



评述与展望

Review and Progress

计算分子生物学：实验分子和基因组生物学与计算技术的结合

Y. Zhang^{1,2}, X. J. Min^{1,2}

1. College of bioinformatics science and technology, Harbin Medical University, China

2. Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, Ohio 44555, USA

 通讯作者: xmin@ysu.edu;  作者

计算分子生物学, 2012 年, 第 1 卷, 第 11 篇 doi: 10.5376/cmb.cn.2012.01.0011

收稿日期: 2012 年 12 月 15 日

接受日期: 2012 年 12 月 15 日

发表日期: 2012 年 12 月 16 日

本文首次以英文发表在 *Computational Molecular Biology* 上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 用中文再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。如果读者对中文含义理解有歧义, 请以英文原文为准。

建议最佳引用格式:

Zhang and Min, 2011. *Computational Molecular Biology: an Integration of Experimental Molecular and Genome Biology with Computational Technology*, *Computational Molecular Biology*, Vol.1, No.1 1-3 (doi: 10.5376/cmb.2011.01.0001)

摘要 计算技术和信息学的产生和发展, 对经典分子生物学研究方法进行了创新和完善。大规模平行标签测序方法, 特别是全基因组方法, 成为已经产生大量生物数据的主流实验方法。计算分子生物学是实验分子和基因组生物学与计算技术的整合。新推出的杂志, 计算分子生物学 (ISSN 1927-5587) 为社会提供了传播的计算分子和基因组生物学这个跨学科领域新发现的平台, 以应对新的挑战, 原始分子数据的产生、数据分析、比较和进化基因组学, 以及应用计算技术的力量的生物技术的应用。

关键词 计算分子生物学; 系统生物学; 基因组和基因组学; 实验分子和基因组生物学; 计算技术

Computational Molecular Biology: an Integration of Experimental Molecular and Genome Biology with Computational Technology

Y. Zhang^{1,2}, X. J. Min^{1,2}

1. College of bioinformatics science and technology, Harbin Medical University, China

2. Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, Ohio 44555, USA

 Corresponding author, xmin@ysu.edu;  Authors

Abstract The emergence and development of computational technology and informatics innovates and complements the research approaches used in the classical molecular biology. Massively parallel signature sequencing approaches, especially genome-wide approaches, become the mainstream experimental approaches, which have generated a large number of biological data. Computational molecular biology is an integration of experimental molecular and genome biology with computational technology. The newly launching journal, *Computational Molecular Biology* (ISSN 1927-5587) provides a platform for the community in computational molecular and genome biology to disseminate new discoveries in this interdisciplinary field to meet new challenges including raw molecular data generation, data analysis, comparative and evolutionary genomics, and applications of biotechnology by applying the power of computational technology.

Keywords Computational molecular biology; Systems biology; Genome and genomics; Experimental molecular and genome biology; Computational technology

计算分子生物学通过结合计算机科学中的数学算法和工具(Brutlag, 1998), 基于分子生物学的背景知识来破解生命的奥秘。计算技术和信息学的产生和发展, 对经典分子生物学研究方法的创新和完善进行了创新和完善。传统的计算生物学

侧重于诸如 RNA 结构预测和序列分析等方面。最近, 高通量实验技术产生的大量数据对了大量相关领域的生物信息学方法引起了巨大的影响方法, 包括序列分析, 蛋白质结构分析, 基因表达, 非编码 RNA, 统计遗传学, 分子进化和计算机辅



助药物设计。特别是在上世纪 80 年代中期出现的基因芯片技术, 具有高通量、并发性、微型化和自动化的优点, 该技术很快被应用在包括药物筛选领域, 新的药物开发和疾病诊断(Weinstein et al., 2002)。DNA 微阵列可以筛选数千个基因的基因表达, 并定量和定性地检测样品中基因表达的差异。基因微阵列和新开发的新一代技术的发展以指数方式产生大量的序列和数字化数据, 使得相关数据分析成为生物学家的瓶颈。然而, 数学和统计原理的应用以及计算机程序有助于解决这些问题(Mychaleckyj, 2007)。

目前的分子生物学, 包括新兴的基因组生物学涉及多个学科, 如生命科学, 医学, 药学和化学, 覆盖从低等生物到高等哺乳动物, 从原核生物到真核生物和其他领域如多生物分子进化, 疾病的诊断, 治疗和预后的分子标记的发现, 药物靶标的设计和预测。致力于生物信息学的研究人员已经开发了基于模式识别, 统计方法和信息技术的用于不同分子水平的高通量分析的算法, 工具和数据库。例如, 研究人员以前开发算法 SAM (微阵列的显着性分析), 其计算每个基因的统计学 D 以测量基于基因表达微阵列数据鉴定差异表达基因的(Tibshirani, 2006)。可以进一步过滤由基因表达微阵列产生的基因表达谱以检测不同样品中的差异表达基因。同时, 聚类分析比较组间和组间基因表达差异, 并提供样品之间的相似性。除了 k 均值聚类和分层聚类算法之外, 可以在聚类分析中使用不同的度量, 包括几何距离, 线性相关系数, 非线性相关系数和互信息。

