

XIANDAIJIYINCAOZUOJISHU

主编·胡福泉

现代

基因操作技术

人民军医出版社

现代基因操作技术

XIANDAI JIYIN CAOZUO JISHU

主 编 胡福泉

副主编 杨 恬

编 者 (以姓氏笔画为序)

万泽生 甘立霞 向明明 刘 昕

李渝萍 杨 恬 余 全 张放鸣

陈忠民 周建新 胡福泉 倪 兵

绘 图 宋文强

人 民 军 医 出 版 社

北 京

图书在版编目(CIP)数据

现代基因操作技术/胡福泉主编. —北京:人民军医出版社,2000.10
ISBN 7-80157-111-8

I. 现… II. 胡… III. 基因-分子生物学-实验-技术 IV. Q343

中国版本图书馆 CIP 数据核字(2000)第 21156 号

人民军医出版社出版

(北京市复兴路 22 号甲 3 号)

(邮政编码:100842 电话:68222916)

潮河印刷厂印刷

春园装订厂装订

新华书店总店北京发行所发行

*

开本:787×1092mm 1/16 • 印张:23.75 • 字数:544 千字

2000 年 10 月第 1 版 2000 年 10 月(北京)第 1 次印刷

印数:0001~5000 定价:39.00 元

(购买本社图书,凡有缺、倒、脱页者,本社负责调换)

内 容 提 要

本书针对国内医学分子生物学实验研究的实际需要,详细介绍了基因操作技术,包括基因分离与基因作图、PCR 条件最佳化及特殊 PCR、PCR 产物的克隆、突变与重组技术、对已知序列两旁未知序列的克隆方法、cDNA 文库的构建与筛选、用差示法和扣除法做 cDNA 分析与克隆、寡核苷酸启动的原位 DNA 合成技术、原位多聚酶链反应技术等。内容先进,实用性强,可供临床和基础医学各专业从事分子生物学实验研究的工作人员参考学习。

责任编辑 姚 磊

前 言

基因操作是一种分子生物学技术,但不等于分子生物学技术,后者有着更宽的范围。然而本书所涵括的内容都是涉及“基因”的技术。本书之所以选择这种侧重点,是因为目前市场上介绍分子生物学一般性原理及通用性技术的书籍已有不少,但涉及基因操作的最“新”技术、更“专”技术的书籍则不多,如本书中所介绍的基因分离与基因作图技术、基因突变(点突变、插入、删除、嵌合与重组)技术、PRINS (oligonucleotide PRimed IN Situ)技术等一般在分子生物学技术书籍中还鲜有涉及。这些技术常常分散在国外一些研究论文或最新专著中,而且这些新技术发展又是层出不穷、日新月异的。为了跟踪这些新技术而且也是成熟的技术,我们邀约了几位新近从国外留学归来或目前还在国外留学的学者——他们都是从事分子生物学实际工作第一线的中青年——编成了此书。希望这本书能够成为广大分子生物学工作者操作案头的一本有用的工具书。

本书内容的特点概括有三:一是在选题、取材方面力求“新”与“专”,而不求面面俱到;二是对于每一种新技术,力求首先将其原理交代清楚,对于大部分技术绘有原理图,以便易于或加深对原理的理解;三是每一技术都独立成为一章,每一章包括了引言(介绍原理)、材料、方法和参考文献四部分,在使用每一技术时,不用参考其它章节,抽出本章就可成为案头操作方案,参照方案即可完成实验。

由于本书所涉及的内容很新,又限于我们的知识有限,所以不可避免地存在不少缺点,甚至错误。我们在每一章中列出了参考文献即是为了弥补这种可能,一旦有错时,便于读者参考原文,但更希望的是读者不吝赐教,请将意见寄主编(重庆市第三军医大学微生物教研室,邮政编码 400038),多谢。

编 者

2000 年 6 月

目 录

第一篇 基因分离与基因作图

第 1 章 基因定位:从 FISH 到序列数据库	1
第 2 章 遗传性状的连锁分析	4
第 3 章 用体细胞杂交进行基因作图	9
第 4 章 用 FISH 技术进行基因作图	12
第 5 章 用 FISH 进行 DNA 位点排序与测距	22
第 6 章 基因物理作图——脉冲场凝胶电泳法	27
第 7 章 疾病基因间区 YAC 重叠群的构建和应用	38
第 8 章 粘粒重叠文库的构建与应用	45
第 9 章 用二核苷酸多态性分析作物理图	52
第 10 章 用多重 PCR 作表达序列标签图谱	57
第 11 章 用外显子捕获法分离基因	61
第 12 章 用直接选择法从基因组分离编码序列	71
第 13 章 用 YAC 杂交筛选法分离 cDNAs	79
第 14 章 差异表达基因的检测与分离	84
第 15 章 化学交联扣除法	89
第 16 章 基因制图、基因分离与数据库访问	96

