

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

Y. Zhang[✉], X. J. Min[✉]

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 email: xmin@ysu.edu; ✉ Author

Computational Molecular Biology, 2011, Vol.1, No.1 doi: 10.5376/cmb.2011.01.0001

Copyright © 2011 Zhang and Min. This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

Computational molecular biology cracks the mystery of life based on the background knowledge in molecular biology by combining mathematical algorithms and tools in computer science (Brutlag, 1998). The emergence and development of computational technology and informatics innovates and complements the research approaches used in the classical molecular biology. Traditionally computational biology focuses on aspects such as RNA structure prediction and sequence analysis. More recently the huge amount of data generated by high-throughput experimental technology attracts a large number of efforts of bioinformatics approaches made to the related areas including sequence analysis, protein structure analysis, gene expression, non-coding RNAs, statistical genetics, molecular evolution and computer-aided drug design. Especially for the gene microarray technology emerged in the mid-1980s, the merits such as high-throughput, concurrency, micromation and automation have quickly

enabled the technology to be applied in areas including drug filtering, new drug development and disease diagnosis(Weinstein et al., 2002). DNA microarray can screen for gene expression of thousands of genes, and detect differences of gene expression among samples quantitatively and qualitatively. The development of gene microarray and the newly developed next-generation technology has produced exponentially a large amount of sequence and digital data, making the related data analysis a bottleneck for biologists. However, the applications of mathematics and statistics principles and the computer programs have helped to solve these problems (Mychaleckyj, 2007).

Current molecular biology, including emerging genome biology, involves of multiple disciplines, such as, life science, medicine, pharmacy and chemistry, with the area covering from lower organisms to higher mammalian animals, from prokaryote to eukaryote and other areas such as multi-organism molecular

Preferred citation for this article:

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)

Received: 09 Feb., 2011 | Accepted: 15 Mar., 2013 | Published: 04 Apr., 2011

evolution, molecular marker discovery of disease diagnosis, treatment and prognosis, drug target design and prediction. Researchers devoted to bioinformatics have developed algorithms, tools and databases for high-throughput analysis in different molecular levels based on pattern recognition, statistical methods, and informatics technology. For instance, researchers previously developed algorithm SAM (significance analysis of microarray) which computed statistics D for each gene to measure difference among samples for identifying differentially expressed genes based on gene expression microarray data (Tibshirani, 2006). The gene expression profile generated by gene expression microarray can be further filtered to detect differentially expressed genes among different samples. Meanwhile, cluster analysis can compare intra- and inter-group gene expression differences and provide similarity among samples. Different measures can be used in cluster analysis including geometry distance, linear correlation coefficients, non-linear correlation coefficients and mutual information in addition to k-means clustering and hierarchical clustering algorithms.

Furthermore, the functionality of proteins depends on the formation and stabilization of the primary, secondary and higher structure of proteins. Researchers developed a large number of tools for protein structure analysis, prediction and visualization and made necessary tools to be available for protein structure and functional analysis (Service, 2010). For instances, SOPMA is a software for protein secondary structure prediction (Geourjon and Deleage, 1995), while SWISS-MODEL has utilities for protein tertiary structure prediction (Schwede et al., 2003).

Recently, massively parallel signature sequencing approaches, especially genome-wide approaches, become the mainstream experimental approaches, which have generated a large number of biological data. RNA-seq data can be filtered to provide gene expression profiles, including non-coding RNA expression profile as well as differentially

expressed coding and non-coding genes (Marioni et al., 2008; Mortazavi et al., 2008).

Now, experimental platforms from DNA, RNA to proteins have generated a large number of data, which call for new data processing and analyzing approaches and advance the rapid development of computational molecular biology (Vallabhajosyula and Raval, 2010). Particularly more and more complete genomes of both prokaryotes and eukaryotes have been sequenced recently that makes comparative analysis feasible at the genome level with advanced computational technology. A large number of new emerging researchers and developers are coming to devote them to this exciting field.

The newly launching journal, Computational Molecular Biology (ISSN 1927-5587), is an open access, peer reviewed journal published online by BioPublisher. The Journal is publishing all the latest and outstanding research articles, letters, methods, and reviews in all areas of Computational Molecular Biology, covering new discoveries in molecular biology, from genes to genomes, using statistical, mathematical, and computational methods as well as new development of computational methods and databases in molecular and genome biology. The papers published in the journal are expected to be of interests to computational scientists, biologists, and teachers/students/researchers engaged in biology, as well as are appropriate for R & D personnel and general readers interested in computational technology and biology. Thus this journal 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.

References

- 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
PMid:8808585
- 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>
PMid:18550803 PMCid:PMC2527709
- 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>
PMid:18516045
- 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
PMid:18450063
- 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>
PMid:12824332 PMCid:PMC168927
- 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>
PMid:16512900 PMCid:PMC1450307
- 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
PMid:20824468
- 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>
PMid:11774349