

基因组学

GENOMICS

杨焕明 编著

科学出版社

北京

内 容 简 介

本书是基因组学的综合性入门教科书,阐述了基因组学的源流和发展史、理念和基本概念、研究领域和核心技术、发展趋势及应用范例,特别是基因组学对生命科学和生物技术的全面贡献;讨论了相关的生命伦理与生物安全问题;从不同角度介绍了基因组学相关技术在生命科学和生物产业中的地位及其对人类社会的影响。

本书可作为已具备遗传学基础的本科生和基因组学相关专业的研究生教材使用,也可供从事与基因组学相关的生物产业,特别是农业与医学等领域的研究人员参考。

图书在版编目(CIP)数据

基因组学 / 杨焕明编著. — 北京: 科学出版社, 2016.10

ISBN 978-7-03-049902-8

I. 基… II. 杨… III. 基因组 IV. Q343.2

中国版本图书馆CIP数据核字(2016)第218026号

责任编辑: 李 敏 刘 超 / 责任校对: 彭 涛

责任印制: 肖 兴 / 封面设计: 李姗姗

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>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2016年10月第 一 版 开本: 889×1194 1/16

2017年4月第三次印刷 印张: 32 1/4

字数: 923 000

定价: 168.00元

(如有印装质量问题, 我社负责调换)

致 谢

本书自 2006 年开始搜集资料，2014 年 9 月开始动笔，集中了很多人的辛劳和心血，特别向他（她）们致以衷心的感谢：

首先，要感谢助理主编夏志和王晓玲，感谢他们从资料收集到最后定稿的全过程中所显示智慧和作出的贡献；

感谢参与本书有关章节初稿撰写的（以汉语拼音为序，下同）阿叁、董伟、方晓东、郭苗苗、胡学达、李波、倪培相、苏夜阳、沈玥、王晓玲、吴婷婷、夏志、赵宏翠、朱师达；

感谢参与本书最终校稿的董伟、刘韧、毛文静、王晓玲、夏志、张国捷、张丽、张秀清、朱燕楠；

感谢为本书构思、设计和手工描绘插图的李胜霆；

感谢参与本书资料收集的蔡锴晔、陈翠、陈姝婧、冯强、高长欣、栗东芳、李贵波、李京湘、李思淼、刘程章、刘汉奎、刘石平、聂超、田娟、王丽敏、吴静、叶辰、叶明芝、曾筱凡、周加利、张建国、张勇、赵倩、赵屹等；

感谢参与本书的组织、协调并提供资源的冯小黎、李生斌、汪建、王俊、徐讯、尹焯、杨碧澄、杨爽等；

感谢为本书提供部分素材与相关信息的华大基因相关部门和工作人员；

当然，还要衷心感谢拨冗阅读初稿，并给予鼓励、支持、指正的巴德年、强伯勤、陈润生、康乐、詹启敏和安锡培等诸位老师，特别是对《基因组学》（征求意见稿）提出宝贵修改意见的郝柏林、高翼之等老师；

衷心感谢对中国及世界的基因组学作出重大贡献的华大基因的几代工作人员及他们的家人，国家人类基因组研究中心（南方、北方）与所有其他中国团队。

此外，笔者特别感谢 Ewan Birney, Lars Bolund, Allen Bradley, Sydney Brenner, George Church, Brian Clark, Francis Collins, David Cox, Antoine Danchin, Richard Durbin, Hans Galjaard, Richard Gibbs, Thomas Hudson, Eric Lander, Marc van Montagu, Michael Morgan, Arno Motulsky, Søren Nørby, Stephen O'Brien, Maynard Olson, Aristides Patrinos, Richard Roberts, Edward Rubin, Yoshiyuki Sakaki, John Sulston, Harold Varmus, Friedrich Vogel, James Watson, Jeans Weissenbach 等国际师友。他们的学术思想和科学贡献，以及对笔者 17 年来的鼓励和指教，是成就本书的重要基石。

最后，还要感谢早逝的父亲和 90 高龄的老母亲、兄弟姐妹和所有其他家人的关怀，特别感谢夫人刘韧博士对本书编著的支持和贡献。

杨焕明

2016 年 9 月 9 日

前 言

15年前，当全球庆祝人类基因组计划(Human Genome Project, HGP)完成人类基因组的序列草图之时，我便坚信基因组学一定会作为一门授课学科，登上高等院校的大雅之堂，也萌生了撰写一本基因组学教科书的念头。

