

Discovering
探索

GENOMICS,
基因组学
PROTEOMICS,
蛋白质组学和
& BIOINFORMATICS
生物信息学

[美] A. 马尔科姆·坎贝尔 著
劳里 J. 海尔 著
孙之荣 主译

A T G C T C A T C G G C A T T G



科学出版社
www.sciencep.com

探索基因组学、蛋白质组学 和生物信息学

[美]A. 马尔科姆·坎贝尔 劳里J. 海尔 著

孙之荣 主译

科学出版社

北京

图字:01-2003-6984号

内 容 简 介

本书围绕对基因组学、蛋白质组学和生物信息学的概括描述,向读者介绍了近几年来生物学以及医学生物学研究方法上的发展,以及这些方法在研究思维上的深远影响。作者以通俗易懂的语言将这三个方面组成一个研究、探索问题的互动平台,使读者对生物信息学有一个系统的认识和了解,利于更深入的研究。内容主要包括基因组序列、基因组变异、基因组表达、DNA芯片、蛋白质组学、全基因组学以及基因组学在医学病例中的应用等,同时穿插问题探讨、数学备忘录等使内容更加丰富,并附有图片光盘利于读者参考。

本书可作生物学专业本科生、研究生的生物信息学教材或教学辅导书,亦可供分子生物学、生物化学、细胞生物学以及医学、药学等领域的科研人员参考。

Simplified Chinese edition copyright 2004 by PEARSON EDUCATION NORTH ASIA LIMITED and SCIENCE PRESS.

Original English language title: Discovering Genomics, Proteomics, and Bioinformatics, by A. Malcolm Campbell and Laurie J. Heyer, Copyright 2003

All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Benjamin Cummings.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Regions of Hong Kong and Macao).

本书封面贴有 Pearson Education 出版集团激光防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

探索基因组学、蛋白质组学和生物信息学/[美]马尔科姆等著;孙之荣主译. —北京:科学出版社,2004.7

ISBN 7-03-012921-0

I. 探… II. ①马…②劳… III. ①基因组-研究②蛋白质-研究③生物信息学-研究 IV. ①Q343.2②Q51③Q811.4

中国版本图书馆CIP数据核字(2004)第010993号

责任编辑:庞在堂 彭克里 孙晓洁/责任校对:张 琪

责任印制:安春生/封面设计:王 浩

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

源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2004年7月第一版 开本:787×1092 1/16

2004年7月第一次印刷 印张:29 3/4

印数:1—3 000 字数:673 000

定价:58.00元(含光盘)

(如有印装质量问题,我社负责调换〈新欣〉)

译者序

21 世纪的生物学研究对每一个生物学工作者来说既是机遇又是挑战。随着人类基因组测序工作的初步完成, 基因组生物学的研究由测序基因组向功能基因组学转移。作为功能基因组学的重要研究领域之一的蛋白质组学, 与基因组学一起, 成为当今生命科学的热点和前沿。基因组学、蛋白质组学这一新兴的研究领域孕育了生物信息学这一新兴交叉学科的产生和发展。基因组学、蛋白质组学和生物信息学密不可分。这本书的作者将基因组学、蛋白质组学和生物信息学放在一起加以介绍, 别具特色。

《探索基因组学、蛋白质组学和生物信息学》围绕基因组学、蛋白质组学和生物信息学, 详细地展示了近几年来生物学及医学生物学研究方法上的革命, 以及它们在研究思维方式上的深远影响。

《探索基因组学、蛋白质组学和生物信息学》一书涵盖了现代生物学的大部分研究领域。作者将本书设计成一个探索基因组学和蛋白质组学问题的互动平台, 能够帮助读者学会假设、推理、分析、建立模型等研究现代生物学的特定模式的能力。所以这本书无论是对教学人员, 还是对科研工作者都是一本适用的参考书。

本书的翻译工作是清华大学生物科学与技术系生物信息研究所的师生集体完成的。参加本书翻译工作的有孙之荣教授、柳树群博士, 以及郑家顺、季星来、夏雪峰、资治科、李薇、陈虎、张扬和沈金诚等。全书由孙之荣教授负责组织和审校工作。希望本书的翻译和出版将对我国生物信息学以及基因组生物学、蛋白质组生物学的发展具有良好的促进作用。

