Contents

Contributors xi

Foreword xiii
Muin J. Khoury

Acknowledgments xv

1 **The Evolving Field of Genetic Epidemiology** 1
Melissa A. Austin, M.S., Ph.D., and Terri H. Beaty, Ph.D.
1.1 Introduction 1
1.2 Assessing Genetic Influences on Disease, Human Genetics Concepts, and Genomic Technology 2
1.3 Family-based Study Designs: Linkage, Exome Sequencing and Case–Parent Trios 3
1.4 Genetic Association Studies of Common and Rare Variants, Large-scale Collaborations, and Population Stratification 4
1.5 Gene–Environment Interactions, Epistasis, and Non-Mendelian Genetics 5
1.6 Software and Data Resources 6
1.7 Ethical Issues and Translational Genetic Epidemiology 7
1.8 Conclusion 8

2 **Assessing Genetic Influences on Diseases and Risk Factors** 10
Melissa A. Austin, M.S., Ph.D.
2.1 Introduction 10
2.2 Familial Aggregation Studies 10
2.2.1 Familial aggregation study designs 11
2.3 Heritability Analysis 14
2.3.1 Polygenic model 15
2.3.2 Estimating heritability using twins 17
2.3.3 Assumptions, biases, and misconceptions in heritability analysis 17
2.4 Conclusion: Heritability and the Polygenic Model in the Genomic Era 22

3 **Genetic Concepts and Genomic Technology for Genetic Epidemiology** 25
Melissa A. Austin, M.S., Ph.D.
3.1 Introduction 25
3.2 Mendelian Inheritance and Complex Traits 26
3.2.1 Mendel's laws 26
3.2.2 Modes of inheritance for single genes 27

3.2.3 Online Mendelian Inheritance in Man (OMIM®) 27
3.2.4 Complex traits 27
3.3 Hardy–Weinberg Principle 29
3.3.1 Deviations from Hardy–Weinberg equilibrium 31
3.4 Genetic Code, Gene Structure, and Genetic Markers 31
3.4.1 Exons, introns, and untranslated regions 31
3.4.2 Single nucleotide polymorphisms (SNPs) and structural variations 32
3.5 Genetic Linkage and Linkage Disequilibrium 33
3.5.1 Genetic linkage and recombination 33
3.5.2 Linkage disequilibrium and haplotypes 34
3.5.3 Tag SNPs 37
3.5.4 SNP genotyping platforms 37
3.6 DNA Sequencing Technologies 38
3.6.1 Three generations of DNA sequencing 39
3.6.2 Exome sequencing 40
3.7 Study Designs for Common and Rare Genetic Variants 42
3.8 Conclusion 42

4 **Family Studies in Genetic Epidemiology: From Linkage to Exome Sequencing** 46
Melissa A. Austin, M.S., Ph.D., Terri H. Beaty, Ph.D., and Janet L. Stanford, M.P.H., Ph.D.
4.1 Linkage Analysis 46
4.1.1 Basics of LOD score linkage analysis 46
4.1.2 Non-parametric linkage analysis 50
4.2 Family-based Association Studies 55
4.2.1 Linkage and association in families 55
4.2.2 Case–parent trios and the transmission disequilibrium test (TDT) 56
4.2.3 Extensions of the TDT 58
4.3 Conclusion 64

5 **Genetic Association Studies** 68
Melissa A. Austin, M.S., Ph.D., Bruce M. Psaty, M.D., Ph.D., and Stephen M. Schwartz, Ph.D., M.P.H.
5.1 Introduction to Genetic Association Studies 69
5.2 Study Designs for Candidate Gene Studies 69
5.2.1 Practical considerations for selecting candidate genes 69
5.3 Complexities of Interpreting Genetic Association Results 72
5.3.1 Direct, causal relationship 72
5.3.2 Indirect association due to linkage disequilibrium 73
5.3.3 False-positive associations 74
5.3.4 False-negative associations and power 75
5.4 Basics of Genome-wide Association Studies (GWAS) 75
5.4.1 GWAS case–control studies 76
5.4.2 Data analysis of GWAS 76