15年过去了，大家认为理所当然堪当此重任者——基因组学的实践者和重要贡献者、国际人类基因组测序协作组(International Human Genome Sequencing Consortium, IHGSC)的主要成员居然无一人动笔，至少迄今没有一人完成这一工作，我自己也没有将此心愿付诸行动，尽管在很多国家基因组学已成为高校的选修或必修课。也许最可能的解释是：他们或陶醉于基因组学一个又一个的新突破；或穷于应付或乐于应答该领域一个又一个的新挑战；或者担心这样的教科书只能招惹非议；最为担心的恐怕还是——这样一个日新月异的研究领域，所有的教科书都难以逃脱“刚一出版就已过时”的命运。此类实例屡见不鲜。

15年来，我们矢志不移，使中国屹立于基因组学的世界民族之林。今天，我们既然要编写这样的教科书，就一定要写好：以我们的亲身经历，写出既反映基因组学的世界潮流和发展趋势，又体现中国特色和反映中国贡献的教科书。我们要共同努力、一起建立并完善基因组学的教学体系，正如合作建立我们的基因组学的研究体系那样。

我们与读者分享的首先是基因组学的理念，其次才是围绕这一理念和核心技术的知识和经验。

我们力图使本书成为一本“好读、好懂、好记、好学、好教”的入门教科书，特别注重科学性、系统性、可读性和趣味性相结合。我们明白，一本好的教科书的编写是一项长期的艰巨任务，本书打算“修订不止，再版不已”，计划将根据学科的发展，隔年或数年修订再版，与时俱进地反映这一领域的进展、基本概念的更新；同时，也为了督促编者自己认真勘误，有错必纠。

“学而知之”乃求学之要旨——不管从师求学还是无师自通，一本好的教科书必不可少；“悟而知之”乃学习生命科学之优势——以“我是生命”来领悟基因组学；“习而知之”——“学而时习之，不亦乐乎？”乐在把书读薄。

写书难，写教科书更难，写一门新兴而又在快速发展的学科教科书更是难上加难。一本受欢迎的教科书的形成过程，正是编与读互动、学与教相长的过程。本书的命运如何，将取决于使用者与编写者现在开始的共同努力。

此为前言，言犹未尽，详见编写说明。

杨焕明

2016年9月9日，于中国深圳

编写说明

《基因组学》的编写是一次新的艰难尝试。

为使读者对我们的构思和本书的构架、体系及写作特点有更全面的了解，也便于对全书提出更有针对性的批评与建议，我们特参照享誉全球、久版不衰的《基因的分子生物学》（Watson 等著，中文第 7 版，科学出版社，2015）的相似方式，对本书的编写作较大篇幅的详细说明。

框架与构思

参照《基因的分子生物学》，本书将通常的目录拆分为简目与细目。简目仅有一页，可一目了然地通观全书的构架，了解我们对基因组学这一学科的科学体系与教学体系的构思。细目共有 14 页，接近《基因的分子生物学》的“内容详注”，可作为全书的复习提纲。

本书共分五部分。

第一部分——基因组学概论分为两篇。

第一篇《基因组学》导读不同于一般的导论。导论是一书的入门之“导”，而导读则是全书的阅读之“导”。作为入门教科书，本书的导读共有九节，从不同的角度，概括了基因组学这一学科的定义、理念、源流、发展趋势及应用范例，以及基因组学相关技术在生命科学和生物产业中的地位及其对人类社会的影响，特别是必须重视的相关生命伦理学 (bioethics) 原则和应该了解的中国的贡献。其内容并没有与本书的框架一一对应，而其中发展趋势和新近进展的部分，将是再版的主要更新内容。

第二篇基因组学的发展史主要描述 HGP 这一基因组学的第一次实践，这是史实。尽管“基因组”一词问世于 1920 年，“基因组学”命名于 1987 年，但真正意义的基因组学是随着 HGP 的讨论才开始的。正因为如此，本书以 1984 年美国 HGP 的第一次讨论会开始，特别以笔者自己亲身参与的体会，列举了这一计划的学术源流、目标变迁、技术路线改进、模式生物选取，以及人类基因组序列草图和精细图的完成，特别是中国的参与和贡献。