孙之荣

清华大学生物科学与技术系

序

一个研究生命体及其组成的新方法和新技术的问世常常会导致生物学新领域的产生。显微镜和细胞切片技术的发展使细胞生物学兴起，X射线晶体衍射和后来的核磁共振技术孕育了三维图像的结构生物学，重组DNA技术使分子生物学领域的许多技术获益。当这些新学科成熟后，它们的方法将深入影响人们对生物学主要问题的思考方式，这些分支将保持它们固有的某些特点并持续发展。

人类基因组计划吸引了科学家和公众的共同关注。2000年6月26日，比尔·克林顿总统在与美国人类基因组研究所和赛莱拉公司（Cerela Genomics）的首脑会晤时宣布：“我们在这里庆祝人类基因组草图的首次测绘完成，毫无疑问，这张草图是人类有史以来最重要、最令人叹为观止的图谱。”对于生物学家来说，人类基因组草图的首次测绘意味着我们已经得到了几十亿对碱基DNA序列。在这里，科学家面对的不仅仅是人类基因组，还有许多模式生物，比如酵母、大肠杆菌、线虫、果蝇和拟南芥的基因组。如何将这些序列综合起来诠释DNA序列，包括从DNA转录得到的RNA，以及从RNA翻译而来的蛋白质，这迫切需要发展新的生物技术和研究方法。

《探索基因组学、蛋白质组学和生物信息学》详细地向我们展示了近几年来研究方法论上的革命，以及它们在生物学研究思维方式上的深远影响。本书对诸如DNA测序、多态性检测、基因阵列、基因敲除以及蛋白质相互作用检测等生物技术进行了深入的阐述。这些技术在特定生物学问题上的应用，为解决生物学上从发育到人类疾病等重要问题提供了参考模型。对大规模数据的积累如何导致基因和蛋白质网络模型的建立，以及细胞行为模型产生的复杂过程，本书也做了详尽的描述。每一章的一系列“问题探讨”引导读者去思考有关生物学的科学、伦理、法律和社会问题。

基因组学和蛋白质组学与生物信息学密不可分，大规模DNA测序以及RNA和蛋白质组研究计划产生的海量数据必须依靠生物信息学工具来处理和分析。本书描述了若干数据库和算法，表明了基因组学、蛋白质组学与生物信息学的密切关系。书中的“数学备忘录”部分就算法的数学原理提供了严格意义上的说明。

这些新方法是否已经导致了新的思维方式？从某种意义上讲，答案是肯定的，因为从基因组和蛋白质组数据中推测出了一些我们以前并不清楚的结论。这里，可能会有人争辩，生物学家一直在研究“系统生物学”，改变的只是研究工具而不是思维方式。以前一次只能检测一个转录子的时候，可以在不同的条件下分析RNA以测定单个基因的功能，现在应用DNA芯片，科学家可以同时检测整个生物体的所有转录子。本书将引导读者进入一个新的领域，在这里，你将会了解到，这些实验和计算工具组成的只是“单纯的生物学”，而不是“系统生物学”的全部和精髓。

Stan Fields
华盛顿大学

前 言

基因组学这个名词来自于基因组，基因组指的是生物体内所有 DNA 的总和，基因组学是研究基因组的学科。当基因组和基因组学的概念流行起来后，大量以 omes 和 omics 结尾的名词开始在出版物中频频出现。《探索基因组学、蛋白质组学和生物信息学》不仅仅是一本 omes 的书，因为该领域已经不局限于基因组学的狭隘定义，正如本书所体现的，基因组学包括细胞内各种分子（比如 DNA、蛋白质、脂质和糖等）的相互作用。本着探索的精神，我们将在本书中探讨一系列引发生物学研究方式变革的工具和问题。

《探索基因组学、蛋白质组学和生物信息学》采取了许多教师成功运用过的两条教学原则：教授感兴趣的问题并在求知欲的基础上引出新的知识。这其实是人们获取新信息的途径，也是引导学生学习的最好方法，这样他们才会有求知的欲望并牢记所学的知识。《探索基因组学、蛋白质组学和生物信息学》从文献中提取出研究例子，在解答这些研究案例所遇到的各种问题的同时，探索科学的内容。本书涵盖了现代生物学的大部分研究领域如基因组学（全基因组和基因组变异）、蛋白质组学和生物信息学。

