**EPI 293: Analysis of Genetic Association Studies**

**(2024 Winter)**

**Instructor**

Liming Liang, Ph.D.  
Professor of Statistical Genetics  
Department of Epidemiology and Department of Biostatistics  
Email: [lliang@hsph.harvard.edu](mailto:lliang@hsph.harvard.edu)  
Web: <https://www.hsph.harvard.edu/liming-liang/>   
Assistant: Mars Murray ([mariemurray@hsph.harvard.edu](mailto:mariemurray@hsph.harvard.edu))   
HSPH office: Building 2, 249A  
Online Office: <https://harvard.zoom.us/my/liang>

**Teaching Assistant**

Buu Truong <btruong@hsph.harvard.edu>

**Office hours**

Buu: Thursday, 4:15pm-5:15pm (Building 2, Room 205)  
Liming: by appointment, please email Liming or Mars Murray.   
**Emails for questions are always welcome and students have the highest priority!**

**Course site**

<https://canvas.harvard.edu/> and then select **Analysis of Genetic Association Studies, 2024 Winter**. All course documents will be distributed under corresponding folders accessible through the “**Files**” item on the left menu on Canvas. Required data for labs and projects will be distributed on the FAS cluster (instruction provided during the labs).

**Slack workspace**

**EPI 293 1 (Wi24):** <https://epi-293-w24-dc2.slack.com> All students will be invited to the Slack workspace. Questions and discussions related to the course are always encouraged.

**Lectures and Lab sections**

All students are expected to attend all lectures and all labs in person. Please see the last page for exact date/time 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 and genomic 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 genomics and other omics studies.

**Course note**Familiarity with R and UNIX computing environment are highly recommended. Source codes for all lab projects will be provided. Students are encouraged to discuss course prerequisites with the instructor.

**Course Prerequisites**BST201 and (BST210 or BST213 or EPI204) and (ID200 or EPI200 or EPI201 or EPI505 or EPI500 or ID201); may not be taken concurrently.

**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.*

**Policy regarding the use of Generative AI:** you may treat generative AI as one of your classmates, who might or might not provide correct answer.

**Data sharing:** please do not post course materials on public repositories, e.g., GitHub.

**Course Evaluations**Course evaluations for the Fall, WinterSession and Spring public have returned to the school’s usual policy. The school has implemented a new student course evaluation system platform, called Blue Explorance, and is now integrated into Canvas. *Constructive comments and suggestions from students are always appreciated and have made the course evolved.*

|  |  |  |
| --- | --- | --- |
| **Topics** | **Date** | **Time** |
| **Lecture 1**: Syllabus; Important concepts in molecular and population genetics | Jan 08 (M) | 2:00-5:15 PM |
| Lab 1: Introduction to genetic data format, software tools and Unix environment | Jan 09 (T) | 9:00-11:15 AM |
| **Lecture 2**: Haplotype, linkage disequilibrium and statistical review for association studies | Jan 09 (T) | 2:00-5:15 PM |
| Lab 2: SNPs tagging, LD computation and haplotype phasing | Jan 10 (W) | 9:00-11:15 AM |
| **Lecture 3**: Genome-wide association studies and genotype imputation | Jan 10 (W) | 2:00-5:15 PM |
| Lab 3: Let’s analyze a GWAS dataset | Jan 11 (R) | 2:00-4:00 PM |
| **Lecture 4**: Population stratification and genetics simulation studies | Jan 12 (F) | 2:00-5:15 PM |
| **Lecture 5**: Gene-environment, gene-gene interaction and “pathway” analyses | Jan 16 (T) | 2:00-5:15 PM |
| Lab 4: Tools to account for population stratification, simulation of genetic sequence data | Jan 17 (W) | 9:00-11:00 AM |
| **Lecture 6**: Association analysis for quantitative traits and microarray data | Jan 17 (W) | 2:00-5:15 PM |
| Lab 5: Gene expression and DNA methylation data processing and analysis, tools for sequencing data analysis | Jan 18 (R) | 2:00-4:00 PM |
| **Lecture 7**: Rare variants and next-generation sequencing data analysis; Course review | Jan 19 (F) | 2:00-5:15 PM |

**2024 Winter Schedule**