HRM 728: Genetic Epidemiology & Biostatistics
Fall – 2019
9:00 am – 12:00 pm

Class Room: HSC 1J10

Course Coordinator: Dr. Sonia Anand
Office**:** MDCL Rm. 3204
Phone**:** 905-525-9140 Ext. 21523
Email: anands@mcmaster.ca

Other Contacts for the Course:
Logistics**:** Mrs. Kathy Stewart
Phone**:** 905-525-9140 Ext. 21523
Email**:** kmstew@mcmaster.ca

PLINK and Dataset help for Students:
Logistics: Dr. Amel Lamri
Email: amel.lamri@phri.ca

Course Material:
Reference Text Book: A Statistical Approach to Genetic Epidemiology: Concepts and Applications (Hardcover), by Andreas Ziegler (Author), Inke R. Koenig (Author)

Courseware: Links to all reference articles in the course outline have been posted on avenue.

Student Evaluation:

- Class Participation **15%** - student's evaluation is based on their class participation. In order to be consistent between instructors, a student evaluation sheet has been provided to you. Please complete this and forward to Kathy Stewart kmstew@mcmaster.ca

- Mid-term Assignment **25%** - student's evaluation will be based on a mid-term assignment (5-page review of an area in genetic epidemiology) that the student will complete and hand over on **Friday, October 11th 2019.**

- Independent Study **35%** - student's evaluation will be based on an independent study (analysis and interpretation of publicly available dataset) presentation done during the last session of the course **(Friday, December 13th, 2019)**

Final Exam = **25%**

HRM 728 (Fall 2019) | Room List & Course Timetable

Last Updated: August 21, 2019 **Health Research Methodology Program**

Course Title:	Genetic Epidemiology & Statistics	Dates:	Fridays | September 2019 to December 2019
Course Coordinator:	Sonia Anand	Time:	9:00am – 12:00pm

Date	Lecture Room (Room booked 8:30-12:30)	Session	Instructor	Content
Sept 6	HSC 1J10	1	Sonia Anand/Amel Lamri	Key concepts in genetic epidemiology
Sept 13	HSC 1J10	2	Amel Lamri	Key concepts in genetic epidemiology (Part II)
Sept 20	HSC 1J10	3	Russell de Souza	Gene diet interactions
Sept 27	HSC 1J10	4	Mark Loeb	Genetic association studies and GWAS
Oct 4	HSC 1J10	5	David Meyre	Analytical challenges in genetic association
Oct 11	HSC1J10	6	Joseph Beyene	Statistical issues in genetic studies
Oct 11	HSC 1J10	6	**Mid-Term Assignment Due in Class**	
Oct 18	HSC 1J10	7	David Meyre	Functional genomics – review of gene expression and molecular biology techniques and interpretation
Oct 25	HSC 1J10	8	Guillaume Paré	Betwixt and between; common and rare variants in human disease
Nov 1	HSC 1J10	9	Amel Lamri	Quality control of Genetic data using PLINK
Nov 8	HSC 1J10	10	Guillaume Pare	Epigenetics and the potential influence of epigenetic variation on the occurrence of phenotypic characterization and disease causation
Nov 15	HSC 1J10	11	Zena Samaan	The application of genetic studies in specific disorders: psychiatric genetics
Nov 22	HSC 1J10	12	Kevin Zbuk	Pharmacogentics
Nov 29	HSC 1J10	13	Amel Lamri/Russell de Souza	Wrap up Session – First hour final exam, hours 2-3 wrap up
Dec 13	HSC1J10	14	Sonia Anand	Final Assignment – Independent Study Presentations

Target Students: HRM Graduate Students or other Health Sciences Students
This will be a half course offered through the department of CE&B comprised of 12-13 seminars.

Course Evaluation:

- Based on Participation and Completion of course assignments.
- The mid-term assignment should be a research protocol that outlines:
    - background
    - hypothesis
    - objectives
    - design and Methods of study
    - Proposed quality control checks and analysis
    - Potential for Bias
    - anticipated challenges
    - References - up to 12.
- The final assignment involves the analysis and interpretation of a publicly available genetics dataset.

Weeks

1. Key concepts in genetic epidemiology
Presenter: Dr. Sonia Anand and Dr. Amel Lamri

Reference Paper:
Anand SS, Meyre D, Pare G, et al. Genetic Information and the Prediction of incident Type 2 Diabetes in a High- Risk Multi-Ethnic population: The EpiDREAM Genetic Study. Diabetes Care 2013 Sep; 36(9):2836-42.

Brief Summary:

Classical epidemiology deals with disease patterns and factors associated with disease causation, with the ultimate aim of preventing disease. Molecular epidemiologic studies measure exposure to specific substances (DNA adducts) and early biological responses (somatic mutations). They evaluate host characteristics (genotype and phenotype) that mediate responses to external agents. Furthermore, they use markers of specific effects (like gene expression) to refine disease categories such as heterogeneity, etiology and prognosis.

