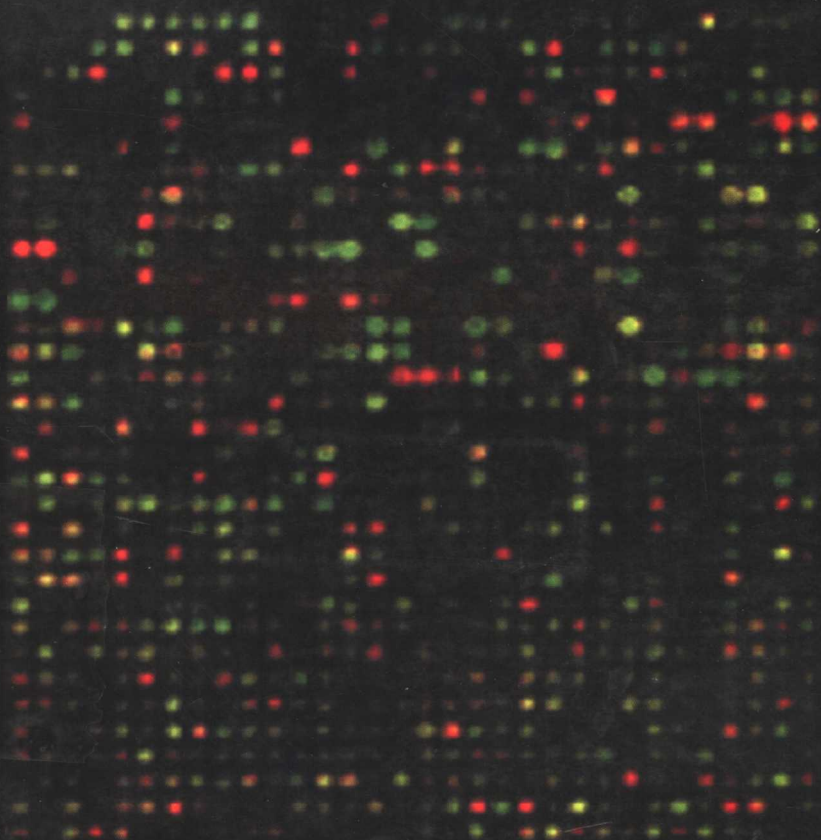


韩金祥 等 编译

DNA芯片

实验操作指南



山东大学出版社

DNA 芯片——实验操作指南

韩金祥等 编译

山东大学出版社

图书在版编目(CIP)数据

DNA 芯片:实验操作指南/韩金祥等编译. — 济南:山东大学出版社, 2001. 10

ISBN 7-5607-2215-6

I. D...

II. 韩...

III. 脱氧核糖核酸 - 实验 - 指南

IV. Q523-62

中国版本图书馆 CIP 数据核字(2001)第 072579 号

山东大学出版社出版发行

(山东省济南市山大南路 27 号 邮政编码:250100)

山东省新华书店经销

青岛胶南印刷厂印刷

850×1168 毫米 1/32 7.5 印张 插页 6 206 千字

2001 年 10 月第 1 版 2001 年 10 月第 1 次印刷

印数:1-4000 册

定价:20.00 元

版权所有,盗印必究

凡购本书,如有缺页、倒页、脱页,由本社发行部负责调换

本书编译人员名单

主 编 译: 韩金祥

副主编译: 王世立 鲁艳芹 李腾国 梁 浩
张忠玲 张效萌 黄海燕 潘光锦

编 译 者 (按姓氏笔画为序):

王世立	闫 实	朱有名	刘 毅
宋长征	孙晓阳	李腾国	李惠卿
张忠玲	张效萌	赵 建	梁 浩
黄海燕	鲁艳芹	韩金祥	潘光锦
潘继红			

序 一

基因芯片技术是 20 世纪 90 年代后期创建的一项分子生物学技术,它是将成千上万的基因探针按一定的阵列集成于固体基片,从而达到同时平行分析大量基因突变、表达等目的。可广泛应用于基因组学、临床诊断、新药筛选、法医学、环境监测、食品卫生等诸多领域,因而具有巨大的市场潜力,备受各国科学家乃至各国政府高度关注。

