
Preface

The idea of this book has been excogitated over the past few years from my own research endeavor to apply an integration of bioinformatics, genomic, and genetic approaches in the identification of genetic and biochemical markers in cardiopulmonary diseases. To this project I have brought my five years as director of the Gene Expression Profiling Core at The Johns Hopkins Center of Translational Respiratory Medicine, my three-year stint as a coordinator of the Affymetrix User Group monthly meeting at The Johns Hopkins University medical institutions and two years as a director of the Molecular Resource Core in the National Heart, Lung, and Blood Institute-funded Program Project Grant on Cytoskeletal Regulation of Lung Endothelial Pathobiology at the University of Chicago, as well as my experience in obtaining my R01 grant award and involvement in a successful Specialized Centers of Clinically Oriented Research application on "Molecular Approaches to Ventilator-Associated Lung Injury," directed by Dr. Joe G.N. Garcia, from the National Institutes of Health. The idea for this book further crystallized by listening to the enlightening advice at the System Biology Symposium held at the University of Chicago in October 2005 from Dr. Phillip A. Sharp, professor of biology from Massachusetts Institute of Technology and a 1993 Nobel laureate (for discovering gene splicing). He pointed out that to be successful in the omics age, every biological or biomedical researcher should know some bioinformatics. I fully concurred with Dr. Sharp's comments and developed a book proposal. Dr. Sunil Nair, a perceptive publisher from Taylor & Francis, was first to hand me a book contract, which I happily accepted.

Bioinformatics is emerging as an ever-evolving new branch of science in which computer tools are applied to collect, store, and analyze biological data to generate new biological information. Over the past few years, major progress in the field of molecular biology, coupled with rapid advances in genomic technologies, has led to an explosive growth in the biologi-

cal information distributed in a variety of biological databases. Currently, genome resources from a number of species are available at the National Center of Biological Information Website (<http://www.ncbi.nlm.nih.gov/Genomes/index.html>). This list is being expanded at an unprecedented pace. A challenge facing researchers today is how to piece together and analyze this plethora of data to make new biological discoveries and gain new unifying global biological insights. This led to the absolute requirement for biologists and medical researchers to obtain a reasonable amount of knowledge on computational biology (bioinformatics), i.e., applying computational approaches to facilitate the understanding of various biological processes such as a more global perspective in experimental design and the ability to capitalize on database mining — the process by which testable hypotheses are generated regarding the function or structure of a gene or protein of interest by identifying similar sequences in better characterized organisms. Equally important is that computer gurus need to have some basic understanding of biological problems in order for them to efficiently execute their computer skills in the field of bioinformatics.

Biologists are usually not extensively trained with computer skills, and computer experts rarely have a biology background. One of the goals of this book is to bridge or shorten the knowledge gap between biologists and computer specialists to make better and more efficient application and development of bioinformatics.

This book will cover the most state-of-the-art bioinformatics applications a biologist needs to know. Part I will focus on genome and DNA sequence analysis with chapters on genome analysis, common DNA analysis tools, phylogenetics analysis and SNP and haplotype analysis. Part II will center on transcriptome and RNA sequence analysis, with chapters on microarray, SAGE, regulation of gene expression, miRNA, and siRNA. Part III will present widely applied programs or tools in proteome, protein sequences, protein functions, and functional annotation of proteins in murine models, and Part IV will introduce the most useful basic biocomputing tools in chapter on the application of programming languages in biology, Website and database design, and interchanging data between Microsoft Excel and Access. Abbreviations, a selected glossary, and Websites for selected bioinformatics software are listed in the appendices for reference.

Our goal is to assimilate the most current bioinformatic knowledge and tools relevant to the omics age into one cohesive, concise, self-contained book accessible to biologists, to computer scientists, and to motivated non-specialists with a strong desire to learn bioinformatics. The twenty-five

contributors to this book have been recruited from nine world-renowned institutions in six countries: Johns Hopkins University, University of Chicago, Ohio State University, Georgetown University in the United States; the University of Toronto in Canada; the Swiss Institute of Bioinformatics in Switzerland; Okayama University in Japan; Wuhan University in China; and the International Centre for Genetic Engineering and Biotechnology in India. In each chapter, a theoretical introduction of the subject is followed by the exemplar step-by-step tutorial to help readers both to have a good theoretical understanding and to master a practical application. Although all the chapters were contributed by experts in their respective fields, the book was written to be easily understood. Complex mathematic deductions and jargon were avoided or reduced to a minimum. Any novice, with a little computer knowledge, can learn bioinformatics from this book without difficulty. In the overview of each chapter, several authoritative references were given, so that more experienced readers may explore the subject in depth and find new horizons. The target readers of this book are biologists but computer specialists may find a fruitful, comprehensive, and concise synopsis of biological problems to tackle in each chapter.

This book is the collective efforts of editorial staff and other individuals. I am deeply indebted to all contributing authors for their tremendous efforts to finish their work on time, and for their gracious tolerance of my repeated haggling for revisions. We are grateful to the editorial staff for their tireless and meticulous work on accuracy and detail. We want to thank several other unnamed people who have helped us along the way for their valuable guidance and innumerable improvement suggestions. We apologize to many colleagues whose work was not covered in this book due to space limitations.

We welcome any criticism and suggestions for improvement so that they may be incorporated into the next edition.

Shui Qing Ye