Genetic epidemiology overlaps with molecular epidemiology. It is the epidemiological evaluation of the role of inherited causes of disease in families and in populations; it aims to detect the inheritance pattern of a particular disease, localize the gene, and find a marker associated with disease susceptibility. Gene-gene and gene-environment interactions are also studied in genetic epidemiology of a disease. Genetic epidemiology is “a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations” (Morton NE, 1982).

In this introductory seminar, the central concepts and issues in modern genetic epidemiology will be reviewed. Students will be provided with an overall framework for investigating the role of genetic determinants and the causation of complex diseases such as type 2 diabetes.
The framework will outline the integration of modern genetics and population science, which came together to form the field of genetic epidemiology. Students will be provided with the basic concepts and vocabulary needed to understand discoveries of methods and error in the field of genetic epidemiology.

2. Key concepts in genetic epidemiology (Part II)
Presenter: Dr. Amel Lamri

Reference Papers:
Manolio TA et al. Opportunities, resources, and techniques for implementing genomics in clinical care. Lancet. 2019 Aug 10;394(10197):511-520.

The 1000 genomes Project Consortium. A global reference for human genetic variation. Nature 2015 Oct 1;526(7571):68-74

Brief Summary:

In this seminar, students will be introduced to more advanced concepts in the field of genetic epidemiology and genomic medicine including sequencing and genotyping technologies, GWAS meta-analysis, polygenic risk scores, and Mendelian randomizations. Practical examples will be given throughout the course to illustrate each concept. Students will also review examples of common mistakes that might bias the results of genetic association tests and how to deal with genotyping errors, genotype calling errors, misuse of genome builds and strand issues.

3. Gene diet interactions
Presenter: Dr. Russell de Souza

Reference Papers:
Langenberg C, Sharp C, et al. Cene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. PloS Med 11(5)

Mattei J, Qi Q, et al. TCF7L2 genetic variants modulate the effect of dietary fat intake on changes in body composition during a weight-loss intervention. Am J Clin Nutr 2012;96:1129–36.

Ottman R. Theoretical Epidemiology. Gene–Environment Interaction: Definitions and Study Designs. PREVENTIVE MEDICINE 25, 764–770 (1996)

Brief Summary:

Dr. Russell de Souza will discuss gene-diet interactions. The overview will begin with a discussion aimed at appreciating how difficult it is to ascertain whether diet causes a disease using current approaches. This will provide the motivation for our discussion of why it is

important to study gene-diet interactions, and a review of the concept of interaction. We will review some "exemplar" studies of gene-diet interactions, and assess their methodological strengths and limitations, and end with a discussion of public health implications of gene-diet interaction studies.

4. Genetic association studies and GWAS
Presenter: Dr. Mark Loeb

Reference Papers:
Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex disease. Lancet 2001 Oct 20;358(9290):1356-1360.

Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence
complex diseases. Nat.Genet. 2005 Apr;37(4):413-417.
(A study of inter-genetic interactions both gene to environmental interactions and gene-gene interactions.)

Loeb M, Eskandarian S, Rupp M, Fishman N, Gasink L, Patterson J, Bramson J, Hudson T, Lemire M. Genetic Variants and Susceptibility to Neurological Complications Following West Nile Virus Infection. The Journal of Infectious Diseases 2011;204:1031-7

Brief Summary:

Genetic epidemiology of complex diseases
The term "complex trait/disease" refers to any phenotype that does not exhibit classical Mendelian inheritance attributable to a single gene; nevertheless, they may exhibit familial tendencies (familial clustering, concordance among relatives). The contrast between Mendelian diseases and complex diseases involves more than just a clear or unclear mode of inheritance. In Mendelian diseases, the risk to relatives decreases by a factor of ½ with each degree of relationship (from first to second to third degree) but in complex diseases the risk decreases more rapidly (Risch, 1990a). Other hallmarks of complex diseases include known or suspected environmental risk factors; seasonal, birth order, and cohort effects; late or variable age of onset; and variable disease progression. Many complex diseases are hard to diagnose accurately; even quantitative traits such as hypertension often involve sizable measurement errors (Guo, 2000b). Ultimate analysis of complex traits requires sophisticated statistical designs incorporating all genetic and nongenetic variables, their interactions, and familial correlations. In general, linkage is harder to show in a complex disease than a Mendelian disorder (Risch, 1992). A complex disease can be modeled in two different ways: (1) an additive model, closely approximates genetic heterogeneity, is characterized by no interlocus interaction, and (2) a multiplicative model, representing epistasis (interaction) among loci (Risch, 1990a).