第二部分——基因组学的方法学也分为两篇，是本书最重要的部分。

这一部分的重要特点是开宗明义，将测序仪与其他的测序关键技术作了明确的切割。测序作为一项技术，测序仪只是其中的一部分，还包括诸多更为重要的关键技术。特别是对大学生和研究生来说，了解这些机器之外的关键技术，如上机前准备工作与所有相关技术，尤其是结合很多相关技术的基因组序列的组装和分析，要比按照说明书操作机器更为重要。

第一篇 DNA 测序分为两章，第一章（测序仪）介绍测序仪的历史性技术突破、市场化的仪器和基本原理，以及“下一代”测序仪的发展趋势。第二章（关键技术）除了介绍策略与方案、测序材料外，还特别介绍了关键技术的发展方向。

第二篇序列的组装和分析的前两章（序列的组装、基因组概貌分析）内容包括下机序列 (reads) 的处理与质控、序列的组装，以及基因组概貌和变异等初步分析。本书主要列举分析的实例，以及部分开源软件工具的使用方法，并没有对所有的生物信息学技术作详尽的介绍。而第三章（基因组比较分析）只是例举了演化基因组与 META 基因组，显然不尽人意。

第三部分 —— 基因组的生物学是本书的另一重要部分和特色，即在基因组序列和基因组学研究的基础上，生命科学揭开的新篇章。

这一部分主要是以物种为基础，着力体现一个物种的基因组序列所带来的该物种生物学各方面研究的革命性变化。物种是所有生物学研究的起点，也是基因组学及其他组学研究的起点。本书试图从一直跟踪的已测序的诸多物种中选取一些有代表性的，如广为熟知的模式生物和具有重要科学和生态意义及经济价值的物种，来介绍基因组序列基础上的生物学新发现，也建议读者作进一步的选择。对照浩如烟海的新文献、日新月异的新知识，这部分内容也不尽人意，反映了笔者知识的局限性，这都将在再版时反思、再议、再补、再改。而所有跨物种、跨组学的分析，特别重要的是生物学的灵魂——演化，暂且穿插在各物种基因组的分析之中，也有待补充并将其在这一知识体系中更好定位。

第一篇人类基因组中的第三章（临床基因组学）就是这一构思的尝试。本章开门见山便是基于序列数据的检测技术。这是基因组学及测序技术给医学临床带来的重大影响。在将来的再版中还会作更多的调整与充实。本章中的法医基因组，则因其特殊性而单独阐述，而外饰基因组学和 META 基因组学等新兴学科，由于已写入技术部分，在这里没有占很大篇幅。

第二篇动物基因组将按动物界的系统分类选择脊椎动物门和无脊椎动物门以及为人熟知的其他门类的典型动物作简要介绍。

第三篇植物基因组将植物人为地分为模式植物、主粮类、蔬菜类、瓜果类、经济类、花卉类和单细胞植物类，并选择典型的物种作简要的介绍。

第四篇微生物基因组则分为真菌、细菌、病毒和环境微生物群，并选择代表物种简要介绍。物种甚多，描述甚简，仅供选择使用。

第四部分 —— 基因组的设计和合成之所以没有使用常见的合成生物学一词，是由于在现阶段尚未真正实现从头（from scratch）合成一个新的生物体的合成生物学。而“合成”一词易被误解成仅仅是化学与工程的范畴。为了突出全基因组“设计”在这一领域的核心地位，故使用了“设计和合成”这两个术语的结合，而所有其他方面则概括成“基因组组成”。合理与接受程度如何，有待实践的检验及笔者和读者的共同努力，并将在再版时，作较大的修改和补充，特别是扩大基因组编辑的新近进展。

第五部分 —— 基因组伦理学是基于笔者十几年的工作经历和体会，深感理解生命伦理学双重使命——为生命科学和生物技术“正名顺言”（不发展生命科学和生物技术是最“不伦理”的）、“鸣锣开道”（生命伦理学要呼唤科学技术的创新）和“保驾护航”（生命伦理学要保证生命科学和生物技术的正确方向）的极端重要性。主要内容是笔者长期担任联合国教科文组织（The United Nations Educational, Scientific and Cultural Organization, UNESCO）的国际生命伦理委员会（International Bioethics Committee, IBC）与政府间生物伦理委员会（International Governmental Bioethics Committee, IGBC），以及联合国人权委员会（Commission on Human Rights, UN）等国际组织伦理小组成员的所得和体会。

