

易于使用

值得信赖

专业权威

# 蛋白质网络与途径分析

## Protein Networks and Pathway Analysis

Yuri Nikolsky and Julie Bryant

实验室解决方案

# Protein Networks and Pathway Analysis

## 蛋白质网络与途径分析

Edited by

Yuri Nikolsky and Julie Bryant

GeneGo, Inc. • Encinitas, CA • USA

科学出版社

北京

图字：01-2012-6184 号

This is an annotated version of

**Protein Networks and Pathway Analysis**

Edited by Yuri Nikolsky and Julie Bryant.

Copyright © Humana Press, a part of Springer Science + Business Media, LLC 2009

ISBN: 978-1-60761-174-5

All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

This reprint has been authorized by Springer-Verlag (Berlin/Heidelberg/New York) for sale in the People's Republic of China only and not for export therefrom.

本版本由 Springer 出版公司（柏林/海德堡/纽约）授权，仅限在中华人民共和国境内销售，不得出口。

#### 图书在版编目(CIP)数据

蛋白质网络与途径分析=Protein Networks and Pathway Analysis:  
英文 / (美) 巴斯 (Nikolsky, Y.) 等主编. —北京: 科学出版社, 2013. 1

(实验室解决方案)

ISBN 978-7-03-035928-5

I. ①蛋… II. ①巴… III. ①蛋白质—基因组—研究—英文  
IV. ①Q51

中国版本图书馆 CIP 数据核字 (2012) 第 258930 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: 27

字数: 637 000

定价: 145.00 元

(如有印装质量问题, 我社负责调换)

## 导 读

人类基因组计划完成后，海量的功能基因组、蛋白质组和代谢组等“组学”数据已开始涌现，基因或蛋白质的序列信息并不能解释全部的生命现象。生物学研究的热点由对单个基因或蛋白质的局部性研究，转移到以全部基因、蛋白质以及代谢产物为研究对象的“组学”研究，即整体性研究。“组学”技术的快速发展把分子生物学推入了系统生物学时代。作为人类基因组计划的发起人之一的美国科学院院士莱诺伊·胡德（Leroy Hood）认为“系统生物学将是 21 世纪医学和生物学的核心驱动力”。2002 年 3 月，美国《科学》周刊登载了系统生物学专集，该专集导论中的第一句话这样写道：“如果对当前流行的、时髦的关键词进行分析，那么人们会发现‘系统’高居在排行榜上。”

那么，什么是系统生物学？系统生物学是在细胞、组织、器官和生物体整体水平研究结构和功能各异的分子及其相互作用，并通过计算生物学来定量描述和预测生物功能、表型和行为。系统生物学将在基因组序列的基础上完成由生命密码到生命过程的研究，这是一个逐步整合的过程，由生物体内各种分子的鉴别及其相互作用的研究到途径、网络和模块，最终完成整个生命活动的路线图。系统生物学的特点是把水平型研究（基因组学、蛋白质组学和其他“组学”的研究，即以单一的手段同时研究成千上万个基因或蛋白质）和垂直型研究整合起来，成为一种“三维”的研究。此外，系统生物学还是典型的多学科交叉研究，它需要生命科学、计算机科学、信息科学和数学等学科的共同参与。

蛋白质是行使生物功能的重要生物大分子。蛋白质相互作用网络几乎参与了生命活动过程的各个环节。从遗传物质的复制、基因的表达调控到细胞间的信号转导、细胞的代谢过程以及病原微生物的致病、宿主对病原微生物的免疫，蛋白质相互作用都在其中扮演着重要的角色。利用计算机强大的数据处理功能，生物学家们可以把生物实验和理论预测结合起来，探究一些人的分析能力所不能解决的问题。同时计算机分析后产生的一些新理论又可以通过生物实验得以验证。对生物体系统而言，生理上分子间的相互作用网络可以使用“结点-边”的计算机语言来表述（用从实验数据获得的基因或蛋白质作为节点，用它们之间的联系作为边），如蛋白质-蛋白质、蛋白质-核酸、蛋白质-代谢物相互作用网络等。对于蛋白质相互作用网络，可以构建一个双向网络来表述，而对于代谢过程中不可逆的反应，可以构建一个有向网络。这样就可以通过计算机构建生物网络来模拟生物体系统，探索生物体系统的生理活动。近年来，生物网络的研究发展非常迅速，特别是集中于探索活细胞中分子的行为和它们之间的相互作用。例如现在的药物设计不只是停留在一个蛋白一个配体这样的传统模型，而认识到了机体被配体刺激后激活的响应机制是一个网络过程。从网络或者系统生物学水平通过计算机模拟、预测药物相互作用网络并与实验相结合的方法去思考药物设计的效果以及副作用。系统分析大量蛋白质在生物系统中的相互作用关系，了解疾病等特殊生理状态下生物信号和物质代谢的反应机制，以及了解蛋白质之间的功能联系都有重要的科学价值和应用前景。

