



生命科学前沿及应用生物技术

基因组学 核心实验方法

[英] Mike Starkey Ramnath Elaswarapu 著

于 军 主译

GENOMICS: ESSENTIAL METHODS



科学出版社

生命科学前沿及应用生物技术大系·典藏版

基因组学：核心实验方法

Genomics: Essential Methods

〔英〕Mike Starkey Ramnath Elaswarapu 著

于 军 主译

科学出版社

北 京

图字：01-2011-4859 号

内 容 简 介

应用生物技术大系和现代生命科学前沿系列图书分别被列为“十一五”和“十二五”国家重点图书出版规划项目。本丛书针对生命科学领域前沿重点发展方向以及应用生物技术领域的新成果、新思路、新方法和新技术，全面展示了其最新的发展动态，涵盖了基础理论和主要技术方法，呈现了新的概念与理论、技术，在更深层次上阐明了生命的本质规律，给人们提供了新的认识生命本质的手段，也为生物技术服务于人类开辟了新的途径。涉及领域包括生物医药、干细胞技术、工业微生物学、蛋白质及蛋白质组、系统生物学、合成生物学、生物材料、农业生物技术、环境生物技术、海洋生物技术、生物资源与安全等。

All rights reserved. Authorized translation from the English language edition published by John Wiley & Sons Limited. Responsibility for the accuracy of the translation rests solely with Science Press Ltd. And is not the responsibility of John Wiley & Sons Limited. No part of this book may be reproduced in any form without the written permission of the original copyright holder, John Wiley & Sons Limited.

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal

图书在版编目(CIP)数据

生命科学前沿及应用生物技术大系：典藏版/舒红兵等编著. —北京：科学出版社, 2016

ISBN 978-7-03-047487-2

I. ①现… II. ①舒… III. ①生命科学—研究②生物工程—研究 IV. ①Q1-0②Q819

中国版本图书馆 CIP 数据核字(2016)第 043876 号

责任编辑：王 静 李 悦

责任印制：张 伟 / 封面设计：刘新新

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>

北京厚诚则铭印刷科技有限公司印刷

科学出版社发行 各地新华书店经销

*

2016 年 7 月第 一 版 开本：787×1092 1/16

2016 年 7 月第一次印刷 印张：2108

字数：49 985 000

定价：8900.00 元

(如有印装质量问题，我社负责调换)

《新生物学丛书》专家委员会成员名单

主 任：蒲慕明

副主任：吴家睿

专家委员会成员(按汉语拼音排序)：

昌增益	陈洛南	陈晔光	邓兴旺	高 福
韩忠朝	贺福初	蒋华良	金 力	李家洋
林其谁	马克平	孟安明	裴 钢	饶 毅
饶子和	施一公	舒红兵	王 琛	王梅祥
王小宁	吴仲义	徐安龙	薛红卫	詹启敏
赵国屏	赵立平	钟 扬	朱 桢	

译者名单

主 译 于 军

翻译人员(按姓氏汉语拼音排序)

李欣刚 肖景发 吴佳妍 孟庆姝

《新生物学丛书》

丛书序

当前,一场新的生物学革命正在展开。为此,美国国家科学院研究理事会于 2009 年发布了一份战略研究报告,提出一个“新生物学”(New Biology)时代即将来临。这个“新生物学”,一方面是生物学内部各种分支学科的重组与融合,另一方面是化学、物理、信息科学、材料科学等众多非生命学科与生物学的紧密交叉与整合。

在这样一个全球生命科学发展变革的时代,我国的生命科学研究也正在高速发展,并进入了一个充满机遇和挑战的黄金期。在这个时期,将会产生许多具有影响力、推动力的科研成果。因此,有必要通过系统性集成和出版相关主题的国内外优秀图书,为后人留下一笔宝贵的“新生物学”时代精神财富。

科学出版社联合国内一批有志于推进生命科学发展的专家与学者,联合打造了一个 21 世纪中国生命科学的传播平台——《新生物学丛书》。希望通过这套丛书的出版,记录生命科学的进步,传递对生物技术发展的梦想。

《新生物学丛书》下设三个子系列:科学风向标,着重收集科学发展战略和态势分析报告,为科学管理者和科研人员展示科学的最新动向;科学百家园,重点收录国内外专家与学者的科研专著,为专业工作者提供新思想和新方法;科学新视窗,主要发表高级科普著作,为不同领域的研究人员和科学爱好者普及生命科学的前沿知识。