作者将本书设计成一个探索基因组学和蛋白质组学问题的互动平台。插图包含了许多数据，读者可以自行挖掘其中更多的内在信息；“在线数据库”鼓励读者像真正做研究一样使用数据库进行实时探索；“问题探讨”使读者专注于关键信息，并促使读者应用文章和图片中的工具和信息去思考。一般的教科书提供的是应付测试所需记忆的事实和细节，而基因组学要求读者学会分析、假设、推理、建立模型；本书正是为了让读者具备这种能力而设计。

21 世纪的生物学研究对每一个生物学工作者来说既是机遇又是挑战。本书的特色在于：读者必须反复使用计算机去获取最新信息。为了让读者深入理解基因组学、蛋白质组学和生物信息学，本书列举了大量的真实而有说服力的研究事例，激发读者的学习兴趣。

写作风格

文体通俗易懂，虽然引入了一些新名词，但尽量避免了不必要的术语。

问题探讨

通过“问题探讨”来引导读者思考，问题探讨不是在每章末尾才提出，而是放在每个研究案例之中。问题探讨让读者的注意力集中在实验设计、数据解释、数据对观点的支持等关键概念上，分析从文献中提取出来的数据，让读者得出自己的结论，而作者只是引导读者去寻找相应数据。为了能够回答其中的一些问题，读者必须自己使用在线数据库，这些数据库是经常更新的。

所有的问题探讨都可以从与本书配套的网站上得到，便于读者与在线资源互动，读

者也可以通过电子邮件把答案送给自己的导师批阅。

问题探讨

- 16. 打开 *S. cerevisiae* Genomic View 网页, 寻找本研究测定的 NORF 位置(第 4 号染色体基部 1 236 754 区域), 寻找与注释基因重叠的插入位点。
- 17. 打开 SGD Gene/Sequence Resource 页面, 找到第 4 号染色体上 1 236 454~1 237 054 位置, 点击“6-Fram translation”。mTn 插入发生在这 600 个碱基的约第 300 位置。那么该位置可能编码多少个蛋白质呢?

相关方法

Quick PDB

结构

环加氧酶

序列

未鉴定蛋白

相关链接

保守域

PDB

PREDATOR

工具菜单

工具菜单贯穿每一章, 提醒读者使用于本书配套的网站资源。在该网站上读者可以阅读更多更深入的方法描述, 获取序列信息, 观看三维结构图, 链接到相关网站。这些工具帮助你更好地参与到理解基因组学核心内容的互动过程中来。

数学备忘录

许多细胞/分子领域的生物学家在他们的工作中用到的数学不多, 但是基因组学、蛋白质组学和生物信息学正在改变这个事实, 这些领域都非常依赖数学。为了让读者更好地理解数据分析的过程和数学在生物学研究中扮演的角色, 我们设立了数学备忘录, 以便读者了解生物和数学之间的联系。数学备忘录(如下例)用研究案例作为统计分析、概率和计算方法的简明教程。

数学备忘录 6.2 Sup35 是蛋白质相互作用网络的中心蛋白质吗?

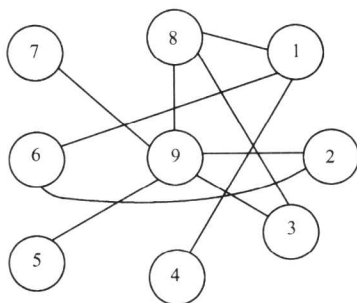
图 6.16 中, Sup35 看起来位于相互作用蛋白质的网络中心。但是, 在这样一个复杂的网络中, 观察者看到的很可能是假象。因此, Sup35 与网络中的其他蛋白质相比是否能与绝大多数蛋白质发生相互作用, 还需要进一步的量化处理。

数学家将图 6.16 的相互作用蛋白网络叫做图。在图论里, 节点间的连线被称为弧或边。定向图的边上具有箭头, 以指明两个节点间的信息流动方向。在图 6.16 中, 节点的连接和关系是相当重要的, 但是, 因为边上没有箭头, 所以并没有给出方向信息。与节点接触的边的数目叫做节点的度。用图论的术语讲, 目前的问题在于节点 Sup35 是否具有比图中“普通”节点大得多的度。如果这样, 研究者便能确定 Sup35 的确在整个蛋白质相互作用网络中发挥中心作用。