如何处理大量的“组学”数据，并找到其中的网络及调控关系是所有研究人员面临的一个重大难题。《蛋白质网络与途径分析》（*Protein Networks and Pathway Analysis*）弥

补了国内这个领域的空白，它是分子生物学方法中难得的一本好书。本书由来自于欧美的学术领域、政府研究机构、医药产业和生物信息学公司的一线从业人员撰写，内容具有专业性和前沿性，全面广泛地介绍了系统生物学实验和函数数据分析的具体分析方法。本书的内容分为三个部分。第一部分阐明蛋白质、化合物和基因间的相互作用以及蛋白质相互作用的手工注释。第二部分是关于函数分析的专业工具介绍，包括 PANTHER 和 Cytoscape 等，并对复杂的、多数据类型及多工具分析提供了相关的解决方案。第三部分介绍了函数分析的应用，例如在高通量 siRNA 筛选、miRNA、疾病方向生物信息学分析、转化医学和生物信息学、高密度等位基因相关数据的网络和途径分析等方面的应用。由浅入深地讲述了函数分析方法及其在药物开发、转化医学和基础医学等领域的应用，适合该领域的研究人员、技术人员和研究生参考阅读。

本书的作者 Nikolsky 博士，GeneGo 公司总裁，在生命科学领域有着数十年的工作经验。在进入 GeneGo 公司之前，曾担任 ChemDiv 公司的总裁。在此之前，他与其他研究人员共同创立了 Integrated Genomics 公司（位于芝加哥的一个比较基因组学及生物信息学公司）并担任业务发展部副总裁。Nikolsky 博士还是 Thermogen 公司的资深科学家，多年的研究和工作经历使 Nikolsky 博士对蛋白质、小分子化合物等大量的“组学”数据的分析方法有着独特的见解和丰富的经验。本书中每个章节从摘要、背景介绍、材料设备、方法步骤、分析解释到参考文献一应俱全，利用一个实验案例详细地介绍了常用系统生物学网络与途径分析的方法。尤其是实际操作方面，从实验设计、执行到结果分析，研究人员可以参照书中的步骤一步一步地完成，这对于初学者来说具有很好的指导作用。有的章节还列出了需要注意的问题，这些问题都是前人经验的沉淀，也往往是实验成败的关键因素。熟练掌握这些经验往往可以取得事半功倍的效果。本书通过理论结合实践的讲述来帮助研究人员快速地学习和掌握“组学”实验方法和数据分析技术，快速找到合适的解决方案。这对于“组学”领域研究人员来说确实是一本重要而且实用的参考资料。当然，由于函数数据分析是一门快速发展的学科，几乎每周都会有新的软件和算法出现，本书不能完全收录新出现的技术，这些技术只能从新的文献中去补充学习。最后，真心的希望本书能给读者打开一道通往“组学”时代的大门。

高友鹤 尹剑锐

## 前 言

目前，在谷歌上用“systems biology (系统生物学)”作为检索词，可以检索到 1400 万条信息，用“pathway analysis (途径分析)”作为检索词，可以检索到 200 万条信息。这些数据深刻地说明了系统数据分析 (systems data analysis) 在现代生物科学中无处不在。近年来，任何关于微矩阵的表达和 SNP 的文章都会有一个与途径、本体富集 (ontology enrichment) 或者生物网络有关的章节。尽管这是一个新兴领域，但是系统生物学已经被广泛地应用于基础研究、临床前药物研究到转化医学以及个体化的医疗保健等领域。

系统生物学在维基百科中的概念是：在生物系统中采用整合 (integration) 而不是还原 (reduction) 的新视角，系统地研究复杂的相互作用。从实际的角度来看，这意味着我们要将积累的生物学知识整合成可用计算机阅读的格式，还要创造出一些工具，利用这些工具可以将这些格式化了的信息应用到生物实验和化学实验的数据分析中。从 20 世纪 70 年代开始，生物化学成为第一个将信息录入数据库的学科，例如 BRENDA、EMP/MPW 和 KEGG 数据库。多年来，科学家把调节和信号组成的部分以蛋白质相互作用数据库 (例如 HPRD 和 BIND 数据库) 的形式增加到生物化学中。同时，关于细胞生物过程和蛋白质功能的理解也有一定发展，这其中最出名的就是 GO 数据库。