如果说科学出版社是一个“支点”,这套丛书就像一根“杠杆”,那么读者就能够借助这根“杠杆”成为撬动“地球”的人。编委会相信,不同类型的读者都能够从这套丛书中得到新的知识信息,获得思考与启迪。

《新生物学丛书》专家委员会

主 任:蒲慕明

副主任:吴家睿

2012 年 3 月

译者前言

《基因组学:核心实验方法》一书主要阐述了基因组学及其衍生学科的各种关键技术,涉及内容广泛,涵盖实验手段和计算机模拟等方法。尤其是其中的实验方法都是依据专家撰写的实验报告而建立的,并来自于实验室里以应用新技术进行研究的第一线数据。特点是实用性强、可靠性强、专业指导性强。

由于基因组学的影响力日益扩大,新技术不断产生,海量数据迅速积累,基因组学的核心实验方法已经被应用到生命科学研究的各个领域当中。越来越多的研究者在试图利用基因组学知识和成果的同时,也越来越重视一系列可靠的研究方法和研究手段的掌握。因此,本书非常适合于从事基因组学及其在生命科学领域的各个衍生学科研究的研究生和青年学者阅读。他们不仅可以在这里了解到实验技术的详细步骤,还可以掌握实验技术的根本思想和原理,有助于进一步提出其他的替代方法。

衷心感谢参加翻译的肖景发博士、吴佳妍博士、孟庆姝博士和李欣刚博士为本书所作的贡献,他们都是多年从事基因组学前沿工作的青年学者,具有很好的语言能力和专业水平。

在译校过程中,虽力求忠于原意、通顺信达,但限于水平有限,谬误之处在所难免,敬希读者批评指正。

于 军

2011年11月11日

前 言

随着人类基因组测序计划完成,基因组学的影响力迅速扩大。目前,由于新技术的产生以及海量数据的积累,使得以前难以想象的大规模实验具有可行性,进而使人们对生物系统的认识发生了巨大改变。

后基因组时代已经拉开序幕,它提供更多对新技术进行相关学术研究和商业开发的机会,越来越多的研究者在试图利用基因组学的知识和成果时,发现掌握一系列可靠的研究方法尤为重要。基于此想法,我们依据主要在实验室以应用新技术进行研究的专家们撰写的实验报告而建立了一系列研究方法。

本书主要介绍了实验技术详细步骤,同时还阐明了技术的根本思想,有助于研究者提出其他可替代的方法。本书和其他的工具书所不同的地方在于,所有作者在介绍实验步骤的同时投入大量精力来阐述技术的原理,因此大大增加了本书的实用价值。

本书的宗旨在于阐述基因组学及其衍生学科的关键技术,并不是广泛收集基因组学实验技术本身。这些技术涉及内容非常广泛,主要包括遗传变异探测技术、mRNA 表达谱和基因组 DNA 拷贝数分析、通过实验或计算机模拟进行蛋白质分析,以及基因疗法的应用等。

第 1~4 章主要涉及基因组分析方法的程序,包括检测染色体拷贝数变化的几种策略。样本数据处理的流程中(第 1 章和第 3 章)充分考虑了进行研究中新鲜组织的难收集性和组织病理学部门收集的组织资源库太大等问题。单核苷酸多态性识别技术(第 2 章)是在疾病易感性和药物基因组学研究方面应用高精度全基因组关联分析方法(第 4 章)的重要工具。

第 5~7 章介绍了基因表达分析技术,转录组分析的发展趋势是利用 RNA 扩增技术得到更为严格定义的细胞群体的表达谱。一方面,由于 RT-PCR 技术(第 6 章)具有特异性和灵敏性强等特点,其被广泛应用于 RNA 的检测和量化;另一方面,许多应用需要外源基因在体内进行表达,第 7 章也介绍了一系列体内系统转移基因的新方法。

第 8 章详细描述了利用酵母双杂交方法进行蛋白质相互作用的研究,并且获取高度可信的蛋白质相互作用的数据信息,蛋白质-蛋白质相互作用研究有助于理解这些基因的生物功能和阐明生化反应途径。确定基因功能是功能基因组学的关键,第 9 章和第 10 章阐述功能基因组学的研究方法。

基因疗法是通过对与致病机制相关的缺陷基因进行修饰而达到治疗疾病的目的,其逐渐被认为是疾病治疗的一种可行性途径,第 11 章和第 12 章着重解决相关问题并提出一般性策略。

