

Contents

Preface	v
Acknowledgments	viii
1 Genetic Diversity	1
1.1 Genetic Terminology	2
The use of the terms genes and alleles varies, though they do have precise definitions	2
1.2 Genetic Variation	5
Genetic variation can be measured by several methods	6
Alleles on the same chromosome are physically linked and inherited as haplotypes	7
Linkage disequilibrium promotes conservation of haplotypes in populations	8
1.3 Genetics and Evolution	9
Mutation is the major cause of genetic variation	10
Genetic variation caused by mutation alters allele frequencies in populations	11
Migration and dispersal cause gene flow	12
Allele frequencies can change randomly via genetic drift	13
The thrifty gene hypothesis	16
Natural selection acting on different levels of fitness affects the gene pool	16
1.4 Calculating Genetic Diversity: Determining Population Variability	19
Genotype and allele frequencies illustrate genetic diversity	19
Allele frequency refers to the numbers of alleles present in a population	20
Heterozygosity provides a quantitative estimation of genetic variation	21
The HWP is a complex but essential concept in population genetics	21
Calculating expected genotype frequencies using the HWP	22
Different populations may have different allele frequencies	22
1.5 Population Size and Structure	26
Breeding population size is important in evolution	26
Genetic variation is not always uniform in a population	26
Wahlund's principle	27
1.6 The Mitochondrial Genome	27
1.7 Gene Expression and Phenotype	29
Genetic variation is manifested in the phenotype	29
Phenotypes are influenced by the environment	29
1.8 Epigenetics	30

1.9	Genomic Imprinting	31
	Conclusions	31
	Further Reading	33
2	Defining Complex Disease	35
2.1	Definition of a Genetically Complex Disease	36
	To fully understand complex disease it is important to deconstruct this definition	36
2.2	Chromosomal Diseases	40
	Changes in chromosome number cause serious genetic diseases	40
	Changes in chromosome structure can cause serious illness	41
2.3	Mendelian Diseases	43
	Mendelian diseases involve a single gene and show simple patterns of inheritance	43
	Penetrance is an important difference between Mendelian and complex diseases	45
	Some diseases have both Mendelian and complex characteristics	47
	Modifier genes may also confuse the picture	47
	Mendelian traits can be studied in families	47
	There are complications to Mendelian diseases	48
2.4	Variation in The Mitochondrial Genome is Associated with Disease	50
	Variation in the mtDNA has been widely associated and linked with many different diseases	52
2.5	De Novo Mutations and Human Disease	53
2.6	Three Different Types of Complex Disease	54
	Studying complex disease is different from studying Mendelian disease	54
	Monogenic complex diseases involve a single risk allele	55
	Oligogenic complex diseases involve several alleles	56
	Polygenic complex diseases involve many risk alleles	56
2.7	Alzheimer's Disease May be a Monogenic Complex Disease	57
2.8	HSCR – An Oligogenic Complex Disease	58
	Sporadic HSCR illustrates the oligogenic model for complex disease	58
2.9	Crohn's Disease is Mostly a Polygenic Complex Disease	61
	Early studies of Crohn's disease suggested a number of locations for risk alleles	61
	Genetic variations in the human equivalent of the plant <i>nod2</i> gene (<i>CARD15</i>) were the first identified and confirmed Crohn's disease risk alleles	62
	Genetic variations in other immune regulatory genes are important risk factors in Crohn's disease	64
	The Wellcome Trust Case Control Consortium (WTCCC1): Crohn's disease	64
	The current number of risk alleles for Crohn's disease may be as high as 163	66
2.10	Applying Disease Models to Populations	67
	Conclusions	67
	Further Reading	69
3	How to Investigate Complex Disease Genetics	73
3.1	Planning Stage 1: Gathering the Basic Knowledge	73
	Incidence and prevalence are measures of how common a disease is	74

