Advanced Population and Medical Genetics
EPI511, Spring semester, 2023
Tue/Thu 9:45-11:15am, Kresge 907

Instructor

Alkes Price, Professor (aprice@hsph.harvard.edu). Office Hours:
Thu 11:15am-12:15pm, Building 2, Room 211 (Jan 26 - May 4, excluding Mar 16)

Teaching Assistant

Jordan Rossen, Doctoral student (jordanrossen@g.harvard.edu). Office Hours:
Mon 1:45-2:45pm + Fr 10-11am, Building 2, Room 209 (Jan 27 - Apr 28, excluding Mar 13,17)

Prerequisites

- BST273 (Introduction to Programming), or equivalent programming experience in Python
- BST227 or EPI507 or EPI293, or equivalent experience in genetics

Course Description

This course will cover quantitative topics in human population genetics and applications to medical genetics, including the HapMap project, linkage disequilibrium, population structure and admixture, population stratification, fine-mapping, heritability, genetic risk prediction, and mixed model association. The course is aimed at Epidemiology and Biostatistics students with a strong interest in statistical genetics, and is included in the Biostatistics Advanced Doctoral Core and Biostatistics Master's core. The course will emphasize hands-on analysis of large empirical data sets, thus requiring prior experience with a general-purpose high-level programming language such as Python. After taking this course, each student will have the experience and skills to develop and apply statistical methods to population genetic data.

Course Objectives

After taking this course, the student will be able to:

- Critically analyze large empirical data sets using a high-level programming language (Python).
- Apply fundamental concepts in population and medical genetics such as linkage disequilibrium, population structure and admixture, population stratification, fine-mapping, and heritability.
- Develop and apply statistical methods to population genetic data.

Texts and Reading Materials

Lecture notes and links to relevant scientific papers will be provided on the course website.

Outcome Measures and Grading

Substantial and informed student participation in class discussions is expected of all students; this includes reading the required advance reading and asking/answering questions in class. The final grade for this course will be based on Experiences 1-6 (60%) + Research Paper (40%).

Experiences:

At the heart of this course are 6 Experiences: biweekly take-home projects in which students apply fundamental concepts from the reading, lecture and discussion parts of the course to analyze empirical data sets. The Experiences will determine 60% of the grade for this course. Computer code written in Python (iPython notebook format preferred), and its output, are to be submitted via the Experiences dropbox on the course www site.

Research Paper:

In addition, each student will write a short research paper (1,000-1,500 words) describing their scientific results on a project of their choice. A list of suggested project topics will be provided. The short research paper will determine 40% of the grade for this course.

Harvard Chan Policies and Expectations:

Inclusivity Statement

Diversity and inclusiveness are fundamental to public health education and practice. It is a requirement that you have an open mind and respect differences of all kinds. I share responsibility with you for creating a learning climate that is hospitable to all perspectives and cultures; please contact me if you have any concerns or suggestions.

Academic Integrity

Each student in this course is expected to abide by the Harvard University and the Harvard T.H. Chan School of Public Health Codes of Academic Integrity. All work submitted to meet course requirements is expected to be a student's own work. In the preparation of work submitted to meet course requirements, students should always take great care to distinguish their own ideas and knowledge from information derived from sources.

Students must assume that collaboration in the completion of assignments is prohibited unless explicitly specified. Students must acknowledge any collaboration and its extent in all submitted work. This requirement applies to collaboration on editing as well as collaboration on substance.
For this course, collaboration is allowed in the following instances:
• Experiences 1-6: OK to discuss Experiences with your colleagues, but each piece of code you write should be your own.
• Research Paper: OK to discuss Research Paper with your colleagues, but each piece of code you write should be your own and each piece of text you write should be your own.

Should academic misconduct occur, the student(s) may be subject to disciplinary action as outlined in the Student Handbook. See the Student Handbook for additional policies related to academic integrity and disciplinary actions.

Accommodations for Students with Disabilities

Harvard University provides academic accommodations to students with disabilities. Any requests for academic accommodations should ideally be made before the first week of the semester, except for unusual circumstances, so arrangements can be made. Students must register with the Local Disability Coordinator in the Office for Student Affairs to verify their eligibility for appropriate accommodations. Contact the OSA (studentaffairs@hsph.harvard.edu) in all cases, including temporary disabilities.

Course Evaluations

Constructive feedback from students is a valuable resource for improving teaching. The feedback should be specific, focused and respectful. It should also address aspects of the course and teaching that are positive as well as those which need improvement.

Completion of the evaluation is a requirement for each course. Your grade will not be available until you submit the evaluation. In addition, registration for future terms will be blocked until you have completed evaluations for courses in prior terms.