特色与不足

本书的特色也正是本书的不足和风险。

1) 重理念而不只是讲授知识。还记得巴德年先生有一次对我们青年教师说：“三流教授讲书本，二流教授讲知识，一流教授讲理念”（尽管本书绝非一流）。在人手一机（手机或计算机）的信息时代，所有知识都是“现成的”，几乎都在网上可以找到，很多人都有“无师自通”的能力，这并不影响一本好的教科书的存在意义，更加突出了“理念”的重要性。集笔者十几年之经历，学习基因组学——掌握理念尤为重要。有理念才能矢志不移，才能掌握知识，才能举一反三，游刃有余。本书所有知识部分的介绍，都是为了证明并分享基因组学的理念；也是为了激发读者的兴趣，挖掘读者的知识潜能，培养读者的创

新意识；同时也为正在构思的网络版留下更多的空间。正因为此，即使在“基因组的生物学”这样的重要部分中，我们也是惜墨如金，点到为止。

作为一门新学科，基因组学博大精深，涉及生命科学的方方面面。本书主要是供刚学完普通遗传学的本科生及基因组学相关专业的研究生作为基因组学的综合性入门教科书。其主要目的是在掌握理念的基础上，了解基因组学的基本要点。正因为如此，本书的正文仅有易读、好记的要点（约占全书篇幅的五分之一），其他拓展性的知识约占五分之四的篇幅；而为了顾及正文内容叙述的完整性与连贯性，全部以小一号的字体呈现。当然，本书的风险或许在此。笔者的初衷是宁失为“过浅”，也不愿“过深”而有悖于入门教科书。

2) 讲源流而不重复相关基础。基因组学的基础几乎涉及生命科学所有学科，特别是遗传学和分子生物学。正因为这样，建立基因组学的教学体系至关重要。而综观许多国内外的教科书或专著，尽管以基因组学为题，但究其内容，一半以上的篇幅实际是地地道道的遗传学和分子生物学。本书忍痛割爱，作了风险最大的尝试——将它们基本删除或作“入框”处理，以突出基因组学的源流和精髓。而“框”的设计和撰文，则侧重历史性、相关性和趣味性，并不要求读者全部仔细阅读，仅供使用时选择。据笔者四十余年各类不同的教学经验，本书特别重视专用名词或术语的准确定义及其内涵和外延，分析这一术语与另一相近术语的异同，而不泛泛而论使学生不得要领。正是出于这样的考虑，本书特别重视“术语学 (terminology)”与基本概念。在书后另附“常用术语与缩写简释”，以补正文中说明之不足。也同样待议、待改及完善。

3) 论原理而不详谈技术细节。基因组学也是一门理论与实验相结合的学科。除了其核心技术——测序技术以外，还涉及很多其他生物和理化技术。本书重在介绍这些技术的基本原理，而不是描述其操作细节。更多的“入框”处理也体现“授人以鱼不如授人以渔”的理念。这里的“渔”，不只是渔具与鱼饵，而是有关“渔”的理念、战略（包括环境变化、天时气候、地域水势，各种“渔”的手段，如垂钓、撒网、拖网、“竭泽而渔”，以至这些手段的选择、改进与创新）。正是出于这样的考虑，本书用较大的篇幅介绍常用技术包括生物信息学分析技术的基本原理和操作要点，特别是在第二部分之中。

4) 倡甚解而不推荐过多文献。这或许是本书的另一风险。作为教科书，理应鼓励学生广泛阅读原始文献。但作为过来之人，深知学生们所修科目甚多、学习任务繁重。列举过多的参考文献，要求学生阅读并真正精读并不实际，也会流于形式。真正精选必读文献也绝非易事。因此本书期望读者把主要精力放在对本书正文的甚解上，而不在正文中直接关联参考文献。但为了一定程度的平衡，将所有必要的参考文献，特别是历史性的经典文献，加以汉语主题词的提示，分篇分章列于全书之后，并不要求学生必读。

5) 抒己见而又博采百家。这是笔者的又一纠结。教科书与在课堂上授课一样，一定有也应该有自己的学术观点，这又一定涉及国内外历史上的一些敏感而现在已无必要的学术争议。既使一些常用且颇多争议的中译名，例如“META 基因组”在本书中暂且直接使用英文缩写。epigenomics 则摒弃易于误导的“表观基因组学”，而采用“外饰基因组学”。本书在博采百家的同时，力求培养学生的分析能力，形成有自己独立见解的能力。

