



华夏英才基金学术文库

主 编 罗建红
副主编 朱丽君

医学分子细胞生物学

研究策略与技术原理



科学出版社



华夏英才基金学术文库

医学分子细胞生物学

研究策略与技术原理

主 编 罗建红
副主编 朱丽君

科学出版社

北 京

内 容 简 介

本书从生物医学相关前沿领域的研究主题切入,着重介绍所涉及的分子生物学和细胞生物学的技术原理与研究策略。全书共分14章,内容涉及基因组学、基因克隆、基因的表达与调控、非编码小RNA、蛋白质组学、蛋白质结构、细胞显微成像、细胞内信号转导及组织细胞电生理学等分子和细胞生物学的核心技术。另外,我们还选择了5个重要的技术专题,包括动物基因修饰、肿瘤生物标志物筛选、分子药物靶点、干细胞及基于网络的生物信息利用,旨在帮助读者尽快掌握生物医学重要前沿领域的研究策略和技术原理,从而采用合适的方法和技术解决相关的科学问题。

本书适于作为现代生物医学相关专业研究生的实验课程教材或研究参考书,也可作为生物学、农学、医学及药学等专业高年级本科生,以及生物医学领域科研工作者的科研参考书。

图书在版编目(CIP)数据

医学分子细胞生物学:研究策略与技术原理/罗建红主编. —北京:科学出版社,2012

华夏英才基金学术文库

ISBN 978-7-03-023811-5

I. 医… II. 罗… III. 医学-分子生物学-细胞生物学 IV. Q7

中国版本图书馆CIP数据核字(2008)第207321号

责任编辑:夏 梁 刘 晶/责任校对:张凤琴

责任印制:钱玉芬/封面设计:陈 敬

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2012年2月第一版 开本:787×1092 1/16

2012年2月第一次印刷 印张:20 1/4

字数:480 000

定价:58.00元

(如有印装质量问题,我社负责调换)

编委会名单

主 编 罗建红

副主编 朱丽君

编 委 (按姓氏汉语拼音排序)

白玉杰	浙江大学医学院 生物化学与遗传学系
陈 智	浙江大学医学院 附属第一医院
丁 玲	浙江大学医学院 解剖与细胞生物学系
金静华	浙江大学医学院 神经生物学系
来茂德	浙江大学医学院 病理与病理生理学系
刘 伟	浙江大学医学院 生物化学与遗传学系
罗建红	浙江大学医学院 神经生物学系
欧阳宏伟	浙江大学医学院 解剖与细胞生物学系
祁 鸣	浙江大学医学院 生物化学与遗传学系
邱 爽	浙江大学医学院 神经生物学系
邵吉民	浙江大学医学院 病理与病理生理学系
沈 静	浙江大学医学院 病理与病理生理学系
沈 颖	浙江大学医学院 神经生物学系
唐修文	浙江大学医学院 生物化学与遗传学系
王琳琳	浙江大学医学院 生物化学与遗传学系
徐 静	浙江大学医学院 病理与病理生理学系
杨 巍	浙江大学医学院 神经生物学系
周天华	浙江大学医学院 生物化学与遗传学系
朱海红	浙江大学医学院 附属第一医院
朱丽君	浙江大学医学院 神经生物学系

前 言

分子生物学和细胞生物学研究技术已渗透到医学生物学研究的各个领域，它们不断在分子和细胞水平揭示出生命过程的机制；此外，由于微观研究技术的趋同，使医学生物学的传统学科界限逐渐模糊甚至交叉融合。尽管多层次技术切入的研究日益成为生物医学研究的手段，但分子和细胞生物学技术却是枢纽性的。同时，生物医学领域新的发现又催生了新的方法和技术，而且受需求的驱动又迅速地应用于转化医学的研究。一些改变医学实践模式的重要研究计划正在进行，这种研究技术和科学问题相互共生发展且日益受到医学应用牵引的态势可以说是现代生物医学的一个重要特征。