Common susceptibility alleles in rare complex diseases
One popular hypothesis proposes that the genetic factors underlying common diseases will be alleles that are themselves quite common in the population at large (Lander, 1996; Chakravarti, 1999); Pritchard, 2001). When several different loci contribute to a phenotype (such as a complex

disease), it is likely that the alleles at loci responsible for such interactions have high frequencies in populations (Carlson, 2004). If, for example, six genes contribute equally to a disease with an incidence of 1.5%, each susceptibility allele must have a population frequency of around 50%. Thus, modest-risk gene variants involved in polygenic diseases are often likely to be normal alleles from unsuspected loci that have relatively high frequencies (Reich, 2001). The identification of normal polymorphisms is of great importance for medical genetics (Cavalli-Sforza, 1998). However, rare coding region alleles are commonly deleterious and their contribution to the development of complex diseases is obvious (Kryukov et al, 2007; see also Ropers, 2007).

Concepts reviewed in this Seminar include the importance of study design and potential threats to validity including sample recruitment and selection, genotyping error, potential errors in data analysis, the importance of replication, and population structure will be reviewed.

5. Experimental and analytical challenges in genetic association studies
Presenter: Dr. David Meyre

Reference Paper:
Li A, Meyre D. Challenges in reproducibility of genetic association studies: lessons from the obesity field. International Journal of Obesity (2013) 37, 559-567

Brief Summary:

A robust replication of initial genetic association findings has proved to be difficult in human complex diseases. An obvious cause of non-replication in genetic association studies is the initial report of a false positive result, which can be explained by a non-heritable phenotype, insufficient sample size, improper correction for multiple testing, population stratification, technical biases, insufficient quality control or inappropriate statistical analyses. Replication may, however, be challenging even when the original study describes a true positive association. The reasons include underpowered replication samples, gene × gene, gene × environment interactions, genetic and phenotypic heterogeneity and subjective interpretation of data. We will address classic pitfalls in genetic association studies and provide guidelines for proper discovery and replication genetic association studies.

6. Statistical issues in genetic studies
Presenter: Dr. Joseph Beyene

Reference Paper:
Hattersley AT, McCarthy MI. What makes a good genetic association study? Lancet 2005 Oct 8;366(9493):1315-1323.
(A review of the potential bias in genetic association studies)

Brief Summary:

Association studies focus on population frequencies, whereas linkage studies focus on concordant inheritance. One may be able to detect linkage without association when there are many independent trait-causing chromosomes in a population (i.e., no LD of the disease causing allele

to a specific marker nearby). Association without linkage is observed when an allele explains only a minor proportion of the variance for a trait, so that the allele may occur more often in affected individuals but does a poor job of predicting disease status within a pedigree (Lander & Schork, 1994). Association is usually with a "susceptibility" locus, which increases the probability of contracting the disease but is not "necessary" or "sufficient" for disease expression. In this case, the marker will not show linkage in families. If an association is, however, with a marker in LD with a "necessary" locus for disease development, then there will be evidence for linkage in family data (Greenberg, 1993; Greenberg & Doneshka, 1996). Linkage analysis is not useful for finding loci that are neither necessary nor sufficient for disease expression (so-called susceptibility loci).

Association studies have several practical advantages over linkage studies. As opposed to linkage studies, families with multiple affected individuals are not required and no assumptions are made about the mode of inheritance of the disease. In addition, association studies have considerable statistical power to detect genes of weak effects unlike linkage studies in families (Risch, 1996; Morton, 1998; Risch, 2000). Most significant factors independently associated with increased success in linkage studies are (a) an increase in the number of individuals studied and (b) study of a sample drawn from only one ethnic group (Altmuller, 2001). For association studies, large datasets, small *P* values and independent replication of results are important for reproducible results (Editorial, Nat Genet 1999; Dahlman, 2002). Use of ancestral haplotype groups in association studies (evolutionary-based association study design) is another way to increase power (Templeton, 1987; 1995; 2000; Schork, 1998; Seltman, 2003; Fejerman, 2004; Tzeng, 2005).

Possible ascertainment problems in case-control studies in genetic epidemiology:

- The sample should be representative of all cases. Inclusion of those identified at a hospital clinic may or may not be appropriate. They should be unrelated, incident (as opposed to prevalent) and consecutively diagnosed ones. If the prevalence of the disease is known, this would give an idea for the completeness of ascertainment (for a rare disease).
- If the disease requires medical attention only in some cases, recruitment from a hospital will be selective (usually for severe cases).
- If there is a survival effect of the disease (as in Alzheimer's disease and ApoE), and if the associated allele also modifies the risk of death from competing causes, the age-dependent frequencies will be different. In this case, age-matching of cases and controls becomes particularly important.
- The controls should be comparable to cases except for having the disease. Local, contemporary controls should be selected via the same routes as the cases. For relevant diseases, age- and sex-matching (by frequency or one-to-one) may be important. Self-selected controls, such like volunteer marrow or blood donors, are not ideal. Controls that are truly population-based controls are more acceptable.