Incidence and prevalence can be very different or very similar depending on the prognosis for the disease	75
Incidence and prevalence of disease may vary in different populations	76
What is the evidence for a genetic component to the disease?	76
What is known about the disease pathology?	80
Before we get down to the hard business of study planning there are one or two other questions that it is important to ask	82
3.2 Planning Stage 2: Choosing a Strategy	84
Two basic strategies for identifying risk alleles in complex disease	84
In terms of the history of genetic studies in complex disease there are two main periods: pre- and post-genome	84
Each of these two strategies has a substrategy	88
3.3 Good and Bad Practice	93
Accurately identifying true disease susceptibility alleles in GWAS (and other association studies) is dependent on sample size	93
Case selection can introduce bias into a study	94
It is important to consider whether we are studying a disease, a syndrome, or a trait within a disease subgroup	94
Selection of appropriate controls is equally important in any study	94
Errors in the laboratory and in sample handling can also introduce bias into a study	96
Statistical analysis is the key in any study of complex disease	96
SNP chip selection is an important factor to consider in study design	96
Unfortunately publication bias does occur	97
Replication in an independent sample is crucial for all association studies, especially GWAS	98
3.4 New Technologies and the Future	100
The technological advances of the past decade have had a major impact on research into the genetics of complex disease and the rate of change is going to increase	100
New developments will come from the ENCODE project, and will also involve more epigenetics and imputation analysis	100
The real debate about the future of complex disease research lies not in the genetics itself, but downstream from the genetics	101
Conclusions	101
Further Reading	103
4 Why Investigate Complex Disease Genetics?	105
4.1 Why Do We Investigate Complex Disease?	106
Complex diseases do not conform to simple patterns of inheritance	106
The HGP in research into genetically complex disease	107
4.2 Disease Diagnosis	108
Early studies on the genetics of ankylosing spondylitis indicated what could be achieved in terms of differential diagnosis in the post-genome era	108
Genetic associations in complex disease confer small risks	110
4.3 Patient Treatment/Management and Care	110
Identifying risk alleles that predict onset of complex diseases may enable patients to make beneficial lifestyle changes	111

	Predicting disease severity through genetic analysis may have clinical significance in terms of patient management	111
	Common genetic variation may predict response to treatment and be critical in patient care	113
	Onset, severity, and response to treatment are all part of patient management	113
4.4	Disease Pathogenesis	114
	Early studies offered potential insight into the biology of ankylosing spondylitis	115
	Later GWAS have offered even further insight into the biology of ankylosing spondylitis	116
	Rheumatoid arthritis has many strong genetic associations, some of which can be used to help us unravel the pathogenesis of this disease	116
	Bipolar disease is a disease for which there are many weak genetic associations, but few strong consistent associations	122
	Coronary artery disease is the most common cause of death in the developed world	127
4.5	What about the Other Diseases?	136
	Conclusions	137
	Further Reading	138
5	Statistical Analysis in Complex Disease: Study Planning and Data Handling	141
5.1	Linkage Analysis	142
	The LOD score is a measure of significance of linkage between a trait and a marker allele	144
5.2	The Basic Statistical Concepts of Association Analysis and their Application in Study Design	145
	In statistical terms, there are two different hypotheses to consider in the analysis of genetic association studies: a null hypothesis and an alternative hypothesis	146
5.3	Statistical Error, Power, and <i>P</i> Values	146
	Making the right decision and avoiding errors in the hypothesis testing	146
	The likelihood of detecting a significant difference in an association study is directly related to sample size	147
	Probability (<i>P</i>) values are simply statements of the probability that the observed differences between two groups could have arisen by chance	148
5.4	The Basic Statistical Considerations for Analysis of Case Control Association Studies and their Application to Data Collection and Analysis	151
	Departures from HWE can have different causes	151
	Pearson's χ^2 and Fisher's exact test are used to assess the departure from the null hypothesis	152
	Fisher's exact test calculates the exact probability (<i>P</i>) of observing the distribution seen in the contingency table	154
	The Cochran–Armitage test looks for a trend for a difference between cases and controls across the ordered genotypes in the table	155
	There is no simple answer to the question of which test to choose	157
	Data may also be analyzed assuming a predefined genetic model	157
	Logistic regression is frequently used in association studies	160
	The pitfalls and problems of GWAS	162
5.5	How to Interpret a GWAS	165
	There are several ways to interpret statistically significant genetic associations	165
	There are several diagnostic plots that can be used for the visualization of genome-wide association results	165

