科爱传播
KE'AI COMMUNICATIONS

基因组学、转录组学与代谢组学

本书文章选自 *Trends in Biochemical Sciences, Trends in Biotechnology, Trends in Genetics & Trends in Pharmacological Sciences*

NEW FOCUS in Life Sciences

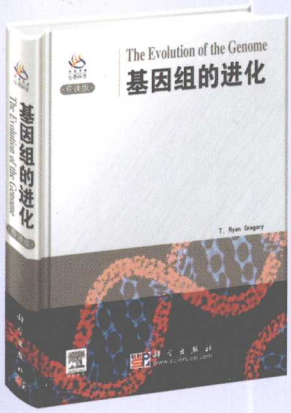
生命科学新视野 1

ELSEVIER
名刊名篇选编

科学出版社
www.sciencep.com

生命科学
未来世界与代谢生物

NEW FOCUS
Life Sciences



定价：96 元
ISBN: 978-7-03-018214-2
出版日期：2007 年 1 月

The Evolution of the Genome

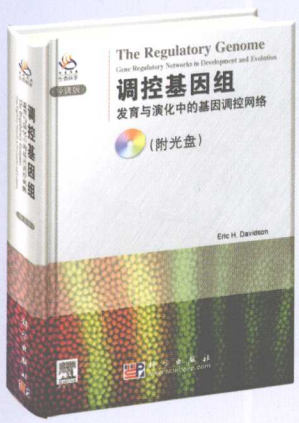
《基因组的进化》（导读版）

原著：T. Ryan Gregory
导读：王文（中国科学院昆明动物研究所）

随着基因组研究技术的迅速发展，我们已经积累了海量的基因组数据，而且未来还会产生更为庞大的基因组数据。如何解读这些海量数据，理解生命的相关本质，成了这一后基因组学时代的一个巨大挑战。基于比较方法的进化基因组学研究无疑将成为解读这些海量数据的一个十分有效的手段。但是，有关基因组进化研究的历史、现状和未来一直都缺乏一个完整的描述，有关其研究内容也十分模糊。

本书通过邀集一批相关领域的学者，包括美国科学院院士 Margaret Kidwell 和亚利桑那大学的著名分子进化学家Sudhir Kumar等著名科学家，首次全貌式地介绍了基因组进化研究的各个方向和领域。内容涉及基因组组成、结构及其进化，基因组进化与一些重要生物学问题的关系，以及通过比较基因组学研究理解基因组结构和功能进化的型式和机制等。

这是一本特别适合那些想了解或进入进化基因组学领域的研究生和博士后的书。但同时对于基因组学领域的专业研究人员和科学家也是一本很好的参考书。



定价：50 元
ISBN：978-7-03-018215-9
出版日期：2007 年 1 月

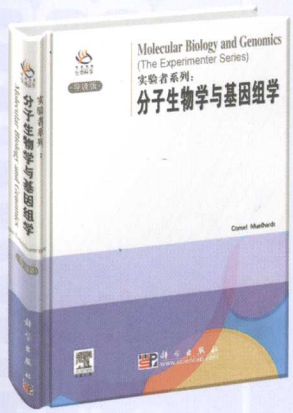
The Regulatory Genome

《调控基因组》（导读版）

原著：Eric H. Davidson
导读：毛炳宇（中国科学院昆明动物研究所）

本书将基因组学、胚胎学和演化生物学的关键概念进行了很好的综合，很好地揭示了动物发育调控中基因调控序列结构与功能的关系，以及由这些调控序列构成的基因调控网络所初步显现的基本性质。本书的目标在于，通过一些无可辩驳的证据，说明发育的动因从根本上讲存在于对基因空间表达的顺式调控之中。而发育只是由很多调控基因组成的调控系统的输出结果。这就为躯体构造的演化机制提供了新的思路。

* 本书作者被誉为“破解发育之谜的真正的先驱之一”。
* 第一本侧重因果关系、以基因组知识系统为基础的教科书。
* 第一次向我们展示了细胞调控的“生物计算机”……在后基因组时代，这是一本划时代的书，标志着原理性的、量化生物学时代的到来。



定价：50 元
ISBN：978-7-03-019389-6
出版日期：2007 年 7 月

The Experimenter Series : Molecular Biology and Genomics

《实验者系列：分子生物学与基因组学》（导读版）

原著：Cornel Muelhardt
导读：刘斌 研究员（中国科学院北京基因组研究所）

目前，实验室人员面临的一大困惑是如何运用正确的工具和方法来阐释基因组所编码的相关蛋白质的结构，功能和表达解释。德国马尔堡大学的Muelhardt所著的分子生物学与基因组学为大家提供了很多有益的实验室小窍门和技巧，能够大大提高实验的成功率和准确性。这本实验室指南简单明快地介绍了分子生物学和基因组学中所涉及的各种现代方法，和各种常规方法，并讨论每种方法的优点和不足。

作为“实验者”系列中的一本，本书秉承了丛书一贯的风格，依然是一本非常实用的实验室手册，以简捷明快的风格，用 100 多个图表，为广大的分子生物学和基因组学实验室工作人员提供了大量的实用小贴士来补充和完善基础知识，从而大大提高实验的准确性和成功率。

本书适用于分子生物学和基因组学领域的实验室工作人员，科研工作者，教师和高年级本科生，研究生使用。

New Focus in Life Sciences

基因组学、转录组学与代谢组学

1

生命科学新视野

科　爱　传　播
KE'AI COMMUNICATIONS

图字：01-2007-3635 号

图书在版编目（CIP）数据

基因组学、转录组学与代谢组学＝Genomics,trans-criptomics and metabolomics：英文/（意）斯卡瑟里（Scarselli，M.）等编著. —北京：科学出版社，2007
（生命科学新视野：1）
ISBN 978-7-03-019491-6

Ⅰ.基… Ⅱ.斯… Ⅲ.①基因组－研究－英文②基因－转录（分子生物学）－研究－英文③代谢－研究－英文
Ⅳ.Q343.1 Q753 Q493.1

中国版本图书馆 CIP 数据核字（2007）第 116628 号

责任编辑：田慎鹏 贾明月
责任印制：钱玉芬
地　　址：北京市东城区东黄城根北街 16 号 1-515
邮政编码：100717
电子邮箱：keai@mail.sciencep.com
网　　址：http://www.kbooks.cn

