科爱传播
KE AI COMMUNICATIONS

导读版

实验室解决方案

易于使用

值得信赖

专业权威

# 组学数据生物信息学

## 研究方法与实验方案

# Bioinformatics for Omics Data

## Methods and Protocols

Bernd Mayer

科学出版社

实验室解决方案

# Bioinformatics for Omics Data

## Methods and Protocols

# 组学数据生物信息学

## 研究方法与实验方案

Edited by
Bernd Mayer
*emergentec biodevelopment GmbH, Vienna, Austria*

科学出版社
北京

**图字：01-2012-6186 号**


ISBN: 978-1-61779-026-3



**图书在版编目(CIP)数据**

组学数据生物信息学：研究方法与实验方案＝Bioinformatics for Omics Data：英文/（奥）迈尔（Mayer，B.）主编．—北京：科学出版社，2013.1
（实验室解决方案）
ISBN 978-7-03-035930-8

Ⅰ.①组… Ⅱ.①迈… Ⅲ.①生物信息论—研究方法—英文②生物信息论—实验方法—英文 Ⅳ.①Q811.4-3

中国版本图书馆 CIP 数据核字（2012）第 258903 号

责任编辑：李小汀 田慎鹏 / 责任印制：钱玉芬
封面设计：耕者设计工作室

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

北京佳信达欣艺术印刷有限公司 印刷
科学出版社发行 各地新华书店经销

*

2013 年 1 月第 一 版 开本：787×1092 1/16
2013 年 1 月第一次印刷 印张：37 3/4
字数：892 000

**定价：178.00 元**
（如有印装质量问题，我社负责调换）

# 中文导读

《组学数据生物信息学：研究方法与实验方案》（*Bioinformatics for Omics Data: Methods and Protocols*）一书是“分子生物学方法”（*Methods in Molecular Biology*）系列丛书之一，由国际著名出版集团斯普林格（Springer）下属 Humana 出版社于 2011 年出版。

本书是目前国际上已经出版的第一本高通量组学数据生物信息学英文著作，全面介绍基因组、转录组、蛋白组、代谢组等各种组学数据分析方法，由奥地利维也纳大学贝恩德·迈尔（Bernd Mayer）教授主编。参加编写本书的作者共 94 位，来自奥地利、德国、意大利、法国、英国、西班牙、荷兰等欧洲各国以及美国、日本的科研机构、大专院校、公司企业。他们多为从事生物信息学特别是组学数据分析的第一线研究开发人员，其中奥地利学者居多，共 21 位。美国印第安纳大学张帆（Fan Zhang）和陈越（Jake Y. Chen）、乔治敦大学胡章志（Zhang-Zhi Hu，原书目录中误写为 Zggang-Zhi Hu）、特拉华大学黄宏展（Hongzhan Huang）等海外华人学者参加了本书编写。

全书共分三篇，即“基础篇”、“专题篇”和“应用篇”。第一篇共分八章，第一章“组学技术、数据和生物信息学原理”是全书的绪论，两位作者均为欧洲生物学信息研究所（European Bioinformatics Institute，EBI）工作人员，其中 Maria V. Schneider 专门从事生物信息学用户培训。本章对基因组、转录组和蛋白组等主要组学数据分析中所用的生物信息技术做了简单介绍，通俗易懂，文笔流畅，建议读者在阅读本书其他章节之前，可先浏览一下本章主要内容。本章正文和注解部分列出了许多相关网址，包括欧洲生物信息学研究所、美国国家生物技术信息中心（National Center for Biotechnology Information，NCBI）、英国 Sanger 研究所（Wellcome Trust Sanger Institute）等国际著名生物信息和基因组中心，Ensembl 等基因组浏览器，小鼠、果蝇、线虫等各种模式生物基因组数据库，以及核酸和蛋白质序列，蛋白质家族、功能、结构等数据库。因此，本章标题译为“常用组学数据分析技术和网络资源”，也许更加贴切。