第二篇 PCR 条件最佳化及特殊 PCR

第 17 章 PCR 基本原理及关键问题	109
第 18 章 Touchdown PCR 及 Stepdown PCR——一步法达到条件最佳化的 PCR 技术	116
第 19 章 长靶序列的 XL PCR 扩增	120
第 20 章 用 Tub DNA 聚合酶作长序列 PCR 扩增	125
第 21 章 高 GC 含量模板的扩增——PCR 反应中的溶剂效应	129
第 22 章 RT-PCR 克隆 cDNA 大片段	132

第三篇 PCR 产物的克隆

第 23 章 用 T4 DNA 聚合酶生成可克隆 PCR 产物	135
第 24 章 PCR 产物的非酶连接法快速克隆	138
第 25 章 PCR 产物的 T/A 克隆法	142
第 26 章 T-载体的构建	146
第 27 章 用 AT-接头做 PCR 产物的修饰和定向克隆	152

第四篇 突变与重组技术

第 28 章	用重组 PCR 做重组和定点突变·····	157
第 29 章	DNA 的体外重组与突变·····	161
第 30 章	用 PCR 法做插入、删除和嵌合连接·····	165
第 31 章	用 Megaprimer PCR 进行基因突变和基因融合·····	168
第 32 章	快速高效的突变方法——PCR-管法·····	173
第 33 章	用耐热连接酶及 PCR 生成突变分子·····	176
第 34 章	用 PCR 突变接头探查基因转录调控原件·····	179
第 35 章	用翻转 PCR 进行序列倒转·····	182
第 36 章	用 GB-D 聚合酶及平板快速筛选法做定点突变·····	185
第 37 章	用 SELEX 组合化学法分离高亲和性的核酸配体·····	190

第五篇 对已知序列两旁未知序列的克隆方法

第 38 章	cDNA 末端的快速扩增·····	197
第 39 章	用单侧特异性引物扩增基因的调控区序列·····	201
第 40 章	末端修饰 PCR 扩增相邻未知 DNA 序列·····	205
第 41 章	用 mRNA 5'端加上一段确定序列用于克隆全长 mRNA·····	212
第 42 章	用 Step-Out PCR 在 DNA 克隆中做快速定向步移·····	220
第 43 章	反向 PCR 快速扩增 cDNA 末端·····	224
第 44 章	用锚定 PCR 方法从基因文库中快速扩增基因末端·····	228
第 45 章	用外显子扩增法从酵母人工染色体中分离基因编码序列·····	231

第六篇 cDNA 文库的构建与筛选

第 46 章	从少量细胞制备 cDNA 文库·····	239
第 47 章	重组 DNA 文库的非放射性 PCR 快速筛选法·····	245
第 48 章	用 PCR 筛选 cDNA 文库·····	249
第 49 章	噬菌体 DNA 亚库的构建及 PCR 筛选·····	252
第 50 章	用简并引物 PCR 克隆基因家族中的成员基因·····	257
第 51 章	用含肌苷的简并引物扩增分离低相似性基因·····	263

第七篇 用差示法和扣除法做 cDNA 分析与克隆

第 52 章	用分子选择法对 cDNA 序列表现度作归化处理·····	267
第 53 章	用磁珠技术和 PCR 做扣除 cDNA 克隆·····	274
第 54 章	扣除 cDNA 的 PCR 再生·····	283
第 55 章	用 PCR 对 cDNA 库进行差示筛选·····	290
第 56 章	用 DDRT-PCR 鉴定和克隆差异表达的基因·····	295

第八篇 寡核苷酸启动的原位 DNA 合成技术

第 57 章	寡核苷酸启动的原位 DNA 合成·····	301
--------	-----------------------	-----

第 58 章	特异染色体的 PRINS	305
第 59 章	染色体寡核苷酸 PRINS 标记 DNA 的亮视野显微镜检测	308
第 60 章	PRINS 方法分析精细胞的非整倍性	314
第 61 章	冰冻组织切片上 PRINS DNA 合成	318
第 62 章	多序列寡核苷酸启动原位 DNA 合成	322
第 63 章	在伸展的染色质样品上进行寡核苷酸 PRINS DNA 合成	325
第 64 章	免疫组化-PRINS 联合应用同时检测细胞表型和基因组参数	328
第 65 章	染色体 PRINS DNA 标记与间接免疫细胞化学的联合应用	332
 第九篇 原位多聚酶链反应技术		
第 66 章	原位多聚酶链式反应技术及其应用	337
第 67 章	直接原位单拷贝 PCR	346
第 68 章	在细胞和石蜡切片上检测低拷贝 DNA 序列的原位 PCR 方法	350
第 69 章	冰冻切片上的逆转录原位 PCR	356
第 70 章	HIV-1 前病毒 DNA 的原位 PCR 检测及流式细胞仪分析	360
第 71 章	细胞内 mRNA 的原位 PCR 扩增	364
第 72 章	组织切片上核酸序列的 PCR 扩增——定位原位扩增	369