科学出版社 出版
双青印刷厂印刷
北新华文设计制作

定　价：50.00 元

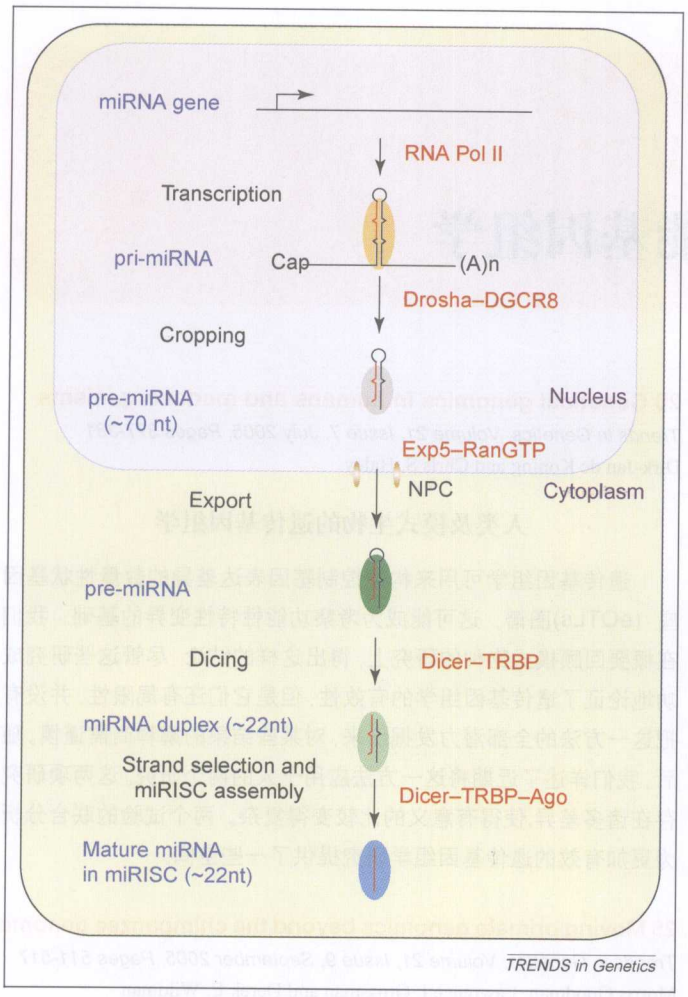
代谢组学

Source journals:

Trends in Biotechnology	ISSN 0167-7799
Trends in Pharmacological Sciences	ISSN 0165-6147
Trends in Genetics	ISSN 0168-9525
*Trends in Biochemical Sciences***	ISSN 0968-0004

***Trends in Biochemical Sciences* (TIBS) is an official publication of the International Union of Biochemistry and Molecular Biology (IUBMB) and is published by Elsevier Limited.

基因组学
——比较／功能基因组学



同时也回顾了有关miRNA基因结构与生物起源机制的知识。本文的重点是动物的miRNA。

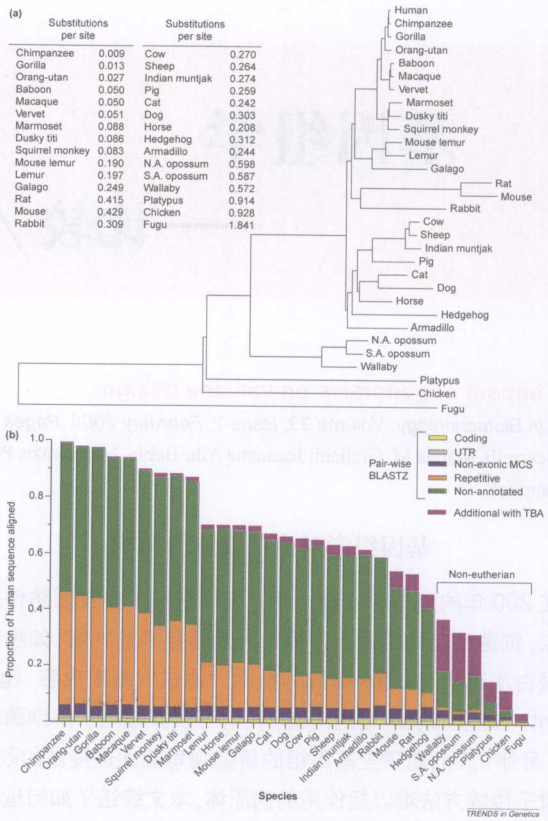
41 Differences between pair-wise and multi-sequence alignment methods affect vertebrate genome comparisons

Trends in Genetics, Volume 22, Issue 4, April 2006, Pages 187-193

Elliott H. Margulies, Christina W. Chen and Eric D. Green

影响脊椎动物基因组比较的成对及多重序列比对方法间的差异

对多个基因组序列进行完整而精确的比对是复杂而容易出错的过程，对来源于高度分化物种的序列比较而言尤为如此。在本文中我们证明，在比对(或“捕获”)系统发生上距离较远的（即由相对较长的进化分支距离所分隔的）的脊椎动物的所有可得直系同源序列时，多重序列比对的方法要远优于成对序列对比的方法。只有在多重序列比对时，才可得到最多的比对结果。这样的多重序列比对中包含大量的在成对序列比对中无法获得的外显子和高度保守的非外显子序列，从而说明序列比对方法在比较基因组研究中的重要性。



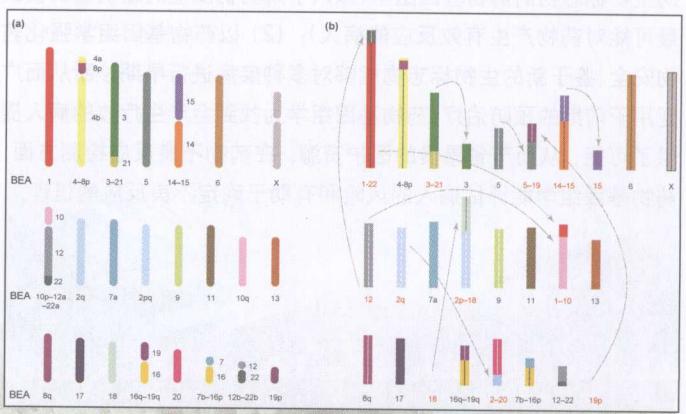
48 Dissecting the mammalian genome - new insights into chromosomal evolution

Trends in Genetics, Volume 22, Issue 6, June 2006, Pages 297-301

Terence J. Robinson, Aurora Ruiz-Herrera and Lutz Froenicke

解析哺乳动物的基因组——染色体进化的新视点

一项比对了八种哺乳动物基因组数据的最新研究，提供了某些染色体结构全新的详实信息，这些染色体包含了真哺乳动物的祖先的染色体组型。分析认为，进化的突破点聚集在富含着丝点的“热点区域”，而那些更普遍出现的人类癌症相关的突破点倾向于和进化突破点共存。



53 Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*

Trends in Genetics, Volume 22, Issue 11, November 2006, Pages 597-602

Andrew H. Paterson, Brad A. Chapman, Jessica C. Kissinger, John E. Bowers, Frank A. Feltus and James C. Estill

在拟南芥、水稻、酵母与河豚中的独立全基因组复制事件后，多种基因和功能区基因家族具有收敛性

基因组复制是产生新基因的潜在的良好资源，但新基因的产生需要长时间的进化。我们已经发现了一组“抗复制”基因，这些基因在多次独立的基因组复制事件后，经过收敛性的还原，已经恢复到单基因的状态。“抗复制”基因到单基因状态的恢复对于多倍体世系的长期延续是很重要的。被子植物表现出比其它古老多倍体植物更频繁的多倍化和更高程度的复制基因的保留，使得其成为进一步更好研究抗复制基因的合适对象。

59 Did brain-specific genes evolve faster in humans than in chimpanzees?

Trends in Genetics, Volume 22, Issue 11, November 2006, Pages 608-613

Peng Shi, Margaret A. Bakewell and Jianzhi Zhang

脑特异基因在人类中的进化比在黑猩猩中快吗?