相关链接

DIP 数据库

解决图 6.16 问题的方法是建立一个图概率模型, 称之为随机图。在这个模型里, 使用一定的概率画出每对节点之间的边, 每个边独立于其他边。高级概率理论以及图论提供了精确测定超过一定数量的大图中最大度的概率方法。但是, 我们有一个简单的假设, 即度数近似于正态分布(见数学备忘录 8.1)因此, 我们能估计出 Sup35 节点的度是否显著大于度数的平均期望值和标准误差。由于图 6.16 作为例子讲解过于复杂, 我们来看一下下面这个小网络图:



在该图中,节点4、5、7的度为1;节点2、3、6的度为2;节点1和8的度是3;节点9的度是5。平均度则是这9个度数的平均值(2.22),9个度数的标准偏差是1.23,用它来衡量这9个节点的度是否正常。特别地,如果节点9的度大于平均度的程度为标准偏差的两倍,那么便可认为节点9具有非正常的大度。由于 $5 > 4.68 = 2.22 + 2 \times 1.23$,因此我们说节点9具有非正常的大度。同样的方法可以应用于图6.16,以量化方法测定 Sup35 在蛋白质网络图中是否发挥中心作用。

插 图

丰富详细的插图再现并扩展了正文提供的原始信息,如下例所示。通过分析插图,读者可以增进理解并获取更多信息。

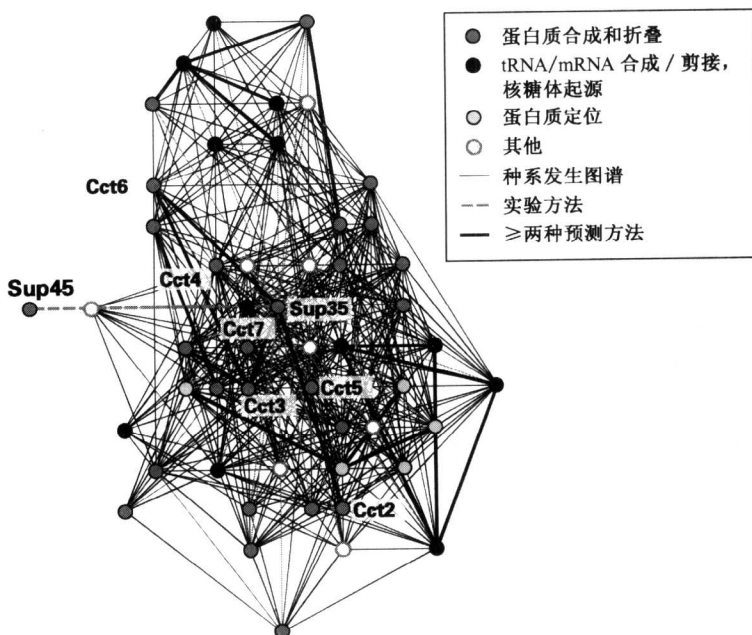


图 6.16 高置信度(粗线)和低置信度(细线)蛋白质相互作用网络
该网络以酵母朊病毒蛋白 Sup35 为中心,网络中有许多蛋白质与蛋白质合成相关,具体包括与核糖体、蛋白质折叠、分拣、修饰、定位有关的蛋白质。在该模型中,节点表示蛋白质,边的粗细则与蛋白质相互作用强度呈正相关。