The pitfalls of conventional epidemiologic studies, such as selection bias, information bias, and confounding, apply equally to molecular epidemiological research (Vineis & McMichael, 1998; Campbell, 2002; Boffetta, 2003). Association studies also face additional problems unique to genetic studies (Olson, 2000; Cordell, 2000; Elbaz & Alperovitch, 2002; Lee & Ho, 2003; Morimoto, 2003; Potter, 2003). In genetic association studies, missing data may be distributed differentially between cases and controls and may generate spurious associations (Clayton, 2005).

This bias may be due to having subsets of DNA samples extracted using different chemistries that influence the performance of the assay differentially. See *Pitfalls in Genetic Association Studies*.

One potential problem is that estimates of genetic effect are subject to confounding when cases and controls differ in their ethnic backgrounds (population stratification bias or confounding by ethnicity). This can occur when both disease risk and genetic mutation frequencies vary among ethnic groups (Thomas, 2002; Wacholder, 2002; Cardon, 2003). To avoid the problem of population stratification bias, matching cases to controls on ethnic background, stratification, multidimensional scaling, family-based association studies or genomic controls (Devlin, 1999; Pritchard, 1999) can be used.

7. An introduction to functional genomics and systems
Presenter: Dr. David Meyre

Reference Paper:
Ichimura A, Hirasawa A, et al. Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. Nature Vol 483, March 2012;350

Brief summary:

Functional genomics integrates information from various molecular methodologies to gain an understanding of how DNA sequence is translated into complex information in a cell (DNA → RNA → Proteins → biological process). Systems biology is the computational and mathematical modeling of complex biological systems. We will describe the experimental tools and methods available to understand the complex relationship between genotype and phenotype.

8. Betwixt and between; common and rare variants in human disease
Presenter: Dr. Guillaume Pare

Reference Paper:
Lek, Karczewski K, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 18 August 2016;Vol 536:285-292

Brief Summary:

Strong evidence suggests that rare mutations of severe effect are responsible for a substantial portion of complex human disease. Evolutionary forces generate vast genetic heterogeneity in human illness by introducing many new variants in each generation. Current sequencing technologies offer the possibility of finding rare disease-causing mutations and the genes that harbour them.

The steps in genetic epidemiologic research are:

1. Establishing that there is a genetic component to the disorder.
2. Establishing the relative size of that genetic effect in relation to other sources of variation in disease risk (environmental effects such as intrauterine environment, physical and chemical effects, behavioral and social aspects).

3. Identifying the gene(s) responsible for the genetic component.

All of these can be achieved either in family studies (segregation, linkage, association) or in population studies (association).

General methods employed in genetic epidemiology:

Genetic risk studies: What is the contribution of genetics as opposed to environment to the trait? Requires family-based, twin/adoption or migrant studies.

Segregation analyses: What does the genetic component look like (oligogenic: "few genes each with a moderate effect", polygenic: "many genes each with a small effect" etc.)? What is the model of transmission of the genetic trait? Segregation analysis requires multigenerational family trees preferably with more than one affected member.

Linkage studies: What is the location of the disease gene(s)? Linkage studies screen the whole genome and use parametric or nonparametric methods such as allele sharing methods (affected sibling-pairs method) with no assumptions on the mode of inheritance, penetrance or disease allele frequency (the parameters). The underlying principle of linkage studies is the co-segregation of two genes (one of which is the disease locus).

Association studies: What is the allele associated with the disease susceptibility? The principle is the coexistence of the same marker on the same chromosome in affected individuals (due to linkage disequilibrium). Association studies may be family-based (transmission / disequilibrium test - TDT; also called transmission distortion test) or population-based. Alleles, haplotypes or evolutionary-based haplotype groups may be used in association studies (Clark, 2004; Tzeng, 2005). More recently, genome-wide association studies (GWAS) have become possible (Clark, 2005; Wang, 2005; Pearson, 2008; McCarthy, 2008; WTCCC GWAS; recent GWAS in OEGE). The samples needed for these studies may be nuclear families (index case and parents), affected relative pairs (sibs, cousins, any two members of the family), extended pedigrees, twins (monozygotic and dizygotic) or unrelated population samples.

This seminar will discuss methods of design and analysis of genetic association studies and review the rationale behind them. The similarities and differences between genetic association studies and classical epidemiologic studies of environmental risk factors will be highlighted. Association studies differ from linkage studies and these differences will also be reviewed. Issues of design, statistical analysis and interpretation will be discussed.

9. Quality control of genetic data using PLINK
Presenter: Dr. Amel Lamri

Reference Paper*:*

Anderson CA1, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564-73.

Brief Summary:
In this session, students will learn to perform basic quality controls for genetic data using the PLINK software. These will include: Identification of individuals with discordant sex information; Identification of individuals with elevated missing data rates or outlying heterozygosity rate; Identification of duplicated or related individuals; Identification of individuals of divergent ancestry and ethnic outliers; Identification of all markers with an excessive missing data rate; test markers for different genotype call rates between cases and controls; Identification of markers that fail Hardy-Weinberg equilibrium test.