函数分析 (functional analysis) 与高通量的，或者说“组学”的实验生物学是不可分割的，从 20 世纪 90 年代后期得到了快速的发展。在基因组规模内，对数以千计的嘈杂组学数据进行分析对于当时的湿实验室<sup>①</sup>的研究者来说是很难的，而且统计学工具仅仅能帮助我们减少数据的复杂性，却不能帮助我们理解潜在的生物学原理。渐渐地，生物信息学家和湿实验室生物学家开始用已有的途径和过程信息的数据库实现实验数据的绘制 (map) 和优先化 (prioritization) (富集分析)。后来从多年的图论和物理学研究中借鉴了一些经验，生物学网络被加入到分析工具箱里。

本书汇编了函数分析方法及其应用，由学术领域、政府研究机构、医药产业和生物信息学公司的一线从业人员编写。该书的内容分为三个部分。第一部分阐明蛋白质、化合物和基因间的相互作用。第二部分主要介绍了与函数分析相关的工具。第三部分讲述了函数分析的应用。

二元的相互作用是细胞内蛋白质、基因和 RNA 行使功能的基本单位。在函数分析中，这些相互作用是计算机模拟的细胞生命即网络和途径中的重要组成成分。这些函数分析结果的质量取决于网络和途径组成的相互作用的质量。因此，揭示相互作用的方法具有重要的意义。在这里，我们列举了一些著名的湿实验室和干实验室<sup>②</sup>获得相互作用信息时用到的工具。我们可以通过生物学实验或者已经发表的实验数据等多种方式获得蛋白质的相互作用信息。湿实验室工具包括可以全面分析蛋白质与蛋白质之间物理上的相互作用的蛋白

---

① 湿实验室：主要指实验室内的实验研究

② 干实验室：主要指计算机模拟和理论分析

质芯片 (Invitrogen 公司), 和一个可以更准确检测转录因子和基因启动子之间的相互作用的芯片 (A-Star 公司, 新加坡)。在基于计算机的方法中, 不同形式的文本挖掘和自然语言处理起到了越来越重要的作用。Linguamatics 公司是一个专注于从 PubMed 和其他数据库中进行文本挖掘相互作用信息并且发展迅速的公司, 这里的研究人员用一个章节介绍了与文本挖掘相关的内容。当然, 我们可以在实验数据里挖掘出相互作用信息, 然而手工注释的过程是缓慢的、耗费精力的但却是高质量的。在本书中专门介绍注释技术的专家来自两个方面, 一部分专家来自生物学领域, 以蛋白质-蛋白质以及蛋白质-DNA 或者与 RNA 的相互作用和代谢为研究重点 (Metalogic 公司, 俄罗斯), 另一部分专家来自化学领域, 以蛋白质-化合物的相互作用为研究重点 (GVK 公司, 印度)。

多数函数分析的方法 (并不是全部) 可分为两类: 基因名称表富集分析 (EA) 和生物学网络。第一类包含许多步骤, 我们利用这些步骤计算出我们感兴趣的数据集在特定的途径、细胞学过程、蛋白质功能、疾病靶向和生物标记物中的富集程度。基因名称表富集分析的结果是在数据内部对于实体进行基于特定打分方法的定量排序。在这里, BMS 公司研究人员介绍了关于基因名称表富集分析中经典的 GSEA 方法; 斯坦福国际研究院 (SRI International) 的研究人员用一个章节详细地描述了生物本体是基因名称表富集分析实体的原始资料。生物网络用动态的方式表示数据, 它用从实验数据获得的基因和蛋白质作为结点, 用它们之间的联系作为边。与静态的途径和流程不同的是, 网络是动态的并且在数据库中是唯一的, 提供了较高的分辨率。正因为如此, 网络是一个提出新假设时可以选择的工具。关于网络分析方法的六个章节构成了本书中分析部分的绝大多数内容。Merck 公司科技工作者撰写的章节全面综述了网络生物学, 包括其概念和用网络对基因名称表富集分析时所用的工具。在讲述 bioPixies 软件的章节中, 研究人员总结了概率网络图在分析大量固有复杂组学数据库中的应用, 并且介绍了一个由普林斯顿大学 (Princeton University) 开发的著名的贝叶斯网络包: bioPixies。GeneGo 公司提供了一个基于知识的网络生成工具包, 作为 MetaCore 数据挖掘套件的一部分。网络分析中的下一步, 即动态网络模型, 由俄罗斯系统生物学研究所 (the Institute of Systems Biology) 和英国的爱丁堡大学 (University of Edinburg) 一个合作小组提供。来自德国慕尼黑 Biomax 公司的小组描述了一个在 BioXM 软件的知识管理环境中实现网络分析的语义法, 这个语义法与众不同并且有着强大的功能。网络分析集还包括一个章节, 由西雅图系统生物研究所 (Institute of Systems Biology) 的科技人员详细地介绍了非常流行且开源的网络可视化工具 Cytoscape 和与其快速增长的一套应用程序插件。函数分析部分还用了一个章节介绍了一个由英国 Infosense 公司开发的非常实用的集成软件包: KDE, 其目的是从众多函数分析应用中建立强大且灵活的分析工作流程。

