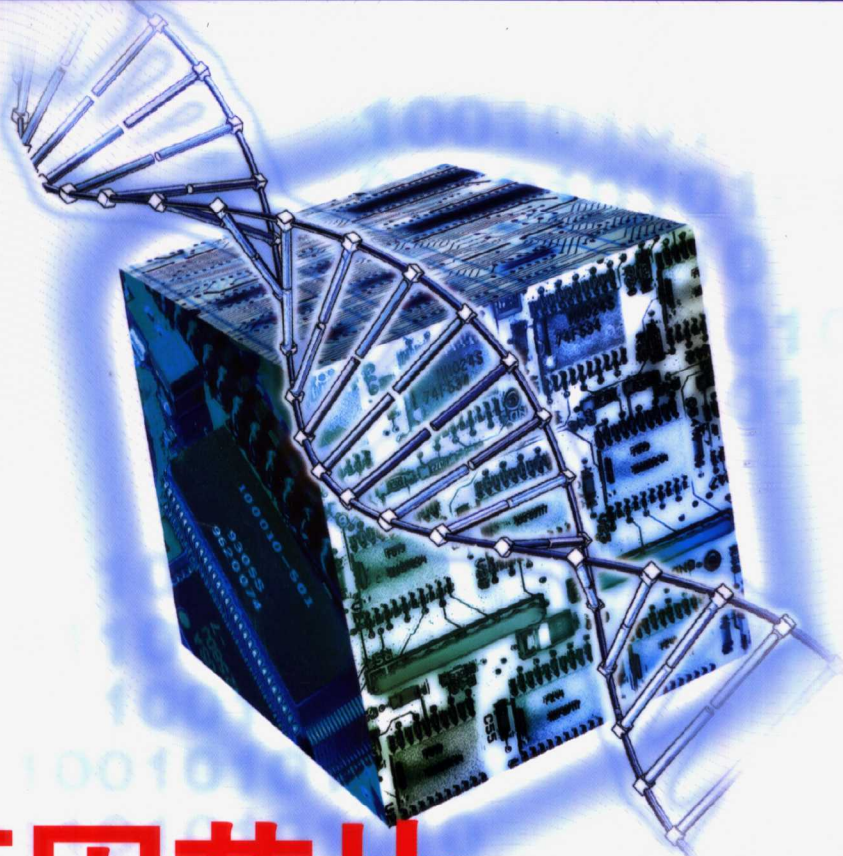


国家科普知识重点图书

高新技术科普丛书



基因芯片技术

——解码生命

李瑶 主编

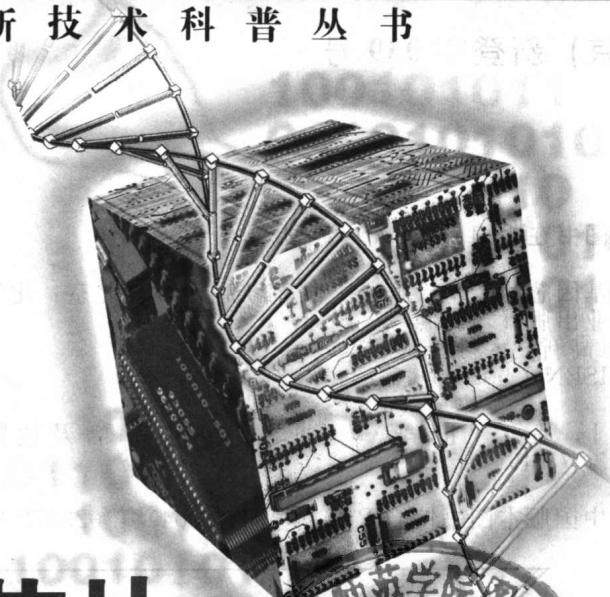
Chemical Industry Press



化学工业出版社

国家科普知识重点图书

高新技术科普丛书



基因芯片技术★

——解码生命

李 瑶 主编



化学工业出版社

· 北 京 ·

MAK03/4

(京)新登字 039 号

图书在版编目 (CIP) 数据

基因芯片技术: 解码生命/李瑶主编. —北京: 化学工业出版社, 2004.3
(高新技术科普丛书)
ISBN 7-5025-5415-7

I. 基… II. 李… III. 基因-芯片-普及读物
IV. Q78-49

中国版本图书馆 CIP 数据核字 (2004) 第 023372 号

高新技术科普丛书

基因芯片技术——解码生命

李 瑶 主编

责任编辑: 周 旭

文字编辑: 周 侗

责任校对: 陈 静

封面设计: 关 飞

*

化学工业出版社出版发行

(北京市朝阳区惠新里 3 号 邮政编码 100029)

发行电话: (010) 64982530

<http://www.cip.com.cn>

*

新华书店北京发行所经销

北京云浩印刷有限责任公司印刷

三河市海波装订厂装订

开本 720 毫米×1000 毫米 1/16 印张 14½ 彩插 2 字数 191 千字

2004 年 5 月第 1 版 2004 年 5 月北京第 1 次印刷

ISBN 7-5025-5415-7/Q·93

定 价: 28.00 元

版权所有 违者必究

该书如有缺页、倒页、脱页者, 本社发行部负责退换

《高新技术科普丛书》编委会

主任

路甬祥 中国科学院院长，中国科学院院士，
中国工程院院士

委员

汪家鼎 清华大学教授，中国科学院院士
闵恩泽 中国石油化工集团公司石油化工科学研究院教授，