蛋白质组分析通常被认为是转录组研究的延伸,最后一章着重介绍基因组学专家如何分析蛋白质表达谱。这一章和前面的章节风格有所不同,主要描述应用蛋白质组学技

术时遇到的相关问题的解决策略。

我们衷心希望本书中关于这些实验技术和基本原理的阐述能够帮助该领域中初级和资深的研究工作者顺利完成实验。

最后谢谢所有作出贡献的作者 David Hames、Clare Boomer 和 Jonathan Ray 以及 Wiley-Blackwell 出版社的工作人员,最重要的是他们的家人在编辑本书期间给予的支持。

Mike Starkey
Ramnath Elaswarapu

目 录

《新生物学丛书》丛书序

译者前言

前言

1	基因拷贝数量变化的高精度分析	1
1.1	简介	1
1.2	方法和途径	2
1.2.1	寡核苷酸比较基因组杂交(aCGH)芯片	2
1.2.2	单核苷酸多态性比较基因组杂交(aCGH)芯片	13
1.2.3	多重连接探针扩增	17
1.3	疑难解答	25
	参考文献	26
2	遗传图谱中的多态性标记识别	29
2.1	简介	29
2.2	方法和途径	30
2.2.1	基因变异的知识库	30
2.2.2	基因变异的靶向重测序	31
2.3	疑难解答	40
2.3.1	引物设计	40
2.3.2	PCR 扩增	41
2.3.3	二进制文件的使用	41
2.3.4	Phred/Phrap 软件	41
	参考文献	41
3	基于 SNP 芯片的基因分型和杂合性缺失分析	44
3.1	简介	44
3.2	方法和途径	44
3.2.1	芯片	44
3.2.2	基因分型	45
3.2.3	连锁关联分析	45
3.2.4	甲醛固定石蜡包埋组织	46
3.2.5	杂合性缺失	52
3.3	疑难解答	56
	参考文献	57

4 复杂性状的基因定位	60
4.1 简介	60
4.2 方法和途径	61
4.2.1 关联分析方法:随机样本	61
4.2.2 关联方法:基于家系样本	72
4.2.3 连锁分析:使用 LOD 值作为参数的分析方法	73
4.2.4 连锁分析:非参数方法	74
4.2.5 总结	75
4.3 疑难解答	75
4.3.1 合并数据集	75
参考文献	76
5 针对单细胞敏感性的 RNA 扩增技术	82
5.1 简介	82
5.1.1 RNA 扩增的目的	82
5.1.2 扩增的方法	84
5.2 方法和途径	89
5.2.1 T7RNA 聚合酶的体外转录	89
5.2.2 全局 RT-PCR	96
5.3 疑难解答	103
参考文献	104
6 表达谱分析的实时定量聚合酶链反应技术	107
6.1 简介	107
6.2 方法和途径	108
6.2.1 样本筛选	108
6.2.2 提取 RNA	109
6.2.3 临床样本和环境样本	112
6.2.4 逆转录	114
6.2.5 SYBR Green I 染料检测的荧光定量 PCR	118
6.2.6 寡核苷酸探针标记的定量 PCR	122
6.2.7 定量方法	124
6.2.8 实时定量 PCR 的标准化	127
6.3 疑难解答	127
6.3.1 未扩增、扩增量少、扩增起步晚	127
6.3.2 缺少模板组或阴性对照组	130
6.3.3 无逆转录酶对照组	130
6.3.4 形成引物二聚体	131
6.3.5 SYBR Green I 溶解曲线出现多峰	131
6.3.6 在样本重复实验 3 次,至少 5 倍对数稀释情况下,相关系数 <0.98 ,其标准曲线	