第一篇 基因分离与基因作图

第1章 基因定位:从 FISH 到序列数据库

一、确定和分离目标染色体

人类的每一个基因都定位在 23 对染色体之中,同一对染色体上所有的基因构成了连锁群。待克隆的目标基因,其表型和其所在的染色体上的其它基因、表型或遗传标记可以形成一定的连锁关系。根据这种关系,人们可以将目标基因定位于特定的一对染色体上。现在人类的每一对染色体上已有很多遗传标记被标定。这些遗传标记可以是已知的可以被观察到的表型,也可以是位置已知的 DNA 分子标记。分析目标基因与这些遗传标记的连锁关系就可以把目标基因定位在一个染色体的特定位置。目前对人类 23 对染色体的每一个带都已分离出一些分子标记,所以一旦一个目标基因被定位在某个染色体的一个特定的带上,利用这个带的 DNA 分子标记,筛选这一区域的候选基因,就可以把目标基因克隆出来。这种先将目标基因定位、再克隆的技术路线称为定位克隆。定位克隆的第一步就是要将目标基因定位在一个染色体上。常用的方法有家系调查法、体细胞杂交法以及以核酸杂交技术为基础的各种方法。这些方法各有自己的特点。家系调查是最古老的遗传学研究方法,其理论和方法都较为成熟,是基因定位中基础性的和不可缺少的工作。对某一致病基因来说,如能收集到信息量足够多的家系资料,就可把这一致病基因的表型与一些特定的遗传标记进行连锁分析。如果两个遗传标记独立分离的概率小于 $1/1\ 000$ 或 LOD 值大于 3,就认为它

们是连锁的。对于呈现明显孟德尔规律的单基因遗传病,连锁分析是定位其基因的有效方法。目前人们已经开始利用连锁分析的方法定位多基因遗传病的基因。

将人类体细胞和小鼠体细胞融合,可以形成杂种细胞。杂种细胞在传代过程中会随机丢失人体细胞的染色体,最后在一株杂种细胞中可能只含有几条或一条人的染色体,或是某个染色体的一部分。利用一套分别包含了人全部染色体的杂种细胞,就有可能把某个基因定位在特定的染色体上。如所有含人类 1 号染色体的杂种细胞都表现出肽酶 C 的活性;而不含 1 号染色体的杂种细胞均无肽酶 C 的活性,就可以把肽酶 C 的基因定位在 1 号染色体上,因为鼠的体细胞无肽酶 C 的活性。除了利用生化遗传标记外,还可利用核酸杂交和 PCR 的方法在这些杂种细胞中进行基因定位。为使基因的定位能更精确,还可用高能射线处理杂种细胞,使其中的染色体断裂成若干个片段,人们就可以利用遗传标记,将目标基因定位在染色体的某一区段内。杂种细胞还可用来进行遗传互补实验。

在基因的染色体定位中,还有一项十分重要的技术——染色体荧光原位杂交技术(fluorescence *in situ* hybridization, FISH)。FISH 是一项十分敏感的技术,可以检测出非整倍体、基因扩增、微小的染色体重排、易位等染色体变异。在肿瘤遗传学的研究中,

可以利用 FISH 来发现一些与肿瘤发生、发展相关的染色体变异,进而找出肿瘤相关基因。近年来人们还发展了比较基因组杂交(comparative genomic hybridization, CGH)

的技术,即用同一种生物异常细胞的基因组作探针,与正常细胞基因组杂交,通过对杂交信号的分析,就可以发现两者之间的不同,找出有关的基因。

二、基因组 DNA 大片段的克隆和物理图谱的构建

一旦确定了致病基因的染色体定位,接下来的工作就是采取从大到小,从粗到细的策略,克隆有关位点的基因组 DNA 大片段,构建其物理图谱,确定各种标记间的距离等。通过构建物理图谱,就可以确定两个遗传标记或位点间的实际距离。在这方面的工作中,脉冲场电泳(PFGE)结合稀有位点限制性内切酶是一个十分有用的方法。在可以表达的基因中常有一些富含 GC 的序列,这些序列称为 GpC 岛。在寻找基因时可将这些 GpC 岛当作基因的标志。而许多稀有限制性内切酶的识别位点就在这些 GC 岛中。PFGE 还可用来确定重叠群内克隆间重叠的区域以及比较正常和异常基因组之间的差异,用细胞遗传学的方法是无法检测到这些差异的。