5.5 Large-scale Collaborations: The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium 77
5.5.1 The CHARGE Consortium 80
5.5.2 CHARGE goals and organization 81
5.5.3 Genome-wide genotyping methods 81
5.5.4 Design of CHARGE analyses 81
5.5.5 Phenotype working groups 82
5.5.6 Analysis methods 82
5.5.7 Productivity 83
5.5.8 Collaborations with non-CHARGE studies and other consortia 83
5.5.9 Working-group model as innovation 83
5.5.10 A new era of consortia 84
5.5.11 CHARGE: past, present, and future 84
5.5.12 Facilitating large-scale collaborations 85
5.6 Case–Parent Trio Designs in GWAS 85
5.7 Heritability and Allelic Architecture of Complex Traits 85
5.7.1 Missing heritability in GWAS 85
5.8 Study Designs and Statistical Analysis of DNA Sequencing Data 89
5.8.1 Study designs for rare genetic variants 90
5.8.2 Methods for detecting association with rare variants 91
5.9 Conclusion 92

6 Population Stratification in Genetic Association Studies 97
Barbara McKnight, Ph.D.
6.1 General Concepts 97
6.1.1 Rationale for GWAS 97
6.1.2 Population stratification confounding 98
6.2 Detecting Population Stratification 100
6.3 Statistical Methods to Adjust for Global Population Stratification Confounding 103
6.3.1 Genomic control 103
6.3.2 Ethnicity adjustment 105
6.3.3 Principal components adjustment 105
6.3.4 Estimated ancestry adjustment 106
6.3.5 Estimated correlation methods 107
6.4 Statistical Methods to Adjust for Local Population Stratification Confounding 109
6.4.1 Mixing of populations and local population structure 109
6.4.2 Lactase persistence gene 110
6.4.3 Controlling for local ancestry 111
6.5 Conclusion 111

7 Gene–Environment Interactions and Epistasis 115
Melissa A. Austin, M.S., Ph.D., and Ruth Ottman, Ph.D.
7.1 Why Study Interaction? 115
7.2 Models of Interaction 116
7.3 Statistical Interaction and the Problem of Measurement Scale 119

7.3.1 Statistical analysis of interaction for cohort and case–control studies 119
7.3.2 Multiplicative and additive measurement scales 121
7.3.3 Causal modeling applied to interaction 122
7.4 Case-only Study Design 126
7.4.1 Considerations for case-only studies 126
7.5 Pharmacogenomics 128
7.6 Gene–Environment-wide Interaction Studies (GEWIS) 128
7.6.1 Harmonization of environmental data 128
7.6.2 GEWIS data analysis approaches 130
7.7 Epistasis: Interactions Between Genes 132
7.8 Conclusion 134

8 Non-Mendelian Genetics **138**
Stephen M. Schwartz, Ph.D., M.P.H.
8.1 Introduction 138
8.2 Mitochondrial Genetics 139
8.3 *De Novo* Variation 141
8.4 Parental and Parent-of-origin Effects 143
8.5 DNA Methylation: An Epigenetic Mechanism Underlying Phenotypic Variation 146
8.6 Conclusion 149

9 Software and Data Resources for Genetic Epidemiology Studies **154**
Timothy A. Thornton, Ph.D.
9.1 Introduction 154
9.2 Software for Genetic Data Analysis 155
9.2.1 R software 155
9.2.2 PLINK 155
9.2.3 Software for association testing in samples from structured populations 155
9.2.4 Individual ancestry estimation software 156
9.2.5 Relatedness estimation software for GWAS 158
9.2.6 Software for power calculations in genetic association studies 159
9.2.7 Genotype imputation software 160
9.2.8 An alphabetical list of genetic analysis software 160
9.3 Human Reference Panels and Resources 160
9.3.1 International HapMap Project (HapMap) 160
9.3.2 1000 Genomes Project 161
9.3.3 Human Genetic Diversity Panel (HGDP) 162
9.3.4 Genome Variation Server 162
9.4 Genotype and Phenotype Repositories 163
9.5 Conclusion 163

10 Ethical Issues in Genetic Epidemiology **166**
Stephanie M. Fullerton, D.Phil., and Kelly Edwards, Ph.D.
10.1 Introduction 166
10.2 Current Regulatory Environment 166

10.3 Trade-offs of Data De-identification 168
10.4 Informed Consent and Respect 169
10.5 Research Governance 171
10.6 Conclusion 172

11 Public Health and Clinical Applications of Genetic Epidemiology **176**
Marta Gwinn, M.D., M.P.H., and W. David Dotson, Ph.D.
11.1 Introduction 176
11.2 T1: From Discovery to Candidate Health Application 177
11.3 T2: From Health Application to Evidence-based Guidelines 180
11.4 T3: From Guidelines to Health Practice 182
11.5 T4: From Practice to Population Health Impact 184
11.6 Conclusion 186

Index **193**

The color plates can be found following page 96.