本篇第八章英文原文为“The Use and Abuse of-*Omes*”，按字面意思直译，应为“‘组’一词的使用和滥用”。根据实际内容，我们将其译为“‘组学’术语的恰当使用”。作者对 PubMed 文献摘要数据库进行了检索，并列表说明检索结果（原书 Table 1）。截至 2010 年 1 月，已有 29 种“组”或“组学”名词术语出现在文献中。这 29 种名词可分为三类，第一类即大家熟知的基因组（Genome）、蛋白组（Proteome）和转录组（Transcriptome）。这三个名词术语在文献中出现时间较早，使用频繁。如“基因组”一词 1943 年就开始使用，在文献标题或摘要中出现近 19 万次。第二类包括代谢组（Metabolome）、相互作用组（Interactome）、表观基因组（Epigenome）、糖组（Glycome）、脂质组（Lipidome）等，多为最近十多年才开始使用，出现次数几十次到几百次不等。而文中未提到的外显子组（Exome）测序技术则是最近三年才开始使用。预计在未来几年中，这些领域的相关研究会进一步深入，文献报道也会增多。而第三类则仅在文献中出现过几次，有的仅出现一两次，如折叠组（Foldome）、信号组（Signalome）等。

这些冠以“组”或“组学”的新名词的出现，固然说明了生命科学研究已经进入以数据驱动的研究阶段，却也难免有赶时髦的“嫌疑”。例如，“信号组”一词只在2011年发表的一篇论文中出现过，文中将它定义为“与信号转导相关并对某个生物信号产生反应的所有基因、蛋白质和配体”。显然，基因组、蛋白组等研究方向比较明确，研究手段也日趋成熟，而“信号组”、“折叠组”等名词术语所指的研究对象尚不明确，研究方法也尚未普及。

此外，同为高通量组学数据，其代表的生物学意义、数据获取方法和分析手段，却很不相同。以转录组数据为例，同一物种的不同个体、同一个体的不同发育阶段、同一发育阶段的不同组织、同一组织在不同环境下的基因表达，都有所不同。搞清组学数据产生的生物学背景和研究目的，是准确理解组学数据分析结果的前提。本章作者特别指出目前组学研究生物学背景知识的缺乏，研究方法的局限性和存在的问题，以及组学数据分析时可能遇到的困难等。建议读者在开始进行新课题研究或在研究过程中遇到困难时，带着问题阅读本章。

本书第二篇为“专题篇”，共包括八章，就基因组、转录组、蛋白组等不同组学数据分析方法，分别进行介绍。第九章“高通量序列数据的计算分析”对二代测序的基本方法做了简单介绍，包括Roche-454 GS FLX、Illumina/Solexa HighSeq 2000、SOLiD 3.0、Helicos等几种常用测序平台，并列表比较它们各自的优缺点（原书Table 1）。应该指出，DNA测序技术发展日新月异，尽管离本章完稿时间刚一年多，表中所列几种测序技术又有了更新和提高。Illumina HiSeq 2000单次运行通量已经从200GB提高到600GB，而新推出的Illumina MiSeq则适用于小型实验室，可在20多小时内完成一次运行。与此同时，美国Life Technoligies公司于2012年推出的Ion Torrent Proton可在2小时内完成一次运行，其通量可达10GB。此外，基于单分子测序技术的第三代测序仪PacBio已经投放市场（表一）。