在构建大尺度的物理图谱时,YAC 是必用的克隆载体,因为其克隆容量可以达到 1 Mb(0.3~1.2 Mb)。插入片段越大,所用的遗传标记就越少,用来组成重叠群的克隆也越少。如果有足够的遗传标记,也可选用 BAC、Pls 和 Cos 质粒。尽管它们的克隆容量较小,但它们的可操作性比较好,较为稳定,发生嵌合的现象也较少。分离出的基因组 DNA 大片段,可以进一步用序列标签位点(sequence tagged-sites, STSs)或重复序列指纹的方法来确定克隆的大小及重叠的程度。经过这样分析的克隆又可作为探针在染色体的有关位置作染色体步移,直至获得含目的基因的克隆。除了用来克隆目的基因外,这些 DNA 大片段还可用来分离分子标记,作 FISH 和其它杂交实验的探针等。

三、分离基因

用遗传连锁分析和物理图谱的方法可以将目的基因定位在 0.5~1 个厘摩的水平,大约相当于 1 Mb 的距离。在这样的范围内通常有 30~50 个基因。分离这些基因,从中找出致病基因是目前多数定位克隆工作的制约瓶颈。用哪种策略进行这些工作,取决于各个实验室的条件。克隆的策略可分为两种,一种是以杂交技术为基础而设计的,另一种是以基因组 DNA 序列中与 RNA 剪接有关序列的功能为基础设计的。

外显子捕捉法依据的是在内含子两端都有保守的接头序列设计的。这种方法不必去考虑结构基因表达的组织特异性和表达的水平。基因组的转录序列可以用 cDNA 探针池与基因文库阵列杂交的方法来分离,也可

用基因组 DNA 的克隆作探针,与 cDNA 文库杂交。采用 cDNA 探针的方法,可以把基因组 DNA 固定在膜上,也可以把基因组 DNA 生物素化或是进行液相杂交。这些方法需要有关目标基因表达水平的资料。这些基于杂交方法的技术通常获得的都是部分外显子的片段,这些片段可用作转录序列标签位点(expressed sequence tagged-sites, ESTs)或分离全基因的探针。

另一种分离目的基因的策略是利用在结构基因的 5'端通常都有 GpC 岛的特点设计的。对 GenBank 中 375 个基因的分析表明,百分之百持家基因(housekeeping gene)的序列中含有 GpC 岛,40%的组织特异表达基因的序列中也含有 GpC 岛。利用稀有限制性

内切酶,PCR 或探针杂交的方法就可以从基因文库中分离出含 GpC 岛的序列。同一种细胞在不同状态时会有不同基因转录,消减杂交和差异显示就是利用这一点发展起来的分离目的基因的技术。

对模式生物基因组的研究也会为人类基因的克隆提供非常有意义的信息。由于小鼠和人类都是哺乳动物,所以研究小鼠基因组对人类基因组的研究来说有很大意义,目前已可将人类基因组划分成 68 组与小鼠基因组同源的序列。

毫无疑问,用定位克隆方法克隆的基因,数目会与日俱增。并且随着克隆基因和分子标记数目的增加,基因的定位也会日益容易。生物信息学的发展,各种数据库的建立和内容的增加也将极大地帮助基因的克隆,目前已有人提出基于数据库的基因克隆。

(刘 昕)

参 考 文 献

- 1 Collins FS. Positional cloning moves from perdictional to traditional. *Nat Genet*, 1995;9:347—350
- 2 Walter MA, Goodfellow PN. Radiation hybrids; irradiation and fusion gene transfer. *Trends Genet*,1993,;9: 352—356
- 3 Strausberg RL, Feingold EA, Klausner RD, et al. The mammalian gene collection. *Science*, 1999;286:455—457
- 4 Collins FS. The human genome project and the future of medicine. *Ann N Y Acad Sci*,1999; 882:42—55
- 5 Chapple JP, Harcastle AJ, Kurzik—Dumke U et al. Assignment(1) of the neuronal cochaperone, HSJ1, to human chromosome bands 2q32—>q34 between D2S295 and D2S339 by in situ hybridization and somatic cell and radiation hybrids. *Cytogenet Cell Genet*,1999;86:62—63
- 6 Kallioniemi A, Kallioniemi OP, Pinkel D, et al. Comparative genomic hybridization; a rapid new method for detecting and mapping DNA amplification in tumors. *Semin Cancer Biol*,1992;4: 41—46
- 7 Berry R, Stevens TJ, Sikela JM, et al. Gene—based sequence — tagged — sites (STSs) as the basis for a human gene map. *Nat Genet*,1995; 10:415—423