脑的容量、结构与机能是人类区别于其它灵长类最显著的特征之一。最近的一项研究指出，在人类起源的过程中，神经系统相关的基因存在着广泛的序列加速进化的现象。为检验这种假说，我们对那些主要或特异性的在脑中表达的基因，以及对脑的功能起重要作用的基因进行了全基因组范围的分析，后者是通过脑特异性的五种不同特征而确定的，虽然这五类脑特异性的基因之间很少有重叠，但其中任何一个都不支持人类加速进化这个假说。相反的，有些数据表明，相比于其它基因，人类脑特异性基因的非同义替换率比猩猩要低得多。我们的结果证明，人脑的独有特征并不是由于许多蛋白质的大量氨基酸适应性突变而产生的。

65 Plant bioinformatics: from genome to phenome

Trends in Biotechnology, Volume 22, Issue 5, 1 May 2004, Pages 232-237

David Edwards and Jacqueline Batley

植物生物信息学：从基因组到表型组

随着生物技术的进步，产生了大量多样的生物学信息，这些海量的信息也导致了生物信息学领域的发展和进化。这一相对新的领域既促进了基因组和后基因组数据的分析，又加速了来自转录物组学、蛋白质组学、代谢物组学和表型组学等相关领域信息的整合。这样的整合使得人们能够鉴定基因和基因产物，并解释基因型和表型之间的功能关系，从而使由基因组到表型组的全系统分析成为可能。随着植物生物技术价值和通量的增加，正有人号召利用生物信息学对不断扩张的“组学”技术得到的各种数据进行整合。

转录组学与微阵列

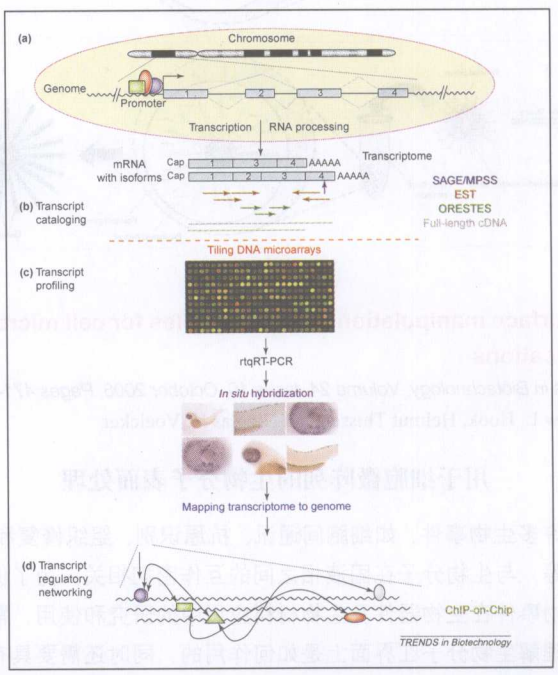
71 Interrogating the transcriptome

Trends in Biotechnology, Volume 22, Issue 1, January 2004, Pages 23-30

Yijun Ruan, Pierre Le Ber, Huck Hui Ng and Edison T. Liu

转录组的整合研究

人类和多个模式生物的基因组序列为控制生物性状特征的基础遗传和进化模板提供了一些重要的观点。然而决定生物表型真正的效应因子是重要的下游信息元件，这些元件首先发现于转录组中，随后在蛋白质组中也有发现。因此有必要设计合理的实验方案来充分了解转录组的特征。这是一项极具复杂性的挑战，因为细胞分化发育的过程中，成千上万的基因以不同的水平表达，且具有不同的转录本。本文就转录组特征研究的近期主要技术进展作一综述，由于内容广泛，我们特别集中评述转录本的研究方法而不是运用电脑虚拟方法确定转录本，我们还对转录网络进行了简述。



79 Microarrays - status and prospects

Trends in Biotechnology, Volume 22, Issue 12, December 2004, Pages 630-637

Srivatsa Venkatasubbarao

微阵列的现状和前景

微阵列已经成为生命科学研究人员极其重要的研究工具，且已开始在诊断，治疗和监测等方面应用。本文对原位合成微阵列、显微点样微阵列、微珠阵列、流式微阵列、光导纤维阵列和纳米条形码等进行了详细描述，同时评述了微阵列的发展以及诊断微阵列的研发所面临的挑战和问题。表中附有各种微阵列的销售商信息，还介绍了商业化销售的预做微阵列、微珠阵列以及发展中的特殊微阵列产品。

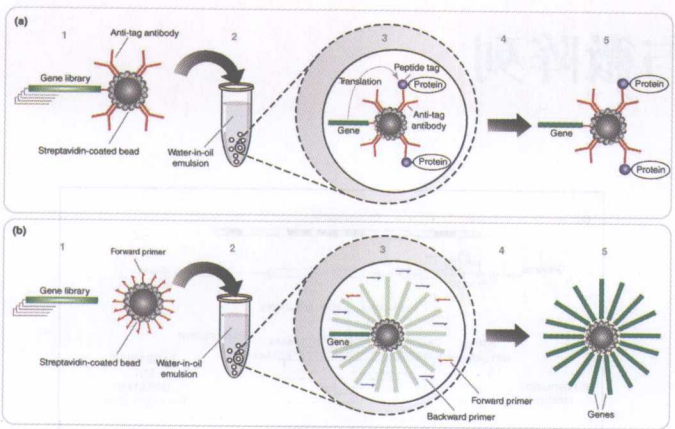
87 Miniaturising the laboratory in emulsion droplets

Trends in Biotechnology, Volume 24, Issue 9, September 2006, Pages 395-402

Andrew D. Griffiths and Dan S. Tawfik

乳滴中的微型实验室

活细胞中的分区使得生化和遗传分析既可微型化又可同时进行。体外分区分析（IVC）在油包水乳液中形成显微水滴空间分区，为分相反应提供了又一策略。这些微小空间仅有细胞大小的容量（可小于10^{-15}升），能够自由的确定和调节其中的成分以及分区数量的多少（每毫升溶液中可高达10^{10}个乳滴），这为新的超高通量的无细胞技术提供了基础。本文描述了蛋白质和RNAs的体外进化、无细胞克隆及测序、遗传学、基因组学和蛋白组学等研究领域中IVC的应用范围和潜力。



95 Surface manipulation of biomolecules for cell microarray applications

Trends in Biotechnology, Volume 24, Issue 10, October 2006, Pages 471-477

Andrew L. Hook, Helmut Thissen and Nicolas H. Voelcker

用于细胞微阵列的生物分子表面处理

许多生物事件，如细胞间通讯、抗原识别、组织修复和DNA转运等，与生物分子在固液相之间的互作密切相关。为了促进这些生物事件在生物器件及生物材料应用中的研究和使用，需要正确的理解生物分子在界面上是如何作用的，同时还需要具有在时间和空间区间上操作界面上生物分子的能力。对细胞微阵列而言尤其如此，此时细胞内各种生命过程必须严格控制以使这些器件的效率和产出达到最大化。尤其有趣的是转染细胞的微阵列（TCMs），这种技术能够在活细胞中高通量的分析基因功能，极大地拓宽了微阵列基因组分析的领域。本文对该领域的近期研究做一综述，讨论了对DNA、蛋白质和其他生物分子进行时空表面处理的基础和应用研究，及其对TCMs的意义。

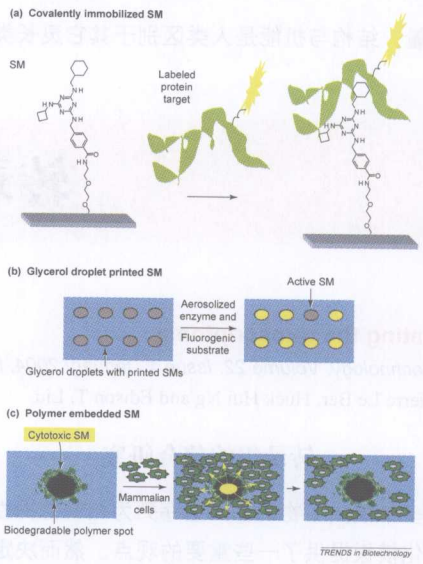
102 Small molecule microarrays: from proteins to mammalian cells - are we there yet?