在函数分析的应用部分, 我们从用户的角度 (无论是基础研究还是药物开发/管理领域) 选择了一些函数分析的方法。来自于转化基因组研究院 (the Translational Genomics Institute) 的小组人员提出了一个先进的技术, 这个技术把 siRNA/化合物的筛选和途径数据分析结合起来应用到癌症研究领域。费城的天普大学 (Temple University) 和意大利几个研究型大学组成的小组描述了大规模的 miRNA 研究中网络分析的应用。斯克利普斯基因组研究所 (Scripps Genomics Research) 和加州大学圣地亚哥分校 (UCSD) 的人类

遗传学家小组提供了在 SNP 遗传相关性研究中途径工具的用法。还有两个章节专门介绍了疾病研究中函数工具和数据库的大规模应用。囊性纤维化治疗基金会 (Cystic Fibrosis Foundation Therapeutics) 领导的小组开发了一个以疾病为中心的数据分析环境 MetaMiner, 这个数据分析环境结合了详细疾病知识库和包含基因名称表富集分析、网络和相互作用组工具的工具箱。在 Strategic Medicine 公司小组撰写的转化医学相关章节中加入了临床上对途径分析的理解。最后, 美国国家毒理学研究中心 (食品及药物管理研究所 FDA) 研究小组提供了一款软件套件: ArrayTrack, 它被美国食品及药物管理研究所研究人员和新药物应用评估人员广泛使用。这个套件最初被用于芯片表达分析, 后来它又结合了许多公共统计工具和商业途径工具而被应用于其他的组学分析。

函数数据分析 (Functional data analysis) 作为一门快速发展的学科, 几乎每周都会出现新的网络算法和软件工具, 其应用范围也随着每一个新的 DNA、RNA 和蛋白质实验进入市场而扩大。因此, 我们不能, 也没有打算在本书中收录所有的工具。相反, 我们试着把焦点集中在已建立的方法和一些在市场上常见的软件包, 让读者对这个新兴领域中的议题和应用有更广泛的了解。

(高友鹤 尹剑锐 译)



---

## Preface

Recent Google queries for “systems biology” and “pathway analysis” fetched over 14 and 2 million entries, respectively. These numbers speak volumes about the ubiquity and popularity of systems data analysis in modern bioscience. These days, any microarray expression or SNP-based article would feature a chapter on pathways, ontology enrichment, and/or biological networks. Although a very young field, the applications of systems biology now spread widely from basic research and pre-clinical drug discovery to translational research and personalized health care.

Systems biology “focuses on the systematic study of complex interactions in biological systems, thus using a new perspective (integration instead of reduction) to study them” (Wikipedia). From a practical view point, it means integration of accumulated biological knowledge in a computer-readable format and the creation of tools for applying this structured information to the analysis of biological and chemical experimental data. Starting in the 1970s, biochemistry was the first field from which information was put into databases such as BRENDA, EMP/MPW, and, later, KEGG. Over the years, regulation and signaling components were added to biochemistry in the form of protein interaction databases such as HPRD and BIND, and comprehensive ontologies of cellular processes and protein functions were developed, the best known of which is Gene Ontology (GO).