中国科学院院士，中国工程院院士
袁 权 中国科学院大连化学物理研究所研究员，中国科学院院士
朱清时 中国科学技术大学教授，中国科学院院士
孙优贤 浙江大学教授，中国工程院院士
张立德 中国科学院固体物理研究所研究员
徐静安 上海化工研究院（教授级）高级工程师
冯孝庭 西南化工研究设计院（教授级）高级工程师

《基因芯片技术——解码生命》

参 编 人 员

主 编

李 瑶 (复旦大学生命科学院遗传研究所遗传工程国家重点
实验室)

副主编

付旭平 (复旦大学生命科学院遗传研究所遗传工程国家重点
实验室)

陈 沁 (上海博星基因芯片有限责任公司)

吴 海 (复旦大学生命科学院遗传研究所遗传工程国家重点
实验室)

李成涛 (上海博星基因芯片有限责任公司)

序

数万年来，人类一直在了解、开发、利用我国周围的自然界，同时不断地认识着自身，科学技术也从一开始就随着人类的生存需求而产生和发展着。人类发展史充分验证了邓小平“科学技术是第一生产力”的论断。科学技术的发展，促进了人类文明和社会的发展。

21 世纪是信息时代，21 世纪是生命科技的世纪，21 世纪是新材料和先进制造技术迅速发展和广泛应用的时代，21 世纪是高效、洁净和安全利用新能源的时代，21 世纪是人类向空间、海洋、地球内部不断拓展的世纪，21 世纪是自然科学发生重大变革、取得突破性进展的时代。科学技术的发展、新技术的不断涌现，必将引起新的产业革命，对我国这样的发展中国家来说，既是挑战，也是机遇，而能否抓住发展机遇，关键在于提高全民族的科学文化水平，造就一支具有科学精神、懂得科学方法、具有知识创新和技术创新能力的高素质劳动者队伍。所以，发展教育和普及科学知识、弘扬科学精神、提倡科学方法是我们应对世纪挑战的首要策略。为此，1999 年 8 月，江总书记在视察中国科学院大连化学物理研究所时进一步强调了科普工作的重要性：“在加强科技进步和创新的同时，我们应该大力加强全社会的科学普及工作，努力提高全民族的科学文化素质。这项工作做好了，就可以为科技进步和创新提供广泛的群众基础。”