10. Epigenetics and the potential influence of epigenetic variation on the occurrence of phenotypic characterization and disease causation
Presenter: Dr. Guillaume Pare

Reference Paper:
Suzuki MM, Bird A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature Reviews, Genetics. 9, June, 465-476.

Brief Summary:

DNA methylation and chromatin remodeling
Because the phenotype of a cell or individual is affected by which of its genes are transcribed, heritable transcription states can give rise to epigenetic effects. Gene expression is regulated on several levels. One way that genes are regulated is through chromatin remodeling. If the way that DNA is wrapped around the histones changes, gene expression can change as well. Chromatin remodeling is initiated by one of two methods:

1. **Post translational modification of histone amino acids:** Post translational modifications change the shape of the histone sphere. DNA is not completely unwound during replication. It is possible, then, that modified histones may be carried into each new copy of the DNA. Once there, these histones may act as templates, initiating the surrounding new histones to be shaped in the new way. By altering the shape of the histones around it, these modified histones would ensure that a differentiated cell would remain differentiated, and revert into a stem cell.
2. **Addition of methyl groups to the DNA:** Methyl groups are added to CpG sites, in order to convert cytosine to 5-methylcytosine. In effect, the gene is turned off.
   Chromosomal regions can adopt stable and heritable alternative states resulting in bistable gene expression without changes to the DNA sequence. Epigenetic control is often associated with alternative covalent modifications of histones. The stability and heritability of states of larger chromosomal regions are often thought to involve positive feedback, in which modified nucleosomes recruit enzymes that similarly modify nearby nucleosomes. Because DNA methylation and chromatin remodeling play such a central role in many types of epigenic inheritance, the word "epigenetics" is sometimes used as a synonym for these processes. However, this can be misleading. Chromatin remodeling is not always inherited, and not all epigenetic inheritance involves chromatin remodeling.

In this seminar, examples of epigenetic phenomenon influencing these outcomes will be reviewed. The potential limitations of techniques used for epigenetic analyses including issues such as reproducibility or epigenetic measures will also be discussed.

11. The application of genetic studies in specific disorders: psychiatric genetics
Presenter: Dr. Zena Samaan

Reference Papers:

Sullivan P, Daly M, O'Donovan M, et al. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet;13(8):537-551

Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature Genetics Sept 2013; 45, 984–994

Brief Summary:

Genetic code does not only determine one's disease susceptibility, morphological features or longevity, the genetic code can also influence human behaviour, being a happy or sad character, staying single or having a partner, developing a psychiatric disorder such as depression or addiction.

In this session, we will cover the basics of heredity, nature and nurture interactions and provide examples of psychiatric disorders with significant genetic bases.

12. Pharmacogenetics
Presenter: Dr. Kevin Zbuk

Reference Papers:

Pare G, Mehta S, Yusuf S, Phil D, Anand S, Connolly S, Hirsh J, Simonsen K, Bhatt Deepak, Fox K, Eikelboom, J. Effects of CYP2C19 Genotype on Outcomes of Clopidogrel Treatment. New England Journal of Medicine. 2010; 363: 1704-14.

Roden DM, et al. Pharmacogenomics. Lancet. 2019 Aug 10;394(10197):521-532.

Brief Summary:

Pharmacogenomics is a science that examines the inherited variations in genes that dictate drug metabolism and response. It explores the ways these variations can be used to predict whether a patient will have a positive, negative, or absent response to a drug. The way a person responds to a drug, whether positively or negatively, is a complex trait that is influenced by many different genes. In the past, scientists found it difficult to develop genetic tests to predict drug response because the associated genes were not yet known. Once scientists discovered that people's genes show small variations in their nucleotide content, all of that changed; genetic testing for predicting drug response is now possible. Currently, there is still no simple way to determine how individuals will respond to a medication. Thus, pharmaceutical companies are limited to developing drugs using a "one size fits all" system.

SNP screenings will benefit drug development and testing. Using, pharmacogenomics screening, pharmaceutical companies will be able to identify individuals for whom the drug would be harmful or ineffective. By excluding these people from clinical trials, the drug will be more likely to demonstrate its usefulness in a population group. It will thus increase its chances for success in the marketplace. Pre-screening subjects should also allow clinical trials to be smaller, faster, and therefore less expensive; therefore, the consumer could benefit in reduced drug costs. Finally, the ability to assess an individual's reaction to a drug before it is prescribed will increase a physician's confidence in prescribing the drug and the patient's confidence in taking the drug. This in turn should encourage the development of new drugs tested in a similar manner.

In this seminar, the emerging discipline of pharmacogenomics/genetics will be discussed, and examples of pharmacogenomics used in medicine today (i.e. genotype testing and the use of warfarin) will be provided.