光盘上的图片

书中所有图片都能在随书光盘中找到,有些图片因为是彩图,在书中可能显示不清晰,可以在光盘中找到这些彩图。本书对这些图片做了分类,以用于课堂讲解和讨论。

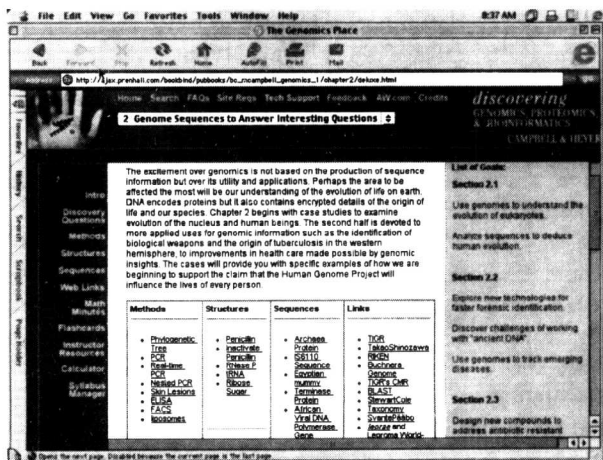
从遗传学到基因组学

人们通常认为基因组是高容量的遗传学,但是一个基因组比起它的各部分之和更为复杂,我们必须使用新方法和新思路进行研究。最后一篇(第四篇)包含了3章的内容,帮助读者从遗传学到基因组学过渡,指出了以前“一个基因、一个蛋白、一个表型”这种看法的错误。研究案例包含了真实的数据,以便于读者理解和发现“细胞网络”的相互连接。这种探索方法将培养读者科学的分析数据、建立模型的能力。

工 具

《探索基因组学、蛋白质组学和生物信息学》的配套网站是一个帮助学习基因组学的良好工具。

- 方法:这些网页解释了分子和基因组学方法是如何运用并产生了什么样的数据,如果读者以前没有学过相关知识,它们将提供相关的背景知识作为补充。
- 序列:在多个问题探讨中,读者要运用在线的生物信息学工具分析蛋白质或DNA序列,为了节省读者的时间并避免排版错误,所有序列都在网页上提供,读者可以复制粘贴到分析工具中。
- 结构:三维结构对每个蛋白质来说都是非常重要的。其中带有有声教程的网页有助于读者更好地理解蛋白质的结构特征。
- 相关链接:每章都有两种链接。第一种链接直接连到在线数据库和生物信息学分析工具的相关网址,例如,NCBI和PDB;第二种链接连接到了相关实验室的主页上,当读者对某个案例的研究内容格外感兴趣时可以使用该种链接。



致辅导员

电子版的《教师手册》(www.geneticsplace.com)提供彩图。例如,如果要求学生画一幅图表或者做一幅线路图,电子版将提供一个可能答案。这些图片加上光盘中的图片,可用于讲课时做幻灯投影。

致 谢

Davidson 学院的同事给了大力支持。Clark Ross, Bobby Vagt, Verna Case 的支持和领导是这本书成书的基础。许多系的同事给予了作者多种形式的支持, 他们是 Karen Bernd, Karen Hales, Dave Wessner, Barbara Lom, Pam Hay, Julio Ramirez, David Brown, Jeanne O'neill, Suzanne Churchill, Chris Paradise, Mur Muchane, Peggy Maiorano 和 Betty Hartsell。感谢 Waksman 基金、国家科学基金、Duke 基金和南部大学联盟提供资金支持。最重要的, 我要感谢我的 14 位勇敢的学生, 本书的第一稿就是他们写的, 他们是: Sean Burke, Amber Hartman, Ben Havard, Julie Hwang, Dennis Jones, Peter Lowry, Jennifer Madden, Emily Oldham, La Powell, Lisa Robinson, Elizabeth Sellars, Liz Shafer, J. D. Willson 和 Marisa Wilson。

本书第一稿成稿之时, 我正在西雅图度假, Mary Claire King、Maynard Olson 和 Leroy Hood 在这段时间给予了极大支持。华盛顿基因组中心的多位成员给了我技术上和精神上的支持。系统生物学研究所 (ISB) 的人们建立了一个极好的学习基因组学和蛋白质组学的环境。另外, Eric Davidson 在 ISB 年会上的演讲给了我很大启发。同时也感谢全世界这领域的研究者, 因为能够使我免费共享他们的工作成果并且受到其热忱的帮助, 能够成为这个知识充分共享的社会的一个成员令我感到非常欣慰。就个人而言, 我在西雅图的朋友和他们的家庭使我感到宾至如归, 特别是 Hope-Young 一家、Hill 一家、Ginger Armbrust 和 Susan Francis, Walker 一家, 以及在 Meadbrook Co-op 幼儿园的 20 个家庭。