为了普及和推广高新技术，化学工业出版社组织几位两院院士和专家编写了《高新技术科普丛书》。本套丛书的特点是：介绍当今科学产业中的一些高新技术原理、特点、重要地位、应用及产业化的现状与发展前景；突出“新”，介绍的新技术、新理论和新方法不仅经实践证明是成熟、可靠的，而且是有应用前景的实用技术；力求深入浅出，图文并茂，知识性、科学性与通俗性、可读性及趣味性的统一，并充分体现科学思想和科学精神对开拓创新的重要作用。

《高新技术科普丛书》涉及与我国经济和社会可持续发展密切相关的高新技术，第一批9个分册包括绿色化学与化工、基因工程技术、纳米技术、高效环境友好的发电方式——燃料电池、最新分离技术（如超临界流体萃取、吸附分离技术、膜技术）、化学激光、生物农药等。本套丛书以后还将陆续组织出版多种高新技术分册。相信该套科普丛书对宣传普及科技知识、科学方法和科学精神，正确地理解、掌握科学，提高全民族的素质将会起到积极的作用。

洛育祥

2000年9月

前言

探索自然的奥秘是人类发展的重要原因和动力，当今社会生命科学的飞速发展，是因为和人类自身的命运息息相关。人们渴望了解自我，提高生命质量，健康长寿并远离病痛。毫无疑问，解码生命、揭示生命发生发展的规律是我们这个时代的神圣使命之一。

生命现象是自然界的一大奇迹，生命里最本质的内容就是遗传，即亲代的性状在子代得到表现。从 1866 年遗传学之父孟德尔（Mendel）提出遗传因子开始，人类对生命的认识进入理性阶段。随着研究手段的不断深入，1953 年沃森-克里克（Watson-Crick）提出核酸 DNA 分子的双螺旋结构，把生命科学引入分子研究水平。生物大分子核酸和蛋白质是生命的基础，是生命运动的执行与体现，也是生命科学研究的主干。于是，有了著名的全球性的研究项目：人类基因组计划（human genomic project, HGP）。随着 HGP 的测序完成，人们已经做好充分的准备去迎接后基因组时代——蛋白质组学的到来。

生命科学一直是一门技术性很强的学科，研究领域的每一次飞跃都得益于技术和方法学上的重大突破。起始于 HGP 应用研究的基因芯片技术又给了人们这样的期望。本书就基因芯片技术和生命研究的关系及发展做一介绍，将这一重要的内容介绍给大家，希望能引导更多的有志之士参与到这一领域，一起见证人类科技发展的辉煌篇章。

由于篇幅的限制，本书着重介绍基因芯片的几个主要方

面，第1章概括介绍了基因芯片相关的基本知识，第2章介绍基因芯片的制备和检测技术，第3章介绍基因芯片在功能基因组研究中的应用，第4章是与基因芯片相关的生物信息学介绍，第5章重点介绍基因芯片在基因组水平检测中的应用。由于该领域进展迅速，可能有些内容没有全部包括；同时我们深知自己的水平有限，错误和疏漏在所难免，欢迎热心的读者提出批评。

在本书的编写过程中，我们得到了裘敏燕、戴建凉、靳超、徐小东和刘三震的热情帮助，在此表示深深的谢意。

2004年2月
编著者

目 录

第1章 基因、基因组与生物芯片	1
1.1 核酸	3
1.2 核酸检测方法	6
1.3 基因和基因组	16
1.4 生物芯片技术概述	32
 第2章 基因芯片制备和检测技术	 39
2.1 基因芯片分类	41
2.2 基因芯片的制备	46
2.3 图片获取	67
 第3章 表达谱基因芯片	 77
3.1 表达谱基因芯片概述	79
3.2 表达谱基因芯片原理及流程	80
3.3 表达谱基因芯片应用——功能基因组学研究 ...	98
 第4章 基因芯片生物信息学	 113

4.1	图像处理	115
4.2	数据的预处理	122
4.3	数据的分析	130
4.4	调控元件与基因网络	142
4.5	基因芯片数据库	144
4.6	基因序列分析	156