6) 论贡献而不分中外老幼。本书首开先河，大量引用了中国科研团队的工作，特别是配用了很多年轻中国人的面孔。乍一看本书，一些读者或颇有微词，但如果仔细阅读有关章节对他们工作的描述，一定会觉得非常自然——中外并蓄、“老幼无欺”。当然，所有的引用仅为例举，而不是全面、平衡的评估，“文责自负”。本书使用的照片数量，还远远比不上《基因的分子生物学》。我们明白这一尝试的风险，乐意在再版时更换、增补。

此外，本书对一些教材编写的具体问题也作了小试。对于诸多英语人名，第一次出现时皆为英文全名，而后仅用英文姓氏。地名、机构名称也作相似处理。对于常用的中译专业名词，仅在第一次出现时附上英文全名和中译名。而常用的 HGP 等缩写词，则在第一次出现时缩写词在先，然后在括号中用英文及中文全称，再次出现时就使用英文缩写词。

仿照《基因的分子生物学》已受广大读者欢迎的编排，我们也在每页的外侧（双页左侧，单页右侧）留出空白。以供学生在听课时作简略的笔记，复习时作每页的归纳或眉批。

尽管笔者和参与编写本书的全体人员已尽了很大努力，力求完美，宁晚勿滥，而必须承认的事实仍是眼高手低。作为首版，本书的各类问题比比皆是，每次再阅，不禁汗颜。请诸位读者、老师和同学严格要求，多予指正。我们仍将征求对本书勘误的修改和评论意见，既包括对本书的框架设计及有关章节中要点的描述，也包括选字、遣词、造句、插图等的具体修改意见，以及概念的更新、补充和对新进展的及时反映。正出于这一目的，本书曾在2015年与2016年先后两次先出“征求意见稿”仅供内部交流和师友试用，以期再求指正、抛砖引玉。本版正是在整合诸位师长的勘误与雅正的基础上，也根据笔者自己试用的体会，作了较大程度的修改。

有关本书的图、表，若有涉及知识产权等问题，请与出版社联系，我们定将在再版时作相应的删改。

杨焕明

2016年9月9日

简 目

基因组学

第一部分 基因组学概论

- 第一篇 《基因组学》导读
- 第二篇 基因组学的发展史

第二部分 基因组学的方法学

- 第一篇 DNA 测序
- 第二篇 序列的组装和分析

第三部分 基因组的生物学

- 第一篇 人类基因组
- 第二篇 动物基因组
- 第三篇 植物基因组
- 第四篇 微生物基因组

第四部分 基因组的设计和合成

第五部分 基因组伦理学

GENOMICS

Part I. An Introduction to Genomics

- Chapter 1. A Guide to *GENOMICS*
- Chapter 2. History of Genomics

Part II. The Methodology of Genomics

- Chapter 1. DNA Sequencing
- Chapter 2. Sequence Assembly and Analysis

Part III. The Biology of Genomes

- Chapter 1. Human Genomes
- Chapter 2. Other Animal Genomes
- Chapter 3. Plant Genomes
- Chapter 4. Microbial Genomes

Part IV. Redesign and Synthesis of Genomes

Part V. Bioethics in Genomics

目 录

第一部分 基因组学概论

第一篇 《基因组学》导读	3
第一节 基因组学的一般定义	3
一、基因组学和基因组	3
二、基因组学的分科及与其他学科的关系	5
第二节 基因组学的两个理念	5
一、生命是序列的	5
二、生命是数字的	6
第三节 基因组学的三个源流	7
一、演化学说	7
二、细胞学说	8
三、基因学说	10
第四节 测序技术的四个突破	14
一、直读	14
二、自动化	14
三、规模化	14
四、MPH	15
第五节 基因组学发展的五大趋势	15
一、重绘“生命之树”	15
二、群体基因组分析	16
三、个体基因组分析	16
四、“跨组学”分析	17
五、基因组的生物学	17
第六节 基因组学应用的六个方面	17
一、外显子和全外显子组测序——单基因性状与遗传病	17
二、全基因组测序——复杂性状与常见疾病	18
三、单细胞测序——基因组异质性	18
四、META 基因组测序——微生物及病原基因组	18
五、微(痕)量 DNA 测序——无创检测、法医和古 DNA 研究	19
六、“数据化”育种与生物条码	19
第七节 正在改变世界的七项技术	19
一、基因组编写	19
二、基因组编辑	20
三、干细胞与 iPS	22
四、动物克隆	23