**表一　常用高通量测序平台比较**

| 测序平台 | 通量（Gb） | 时间 | 读长（bp） | 插入（bp） | 错误率（%） |
|---|---|---|---|---|---|
| Illumina HiSeq 2000 | 600 | 11天 | 100 | 700 | 0.26 |
| Illumina GAIIx | 50 | 10天 | 150 | 700 | 0.76 |
| Illumina MiSeq | 7 | 27小时 | 250 | 700 | 0.80 |
| LifeTech Ion Torrent PGM 314 | 0.02 | 2小时 | 200 | 250 | 1.70 |
| LifeTech Ion Torrent PGM 316 | 0.2 | 2小时 | 200 | 250 | 1.70 |
| LifeTech Ion Torrent PGM 318 | 1 | 2小时 | 200 | 250 | 1.70 |
| LifeTech Ion Torrent Proton | 10 | 2小时 | 200 | 250 | 1.70 |
| Roche 454（GS FLX） | 0.45 | 10小时 | 450-600 | 1000 | 0.01* |
| Roche 454（GS FXL+） | 0.70 | 24小时 | 700-1000 | 700 | 0.01* |
| Roche 454 Junior | 0.035 | 10小时 | 400 | 700 | 0.01* |
| PacBio RS | 0.1 | 2小时 | 1500 | 10000 | 13.0 |

* 15倍测序覆盖度时一致性序列错误率

Nature Biotechnology 2012 年 3 月发表的论文（PubMed 号 22522955）和 BMC Genomics 2012 年 7 月发表的论文（PubMed 号 22827831），分别对 Roche454、Illumina、Ion Torrent 和 PacBio 等目前常用的几种测序方法进行了比较，指出了它们的优缺点和适用范围。

Oxford Nanopore 公司于 2012 年 2 月宣布即将投入使用的单分子测序仪成本更低、精度更高。而其他各种第三代测序仪也正在加紧研制或已经开始生产样机。《遗传学和基因组学杂志》（*J Genet Genomics*，原遗传学报）2011 年 3 月刊载的综述（PubMed 号 21477781）列出了十多种正在研制或开发的第三代测序方法和平台。相信本书出版后不久，高性能、低价格的新一代测序仪将逐步开始投放市场。届时，基因组、转录组、表观基因组等高通量测序，不再是某些测序公司和大型测序中心的专门业务，而是普通生物学实验室的常规手段。郝柏林院士在“基因组测序永无止境的根本原因”一文中指出，根据香农第三定理，“从自然界抽提出来的生物学符号序列，不是随机序列，而是属于同等长度或更长的序列集合中的非典型序列子集合，对它们几乎要一条一条地具体研究”。这就意味着，自然界现有各种物种、同一物种的各个个体、同一个体的不同组织或不同发育阶段，同一组织的不同生理和病理状态，其基因组、转录组、表观基因组序列各不相同，需要逐条测定，逐个研究。而这些海量数据的处理，必须依赖于生物信息学理论、方法和工具。

本书第三篇为“应用篇”，共包括十章，前五章为基本方法，每章各有重点，包括计算分析流程、数据整合、网络推断、文献挖掘等。后五章则为具体应用，包括临床相关组学数据分析、病理过程分析和癌症等疾病相关分子靶鉴别等。作者结合自己研究课题实例，说明如何从多种不同类型的海量组学数据中，根据一定的分析流程和统计学方法，逐步提取所需要的信息。

需要说明的是，本书既不是一本入门教科书，也不是针对某一专题而编写的专著，而是由生物信息学研究、开发和应用人员，根据各自的专业特长和工作经验，就各自熟悉的组学专题，分别撰写的综述，是目前能够见到的最为全面的高通量数据分析相关著作。当然，高通量数据种类繁多，新的研究方法和数据类型不断产生。本书不可能包罗万象，如基于结构的高通量药物筛选和设计、基因组 DNA 甲基化和外显子组测序等方法和数据分析技术，本书没有或几乎没有提及。

本书中文目录中已将各章标题译成中文，中文前言中又分别对各章梗概做了简明扼要的介绍，此处不予赘述。总体说来，本书各篇各章自成体系，读者可以根据自己的基础和需求，各取所需，而不必按步就班、逐篇逐章阅读。此外，本书基本上按照“分子生物学方法”丛书的统一模式撰写，即各章均由“引言—材料—方法—注解”四部分组成。这种模式对有些章节并不完全适用，阅读时请稍加注意。

