BSTA 787: Methods for Statistical Genetics and Genomics in Complex Human Diseases
Spring 2022

Description: This is an introductory course for graduate students in Biostatistics, Epidemiology, Computational Biology and other BGS disciplines, which will cover statistical methods for the analysis of genetic and genomic data. Topics covered include twin studies, linkage analysis, population-based and family-based association, haplotype analysis, genome-wide association studies (GWAS), analysis of copy number and structural variations, next-generation sequencing data (DNA), analysis of bulk RNA-seq data, single-cell RNA-seq data, single-cell multi-omics data integration, and spatial transcriptomics data. Students will be exposed to the latest methodology and computer tools in statistical genetics and genomics.

Prerequisites: Introductory graduate-level courses in statistics or biostatistics (e.g., BSTA 630-631, EPID 520-521), or permission of the instructor.

Instructors: Rui Feng, PhD, ruifeng@upenn.edu, 209 Blockley Hall
Mingyao Li, PhD, mingyao@upenn.edu, 213 Blockley Hall

Lecture Time: 1:45pm-3:15pm Tuesday and Thursday

Location: 701 Blockley Hall

Online Zoom Link:
https://pennmedicine.zoom.us/j/91513252047?pwd=ZDhqcE1GcmppaXVFUjNKdDdqalBKQT09

Office Hour: 3:15pm-4:15pm Thursday or by appointment

Evaluation:

1. Two homework assignments (25% of the final grade). Submit your homework on Canvas in word or pdf. Late homework is NOT accepted.
2. Two paper presentation (35% of the final grade). Students will form 2-person teams and each team will present two papers, one in the 1st half and the other in the 2nd half of the semester. The presenters are expected to clearly explain the content of the assigned paper and lead discussions (about 75 minutes) during class. All non-presenters are required to submit their questions in Zoom before each presentation starts.
3. Final project (20% of the final grade). Students will submit a written report (at least 2,000 words in the format of a scientific article with Abstract, Introduction, Methods, Results, and Discussion) at the end of the semester. Students can 1) analyze real data using existing methods or new methods proposed by yourself, or 2) evaluate existing methods by simulations and applications to real data, or 3) do a comprehensive literature review on a specific topic. You are welcome to discuss your project with the instructor. The abstract of the project is due on April 1, 2022. The final project is due on May 6, 2022.
4. Lecture attendance and class participation (20% of the final grade).

Recommended (not Required) Textbook:

1. A Statistical Approach to Genetic Epidemiology: Concepts and Applications. Andreas Ziegler, Inke R. Koenig (2006).

2. Handbook on Analyzing Human Genetic Data: Computational Approaches and Software. Shili Lin, Hongyu Zhao (2010). http://www.springerlink.com/content/978-3-540-69263-8

3. Statistical Analysis of Next Generation Sequencing Data. Somnath Datta, Daniel Nettleton (2014). http://www.springer.com/us/book/9783319072111.

4. Computational Methods for Single-Cell Data Analysis. Guo-Cheng Yuan (2019). https://www.springer.com/gp/book/9781493990566.

5. Museum of Spatial Transcriptomics. Lambda Moses, Lior Pachter (2021). https://www.biorxiv.org/content/10.1101/2021.05.11.443152v2