Trends in Biotechnology, Volume 23, Issue 6, June 2005, Pages 271-274

Gabriela Chiosis and Jeffrey L. Brodsky

小分子微阵列：从蛋白质到哺乳动物细胞——离目标还有多远？

Stockwell及其同事最新的文章证实了小分子微阵列技术（SMMs）的飞跃发展。通过在生物可降解聚合体中注入小分子形成微阵列，作者首次证明SMMs能够被用于细胞水平的研究。这项技术的改善为使用SMMs在哺乳动物细胞中进行高通量筛选提供了方法。



106 Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping

Trends in Genetics, Volume 21, Issue 8, August 2005, Pages 466-475

Thomas E. Royce, Joel S. Rozowsky, Paul Bertone, Manoj Samanta, Viktor Stolc, Sherman Weissman, Michael Snyder and Mark Gerstein

寡核苷酸嵌合微阵列分析转录本的相关问题

传统微阵列使用与已知基因互补的探针来量化在两种或更多情况中差异基因的表达。与传统基因表达微阵列不同，嵌合微阵列的探针选择是基于全基因组序列，按一定间隔选择确定转录本的有无。这意味着偏差的均一，使用中与实验方法无关。本文介绍了嵌合微阵列分析中研究人员所面临的信息学挑战，同时描述了该项新技术分析的初步方法。

116 Evidence that functional transcription units cover at least half of the human genome

Trends in Genetics, Volume 20, Issue 5, 1 May 2004, Pages 229-232

Marie Sémon and Laurent Duret

功能性转录单元至少占人类基因组一半的有关证据

转录组分析发现，大部分人类基因组是转录的。然而，其中许多转录本可能是无功能的。为了将功能性转录单元（FTUs）和假转录本区分开，我们寻找了抗突变选择压力的标志，这些突变会降低转录水平。我们分析了转座子的分布，其在FTUs中是反向选择的。我们发现这些特点足以预测一段序列是否转录，如果转录的话其方向如何。我们的结果显示，FTUs至少组成基因组内涵的50%，其中大约三分之一的转录本不编码蛋白质。

代 谢 组 学

120 Metabolomics by numbers: acquiring and understanding global metabolite data

Trends in Biotechnology, Volume 22, Issue 5, 1 May 2004, Pages 245-252

Royston Goodacre, Seetharaman Vaidyanathan, Warwick B. Dunn, George G. Harrigan and Douglas B. Kell

数字中的代谢物组学：获取和理解整体代谢物数据

在后基因组时代，为了验证药物治疗的潜在靶标、发现疾病新的生物标志，给孤独基因进行功能定位已成为一种特殊需求。代谢组学是一个新兴的领域，是其它“组学”的补充，并且正展现出独特的优势。同转录组学或者蛋白组学一样，典型的代谢指纹或者代谢组学实验可能产生数千个数据点，其中可能仅仅少部分对所研究的问题有所帮助。因而，从这些数据中提炼出最有意义的的元素，就成为产生有机制性或阐释性作用的新知识的关键步骤。

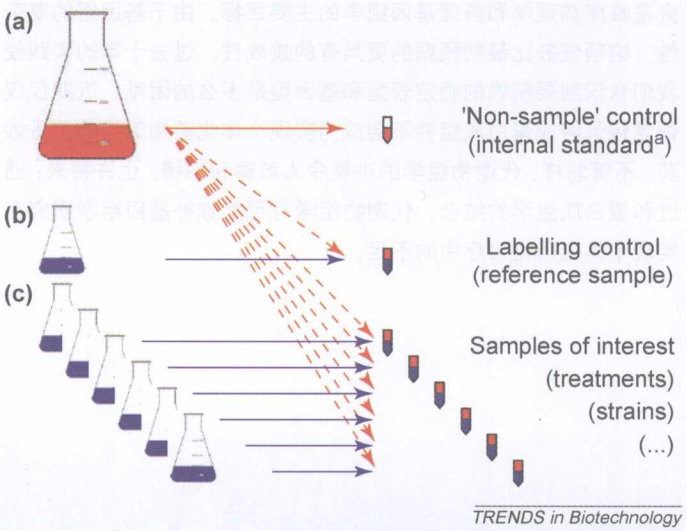
128 Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling

Trends in Biotechnology, Volume 23, Issue 1, January 2005, Pages 28-33

Claudia Birkemeyer, Alexander Luedemann, Cornelia Wagner, Alexander Erban and Joachim Kopka

代谢物组学分析：稳定同位素体内标记在代谢物谱中的应用潜力

代谢物组学分析技术还处于早期发展阶段，与基因组、转录组和蛋白质组分析不同，代谢物组学分析需要处理一个高度变化的范畴中的生物分子。因此，不同分析平台的结合对全面进行代谢物组学研究是必须的。每个平台只包含代谢物组学的一个部分。为了建立多并行的技术，有必要使每一种测量的代谢物全面标准化。标准化的最佳方式是对每种代谢物添加特殊的稳定同位素标记的化合物，同位化合物。初一看，这个建议由于时间和成本的关系似乎有点不切实际。解决这个问题的可能办法在这篇文章得以论述。使用稳定同位素进行体内饱和标记，可生物合成不同质量标记的代谢产物混合物，这样可以通过不同质量同位化合物的比率来研究高度标准化的代谢物谱。



134 The next wave in metabolome analysis

Trends in Biotechnology, Volume 23, Issue 11, November 2005, Pages 544-546

Jens Nielsen and Stephen Oliver

代谢物组学分析的下一个浪潮

一个细胞的代谢物组学表现了其它功能基因组水平（例如转录组和蛋白质组）信号的扩增和整合。尽管这使代谢物组学成为表型高通量分析的一个有力工具，但是由于和基因组缺乏直接关

联，导致代谢物组学在数据解释上出现困难。然而，功能基因组学中已有例子说明，代谢物组学在阐释其它情况下沉默的突变表型上有所作用。尽管有成功的例证，我们相信未来的代谢组学研究仍然要聚焦在已知代谢物浓度的精确测量上。代谢物组学的研究团体必须开发出典型培养条件细胞中代谢物浓度的数据库，以便代谢物组学数据能够与其它功能基因组学数据完全整合，并为系统生物学的发展作出贡献。