第2章 遗传性状的连锁分析

(一) 复杂遗传现象

遗传异常按经典遗传学可分为单基因和多基因遗传病,其中单基因遗传病又可分为显性和隐性两类。还有许多遗传病存在复杂遗传现象。造成复杂遗传现象的原因有多种,如多基因遗传;基因型和表型之间的非共线性所导致的不准确的表型分类;外显不全或年龄依赖性外显;拟表型,遗传异质性和寡基因遗传等。多基因控制的遗传性状多是一些数量性状;外显不全或年龄依赖性外显是指获得显性致病基因的个体并不发病或发病的机会与年龄相关;拟表型指一个个体尽管没有获得致病基因,却由于环境因素或其它基因的突变而表现出该致病基因所致疾病的症状;遗传异质性指一组基因中的任意一个发生突变,都会产生相同的表型;寡基因指一个特定表型的出现是一组基因同时突变的结果。

(二) 复杂遗传现象分析方法

遗传分析的主要目的是判断一个性状形成的过程中有无遗传因素的参与,这些遗传因素作用的大小和彼此间的关系以及它们在基因组中的定位。由于存在基因型和表型之间的非共线性,使得复杂遗传模式的分析较为困难。近年来发展了一些分析复杂遗传模式的遗传流行病学方法。应用这些方法可以估计环境因子对表型的影响,多基因的存在及其相互作用等。

1. 家族聚集 尽管环境因素也会在一个家族中造成多个成员患病,但家族聚集现象仍可认为是遗传病的明显特征之一。由遗传因素造成的家族聚集,其发病个体的血缘亲属的发病风险要高于普通人群的个体。对于有家族聚集发病的群体来说,要精确估计发病个体血缘亲属的发病风险,群体的发病率和流行率,有赖于大规模的流行病学调查

和家系调查。

相对风险因数 λ_R 是一个衡量在家族聚集现象中遗传因素对疾病表型影响大小的参数。它是用呈现家族聚集发病家族中发病个体亲属的发病率除以普通人群中该疾病患者亲属的发病率。下标 R 指亲属的种类,如 λ_o 是指子女等后代亲属; λ_s 指兄妹等同胞亲属。有研究表明,伴有高 λ_s 的性状,如 λ_s 高于 10,其遗传定位较为容易;而 λ_s 如果低于 2,则该性状的遗传定位就比较困难。普通人群中,相对风险因数很低,而呈孟德尔方式遗传的性状,其相对风险因数非常高。

2. 孪生子研究 孪生子研究是人类和医学遗传学研究的重要方法。比较单卵双生和二卵双生数量性状的相关系数和质量性状的相似率,就可以得出遗传因素是否是该性状的形成的主要原因。如果一个性状在单卵双生间差异较小,而在二卵双生间差异较大,那么这个性状多半是由遗传因素决定的。

3. 分离分析 分离分析包括分离率的估计以及比较期望分离率和观察分离率的吻合程度等参数。运用分离分析方法分析按孟德尔方式遗传的性状,可以确定该性状的遗传方式。分析复杂遗传现象时所用的分离分析方法称为综合分离分析。这一分析方法可以用来判断在一个确定的家族中,某一性状是否是多基因遗传的,有无主基因,最有可能而参数最少的模型是什么等问题。在用 LOD 法进行连锁分析时也需要用分离分析方法获得的遗传模型。

分离分析对原始数据的偏差十分敏感,如果像进行一般的连锁分析那样,只选取有多个成员受影响的家系,往往会得出该性状呈孟德尔方式遗传的错误结论,也会高估该基因的频率和外显率。多位点分离分析在分析数量性状时是一个十分有用的工具。

(三) 参数连锁分析

1. 简介 连锁分析主要要阐明性状的遗传方式和决定特定性状的基因的染色体定位。目前,有两类不同的连锁分析方法,一种是基于家系资料的参数连锁分析,另一种是基于同胞对法的非参数分析。

连锁分析是基于这样一个事实:同一条染色单体上的基因是连锁遗传的,基因间的距离越长,同源染色体间发生交换、产生新的单倍体(重组)的可能性越大。在性细胞的减数分裂过程中,同源染色体发生交换的频率称为重组率(θ)。两基因座位的距离越远,重组率越高,若重组率超过 50%,表明这两个座位不在一个染色体上。因此依据两个基因

座位之间重组率的高低就可以估计它们之间在染色体上的距离,如果重组率在 1%($\theta=0.01$),则两个座位间的距离称为一个图距或一个分摩(cM)。通常只有两个座位的距离在 25 cM 以内时两座位间的距离才和重组率呈线性关系。

用统计学方法分析两座位是否连锁时,两个基本的假设是:两座位连锁和两座位不连锁(H_0 , null hypothesis, 无效假设)。对于第一个假设, $0 \leq \theta < 0.5$;对于第二个假设, $\theta=0.5$ 。参数连锁分析的主要内容是根据一个家系内两个座位重组率的观测资料,计算并比较两个假设的概率,得出两个座位连锁或不连锁的结论。

