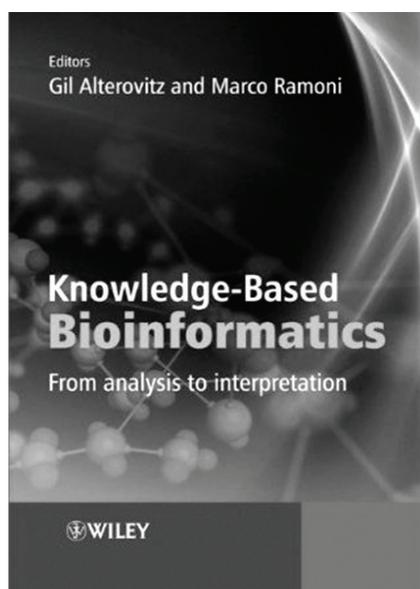


# Knowledge-Based Bioinformatics: From Analysis to Interpretation, Edited by Gil Alterovitz & Marco Ramoni

Eun-Young Kim, MD, PhD

Department of Clinical Pharmacology, Inje University Busan Paik Hospital  
eykim@inje.ac.kr



By Gil Alterovitz & Marco Ramoni  
John Wiley & Sons Ltd.; 2010  
Price: \$75.00  
ISBN: 978-0-470-74831-2

Two editors, Gil Alterovitz and Marco Ramoni, are faculty members in Children's Hospital Informatics Program, Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School. The book was edited based on their research experiences in the field of bioinformatics. Editors wrote that the book is the result of an international effort including the contributors from 19 institutions in 7 countries in the preface.

This book consists of four sections, *Knowledge-Driven Approaches*, *Data-Analysis Approaches*, *Gene and Protein Information*, and *Biomolecular Relationships and Meta-Relationships*. It covers extensive topics from knowledge-driven approaches, data analysis methods, to application in the field of bioinformatics and systems biology. It provides guidance on dealing with large volumes of data in knowledge-based analysis and helps to interpret genomic, proteomic and metabolomic data to enhance biomedical research.

The first section introduces knowledge-driven approaches such as knowledge representation, data mining, visualization, and ontology-based analysis to discover, predict, or aggregate information from large datasets. It also introduces not only for online databases of biomedical literature and experimental data but also for implemented tools. The implemented examples are variety of data on genome-scale, gene-expression, and proteins. In chapter 4, they discuss about the design of knowledge bases, which is not an easy task requiring communication between resource developers and end users to achieve a successful outcome.

In the second section, *Data-Analysis Approaches*, the book introduces analytic methods from classical statistical learning methods to automatic text mining analysis including Bayesian methods. Chapter 5, *Classical statistical learning in bioinformatics*, explains clustering methods, principal components analysis, multidimensional scaling, and supporting vector machines as a machine learning method in terms of

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2010 The Korean Society of Medical Informatics

classification and prediction using biomedical data. Chapter 6, *Bayesian methods in genomics and proteomics studies*, introduces the basic theory and applications in many different concepts and data using the methods. Chapter 7, *Automatic text analysis for bioinformatics knowledge discovery*, focuses on knowledge discovery from enormous numbers of biomedical literature using natural language processing technology and its evaluation.

The third section, *Gene and Protein Information*, introduces the basis of gene ontology (GO) functional annotation with its application and limitations. The section reviews recent progress in genome annotation and indicates that GO is an example of value-added biocuration effort. It also introduces batch-learning self-organizing map (BLSOM) adapted for genome informatics, which is suitable for phylogenetic estimation for novel gene sequences. Another example of the topics included is large-scale metagenomic analyses using recently released next-generation sequencers for improving the accuracy of phylogenetic estimation with BLSOM for shorter sequences up to 200 bases.

The fourth section, *Biomolecular Relationships and Meta-Relationships*, touches molecular networks, biological pathways, and challenges in genotype to phenotype relationships. Chapter 11, *Molecular network analysis and applications*,

reviews four types of commonly conducted network analysis, topology analysis, motif analysis, modular analysis, and network comparison, on large-scale molecular networks and their applications. Chapter 12, *Biological pathway analysis: an overview of Reactome and other integrative pathway knowledge bases*, introduces knowledge based human biological pathways with databases such as Reactome, KEGG, and WikiPathways. The last chapter 13, *Methods and challenges of identifying biomolecular relationships and networks associated with complex diseases/phenotypes, and their application to drug treatments*, addresses molecular type to phenotype relationship using pathway knowledge bases and systems biology approaches to bridge into clinical use.

The book emphasizes the importance of knowledge-driven information. In recent decades, there have been exponential growths of biomedical knowledge from biochemical and genetic research although little has been used for clinical application. Thus knowledge-driven application might be one of ways to solve the problem. Therefore, this book covering knowledge-driven bioinformatics from the concept to the application is useful for biologists, clinicians, computer scientists, bioinformaticians, and graduate students studying in the field.