不可信	131
6.3.7 不稳定的扩增曲线图或大幅度的井间变化	131
参考文献	132
7 哺乳动物细胞中的基因表达	137
7.1 简介	137
7.1.1 人工染色体和转基因技术	138
7.1.2 基因转移和表达	139
7.1.3 位置效应和核染色质	139
7.1.4 组织特异性调控元件	139
7.1.5 持续表达和染色质绝缘体	139
7.2 方法和途径	140
7.2.1 哺乳动物细胞的位点特异性染色体重组	140
7.2.2 质粒要求	142
7.2.3 染色体转移	144
7.3 疑难解答	149
参考文献	149
8 酵母双杂交在分析大量蛋白质相互作用中的应用	152
8.1 概述	152
8.2 方法和途径	154
8.2.1 建立大量“诱饵”或“猎物”蛋白克隆	154
8.2.2 生成兼容性重组插入用于缺口修复克隆	156
8.2.3 缺口修复反应	157
8.2.4 阳性转化株鉴定	159
8.2.5 酵母菌落 PCR	159
8.2.6 “诱饵”与“猎物”克隆自激活检测	161
8.2.7 靶向矩阵法 Y2H 筛选	162
8.3 疑难解答	166
参考文献	166
9 蛋白质功能预测	168
9.1 引言	168
9.2 方法和途径	168
9.2.1 注释策略	169
9.2.2 多个蛋白质识别系统的应用	171
9.2.3 序列同源性	172
9.2.4 系统发生关系	175
9.2.5 序列衍生的功能和化学性质	177
9.2.6 蛋白质-蛋白质相互作用图谱	178
9.3 故障排查	180

参考文献	181
10 通过基因工程小鼠阐释基因功能	186
10.1 引言	186
10.2 方法和途径	187
10.2.1 小鼠目标基因剔除原理	187
10.2.2 小鼠基因打靶策略	189
10.2.3 通过重组工程从 BAC 中获得 DNA	191
10.2.4 胚胎干细胞和胚胎成纤维细胞培养	195
10.2.5 嵌合体配对和下游的应用	215
10.3 疑难解答	216
参考文献	217
11 基因转移的载体系统	220
11.1 引言	220
11.2 方法和途径	220
11.2.1 理想的基因治疗载体	220
11.2.2 质粒设计	222
11.2.3 病毒载体	222
11.2.4 非病毒 DNA 载体	232
11.2.5 鉴定非病毒载体的物理性质	235
11.2.6 优化体外基因传递	236
11.2.7 优化方案	239
11.2.8 报告基因和分析	240
11.2.9 细胞毒性分析	240
11.2.10 非病毒载体的未来发展	240
11.3 疑难解答	241
11.3.1 一般问题	241
参考文献	242
12 基因治疗策略:构建 AAV 特洛伊木马	249
12.1 简介	249
12.1.1 基因治疗的常规策略:基本方法	249
12.1.2 基因治疗策略:基因转染细胞	253
12.1.3 病毒载体	254
12.1.4 重组 AAV 病毒的生产、纯化和滴度检测	256
12.2 方法和途径	257
12.3 疑难解答	267
参考文献	267
13 蛋白质组学技术简介	270

13.1 简介.....	270
13.2 方法和途径.....	271
13.2.1 基于凝胶的策略	271
13.2.2 LC/MS 策略	274
13.2.3 基质辅助激光解吸电离(MALDI)成像和概览	276
13.3 故障诊断.....	277
13.3.1 若干分解出来的特征和修饰	277
13.3.2 样品消耗、蛋白质识别及覆盖深度	278
13.3.3 统计强度	279
13.3.4 结论	279
参考文献.....	280
索引.....	284

1 基因拷贝数量变化的高精度分析

Mario Hermesen¹, Jordy Coffa², Bauke Ylstra³, Gerrit Meijer², Hans Morreau⁴,
Ronald van Eijk⁴, Jan Oosting⁴ and Tom van Wezel⁴

¹Department Otorrinolaringología, Instituto Universitario de Oncología del Principado de Asturias, Oviedo, Spain

²Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands

³Microarray Facility, VU University Medical Center, Amsterdam, The Netherlands

⁴Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

1.1 简介

1992 年 Kallioniemi 等首次提出了比较基因组杂交技术 (comparative genomic hybridization, CGH), 这标志着从全基因组角度分析 DNA 拷贝数变化的开始 [1]。该技术的基本原理是, 用不同的荧光染料通过缺口平移法分别标记待测的全基因组 DNA 样本和对照样本 DNA 制成的探针, 并与正常人的间期染色体进行共杂交, 用在染色体上显示的待测样本与对照样本的荧光强度的不同来反映待测样品基因组 DNA 表达状况的变化, 再借助于图像分析技术可对染色体拷贝数量的变化进行定量研究。

表面上看这项实验技术对非专门从事染色体组技术的实验室是非常困难, 但自从介绍该实验详细过程的文章 [2] 刊登之后, CGH 技术被广泛应用于各项研究中, 特别是在癌症遗传学研究方面。使用福尔马林固定石蜡包埋 (formalin-fixed and paraffin-embedded, FFPE) 方法处理 DNA 样品为后续基于临床数据研究肿瘤奠定了基础, 可以在肿瘤发生、扩散和转移等不同时期探测基因的变化 [3]。