Linkage disequilibrium is a useful tool in association studies provided you know how to handle it	167
The ability to detect a significant association through linkage disequilibrium can increase the power of an association study	167
Most association analyses identify multiple SNPs, other genetic variants, and haplotypes	170
Conclusions	171
Further Reading	172
6 The Major Histocompatibility Complex	175
6.1 Histocompatibility	176
The idea of histocompatibility first started with blood groups	176
The MHC-encoded HLA antigens are the second major histocompatibility group	176
Naming the HLA antigens and alleles up to and including the early molecular genotyping era	177
The current naming system for HLA alleles and genes allows for a greater level of resolution to be reported	183
The MHC encodes a cornucopia of genetic diversity within the HLA genes	184
Comparing the levels of genetic diversity at DR with those at DQ can make DQ look like a poor relation	185
HLA class II molecules can be expressed in <i>trans</i> or in <i>cis</i>	187
The final groups of genes that need to be considered are those called pseudogenes, gene fragments, and null alleles	188
6.2 The Extended Human MHC MAP	189
6.3 Molecular Structure of HLA Class I and Class II	191
X-ray crystallography of HLA-A2 revealed the full structure and much about the function of HLA class I	191
The X-ray crystallography structure of HLA class II structure revealed the critical difference between class I and class II	192
6.4 Immune Function of HLA Class I and Class II	193
Class I molecules have distinct features	193
HLA class II is different to class I	193
HLA class I and class II have important similarities	194
HLA class I and antigen engagement in the cell is different from HLA class II	194
HLA class II and antigen engagement in the cell is different	195
6.5 HLA Class I and Disease	196
Hemochromatosis is an example of a Mendelian disease which maps within the xMHC	196
Psoriasis proves the point that <i>HLA-C</i> is an important locus to consider in genetic studies of the MHC	196
Type I versus type II psoriasis	197
Before we leave HLA class I we need to consider Bw4 and Bw6	197
6.6 HLA Class II and Disease	197
Severe or cataplectic narcolepsy has one of strongest HLA associations ever reported	197
There are different functional interpretations of the HLA association with narcolepsy	198
Multiple sclerosis is a disease with a strong genetic association with HLA class II	199
HLA class II and autoimmune liver disease	201

AIH is a relatively rare classical autoimmune disease of the liver	202
PSC is not a classical autoimmune disease	207
PBC is an autoimmune liver disease with a genetic component	210
6.7 Comparing the HLA Associations of the Three Liver Diseases	213
6.8 Non-HLA MHC Genes and Disease	213
The MHC class III region complement, <i>MICA</i> , and <i>TNFA</i> genes in complex disease	214
6.9 A Single Gene or a Risk Portfolio	216
A single gene may explain MHC- encoded genetic susceptibility to disease	216
Alternatively there is always room for a second bite of the cherry: a multihit hypothesis	217
6.10 How to Compare and Critically Evaluate Contrasting Studies	218
Knowing history is important when we critically review and design studies	218
Conclusions	219
Further Reading	221
7 Genetics of Infectious Disease	223
7.1 The Infection Process and Disease	223
Mechanisms of infection vary widely but common steps in the process can be identified	224
The immune response combats infectious disease	224
Individuals infected by the same pathogen may experience different outcomes	225
7.2 Heritability of Resistance and Susceptibility to Infectious Disease	225
Different populations infected by the same pathogen may experience different outcomes	225
Leprosy and tuberculosis were once believed to be inherited diseases	226
Adoption studies indicate that susceptibility to infectious disease has a heritable component	227
Rare monogenic defects in immunity can cause primary immune deficiencies	227
7.3 Identifying Alleles that Affect Risk of Susceptibility and Resistance to Infectious Disease	228
Risk alleles can be identified using a hypothesis-driven or genome- wide approach	228
The outcome of infectious disease being tested must be clearly defined	229
7.4 Malaria	229
The life cycle of the <i>Plasmodium</i> protozoa is complex	229
Hemoglobinopathies confer resistance to malaria	230
Haldane's malaria hypothesis proposed that thalassemia confers protection against malaria	232
Allison demonstrated that sickle cell trait confers resistance to <i>P. falciparum</i>	232
Studies on Pacific Island populations provided experimental evidence that thalassemia confers protection from malaria	233
The mechanism of resistance to malaria conferred by hemoglobinopathies is still not fully understood	233
Resistance to malaria conferred by HbS and thalassemia is a complex genetic trait	235
Other malaria resistance alleles have been identified via epidemiological or hypothesis-driven studies	235
GWAS suggest that polymorphisms in immunity-related genes may affect outcome of <i>Plasmodium</i> infection	235
GWAS searching for malaria resistance alleles highlight the challenges of GWAS in African populations	235

