EPI 293: Analysis of Genetic Association Studies

Instructor
Liming Liang, Ph.D.
Associate Professor of Statistical Genetics
Department of Epidemiology and Department of Biostatistics
Email: liang@hsph.harvard.edu
Web: http://www.hsph.harvard.edu/faculty/liming-liang/
Office Phone: 617.432.5896
Office: Building 2, Room 207, HSPH

Teaching Assistant and Office Hour
Qianyu Yuan, qyyuan@g.harvard.edu
Jiaxuan (Jessie) Liu, j_liu@g.harvard.edu

Course site
https://canvas.harvard.edu/ and then select Analysis of Genetic Association Studies, 2019-2020 Winter.

Office hours
Qianyu Yuan: Tuesday, 1pm-2pm, Kresge 202A
Jiaxuan (Jessie) Liu, Thursday, 1pm-2pm, Kresge 202A
Liming Liang: by appointment, also welcome to stop by whenever office is not occupied

Emails for questions are always welcome and students have the top priority!

Lectures and Lab sections
All students are expected to attend all lectures and all labs. Please see the last page for exact date and location for each lecture and lab.

Objectives
At the end of this course students will grasp Concept and Theory, Methods and Software Tools needed to critically evaluate and conduct genetic association studies in unrelated individuals and family samples, including: basic molecular and population genetics, marker selection algorithms, haplotyping, multiple comparisons issues, population stratification, genome-wide association studies, genotype imputation, gene-gene and gene-environment interaction, analysis of
microarray data (including gene expression, methylation data analysis, eQTL mapping), next-generation sequencing data analysis and genetics simulation studies. Useful software tools will be introduced and practiced in labs and projects. Students interested in methodology development will find interesting research topics to pursue further. Students interested in application will learn cutting-edge methods and tools for their ongoing projects. Course materials will be updated according to the fast-growing areas of genetics, genomics, epigenomics and metabolomics.

Course note: Familiarity with SAS or S-PLUS/R and UNIX computing environment are highly recommended. Students are encouraged to discuss course prerequisites with the instructor.

Course Prerequisites
May be taken concurrently: BST 201(+alternative options) and EPI 201 (+alternative options) and PHS2000A

May not be taken concurrently: BST 210 or BST 213 or EPI 204

Texts and Reading Materials
Lecture notes and suggested readings are available at the course website.

Outcome Measures

Final project. Students will design and analyze a hypothetical genetic association study and summarize their methods and results in a brief report. Project details will be distributed at the course website.

Class participation. Every student will participate in class discussion, attend labs and complete an assignment for each lab.

Grading Criteria
Students will be graded on the final project (50%), lab assignment (30%) and class participation (20%). For homework and project report, providing only figures and tables without appropriate explanation would not receive credit.

Discussion of homework and final project are encouraged but students should first work out their solution independently, appropriately acknowledge contribution from others and submit their own work. No outsourcing.

Course Evaluations
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. Constructive comments and suggestions from students are always welcome and have made the course evolved.
<table>
<thead>
<tr>
<th>Topics</th>
<th>Date</th>
<th>Time</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lecture 1:</strong> Syllabus; Important concepts in molecular and population genetics</td>
<td>Jan 13 (M)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td>Lab 1: Introduction to genetic data format, software tools and Unix environment</td>
<td>Jan 14 (T)</td>
<td>9:00-11:00 AM</td>
<td>Kresge 202B</td>
</tr>
<tr>
<td><strong>Lecture 2:</strong> Haplotype, linkage disequilibrium and statistical review for association studies</td>
<td>Jan 14 (T)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td>Lab 2: SNPs tagging, LD computation and haplotype phasing</td>
<td>Jan 15 (W)</td>
<td>9:00-11:00 AM</td>
<td>Kresge 202B</td>
</tr>
<tr>
<td><strong>Lecture 3:</strong> Genome-wide association studies and genotype imputation</td>
<td>Jan 15 (W)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td>Lab 3: Let’s analyze a GWAS dataset</td>
<td>Jan 16 (R)</td>
<td>2:00-4:00 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td><strong>Lecture 4:</strong> Population stratification and genetics simulation studies</td>
<td>Jan 17 (F)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td><strong>Lecture 5:</strong> Gene-environment, gene-gene interaction and “pathway” analyses</td>
<td>Jan 21 (T)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td>Lab 4: Tools to account for population stratification, simulation of genetic sequence data</td>
<td>Jan 22 (W)</td>
<td>9:00-11:00 AM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td><strong>Lecture 6:</strong> Association analysis for quantitative traits and microarray data</td>
<td>Jan 22 (W)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td>Lab 5: Gene expression and DNA methylation data processing and analysis, tools for sequencing data analysis</td>
<td>Jan 23 (R)</td>
<td>2:00-4:00 PM</td>
<td>Kresge 202A</td>
</tr>
<tr>
<td><strong>Lecture 7:</strong> Rare variants and next-generation sequencing data analysis; Course review</td>
<td>Jan 24 (F)</td>
<td>2:00-5:15 PM</td>
<td>Kresge 202A</td>
</tr>
</tbody>
</table>