表 2-1 Treacher Collins 综合征与 D5S210 的 Lods 值

	重 组 率(θ)				
	0.01	0.05	0.10	0.20	0.30
家系 1	0.97	1.49	1.55	1.34	0.94
家系 2	0.42	0.39	0.36	0.28	0.20
家系 3	1.78	1.65	1.49	1.13	0.73
家系 4	1.19	1.11	1.02	0.82	0.59
家系 5	0.29	0.26	0.22	0.13	0.06
家系 6	0.59	0.54	0.47	0.32	0.17
家系 7	3.01	2.77	2.46	1.80	1.14
家系 8	0.29	0.26	0.21	0.13	0.06
总 计	8.54	8.47	7.78	5.95	3.89

在进行连锁分析时目前最常用的是 Lods 法,即对数优势记分法(log odds score method)。应用这一方法先要在每一个家系中计算两个座位不连锁($\theta=0.5$)的概率;然后计算两个座位以不同重组率相连锁的概率,重组值的取值范围通常在 0.01~0.5 之间;用两座位连锁的概率除以两座位不连锁的概率,就得到连锁优势概率比(odds ratio)。对连锁优势概率比取对数,就得到对数连锁优势分值(log odds score, Lods)。对于孟德尔性状来说,如果连锁的概率比否定连锁的概率高 1 000 倍或 1 000 倍以上,即连锁优势概率比 $>1\,000$ ($Lods > 3$),就可以认

为两座位是连锁的。表 2-1 是 8 个 Treacher Collins 综合征家系致病基因与 5 号染色体上的一个 DNA 多态标记 D5S210 的 Lods 值表。

从表 2-1 中数据可以看出,当重组率设定在 1%时,Lods 值达 8.54,表明 Treacher Collins 综合征的致病基因与 D5S210 紧密连锁,两者的距离在 1 cM 以内。

Lods 值的大小取决于家系的大小和连锁标记的信息量。一个标记的杂合度大于 0.7,就可以用于连锁分析。人类基因组的全长,男女性平均大约是 3 500cM,因此理论上用 350 个平均图距 10cM 的遗传标记就可以

进行全基因组的连锁分析。实际应用时需要根据标记的信息量、家系的大小和结构以及致病基因的遗传方式来确定标记的密度。

如果连锁分析的对象是两个遗传标记,则参数连锁分析可以得到十分可靠的结果,因为遗传标记在家系中以共显性方式遗传,并且其基因型和表型是一一对应的。对呈孟德尔方式遗传的性状进行参数连锁分析也可以获得较好的结果,但个别情况下需要考虑外显不全对家系资料的影响。有时在一组家系资料中只要有一个家系的 Lods 值大于 3,也可以得出明确的连锁的结论。

2. 复杂性状 在复杂性状的遗传过程中,由于表型和基因型的非共线性,使得参数连锁分析往往得出不正确的结论,所以参数连锁分析方法一般不用于复杂性状的连锁分析。如前所述,表型和基因型的非共线性主要由以下原因造成:对家系成员的误诊、不完全外显、拟表型、遗传异质性及寡基因遗传。尽管如此,在复杂性状的研究中,参数连锁分析并非毫无用处。已有一些应用参数连锁分析研究复杂性状获得成功的例子。注意以下几个方面的问题有助于提高参数连锁分析研究复杂性状的准确性。

在复杂性状的研究中,如果将外显率估计过高,就会丢失来自正常携带者的信息。在分析存在外显不全和年龄依赖性外显的性状时可以人为地将外显率定得很小,以便进行唯患病者分析。这样可以减少对正常携带者重组的错误估计,但会增加观察数量。还

可以通过变换诊断方法对不明确的表型进行判断。

可以先通过表型的连锁分析建立遗传模型,在此基础上再用相应的 DNA 标记进行连锁分析。如血浆 IgE 水平调控基因的定位就是先将 IgE 水平与哮喘进行连锁分析,将 IgE 代谢调控基因定位在 5 号染色体长臂后,进一步用相应的 DNA 标记进行了该基因的精细定位。

在临床上可以根据疾病的临床表现,如发病年龄、病情等把病人分为不同的系谱进行连锁分析。这方面一个成功的例子是早发性乳腺癌基因 BRCA1 的定位。尽管如此,高度的遗传异质性还是常常会导致连锁分析的失败。尽量选择相似的家系可以提高分析的成功率。当性状是由两个基因决定时,有时连锁分析会得出单个位点连锁的结论。此时用双位点模型可以显著增加连锁的可信度。如在荷兰一个哮喘家系作 DNA 标记 D5S436 与血浆 IgE 水平的连锁分析时,用单位点隐性遗传模型计算出的 LOD 值是 3.0。随后用分离分析的方法发现,该家系的资料更符合双位点隐性遗传的模型。改用双位点隐性遗传模型后计算出的 LOD 值是 4.6。

