M21-515 Fundamentals of Genetic Epidemiology
Summer 2017

Revised: 6/25/2017

Course Masters	Treva Rice, Ph.D. (treva@wustl.edu) Yun Ju Sung, Ph.D. (yunju@wustl.edu)
Grading	Computer Lab Assignments 30% Daily Quiz 30% Midterm Exam (covers lectures / homework 1-5) 20% Final Project 20% Final Grade (+/– letter grades)
Software	Software packages include: (1) R (http://www.r-project.org/) (2) PEDSTATS, Merlin and QTDT (http://www.sph.umich.edu/csg/abecasis/software.html) (3) PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/)
Textbook	Austin, MA. ***Genetic Epidemiology: Methods & Applications*** 2013, CABI: Oxfordshire, UK.
Format	2-week INTENSIVE course Morning lecture: 9:00 am – 12:00 noon Afternoon computer lab: 1:30 pm – 4:00 pm 10-minute break ~ every hour
1st Homework Assignment	Due first day of class, Textbook (Chapters 1-2) On-line genetics tutorial, Chapters 3, 4, and 6 (http://anthro.palomar.edu/tutorials/biological.htm)
Prerequisite	1) Knowledge of R programming, either having taken the MSIBS summer R-course or experience in R-programming; 2) experience in Linux/Unix operating system.

The core competency for the Fundamentals of Genetic Epidemiology course (M21-515) is for students to understand basic concepts, methods and analytical approaches in genetic epidemiology.

Learning objectives are to

- Understand familial resemblance, heritability and family study designs
- Appreciate maximum likelihood methods and hypothesis testing
- Be aware of selected molecular and population genetics principles, including Hardy-Weinberg Equilibrium
- Grasp the basic concepts and principles underlying genetic linkage and association

- Be able to perform analysis in heritability, linkage and association using selected software and critically evaluate and interpret the corresponding results

Additional Information

1. Quiz based on lecture/assignment from previous day.
2. No make-up quizzes unless pre-arranged before day of quiz.
3. Homework (reading) due BEFORE day of assignment
4. Computer lab (practicum) due on day following assignment BEFORE lecture.
5. Five points per day deducted for lateness in submitting practicum assignment.
6. Final project consists of oral presentation and one-page written report.

Additional Reading

1. Heritability:

- Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. Nature Reviews Genetics 2008 9(4):255-266.
- Tenesa A, Haley CS. The heritability of human disease: estimation, uses and abuses. Nature Reviews Genetics 2013 14(2):139-149.
- Eichler EE, Flint J. Gibson G, Kong A, Leal SZ, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics 2010; 11:446-450.
- Speed D, Balding DJ. Relatedness in the post-genomic era: Is it still useful? Nature Reviews Genetics 2015; 16:33-44.

2. Molecular Genetics:

- Allison LA. Fundamental Molecular Genetics. 2007 Blackwell Publishing (Last accessed 9/30/2014: https://molbiomadeeasy.files.wordpress.com/2013/09/fundamental_molecular_biology.pdf).

3. Markers:

- Van Eenennaam A. Basics of DNA markers and genotyping. June 2009 (last accessed 9/30/2014: http://animalscience.ucdavis.edu/animalbiotech/Outreach/Basics_of_DNA_Markers_and_Genotyping.pdf).
- Morin PA, Luikart G, Wayne RK. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 2004 19(4): 208-216.

4. Huntington disease:

- Bates GP. History of genetic disease: The molecular genetics of Huntington disease—A history. *Nature Reviews Genetics* 6, 766–773 (2005)