Mark Schena 教授是基因芯片的创始人,他主编的 *DNA Microarrays* 一书是国际上第一部关于基因芯片技术的专著。山东省医学科学院韩金祥教授等根据我国的实际情况和基因芯片技术近几年的发展以及他实验室的研究成果,以这本专著为蓝本,编译成《DNA 芯片——实验操作指南》一书,该书系统地介绍了基因芯片的制备、检测,以及在基因表达、基因多态性分析、基因克隆和筛选、病原体检测、新药筛选等方面的应用,内容新颖,叙述系统,层次分明,原理与技术结合,理论和实践结合,对于系统学习和掌握基因芯片技术,以及开展相关课题研究者均属实用,是向我国从事分子生物学研究的读者奉献的一部难得的好书。对此,我深感欣慰,并向他们表示热烈地祝贺。我相信,本书的出版对我国基因芯片的研究和发展必将起到重要的推动作用。

中国医学科学院

侯云德院士

2001 年 8 月于青岛



序 二

分子生物学是现代生命科学的脊梁,如果从 1953 年 Watson 和 Crick 首次阐明 DNA 双螺旋结构为起点,它已走过半个世纪的光辉历程。以基因克隆和重组 DNA 技术为中心的基因工程技术给人类社会带来的既广泛又深刻的影响,可谓史无前例,举世瞩目。

在分子生物学飞快发展的几十年中,相继诞生了基因组学(Genomics),蛋白质组学(Proteomics)和生物信息学(Bioinformatics)等潜力巨大的新兴学科,它们各有侧重,又密不可分,共同主宰着整个现代生命科学。在分子生物学的发展历程中,不断产生重大的技术革新,产生出崭新的技术;这些技术成为生命科学发展的原动力,改变着相关领域的面目,例如,聚合酶链式反应技术(PCR),使原来作坊式手工操作 DNA 技术,焕然一新发展成自动化和半工业化的水平。不难设想,如果没有 PCR 技术的发明和应用,像人类基因组这样涉及亿万个碱基对的巨大工程的完成,也许是几十年以后的事情了。

生物芯片(Biochip)技术是分子生物学里的又一项里程碑式的重大发展。这一技术是最近几年崭露头角的一项新技术,它是一个典型的学科间交叉的新领域。集 DNA 微阵列技术与微电子、微加工技术之大成,它把生命科学中不连续的分析过程集成在硅芯片或玻璃芯片为载体的微型生物化学分析系统里以达到完成对细胞、蛋白质、基因及其生物组分的准确、快速和大信息量检测,做到高通量、微型化和自动化。显然这是继大规模集成电路之后的

又一次具有深远意义的科学技术革命。这个新兴技术的研究开发及其产业化,将对疾病诊断、药物筛选、农作物优育和优选、环境检测和防治、食品卫生监督乃至司法鉴定等技术发展起到巨大的推动作用。无疑,它会在生命科学大发展的 21 世纪,新兴生物技术产业化中起到带头兵的作用。

我以非常欣慰和激动的心情浏览了由山东省医学科学院韩金祥研究员为首编译的这本《DNA 芯片——实验操作指南》。本书从基因芯片原理、制备操作到应用分析都作了由浅入深的全面介绍,是一本颇具特色、非常适用的参考书。在有幸参观了本书译者的实验室之后,我看到的不仅是装备精良的现代化实验室,更可爱的感受是一个意气勃发、技术娴熟、雄心勃勃的年轻科学家群体,他们那种脚踏实地的科学作风,令人振奋。

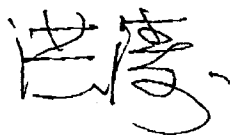
我相信,这本书的出版将对我国方兴未艾的基因芯片的研究和应用起到积极推动作用。虽然译者谦称该书成文仓促,颇多不足,但我认为瑕不掩瑜。因之,我愿应邀为本书作序,向广大同道举贤推荐,愿我国在基因芯片技术上审势夺时,早日取得重要进展!