3. 多标记定位 联合应用多个标记,可以增加家系的信息量,对一个特定标记不能提供连锁信息的家系可能会给这个标记旁侧的标记提供连锁信息。表 2-2 显示了对乳腺癌患者按发病年龄分组后用多个标记进行连锁分析的结果。

表 2-2 乳腺癌患者多标记的连锁分析

	图					距*						
	D17S78					D17S41		D17S74				
家系	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.184	0.208	
≤45	2.83	3.09	3.30	3.47	3.57	3.41	4.46	4.60	5.24	5.41	5.24	
46~51	-0.30	-0.07	0.01	0.03	-0.05	-0.20	-1.58	-2.71	-9.14	-5.61	-4.24	
>51	-6.70	-5.80	-5.51	-5.52	-5.89	-6.98	-6.60	-7.94	-15.21	-8.94	-6.79	

* D17S78 与 D17S41 间的重组率是 10%, D17S41 与 D17S74 间的重组率是 6%

由表 2-2 可以看出, BRCA1 与 D17S74 的连锁要更紧密, 同时显示只有发病年龄小于 45 岁的家系才表现出明显的连锁。多标记分析对基因型和表型的错误比较敏感, 也要求完整的数据。可以用连锁不平衡进一步确定致病基因的位点。这种分析方法由于增加了初始者效应的概率, 对分析来自封闭人群的家系特别有效。

4. 注意事项 标记位点等位基因频率的错误估计会导致假阳性的连锁分析结果, 特别是在分析亲本等位基因频率不确定的家系时更是如此, 因为过低估计等位基因的频率会得出错误的连锁信息。例如表兄弟姐妹共同拥有“罕见”的等位基因, 一般认为是存在连锁的标志。实际上如果表亲的祖父母若是这个“罕见”等位基因的纯合子, 则表亲可能只是获得了这个等位基因的不同拷贝。因此必须认真检查观测样本等位基因频率的数据。除了对等位基因频率的错误估计可以导致假阳性的错误外, 为了明确家系遗传方式进行的重复观察也会增大假阳性出现的机会, 有时使连锁的结果难以解释。

(四) 非参数连锁分析方法

1. 简介 参数连锁分析依赖于性状的遗传方式, 非参数连锁分析 (allele-sharing methods, 等位基因共享法) 则不依赖于性状的遗传方式, 只需简单的观测标记和性状是否分离。最简单的非参数连锁分析方法是同胞对法, 又称患病同胞对法。这一方法是基于这样的假设: 从一个共同祖先来的多态标记等位基因在同胞对中的分布遵从孟德尔分离和自由组合定律。如果一个致病基因与所观察的多态标记不连锁, 不同家系中同胞对间多态标记等位基因的组合 (基因型) 与该致病基因的表型是相互独立的; 如果一个致病基因与所观察的多态标记相连锁, 它们就有共同传递给子女的倾向, 使得患病同胞对中

某些多态标记等位基因的分布偏离孟德尔定律。患病同胞拥有的来自共同祖先多态标记等位基因称为同源等位基因 (alleles identical by descent, IBD)。

2. 质量性状的非参数连锁分析 患病同胞对等位基因共享法。在参数连锁分析中, 外显不全、误诊和复杂性状等会使结果出现较大的偏差。此时可以考虑运用患病同胞对法, 这一方法的原理也可以用于分析非患病同胞对。如上所述, 如果致病基因与多态标记之间不连锁, 患病同胞对间 IBD 的分布将与表型无关, 并遵从孟德尔定律按两种纯合子各为 0.25, 杂合子为 0.5 的频率分布。如果致病基因与多态标记连锁, 则在患病同胞对中多态标记等位基因的分布将偏离上述频率。有多种分析上述偏离的统计学方法, 其中最为简单的是均数比较, 这一方法检验一系列 IBD 在同胞对中的频率是否偏离 0.5。

最近发展了用似然比率 (likelihood) 来估计患病同胞对多态标记等位基因分布的方法。这一方法中通过计算观察到的患病同胞对 IBD 似然度与在无效假设下患病同胞对 IBD 似然度两者的比值获得 LOD 值, 根据 LOD 值判断多态标记与致病基因是否连锁。

通过分别观察父母双亲与患病同胞对患病同胞对 IBD 的分布, 患病同胞对法还可用来估计致病基因的亲本来源。这对线粒体遗传和有基因组印迹发生的遗传性状的分析极为有用。