作为一名刚进入生物医学研究领域的研究生，除了在基础理论学习和文献阅读基础上提出要解决的科学问题外，接下来的问题可能就是如何采用合适的方法和技术来解决所提出的科学问题。为了帮助研究生更好地掌握分子生物学和细胞生物学的研究策略与研究技术并应用于解决具体问题，我们编写了本书。本书有别于传统的以系统阐述分子生物学和细胞生物学理论为主的教科书，也有别于详细介绍技术本身的方法学书籍和实验手册，我们试图按照利用分子生物学和细胞生物学技术进行研究的一些重要主题来介绍研究策略和相关的技术原理，从而使研究生和年轻的生物医学研究者对某一研究领域的研究在策略和方法学上能尽快有一个宏观的了解。这些主题涵盖了分子、细胞和整体动物等不同层次，内容涉及基因组学、基因克隆、基因的表达与调控、非编码小 RNA、蛋白质组学、蛋白质结构、细胞显微成像、细胞内信号转导及组织细胞电生理学等分子生物学和细胞生物学的核心技术。另外，我们还选择了 5 个重要的技术专题，包括动物基因操作、肿瘤生物标志物筛选、分子药物靶点、干细胞及基于网络的生物信息利用。

本书酝酿良久，这次正式出版的书稿基本上脱胎于浙江大学医学院十几年来多次改版的分子生物学实验技术课程讲义。这次的编著工作更是得到了浙江大学医学院许多新引进青年教授和高年级研究生的积极响应与支持，他们在各自专业上的造诣也凝结在本书的字里行间。在此，我对全体参与编著工作的老师和研究生所付出的辛勤劳动表示衷心的感谢，同时也对科学出版社在出版过程中给予的帮助和支持表示感谢！最后，我自己实验室的老师和同学对书稿进行了认真的阅读和校对，在此一并表示谢意！

由于专业水平所限，本书在内容或撰写上存在不足甚至谬误之处在所难免，惶恐之余，也恳求读者的指正。

罗建红

2011 年 5 月

目 录

前言

第一章 基因组学	1
第一节 概述	1
一、基因组学基本概念	1
二、人类基因组计划	3
第二节 研究策略及基本原理	5
一、遗传或连锁图谱	5
二、物理或分子图谱	7
三、基因组测序	9
四、基因芯片技术	17
五、基因克隆策略	19
第三节 相关实验技术及应用举例	21
一、RFLP 作图操作举例	21
二、放射性杂交文库构建	22
三、全自动荧光标记法测序	22
四、基因组测序举例	23
五、基因芯片的操作举例	24
第二章 基因克隆及表达技术	27
第一节 概述	27
第二节 基于细胞的 DNA 分子克隆	28
一、基于细胞的 DNA 克隆的原理和方法	28
二、DNA 文库	33
三、克隆载体系统	35
第三节 克隆基因的表达	40
一、表达型克隆的原理和基本条件	41
二、克隆基因在原核细胞中的表达	45
三、克隆基因在真核细胞中的表达	47
四、克隆基因表达产物的检测	55
第三章 真核细胞基因转录起始调控研究	57
第一节 概述	57
一、顺式作用元件	57
二、反式作用因子	58
第二节 真核基因转录起始调控的研究策略	59