7.5	HIV-1	237
	C-C chemokine receptor 5 (CCR5) acts as a co-receptor for HIV-1 in the early stages of infection	237
	Some individuals are naturally resistant to HIV infection	238
	A 32-bp deletion in the <i>CCR5</i> gene confers resistance to HIV-1 infection	239
	Selection pressure by HIV-1 cannot account for the high frequency of <i>CCR5-Δ32</i> in the northern European population	240
	<i>CCR5-Δ32</i> affects the outcome of infection by West Nile virus	240
	<i>CCR5-Δ32</i> cannot account for all HIV-1 resistance	241
	CCR5 promoter polymorphisms affect HIV-1 control	242
	CCR5/CCR2 haplotypes have a complex effect on HIV-1 control	242
	Polymorphisms in chemokine receptor ligand genes influence HIV-1 control	244
	Polymorphisms in HLA genes affect outcome of HIV infection	245
	HLA class I homozygosity is not always bad news	246
	GWAS confirms the protective role of <i>HLA-B</i> in HIV-1 infection	247
	Amino acids in the HLA-B binding groove are associated with HIV-1 control	248
	GWAS revealed, for the first time, association of HLA-C with HIV-1 control	248
	Some SNPs previously implicated in HIV-1 control have not yet been confirmed by GWAS	249
	Conclusions	249
	Further Reading	251
8	Pharmacogenetics	253
8.1	Definition and a Brief History of Pharmacogenetics	254
8.2	Cytochrome P450	255
	There is a clear relationship between genotype and phenotype for several forms of cytochrome P450	255
	The conversion of the analgesic drug codeine, which is administered as a pro-drug and is activated to morphine by CYP2D6, is of clinical importance	258
	The cytochrome P450 CYP2C9 metabolizes warfarin – a very widely used drug	259
	CYP2C19 activates clopidogrel – a drug widely used to prevent strokes and heart attacks	259
8.3	Other Drug-Metabolizing Enzymes and Transporters	261
	For phase II conjugation reactions, the UDP glucuronosyltransferase family makes the largest contribution	261
	Methyltransferases are also important in phase II drug metabolism	261
	Polymorphisms in drug transporters also play a role in pharmacogenetics	263
8.4	Drug Targets	263
	The relationship between VKOR and coumarin anticoagulants is one of the most consistently reported genetic associations involving drug targets unrelated to cancer	263
	The efficacy of β -adrenergic receptor agonists widely used in the treatment of allergies may also be genetically determined	264
8.5	Adverse Drug Reactions	266
	HLA genotype is a potent determinant of susceptibility to several different types of adverse drug reactions	266

The anti-human immunodeficiency virus (HIV-1) drug Abacavir gives rise to hypersensitivity in some patients	267
Drug-induced liver injury is a rare, but clinically important problem	267
There are many other susceptibility factors for serious adverse drug reactions	269
Adverse reactions to commonly used statins provide a key example of non-HLA-related adverse drug reactions	270
Cardiotoxicity reactions to drugs do not appear to involve an immune or inflammatory response	270
Conclusions	271
Further Reading	273
9 Cancer as a Complex Disease: Genetic Factors Affecting Cancer Susceptibility and Cancer Treatment	275
9.1 Defining Cancer	278
9.2 Cancer as a Complex Disease	280
Early studies of cancer found evidence of genetic associations with risk	280
GWAS has revolutionized the search for cancer-promoting alleles in non-familial cancers	281
9.3 Genetic Risk Factors for Particular Cancers Detected by GWAS	281
GWAS has identified a number of biologically plausible genetic risk factors for breast cancer	281
Novel insights into lung cancer involving the target for nicotine were detected by GWAS	283
A large number of genetic risk factors for prostate cancer have been revealed by GWAS	283
9.4 General Cancer Risk Loci Detected by GWAS	285
9.5 Previously Established Cancer Risk Factors Confirmed by GWAS	286
Alcohol, smoking, and chemical exposure increase the risk of cancer	286
9.6 Individualizing Drug Treatment Based on Tumor Genotype	288
Newly developed drugs inhibit the function of mutated proteins in cancer cells	288
Specific antibodies can target tumor-specific proteins and inhibit tumor growth	289
Epigenetic changes in the tumor involving methylation may affect response to conventional drug treatments	290
Gene expression profiling may enable personalized cancer treatment	290
Before we close the chapter on cancer it is important to recognize that there are many forms of this disease	291
Conclusions	292
Further Reading	294
10 Genetic Studies on Susceptibility to Diabetes	295
10.1 Diabetes Mellitus	295
10.2 Genetics of T1D	297
10.3 Early Genetic Studies in T1D	297
HLA class II genotype is the strongest genetic risk factor for T1D	298
Not all of the risk for T1D above may be associated with the <i>DQB</i> allele or HLA class II	299
Other genetic risk factors for T1D include the genotype for the insulin gene	300
Candidate gene studies have identified a number of other non-MHC associations with T1D	300