五、大数据与生物库	24
六、表型组分析	24
七、“组学”与相关技术	25
第八节 基因组伦理学的八个方面 (HELPCCESS)	25
一、人文 (H, Humanity)	25
二、伦理 (E, Ethics)	25
三、法律 (L, Law)	25
四、公众关系和决策 (P, Public-relationship/Policy-making)	26
五、文化 (C, Culture)	26
六、经济和教育 (E, Economy/Education)	26
七、生物安全和防护 (S, Safety/Security)	26
八、社会 (S, Society)	26
第九节 中国对基因组学的九大贡献	27
一、1% 国际人类基因组计划 (HGP)	27
二、10% 国际单体型图计划 (HapMap 计划)	28
三、国际千人基因组计划 (G1K 计划)	28
四、国际癌症基因组计划 (ICGP)	29
五、水稻和家蚕基因组计划	30
六、META 基因组	31
七、“生命之树”与动植物基因组	31
八、第二代酵母全基因组设计和合成	31
九、倡导“HGP 精神”和“合作”的文化	32
第二篇 基因组学的发展史	33
第一节 HGP 的起源	33
一、HGP 的讨论和启动	33
二、HGP 的国际化	35
三、HGP 的学术源流	40
第二节 HGP 的目标和技术路线	42
一、HGP 的技术目标	42
(一) 遗传图	43
(二) 物理图	43
(三) 转录图	45
(四) 序列图	45
二、HGP 的技术路线	45
第三节 HGP 中的模式生物	46
一、大肠杆菌	47
二、酿酒酵母	47
三、秀丽线虫	48
四、黑腹果蝇	49
五、拟南芥	50
六、河豚鱼	51
七、小鼠	51
第四节 HGP 的完成与后续计划	52



一、HGP 的完成	52
二、HGP 的后续计划	54
(一) 国际 HapMap 计划	55
(二) ENCODE 计划	56
(三) 国际千人基因组计划 (G1K 计划)	56
(四) 国际癌症基因组计划 (ICGP)	57
(五) 百万人基因组计划与一亿人基因组建议 (G1M 与 G 100M) ...	58
(六) 其他国际合作计划	58
第五节 HGP 的意义和影响	61
一、创造了一种新的文化: 合作	61
二、催生了一门新的学科: 组学	61
三、提供了一个新的技术: 测序	61

第二部分 基因组学的方法学

第一篇 DNA 测序	67
第一章 测序仪	67
第一节 测序技术的四个突破	68
一、第一个突破: 直读	68
(一) 前直读法	68
(二) 直读法	69
1. SBC 法	70
2. SBS 法	71
二、第二个突破: 自动化	74
三、第三个突破: 规模化	77
四、第四个突破: MPH 测序	80
(一) 焦磷酸测序	81
1. 454	81
2. Ion Torrent/Proton	83
(二) 新一代 SBS	85
1. Illumina	85
2. BGISEQ-500	88
3. Max-Seq	90
4. LaserGen	91
(三) 连接测序	92
第二节 下一代测序仪及其基本原理	94
一、单分子测序	94
(一) HeliScope	94
(二) PacBio	94
二、纳米孔测序	97
(一) 物理纳米孔	98
(二) 生物纳米孔	100
三、杂交测序	101

四、显微测序	102
第二章 关键技术与方法	104
第一节 Southern 测序	105
一、测序策略	105
(一) 点测序	105
1. PCR 扩增	106
2. 质粒克隆	108
(二) 靶区域测序	109
(三) 外显子组测序	111
(四) 全基因组测序	112
(五) 单细胞测序	113
(六) META 测序	114
(七) Indexing	114
二、基因组概貌评估	115
(一) 评估参数	115
1. 基因组大小	115
2. 测序深度	116
(二) 评估方法	116
1. 基因组速览	116
2. K-mer 分析	116
三、测序的技术路线	120
(一) 定位克隆霰弹法	120
(二) 全基因组霰弹法	121
四、测序材料	121
(一) 常用材料	122
1. 细胞培养物和外周血细胞	122
2. 其他组织细胞	122
(二) 特殊材料	122
1. 单细胞或痕量 DNA	122
2. 石蜡包埋材料	123
3. 古 DNA 样本	123
4. META 样本	124
(三) DNA 的制备、纯化和质检	124
1. 盐析法: 血液样本	125
2. CTAB 法: 植物组织、真菌类样本	125
五、测序文库	125
第二节 Northern 测序	126
一、测序材料及 RNA 模板制备	126
二、RNA 的测序和分析	129
(一) 数字化表达谱	129
(二) 全转录组	130
1. 已有基因组参考序列的转录组分析	131
2. 无基因组参考序列的转录组分析	132