一、启动子的分析	60
二、转录因子的分析	63
第三节 真核基因转录起始调控的相关技术原理	69
一、启动子序列的克隆	69
二、启动子及转录因子的鉴定	73
第四节 应用与实例	79
第四章 非编码小 RNA	85
第一节 非编码小 RNA 概述	85
一、RNAi 概述	85
二、microRNA 概述	87
三、miRNA 和 siRNA 的比较	89
第二节 非编码小 RNA 的研究策略	90
一、RNAi 的研究策略	90
二、miRNA 的研究策略	92
三、总结	94
第三节 非编码小 RNA 应用领域及其技术原理	94
一、RNAi 技术的应用及其原理	94
二、miRNA 的作用及其原理	95
第五章 蛋白质组学	101
第一节 概述	101
一、蛋白质组学的概念及其发展史	101
二、蛋白质组学的特点	102
三、蛋白质组学的研究方法和发展趋势	103
第二节 研究策略	103
一、定量蛋白质组学	104
二、蛋白质组翻译后修饰研究策略	110
三、蛋白质的相互作用研究策略	117
四、亚蛋白质组学研究策略	122
第三节 技术原理	123
一、蛋白质样品制备技术	123
二、蛋白质样品分离技术	127
三、蛋白质样品鉴定技术——质谱技术	129
四、蛋白质组生物信息学	131
第六章 生物大分子结构研究技术	135
第一节 概述	135
第二节 研究策略、技术原理和应用	136
一、X 射线晶体衍射分析	136
二、核磁共振技术	139

三、电镜及电镜图像的三维重构	142
四、生物大分子的三维结构数据库及三维结构预测、模拟和分析	144
第三节 应用举例	150
第七章 细胞成像	154
第一节 概述	154
第二节 荧光团	156
一、有机染料分子	156
二、荧光蛋白	157
三、量子点	161
第三节 荧光显微镜	162
一、激光共聚焦扫描显微镜	162
二、广场落射式荧光显微镜	162
三、全内反射性荧光显微镜	162
四、双光子或多光子激光扫描显微镜	163
五、多光谱激光扫描显微镜	163
第四节 荧光显微术	164
一、荧光蛋白生物探测器	164
二、用光漂白技术来研究活细胞内蛋白质的运动和分子动力学: FRAP、iFRAP 和 FLIP	165
三、荧光共振能量转移	168
四、荧光寿命成像显微术	170
五、荧光相关光谱	171
六、双分子荧光互补和绿色荧光蛋白的环形重排	173
七、可被光激活的笼锁蛋白及发色团协助的激光失活的应用	174
八、荧光颗粒显微术	175
第五节 单分子检测技术	176
一、单分子成像技术	177
二、单分子纳米操作技术	179
三、单分子成像与纳米操作技术的结合	180
第八章 组织细胞电生理学技术	182
第一节 电生理技术概述	182
一、发展概况	182
二、电生理测量技术	182
第二节 细胞外记录技术	184
一、细胞外记录——根据胞外电位推断神经元活动的电生理记录	184
二、心电图	185
三、脑电图	189
第三节 细胞内记录技术	191

一、细胞内记录的设计	191
二、脑切片上的细胞内记录	192
第四节 膜片钳记录	193
一、膜片钳技术设计	193
二、离子通道的几种记录模式	194
三、膜片钳技术的基本操作步骤	196
四、研究应用与举例	198
第五节 RT-PCR 和膜片钳技术	198
一、基本原理	198
二、应用和举例	199
第六节 膜片钳和钙成像技术	200
一、基本原理	200
二、设备和实验步骤	200
三、测定的准备工作	202
四、测定的实际操作	203
五、应用与举例	204
第七节 脑片的胞体、树突和轴突上的膜片钳记录	204
一、树突或轴突记录的确认	204
二、树突记录的实例	205
三、讨论和结论	206
第八节 爪蟾卵母细胞膜片钳记录技术	206
一、基本原理	206
二、实验基本操作	207
第九章 细胞内信号转导的研究	209
第一节 细胞信号转导的基本原理	209
一、胞外信号分子与特异性受体结合	209
二、胞外信号可经过长程或短程起作用	210
三、每个细胞应答特定的细胞外信号分子	210
四、相同信号分子能诱导不同靶细胞的不同反应	211
五、一氧化氮能直接进入靶细胞结合细胞内酶	211
六、主要的三类细胞表面受体	212
七、细胞内信号蛋白起一系列分子开关的作用	212
第二节 信号转导研究相关策略及技术	213
一、蛋白质磷酸化	213
二、蛋白质相互作用	215
第三节 信号通路举例	218
一、MAP kinase 信号通路	218
二、NF- κ B 信号通路	222

