PREFACE

Life follows patterns. The way life has evolved, how it is organized and how it responds to a stimulus are patterns that scientists aim to unravel. Mathematical and computational models that identify or use these biological patterns have a myriad of applications. When scientists sequence a billion-base genome, we look for sequence patterns that tell us how the organism evolved, how the expression of its genes is regulated, and even if the patterns can reveal the three-dimensional structure of a molecule and its interaction with others. The same can be applied when measuring the level of expression of thousands of genes and looking for patterns that tell us how genes are being regulated in a biological condition, or which ones can be used to predict a clinical outcome.

Bioinformatics enables the computational analysis of the patterns generated by scientific experiments. Today, a single laboratory is capable of generating a vast amount of biological data. In addition, there is a wealth of data already available in public databases, which makes the modern life sciences almost dependent on bioinformatics.

Annually, hundreds of computational tools and methods are developed to allow the analysis and integration of large volumes of biological data. This book brings together the state-of-the-art from several fields of bioinformatics—from the automatic identification and classification of viruses to the analysis of the transcriptome of single cells, including artificial intelligence algorithms to discover biomarkers and text mining approaches to help in the interpretation of the findings.

The chapters cover areas of bioinformatics that are different but complementary. The chapters by Annie Lee and Zhou describe different approaches of pattern analysis of biological sequences. The former shows how the WeMine program can be used to identify protein functional regions, to predict pairwise interactions between protein-DNA and protein-protein network, and to find correlations among patterns and residues. The latter describes a new method named Pattern Discovery and Disentanglement that performs association discovery on aligned pattern clusters and residue-to-residue interactions prediction. Gruber's chapter demonstrates the importance of profile hidden Markov models in the classification and discovery of viruses in samples of metagenomics. The chapters by Chen and Haspel show the importance of bioinformatics in genomics. While the former evaluates 23 error-correction methodologies in sequencing data, the latter describes computational methods that detect large-scale structural rearrangements in chromosomes from whole genome and transcriptome sequences, as well as high-throughput chromosome conformation capture data.

The book also covers techniques associated with transcriptome analysis, which measures the expression profile of all genes in a single organism, tissue or cell. The chapter by Moriyama reviews and benchmarks several transcriptome assembly techniques for plants. Since many plants are polyploid and have huge genomes with a lot of repetitive sequences, the de novo transcriptome assembly represents a significant challenge. The chapter by David covers the methods for analysis of single cell transcriptomics, including trajectory analysis, pseudotime assumptions, and network inference. The chapter by Sanchez-Cabo demonstrates that Bayesian inference of probabilistic models is a powerful method to analyze and integrate gene expression from RNA-seq data. Zhang's chapter gives an overview of the machine learning approaches applied to gene expression data which are utilized in biomarker discovery. Finally, the chapter by Candelier focuses on the use of text mining approaches to help interpret the results of transcriptome analysis.

Bioinformatics is permeated in life sciences research. Without bioinformatics we will not be able to fully advance in the understanding and functioning of life, to identify the molecular mechanisms and treatment of diseases, nor to understand the evolutionary processes of life. I hope this book will help the readers to see the importance of bioinformatics in molecular biology.

Helder I. Nakaya, PhD
Department of Clinical and Toxicological Analyses
School of Pharmaceutical Sciences
University of Sao Paulo, Sao Paulo, Brazil
Doi: https://doi.org/10.36255/exonpublications.bioinformatics.2021.pr