本书彩色插图，请读者登陆以下网址下载：http://www.kbooks.cn

组学数据的特征是数据量极大，数据类型繁多、不同类型的数据之间关系复杂。如何从错综复杂的海量数据中提取有用信息，并在此基础上归纳总结成为生物学知识，是组学生物信息学的主要任务。近年来，国内不少从事传统的、以生物学问题为导向的生物学实验室已经开始进行以高通量数据为基础、从中挖掘生物学知识的研究。例如，利用二代测

序平台，对尚无基因组序列的物种进行全基因组从头测序，或对已经完成基因组测序的物种进行基因组重测序。而转录组、外显子组测序，由于成本相对较低，所研究的生物学问题往往比较明确，则更是发展迅速。

对于已经或正要从事高通量组学数据分析的研究人员或博士、硕士研究生，本书是一本难得的参考书。当然，从事高通量组学数据分析，除了具有较好的生物学研究背景外，必须具有基本生物信息学基础，必须熟悉 Linux 操作系统，必须学会独立编写程序，必须具备一定的统计学基础。

感谢科学出版社特别是田慎鹏编辑和李小汀编辑的努力，将本书英文版引进国内。以影印版方式尽快出版，而不是将全书翻译成中文，我认为是切合实际的好办法。本书在中国影印出版得到了原书作者贝恩德·迈尔的支持和鼓励，胡章志、陈越等海外华人撰稿人以及胡松年、赵义、罗洪、颜林林等对译稿和导读提出了很好的修改意见。

本人虽从事生物信息学研究和教学多年，但多为单个基因、蛋白质或基因家族等小批量数据分析，真正涉足高通量数据分析，还不到两年。加上时间仓促，尚未对全书认真阅读。上述导读意见，难免有谬误之处，衷心希望读者指正。

北京大学　罗静初

2012 年 10 月 15 日

# 前　言

本书从多个侧面对组学数据生物信息学做了详尽的介绍。组学数据生物信息学是一个全新的研究领域，是分子生物学、应用信息学和统计学等多个学科的交叉和整合。近年来，生物信息学已成为以数据为驱动的组学研究领域的常规技术，也是组学应用研究人员必须掌握的重要技能。生命科学分析仪器的不断小型化，以及数据测定技术的快速发展，使我们能够同时处理并分析多种细胞组份及其状态，为解释特定生命现象提供了有力的工具。然而，如果不作恰当的处理和分析，组学数据对我们更好理解所研究的生命现象毫无裨益，就连对组学方法产生的海量原始数据的管理，在某种程度上也已经成为一种技术。

生物信息学通常被认为是一门技术科学。实际上，它是以计算机为工具进行生物学研究的交叉科学，即“计算生物学”，或“计算分子生物学”。组学的发展促使生命科学的研究模式发生了革命性的改变，传统的以问题为驱动的研究拓展为从数据出发的探索性研究。目前，组学恰恰处于上述两种研究模式彼此融合的发展阶段。从这一点上说，组学数据生物信息学不只关注数据的管理和分析，更关注生物问题的提出和假设的验证。目前，将数据分析策略整合在一起的生物信息工作流程已经出现，而这些流程本身的复杂性也反映了生物体和生命过程的复杂性。通过对大量调控方式和作用网络的综合分析，我们有可能从更高层次上理解细胞内的生命过程。不言而喻，组学生物信息学正逐步向计算系统生物学过渡，其最终目标是对高度动态的生命现象建立定量模型，用以预测由细胞内各种分子间相互作用产生的复杂生命活动的分子过程及功能。

毋庸置疑，组学数据生物信息学的研究对象十分复杂，它是引领现代分子生物学研究的重要分支学科。希望本书不仅能对组学研究起到具体指导的作用，也能为读者展现这一研究领域极具魅力的美好图景。

本书共分三篇。第一篇介绍核心分析策略、标准分析规范、数据管理指南，以及用于分析组学数据的基本统计方法。第二篇介绍用于基因组、转录组、蛋白质组、代谢组等各种不同组学数据的生物信息学分析方法，包括基本概念和实验背景，以及原始数据预处理和深入分析的基本方法。第三篇则介绍如何利用生物信息学进行组学数据分析的实例，包括人类疾病相关生物标记鉴定和靶标识别等具体例子。

