M21-515 Fundamentals of Genetic Epidemiology
Fall 2019

Revised: 08/29/2019

Course Masters	Treva Rice, Ph.D. (treva@wustl.edu) Yun Ju Sung, Ph.D. (yunju@wustl.edu)
Lab Instructor	Oyomoare Osazuwa-Peters (oosazuwa-peters@wustl.edu)
Grading	Computer Lab Assignments 30% Daily Quiz 30% Midterm Exam 20% Final Project 20% Final Grade (+/– letter grades)
Software	Software packages include: 1. R (http://www.r-project.org/) 2. PEDSTATS, Merlin, QTDT (http://www.sph.umich.edu/csg/abecasis) 3. PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/)
Textbook	Austin, MA. ***Genetic Epidemiology: Methods & Applications*** 2013, CABI: Oxfordshire, UK.
Time	Mon/Wed 9:00 am – 12:00 noon Starting Monday 10/21 and ending Wednesday 12/18
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. Weekly quiz based on lecture/assignment from previous lecture.

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.

Course Syllabus

Day	Topic	Instructor	Reading/Homework	Location
1 10/21/Mon	• Overview of Couse • Overview of GE • Heritability, Family Designs	Rice	On-line tutorial Ch 1 – 2 Additional Reading 1	Lecture
2 10/23/Wed	• MLE • Hypothesis testing • Molecular Genetics (DNA)	Rice	Ch 3 Additional Reading 2	Lecture
3 10/28/Mon	• Phenotype Data QC with PEDSTATS, R	Osazuwa-Peters	**Assignment 1 (due 11/3/Sun)**	Lab
4 10/30/Wed	• Population Genetics (Mendel, Segregation, HWE) • Genetic Markers	Rice	Ch 3 Additional Reading 3	Lecture
5 11/4/Mon	• Heritability using Merlin and QTDT	Osazuwa-Peters	**Assignment 2 (due 11/10/Sun)**	Lab
6 11/6/Wed	**Midterm Exam**			**Lecture**
7 11/11/Mon	• Gene-mapping Example • Introduction to Linkage	Sung	Ch 3, 4 Additional Reading 4	Lecture
8 11/13/Wed	• Genotype Data QC with PEDSTATS	Osazuwa-Peters	**Assignment 3 (due 11/19/Tue)**	Lab
9 11/18/Mon	• Model-based Linkage • Model-free Linkage	Sung	Ch 4	Lecture
10 11/20/Wed	• Linkage Analysis using Merlin	Osazuwa-Peters	**Assignment 4 (due 12/1/Sun)**	Lab
11 11/25/Mon	• Association Studies	Sung	Ch 5	Lecture
12 12/2/Mon	• GWAS using PLINK	Osazuwa-Peters	**Assignment 5 (due 12/8/Sun)**	Lab
13 12/4/Wed	• Population Stratification	Sung	Ch 6	Lecture
14 12/9/Mon	• Analysis of Rare Variants • Data Resources	Sung	Ch 5, 9	Lecture
15 12/11/Wed	**Practice Presentation**			**Lecture**
16 12/16/Mon	**Final Project Presentation**			**Lecture**
17 12/18/Wed	**Final Project Presentation (if more time is needed)**			**Lecture**

Note that the schedule and topics are subject to change.

Suggested 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)

Final Project

Each student will choose one topic from the list provided below. 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:

- Practice presentation (5 minutes) on December 11, Wednesday (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 December 16 (and 18) (in the lecture room). You may have questions at the end of your presentation. All three instructors 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 December 18, 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 165, 2213-2233 (2003)
- Howie B et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genetics 44, 955–959 (2012)

1b. Principal Component Analysis in Human Genetics

- 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)

1d. Mendelian Randomization

- Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 2008; 27: 1133-63.
- Bowden, J. et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization: A framework for two-sample summary data MR. Statistics in Medicine 36, 1783–1802 (2017).
- Lawlor DA. Two-sample Mendelian randomization: opportunities and challenges. International Journal of Epidemiology 2016; 908–915. doi: 10.1093/ije/dyw127.
- Zheng, J. et al. Recent Developments in Mendelian Randomization Studies. Current Epidemiology Reports 4, 330–345 (2017).
- Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ (2018)
- Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human Molecular Genetics (2018). doi:10.1093/hmg/ddy163

1e. LD Score Regression

- https://github.com/bulik/ldsc
- Bulik-Sullivan BK LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genetics Mar;47(3):291-5 (2015)
- Jie Zheng et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33; 272–279 (2017)

1f. 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

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 (2010)
- An integrated map of genetic variation from 1092 human genomes. Nature. 491: 56–65 (2012)
- 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 526: 68-74 (2015)

2c. Rare Variants

- Vikas Bansal et al. Statistical analysis strategies for association studies involving rare variants. Nature Review Genetics (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. Epistasis

- Phillips PC. Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics 9:855-867 (2008)
- Cordell H. Detecting gene-gene interactions that underlie human diseases. Nature Reviews Genetics 10:392-404 (2009)
- Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nature Reviews Genetics 15:722-723 (2014)

2e. Pleiotropy

- Stearns FW. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* 186(3):767-773 (2010)
- Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nature Reviews Genetics 12: 204-213. (2011)
- Salinas YD, Wang Z, DeWan AT. Statistical Analysis of Multiple Phenotypes in Genetic Epidemiologic Studies: From Cross-Phenotype Associations to Pleiotropy. *Am J Epidemiol*. 187(4):855-863 (2018)
- Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits. *Open Biol*. 7(11) (2017)