第十章 动物基因修饰技术	228
第一节 概述.....	228
第二节 研究策略及相关技术原理.....	228
一、非定向动物基因修饰技术	229
二、定向动物基因修饰技术	231
三、相关资料库及互联网资源	234
第十一章 肿瘤生物标志物的筛选策略	236
第一节 肿瘤生物标志物的基本概念.....	236
第二节 肿瘤生物标志物的筛选策略.....	236
一、差异表达基因 (mRNA) 的筛选	236
二、差异表达蛋白的筛选	239
三、蛋白质指纹图谱技术	242
四、生物信息学	243
第三节 肿瘤生物标志物的验证.....	243
一、mRNA 水平验证	244
二、蛋白质水平验证	245
第四节 肿瘤标志物的临床评价.....	246
第十二章 分子药物靶点的研究	248
第一节 概述.....	248
一、分子药物靶点的概念	248
二、分子药物靶点的研究内容	248
三、分子药物靶点的研究策略	249
第二节 分子药物靶点的研究技术原理.....	250
一、基因组学	250
二、蛋白质组学	254
三、遗传联合	255
四、正向遗传学	255
五、反向遗传学	255
第三节 分子靶点的确认.....	257
第十三章 干细胞实验研究	259
第一节 概述.....	259
一、干细胞分类	259
二、干细胞研究历史	261
第二节 干细胞研究策略.....	261
一、干细胞研究材料的获取	261
二、干细胞生物学.....	263
三、干细胞的应用前景	266
第三节 干细胞研究的相关技术原理.....	268

一、胚胎干细胞体外培养技术	268
二、成体干细胞的分离纯化技术	271
三、干细胞的鉴定技术	273
四、干细胞体外分化技术	277
五、干细胞低温保存与复苏	280
第十四章 网络生物信息数据库及相关软件的应用	286
第一节 概述	286
第二节 常用的数据库介绍	286
一、数据库存储格式	286
二、核酸数据库	290
三、蛋白质数据库	290
四、数据库检索方法	292
第三节 数据库及软件的应用举例	297
一、用 GenScan 来预测内含子、外显子及氨基酸序列	298
二、用 ScanProsite 来分析编码的氨基酸的 motif	300
三、用 BLAST 来搜索序列相似蛋白质	300
四、用 COGnitor 来分析直系同源物	300
五、用 MultAlin 来进行多序列比对	301
六、搜索 CDD 数据库, 找到预测蛋白的保守区域并显示此结构域的空间结构	301
七、小结	301
第四节 常用生物信息数据库及软件	302
一、生物信息数据库	302
二、常用软件和网上服务	307
名词索引	310

第一章 基因组学

基因组学在医学研究中的核心内容是运用基因组学的新思路、新技术进行疾病基因的快速克隆，以及高通量、系统化地揭示疾病的分子病理。

第一节 概 述

一、基因组学基本概念

基因组（genome）是指生物体单体中一套完整的遗传物质，即全部基因的总称（包括所有的编码区和非编码区），是特定物种一个细胞内全部 DNA 分子的总和；或者指当同样信息存在于多个拷贝中时，单个拷贝的遗传物质（后者有时也被称为“单倍体基因组”）。

真核生物 99% 以上的 DNA 存在于核基因组（nuclear genome）中，少部分存在于细胞器基因组（organelle genome）中。例如，人类基因组由 3×10^9 个碱基对组成，分布于 22 对常染色体、2 条性染色体和 1 个线粒体上。基因组具有功能性的高级结构，可以根据其理化性质和顺序进行描述。大部分生物的 DNA 可简单地分为只出现一次的单一序列（unique sequence）DNA 和重复序列（repetitive sequence）DNA。大多数基因存在于单一序列 DNA 中，但在中等重复序列 DNA 中也存在与高度保守的多基因家族相对应的基因。从低等原核生物到高等真核生物，基因组大小、复杂度、基因数目、组成、结构及基因在基因组中的组织方式都发生了显著变化（表 1-1）。细菌和一些低等真核生物基因组中绝大部分是可表达的单一 DNA 序列，基因结构复杂且排列密度高，基因内部一般不含内含子。相反，高等真核生物基因组大，大部分为非编码 DNA，基因大小变化很大，常含有多个内含子，且内含子一般大于外显子（表 1-2）。