每个项目都需要一大批人的参与与创新, 这本书从最初模糊的想法到最后成稿能如此迅速, 得益于我的朋友 Michele Sordi 的指导, 没有她就不会有如此硕果。本书的出版也归功于 Peggy Williams 的勤奋和公正。她的劳动对于这样一本新领域中的新书的问世至关重要。另外, 在 John Inglis 的领导下, 冷泉港实验室出版社的编辑们工作非常敬业。在 Benjamin Cummings, 许多人对这本书作出了重要的贡献, 包括 Steve McEntee 和 Mary Ann Tenorio, 他们的艺术才华使得书中的图片显得十分真实, 更方便用于教学。Benjamin Cummings 的 Jamie Sue Brooks、Larry Lazopoulos 和 Left Coast Group 成员的设计方案令人赞叹。感谢 Andrew Ogus 给本书设计外观, 感谢 Yvo Rezebios 为本书设计封面, Michael McArdle 作为出版助理, 很好地使每个人的工作团结合作起来。

最后, 最重要的支持来自于离家最近的人们。这本书是一位非常了解生物学的数学家和一位尽力了解数学的生物学家彻夜合作的结晶, 为此, 特别感谢 Laurie。感谢 Genome Consortium for Active Teaching (GCAT) 的成员们提供了我所需的大学支持体系。Susan、Paulina 和 Celeste 长途跋涉了两次, 给予我继续完成本书的毅力。我的大家庭也给了我极大支持, 没有他们, 我将还呆在工厂里制作地板瓷砖。

——A. Malcolm Campell

感谢数学系的同事们,尤其是系主任 Stephen Davis 和好朋友 Donna Molinek 给予了我无穷的支持和鼓励。我的学生们是源源不断的力量和灵感的源泉,我尤其受惠于计算生物学的学生们: Frank Chemotti, Amber Hartman, Soren Johnson, Jennifer Kawwass, Peter Leese, Rachel Patton McCord, Celia Mendiondo, Emily Oldham, Ma Powell, Talbot Presley, Stanley Prybe, Lang Robertson, Lisa Robinson 和 Megan Shafer。这个多学科交叉的团队制作了一级 Kyte Doolittle 和层次聚类的网页。基因组学和计算生物学的学生都测试了许多数学备忘录,尤其感谢对撰写数学备忘录 6.1 有帮助的那些问题的提出者。

我至今仍然能够从那些引导我深入该领域的老师和朋友们那里获得重要的灵感,他们是: Mike Waterman, Simon Tavaré, Gary Stormo 和 John Williamson。他们与我的朋友们和家人一起,让我能对这本书有所贡献。Malcolm, 一个愿意与新同事分享他的基因组学观点的人,让这一切成为现实。最后,我感谢我的丈夫 Bill,他是如此的耐心和如此的爱我,他给了我最大的支持。

——Laurie J. Heyer

目 录

第一篇 基因组序列

第 1 章 基因组序列的获取和分析	(2)
1.1 基因组学的定义	(2)
1.1.1 什么是基因组学?	(2)
1.1.2 全基因组是如何测序的?	(3)
数学备忘录 1.1 什么是 E-Value?	(6)
1.1.3 为什么数据库中有那么多不完整序列?	(8)
1.1.4 我们如何来理解所有这些碱基的含义呢?	(12)
1.1.5 我们能预测蛋白质的功能吗?	(14)
1.1.6 不同物种中基因的保守性	(16)
1.1.7 怎样才能知道哪个碱基构成了基因?	(19)
1.1.8 基因可以生产多少蛋白质?	(21)
本节小结	(23)
1.2 我们从人类基因组测序草图中学到了什么?	(23)
1.2.1 人类基因组第一草图概况	(23)
1.2.2 人类基因组第一草图的初步结论	(23)
数学备忘录 1.2 你怎样用一条直线来拟合数据?	(26)
1.2.3 我们能否描述一个典型的人类基因?	(27)
1.2.4 什么时候数据会足够多?	(32)
1.2.5 基因组能否在不改变基因序列的情况下调节基因的表达?	(34)
本节小结	(38)
本章总结	(38)
参考文献	(38)
第 2 章 基因组序列解答有趣的问题	(40)
2.1 基因组的进化	(40)
2.1.1 真核生物是如何起源的?	(40)
数学备忘录 2.1 引用计数的差别是否显著?	(43)
2.1.2 人类的起源	(53)