2f. Phenome-Wide Association Studies (PheWAS)

- Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome –phenome relationship using phenome-wide association studies. Nature Reviews Genetics 17:129-145 (2016)

- Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. Nature Reviews Genetics 16:85-97 (2015)
- 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)

2g. Pharmacogenomics

- Veenstra DL. *The Role of Epidemiology in Assessing the Potential Clinical Impact of Pharmacogenomics*. Oxford University Press; 2009. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195398441.001.0001/acprof-9780195398441-chapter-27. Accessed July 5, 2018.
- https://www.genome.gov/27530645/faq-about-pharmacogenomics/
- Pharmacogenomics and Personalized Medicine https://www.nature.com/scitable/topicpage/pharmacogenomics-and-personalized-medicine-643
- https://www.nature.com/subjects/pharmacogenomics

2h. Epigenetics

- Johnstone SE, Baylin SB. Stress and the epigenetic landscape: a link to the pathobiology of human diseases? Nature Reviews Genetics 11; 806-812 (2010)
- Guthman J, Mansfield B. The implications of environmental epigenetics: A new direction for geographic inquiry on health, space, and nature-society relations. Progress in Human Geography. 37:486-504 (2013)
- Taudt A, Colme-Tatche M, Johannes F. Genetic sources of population epigenomic variation. Nature Reviews Genetics (2016)
- Flanagan JM. Epigenome-wide association studies (EWAS): past, present, and future. *Methods Mol Biol*. 1238:51-63 (2015)
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nature Review Genetics* 12(8):529-541 (2011)

3. From Association to Function

3a. From Association to Function

- Stacey Edwards et al. Beyond GWASs: Illuminating the Dark Road from Association to Function. American journal of human genetics. 93: 779-797 (2013)
- Ruth McPherson. From genome-wide association studies to functional genomics: new insights into cardiovascular disease. Can J Cardiol. 29: 23-9 (2013)

3b. 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. 9(4):e1001046 (2011)
- ENCODE. An integrated encyclopedia of DNA elements in the human genome. Nature 489(7414):57-74 (2012)

3c. 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. 12(6):996-1006 (2002)
- 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. 15(5):737-41 (2005)
- Rosenbloom KR, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43 (Database issue):D670-81 (2015)

3d. 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 40 (D1): D930-D934 (2012)
- Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nature Biotechnology 30, 1095–1106 (2012)

3e. Cancer Genomics

- http://cancer.genomics.nih.gov
- Hindorff LA, Gillanders EM, Manolio TA. Genetic architecture of cancer and other complex diseases: lessons learned and future directions. Carcinogenesis. 2011 Jul;32(7):945-54. doi: 10.1093/carcin/bgr056. Epub 2011 Mar 31.
- Cancer Genome Atlas Research Network, Ley TJ et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 368(22): 2059-74 (2013)
- Wartman LD. A case of me: clinical cancer sequencing and the future of precision medicine. Cold Spring Harb Mol Case Stud 1: a000349 (2015)
- White BS, DiPersio JF. Genomic tools in acute myeloid leukemia: From the bench to the bedside. Cancer 120: 1134-44 (2014)
- In Treatment for Leukemia, Glimpses of the Future. https://www.nytimes.com/2012/07/08/health/in-gene-sequencing-treatment-for-leukemia-glimpses-of-the-future.html

4 Biobanks and Diversity

4a. Diversity

- Non-European populations still underrepresented in genomic testing samples. *American Journal of Medical Genetics Part A*. 173(2):296-297
- Hindorff LA, Bonham VL, Brody LC, et al. Prioritizing diversity in human genomics research. *Nat Rev Genet*. 19(3):175-185 (2018)
- Bentley AR, Callier S, Rotimi CN. Diversity and inclusion in genomic research: why the uneven progress? *J Community Genet*. 8(4):255-266 (2017)

4b. eMERGE network

- McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Medical Genomics*. 4:13. doi:10.1186/1755-8794-4-13 (2011)

- Crawford DC, Crosslin DR, Tromp G, et al. eMERGEing progress in genomics—the first seven years. *Frontiers in Genetics* 5:184. doi:10.3389/fgene.2014.00184 (2014)
- Kho AN, Pacheco JA, Peissig PL, et al. Electronic Medical Records for Genetic Research: Results of the eMERGE Consortium. *Science translational medicine*. 3(79):79re1. (2011)

4c. UK Biobank

- http://www.ukbiobank.ac.uk/
- Collins R. What makes UK Biobank special? *The Lancet* 379(9822):1173-1174 (2012)
- Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* 12(3):e1001779 (2015

5. Genetics Software/Programs

5a. 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

5b. EasyQC

- http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/index.html
- Thomas Winkler et al. Quality control and conduct of genome-wide association meta-analyses. Nature Protocols 9, 1192–1212 (2014)

5c. METAL

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

5d. RAREMETAL

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

5e. FunciSNP

- http://bioconductor.org/packages/release/bioc/html/FunciSNP.html
- Simon Coetzee et al. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. Nucl. Acids Res. (2012)

5f. GCTA-a tool for genome-wide complex trait analysis

- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet*. 88(1):76-82 (2011)
- Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol*. 1019:215-236 (2013)
- Mitchell JS, Johnson DC, Litchfield K, et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. *Scientific Reports* 5:12473 (2015)