13. Wrap up Session – First hour final exam, hours 2-3 wrap up
Dr. Russell de Souza & Dr. Amel Lamri

14. Final Assignment - Independent Study Presentations – December 9, 2016
Dr. Sonia Anand

Online Resources

1. Wellcome Trust Centre for Human Genetics: Course on the Design and Analysis of Disease-Marker Association Studies http://www.well.ox.ac.uk/index.html

2. Wellcome Trust Genome Sequence & Variation Course Manual (2003) (Other Course Manuals) http://www.well.ox.ac.uk/index.html

3. Genetic Epidemiology Course Lectures by David Clayton http://www-gene.cimr.cam.ac.uk/clayton/talks/Bristol_2003/

4. EFG: Statistical Genetic Analysis of Complex Phenotypes Course Notes (2005) incl. Case-Control Association Studies by CM Lewis http://www.dorak.info/epi/genetepi.html

5. Statistical Methods in Genetic Epidemiology Course (ST115) by Ivan Iachin http://statmaster.sdu.dk/courses/st115/index.html

6. Introduction to Genetic Linkage and Association Course Notes by Dave Curtis http://www.smd.qmul.ac.uk/statgen/dcurtis/lectures.html

7. Genetic Epidemiology SuperLecture by Kevin Kip http://www.pitt.edu/~super1/lecture/lec16901/index.htm

8. Introduction to Genetic Epidemiology Lecture by Hermine Maes

9. Human Molecular Genetics (Strachan & Read; 1999): Genetic Mapping of Complex Characters & Complex Disease Genetics http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=hmg.TOC&depth=1

10. Centre for Integrated Genomic Medical Research (CIGMR): Statistical Genetic Analysis http://www.medicine.manchester.ac.uk/genomicepidemiology/aboutus/cigmr/

11. GENESTAT: Genetic Association Studies Portal http://ki.se/ki/jsp/polopoly.jsp?d=9600&l=sv

12. CDC Genomics and Disease Prevention Center: Research Methods Publications http://www.cdc.gov/genomics/default.htm

13. Basic Molecular Genetics for Epidemiologists http://jech.bmj.com/cgi/content/abstract/57/6/398

14. Genetic Epidemiology Studies on Twins by Nick Martin http://www.qimr.edu.au/research/labs/nickm/

15. Quantitative Genetics Lecture Notes (slide presentation)
http://www.ndsu.nodak.edu/instruct/mcclean/plsc431/quantgen/qgen1.htm

16. Human Genetics Interactive Learning Exercises
http://psych.colorado.edu/~carey/hgss/hgssapplets/hgssapplets.htm

17. Genetic Calculation Applets by Knud Christensen (including heritability and variance components)
18. Genetic Power Calculator http://pngu.mgh.harvard.edu/~purcell/gpc/

19. NIH National Human Genome Research Institute Programs (HapMap; ENCODE; Genetic Variation) http://www.genome.gov/12010633

20. HapMap Webcast http://www.genome.gov/17015048

21. HapMap User Guide
http://www.hapmap.org/downloads/presentations/users_guide_to_hapmap.pdf

22. SNP@Ethnos: a database of ethnically variant SNPs (Park, 2007)
http://variome.kobic.re.kr/SNPatETHNIC/

23. Online Encyclopedia for Genetic Epidemiology Studies http://www.genes.org.uk/

24. Glossary on Genetic Epidemiology: Basic http://www.genes.org.uk/ & Advanced
http://jech.bmj.com/cgi/content/abstract/57/8/562

25. HuGE Review Handbook
http://www.genesens.net/_intranet/doc_nouvelles/HuGE%20Review%20Handbook%20v11.pdf

26. STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) & Checklists http://www.strobe-statement.org/index.html

References

1. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. Am.J.Hum.Genet. 2001 Nov;69(5):936-950.
http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1274370&blobtype=pdf

2. Bass MP, Martin ER, Hauser ER. Pedigree generation for analysis of genetic linkage and association. Pac.Symp.Biocomput. 2004:93-103. http://helix-web.stanford.edu/psb04/bass.pdf

3. Boffetta P. Molecular epidemiology. J.Intern.Med. 2000 Dec;248(6):447-454.
http://carcin.oxfordjournals.org/cgi/reprint/28/8/1621

4. Campbell H, Rudan I. Interpretation of genetic association studies in complex disease. Pharmacogenomics J. 2002;2(6):349-360. http://www.nature.com/tpj/journal/v2/n6/pdf/6500132a.pdf

5. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet 2003 Feb 15;361(9357):598-604. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T1B-47XWB9C-X&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_version=1&_urlVersion=0&_userid=1067350&md5=967caf158d4a233c27157704e87f7d94

6. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature 2004 May 27;429(6990):446-452. http://www.nature.com/nature/journal/v429/n6990/full/nature02623.html

7. Cavalli-Sforza LL. The DNA revolution in population genetics. Trends Genet. 1998 Feb;14(2):60-65. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TCY-413CK2K-C&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000051241&_version=1&_urlVersion=0&_userid=1067350&md5=4cb167c954992d295ed62d022ba6a60e

