**BST 227: Introduction to Statistical Genetics**

**Fall 2 2023**

**Monday and Wednesday 3:45 - 5:15 pm FXB G13**

**Instructor Information**

**Faculty**

Martin Aryee

Associate Professor, Department of Data Science, Dana Farber Cancer Institute

Associate Professor, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Institute Member, Broad Institute of MIT and Harvard

Email: martin.aryee@ds.dfci.harvard.edu

Office hours: TBD

Sharon Lutz

Associate Professor, Department of Population Medicine, HPHC/Harvard Medical School

Assistant Professor, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Email: [smlutz@hsph.harvard.edu](mailto:smlutz@hsph.harvard.edu)

Office hours: TBD

**Teaching** **Assistants**

Yushan Xu

Email: yushan\_xu@hsph.harvard.edu

Zoom for the office hour (link on canvas)

Office hours: Thursday 1-2pm

Julie-Alexia Dias

Email: juliealexiadias@hsph.harvard.edu

Zoom for the office hour (link on canvas)

Office hours: Friday 1-2pm

**Laboratory Sections**

Mondays 5:30 pm - 7:30 pm iand Wednesdays 5:30 pm - 7:30 pm in FXB G11.

Lab 1A: Introduction to R

Lab 1B:  Biology Refresher

Lab 2: Concepts in linkage/association

Lab 3: GWAS

Lab 4: Rare variants

Though the labs are optional, we encourage you to attend. Note the lab sections are not every week. Lab 1A is an introduction to R. Lab 1B is a biology refresher focused on DNA, transcription, translation, mitosis, and meiosis. Lab 2 on linkage and association corresponds to HW1. Lab 3 on GWAS corresponds to HW2&3. Lab 4 on rare variants corresponds to HW4. If you have never used plink before, we **highly recommend** that you attend Lab 3 since plink is needed for HW3.

**Credits**

2.5 credits

**Course Description**

This course will provide an introduction to statistical methods for genetic studies. The emphasis of this course is on understanding basic concepts and methods and how they are applied in the analysis of population genetic data. The course starts out with an overview of methods for identifying disease-associated variants, and then goes on to cover approaches for interpreting the role of these variants in disease. A background in genetics is not necessary. The basic concepts of human and molecular genetics relevant to the course will be covered in class and the bioblogy refresher lab.

* **Pre-Requisites**

BST210 (Concurrent Enrollment Allowed) or PHS2000A (Concurrent Enrollment Allowed)

**Learning Objectives**

Upon successful completion of this course, you should be able to:

* Understand the basic concepts and methods in the analysis of population genetic data
* Run basic Genome Wide Association Studies (GWAS)
* Understand methods for sequencing data, rare variants, and Epigenome-Wide Association Studies (EWAS)

**Course Readings**: There is no required text book. There are supplemental articles on the canvas website.

* Suggested Textbook:

The Fundamentals of Modern Statistical Genetics, Nan Laird and Christoph Lange, Springer 2011.

* Additional Reference Books

Statistics in Human Genetics by Pak Sham

Statistical Methods for Genetic Epidemiology by Duncan Thomas

**Course Structure**

***Canvas Course Website:*** *https://canvas.harvard.edu/courses/62208*

**Grading, Progress and Assessment**

The final grade for this course will be based on:

* Homework assignments (60%)
* Final project (40%)

**Homework assignments (60%)**

There will be 4 homework assignments due on Fridays to be submitted via the online course dropbox in Canvas by midnight EST. Students should feel free to discuss approaches to solving the problems in working through homework problems, but each student must turn in their own solution, written entirely in his or her own words and not copied from another source. In addition, you should not share your final solutions with another student.

**Final project (40%)**

The class will be divided into teams which will complete a project, write a report, and present their results the last week of class.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Week** | **Date** | **Topic** | **Lecturer** | **Lab/Notes** | **Assignments** |
| Week 1 | 10/23 | Introduction and overview of genetic disease | Dr. Aryee | **Lab 1A:** Intro to R  **Lab 1B:** Biology refresher |  |
| 10/25 | Mendel’s laws & Hardy Weinberg Equilibrium | Dr. Lutz |  |  |
| Week 2 | 10/30 | An overview of linkage and association | Dr. Lutz | **Lab 2:** Concepts in linkage/association |  |
| 11/1 | Population substructure & family-based association tests (FBAT) | Dr. Lutz |  | **HW 1**  Due 11/3 |
| Week 3 | 11/6 | Genome wide association studies (GWAS) | Dr. Lutz | **Lab 3:** GWAS |  |
| 11/8 | Meta-analyses & multiple testing | Dr. Lutz |  | **HW 2**  Due 11/10 |
| Week 4 | 11/13 | Pleiotropy & polygenic risk scores (PRS) | Dr. Lutz |  |  |
| 11/15 | Analysis of rare variants and non-SNP variation | Dr. Aryee |  | **HW 3**  Due 11/17 |
| Week 5 | 11/20 | Variant calling from high-throughput sequencing data | Dr. Aryee | **Lab 4**: Rare variants |  |
| 11/22 | **Thanksgiving Recess – No class** |  | No office hours/ lab 11/23-11/25 |  |
| Week 6 | 11/27 | Using the 3-dimensional organization of DNA to interpret variants | Dr. Aryee |  |  |
| 11/29 | Cancer and somatic variants | Dr. Getz |  | **HW 4**  Due 12/1 |
| Week 7 | 12/4 | Epigenome-wide Association Studies (EWAS) | Dr. Aryee |  |  |
| 12/6 | Editing DNA to interpret variants | Dr. Aryee |  |  |
| Week 8 | 12/11 | Special topics in statistical genetics | Dr. Christensen |  |  |
| 12/13 | Project presentations |  |  | **Project presentations** |

**Course Schedule & Assessment of Student Learning**

* Please note, session topics, and activities may be subject to change during the course