表 1-1 一些生物的基因组数据

物种	基因组大小	复杂度/%*	GC 含量/%	基因数目
λ 噬菌体	45kb	>99	48	100
大肠杆菌	4. 7Mb	99	51	4 100
啤酒酿酒酵母	13. 5Mb	90	41	6 300
裂殖非洲栗酒酵母	20Mb	90		6 000
盘基网柄菌	47Mb	70	23	7 000
秀丽隐杆线虫	100Mb	83		14 000
黑腹果蝇	165Mb	70	39	12 000
河豚	400Mb	>90	44	70 000
担尼鱼	1. 9Gb			70 000

续表

物种	基因组大小	复杂度/%*	GC 含量/%	基因数目
非洲爪蟾	2.9Gb	54	50	70 000
小鼠	3.3Gb	58	41	70 000
人	3.3Gb	64	40	70 000

* 复杂度是以单一序列 DNA 的百分比来表示的。

表 1-2 根据含量和功能分类的真核生物基因组

DNA 类型	定义
根据含量	
单一序列（单拷贝、低拷贝及非重复 DNA）	每个基因组中出现一次或很少次，包括大部分基因和内含子、调节顺序和其他未知功能的 DNA
中等重复序列 DNA	每个基因组中出现 10~10 000 个拷贝，一般代表高度保守的多基因家族的分散重复序列（功能基因和假基因）和转座子，偶尔成簇排列
高度重复序列 DNA	每个基因组中出现 10 000~1 000 000 个拷贝，一般作为随机重复序列被发现，部分超高丰度的转座子也属于此类（如 Alu 元件）
根据功能	
基因 DNA	可表达的 DNA，可进一步分为 mRNA（编码蛋白质）、rDNA、tDNA、snDNA 等，代表不同的基因产物
调节 DNA	调节基因表达（如启动子、增强子）或调节 DNA 功能（如重复起始区、核基质结合区域等）
基因内 DNA，间隔 DNA	内含子和分隔基因的 DNA。靠近着丝粒、端粒等位置的高度重复 DNA，有些在染色体功能中发挥作用
卫星 DNA	
自在 DNA	介导自身在基因组中的复制和生存，如一些卫星 DNA 和转座子
无用 DNA	没有确定功能的 DNA

基因组携带了有关个体生长发育、生老病死的全部遗传信息。人类对自身遗传信息本质的研究和探索已进行了几十年，但从 20 世纪 60 年代开始才真正从 DNA 的组成和分子结构水平认识基因。基因组学（genomics）一词最早于 1986 年 7 月由 T. H. Roderick 等提出，是包括基因组作图、测序及分析的综合学科。基因组学着眼于研究并解析生物体整个基因组的所有遗传信息，完全不同于经典遗传学“局部和分散”的研究方法。人类基因组学研究主要包括两个部分：①基因组 DNA 全部序列的测定；②基因 DNA 序列的识别及其正常功能、基因变异与功能关系的研究。已经完成的人类基因组计划（human genome project, HGP）是基因组研究的里程碑，将破译蕴藏在 DNA 分子中有关人类生存、繁衍和进化的全部遗传信息，不仅有助于了解人类的正常生理功能，还将揭示多种疾病的发病机制，可广泛应用于疾病的诊断、治疗及预防。