数学备忘录 2.2 怎么知道进化树是正确的?	(5 6)
本节小结	(5 9)
2.2 基因组鉴定	(5 9)
2.2.1 我们怎样检查生物武器?	(5 9)
2.2.2 DNA 能存活多久?	(6 3)
2.2.3 肺结核是怎样进入北美的?	(6 5)
2.2.4 怎样鉴定新疾病?	(6 9)
本节小结	(7 4)
2.3 生物医学基因组研究	(7 4)
2.3.1 我们能利用基因组序列制造新疫苗吗?	(7 4)
2.3.2 我们能够开发出新型抗生素吗?	(7 7)
2.3.3 我们能够发明新型药物吗?	(7 9)
2.3.4 我们怎么会会对致命的大肠杆菌感兴趣?	(8 3)
数学备忘录 2.3 我们如何才能判断碱基组成是否不同?	(8 3)
本节小结	(8 5)
本章总结	(8 6)
参考文献	(8 6)
第 3 章 基因组变异	(8 8)
3.1 环境因素案例分析	(8 8)
3.1.1 基因组多样性能影响全球变暖吗?	(8 9)
数学备忘录 3.1 如何衡量遗传变异?	(9 2)
数学备忘录 3.2 群体如何建模?	(9 4)
本节小结	(9 5)
3.2 人类基因组变异	(9 6)
3.2.1 人类基因组有多大差异?	(9 6)
数学备忘录 3.3 所有 SNP 都真的是 SNP 吗?	(9 9)
3.2.2 我们为何关注 SNP?	(100)
3.2.3 SNP 引起疾病的案例	(103)
3.2.4 在非疾病 QTL 中存在由 SNP 导致的变化吗?	(106)
3.2.5 为什么 SNP 如此狂热? 药物基因组学!	(109)
本节小结	(111)
3.3 最终的基因组表型——死亡	(112)
3.3.1 我们为什么衰老?	(112)
3.3.2 延长寿命有隐含的成本吗?	(113)
3.3.3 细菌也有基因组的利弊平衡吗?	(115)
本节小结	(117)
3.4 基因组变异的伦理后果	(117)
3.4.1 基因改造生物有害吗?	(118)
3.4.2 遗传检测有益吗?	(121)
3.4.3 复杂基因组的简单应用存在吗?	(126)

本节小结	(129)
本章总结	(130)
参考文献	(130)
第二篇 基因组表达	
第4章 DNA 芯片基础研究	(134)
4.1 DNA 芯片简介	(134)
4.1.1 我的“家酿”怎么了?	(135)
数学备忘录 4.1 怎样对数据进行处理以避免分数比值?	(143)
数学备忘录 4.2 怎样评估表达模式的相似性?	(144)
数学备忘录 4.3 怎样进行基因聚类?	(146)
4.1.2 从经受外界环境变化的酵母身上,我们能够了解到什么?	(160)
4.1.3 为什么有些基因有很多拷贝而其他基因则没有?	(164)
4.1.4 启动子怎么调控基因表达?	(165)
4.1.5 启动子可以逆向工作吗?	(167)
本节小结	(169)
4.2 DNA 芯片的其他应用	(169)
4.2.1 为什么许多无关的基因会拥有相同的表达谱?	(169)
数学备忘录 4.4 比较基因表达矩阵列有什么用?	(170)
4.2.2 细胞能够对自己的基因加以验证吗?	(174)
本节小结	(180)
本章总结	(180)
参考文献	(181)
第5章 DNA 芯片的应用研究	(183)
5.1 癌症和基因组芯片	(183)
5.1.1 有更好地诊断癌症的方法吗?	(184)
数学备忘录 5.1 什么是签名基因? 如何使用签名基因?	(187)
5.1.2 能否用生物芯片对乳腺癌也进行分类呢?	(191)
5.1.3 癌细胞中的基因组发生了什么样的变化呢?	(195)
本节小结	(199)
5.2 使用 DNA 芯片改进卫生保健	(199)
5.2.1 为什么肺结核疫苗效果减弱?	(200)
5.2.2 这种药真的有效吗?	(203)
5.2.3 我们能否预知何种药物对不同的癌症有效?	(209)
5.2.4 脂肪的积累会发生什么?	(213)
本节小结	(221)
本章总结	(221)
参考文献	(222)
第6章 蛋白质组学	(224)
6.1 前言	(225)