洪 涛

中国工程院院士

中国预防医学科学院

2001 年 9 月 6 日于北京



前 言

被美国总统喻为“政治技术”的生物芯片发展于 20 世纪 90 年代,不到 10 年的时间,生物芯片技术已经显示出巨大的商业价值,备受世界各国政界、科技界及各大企业的关注。《财富》杂志预测,“到 2005 年仅仅在美国生物芯片销售额可达到约 50 亿美元,2010 年有可能上升为 400 亿美元。”

生物芯片主要是指通过平面微细加工技术在固体芯片表面构建的微流体分析单元和系统,以实现细胞、蛋白、核酸以及其他生物组分的准确、快速、大信息量的检测。基因芯片是最重要的一种生物芯片,芯片上成千上万密集排列的基因探针,能够在同一时间内平行分析大量的基因,使人们可以迅速地读取生命的篇章。这一技术广泛应用于:① 寻找和发现新基因、研究基因的功能以及生物体进化、发育、遗传等过程;② 现代医学诊断及法医学;③ 新药的研发;④ 环境监测和农业。这将是继大规模集成电路之后的又一次具有深远意义的科学技术革命。

由 M. Schena 主编、1999 年出版的 *DNA Microarrays* 一书是由目前国际上从事生物芯片研究造诣很深的科学家们共同编著的,是经典的基因芯片实用性书籍,内容十分新颖,图文并茂,体现了基本原理与技术方法并举、理论与实践相结合的特点。我们认为,这部著作不仅是基因芯片技术的实验手册,也是基因芯片技术方面不可多得的优秀教科书,因而,根据我国的实际情况将该书编译成中文介绍给国内同行,对推动我国基因芯片研究与开发具有非常现实的意义。

前 言

山东省医学科学院医药生物技术研究中心、山东省医学分子生物学重点实验室是一个年轻的集体,大家对编译这本书兴致极高,利用节假日和休息日搜集整理了大量的材料。但生物芯片技术是一门崭新的生物技术,有一些词在中文正式出版物中前所未见,而为数更多的术语则使中文译名相当混乱,往往令人无所适从,对这些术语,我们只能用英文注释。总之,这部著作是急就章,而编译者又是青年军,由于知识、能力、精力和水平有限,对基因芯片技术缺乏感性认识,在编译过程中疏忽与错漏在所难免,如能得到同行专家和读者的批评指正,我们将不胜感激。

中国工程院副院长侯云德院士和中国预防医学科学院洪涛院士在百忙中欣然为本书作序,这给予我们极大鼓励,在此表示衷心感谢。

编译者

2001年6月于山东省医学科学院

英 文 缩 写

APD	雪崩光二级管
BSA	牛血清白蛋白
CCD	电荷偶联装置
DEPC	焦碳酸二乙酯
EST	表达序列标签
FITC	异硫氰酸荧光素
LCM	激光显微捕获
NA	数码孔径
PCR	聚合酶链式反应
PMT	光电倍增管
PR	光致抗蚀剂
RDA	呈现性差异分析
RT	室温
SNP	单核苷酸多态
SSH	差减杂交
STR	短串联重复序列

目 录

第一章 基因、基因组、基因芯片	(1)
1. 基 因	(1)
2. 基因组信息	(2)
3. 基因芯片概念	(7)
4. 微阵列应用	(14)
5. 基因芯片与基因药物	(16)
6. 小 结	(17)
参考文献	(18)
第二章 基因芯片的制备技术	(21)
1. 基因芯片制备技术简介	(21)
2. 喷墨技术	(21)
3. 微电子阵列	(41)
4. 小 结	(56)
参考文献	(57)
第三章 微阵列检测技术	(59)
1. 引 言	(59)
2. 样品处理	(69)
3. 微阵列的共聚焦扫描	(71)
4. 波长识别:降低背景图像	(74)
5. 检测器	(77)
6. 信号处理与仪器控制	(78)
7. 仪器性能的度量标准	(81)

8. ScanArray 共聚焦微阵列扫描仪	(85)
参考文献	(88)
第四章 表达数据与生物信息学	(91)
1. 基因/蛋白质	(91)
2. 原始数据传递途径	(97)
3. 参照数据	(102)
4. 分子信息的结构	(103)
5. 医学辞典与供体样品登记	(108)
6. 表达数据的采集	(109)
7. 结 论	(115)
参考文献	(115)
第五章 克隆和筛选差异表达基因	(118)
1. 引 言	(118)
2. RDA	(118)
3. RDA 扣除文库的建立	(128)
4. 微阵列的制造	(131)
5. 微阵列杂交	(132)
6. 用微阵列筛选差异表达基因	(136)
7. 小 结	(137)
参考文献	(137)
第六章 基因多态性分析	(139)
1. 引 言	(139)
2. 阵列的制造	(139)
3. 靶 DNA 的制备	(145)
4. 与阵列杂交	(146)
5. 酶促催化的延伸反应	(148)
6. 方案优化	(153)
参考文献	(155)