经典的 CGH 技术只有 5~10Mb 的染色体带的分辨率。但 1997 年出现的比较基因组杂交芯片 (array comparative genomic hybridization, aCGH) 技术克服了这一“瓶颈” [4, 5]。该技术沿用了经典 CGH 技术的基本思想, 它的改进在于不是使用中期染色体组而是采用基因组 DNA 克隆芯片或者寡核苷酸作为杂交的靶标。aCGH 的优势在于其精度由 DNA 克隆数量或者寡核苷酸数量决定, 另一个优势在于不需要染色体组分型。目前, 主要有两种广泛应用的 aCGH, 分别是寡核苷酸芯片和单核苷酸多态性芯片 (single nucleotide polymorphism, SNP) [6]。

2002 年由 Schouten 等提出并公开发表的多重连接探针扩增 (multiple ligation-dependent probe amplification, MLPA) 技术是另一种 DNA 拷贝数分析技术 [7], 该技

术主要针对那些感兴趣且已知的特异性基因或者染色体区域。MLPA 技术仅需要 20ng 的 DNA 样品就能同时定量 50 个大约 50 个核苷酸长度的不同靶标。MLPA 的另一个优势在于其具有可重复性和特异性等特点，在日常诊断中应用此技术能够保证实验高效率 and 低成本完成。

对 FFPE 样品的基因组表达概况分析应用的重要性逐渐增长，在全世界范围内收集了大量临床跟踪的 FFPE 样品集，然而，FFPE 样品 DNA 的降解程度因其长度、固定方法及样品保存时间的不同而表现出不一致性。本章旨在描述寡核苷酸 aCGH 方法、SNP aCGH 方法和 MLPA 技术，还特别介绍了从 FFPE 样品中获取 DNA 的使用。这些技术主要应用于癌症研究，同时它们也同样适用于人类遗传疾病的 DNA 拷贝数畸变分析。

1.2 方法和途径

1.2.1 寡核苷酸比较基因组杂交 (aCGH) 芯片

第一个全基因组芯片是只包含 2400 个基因克隆的探针，探针的引入主要是通过细菌人工染色体 (bacterial artificial chromosome, BAC) 进行 [8]。寡核苷酸比较基因组杂交芯片和经典的 CGH 相比具有更高的实验精度，对于人类全基因组 30 亿碱基对来说，其精度平均可以达到 1Mb 左右 [1]，相对经典的 CGH 方法提高约 1 个数量级。如果要达到完全覆盖的精度，即把精度再提升一个数量级，大约需要制备 30 000 个 BAC 载体 [9]。然而，制备如此大量 BAC 载体用于 aCGH 实验是非常昂贵和耗时的，由于 BAC 载体较大的克隆容量，BAC aCGH 方法已经达到了其精度的限制。

一些实验室使用测定染色体拷贝数变化的 cDNA 芯片来研究基因表达谱 [10]。尽管 cDNA 芯片方法确实能得到一些有用信息，但是由于实验精度的限制还远远不能和寡核苷酸平台相提并论。寡核苷酸平台具有精度高、灵活性强和性价比高等特点 [6]，还能应用于所有已测序物种的全基因组。寡核苷酸芯片 CGH 和基因表达谱可以用来直接比较 mRNA 表达和 DNA 拷贝数比率。此外，在基因表达谱研究中经常使用或者设计新的寡核苷酸芯片，并且该方法已被广泛地应用。

商业寡核苷酸 aCGH 平台包括 Illumina (60mer)、Operon (70mer)、Affymetrix (25mer)、Agilent (60mer) 和 NimbleGen (45~85bp)，最新的芯片上的寡核苷酸数目可以达到 210 万个 [11]。随着对单个碱基的敏感度水平迅速达到单个 BAC 克隆载体的水平，寡核苷酸 aCGH 平台精度也不断改进。目前，并不是所有的寡核苷酸 aCGH 平台都可以检测到单个碱基的获得或缺失，但是 3~5 个相邻碱基的变化能可靠地检测到 [6, 11]。而且，随着实验技术的改进，目前在长达 50bp 以上的寡核苷酸芯片上进行分析，从 FFPE 肿瘤样本分离的 DNA 可以和新鲜材料 (实验方案 1.1~1.4) 的 DNA 相比拟 (图 1.1)。