Functional analysis is inseparable from high-throughput, or “OMICs,” experimental biology, which has been rapidly evolving since the late 1990s. The “genome-wide”, noisy OMICs assays with thousands of data points were nearly illegible for wet lab researchers at the time, and statistical tools only helped to reduce data complexity but not to understand the underlying biology. Gradually, bioinformaticians and wet lab biologists started to use databases of pathways and processes available to them for mapping and prioritization of experimental data (enrichment analysis). Later, biological networks were added to analysis toolboxes, borrowing from years of research in graph theory and physics.

This book represents a compilation of methods of functional analysis and its applications, written by practitioners from academy, government research organizations, the pharmaceutical industry, and bioinformatics companies. The content is divided into three sections. The first section deals with elucidation of protein, compound, and gene interactions. The second section is devoted to analytical tools of functional analysis. Applications of functional analysis is covered in the third section.

Binary interactions are the elementary units of functional organization of proteins, genes, and RNA in the cell. In functional analysis, these interactions play the role of “building blocks” in our computer-readable approximation of cellular life – pathways and networks. The result of functional analysis is only as good as the quality of interactions that pathways and networks are assembled from. Therefore, the methods of elucidation of interactions have a foremost importance. Here we present several prominent wet lab and dry lab tools for extraction of interaction information. There are

many ways to obtain protein interactions, both from experiments and from already published experimental literature. Wet lab tools include profiling of physical protein–protein interactions with protein arrays (Invitrogen) and one of the most accurate versions of detection of transcription factor – gene promoter interactions (A-Star, Singapore). Among the computer-based methods, different types of text mining and language processing play an increasingly important role. We present a chapter on it by Linguamatics, a fast growing company specializing in text mining interaction information from PubMed and other sources. And, of course, the interactions can be dug out of experimental articles by slow, often painful, but high quality manual annotation, which is a technology all its own. We have chapters on annotation technology from specialists in biology-centered protein–protein, protein–DNA/RNA interactions and metabolism (MetaLogic, Russia), and chemistry-centered protein–compound interactions (GVK, India).

Many (but by no means all) methods of functional analysis can be divided into two groups: gene list enrichment analysis (EA) and biological networks. The first group comprises several procedures for calculation of relative enrichment of the data set of interest into components of certain pathways, cellular processes, protein function, disease targets, biomarkers, etc. EA results in quantitative ranking of ontology entities within the data set(s) based on a certain scoring. We present here the “classical” GSEA method of EA (by BMS) and a detailed chapter on biological ontologies as a source of entities for EA by SRI International. Biological networks represent the data in a dynamic way, using genes and proteins from experimental data points as “nodes” and connections between them as “edges”. Unlike static pathways and processes, networks are dynamic and are unique for the data set, providing a higher resolution analysis. Because of this, networks are a tool of choice for applications which require new hypothesis generation. Six chapters on methods of network analysis comprise the bulk of the analysis section for this book. The chapter from Rosetta/Merck Co. provides a comprehensive overview of network biology, concepts, and tools for gene lists analysis using networks. A chapter from bioPixies reviews probabilistic networks applied for analysis of large, inherently noisy OMICs data sets, and describes bioPixie, a well-known Bayesian network package developed at Princeton University. A knowledge-based network generation toolkit is presented by GeneGo, Inc. as part of its MetaCore data mining suite. The next step in network analysis, dynamic network modeling, is presented by a collaborative team from the Institute of Systems Biology (Russia) and University of Edinburg (UK). A team from Biomax, GmbH (Munich, Germany) describe a very different yet powerful semantic approach to network analysis realized in the BioXM Knowledge Management Environment. The network analysis set is concluded with a detailed chapter on the very popular open source “community” network visualization tool Cytoscape and its fast growing set of application plug-ins (Institute of Systems Biology, Seattle). The analysis section also includes a chapter on a very useful integration package, KDE, by Infosense, Ltd. (UK), which is designed for building robust and flexible analytical workflows from multiple functional analysis applications.