10.4	GWAS Studies in T1D	303
	The 2007 WTCCC1 study was one of the first GWAS in T1D	303
	Following the introduction of GWAS in 2007, research has resulted in the identification of at least 40 further potential T1D alleles	305
10.5	Early Genetics of T2D	306
	There have been different interpretations of the associations with <i>PPARG</i> , <i>KCNJ11</i> , and <i>TCF7L2</i>	307
10.6	GWAS Studies in Type T2D	307
	Examples from the WTCCC1 study	308
	Other risk alleles for T2D from other studies	309
10.7	The Future of Genetics in T2D	310
	Future prospects in T2D research involve genome sequencing	310
	Epigenetics may be important in diabetes	310
10.8	Genetics of Monogenic Diabetes	311
	Conclusions	312
	Further Reading	314
11	Ethical, Social, and Personal Consequences	315
11.1	Defining Ethics	316
	There are philosophical arguments for and against ethical constraint in biomedical research	316
	What are the practical ethical implications in the study of genetics of complex disease?	317
11.2	Ethics in Genetics: What We Can Learn from the Past?	318
	The consequence of the Eugenics Movement and the ideas it spread were extremely bad news for the developing science of genetics	318
11.3	Looking into the Future Use of Genetic Data	321
	Genetic studies of complex disease will have a major impact on clinical medicine	321
	The potential personal impact of data from studies in complex disease is considerable	322
11.4	Who Does the Data Belong to? Interacting with Commerce	329
	Do I own my genome?	330
11.5	Who Should be Able to Access the Data?	332
	Conclusions	332
	Further Reading	334
12	Sequencing Technology and the Future of Complex Disease Genetics	337
12.1	DNA Sequencing: The Past, Present, and Future	338
	The development of DNA sequencing using the Sanger sequencing technique opened the way to sequencing the genome	338
	The new era: next-generation DNA sequencing	340
	The upcoming era: third-generation sequencing	343
12.2	The Future of NGS in Clinical Practice and Research	347
	Using NGS will enable high- resolution genotyping for SNPs in complex disease	348
	Using NGS will enable better identification of CNVs	350
	Sequencing the RNA transcript and the whole transcriptome is an alternative way forward	350

12.3	Whole-Genome Versus Exome Sequencing	350
12.4	The Next Generations of Genome/Exome-Wide Association Studies	351
	Missing and non-genotyped SNP data can be imputed using large databases and known patterns of linkage disequilibrium	352
	GWAS identifies both synthetic (false) associations and direct (real) associations	353
	The importance of linking genotype to phenotype	353
	Genotyping on new arrays provides a focus and higher level of resolution for GWAS	354
	Different forms of NGS technology will impact on how GWAS is used	354
12.5	Epigenetics: A Complimentary Strategy in Complex Disease Studies	357
12.6	Metagenomics and the Bacterial Genome	357
	To put metagenomics into context, we need to consider the impact it may have	359
12.7	Major Ongoing International Genome Projects	360
	HapMap is a project with major significance in current research, especially GWAS	360
	The 1000 Genomes Project has major potential in studies of complex disease	362
	ENCODE will help to link genotype to phenotype in complex disease	363
12.8	Systems Biology	364
	Considering systems biology allows us to look into the future	364
	Conclusions	366
	Further Reading	368
	Glossary	373
	Index	397
	Color Inserts	