137 Can personalized drug therapy be achieved? A closer look at pharmaco-metabonomics

Trends in Pharmacological Sciences, Volume 27, Issue 11, November 2006, Pages 580-586

Daniel W. Nebert and Elliot S. Vesell

个体化药物治疗能否实现——近观药物代谢组学

在1930年和1990年之间，数十个高外显率，主要是单基因的疾病得以发现并描述，使得一些科学家预测个体化药物治疗将为时不远。DNA检测被用于确定个体对毒性和癌症的遗传易感性，由此来鉴别某种药物可能有效的个体以及药物毒性风险会增加的个体。这些实验体现了表型-基因型关联性研究的前沿成果，该研究是临床药理学和药理基因组学的主要目标。由于基因组的复杂性，这项任务比最初预期的更具有的挑战性。过去十年的实践使我们意识到要精确的确定表型和基因型是多么的困难。近期仅仅依靠转录组或蛋白质组并不会成为实现个体化药物治疗的灵丹妙药。不管怎样，代谢物组学的进展令人鼓舞和期待。也许将来，通过和蛋白质组学的结合，代谢物组学有可能弥补基因组学研究在实现个体化药物治疗中的不足。

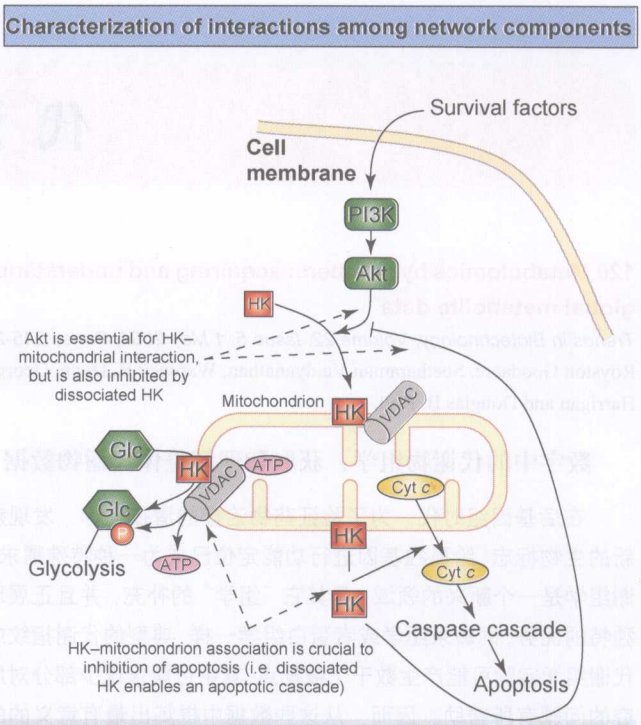
144 Systems analyses characterize integrated functions of biochemical networks

Trends in Biochemical Sciences, Volume 31, Issue 5, May 2006, Pages 284-291

Erwin P. Gianchandani, David L. Brautigan and Jason A. Papin

系统分析鉴定生化网络的整体功能

在过去的一个世纪中，人们已经对代谢、调节和信号传送途径作出了细节描述。随着基因组、蛋白质组和代谢物组的数据逐渐增多，研究这些数据的数学和分析能力也在向前发展，人们发现了途径组分的叠加作用。这些发展都反映了亚细胞生化网络的整合性质。包括计量网络重组在内的多系统分析，为网络成分的代谢、调节和信号传送功能中的量化叠加提供了一组关键的工具。这说明此种整合对于精确地描述生化网络的功能是必不可少的。



The impact of genomics on vaccine design

Maria Scarselli, Marzia M. Giuliani, Jeannette Adu-Bobie, Mariagrazia Pizza and Rino Rappuoli

Chiron Vaccines, Via Fiorentina 1, 53100 Siena, Italy

After 200 years of practice, vaccinology has gained new perspectives for preventing infectious diseases. Sequencing of complete bacterial genomes led to the development of new large-scale technologies, such as bioinformatics, proteomics and DNA microarrays. By examining genetic content, as well as transcription and expression profiles, a more detailed understanding of bacterial pathogenesis can be reached. Moreover, the whole-genome perspective is expected to provide an instrumental contribution to vaccine development, particularly to target those pathogens for which the traditional approaches have failed so far. In this review, we describe how genomic approaches can be used to identify novel vaccine candidates or create safer live-attenuated vaccines.

Over the years, immunization has experienced several milestones, from the early attempts of 'variolation' to modern, genetically engineered vaccines. Since its introduction more than 200 years ago, vaccination has prevented illness and death for millions of individuals every year. However, despite the success of vaccinology, infectious diseases are still the leading cause of death worldwide. All vaccines available are based on killed or live-attenuated microorganisms, on toxins detoxified by chemical treatment, on purified antigens and on polysaccharide conjugated to proteins (**Figure 1**). The knowledge of the pathogenesis of microorganisms, the identification of the main virulence factors and the characterization of the immune response after infection has been fundamental for the design of second-generation vaccines based mainly on highly purified antigenic components [1].

Recently, there have been two major innovations in vaccine design. The first was the advent of modern molecular biology techniques. These techniques gave rise to two efficacious recombinant vaccines: the hepatitis B vaccine, which is based on a highly purified envelope protein [2] and that against *Bordetella pertussis*, based on three highly purified proteins [3]. The latter also pioneered the use of structure–function studies to produce a genetically detoxified toxin that maintained an unaltered antigenic conformation [4].

Until recently, vaccine design was based on conventional approaches using biochemical, microbiological and serological methods to identify single antigen components that can be produced either in pure form from the pathogen cultivated in laboratory conditions or by using recombinant DNA technology. The antigens are then tested for their ability to induce an immune response. This approach can take years or decades, and in many cases fails to identify protective antigens that are relevant *in vivo* during infection. The approach needs the pathogen to be grown *in vitro* and therefore is not applicable to noncultivable organisms.

The second revolution in vaccine design is more recent and is a result of the use of genomic technology. Sequencing of the first bacterial genome *Haemophilus influenzae* marked the beginning of a 'genomic era', changing the landscape of modern biology and leading to a new approach in vaccine design. So far, ~172 complete genomes of pathogens and nonpathogens are available and more than 400 other microorganisms are being sequenced [Genomes OnLine database (GOLD) at **http://wit.integratedgenomics.com/GOLD/**]. It is now possible to determine the complete genome sequence of a pathogen in a short period of time (months) at low cost. Genomic information can then be used to screen the inclusive set of proteins potentially encoded by a microorganism, in search of potential vaccine candidates – an approach known as reverse vaccinology [5].

Moreover, is it now possible to compare the sequences of related bacteria and pathogens against commensals of the same or related species, and even bacteria with different or similar pathogenic profiles, identifying putatively disease-related genes (comparative genomics).

Functional genomics approaches are complementary to *in silico* antigen discovery. These include the large-scale analysis of gene transcription by DNA microarrays, the identification of the whole set of proteins encoded by an organism (proteomics) by two-dimensional gel electrophoresis and mass spectrometry, as well as the comparative genome–proteome technologies.