In the Applications section, we selected methods dealing with functional analysis from the user perspective, both in basic research and drug discovery/regulation. A group from the Translational Genomics Institute (Tgen, Phoenix) presents an

advanced technology on a combined siRNA/compound screening, coupled with pathway data analysis which was used in cancer research. Application of network analysis in large-scale miRNA studies is described by a group from Temple University (Philadelphia) and several research universities in Italy. A team of human geneticists from Scripps Genomics Research and UCSD present usage of pathway tools in SNP genetic associations studies. Two chapters are devoted to larger scale applications of functional tools and databases for disease research. The group led by Cystic Fibrosis Foundation Therapeutics presents MetaMiner (Cystic Fibrosis), a disease-centered data analysis environment which combines a detailed disease knowledge base with a toolbox of EA, networks, and interactome tools. Clinical perspective on pathway analysis is added in the chapter on translational research submitted by Strategic Medicine, Inc. Finally, a group from the National Center for Toxicological Research, an FDA research institute, presents its software suite ArrayTrack, which is widely applied by FDA researchers and new drugs applications reviewers. ArrayTrack, primarily designed for microarray expression analysis and expanded into other OMICs assays, combines a number of statistical tools with public and commercial pathway tools.

Functional data analysis is evolving quickly as a discipline. Novel network algorithms and software tools are published almost weekly, and the scope of applications expands with every new DNA, RNA, or protein assay hitting the market. Therefore, we could not and had no intention to pack as many tools as possible into this volume. Instead, we tried to focus on the established methods and software packages we see in the marketplace every day and provide readers with a broad understanding of issues and applications of this fascinating new field.

*Yuri Nikolsky  
Julie Bryant*

---

## Contributors

- NATALIA BAGROVA • *Institute for Systems Biology SPb, Moscow, Russia*  
JUDITH BANDY • *Linguamatics Ltd, St John's Innovation Centre, Cambridge, UK*  
JOHN BOYLE • *Institute for Systems Biology, Seattle, WA, USA*  
SVETLANA BUREEVA • *Metalogic OOO, Moscow, Russia*  
GREGORY W. CARTER • *Institute for Systems Biology, Seattle, WA, USA*  
AARON N. CHANG • *Regulus Therapeutics, Carlsbad, CA, USA*  
MINJUN CHEN • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
CAMELIA CHIRIAC • *Pharmacopeia, Cranbury, NJ, USA*  
OLEG DEMIN • *Institute for Systems Biology SPb, Moscow, Russia*  
SREENIVAS DEVIDAS • *GVK BioSciences, Pvt. Ltd., Hyderabad, India*  
ALEKSANDER DORODNOV • *Institute for Systems Biology SPb, Moscow, Russia*  
HONG FANG • *Z-Tech Corporation, an ICF International Company, Jefferson, AR, USA*  
ANTONIO GIORDANO • *Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, USA*  
NAIL GIZZATKULOV • *Institute for Systems Biology SPb, Moscow, Russia*  
EKATERINA GORYACHEVA • *Institute for Systems Biology SPb, Moscow, Russia*  
IGOR GORYANIN • *Edinburgh University, Edinburgh, UK*  
YIKE GUO • *InforSense Ltd., London, UK*  
STEPHEN C. HARRIS • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
KLAUS HEUMANN • *Biomax Informatics AG, Martinsried, Germany*  
TATIANA KARELINA • *Institute for Systems Biology SPb, Moscow, Russia*  
JEFF KIEFER • *Pharmaceutical Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ, USA*  
SARAH KILLCOYNE • *Institute for Systems Biology, Seattle, WA, USA*  
EUGENE KIRILLOV • *Vavilov Institute for General Genetics, Russian Academy of Sciences, Moscow, Russia*  
YURIY KOSINSKY • *Institute for Systems Biology SPb, Moscow, Russia*  
VLADIMIR A. KUZNETSOV • *Bioinformatics Institute, Biopolis, A-STAR, Singapore*  
MICHAEL LIEBMAN • *Strategic Medicine, Inc., Kennett Square, PA, USA*  
SASCHA LOSKO • *Biomax Informatics AG, Martinsried, Germany*  
DAWN R. MATTOON • *Invitrogen Corporation, Protein Array Center, Branford, CT, USA*  
SARAH MCQUAY • *Linguamatics Ltd, Cambridge, UK*  
EUGENIY METELKIN • *Institute for Systems Biology SPb, Moscow, Russia*  
HUAIYU MI • *Evolutionary Systems Biology Group, SRI International, Menlo Park, CA, USA*  
DAVID MILWARD • *Linguamatics Ltd, Cambridge, UK*