Course Schedule

Week: Dates	Topic	Recommended advance reading (Optional advance reading in parentheses)
Week 1: Jan 24,26	Introduction + HapMap / 1000 Genomes projects	International HapMap3 Consortium 2010 Nature[1] (Visscher et al. 2017 Am J Hum Genet[2])
Week 2: Jan 31, Feb 2	Linkage disequilibrium	Conrad et al. 2006 Nat Genet[3] (Slatkin 2008 Nat Rev Genet[4])
Week 3: Feb 7,9	Population structure	Novembre et al. 2008 Nature[5] (Tishkoff et al. 2009 Science[6])
Week 4: Feb 14,16	Population admixture	Sankararaman et al. 2008 Am J Hum Genet[7] (Price et al. 2009 PLoS Genet[8])
Week 5: Feb 21,23	Population stratification	Price et al. 2006 Nat Genet[9] (Bulik-Sullivan et al. 2015a Nat Genet[10])
Week 6: Feb 28, Mar 2	Fine-mapping	Maller et al. 2012 Nat Genet[11] (Schaid et al. 2018 Nat Rev Genet[12])
Week 7: Mar 7,9	Natural selection	Galinsky et al. 2016a Am J Hum Genet[13] (Sabeti et al. 2006 Science[14])
Week 8: Mar 21,23	Heritability	Yang et al. 2010 Nat Genet[15] (Lee et al. 2012 Nat Genet[16])
Week 9: Mar 28,30	Genetic risk prediction	Vilhjalmsson et al. 2015 Am J Hum Genet[17] (Khera et al. 2018 Nat Genet[18])
Week 10: Apr 4,6	Mixed model association	Yang et al. 2014 Nat Genet[19] (Loh et al. 2015a Nat Genet[20])
Week 11: Apr 11,13	Rare variant analysis	Karczewski et al. 2022 Cell Genom[21] (Lee et al. 2014 Am J Hum Genet[22])
Week 12: Apr 18,20	Functional interpretation of genetic associations	Finucane et al. 2015 Nat Genet[23] (Gazal et al. 2017 Nat Genet[24])
Week 13: Apr 25 (Tue)	eQTL and transcriptome-wide association studies	GTEx Consortium 2020 Science[25] (Gusev et al. 2016 Nat Genet[26])
Week 13: Apr 27 (Thu)	*Allele-specific analysis for regulatory variant discovery	Grishin et al. 2022 Nat Genet[27] (Knowles et al. 2017 Nat Methods[28])

*guest lecture by Sasha Gusev, Associate Professor, DFCI/Harvard Medical School

Experiences (due by 8:00am each Tue, via Experiences dropbox on course www site.)

Experience 1: due Tue Feb 7
Experience 2: due Tue Feb 21
Experience 3: due Tue Mar 7
Experience 4: due Tue Mar 28
Experience 5: due Tue Apr 11
Experience 6: due Tue Apr 25

Research Paper (due by 8:00am Wed May 10. Please send by email to Alkes Price.)

Each student should choose one topic for the short research paper. Possible topics will be suggested each week, and an aggregate list of suggested topics will be provided on Apr 25. Each student should schedule a 15-minute appointment with the instructor during the week of May 1-5, and should choose and begin work on their topic prior to this meeting. The short research paper will be due by 8:00am Wed May 10. The paper should be 1,000-1,500 words long, and should include an abstract, plus one figure and one table and at least 10 references. Additional subdivision into Introduction, Results, Discussion and Methods sections is optional. For an example of a short research paper, see Lindstrom et al. 2011 Nat Genet[29].

Bibliography

1. The International HapMap 3 Consortium. An integrated haplotype map of rare and common genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
2. Visscher, P.M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* **101**, 5-22 (2017).
3. Conrad, D.F. et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* **38**, 1251-60 (2006).
4. Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-85 (2008).
5. Novembre, J. et al. Genes mirror geography within Europe. *Nature* **456**, 98-101 (2008).
6. Tishkoff, S.A. et al. The genetic structure and history of Africans and African Americans. *Science* **324**, 1035-44 (2009).
7. Sankararaman, S., Sridhar, S., Kimmel, G. & Halperin, E. Estimating local ancestry in admixed populations. *Am J Hum Genet* **82**, 290-303 (2008).
8. Price, A.L. et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* **5**, e1000519 (2009).
9. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
10. Bulik-Sullivan, B.K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
11. Maller, J.B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**, 1294-301 (2012).
12. Schaid, D.J. et al. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
13. Galinsky, K.J. et al. Fast principal components analysis reveals convergent evolution of ADH1B gene in Europe and East Asia. *Am J Hum Genet* (2016).
14. Sabeti, P.C. et al. Positive natural selection in the human lineage. *Science* **312**, 1614-20 (2006).
15. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
16. Lee, S.H. et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* **44**, 247-50 (2012).
17. Vilhjalmsson, B.J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* **97**, 576-92 (2015).
18. Khera, A.V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-24 (2018).
19. Yang, J., Zaitlen, N.A, Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**, 100-106 (2014).
20. Loh, P.R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
21. Karczewski, K.J. et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* **2**, 100168 (2022).
22. Lee, S. et al. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet* **95**, 5-23 (2014).
23. Finucane, H.K. et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat Genet* **47**, 1228-35 (2015).
24. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet* **49**, 1421-7 (2017).
25. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-30 (2020).
26. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48** 245-52 (2016).
27. Grishin, D. et al. Allelic imbalance of chromatin accessibility in cancer identifies candidate causal risk variants and their mechanisms. *Nat Genet* **54**, 837-49 (2022).
28. Knowles, D.A. et al. Allele-specific expression reveals interactions between genetic variation and environment. *Nat Methods* **14**, 699-702 (2017).
29. Lindstrom, S. et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet* **43**, 185-7 (2011).