衷心感谢参与本书各章编写的所有作者，感谢 John Walker 促成了编著本书的设想。书中若有纰漏之处，责任当由本人承担。希望大家阅读愉快。

奥地利　维也纳
贝恩德·迈尔

（罗静初　译）

# Preface

This book discusses the multiple facets of "Bioinformatics for Omics Data," an area of research that intersects with and integrates diverse disciplines, including molecular biology, applied informatics, and statistics, among others. Bioinformatics has become a default technology for data-driven research in the Omics realm and a necessary skill set for the Omics practitioner. Progress in miniaturization, coupled with advancements in readout technologies, has enabled a multitude of cellular components and states to be assessed simultaneously, providing an unparalleled ability to characterize a given biological phenotype. However, without appropriate processing and analysis, Omics data add nothing to our understanding of the phenotype under study. Even managing the enormous amounts of raw data that these methods generate has become something of an art.

Viewed from one perspective, bioinformatics might be perceived as a purely technical discipline. However, as a research discipline, bioinformatics might more accurately be viewed as "[molecular] biology involving computation." Omics has triggered a paradigm shift in experimental study design, expanding beyond hypothesis-driven approaches to research that is basically explorative. At present, Omics is in the process of consolidating various intermediate forms between these two extremes. In this context, bioinformatics for Omics data serves both hypothesis generation and validation and is thus much more than mere data management and processing. Bioinformatics workflows with data interpretation strategies that reflect the complexity of biological organization have been designed. These approaches interrogate abundance profiles with regulatory elements, all expressed as interaction networks, thus allowing a one-step (descriptive) embodiment of wide-ranging cellular processes. Here, the seamless transition to computational Systems Biology becomes apparent, the ultimate goal of which is representing the dynamics of a phenotype in quantitative models capable of predicting the emergence of higher order molecular procedures and functions that arise from the interplay of basic molecular entities that constitute a living cell.

Bioinformatics for Omics data is certainly embedded in a highly complex technological and scientific environment, but it is also a component and driver of one of the most exciting developments in modern molecular biology. Thus, while this book seeks to provide practical guidelines, it hopefully also conveys a sense of fascination associated with this research field.

This volume is structured in three parts. Part I provides central analysis strategies, standardization, and data management guidelines, as well as fundamental statistics for analyzing Omics profiles. Part II addresses bioinformatics approaches for specific Omics tracks, spanning genome, transcriptome, proteome, and metabolome levels. For each track, the conceptual and experimental background is provided, together with specific guidelines for handling raw data, including preprocessing and analysis. Part III presents examples of integrated Omics bioinformatics applications, complemented by case studies on biomarker and target identification in the context of human disease.

I wish to express my gratitude to all authors for their dedication in providing excellent chapters, and to John Walker, who initiated this project. As for any omissions or errors, the responsibility is mine. In any case, enjoy reading.

*Vienna, Austria* *Bernd Mayer*

# Contributors


S. James Adelstein • *Harvard Medical School, Harvard University, Boston, MA, USA*

Shahab Asgharzadeh • *Department of Pediatrics and Pathology, Keck School of Medicine, Childrens Hospital Los Angeles, University of Southern California, Los Angeles, CA, USA*

Stephan J.L. Bakker • *Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands*

Catherine A. Ball • *Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA*

Marco Behr • *emergentec biodevelopment GmbH, Vienna, Austria*

Touati Benoukraf • *Université de la Méditerranée, Marseille, France; Centre d'Immunologie de Marseille-Luminy, Marseille, France; CNRS, UMR6102, Marseille, France; Inserm, U631, Marseille, France*

Andreas Bernthaler • *emergentec biodevelopment GmbH, Vienna, Austria*

Chris Bielow • *AG Algorithmische Bioinformatik, Institut für Informatik, Freie Universität Berlin, Berlin, Germany*

Pierre Cauchy • *Inserm, U928, TAGC, Marseille, France; Université de la Méditerranée, Marseille, France*

