Contents

1 Introduction to Statistical Genetics and Background in Molecular Genetics 1
1.1 Basic Concepts in Genetic Disease 2
1.2 Review of Molecular Genetics 6
1.3 Types of Genetic Variants 9
1.4 Effects of Genetic Variants on Disease 12

2 Principles of Inheritance: Mendel's Laws and Genetic Models 15
2.1 Mendel's Experiments 15
2.2 A Framework for Genetic Models 19
2.3 The Biology Underlying Mendelian Inheritance 24
2.4 Exercises 28

3 Some Basic Concepts from Population Genetics 31
3.1 Estimation of Allele Frequencies 31
3.2 Population Substructure 33
3.2.1 Population Stratification 33
3.2.2 Population Admixture 34
3.2.3 Population Inbreeding 35
3.3 Hardy-Weinberg Equilibrium 36
3.3.1 Testing for HWE 38
3.3.2 Some Causes of the Failure of HWE 39
3.3.3 Measuring the Departure from HWE 41
3.4 Exercises 42

4 Aggregation, Heritability and Segregation Analysis: Modeling Genetic Inheritance Without Genetic Data 45
4.1 Preliminaries 46
4.2 Aggregation Analysis 48
4.2.1 Estimating Recurrence Risk Ratios 51
4.2.2 Further Simplifications 51
4.3 Heritability Analysis 54

4.4 Segregation Analysis 57
4.4.1 Segregation Analysis for Dominant Mendelian Diseases 58
4.4.2 Segregation Analysis for Recessive Mendelian Diseases 62
4.4.3 Summary 63
4.5 Exercises 63

5 The General Concepts of Gene Mapping: Linkage, Association, Linkage Disequilibrium and Marker Maps 67
5.1 Introduction 67
5.2 Genetic Markers and Marker Maps 72
5.3 Testing for Linkage or Association: Basic Concepts 75
5.4 A Formal Definition of Linkage Disequilibrium and Related Measures Used to Describe Linkage Disequilibrium 77
5.5 The Origin and Extent of LD in the Human Genome 81
5.6 The Human Genome and HapMap Projects 82
5.7 Exercises 84

6 Basic Concepts of Linkage Analysis 87
6.1 Basic Approach to Assessing Linkage Between Two Loci 88
6.2 The Direct Counting Method 90
6.3 The Interpretation of LOD Scores 94
6.4 Exercises 95

7 The Basics of Genetic Association Analysis 99
7.1 Testing Association with Dichotomous Disease Traits: Codominant, Recessive and Dominant Models 101
7.2 The Additive Genetic Model: The Alleles Test and the Trend Test 103
7.3 Small Sample and Permutation Tests 106
7.4 Which Mode of Inheritance Should We Assume for Testing? 107
7.5 Estimating Effect Sizes and Confidence Intervals 108
7.6 Examples of Testing Association with Diallelic Markers 109
7.7 The Regression Approach: Extensions to Covariate Adjustment and to Other Phenotypes 111
7.8 Association Analysis with Complex Traits: An Association Between INSIG2 and BMI 114
7.9 Sample Size and Power Considerations for Case-Control Design 116
7.10 Power and Effect Estimation: Testing a Marker in LD with the DSL 120
7.11 Exercises 122

8 Population Substructure in Association Studies 125
8.1 The Impact of Population-Admixture and Stratification on Genetic Association Tests 127
8.2 Genomic Control Approaches 132
8.3 Modeling the Effects of Population Admixture and Stratification 133
8.4 Regression-Based and Principal Component Approaches 133
8.5 Exercises 136

9 Association Analysis in Family Designs 139
9.1 The Trio Design and the TDT 139
9.2 Family Based Association Tests: FBAT 142
9.2.1 Missing Parents 145
9.2.2 Comparative Power for Family-Based and Case-Control Designs 147
9.3 Applications 148
9.3.1 Using FBAT to Obtain the TDT 149
9.3.2 Deriving a TDT for a Recessive Mode of Inheritance 150
9.3.3 Informative Families 150
9.3.4 Codominant Mode of Inheritance 151
9.3.5 Multiallelic Test 151
9.3.6 Using Unaffected Offspring 152
9.3.7 Missing Parental Information 153
9.3.8 Quantitative Traits 156
9.4 Exercises 158

10 Advanced Topics 161
10.1 The Multiple Testing Problem in Association Studies 161
10.1.1 Methods Based on P-Value Adjustment 161
10.1.2 Permutation and Monte Carlo Tests 164
10.2 Other Methods for the Analysis of Multiple SNPs, Including Haplotypes 165
10.3 Gene–Environment/Gene–Drug Interaction 170
10.4 Exercises 174

11 Genome Wide Association Studies 175
11.1 Introduction 175
11.2 Quality–Control for the Genotype Data 176
11.3 Multi-Stage Designs 182
11.4 Testing Strategies for Family-Based Studies 185
11.5 Replication, Non-replications and Meta-analysis 186
11.6 Exercises 189

12 Looking Toward the Future 191

A Basic Concepts of Linkage Analysis (Continued from Chapter 6) 193
A.1 General Issues with Parametric Linkage Analysis 193
A.2 Non-parametric Linkage Analysis 195
A.3 Multipoint Linkage Analysis 199

B A Class of Score Tests for Family Designs 203
Properties of the Score Test 204
Missing Parents 205

C The TDT Tests for Both Linkage and Association (LD) 207

Bibliography 211

Index 219