Reverse vaccinology

The first example in which genomic technology was used for the identification of potential vaccine candidates is the vaccine against *Neisseria meningitidis* serogroup B [6], the major cause of sepsis and meningitis in children and young adults. For several decades, studies using conventional approaches to vaccine research failed to provide an effective and universal vaccine against meningococci B

Corresponding author: Rappuoli, R. (rino_rappuoli@chiron.com).
Available online 18 December 2004

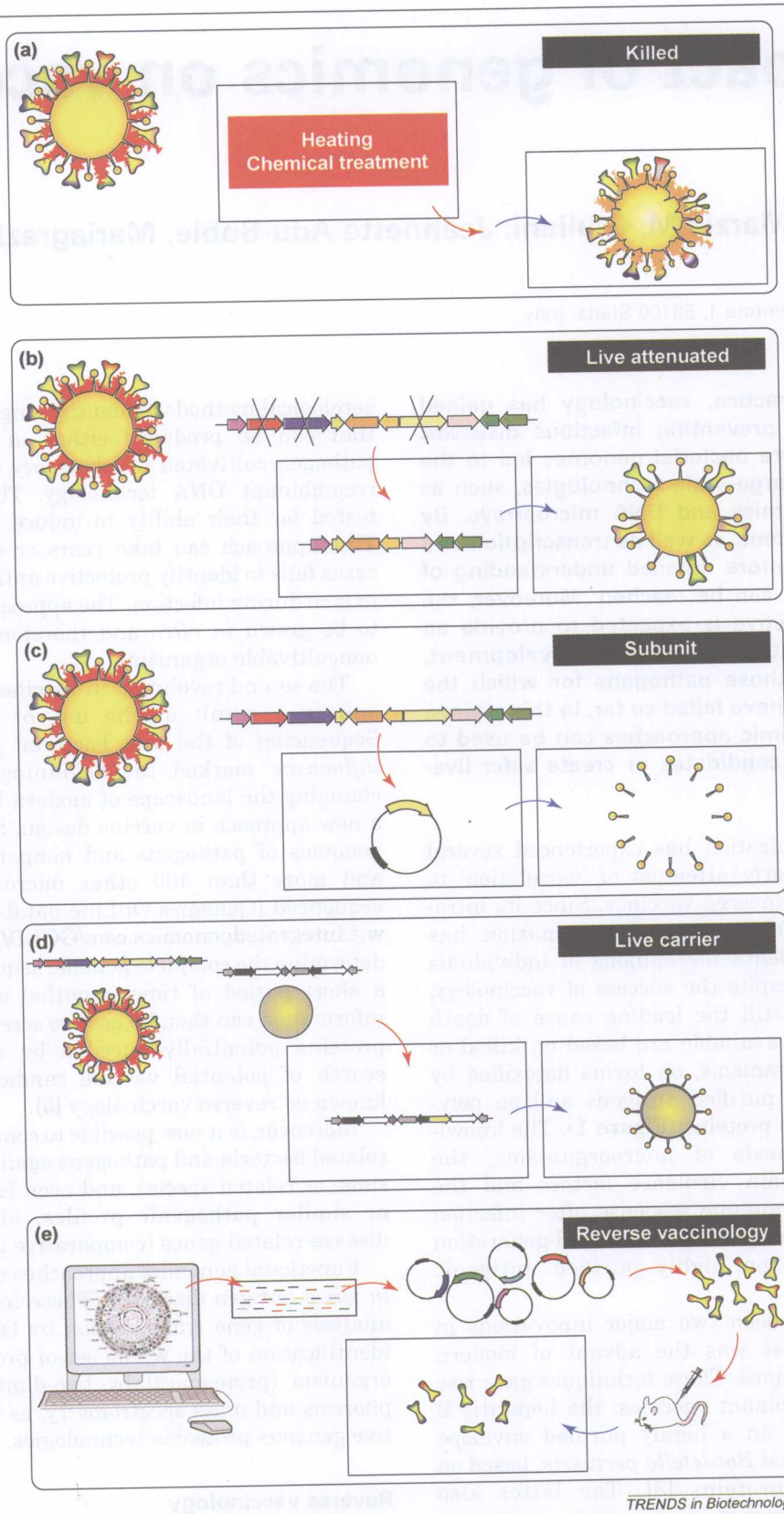


Figure 1. Different approaches to vaccine design. The development of vaccines has been the consequence of the continuous introduction of a series of new technologies. Progress made in molecular immunology and biotechnology in the past two decades has led to a gradual shift from the classical whole-cell vaccines consisting of killed pathogens (**a**) to subunit vaccines (**c**,**e**). Application of DNA recombinant technology has resulted in the production of live attenuated vaccines (**b**) by inactivating genes responsible for pathogenicity. DNA manipulation has also revolutionized the ability to clone, purify and engineer subunit antigens to remove toxic effects while retaining immunogenicity (**c**). Moreover, attenuated or nonpathogenic bacteria and viruses can be engineered to express foreign protein antigens (**d**). The possibility of determining the complete genome sequence of bacteria marked the beginning of a 'genomic era' that opened new perspectives for vaccine design. The entire set of potential antigens can be identified by the analysis *in silico* of the genome sequence. Then potential antigens are cloned, purified and subjected to immunological screening. The whole procedure leads to the identification of a restricted number of vaccine candidates. This approach is called reverse vaccinology (**e**).

(MenB). With the advent of genomics, in collaboration with The Institute of Genomic Research (TIGR; **www.tigr.org**), the complete genome of the virulent strain MC58 was sequenced [7]. Within 18 months of beginning the sequencing, more than 600 potential vaccine candidates had been predicted using computer analysis, and 350 of them were expressed in *Escherichia coli*, purified and used to immunize mice [6]. The antisera obtained were tested using enzyme-linked immunosorbent assay (ELISA) fluorescent-activated cell sorting (FACS) techniques to evaluate the surface localization of the antigens in MenB. In addition, the antisera were tested for bactericidal activity, a property known to correlate with protection in humans. Ninety-one novel surface-exposed antigens were identified, 29 of which induced bactericidal antibodies. Moreover, sequence variability of the potential vaccine candidates was evaluated among multiple isolates and serogroups of *N. meningitidis*, three strains of *Neisseria gonorrhoeae*, as well as one strain each of *Neisseria cinerea* and *Neisseria lactamica*. The antigens identified using the genome approach are different from those identified using conventional approaches. Many of the new antigens include surface-exposed proteins or lipoproteins with a globular structure and without membrane-spanning domains, and many of them are not abundant on the bacterial surface. The approach of reverse vaccinology has identified more potential vaccine candidates than have been identified over the past 40 years. These promising vaccine candidates, able to induce wide strain coverage, are currently under evaluation and have entered into development. In a subsequent study, Grifantini *et al.* [8], using DNA microarray technology, identified several novel MenB antigens that had not been identified by the reverse vaccinology technique. This study therefore showed that DNA microarrays can be used to identify vaccine candidates and complement other genome-mining strategies.

Following the success of MenB, other groups have used this approach to identify vaccine candidates against the main human pathogens. Pathogens that have been studied using reverse vaccinology include *Bacillus anthracis* [9], *Streptococcus pneumoniae* [10], *Staphylococcus aureus* [11,12], *Chlamydia pneumoniae* [13], *Porphyromonas gingivalis* [14], *Edwardsiella tarda* [15] and *Mycobacterium tuberculosis* [16]. It is worth noting that DNA technology is a powerful tool that facilitates the discovery of previously unknown proteins. However, genome projects risk becoming a vast library of information and it is important to characterize these novel antigens not only as vaccine candidates but also in terms of understanding their roles and functions.

