

Preface

During the past few years, there have been enormous advances in genomics and molecular biology, which carry the promise of understanding the functioning of whole genomes in a systematic manner. The challenge of interpreting the vast amounts of data from microarrays and other high throughput technologies has led to the development of new tools in the fields of computational biology and bioinformatics, and opened exciting new connections to areas such as chemometrics, exploratory data analysis, statistics, machine learning, and graph theory.

The Bioconductor project is an open source and open development software project for the analysis and comprehension of genomic data. It is rooted in the open source statistical computing environment R. This book's coverage is broad and ranges across most of the key capabilities of the Bioconductor project. Thanks to the hard work and dedication of many developers, a responsive and enthusiastic user community has formed. Although this book is self-contained with respect to the data processing and data analytic tasks covered, readers of this book are advised to acquaint themselves with other aspects of the project by touring the project web site www.bioconductor.org.

This book represents an innovative approach to publishing about scientific software. We made a commitment at the outset to have a fully *computable book*. Tables, figures, and other outputs are dynamically generated directly from the experimental data. Through the companion web site, www.bioconductor.org/mogr, readers have full access to the source code and necessary supporting libraries and hence will be able to see how every plot and statistic was computed. They will be able to reproduce those calculations on their own computers and should be able to extend most of those computations to address their own needs.

Acknowledgments

This book, like so many projects in bioinformatics and computational biology, is a large collaborative effort. The editors would like to thank the chapter authors for their dedication and their efforts in producing widely used software, and also in producing well-written descriptions of how to use that software.

We would like to thank the developers of R, without whom there would be no Bioconductor project. Many of these developers have provided additional help and engaged in discussions about software development and design. We would like to thank the many Bioconductor developers and users who have helped us to find bugs, think differently about problems, and whose enthusiasm has made the long hours somewhat more bearable.

We would also like to thank Dorit Arlt, Michael Boutros, Sabina Chiaretti, James MacDonald, Meher Majety, Annemarie Poustka, Jerome

Ritz, Mamatha Sauermann, Holger Sltmann, Stefan Wiemann, and Seth Falcon, who have contributed in many different ways to the production of this monograph. Much of the preliminary work on the **MLInterfaces** package, described in Chapter 16, was carried out by Jess Mar, Department of Biostatistics, Harvard School of Public Health. Ms Mar’s efforts were supported in part by a grant from Insightful Corporation.

The Bioconductor project is supported by grant 1R33 HG002708 from the NIH as well as by institutional funds at both the Dana Farber Cancer Institute and the Fred Hutchinson Cancer Research Center. W.H. received project-related funding from the German Ministry for Education and Research through National Genome Research Network (NGFN) grant FKZ 01GR0450.

Seattle
Boston
Cambridge (UK)
Baltimore
Berkeley

Robert Gentleman
Vincent Carey
Wolfgang Huber
Rafael Irizarry
Sandrine Dudoit
February 2005

Bioinformatics and Computational Biology Solutions

Using R and Bioconductor

Gentleman, R.; Carey, V.; Huber, W.; Irizarry, R.; Dudoit,
S. (Eds.)

2005, XIX, 474 p., Hardcover

ISBN: 978-0-387-25146-2