二、人类基因组计划

(一) 人类基因组计划概况

人类基因组计划与曼哈顿计划、阿波罗登月计划并称为人类自然科学史上的三大工程。其主要内容是绘制人类的遗传图、物理图和序列图，其核心是序列图的绘制。HGP 在 15 年内完成人类 23 对染色体的基因组图谱的绘制和 DNA 全序列分析，获得人类认识自我的最重要的生物信息。

HGP 于 20 世纪 80 年代开始酝酿，1989 年美国国会正式批准资助，1990 年 10 月 1 日美国能源署 (DOE) 和 NIH 正式启动该计划，投资 30 亿美元，且每年追加新的资金。除美国外，英国、法国、德国、日本、中国等国家先后加入，使得 HGP 成为一项世界范围的科研项目。HGP 原预计 15 年完成，由于实施过程中大规模测序技术及生物信息分析技术的重大突破，项目进展超过了预期进度，于 2000 年 6 月完成全基因组工作草图，2003 年 4 月完成全部目标，比原计划提前了 2 年。

(二) 人类基因组计划的成果

HGP 完成的测序部分达到全基因组的 99.99% 以上，所建立的人类基因组图谱可理解为人类的“第二张解剖图谱”或“分子解剖图谱”，更为详细和全面地揭示了人类基因组信息：人类全基因组由约 31.647 亿个碱基对组成，分布在 22 对常染色体、2 条性染色体上和 1 个线粒体上；任何两个个体基因组 99.9% 以上的碱基序列相同，差异不到 0.1%；现发现蛋白质编码基因约 3 万个，远少于先前所预测的 10 万个；各染色体含基因数量差别较大，其中 1 号染色体最多 (2968 个)、Y 染色体最少 (78 个)；基因平均长度为 3kb，基因长度差别很大，已测得的最大基因是长达 2.4Mb 的抗肌萎缩蛋白基因；蛋白质编码序列仅占全基因组序列的 1.5% 左右，非编码的重复序列，即所谓的“垃圾 DNA (junk DNA)”超过总基因组的 50%，这些重复序列可能无直接功能，但可能对染色体结构有动力学作用，并可能参与基因改形、制造新基因、基因修饰或改组等，是有待进一步研究的重要领域。

HGP 的科学意义在于：

- (1) 确定基因组中的基因序列和物理位置，为基因产物及其功能的研究奠定基础；
- (2) 分析基因转录及调控元件的结构和位置，从全基因组水平了解基因转录及调控的机制；
- (3) 从整体水平了解染色体结构，不同序列在 DNA 复制、转录和调控中的作用及影响；
- (4) 从三维空间角度研究基因的转录和调控；
- (5) 发现与 DNA 复制、重组相关的序列，研究人类基因组的遗传与进化进程；
- (6) 通过对 DNA 突变、重排和染色体断裂等的研究，了解多种疾病的分子机制；

(7) 比较和研究个体染色体间的多态性,以用于基因诊断、个体识别、组织配型及疾病风险预测等;

(8) 确定人类染色体中转座子、逆转座子和病毒残余序列及其周围序列,了解病毒基因组侵染人体基因组的影响,指导基因治疗病毒载体的设计 and 应用。

HGP 所建立的研究策略、合作机制及技术体系奠定了人类基因组学研究的基础,可应用于微生物、植物、动物等物种的研究,以及医学、农业、工业、环保、能源等多个领域的研究。HGP 的实施促进了生命科学与信息科学、材料科学、精密制造等多学科和技术的交叉融合,带动了一批新技术产业的发展。

(三) 基因组研究的远景

人类基因组图谱将成为疾病预测、预防、诊断、治疗和个体医学研究的分子参照,将是 21 世纪生命科学、基础医学和生物产业的基础。其研究成果将被应用于众多领域,如下所述。