此外, 蛋白质的功能性取决于蛋白质的一级, 二级和更高级结构的形成和稳定性。研究人员开发了大量的蛋白质结构分析, 预测和可视化工具, 并提供了可用于蛋白质结构和功能分析的必要工具(Service, 2010)。例如, SOPMA 是用于蛋白质二级结构预测的软件(Geourjon and Deleage, 1995), 而 SWISS-MODEL 是用于蛋白质三级结构预测的实用程序(Schwede et al., 2003)。

最近, 大规模平行标签测序方法, 特别是全基因组方法, 成为已经产生大量生物数据的主流实验方法。可以过滤 RNA-seq 数据以提供基因表达谱, 包括非编码 RNA 表达谱以及差异表达的编码和非编码基因(Marioni et al., 2008; Mortazavi et al., 2008)。

现在, 从 DNA、RNA 到蛋白质的实验平台已经产生了大量的数据, 这需要新的数据处理和分析方法, 推进计算分子生物学的快速发展(Vallabhajosyula and Raval, 2010)。特别是最近测序了越来越多的原核生物和真核生物的完整基因组, 可用先进的计算技术在基因组水平上比较分析。大量新兴的研究人员和开发人员正在投入到这个令人兴奋领域。

新近推出的《计算分子生物学》(ISSN 1929-4417)是一本基于同行评审、开放取阅以及在线即时发表于 BioPublisher 的期刊。该期刊主要发表计算分子生物学所有领域的最新和杰出的研究论文, 信件, 方法和综述, 涵盖从基因到基因组分子生物学的新发现, 使用统计、数学和计算方法以及新的计算方法和分子和基因组生物学数据库的开发。本刊发表的论文不仅针对计算机科学家、生物学家和从事生物学研究的师生、科研人员的研究兴趣, 而且还适合从事计算机和生物学的研发人员和对计算分子生物学领域感兴趣的本刊读者群。因此, 本期刊为社会提供了传播的计算分子和基因组生物学这个跨学科领域新发现的平台, 以应对新的挑战, 原始分子数据的产生、数据分析、比较和进化基因组学, 以及应用计算技术的力量的生物技术的应用。

参考文献

- Brutlag D.L., 1998, Genomics and computational molecular biology, *Curr Opin Microbiol*, 1: 340-345
[http://dx.doi.org/10.1016/S1369-5274\(98\)80039-8](http://dx.doi.org/10.1016/S1369-5274(98)80039-8)
- G Geourjon C., and Deleage G., 1995, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Comput Appl Biosci*, 11: 681-684
- Marioni J.C., Mason C.E., Mane S.M., Stephens M., and Gilad Y., 2008, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, *Genome Res*, 18: 1509-1517
<http://dx.doi.org/10.1101/gr.079558.108>
- Mortazavi A., Williams B.A., Mccue K., Schaeffer L., and Wold B., 2008, Mapping and quantifying mammalian transcriptomes by RNA-Seq, *Nat Methods*, 5: 621-628
<http://dx.doi.org/10.1038/nmeth.1226>
- Mychaleckyj J.C., 2007, Genome mapping statistics and



-
- bioinformatics, *Methods Mol Biol*, 404: 461-488
http://dx.doi.org/10.1007/978-1-59745-530-5_22
- Schwede T., Kopp J., Guex N., and Peitsch M.C., 2003, SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res*, 31: 3381-3385
<http://dx.doi.org/10.1093/nar/gkg520>
- Service R.F., 2010, Computational biology. Custom-built supercomputer brings protein folding into view, *Science*, 330: 308-309
<http://dx.doi.org/10.1126/science.330.6002.308-a>
- Tibshirani R., 2006, A simple method for assessing sample sizes in microarray experiments, *BMC Bioinformatics*, 7: 106
<http://dx.doi.org/10.1186/1471-2105-7-106>
- Vallabhajosyula R.R., and Raval A., 2010, Computational modeling in systems biology, *Methods Mol Biol*, 662: 97-120
http://dx.doi.org/10.1007/978-1-60761-800-3_5
- Weinstein J.N., Scherf U., Lee J.K., Nishizuka S., Gwadry F., Bussey A.K., Kim S., Smith L.H., Tanabe L., Richman S., Alexander J., Kouros-Mehr H., Maunakea A., and Reinhold W.C., 2002, The bioinformatics of microarray gene expression profiling, *Cytometry*, 47: 46-49
<http://dx.doi.org/10.1002/cyto.10041>