

PREFACE

Massively parallel DNA sequencing and RNA sequencing have become widely available, reducing the cost by several orders of magnitude and placing the capacity to generate gigabases to terabases of sequence data into the hands of individual investigators. These so-called *next-generation sequencing (NGS)* technologies have dramatically accelerated biological and biomedical research by enabling the comprehensive analysis of genomes and transcriptomes to become inexpensive, routine, and widespread. The ensuing explosion in the volume of data has spurred numerous advances in computational methods for NGS data analysis.

This book aims to provide an in-depth survey of some of the most important recent developments in this area. It is neither intended as an introductory text nor as a comprehensive review of existing bioinformatics tools and active research areas in NGS data analysis. Rather, our intention is to make a carefully selected set of advanced computational techniques accessible to a broad readership, including graduate students in bioinformatics and related areas and biomedical professionals who want to expand their repertoire of computational techniques for NGS data analysis. We hope that our emphasis on in-depth presentation of both algorithms and software for computational data analysis of current high-throughput sequencing technologies will best prepare the readers for developing their own algorithmic techniques and for successfully implementing them in existing and novel NGS applications.

The book features 18 chapters authored by bioinformatics experts who are active contributors to the respective subjects. The chapters are intended to be largely independent, so that readers do not have to read every chapter nor have to read them in a particular order. The chapters are grouped into the following four parts:

- Part I focuses on computing and experimental infrastructure for NGS data analysis, including chapters on cloud computing, a modular pipeline for metabolic pathway reconstruction, pooling strategies for massive viral sequencing, and high-fidelity sequencing protocols.

- Part II concentrates on analyses of DNA sequencing data and includes chapters on the classic scaffolding problem, detection of genomic variants, two chapters on finding insertions and deletions, and two chapters on the analysis of DNA methylation sequencing data.
- Part III is devoted to analyses of RNA-seq data. Two chapters describe algorithms and compare software tools for transcriptome assembly: one chapter focuses on methods for alternative splicing analysis and the other chapter focuses on tools for transcriptome quantification and differential expression analysis.
- Part IV explores computational tools for NGS applications in microbiomics. The first chapter concentrates on error correction of NGS reads from viral populations, then two chapters describe methods for viral quasispecies reconstruction, and the last chapter surveys the state of the art and future trends in microbiome analysis.

We are grateful to all the authors for their excellent contributions, without which this book would not have been possible. We hope that their deep insights and fresh enthusiasm will help in attracting new generations of researchers to this dynamic field. We would also like to thank Yi Pan and Albert Y. Zomaya for nurturing this project since its inception, and the editorial staff at Wiley Interscience for their patience and assistance throughout the project. Finally, we wish to thank our friends and families for their continuous support.

ION I. MĂNDOIU

Storrs, Connecticut

ALEXANDER ZELIKOVSKY

Atlanta, Georgia