衡。目前有两种类型的关联研究：候选基因的研究和全基因组关联研究。

基于此原理的遗传统计方法有对隐性遗传模式非常有效的传递不平衡（transmission disequilibrium test，TDT），患者家系对照者分析（affected family-based controls，AFBAC）及单倍型相对风险率分析（haplotype relative risk，HRR）等方法。TDT是在家系内进行关联分析，观察双亲（至少一个是杂合子）将标记位点等位基因传递给患者的频率。TDT的优点有：①可完全消除种族分层引起的误差；②可用于分析父母在基因传递上的差异。

3 候选基因筛查

候选基因筛查（candidate gene screening）的方法包括基因的选择及致病基因的多态性位点选择。可通过样本的病例–对照研究检验这些多态性位点与疾病的相关性。基于基因在疾病中的作用的假设使得相关性分析更值得可信。基因的选择是以前期实验数据或理论机理为基础的。易感基因位点可能会出现在DNA内含子区段，但就目前的知识而言还无法知晓哪一个遗传变异可以有效预测疾病的发生风险。因此，候选基因位点的选择往往是首先考虑从基因内部编码序列或启动子区域入手。

但是，即使候选基因与疾病相关，结果也必须在

不同的人群中进行重复，因为我们对复杂疾病的机制知之甚少，因果关系的假设可能不支持。

4 全基因组关联分析

关联研究是基于“常见疾病，常见变异”（common disease, common variant）的假设, 其基本原理是：扫描整个基因组上的数千标记物（SNP），采用病例和对照的大样本检测并揭示它们可能和疾病的相关性。这将意味着这些标记物可能与易感基因处于连锁不平衡（LD），距离更接近，因此结果可将易感区域精细定位在包含一个或几个感兴趣的基因范围内。目前，国际间协作的Hapmap计划正是基于此而开展，使在大规模人群中对全基因组的标签SNP进行基因分型成为可能[5]。

对于复杂疾病而言，全基因组关联分析（genome-wide association study, GWAS）对于每个微效基因的鉴别较全基因组连锁分析更有效[6]。以单倍型为基础的分析得出的结论将明显优于单个标记的研究[7]。全基因组关联分析在隔离群体中具有更强的检测效力[8]。在过去的5年里，各国科学家开展了对不同疾病的GWAS，涉及肿瘤、心血管系统疾病、内分泌系统疾病、胃肠道疾病、肝脏疾病、眼科疾病、神经精神类疾病、风湿病、皮肤病以及感染性疾病等领域, 并取得了重大成果。GWAS在复杂疾病遗传学研究中取得的成果加深了我们对这些复杂疾病遗传学基础的认识, 为后续对其发病机制研究提供启示, 也为将来的基因诊断和个体化治疗奠定了理论基础。全基因组关联分析在今后将有广阔应用前景。

5 全基因组外显子深度测序

自2005年以来，尽管利用GWAS对多种常见疾病进行了研究，发现和重复验证了近2 000 个SNPs或位点，其中包括以前未检测到的而与疾病密切相关的基因及部分未知基因[1]。但是由于GWAS的结果存在假阳性、假阴性、检测到的单核苷酸多态性很少位于功能区以及对稀有变异和结构变异不敏感等问题, 导致了其应用的局限性。而新一代测序技术的进步, 促进了全基因组测序和全基因组外显子测序的快速发展, 为解决上述问题提供了契机。全基因组外显子深度测序（genome-wide exon deep sequencing）是利用序列捕获技术将全基因组外显子区域DNA捕捉并富集后进行高通量测序的基因组分析方法。目前已有成功利用全基因组外显子测序的方法结合或不结合连锁分析结果研究复杂疾病的案例，例如：Jones等[9]利用外显子组测序研究对卵巢透明细胞癌进行了深入研究，鉴定出4 个在至少两例肿瘤中发生过突变的基因，*PIK3CA*、*KRAS*、*PPP2R1A*和*ARID1A*。其中，*ARID1A*和*PPP2R1A*是新发现的，前者是致癌基因，后者是抑癌基因。随后对这4个基因在42 个病人中进行验证, 结果发现57%的患者中*ARID1A*基因发生了突变。由于*ARID1A*基因编码了染色质重塑中的关键蛋白, 因此推断染色质重塑异常可能与卵巢透明细胞癌的发生有关。Bowden等[10]利用安捷伦SureSelect外显子捕获系统和Genome Analyzer Ⅱx 系统对血浆乙二腈水平无显著性差别的2个家系中的3个患者进行测序，发现ADIPOQ基因的低频突变（1.1%）G45R，能解释17%的西班牙裔美国人的血浆乙二腈水平，63%的家族存在该突变；Bilguvar等[11]运用NimbleGen 2.1M 芯片捕获外显子和Genome Analyzer Ⅱ 测序系统对1例患者测序，发现*WDR62*基因与脑皮质发育异常疾病相关。

目前, 全基因组外显子测序已在孟德尔疾病或罕见综合征的研究中取得了重大突破, 然而在复杂疾病的研究中还刚刚起步，在肿瘤方面的研究较多。由于外显子是与疾病及表型相关的最具特征性的区域, 并且迄今为止较难评价非编码区域对疾病的影响, 所以在全基因组测序费用居高不下的今天, 全基因组外显子测序仍然不失为一个很好的选择。

参考文献

[1] Swaroop A, Branham KE, Chen W, et al. Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits [J]. Hum Mol Genet, 2007, 16 (2): 174–182.

[2] Dawn Teare M, Barrett JH. Genetic linkage studies [J]. Lancet, 2005, 366 (9490) :1036–1044.

[3] Badn er J A, Gershon E S, Gold in L R. Optimal ascertainment strategies to detect linkage to common disease alleles [J]. Am J H um Genet, 1998, 63 (3) : 880–888.

[4] Holmans P. Affected sib pair methods for detecting linkage to dichotomous traits: review of the methodology [J]. Hum Biol, 1998, 70 (6) : 1025–1040.

[5] Con sortium T H. A haplotype map of the hum an genome [J]. Nature, 2005, 437 (7063) : 1299–1320.

[6] Risch N, Merikangas K. The future of genetic studies of complex

human diseases [J] . Science, 1996, 273 (5281) : 1516–1517.

[7] Zhang K, Calabrese P, Nordborg M, et al. Haplotype block structure and its applications to association studies: power and study designs [J]. Am J Hum Genet, 2002, 71 (6) : 1386–1394.

[8] Wright A F, Carothers A D, Pirastu M. Population choice in m app ing genes for comp lex d iseases [J] . Nat Genet, 1999, 23 (4): 397–404.

[9] Jones S, Wang TL, Shih IeM, et al. Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma [J]. Science, 2010, 330 (6001) : 228–231.

[10] Bowden DW, An SS, Palmer ND, et al. Molecular basis of a linkage peak: exome sequencing and family–based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study [J]. Hum Mol Genet, 2010, 19 (20) : 4112–4120.

[11] Bilgüvar K, Özturk AK, Louvi A, et al. Whole-exome sequencing identifies recessive *WDR62* mutations in severe brain malformations [J]. Nature, 2010, 467 (7312) : 207–210.

（责编　孟丽）