(1) 分子医学:疾病的基因诊断、疾病遗传易感性分析及风险预测、新药开发、基因治疗、个体化药物设计等。

(2) 微生物基因组研究:快速诊断及治疗临床病原微生物、开发新能源、监测环境污染、防止生物武器、清除有毒废弃物等。

(3) 职业风险评估:某些接触放射性或有毒物质的特殊职业人员的风险评估。

(4) 生物考古学(bioarchaeology)、人类学(authropology)、进化及人类迁移分析。

(5) 法医学 DNA 鉴定。

(6) 农业、畜牧业及生物工艺:抗病、抗虫、抗旱新作物,健康、高产、抗病家禽、家畜,高营养产品,生物杀虫剂,食物疫苗,清洁或保护环境的植物等。

HGP 的完成仅仅是第一步,只是以测序为主的“结构基因组学”,对人类遗传信息的解读和开发利用才是最终目标,此即“后基因组计划”的任务,也就是对基因组的功能进行探索和利用。在 HPG 完成后,科学家相继提出了药物基因组学、环境基因组学、蛋白质组学等新的学科,并组织 and 实施了一系列旨在开发和利用基因组信息的新研究计划。例如,美国的“人类癌症基因组计划”投资 13.5 亿美元,建立了超过 12 500 个肿瘤标本的基因库,并将肿瘤患者和正常人的基因组进行比对分析,计划为 50 种常见肿瘤绘制 250 幅基因图谱,编制所有致癌突变目录,此计划规模相当于 HGP 的 100 倍。

美国能源署(DOE)提出的基因组与生命科学计划(genomes to life, GTL),通过分析微生物和植物群体在自然环境中的功能特性、重要功能蛋白质的作用及其基因调节和控制,来模仿复杂的生物系统。预期各阶段目标为:10 年内掌握微生物清洁体系的基础知识,到 2020 年在废物和有毒物处理方面节省数十亿美元;到 2040 年了解地球自然碳循环并设计增加碳捕获能力的策略,帮助稳定大气层 CO₂,应对和遏制全球变暖;到 2050 年开发出生物燃料和生物电资源,形成生物能源工业(biohydrogen-based

industry), 为美国能源安全作出贡献。

第二节 研究策略及基本原理

一、遗传或连锁图谱

遗传图 (genetic map) 又称为连锁图 (linkage map), 它用遗传学方法确定基因或 DNA 标志在染色体上的相对位置与遗传距离。遗传距离常用基因或 DNA 片段在染色体交换过程中的分离频率来表示, 单位为厘摩 (cM), 1cM 代表 1% 的交换率。cM 值越大, 两者间的遗传距离越远。

(一) 遗传作图的标记

理想的分子标记应满足以下要求: ①具有较高的多态性; ②共显性遗传, 即利用分子标记可鉴别二倍体中杂合和纯合基因型; 能明确辨别等位基因; ③除特殊位点的标记外, 分子标记均匀分布于整个基因组; ④检测手段简单、快速, 实验程序易实现自动化, 开发成本和使用成本尽量低廉; ⑤在实验室内和实验室间重复性好, 方便进行数据交换。

经典遗传作图标记包括形态学标记、细胞学标记、生化标记等。由于此类标记依赖于蛋白质表达, 只能间接反映基因组信息, 存在多个基因共同影响同一表型、数量及分布有限、标记数量小、特殊遗传材料培育困难等问题, 因此在 20 世纪 80 年代后逐渐被 DNA 遗传标记所取代。这里主要介绍 DNA 标记。

1. 第一代 DNA 标记

20 世纪 80 年代, David Bostein 等提出了限制性片段长度多态性 (restriction fragment length polymorphism, RFLP)。其基本原理是: 由于 DNA 序列差异导致限制性内切核酸酶识别位点的丢失或新增, 用限制性内切核酸酶处理后可产生长度不同的限制性酶切片段, 从而直接显示不同个体在基因组同一位点 DNA 组成的差异。限制性内切核酸酶专一性识别碱基序列, 不同限制性内切核酸酶处理同一 DNA 样品可产生相应的不同 DNA 片段, 因此可提供大量的位点多态性信息。RFLP 是第一种用于遗传作图的 DNA 标记, 其优点包括: ①在染色体上位置相对固定; ②同一亲本及其子代在相同位点的多态性特征一致; ③共显性特征; ④同一电泳可显示不同多态性片段。RFLP 的主要局限性是多样性低、操作费时、需要大量高质量的 DNA 模板等。