6.1.1 所有这些蛋白质能做什么?	(225)
6.1.2 不同的培养条件下需要哪些不同的蛋白质?	(232)
数学备忘录 6.1 如何得知样品中有足够数量的细胞?	(234)
6.1.3 人体缺乏某些蛋白质能够存活吗?	(237)
本节小结	(238)
6.2 蛋白质三维结构	(238)
6.2.1 蛋白质的结构能否揭示它的功能?	(238)
6.2.2 我们能利用结构来发展更好的药物吗?	(241)
6.2.3 单个蛋白质能致命吗?	(242)
本节小结	(244)
6.3 蛋白质相互作用网络	(244)
6.3.1 哪些蛋白质之间发生相互作用?	(245)
6.3.2 如何对蛋白质相互作用进行测量?	(245)
数学备忘录 6.2 Sup35 是蛋白质相互作用网络的中心蛋白质吗?	(249)
6.3.3 有可能在蛋白质组水平上了解蛋白质间的相互作用吗?	(251)
本节小结	(254)
6.4 蛋白质测量	(254)
6.4.1 怎样知道什么蛋白质存在?	(254)
6.4.2 白细胞杀死病原体过程中蛋白质发挥什么作用?	(259)
6.4.3 每个蛋白在细胞中存在多少拷贝?	(261)
6.4.4 我们能制造蛋白质芯片吗?	(269)
6.4.5 所有细胞地位是等同的吗?	(273)
6.4.6 蛋白质组的产物是什么?	(277)
本节小结	(280)
本章总结	(280)
参考文献	(281)

第三篇 全基因组观点

第 7 章 单基因中的基因组回路	(284)
7.1 解析基因组回路	(285)
7.1.1 基因组如何控制个体基因?	(285)
7.1.2 基因如何控制它的转录地点、时间和数量?	(288)
7.1.3 模块 G 的作用是什么?	(295)
7.1.4 能否将工程学和计算机科学中的概念应用到基因当中去?	(307)
本节小结	(310)
7.2 整合单基因回路	(311)
7.2.1 如何描述已知的基因组回路?	(311)
7.2.2 技术提示	(312)
7.2.3 有可能将蛋白质相互作用和 DNA 结合信息做成回路图吗?	(312)
本节小结	(313)

本章总结	(313)
参考文献	(314)
第8章 整合基因组回路	(315)
8.1 简单整合回路	(316)
8.1.1 基因能够形成转换开关并作出转换选择吗?	(316)
数学备忘录 8.1 随机模型怎样应用于细胞过程中?	(317)
8.1.2 人类能够构建基因转换开关吗?	(322)
8.1.3 人类能够利用转换开关构建生理周期时钟吗?	(324)
8.1.4 如果转换开关的噪音很多,那么多细胞生物如何发育呢?	(327)
8.1.5 冗余性:拥有多拷贝基因真的会带来福音吗?	(328)
本节小结	(330)
8.2 复杂整合回路	(330)
8.2.1 回路对于学习很重要?	(331)
数学备忘录 8.2 预测稳态行为可能吗?	(338)
8.2.2 通过分析癌症回路我们能够更好地了解它吗?	(346)
8.2.3 如果回路是相互连接的,那么基因之间是否也是连接的呢?	(350)
本节小结	(355)
本章总结	(355)
参考文献	(356)
第9章 全基因组回路建模	(357)
9.1 基因组学:一个新的前景?	(357)
9.1.1 参与人员:谁在研究系统生物学?	(358)
9.1.2 通讯的质量:系统生物学家提什么问题?	(359)
9.2 我们能够用系统的方法对整个真核生物建模吗?	(359)
9.2.1 基因组学和蛋白质组学的比较	(364)
9.2.2 建立系统模型	(365)
9.2.3 讯息的背景	(368)
9.3 系统生物学会走向繁荣吗?	(369)
本章总结	(370)
参考文献	(370)

第四篇 从遗传学到基因组学:对于医学案例的研究

第10章 我的孩子究竟怎么了?	(372)
10.1 首先要关心的问题	(372)
10.1.1 第一阶段:临床表现	(372)
10.1.2 第二阶段:家庭病史研究	(374)
10.1.3 第三阶段:染色体组型及连锁分析	(374)
10.1.4 第四阶段:DNA 序列分析	(377)
本节小结	(378)
10.2 了解这一疾病的下一步工作	(379)