BSTA 787: Methods for Statistical Genetics and Genomics in Complex Human Diseases
Spring 2026

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, spatial transcriptomics, and other spatial omics 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, CTH03-312
Mingyao Li, PhD, mingyao@pennmedicine.upenn.edu, CTH03-201A, 3600 Civic Center Blvd

Lecture Time: 1:45 - 4:45 pm 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. One paper presentation (**40%** of the final grade). Each student will select and present one paper in either the 1st half or 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 paper and lead discussions (about 60 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 3rd, 2026**. The final project is due on **May 12th, 2026**.
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 2026 Spring Course Schedule (tentative)

#	Date	Content	Instructor
		1st Half focusing on statistical genetics	
1	1/15	Introduction	Xiao
		Linkage analysis	Xiao
2	1/22	Basics of genetic association studies + Haplotype analysis	Xiao
		GWAS	Xiao
3	1/29	Genotype imputation	Guest Lecture (Dr. Quan Sun)
		eQTL + allele-specific expression (ASE)	Xiao
4	2/5	Paper presentation (1): **Genomewide Association Study of Severe Covid-19 with Respiratory Failure.** *New England Journal of Medicine* 383(16):1522-1534 (2020)	Students
		Paper presentation (2): **Fast and efficient QTL mapper for thousands of molecular phenotypes**. *Bioinformatics*. 32(10): 1479-1485 (2016)	Students
		Paper presentation (3): **A framework for detecting noncoding rare-variant associations of large-scale whole-genome sequencing studies.** *Nature Methods* 19(12):1599-1611 (2022)	Students
5	2/12	Microbiome analysis	Guest Lecture (Dr. Xiaowei Zhan)
		Polygenic risk scores	Guest Lecture (Dr. Jin Jin)
6	**2/19**	**No Class**	
7	2/26	Paper presentation (4): **OTTERS: a powerful TWAS framework leveraging summary-level reference data.** *Nature Communications* 14(1):1271 (2023)	Students
		Paper presentation (5): **A robust and efficient method for Mendelian randomization with hundreds of genetic variants.** *Nature Communications* 11, 376 (2020)	Students
		2nd Half focusing on statistical genomics	
8	3/5	RNA-seq	Li
		Single-cell RNA-seq	Li
9	3/19	DNA variant calling	Guest Lecture (Dr. Tristan Hayeck)
		Paper presentation (6): **Fast, sensitive and accurate integration of single-cell data with Harmony.** *Nature Methods* (2019)	Students
10	3/26	Denoising in scRNA-seq and cell type deconvolution	Li
		Spatial transcriptomics	Li
11	4/2	Modeling histology in spatial omics analysis	Li
		Paper presentation (7): **Accurate and efficient integrative reference-informed spatial domain detection for spatial transcriptomics.** *Nature Methods* (2024)	Students
12	4/9	Technology advances in spatial multiomics	Guest Lecture (Dr. Yanxiang Deng)
		Paper presentation (8): **Novae: a graph-based foundation model for spatial transcriptomics data.** *Nature Methods* (2025)	
13	4/16	Multi-modal spatial omics	Li

		Paper presentation (9): **Multimodal AI generates virtual population for tumor microenvironment modeling.** *Cell* (2025)	Students
14	4/23	Paper presentation (10): **Bridging the dimensional gap from planar spatial transcriptomics to 3D cell atlases.** *Nature Methods* (2025)	Students
		Final Project Presentation	Students
15	4/29	Final Project Presentation	Students