3. 数量性状 显然, 同胞对间共同的 IBD 越多, 表型相似的程度越大。可以用回归分析的方法分析某一数量性状与同胞对间共同 IBD 之间的相关关系。用 t 检验或方差分析检验回归系数的统计学意义。

4. 小结 表 2-3 总结了运用遗传分析方法进行致病基因定位的一般过程。

表 2-3 基因分析的遗传学方法

问 题	方 法
是否为遗传性状?	家系调查
如为复杂性状,有无主基因?	分离分析
主基因是显性还是隐性?	
主基因如何定位?	连锁分析
致病基因与某个 DNA 标记是否连锁?	参数分析
在患病同胞对或亲属间是否存在某些等位基因分布的异常?	等位基因共享法
基因的精细定位	多位点分析

(刘 昕)

参 考 文 献

- 1 Khoury MJ, Beaty TH, Cohen HB. Fundamentals of Genetic Epidemiology. New York; Oxford University Press, 1993
- 2 Ott J. Analysis of Human Genetic Linkage. Baltimore, Johns Hopkins University Press, 1992
- 3 Thomson G. Identifying complex disease genes: progress and paradigms. Nat Genet, 1994; 8: 108—110
- 4 Barcellos LF, Klitz W, Field LL, *et al.* Association mapping of disease loci, by use of a pooled DNA genomic screen. Am J Hum Genet, 1997; 61: 734—747
- 5 Luo DF, Bui MM, Muir A, *et al.* Affected—sib—pair mapping of a novel susceptibility gene to insulin—dependent diabetes mellitus (IDDM8) on chromosome 6q25—q27. Am J Hum Genet, 1995; 57: 911—919
- 6 Naiglin L, Clayton J, Gazagne C, *et al.* Familial high myopia: evidence of an autosomal dominant mode of inheritance and genetic heterogeneity. Ann Genet, 1999; 42: 140—146

第3章 用体细胞杂交进行基因作图

一、引言

在基因定位中,人/啮齿类动物体细胞杂交技术是一项非常有用的方法。第一个人小鼠杂交细胞系是利用胸苷激酶(TK)缺陷的小鼠L-细胞和人胚纤维细胞的自发融合的杂种细胞建立的。将上述杂种细胞用HAT培养液培养,非杂交细胞会被杀死。留下的杂种细胞在培养过程中,人的染色体会逐渐丢失,最后,只有含有人17号染色体的杂种细胞存活下来,因为人的TK基因位于17号染色体。由于杂种细胞中人染色体的丢失是

随机的,因此在这个过程中会形成含有不同人类染色体的杂种细胞,把这些细胞分别克隆,就可以制备一套包括全部人类染色体的杂种细胞集。在这套细胞集中的每一株细胞都只含有特定的几条人类染色体,有的细胞可能只含有一条染色体或是某条染色体的一部分。以这样的一套细胞集为材料,结合运用核酸杂交或是PCR技术,就可以很快将基因定位在特定的染色体上。但在这一过程中应注意区别人和小鼠的DNA序列。

二、材料

(一)DNA转印杂交部分

1. 基因组DNA酶切

(1)杂种细胞,人和小鼠的对照细胞。

(2)1×TAE缓冲液:40mmol/L Tris 醋酸, pH 7.8。2mmol/L EDTA, pH 8.0。

(3)EB溶液。

(4)DNA分子量标记:λDNA HindIII 酶切片段。

2. 转膜

(1)0.4mol/L NaOH。

(2)尼龙膜。

(3)吸水纸。

(4)20×SSC。

3. 探针标记

(1)10mg/ml BSA。

(2)寡核苷酸标记缓冲液:下述三种溶液按2:5:3的体积比混合。

溶液A:

1.25mol/L Tris-HCl, pH 8.0;
125mmol/L MgCl₂; 0.18% 2-巯基乙醇;
0.5mmol/L dATP; 0.5mmol/L dTTP。

溶液B:

2mol/L HEPES, pH 6.6。

溶液C:

六聚脱氧核苷酸,溶于3mmol/L Tris-HCl, 0.2mmol/L EDTA, pH 7.0的溶液中,90OD U/ml。

(3) $\alpha^{32}\text{P}$ -dCTP。

(4)Klenow酶。

(5)终止液:20mmol/L NaCl; 20mmol/L Tris-HCl, pH 7.5; 2mmol/L EDTA; 0.25% SDS。

(6)Sephadex G50柱。

(7)20×SSC。

4. 杂交

(1)预杂交液:5×SSC; 53Dehard液; 40mmol/L PB, pH 6.5; 0.1mg/ml 变性鲑鱼精DNA。

(2)杂交仪。

(3)标记好的探针。

(4)杂交液:含10%硫酸葡聚糖的预杂交液。

(5)漂洗液:0.23SSC, 0.1%SDS。