In our efforts to discover surface-exposed proteins of *N. meningitidis* to exploit for vaccine development, we identified and characterized, from the biochemical and functional point of view, three meningococcal adhesins (Adhesion and penetration protein App [17], *Neisseria* adhesin NadA [18] and *Neisseria* Hia/Hsf homolog NhhA [19,20]), two novel lipoproteins (Genome-derived *Neisseria* antigens GNA1870 [21] and GNA33 [22]) and the ADP-ribosyltransferase NarE [23].

App is a member of the autotransporter family homolog to the *Haemophilus* adhesion and penetration protein (Hap) of *H. influenzae* [24]. The recombinant protein has serine protease activity and is able to bind to human epithelial cells. Deletion of the *app* gene in a virulent MenB strain significantly reduces its adherence capability compared with the wild type, suggesting that App is an adhesin, which might have a role in *N. meningitidis* colonization of the nasopharyngeal mucosa [17].

NadA is a surface protein of *N. meningitidis* and is homologous to YadA of enteropathogenic *Yersinia* and to UspA2 of *Moraxella catarrhalis* [25]. The NadA gene was found in strains belonging to many hypervirulent lineages but was absent in *N. gonorrhoeae* and in the commensal species *N. lactamica* and *N. cinerea*. The protein forms high molecular weight oligomers on the surface of *N. meningitidis* and mediates bacterial binding to epithelial cells, suggesting that it might be involved in adhesion [18].

NhhA, a protein showing sequence similarity to Hia and Hsf adhesins of *H. influenzae*, is also likely to have a role in the cell adhesion process. These three proteins present a common modular organization and contain a conserved amino acid motif which is also present in human proteins belonging to the family of cell adhesion molecules (CAMs). The analogy between NhhA and CAMs could suggest that NhhA might mediate the adhesion of meningococcus to human cells by mimicking the cell–cell interaction mechanisms [19].

GNA1870, a surface-exposed protein, is considered to be a top candidate for the development of a new vaccine against meningococcus, being able to induce protective immunity in animals [21,26]. GNA33 is a highly conserved lipoprotein, homologous to a membrane-bound lytic transglycosylase (MltA) of *E. coli*, and its lytic transglycosylase and muramidase activity has been demonstrated [27], indicating that GNA33 is involved in the degradation of both insoluble murein sacculi and unsubstituted glycan strands. The role of GNA33 in pathogenesis and virulence has been further investigated by generating a knockout mutant in *N. meningitidis* serogroup B. The mutant showed retarded growth and altered morphology *in vitro*, as well as being unable to cause bacteremia in the infant rat model [22]. These results showed that GNA33 has an important role in peptidoglycan metabolism, cell separation and virulence.

A profile-based computational approach, assisted by secondary structure predictions, has been the basis for the identification of NarE, a surface ADP-ribosyltransferase in *N. meningitidis*, with structural analogies to *E. coli* heat-labile enterotoxin and cholera toxin [23]. Bacterial ADP-ribosylating exotoxins exert their toxic effects by transferring the ADP-ribose moiety of NAD onto specific eukaryotic target proteins. The possible membership of NarE to this class of virulence factors makes this protein a potentially valuable target for antibiotic therapy as well as an immunogen for vaccine development.

Comparative genomics

In silico whole-genome analysis has the potential to provide the basis for the complete understanding of the

genetics, biochemistry, physiology and pathogenesis of microorganisms.

As the number of complete genome sequences increases, it becomes possible to compare a significant number of genomes from bacteria belonging to closely related species. In particular, the analysis of the genetic variability between pathogens and closely related non-pathogenic microorganisms leads to the rapid identification of the complete set of genes potentially responsible for acquisition of virulence and has two main practical implications in vaccine discovery.

Firstly, it offers a valuable guideline into the search for suitable proteins to use as purified antigens in subunit-based vaccines. Secondly, it can provide the rational basis for a safe and stable attenuation of live vaccine candidates or vectors for vaccine delivery.

The virulence of many pathogens often correlates with the presence of DNA tracts encoding disease-related factors, the so-called pathogenicity islands (PAIs), usually acquired by genetic horizontal transfer (GHT) and absent from nonpathogenic species [28].

Although GHT is a ubiquitous phenomenon common to prokaryotic and eukaryotic phyla, the frequency, mechanisms and extent of GHT vary widely among different microorganisms [29]. Enterobacteriaceae and group A Streptococci (GAS) both represent examples of bacteria subjected to extensive GHT.

Comparison of the genome of pathogenic strains of *E. coli* with the laboratory strain K-12 MG1655 showed evidence of a mosaic-like structure in these genomes, in which the synteny of the common core chromosome is broken by multiple segments of apparently introgressed DNA-carrying directories of virulence-associated genes. Different combinations of these acquired 'islands' confer to each type of *E. coli* its characteristic lifestyle and pathogenic potential [30–32].

Group A Streptococci, human pathogens causing pharyngitis, cellulites, sepsis, necrotizing fasciitis and acute rheumatic fever, represents perhaps the most prominent example of the role of phages in determining genetic variability [33]. Complete GAS genome sequences determined so far detected the presence of a phage population heterogeneous in number, gene content and chromosomal integration position [34–37]. Almost all of the GAS phages contain proven or putative virulence factors, supporting the hypothesis that the mechanism by which variation in virulence is generated in natural GAS populations is via recombination events involving phage-borne toxin genes.

By contrast, in the case of *Listeria* [38], *Corynebacterium* [39], *Bacillus* [40–43], *Clostridium* [44] and *Pseudomonas* [45] genera, comparative genome analysis detected a lower degree of genetic variability between pathogens and avirulent variants, which nevertheless still enabled the identification of strain-specific genes responsible for the pathogenic lifestyle.

The association between GHT and the acquisition of virulence is valuable information for therapeutic intervention. In particular, proteins encoded by disease-related genes form a promising class of vaccine candidates, which underlines the usefulness of comparative genome analyses to tackle outbreaks generated by the emergence of new invasive microorganisms that have recently expanded their offensive arsenal of genes.

Although DNA acquisition is often the main cause of inter- and intraspecies genetic differences, it is not the only determinant of pathogenicity: the genome of *Campylobacter jejuni*, the causative agent of Guillain–Barré syndrome, presents few indications of GHT [46], in contrast to the closely related gastric pathogen *Helicobacter pylori* [47].

The uptake of exogenous DNA is typically a rare event in the case intracellular bacteria such as Chlamydiaceae [41,48,49] and *Mycobacterium* spp. [50]. The inherent isolation of intracellular microorganisms probably provides little chance of horizontal transfer and seems to enhance the tendency to genome reduction as a consequence of adaptation to the host cell environment. In principle, this higher stability of intracellular bacterial genomes enhances the possibility of identifying protein targets for broad-spectrum drugs and vaccines.

Genome sequence analysis also provides crucial insights into the evolution of bacterial pathogenesis, enabling us to decipher a series of different mechanisms evolved by microorganisms to alter their antigenic appearance: phase variation [51], gene duplication [52–54] and tendency to loss of rudimental functions [55,56].