Jake Y. Chen • *Indiana University School of Informatics, Indianapolis, IN, USA*

Stephen A. Chervitz • *Affymetrix, Inc., Santa Clara, CA, USA*

Irina Dalah • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*

Eric W. Deutsch • *Institute for Systems Biology, Seattle, WA, USA*

Ana Dopazo • *Genomics Unit, Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain*

Anatoly Dritschilo • *Lombardi Cancer Center, Georgetown University, Washington, DC, USA*

Daniela Dunkler • *Section of Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria*

David P. Enot • *BIOCRATES life sciences AG, Innsbruck, Austria*

Raul Fechete • *emergentec biodevelopment GmbH, Vienna, Austria*

Pierre Ferrier • *Centre d'Immunologie de Marseille-Luminy (CIML), Marseille, France*

Dawn Field • *NERC Centre for Ecology and Hydrology, Oxford, UK*

Claudio Franceschi • *'L Galvani' Interdept Center, University of Bologna, Bologna, Italy*

Alberto de la Fuente • *CRS4 Bioinformatica, Parco Tecnologico SOLARIS, Pula, Italy*

Srinubabu Gedela • *Stanford University School of Medicine, Stanford, CA, USA*

Gernot Glöckler • *emergentec biodevelopment GmbH, Vienna, Austria*

Martin G. Grigorov • *Nestlé Research Center, Lausanne, Switzerland*

Clemens Gröpl • *Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, Germany*

Bernd Haas • *BIOCRATES life sciences AG, Innsbruck, Austria*

Jörg Hackermüller • *Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig, Germany; Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany*
Hubert Hackl • *Division for Bioinformatics, Innsbruck Medical University, Innsbruck, Austria*
Martin Haiduk • *emergentec biodevelopment GmbH, Vienna, Austria*
Arye Harel • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*
Georg Heinze • *Section of Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria*
Merel E. Hellemons • *Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands*
Steve Hoffmann • *Interdisciplinary Center for Bioinformatics and The Junior Research Group for Transcriptome Bioinformatics in the LIFE Research Cluster, University Leipzig, Leipzig, Germany*
Friedemann Horn • *Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany; Institute of Clinical Immunology, University of Leipzig, Leipzig, Germany*
Zhang-Zhi Hu • *Lombardi Cancer Center, Georgetown University, Washington DC, USA*
Hongzhan Huang • *Center for Bioinformatics & Computational Biology, University of Delaware, Newark, DE, USA*
Mira Jung • *Lombardi Cancer Center, Georgetown University, Washington, DC, USA*
Amin I. Kassis • *Harvard Medical School, Harvard University, Boston, MA, USA*
Oliver Kohlbacher • *Eberhard-Karls-Universität Tübingen, Tübingen, Germany*
Vinod Kumar • *Computational Biology, Quantitative Sciences, GlaxoSmithKline, King of Prussia, PA, USA*
Hiddo J. Lambers Heerspink • *Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands*
Doron Lancet • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*
Paola Lecca • *The Microsoft Research – University of Trento Centre for Computational and Systems Biology, Povo, Trento, Italy*
Yonghong Li • *Celera Corporation, Alameda, CA, USA*
Henrique Lopes de Abreu Madeira • *emergentec biodevelopment GmbH, Vienna, Austria*
Arno Lukas • *emergentec biodevelopment GmbH, Vienna, Austria*
Gert Mayer • *Department of Internal Medicine IV (Nephrology and Hypertension), Medical University of Innsbruck, Innsbruck, Austria*
Harald Mischak • *mosaiques diagnostics GmbH, Hannover, Germany*
Irmgard Mühlberger • *emergentec biodevelopment GmbH, Vienna, Austria*
Thanh-Phuong Nguyen • *The Microsoft Research – University of Trento Centre for Computational and Systems Biology, Povo, Trento, Italy*
Rainer Oberbauer • *Medical University of Vienna and KH Elisabethinen Linz, Vienna, Austria*
Soichi Ogishima • *Department of Bioinformatics, Medical Research Institute, • Tokyo Medical and Dental University, Tokyo, Japan*