Course Syllabus

	Morning (Lecture Room) *9 am to 12 noon*			*Afternoon (Computer Lab)* *1:30 pm to 4 pm*	
Day	**Lecture**	**Instructor**	**Reading Homework**	**Computer Lab Practicum**	**Instructor**
1 7/13/Th	Overview of Couse Overview of GE Heritability, Family Designs	Rice	On-line Tutorial Ch 1 – 2 Additional Reading 1	Heritability, Merlin	Rice
2 7/14/F	MLE Hypothesis testing Molecular Genetics (DNA)	Rice	Ch 3 Additional Reading 2	Phenotype Data QC, PEDSTATS, R	Rice
3 7/17/M	Population Genetics (Mendel, Segregation, HWE) Genetic Markers	Rice	Ch 3 Additional Reading 3	Phenotype Data QC, R ***Review Midterm Exam***	Rice
4 7/18/T	GxE, Non-Mendelian Genetics Ethics, Public Health	Rice	Ch 8, 10, 11	**Midterm Exam** **1:30 pm – 4 pm (Lecture Room)**	
5 7/19/W	Gene-mapping Example Introduction to Linkage	Sung	Ch 3, 4 Additional Reading 4	Genotype Data QC using PEDSTATS Introduction to Merlin	Sung
6 7/20/Th	Model-based Linkage Model-free Linkage	Sung	Ch 4	Linkage Analysis using Merlin	Sung
7 7/21/F	Association Studies	Sung	Ch 5	GWAS using PLINK (Part 1)	Sung
8 7/24/M	Population Stratification	Sung	Ch 6	GWAS using PLINK (Part 2)	Sung
9 7/25/T	Analysis of Rare Variants Data Resources	Sung	Ch 5, 9	**Practice Presentation** **(Lecture Room)**	
10 7/26/W	**Final Project Presentation** **9:00 - noon and 1:30 - 4 pm (Lecture Room)**				

Note that the schedule and topics are subject to change.

Final Project

Each student will choose one topic from the list provided below (from page 4 to page 7). After doing research, reading suggested (and additional) papers on the topic, you will give a 10-minutes oral presentation and submit a one-page written report.

You need to give a

- Practice presentation (5 minutes) on July 25, Tuesday afternoon (1:30 – 4 pm in the lecture room) to show us the rough idea of what you will present the next day. You will have a chance to see how your classmates are doing and will receive some feedback and suggestions from your instructor and TA.
- Final presentation (10 minutes) on July 26, Wednesday (in the lecture room). You may have questions at the end of your presentation. Both instructors and TA will provide scores on your presentation. Also Karen Schwander has kindly agreed to listen to your presentation and give a score.
- One-page written report (by midnight on July 26, Wednesday). I will be responsible for grading your written report.

As we want each topic to be covered by only one student, no two students should choose the same topic. Choose the topic that you want to present at your earliest convenience before others choose.

Here are some guidelines for choosing a topic and preparing a presentation:

- The order of presentation will follow the order of topics that I have created.
- Some topics will be covered during my lecture. You need to present new information, rather than repeating what you have learned from my lecture.
- We expect that
    - Your voice should be loud and clear.
    - Your presentation should be well organized. A 10-minute presentation is very short compared to the many hours you spend researching the topic.
    - You only include the materials that you understand or consider important.
    - Your slides should not be too pretty (Please use simple style to avoid distractions!)

Please educate your classmates and instructors. We have a high expectation and confidence that you will do well!

1. Statistical Methodologies for GWAS Era

1a. Genotype Imputation

- Marchini J and Howie B. Genotype imputation for genome-wide association studies. Nature Review Genetics. 11, 499–511 (2010)
- Li N and Stephens M. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics. 2003 165, 2213-2233 (2003)

1b. Population Stratification

- John Novembre et al. Genes mirror geography within Europe. Nature 456, 98-101 (2008)
- David Reich et al. Principal component analysis of genetic data. Nature Genetics 40, 491 - 492 (2008).
- Alkes Price et al. New approaches to population stratification in genome-wide association studies. Nature Review Genetics. 11, 459–463 (2010)

1c. Meta-analysis

- Sarah Brockwell and Ian Gordon. A comparison of statistical methods for meta-analysis. Statistics in Medicine. 20, 825–840 (2001)
- Paul de Bakker et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122-R128 (2008)
- Evangelos Evangelou et al. Meta-analysis methods for genome-wide association studies and beyond. Nature Reviews Genetics 14, 379–389 (2013)