(三) 单细胞的转录组或表达谱分析	132
第三节 Western 测序	132
一、染色质免疫沉淀测序	133
二、核糖体保护	134
第四节 Eastern 测序	134
一、化学法	135
二、亲和法	136
三、酶切法	137
第五节 其他重要技术	138
一、GWAS	138
二、基因组编辑	140
三、芯片技术	142
(一) 基因表达芯片	142
(二) SNP 芯片	142
(三) CGH 芯片	142
(四) 突变检测芯片	143
(五) 捕获芯片	143
(六) 测序芯片	143
第二篇 序列的组装和分析	143
第一章 序列的组装	143
第一节 测序数据的质控和预处理	143
一、碱基读取	143
二、载体和接头序列的去除	146
第二节 序列的组装	146
一、基于参考序列的组装	147
二、 <i>de novo</i> 组装	148
(一) OLC 法	148
(二) <i>de Bruijn</i> 图法	150
(三) “穷举”法	152
三、多策略结合组装	152
(一) 复杂基因组	152
(二) 基因组精细图	153
四、组装的质控	154
第二章 基因组概貌分析	156
第一节 基因组大小	156
第二节 GC 含量	157
一、GC 含量和染色体带型	158
二、GC 含量和基因密度	160
三、GC 含量和 CpG 岛	161
第三节 重复序列	161
一、重复序列的分类	161
二、重复序列的识别	162

第四节	编码基因	164
一、	编码基因的注释	164
(一)	<i>de novo</i> 预测	164
(二)	同源比对预测	166
二、	基因的功能注释	168
(一)	InterProScan	168
(二)	GO	168
(三)	COG	169
(四)	KEGG	169
第五节	假基因	170
第六节	非编码序列	171
一、	rRNA	171
二、	tRNA	172
三、	miRNA	173
(一)	miRNA 比对预测	174
(二)	miRNA 结构预测	175
(三)	miRNA 靶基因注释	176
第七节	基因组变异	177
一、	单核苷酸多态性	177
二、	拷贝数变异	180
三、	其他结构变异	181
(一)	倒位	181
(二)	易位	181
四、	个体杂合度	182
第三章	基因组比较分析	183
第一节	演化基因组	183
一、	基因家族	183
二、	系统发生	186
三、	同源与直系同源基因	190
四、	基因的选择和演化	191
五、	多基因家族	193
六、	转位因子	195
七、	保守性区域及共线性	196
八、	全基因组加倍	198
第二节	META 基因组	202
一、	16S rDNA	203
二、	环境基因标签和 K-mer	204
三、	参考基因组集和共有核心细菌	205
四、	核心基因家族	207
五、	整合基因集构建	207
六、	META 微生物组群与疾病相关性	208
七、	MWAS	209

第三部分 基因组的生物学

第一篇 人类基因组	214
第一章 人类基因组概貌	214
一、人类基因组大小	216
二、编码基因	218
(一) 编码基因的估计数目	218
(二) 基因序列、外显子和编码序列	219
(三) 编码基因的分布和密度	220
三、假基因	222
四、GC 含量	222
五、重复序列	222
六、CpG 岛和甲基化	223
七、非编码 RNA	223
(一) rRNA	223
(二) tRNA	224
(三) miRNA	224
八、线粒体 DNA 基因组	224
第二章 人类基因组的生物学	226
一、人类基因组多样性与泛基因组	226
二、演化与群体基因组	227
第三章 临床基因组学	228
第一节 基于序列数据的检测技术	229
一、无创产前检测	229
二、植入前检测	233
三、症状前检测	235
四、精准医学和 21 世纪的医学	237
第二节 单基因病及基因定位	240
一、单基因病	240
二、基因定位	241
(一) 遗传分析定位	241
(二) 细胞遗传学定位	242
(三) 原位杂交	243
(四) 外显子组测序	244
第三节 癌症基因组	244
一、癌症与基因组学	244
二、癌症的组学研究	246
(一) 全基因组 MPH 测序与癌症基因组分析	246
1. 点突变与 InDel	247
2. 结构变异	247
3. 基因扩增	247
(二) 单细胞组学技术与癌症基因组异质性	248