BSTA 787: Methods for Statistical Genetics and Genomics in Complex Human Diseases
Spring 2025

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 linkage analysis, association analysis, haplotype analysis, genome-wide association studies (GWAS), genotype imputation, expression quantitative trait locus analysis (eQTL) and allele-specific expression (ASE) analyses, 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 Xiao, PhD, rxiao@pennmedicine.upenn.edu, xiaor@chop.edu, 206 Blockley Hall
Mingyao Li, PhD, mingyao@pennmedicine.upenn.edu, 213 Blockley Hall

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

Location: 252 BRB

Office Hour: by appointment

Evaluation:

1. Two homework assignments (**10%** of the final grade). Submit your homework on Canvas in word or pdf. Late homework is NOT accepted.
2. Two paper presentations (**40%** of the final grade). Students will form a 2-person team and each team will select and present one paper in the 1st half of the semester and another paper in the 2nd half of the semester. The papers will be selected from the list provided by the instructors. 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 questions to Canvas before the presentation starts. Non-presenters are also required to participate in the discussion when their questions are discussed.
3. Final project and presentation (**30%** 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) and present the project (~10 minutes) 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 approved by the instructors. You are welcome to discuss your project with the instructors. The abstract of the project is due on **April 4th, 2025**. The final project is due on **May 13th, 2025**.
4. Class participation and paper presentation discussion (**10%** of the final grade).
5. Lecture attendance (**10%** 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 (2022). https://www.nature.com/articles/s41592-022-01409-2

BSTA 787 2025 Spring Course Schedule

#	Date	Content	Instructor
		1st Half focusing on statistical genetics	
1	1/16	Introduction	Xiao
2	1/21	Linkage analysis	Xiao
3	1/23	Basics of genetic association studies + Haplotype analysis	Xiao
4	1/28	GWAS	Xiao
5	1/30	Genotype imputation	Guest Lecture (Dr. Quan Sun)
6	2/4	Paper presentation (1): **Genomewide Association Study of Severe Covid-19 with Respiratory Failure.** *New England Journal of Medicine* 383(16):1522-1534 (2020)	Students
7	2/6	eQTL + allele-specific expression (ASE)	Xiao
8	2/11	Paper presentation (2): **Fast and efficient QTL mapper for thousands of molecular phenotypes**. *Bioinformatics*. 32(10): 1479-1485 (2016)	Students
9	2/13	Polygenic risk scores	Guest Lecture (Dr. Jin Jin)
10	2/18	Paper presentation (3): **A robust and efficient method for Mendelian randomization with hundreds of genetic variants.** *Nature Communications* 11, 376 (2020)	Students
11	2/20	Paper presentation (4): **MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies.** *Nature Genetics* 56(1):170-179 (2024)	Students
12	2/25	Paper presentation (5): **OTTERS: a powerful TWAS framework leveraging summary-level reference data.** *Nature Communications* 14(1):1271 (2023)	Students
13	2/27	Paper presentation (6): **A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies.** *Nature Methods* 19(12):1599-1611 (2022)	Students
		2nd Half focusing on statistical genomics	
14	3/4	RNA-seq and single-cell RNA-seq	Li
15	3/6	DNA variant calling	Guest Lecture (Dr. Tristan Hayeck)
16	3/18	Single-cell RNA-seq clustering, annotation, and imputation	Li
17	3/20	Paper presentation (7): **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biology 15:550* (2014)	Students
18	3/25	Long read sequencing	Guest Lecture (Dr. Kai Wang)
19	3/27	Cell type deconvolution	Li
20	4/1	Paper presentation (8): **Dictionary learning for integrative, multimodal and scalable single-cell analysis.** *Nature Biotech* (2023)	Students
21	4/3	Spatial transcriptomics	Li
22	4/8	Paper presentation (9) **Search and match across spatial omics samples at single-cell resolution.** *Nature Methods* (2024)	Students
23	4/10	Modeling histology in spatial omics analysis	Li

24	4/15	Paper presentation (10): **Multiscale topology classifies cells in subcellular spatial transcriptomics.** *Nature* (2024)	Students
25	4/18	Paper presentation (11): **Probe set selection for targeted spatial transcriptomics.** *Nature Methods* (2024)	Students
26	4/22	Paper presentation (12): **CelloType: a unified model for segmentation and classification of tissue images.** *Nature Methods* **(2024)**	Students
27	4/24	Final Project Presentation	Students
28	4/29	Final Project Presentation	Students