第七章 反义核酸的筛选	(157)
1. 引 言	(157)
2. 阵列制作	(158)
参考文献	(179)
第八章 基因表达分析	(181)
1. 引 言	(181)
2. 激光显微捕获技术	(181)
3. aRNA 的产生——第一个循环	(187)
4. cDNA 微阵列	(193)
参考文献	(199)
第九章 基因芯片技术在临床诊断中的应用	(201)
1. 在遗传病监测中的应用	(202)
2. 在肿瘤研究中的应用	(203)
3. 生物芯片在病毒检测中的应用	(209)
4. 生物芯片在细菌检测中的应用	(211)
5. 问题和展望	(212)
参考文献	(213)
第十章 新药筛选研究	(215)
1. 新药筛选的发展趋势	(215)
2. 生物芯片在药物靶标筛选中的应用	(217)
3. 生物芯片与药理研究	(221)
4. 生物芯片与毒理学研究	(221)
5. 改变药物作用方式	(223)
6. 对中药物种的鉴定	(223)
7. 应用前景	(224)
参考文献	(224)

第一章 基因、基因组、基因芯片

1. 基 因

1838~1839年,德国植物学家 Schwann, Schleiden 提出:一切植物、动物都是由细胞组成的,细胞是一切动、植物的基本单位,从而确立了生物学史上重要的“细胞理论”。1859年,伟大的英国生物学家达尔文(Charles Darwin)的巨著《物种起源》确立了进化论的概念,打破了上帝造人的传统概念,改变了社会对人类在整个世界中的地位的看法,极大地推动了人类思想的发展。

细胞理论和进化论的结合和发展,产生了以遗传学和生物化学为主要研究内容的现代生物学,其中,生物化学的主要使命是分析细胞的组成成分,弄清楚这些物质与细胞内生命现象的联系。19世纪中叶到20世纪初,组成蛋白质的20种基本氨基酸被相继发现,细胞的其他成分,如脂类、糖类和核酸也相继被当时的科学家所认识和部分纯化。但最重要的生命活动,即细胞成分是如何时代相传的却无法解释。1857年到1864年的7年间,奥地利科学家孟德尔选择了7对差异明显的简单性状,对豌豆的生长进行了仔细的观察研究,提出了著名的遗传统一律和分离规律,1865年发表论著“植物杂交实验”。当时欧美各国科学界几乎无人理睬这一巨大成果,直到他逝世后的1900年,这一理论才被重新发现并得到普遍认可,同时孟德尔被誉为遗传学的奠基人。

在孟德尔遗传学的基础上,美国著名遗传学家 Morgan 通过

对白眼果蝇和红眼果蝇的杂交实验,发现了遗传形状的连锁定律,进而提出了基因学说, Morgan 指出:种质必须由某些独立的要素组成,我们把这些要素称为遗传因子,或者更简单地称为基因。

尽管 Morgan 的基因学说得到了普遍的承认,但当时人们对基因的理解仍然是抽象的、概念化的,缺乏准确的物质内容。而且也不能解释位于细胞核中的染色体和基因是怎样控制细胞质中的各种生化过程,不能解释基因是怎样在细胞繁殖过程中准确地复制和遗传的。当时,许多生物学家都认为蛋白质可能是决定细胞生物学特性和遗传的物质基础。1944 年,美国著名的微生物学家 Avery 通过关于致病力强的光滑型(S 型)肺炎链球菌导致致病力弱的粗糙型(R 型)细菌发生遗传转化的实验证明是 DNA 而不是蛋白质是遗传信息的载体,即 DNA 是基因的物质基础。1953 年, Watson 和 Crick 提出的 DNA 双螺旋模型揭示了基因的立体存在形式,预示出 DNA 复制的规则,标志着分子生物学的创立。这也是基因芯片技术的理论根据。