8. Chakravarti A. Population genetics--making sense out of sequence. Nat.Genet. 1999 Jan;21(1 Suppl):56-60. http://www.nature.com/ng/journal/v21/n1s/pdf/ng0199supp_56.pdf

9. Clark AG. The role of haplotypes in candidate gene studies. Genet.Epidemiol. 2004 Dec;27(4):321-333. http://www3.interscience.wiley.com/cgi-bin/fulltext/109608937/PDFSTART

10. Clark AG, Boerwinkle E, Hixson J, Sing CF. Determinants of the success of whole-genome association testing. Genome Res. 2005 Nov;15(11):1463-1467. http://genome.cshlp.org/cgi/reprint/15/11/1463.pdf

11. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat.Genet. 2005 Nov;37(11):1243-1246. http://www.nature.com/ng/journal/v37/n11/pdf/ng1653.pdf

12. Cloninger CR, Rao DC, Rice J, Reich T, Morton NE. A defense of path analysis in genetic epidemiology. Am.J.Hum.Genet. 1983 Jul;35(4):733-756. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1685729&blobtype=pdf

13. Cordell HJ, Olson JM. Correcting for ascertainment bias of relative-risk estimates obtained using affected-sib-pair linkage data. Genet.Epidemiol. 2000 Apr;18(4):307-321. http://www3.interscience.wiley.com/cgi-bin/fulltext/72001202/PDFSTART

14. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, et al. Parameters for reliable results in genetic association studies in common disease. Nat.Genet. 2002 Feb;30(2):149-150. http://www.nature.com/ng/journal/v30/n2/pdf/ng825.pdf

15. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999 Dec;55(4):997-1004. http://www3.interscience.wiley.com/cgi-bin/fulltext/119062104/PDFSTART

16. Dorak MT. Pitfalls in Genetic Association Studies. 2006; Available at: http://www.dorak.info.libaccess.lib.mcmaster.ca/epi/pitfalls.ppt Accessed 19 Aug, 2008.

17. Dorak MT. Basic Population Genetics. Available at: http://www.dorak.info.libaccess.lib.mcmaster.ca/genetics/popgen.html Accessed 19 Aug, 2008.

18. Elbaz A, Alperovitch A. Bias in association studies resulting from gene-environment interactions and competing risks. Am.J.Epidemiol. 2002 Feb 1;155(3):265-272. http://aje.oxfordjournals.org/cgi/reprint/155/3/265

19. Fejerman L, Bouzekri N, Wu X, Adeyemo A, Luke A, Zhu X, et al. Association between evolutionary history of angiotensinogen haplotypes and plasma levels. Hum.Genet. 2004 Sep;115(4):310-318. http://www.springerlink.com/content/akkcluue7k48kvg1/fulltext.pdf

20. Greenberg DA. Linkage analysis of "necessary" disease loci versus "susceptibility" loci. Am.J.Hum.Genet. 1993 Jan;52(1):135-143. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1682115&blobtype=pdf

21. Greenberg DA, Doneshka P. Partitioned association-linkage test: distinguishing "necessary" from "susceptibility" loci. Genet.Epidemiol. 1996;13(3):243-252. http://libaccess.mcmaster.ca/login?url=http://www3.interscience.wiley.com/journal/69051/abstract?CRETRY=1&SRETRY=0

22. Guo SW, Lange K. Genetic mapping of complex traits: promises, problems, and prospects. Theor.Popul.Biol. 2000 Feb;57(1):1-11. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WXD-45BCF0D-1J&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000051241&_version=1&_urlVersion=0&_userid=1067350&md5=3a346a92b63c618c59c2789b0f02bb04

23. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am.J.Hum.Genet. 2007 Apr;80(4):727-739. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1852724&blobtype=pdf

24. Lander ES. The new genomics: global views of biology. Science 1996 Oct 25;274(5287):536-539. *not available online*

25. Lander ES, Schork NJ. Genetic dissection of complex traits. Science 1994 Sep 30;265(5181):2037-2048. http://www.sciencemag.org/feature/data/genomes/265-5181-2037.pdf

26. Lee WC, Ho YY. Potential for gene-gene confounding bias in case-parental control studies. Ann.Epidemiol. 2003 Apr;13(4):261-266. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T44-488KN3F-8&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000051241&_version=1&_urlVersion=0&_userid=1067350&md5=d1741f52f1c992f282972d5ef522550d

27. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat.Rev.Genet. 2008 May;9(5):356-369. http://www.nature.com/nrg/journal/v9/n5/pdf/nrg2344.pdf

28. Morimoto LM, White E, Newcomb PA. Selection bias in the assessment of gene-environment interaction in case-control studies. Am.J.Epidemiol. 2003 Aug 1;158(3):259-263. http://aje.oxfordjournals.org/cgi/reprint/158/3/259

29. Morton NE. Linkage disequilibrium maps and association mapping. J.Clin.Invest. 2005 Jun;115(6):1425-1430. http://www.jci.org/115/6/1425/pdf

30. Morton NE. Significance levels in complex inheritance. Am.J.Hum.Genet. 1998 Mar;62(3):690-697. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1376937&blobtype=pdf

31. Olson JM, Cordell HJ. Ascertainment bias in the estimation of sibling genetic risk parameters. Genet.Epidemiol. 2000 Mar;18(3):217-235. . http://www3.interscience.wiley.com/cgi-bin/fulltext/71000704/PDFSTART

32. Online Encyclopedia for Genetic Epidemiology Studies. Recent genome-wide association studies. 2008; Available at: http://www.genes.org.uk.libaccess.lib.mcmaster.ca/gwas.shtml Aug 19 2008.