- EKATERINA MOGILEVSKAYA • *Institute for Systems Biology SPb, Moscow, Russia*  
SPYRO MOUSSES • *Pharmaceutical Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ, USA*  
ROBIN E. J. MUNRO • *InforSense Ltd., London, UK*  
CHAD L. MYERS • *Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA*  
TATIANA NIKOLSKAYA • *Vavilov Institute for General Genetics, Russian Academy of Sciences, Moscow, Russia*  
YURI NIKOLSKY • *GeneGo, Inc., St. Joseph, MI, USA*  
ROGER PERKINS • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
KIRILL PESKOV • *Institute for Systems Biology SPb, Moscow, Russia*  
TATIANA PLYUSNINA • *Institute for Systems Biology SPb, Moscow, Russia*  
FENG QIAN • *Z-Tech Corporation, an ICF International Company, Jefferson, AR, USA*  
QIANG Q. QUE • *Pharmaceutical Genomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ, USA*  
EUGENE RAKHMATULIN • *GeneGo, Inc., St. Joseph, MI, USA*  
VALENTIN ROMANOV • *Metalogic OOO, Moscow, Russia*  
GIUSEPPE RUSSO • *Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, USA*  
NICHOLAS J. SCHORK • *University of California, San Diego, CA, USA; Scripps Genomic Medicine, Scripps Health and The Scripps Research Institute, San Diego, CA, USA*  
BARRY SCHWEITZER • *Invitrogen Corporation, Protein Array Center, Branford, CT, USA*  
TATIANA SEREBRYISKAYA • *Metalogic OOO, Moscow, Russia*  
LEMING SHI • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
NATHAN O. SIEMERS • *Research & Development, Bristol-Myers Squibb Company, Pennington, NJ, USA*  
SERGEY SMIRNOV • *Institute for Systems Biology SPb, Moscow, Russia*  
JENNIFER SMITH • *Institute for Systems Biology, Seattle, WA, USA*  
ZHENJIANG SU • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
PAUL THOMAS • *Evolutionary Systems Biology Group, SRI International, Menlo Park, CA, USA*  
CHARLES A. TILFORD • *Research & Development, Bristol-Myers Squibb Company, Pennington, NJ, USA*  
WEIDA TONG • *National Center for Toxicological Research (NCTR), FDA, Jefferson, AR, USA*  
ALI TORKAMANI • *University of California, San Diego, CA, USA; Scripps Genomic Medicine, Scripps Health and The Scripps Research Institute, San Diego, CA, USA*  
OLGA G. TROYANSKAYA • *Department of Computer Science, Princeton University, Princeton, NJ, USA*  
DIANA R. WETMORE • *Cystic Fibrosis Foundation Therapeutics, Bethesda, MA, USA*  
JERRY M. WRIGHT • *Johns Hopkins University, Baltimore, MD, USA*

HONGWEI H. YIN • *Pharmaceutical Genomics Division, Translational Genomics  
Research Institute (TGen), Phoenix, AZ, USA*

ROMAN ZUEV • *Vavilov Institute for General Genetics, Russian Academy of Sciences,  
Moscow, Russia*

SVETLANA ZVEREVA • *Metalogic OOO, Moscow, Russia*

# 目 录

前言 .....	v
撰稿人 .....	ix

## 第一部分：相互作用

1. 用 Linguamatics 公司研发的 I2E 软件从发表的文献中挖掘蛋白质相互作用 .....	3
Judith Bandy, David Milward, and Sarah McQuay	
2. 基因组规模实验中转录因子与 DNA 结合的相对亲和力、特异性和敏感度 .....	15
Vladimir A. Kuznetsov	
3. 抑制因子-靶标数据的管理：步骤和在途径分析上的作用 .....	51
Sreenivas Devidas	
4. 用功能蛋白质芯片描绘蛋白质相互作用网络 .....	63
Dawn R. Mattoon and Barry Schweitzer	
5. 蛋白质相互作用的手工注释 .....	75
Svetlana Bureeva, Svetlana Zvereva, Valentin Romanov, and Tatiana Serebryiskaya	

