Contents

Preface, ix

Acknowledgments, xiii

Editor, xv

Contributors, xvii

SECTION I **Commonly Used Tools for Big Data Analysis**

CHAPTER 1 ▪ Linux for Big Data Analysis 3
SHUI QING YE AND DING-YOU LI

CHAPTER 2 ▪ Python for Big Data Analysis 15
DMITRY N. GRIGORYEV

CHAPTER 3 ▪ R for Big Data Analysis 35
STEPHEN D. SIMON

SECTION II **Next-Generation DNA Sequencing Data Analysis**

CHAPTER 4 ▪ Genome-Seq Data Analysis 57
MIN XIONG, LI QIN ZHANG, AND SHUI QING YE

CHAPTER 5 ▪ RNA-Seq Data Analysis 79
LI QIN ZHANG, MIN XIONG, DANIEL P. HERUTH, AND SHUI QING YE

CHAPTER 6 ▪ Microbiome-Seq Data Analysis 97
DANIEL P. HERUTH, MIN XIONG, AND XUN JIANG

CHAPTER 7 ▪ miRNA-Seq Data Analysis 117
DANIEL P. HERUTH, MIN XIONG, AND GUANG-LIANG BI

CHAPTER 8 ▪ Methylome-Seq Data Analysis 131
CHENGPENG BI

CHAPTER 9 ▪ ChIP-Seq Data Analysis 147
SHUI QING YE, LI QIN ZHANG, AND JIANCHENG TU

SECTION III **Integrative and Comprehensive Big Data Analysis**

CHAPTER 10 ▪ Integrating Omics Data in Big Data Analysis 163
LI QIN ZHANG, DANIEL P. HERUTH, AND SHUI QING YE

CHAPTER 11 ▪ Pharmacogenetics and Genomics 179
ANDREA GAEDIGK, KATRIN SANGKUHL, AND LARISA H. CAVALLARI

CHAPTER 12 ▪ Exploring De-Identified Electronic Health Record Data with i2b2 201
MARK HOFFMAN

CHAPTER 13 ▪ Big Data and Drug Discovery 215
GERALD J. WYCKOFF AND D. ANDREW SKAFF

CHAPTER 14 ▪ Literature-Based Knowledge Discovery 233
HONGFANG LIU AND MAJID RASTEGAR-MOJARAD

CHAPTER 15 ▪ Mitigating High Dimensionality in Big Data Analysis 249
DEENDAYAL DINAKARPANDIAN

INDEX, 265