

Preface

The last 10 to 15 years have seen a great surge of statistical research motivated by data from high-throughput genomic assays. Microarray technology, in particular, has driven many statistical advances in high-dimensional data analysis. Novel and older concepts of error rate control for testing multiple hypotheses, various adaptations of empirical Bayes methods, penalized and shrinkage regression techniques, etc., have all found new glory. It is not unjust to say that statisticians have done their fair share of work in developing the field of array-based bioinformatics and biomedical research.

In the last five or so years, next generation sequencing (NGS) technology has been changing the face of biomedical research, replacing the old (micro)-array technology in many ways. Overall, NGS offers a more accurate and cost effective means of studying a variety of genomic signals with a wide range of applications. With any new high-throughput technology come new data analytic challenges. Statisticians are in a unique position to make a difference in this exciting new interdisciplinary area by offering valid methods for studying signals in noisy data and the means to compare signals across multiple experimental conditions. Statistical methods for this relatively new data type have been sufficiently developed to warrant compilation of this book, which in turn may generate further interest in NGS technology amongst statisticians and lead to additional advances in the field.

The idea of editing a volume on statistical methods for analyzing NGS data first came to us about two years ago. We discussed the possibility during a conference on NGS data analysis that took place at Iowa State University. Subsequently, we approached several prominent researchers with extensive experience in the area. We were fortunate to have received overwhelming support and commitment from many individuals and their research teams. As a result, we now have this exciting volume consisting of twenty chapters written by statistical experts with first-hand knowledge in the field of NGS data analysis.

The first chapter of the book provides an introduction and an overview of NGS technologies, statistical challenges, and data analysis techniques. The next six chapters discuss design issues and inferential techniques for analyzing gene expression data as measured by next generation sequencing of RNA (RNA-seq).

Mapping of expression QTL is discussed next, followed by normalization of RNA-seq data. Statistical clustering and classification methods specially geared toward RNA-seq data are covered in Chaps. 10 and 11. Chapters 12 and 13 present different aspects of isoform detection using RNA-seq data. Another important NGS data type, CHIP-seq, is covered in the next two chapters. Other specialized applications of NGS technology—such as genotype calling, metagenomic analysis, detection of copy number changes and other structural variations, analysis for paired samples, and analysis of rare variants—are discussed in the last five chapters of this book.

This volume has been written primarily for statisticians who are interested in conducting methodological research in this area. No prior knowledge of NGS or genomics is assumed. Most of the required concepts from genomics and biochemistry have been explained, and references have been provided for a deeper understanding of such concepts. Scientists and practitioners dealing with NGS data will also find this book useful. Powerful software tools for NGS data analysis are illustrated in several chapters. Also, many chapters from this book could be used in a one to two semester graduate-level course in statistical bioinformatics.

We wish to thank the outstanding researchers who provided chapters for this book. We appreciate their hard work and their willingness to make the revisions we requested. Reading their work has enhanced our knowledge of the field, and we hope many other readers will benefit from the contributions of the authors.

Louisville, KY, USA
Ames, IA, USA

Somnath Datta
Dan Nettleton

Statistical Analysis of Next Generation Sequencing Data

Datta, S.; Nettleton, D. (Eds.)

2014, XIV, 432 p. 87 illus., 68 illus. in color., Hardcover

ISBN: 978-3-319-07211-1