Sandra Orchard • *EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton,Cambridge, UK*
Antonio Ortega • *Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA*
Helen Parkinson • *EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
Paul Perco • *emergentec biodevelopment GmbH, Vienna, Austria*
Shmuel Pietrokovski • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*
Roger Pique-Regi • *Department of Human Genetics, University of Chicago, Chicago, IL, USA*
Corrado Priami • *The Microsoft Research – University of Trento Centre for Computational and Systems Biology, Povo, Trento, Italy*
Sonja J. Prohaska • *Department of Computer Science and Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany*
John Quackenbush • *Department of Biostatistics, Dana-Farber Cancer Institute, Boston, MA, USA*
Paola Quaglia • *The Microsoft Research – University of Trento Centre for Computational and Systems Biology, Povo, Trento, Italy*
Johannes Rainer • *Bioinformatics Group, Division Molecular Pathophysiology, Medical University Innsbruck, Innsbruck, Austria*
Kristin Reiche • *Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany*
Knut Reinert • *AG Algorithmische Bioinformatik, Institut für Informatik, Freie Universität Berlin, Berlin, Germany*
Anna T. Riegel • *Lombardi Cancer Center, Georgetown University, Washington, DC, USA*
Phillipe Rocca-Serra • *EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
Simon Rogers • *Inference Research Group, Department of Computing Science, University of Glasgow, Glasgow, UK*
Peter Rossing • *Steno Diabetes Center Denmark, Gentofte, Denmark*
Marilyn Safran • *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*
Fátima Sánchez-Cabo • *Genomics Unit, Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain*
Susanna-Assunta Sansone • *EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
Maria V. Schneider • *EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
Katharina Schutt • *Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany; Institute of Clinical Immunology, University of Leipzig, Leipzig, Germany*
Dov Shiffman • *Celera Corporation, Alameda, CA, USA*
Peter F. Stadler • *Department of Computer Science and Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany*

Christian J. Stoeckert Jr • *Department of Genetics and Center for Bioinformatics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
Hiroshi Tanaka • *Department of Computational Biology, Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan*
Chris F. Taylor • *EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK*
Ronald Taylor • *Computational Biology & Bioinformatics Group, Pacific Northwest National Laboratory, Richland, WA, USA*
Alberto Termanini • *'L Galvani' Interdept Center, University of Bologna, Bologna, Italy*
Paolo Tieri • *'L Galvani' Interdept Center, University of Bologna, Bologna, Italy*
Zlatko Trajanoski • *Division for Bioinformatics, Innsbruck Medical University, Innsbruck, Austria*
Kerstin Boll • *Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany; Institute of Clinical Immunology, University of Leipzig, Leipzig, Germany*
Melissa Warden • *Department of Pediatrics and Pathology, Keck School of Medicine, Childrens Hospital Los Angeles, University of Southern California, Los Angeles, CA, USA*
Klaus M. Weinberger • *BIOCRATES life sciences AG, Innsbruck, Austria*
Anton Wellstein • *Lombardi Cancer Center, Georgetown University, Washington, DC, USA*
Martin Wiesinger • *emergentec biodevelopment GmbH, Vienna, Austria*
Julia Wilflingseder • *Medical University of Vienna and KH Elisabethinen Linz, Vienna, Austria*
Cathy H. Wu • *Center for Bioinformatics & Computational Biology, University of Delaware, Newark, DE, USA*
Yongliang Yang • *Department of Radiology, Harvard Medical School, Harvard University, Boston, MA, USA; Center of Molecular Medicine, Department of Biological Engineering, Dalian University of Technology, Dalian, China*
Dick de Zeeuw • *Department of Nephrology, University Medical Center Groningen, Groningen, The Netherlands*
Fan Zhang • *Indiana University School of Informatics, Indianapolis, IN, USA*

# 目　录



（罗静初　译）

# Contents

# Part I

## Omics Bioinformatics Fundamentals