寡核苷酸 aCGH 和其他 aCGH 的原理一样，都是采用不同标记的肿瘤组织的 DNA (实验方案 1.5) 和正常组织的 DNA 在寡核苷酸芯片上共杂交 (实验方案 1.6)，通过专用扫描器和数字图像处理软件计算出芯片上每一点的两种样本发出的荧光强度比值。

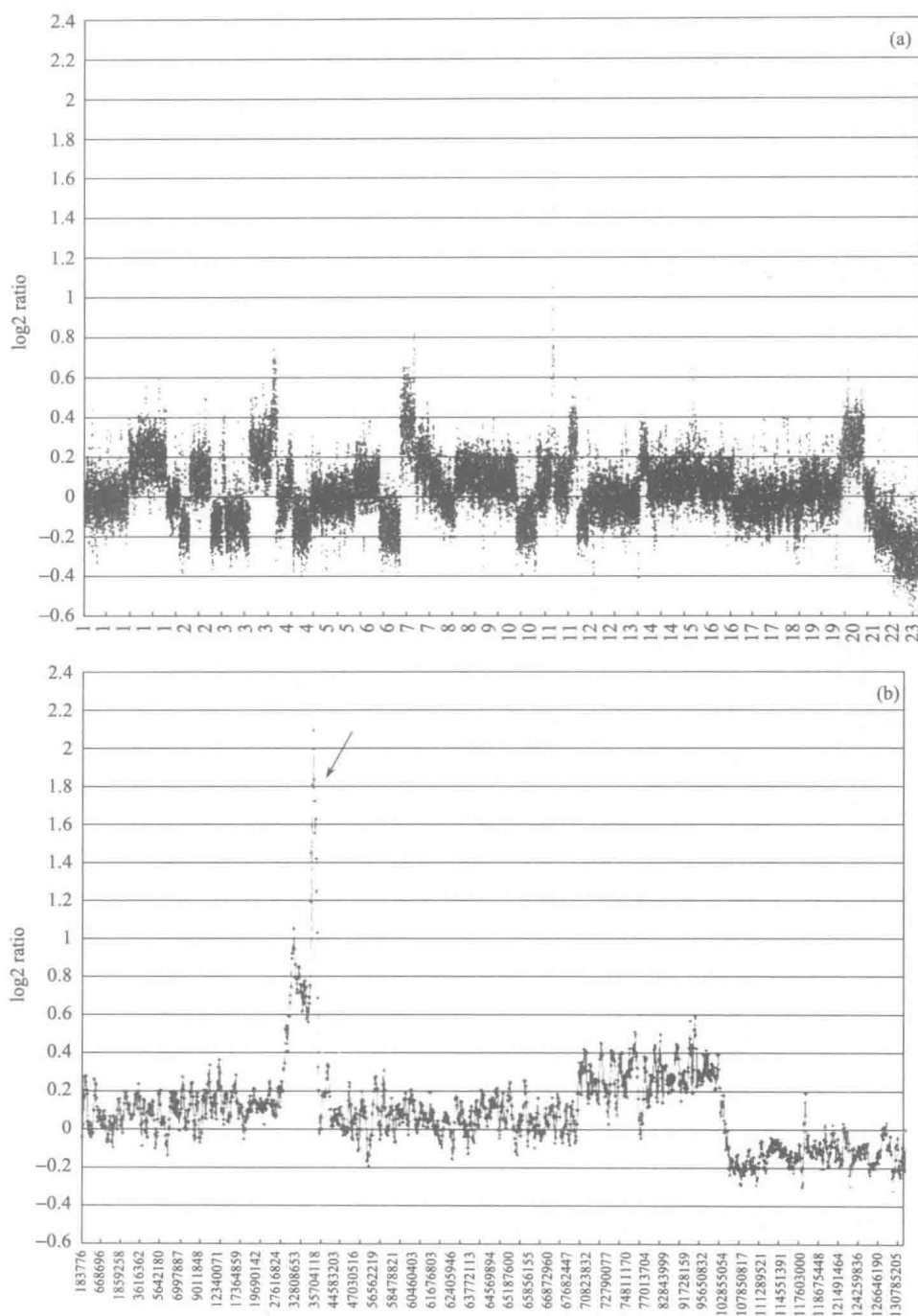


图 1.1 从 FFPE 抽取的肿瘤 DNA 的寡核苷酸 aCGH 分析实例。(a) 所有寡核苷酸按在染色体上的位置排序。(b) 11 号染色体的寡核苷酸, 观测到一个显著的扩增 (箭头所示)。