The discovery of such a panoply of bacterial strategies devoted to shielding crucial proteins from attack by the immune system emphasizes that a good vaccine candidate has to be conserved and stably maintained within the natural population of the target pathogen.

Recently, increasing emphasis has been placed on genome sequencing of closely related organisms [57], which can, in principle, be used to monitor gene content variability within a single species. In particular, some examples of genome comparison between different clinical isolates have been reported [58–60]. Interestingly, these studies have contributed to a better definition (and in some cases a revision) of the profile of genetic variability of these pathogens. The comparative genome analysis of *H. pylori* strain J99, isolated in the USA in 1994 from a patient with a duodenal ulcer, and strain 26695, recovered before 1987 in the UK from a patient with gastritis, revealed a limited number of strain-specific genes and an overall genome organization quite similar between the two strains [58]. These results, somewhat unexpected in a pathogen previously believed to be subject to extensive genomic and allelic diversity, suggested that, in addition to differences in bacterial gene expression, human host factors might also have a significant and perhaps previously underestimated role in determining the severity of *H. pylori* infection.

By contrast, from a comparison of the genome sequence of *M. tuberculosis* clinical strain CDC1551 with the laboratory-adapted H37Rv emerged a more extensive variability than had been initially anticipated [59]. Such divergence could have important implications for pathogenesis and immunity, as it is likely that these variable proteins are involved in interactions with the host.

The case of *Mycobacterium* highlights the relevance of detecting genetic diversity in pathogenic strains.

Whole-genome sequencing is undoubtedly the most powerful method, in that it provides the most precise information. In practice, it is not actually feasible to sequence a panel of isolates that is wide enough to fully monitor a circulating bacterial population, and the available genome sequences actually refer to a limited number (often only one) of strains. Moreover, the choice of strains for genome sequence projects is contentious because it is unlikely that a single strain can be sufficiently paradigmatic for the biology and pathogenic potential of a species. It is therefore imperative that the identification of candidate antigens obtained by genome sequence analysis is validated and extended by the contribution of other high-throughput techniques.

DNA microarray is a powerful technology for investigating genome diversity and relatedness from a comparative genomics perspective [61]. In particular, the comparative genomic hybridization (CGH) approach circumvents the need for sequencing multiple closely related genomes. Briefly, through differential hybridization of probes generated from a test genomic DNA, for which the sequence is available, CGH reveals regions of loss (or retention) with respect to the reference strain. Owing to its intrinsic technical limitations, CGH analysis necessarily fails to detect acquisition events with respect to the reference strains. However, the increasing number of CGH studies directed towards detecting gene presence or absence profiles in bacterial populations [61,62] permits a more accurate evaluation of the genetic stability of a great number of bacterial pathogens.

Functional genomics

Based on the availability of the entire genome sequences of an organism, new disciplines of molecular biology have emerged. These techniques have the potential to accelerate the process of identifying protective protein antigens as subunit vaccine targets as well as validating and extending the range of available candidate antigens (Figure 2). Moreover, by providing a functional correlation for genes and proteins with phenotypes such as the presence and mode of pathogenicity, they offer new perspectives for the production of attenuated mutants which might be used as live vaccines or delivery systems for heterologous antigens.

Microarray expression technology

Following the expression profiling of mRNA, the transcriptome (i.e. the complete set of transcripts of an organism) can be monitored through a series of specific standardized *in vitro* conditions of growth, to identify genes that are differentially expressed in response to environmental modifications. Consequently, the use of this technology has been exploited for the study of infectious diseases by analyzing global variations in gene expression occurring during infection on both sides of the host–pathogen interaction [59].

Microarray-based expression studies provide a strong contribution to the understanding of how a pathogen orchestrates responses to the host environment. The transcriptional changes in *N. meningitidis* were investigated from meningococci incubated in human serum as

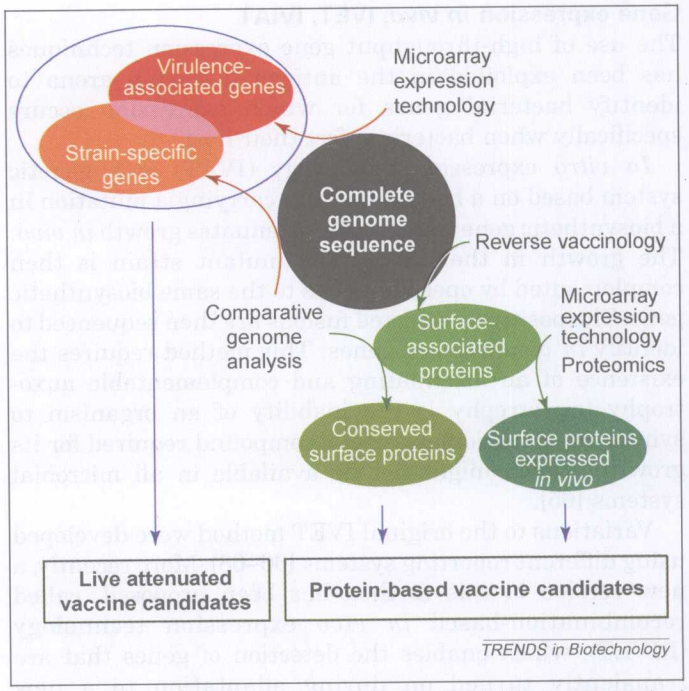


Figure 2. Interplay between the different rationales for applying genome-derived technologies to vaccine design. The rational design of protein-based vaccines can exploit complete bacterial genome sequencing by predicting the cellular localization of all the putative gene products, using a variety of sophisticated computer programs. It becomes possible to choose *a priori* potentially surface-exposed proteins that form a first group of vaccine candidates. This set of potential antigens can be validated by using DNA microarray technology and/or two-dimensional gel electrophoresis profiles to verify whether such proteins are effectively expressed by the microorganism. The elucidation of complete bacterial genome sequences also gave a new impetus to the development of live attenuated vaccines. By analyzing differences in gene content and sequence variability between pathogenic and harmless microorganisms, a set of virulence-specific genes can be identified *in silico*. Functional genomics represents an alternative approach to identifying genes essential for survival and fitness within the host, as well as a valuable tool for verifying the results of computer predictions. The wealth of information obtainable is expected to greatly enhance the ability to engineer safer and more effective live attenuated vaccines.

well as adherent to human epithelial and endothelial cells [63]. The authors of this study discovered a wide range of surface proteins that are induced under *in vivo* conditions and that could represent novel candidates for a protein-based vaccine for meningococcal diseases.

Understanding the mechanism of protection of a vaccine is important for developing a new generation of vaccines. Recently, Mueller *et al.* [64] used gene expression profiling and immunohistochemical analysis to elucidate the mechanism of protection of a whole-cell sonicate vaccine of *Helicobacter felis* in mice.

Ideally, microarray expression profiles would reveal the complex interaction between host and pathogen during the various infection steps and monitor the expression of virulence genes. Unfortunately, most microarray studies have been initiated from organisms grown *in vitro* because of the lack of efficient methods for differentially extracting bacterial RNA from tissues. Considering the influence of growth conditions on microorganism gene expression, it should be pointed out that the expression results cannot always reproduce the real *in vivo* situation. Because of these considerations, traditional biological studies remain important.