## 第二部分：分析

6. 基因集富集分析 .....	99
Charles A. Tilford and Nathan O. Siemers	
7. PANTHER 途径：一个整合了数据分析工具且基于本体的途径数据库 .....	123
Huaiyu Mi and Paul Thomas	
8. 采用网络分析优化排序影响途径的基因 .....	141
Aaron N. Chang	
9. 从多样的功能基因组数据中发掘生物学网络 .....	157
Chad L. Myers, Camelia Chiriac, and Olga G. Troyanskaya	
10. 在基于知识的集成平台上对组学数据及小分子化合物的函数分析 .....	177
Yuri Nikolsky, Eugene Kirillov, Roman Zuev, Eugene Rakhmatulin, and Tatiana Nikolskaya	
11. 动力学模型作为一种整合多层次动态实验数据的工具 .....	197
Ekaterina Mogilevskaya, Natalia Bagrova, Tatiana Plyusnina, Nail Gizzatkulov, Eugeny Metelkin, Ekaterina Goryacheva, Sergey Smirnov, Yuriy Kosinsky, Aleksander Dorodnov, Kirill Peskov, Tatiana Karelina, Igor Goryanin, and Oleg Demin	
12. Cytoscape：用于网络建模的一个基于社区的框架 .....	219
Sarah Killcoyne, Gregory W. Carter, Jennifer Smith, and John Boyle	
13. 用语义数据集成和知识管理表示生物网络相关性 .....	241
Sascha Losko and Klaus Heumann	

14. 复杂的、多数据类型及多工具分析的解决方案：运用工作流程与流水线方法的原则及应用 ..... 259  
Robin E. J. Munro and Yike Guo

### 第三部分：应用

15. 高通量 siRNA 筛选结合化合物筛选作为一种干扰生物系统以及识别目标途径的方法 ..... 275  
Jeff Kiefer, Hongwei H. Yin, Qiang Q. Que, and Spyro Mousset
16. 用高密度等位基因关联数据进行途径和网络的分析 ..... 289  
Ali Torkamani and Nicholas J. Schork
17. miRNAs：从生物起源到网络 ..... 303  
Giuseppe Russo and Antonio Giordano
18. MetaMiner (CF)：疾病导向的生物信息学分析环境 ..... 353  
Jerry M. Wright, Yuri Nikolsky, Tatiana Serebryiskaya, and Diana R. Wetmore
19. 转化研究与生物医学信息学 ..... 369  
Michael Liebman
20. ArrayTrack：一个美国食品及药物管理局 (FDA) 和公共基因组工具 ..... 379  
Hong Fang, Stephen C. Harris, Zhenjiang Su, Minjun Chen, Feng Qian, Leming Shi, Roger Perkins, and Weida Tong
- 索引 ..... 399

(高友鹤 尹剑锐 译)



---

# Contents

<i>Preface</i> . . . . .	<i>v</i>
<i>Contributors</i> . . . . .	<i>ix</i>
SECTION I: INTERACTIONS	
1. Mining Protein–Protein Interactions from Published Literature Using Linguamatics I2E . . . . . <i>Judith Bandy, David Milward, and Sarah McQuay</i>	3
2. Relative Avidity, Specificity, and Sensitivity of Transcription Factor–DNA Binding in Genome-Scale Experiments . . . . . <i>Vladimir A. Kuznetsov</i>	15
3. Curation of Inhibitor-Target Data: Process and Impact on Pathway Analysis . . . . . <i>Sreenivas Devidas</i>	51
4. Profiling Protein Interaction Networks with Functional Protein Microarrays . . . . . <i>Dawn R. Mattoon and Barry Schweitzer</i>	63
5. Manual Annotation of Protein Interactions . . . . . <i>Svetlana Bureeva, Svetlana Zvereva, Valentin Romanov, and Tatiana Serebryiskaya</i>	75
SECTION II: ANALYSIS	
6. Gene Set Enrichment Analysis . . . . . <i>Charles A. Tilford and Nathan O. Siemers</i>	99
7. PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools . . . . . <i>Huaiyu Mi and Paul Thomas</i>	123
8. Prioritizing Genes for Pathway Impact Using Network Analysis . . . . . <i>Aaron N. Chang</i>	141
9. Discovering Biological Networks from Diverse Functional Genomic Data . . . . . <i>Chad L. Myers, Camelia Chiriac, and Olga G. Troyanskaya</i>	157
10. Functional Analysis of OMICs Data and Small Molecule Compounds in an Integrated “Knowledge-Based” Platform . . . . . <i>Yuri Nikolsky, Eugene Kirillov, Roman Zuev, Eugene Rakhmatulin, and Tatiana Nikolskaya</i>	177
11. Kinetic Modeling as a Tool to Integrate Multilevel Dynamic Experimental Data . . . . . <i>Ekaterina Mogilevskaya, Natalia Bagrova, Tatiana Plyusnina, Nail Gizzatkulov, Eugeniy Metelkin, Ekaterina Goryacheva, Sergey Smirnov, Yuriy Kosinsky, Aleksander Dorodnov, Kirill Peskov, Tatiana Karelina, Igor Goryanin, and Oleg Demin</i>	197