

# Preface

The purpose of this book is to collect in a single volume the essentials of high throughput sequencing data analysis. These new technologies allow performing, at an unprecedented low cost and high speed, a panoply of experiments spanning the sequencing of whole genomes or transcriptomes, the profiling of DNA methylation, and the detection of protein–DNA interaction sites, among others. In each experiment a massive amount of sequence information is generated, making data analysis the major challenge in high throughput sequencing-based projects. Hundreds of bioinformatics applications have been developed so far, most of them focusing on specific tasks. Indeed, numerous approaches have been proposed for each analysis step, while integrated analysis applications and protocols are generally missing. As a result, even experienced bioinformaticians struggle when they have to discern among countless possibilities to analyze their data. This, together with a lack of enough qualified personnel, reveals an urgent need to train bioinformaticians in existing approaches and to develop integrated, “from start to end” software applications to face present and future challenges in data analysis.

Given this scenario, our motivation was to assemble a book covering the aforementioned aspects. Following three fundamental introductory chapters, the core of the book focuses on the bioinformatics aspects, presenting a comprehensive review of the methods and programs existing to analyze the raw data obtained from each experiment type. In addition, the book is meant to provide insight into challenges and opportunities faced by both, biologists and bioinformaticians, during this new era of sequencing data analysis.

Given the vast range of high throughput sequencing applications, we set out to edit a book suitable for readers from different research areas, academic backgrounds and degrees of acquaintance with this new technology. At the same time, we expect the book to be equally useful to researchers involved in the different steps of a high throughput sequencing project.

The “newbies” eager to learn the basics of high throughput sequencing technologies and data analysis will find what they yearn for specially by reading the first introductory chapters, but also by obviating the details and getting the rudiments of the

core chapters. On the other hand, biologists that are familiar with the fundamentals of the technology and analysis steps, but that have little bioinformatic training will find in the core chapters an invaluable resource where to learn about the different existing approaches, file formats, software, parameters, etc. for data analysis. The book will also be useful to those scientists performing downstream analyses on the output of high throughput sequencing data, as a perfect understanding of how their initial data was generated is crucial for an accurate interpretation of further outcomes. Additionally, we expect the book to be appealing to computer scientists or biologists with a strong bioinformatics background, who will hopefully find in the problematic issues and challenges raised in each chapter motivation and inspiration for the improvement of existing and the development of new tools for high throughput data analysis.

Naiara Rodríguez-Ezpeleta  
Michael Hackenberg  
Ana M. Aransay

Bioinformatics for High Throughput Sequencing

Rodríguez-Ezpeleta, N.; Hackenberg, M.; Aransay, A.M.  
(Eds.)

2012, XI, 255 p. 29 illus., 25 illus. in color., Hardcover

ISBN: 978-1-4614-0781-2