33. Pearson TA, Manolio TA. How to interpret a genome-wide association study. JAMA 2008 Mar 19;299(11):1335-1344. http://jama.ama-assn.org/cgi/reprint/299/11/1335

34. Potter JD. Epidemiology, cancer genetics and microarrays: making correct inferences, using appropriate designs. Trends Genet. 2003 Dec;19(12):690-695. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TCY-49VC7CF-1&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_version=1&_urlVersion=0&_userid=1067350&md5=bbbd12b09b36c15b2927db6bdf9e0150

35. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am.J.Hum.Genet. 2001 Jul;69(1):1-14. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1226024&blobtype=pdf

36. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am.J.Hum.Genet. 1999 Jul;65(1):220-228. http://pritch.bsd.uchicago.edu/publications/strat.pdf

37. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001 Sep;17(9):502-510. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TCY-43RRTX1-T&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000051241&_version=1&_urlVersion=0&_userid=1067350&md5=860287b869984926d6993079ddcd18cb

38. Risch N. Genetic linkage: interpreting lod scores. Science 1992 Feb 14;255(5046):803-804. *not available online*

39. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. Am.J.Hum.Genet. 1990 Feb;46(2):222-228. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1684987&blobtype=pdf

40. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996 Sep 13;273(5281):1516-1517. *not available online*

41. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000 Jun 15;405(6788):847-856. http://www.nature.com/nature/journal/v405/n6788/pdf/405847a0.pdf

42. Ropers HH. New perspectives for the elucidation of genetic disorders. Am.J.Hum.Genet. 2007 Aug;81(2):199-207. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1950818&blobtype=pdf

43. Schork NJ, Cardon LR, Xu X. The future of genetic epidemiology. Trends Genet. 1998 Jul;14(7):266-272. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TCY-3V8TMDH-B&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_version=1&_urlVersion=0&_userid=1067350&md5=f0e081d644796b5ce53c71c52fe5b05c

44. Seltman H, Roeder K, Devlin B. Evolutionary-based association analysis using haplotype data. Genet.Epidemiol. 2003 Jul;25(1):48-58. http://www3.interscience.wiley.com/cgi-bin/fulltext/104538134/PDFSTART

45. Templeton AR. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. Genetics 1995 May;140(1):403-409.

46. Templeton AR, Boerwinkle E, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics 1987 Oct;117(2):343-351. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1206565&blobtype=pdf

47. Templeton AR, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF. Cladistic structure within the human Lipoprotein lipase gene and its implications for phenotypic association studies. Genetics 2000 Nov;156(3):1259-1275. http://www.genetics.org/cgi/reprint/117/2/343

48. Thomas DC, Witte JS. Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol. Biomarkers Prev. 2002 Jun;11(6):505-512. http://cebp.aacrjournals.org/cgi/reprint/11/6/505.pdf

49. Tzeng JY. Evolutionary-based grouping of haplotypes in association analysis. Genet.Epidemiol. 2005 Apr;28(3):220-231. http://www3.interscience.wiley.com/cgi-bin/fulltext/109929608/PDFSTART

50. Vineis P, McMichael AJ. Bias and confounding in molecular epidemiological studies: special considerations. Carcinogenesis 1998 Dec;19(12):2063-2067. http://carcin.oxfordjournals.org/cgi/reprint/19/12/2063

51. Wacholder S, Rothman N, Caporaso N. Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer. Cancer Epidemiol.Biomarkers Prev. 2002 Jun;11(6):513-520.http://cebp.aacrjournals.org/cgi/reprint/11/6/513

52. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat.Rev.Genet. 2005 Feb;6(2):109-118. http://www.nslij-genetics.org/ld/wang05.pdf

53. Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. Trends Genet. 2002 Jan;18(1):19-24. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TCY-44N9B08-C&_user=1067350&_rdoc=1&_fmt=&_orig=search&_sort=d&view=c&_acct=C000051241&_version=1&_urlVersion=0&_userid=1067350&md5=3e6334d25eac91558a55eede065bdd5a

54. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007 Jun 7;447(7145):661-678. http://www.nature.com/nature/journal/v447/n7145/pdf/nature05911.pdf

55. Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. Nat.Genet. 1999 Dec;23(4):397-404. http://www.nature.com/ng/journal/v23/n4/pdf/ng1299_397.pdf