第5章 寡核苷酸芯片与 DNA 检测 171

5.1	寡核苷酸芯片的方法学	173
5.2	寡核苷酸芯片技术的应用现状	179
5.3	生物芯片的发展和应用趋势	213

第1章

基因、基因组与生物芯片

核酸

核酸检测方法

基因和基因组

生物芯片技术概述

1.1 核酸

1.1.1 核酸是生命遗传信息的载体

生命是世界上物质存在和发展的高级形式。核酸和蛋白质两大类大分子是构成生命形态结构和生理功能最重要的物质基础，二者分子结构复杂，分子量巨大，分子中载有生命活动的信息，担负着各种各样的生理功能。在生命有机体中核酸又是蛋白质合成的编码器，因此核酸在生命的生长、发育、繁殖和疾病的控制中起着举足轻重的作用。

核酸最初从细胞核中分离出来，具有强酸性，后来发现在细胞质以及线粒体、叶绿体等细胞器中都有核酸，它是遗传变异的物质基础，控制着生命活动的有序性。

1.1.2 核酸的化学组分和种类

核酸是由许多个单核苷酸通过磷酸二酯键聚合而成的高分子有机化合物。一个单核苷酸是由一分子五碳糖、一分子磷酸和一分子含氮有机碱（碱基）缩合而成的。

因组成核酸分子的五碳糖分为核糖或脱氧核糖，因此核酸可分为核糖核酸（RNA）和脱氧核糖核酸（DNA），所有生物细胞都含有这两类核酸。

含氮有机碱分为嘧啶碱和嘌呤碱两类。嘧啶碱有胞嘧啶（cytosine, C）和胸腺嘧啶（thymine, T），嘌呤碱有腺嘌呤（adenine, A）和鸟嘌呤（guanine, G）。遗传信息就是储存在组成核酸的 A、T、C、G 这四种碱基的不同排列顺序中。

1.1.3 DNA 和 RNA 的结构

DNA 分子通常以两股反向平行互补的脱氧多核苷酸长链的形式存在，围绕一个主轴向右形成的双螺旋结构分子，即所谓的 Watson-Crick 模型。

脱氧核糖与磷酸残基在每股长链的外侧，构成 DNA 分子的骨架。碱基位于长链的内侧，A、T、C、G 四种碱基的排列方式称为碱基序列 (base sequence)。其中两股长链的碱基总是按照 A : T、C : G 通过氢键互补配对的，其中 A : T 之间有两个氢键，C : G 之间有三个氢键，这一规律称为碱基的配对原则 (图 1-1)。

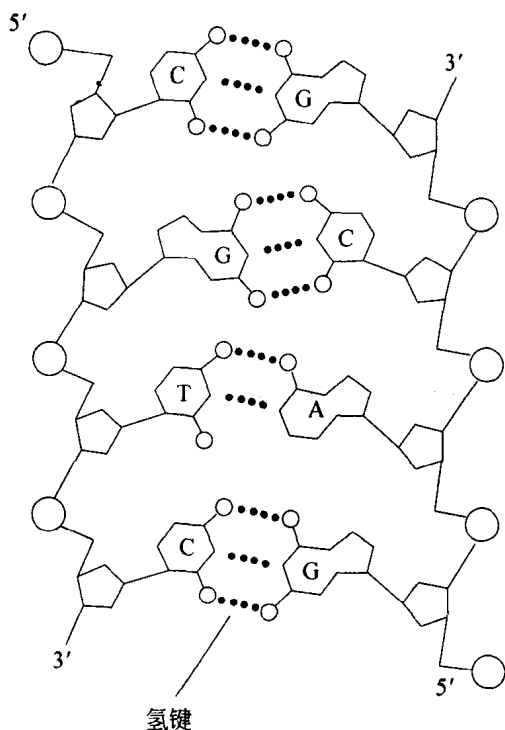


图 1-1 DNA 的双螺旋结构

多数 RNA 分子的结构为单链多核苷酸长链，有的 RNA 单链的某些区段能够自身互补折叠形成局部双链，由于配对的位置和方向的不同，使 RNA 的高级结构较 DNA 复杂，如常见的发夹结构和颈