2. Beyond GWAS

2a. Missing Heritability

- Brendan Maher. Personal genomes: The case of the missing heritability. Nature 456:18-21 (2008)
- Terri Manolio et al. Finding the missing heritability of complex diseases. Nature. 461:747-753 (2009)
- Greg Gibson. Hint of hidden heritability in GWAS. Nature Genetics 42, 558–560 (2010)

2b. 1000 Genomes Project

- http://www.1000genomes.org/
- The 1000 Genomes Project. A map of human genome variation from population-scale sequencing. Nature 467:1061-1073. October 2010
- An integrated map of genetic variation from 1092 human genomes. Nature. 491: 56–65. November 2012

2c. Rare Variants

- Vikas Bansal et al. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet. (2010)
- Elizabeth Cirulli and David Goldstein. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nature Reviews Genetics 11, 415-425 (2010)
- Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet. 95:5-23 (2014)

2d. PheWAS (reverse of GWAS)

- Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome –phenome relationship using phenome-wide association studies. Nature Reviews Genetics 2016 17:129-145.
- Brookes AJ, Robinson PN. Human genotype-phenotype databases: Aims, challenges and opportunities. Nature Reviews Genetics 2016
- Verma SS, Frase AT, Verma A, Pendergrass SA, Mahony S, Haas DW, Ritchie MD. PHENOME-WIDE INTERACTION STUDY (PheWIS) IN AIDS CLINICAL TRIALS GROUP DATA (ACTG). Pac Symp Biocomput. 21:57-68 (2016)
- Pendergrass SA, Verma A, Okula A, Hall MA, Crawford DC, Ritchie MD. Phenome-Wide Association Studies: Embracing Complexity for Discovery. Hum Heredity 79:111-123 (2015)

2e. Machine Learning in Genetics

- Machine learning applications in genetics and genomics: http://www.nature.com/nrg/journal/v16/n6/abs/nrg3920.html
- From Statistical Genetics to Predictive Models in Personalized Medicine: http://videolectures.net/nipsworkshops2011_personalized_medicine/
- Regularized Machine Learning in the Genetic Prediction of Complex Traits: http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004754
- Big Data Analysis Using Modern Statistical and Machine Learning Methods in Medicine: http://einj.org/journal/view.php?number=473

3. From Association to Function

3a. ENCODE Project

- https://www.encodeproject.org/
- http://www.nature.com/encode/#/threads
- A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011;9(4):e1001046.
- ENCODE. An integrated encyclopedia of DNA elements in the human genome. Nature. Sep 6 2012;489(7414):57-74.

3b. UCSC Genome Browser

- http://genome.ucsc.edu/

- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006.
- Kent WJ, Hsu F, Karolchik D, Kuhn RM, Clawson H, Trumbower H, Haussler D. Exploring relationships and mining data with the UCSC Gene Sorter. Genome Res. 2005 May;15(5):737-41.

3c. HaploReg

- http://archive.broadinstitute.org/mammals/haploreg/haploreg.php
- Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res (2012) 40 (D1): D930-D934.
- Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nature Biotechnology 30, 1095–1106 (2012) doi:10.1038/nbt.2422

4. Genetics Software/Programs

4a. ABEL Suite

- http://www.genabel.org/
- Yurii Aulchenko et al. GenABEL: an R library for genome-wide association analysis. Bioinformatics (2007) 23 (10): 1294-1296.
- Yurii Aulchenko et al. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 2010, 11:134

4b. METAL

- http://genome.sph.umich.edu/wiki/METAL_Documentation
- Cristen Willer et al. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010 Sep 1; 26(17): 2190–2191.

4c. RAREMETAL

- http://genome.sph.umich.edu/wiki/RAREMETAL_Documentation
- Shuang Feng et al. RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics. 2014 Oct; 30(19): 2828–2829.
- Dajiang Liu et al. Meta-analysis of gene-level tests for rare variant association. Nature Genetics 46, 200–204 (2014)