2. 第二代 DNA 标记

简单序列长度多态性 (simple sequence length polymorphism, SSLP) 是指基因组中重复序列的可变性排列, 由于重复次数不同而表现为 DNA 序列的长度变化。SSLP 主要包括: ①小卫星序列 (minisatellite), 又称可变串联重复 (variable number of tan-

direct repeat, VNTR), 重复单位长度一般为 15~65bp; ②微卫星序列 (microsatellite) 或简单串联重复 (simple tandem repeat, STR), 重复单位长度为 1~6bp, 重复次数为 10~80 次。由于 STR 在整个基因组中分布更广、密度更高, 且便于进行 PCR 分析, 因此较 VNTR 应用更为普遍。人类基因组作图应用最多的 STR 为 AG, 主要分布在 5' 和 3' 非翻译区及内含子中。

SSLP 可用酶切后电泳、特异性探针杂交, 或采用 PCR 扩增后直接电泳观察产物条带的迁移率变化进行判断。

3. 第三代 DNA 标记

单核苷酸多态性 (single nucleotide polymorphism, SNP) 是指在基因组水平上由单个核苷酸的变异所引起的 DNA 序列多态性, 即基因组内特定核苷酸位置上存在两种不同的碱基, 其中最少数一种在群体中的频率不小于 1%。它是人类可遗传的变异中最常见的一种, 占有已知多态性的 90% 以上。SNP 所表现的多态性只涉及单个碱基的变异, 可由单个碱基的转换 (transition) 或颠换 (transversion) 所致。理论上讲, SNP 既可能是二等位多态性, 也可能是三或四等位多态性, 但实际上后两者非常少见, 几乎可以忽略。因此, 通常大部分 SNP 都是二等位多态性。SNP 作为一种碱基的替换, 大多数为转换, 颠换较少, 转换与颠换之比大约为 2:1。换句话说, 转换型 SNP 约占 2/3, 颠换型约为 1/3。

SNP 作为分子标记具有以下优势。

(1) 数量多, 分布广泛。人类基因组平均每 100~300 个核苷酸就有一个 SNP, 人类基因组 30 亿个碱基中共有 1000 万~3000 万个 SNP, 至少 93% 的人类基因都存在 SNP。

(2) 适于快速、规模化筛查。SNP 通常为二态性标记, 即二等位基因 (biallelic)。在基因组筛选中 SNP 往往只需 +/− 的分析, 易于进行自动化筛选或检测。

(3) 突变率低, 稳定性好。与串联重复的微卫星序列相比, SNP 高度稳定, 而前者的高突变率容易引起对人群的遗传分析出现困难, 尤其是处于编码区的 cSNP。

(4) 可直接作为疾病分析遗传标记。部分位于基因内部的 SNP 可能会直接影响产物蛋白质的结构或表达水平。SNP 本身可能就是疾病遗传机制的候选位点。

SNP 自身的特性决定了它比其他两类多态标记 (RFLP 和 SSLP) 更适合于对复杂性状疾病的遗传解剖和基于群体的基因识别等方面的研究, 它是目前最为理想的 DNA 标记。

(二) 遗传作图方法

遗传作图 (genetic mapping) 的基础是孟德尔的遗传学原理及 Bateson 和 Punnett 建立的连锁分析方法。细胞减数分裂过程中, 每条同源染色体复制后并非立即分开, 而是彼此靠近, 同源区段并排形成二价体, 此时并列的同源染色体臂之间发生机械断裂, 彼此交换 DNA 区段, 此过程称为交换 (crossing-over) 或重组 (recombination)。假如交换随机发生, 两个相互邻近基因因交换而分离的频率低于距离较远的两个基因; 换言