环结构。某些 RNA 病毒如呼吸道肠道病毒是双链 (±) RNA 病毒。

真核生物的 RNA 分子根据其结构和行使的生物学功能的不同可分为信使 RNA、转移 RNA 及核糖体 RNA。

1.1.4 DNA 与 RNA 的功能

DNA 与 RNA 在结构上的差异决定了它们具备不同的功能。

1.1.4.1 DNA 的主要生物学功能

(1) **编码储存遗传信息** 在染色体的特定位置,按特定碱基顺序排列的一段 DNA 构成了带有某种生物遗传信息的基因。遗传信息量与组成 DNA 的碱基对的数量成正比,越复杂高等的生物,它们的 DNA 的含量越高。人的染色体 DNA 就含有 30×10^8 bp (碱基对),约含 3 万多个基因。而低等生物的大肠杆菌就只有 4.7×10^6 bp,含有约 3500 个基因。

(2) **自我复制** DNA 能以自身分子作模板,自我复制合成新的 DNA。复制时依靠 DNA 合成酶系统,以 DNA 解链后的每条单链为模板,按照 A=T、C≡G 的碱基配对原则,利用单核苷酸合成与自身结构完全相同的两个 DNA 分子。这种半保留的复制方式严格确保了遗传物质的稳定性和遗传信息的忠实性。

(3) **突变和畸变** 构成基因组分的核苷酸的种类、数量和排列顺序发生变化,就会引起基因突变和染色体畸变。一方面半保留复制保证了物种的遗传稳定性,可是另一方面,生物体在内外环境的影响下 DNA 分子发生了碱基的置换、缺失、插入或融合,就会导致指导合成的蛋白质的氨基酸发生改变,从而引发所谓的基因突变。大片段遗传物质的改变甚至是整条染色体的结构和数目的改变就导致了染色体畸变。基因突变和染色体畸变是导致生物遗传性疾病的原因。

(4) **转录和翻译** DNA 分子可以以一条信息链为模板互补合成不同的 RNA 分子,再通过 RNA 分子将 DNA 分子中的遗传信息表达为相应的遗传性状。

1.1.4.2 RNA 的主要生物学功能

20 世纪 80 年代 RNA 的研究揭示了 RNA 功能的多样性,它不仅仅是遗传信息由 DNA 到蛋白质的中间传递体,虽然这是它的核心功能。归纳起来, RNA 有 5 类功能:①控制蛋白质的合成;②作用于 RNA 转录后的加工和修饰;③对基因表达与细胞功能的调节;④生物催化功能;⑤遗传信息的加工与进化。

此外不同 RNA 分子其生物学功能又各不相同。信使 RNA (messenger RNA, mRNA) 的作用是从细胞核 DNA 分子上转录遗传信息,带到细胞质中的核糖体上,作为合成蛋白质的指令;转移 RNA (transfer RNA, tRNA) 的作用是在遗传信息表达过程中运输活化氨基酸到核糖体的特定部位,并把 mRNA 的碱基顺序翻译为多肽链的氨基酸顺序,每种 tRNA 只能特异性地连接和转运一种氨基酸;核糖体 RNA (ribosomal RNA, rRNA) 是构成核糖体的重要成分,与蛋白质相结合后维系核糖体的立体结构。另外还有与 mRNA 的前体加工有关的小分子 RNA (snRNA),以及在 mRNA 的翻译过程中起调节因子作用的反义 RNA (antisense RNA)。

总之,生物机体通过 DNA 复制而将遗传信息由亲代传递给子代;通过 RNA 转录和翻译而使遗传信息在子代得以表达。

1.2 核酸检测方法

由于对基因的组成成分——核酸 (DNA/RNA) 检测的实验技术不断提高,对基因乃至生命研究也逐渐深入。核酸的主要检测方法包括分光光度法、琼脂糖及聚丙烯酰胺凝胶电泳法、测序法、分子杂交法等。这些方法从不同的角度揭示了核酸的理化特性、带电