2. 基因组信息

2.1 脱氧核糖核酸(DNA)

基因组(genome)是指人类细胞中所有遗传信息的总和。基因组信息是由脱氧核糖核酸(DNA)携带的。基因组序列的大小和组成决定着该生物体的结构与功能。一般地讲,基因组的复杂性 with 相应生物体的复杂性是成比例的。比较简单的生物体如细菌,其基因组含有 100 万~500 万个碱基,而包括人类在内的哺乳类生物体基因组大约含有 30 亿个碱基;较复杂的生物体,比较典型的如蠕虫和昆虫,其基因组通常含有 1 亿~2 亿个碱基。

基因组被分成不同的区段,称作染色体。大肠杆菌含有一环状染色体;而酵母菌含有 16 条线状染色体,人类基因组是由 22 条

常染色体和 1 对性染色体组成的,这 24 条染色体包含了人类基因组中的全部 30 亿碱基。

基因组 DNA 是一种由 4 种碱基(A, G, C 和 T)组成的双链聚合体,这种双链结构以磷酸核糖为骨架(表 1-1),其碱基顺序称作 DNA 的一级序列结构。基因组计划的目的是确定 24 条染色体中全部 30 亿个碱基的一级序列。基因组序列又称结构基因组,它可为生物芯片业提供序列信息。

基因组 DNA 的两条链是通过每条链上互补碱基间氢键的相互作用而连接在一起的。碱基的化学结构显示, A 与 T 相互作用(而不与其他任何碱基作用), G 与 C 相互作用(而不与其他任何碱基作用)。DNA 固有的内部双链以及碱基对相互作用的精确性在杂交过程中被利用。杂交是两条互补的单核苷酸链形成一条稳定的双股螺旋^[1]。杂交反应可以在溶液中两个互补分子间发生。或在一个溶液的分子和固定在固相载体上的互补分子间发生,因为 DNA 芯片分析是利用了标有荧光的单链分子与 DNA 芯片上的单链序列的杂交反应,所以杂交是生物化学过程,这是 DNA 微阵列的基础。

生物体基因组含有蛋白编码区和非编码区两部分(表 1-1),包括外显子和内含子,启动子和基因调控区、复制起点、端粒以及无功能的基因间 DNA。基因复制的多少和染色体的数目或多倍体以及该 DNA 一级序列结构中单个碱基的差异都可以用基因组分析进行定量检测。因遗传而导致的单个碱基的改变,通常称为多态性;而生物体在其生命过程中获得的单个碱基的改变则称为突变。总之,基因组分析通过描述基因及基因调控区精确的核苷酸序列,反映了一个生物体遗传潜能(genetic potential)的精细图,但这种 DNA 水平的基因组分析并不能提供一种检测基因表达的方法,因为基因表达是编码序列的 RNA 及蛋白质合成的过程。

表 1-1 基因组信息容量

信 息	脱氧核糖核酸	核糖核酸	蛋白质
遗传图	完整	部分	部分
基因组表现度 ^a	100 %	3 %	3 %
类型	编码和非编码	编码	编码
形式	双链	单链	折叠
构成成分	核酸	核酸	氨基酸
杂交	有	有	无
碱基	A, G, C, T	A, G, C, U	无
总结构元件	4	4	20
多倍体效应 ^b	有	有	有
转录效应 ^b	无	有	有
翻译效应 ^b	无	无	有
生化修饰	甲基化	聚腺苷酸化帽结构	磷酸化酰基化
微阵列分析	有	有	无 ^c

a. 人类。

b. 第二效应除外。

c. 已发表的。

一般认为,来自同一生物体的所有细胞含有完全相同的基因组,来自同种而不同个体的基因组约 99.9% 完全相同。对生物体庞大的基因组来说个体间 0.1 % 的多态性发生率是很重要的^[2]。根据任何两个人体基因组的全序列分析,预计约有 300 万种多态性(30 亿碱基 $\times$ 0.1%)。如果单个碱基的改变发生在蛋白质编码区,多态性可改变该蛋白质的序列,进而改变该基因产物的生物化学活性。假定人类基因组有约 3% 的编码序列,约 1% 的多态性改