

BSTA 787: Methods for Statistical Genetics and Genomics in Complex Human Diseases Spring 2024

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@penncmedicine.upenn.edu, xiaor@chop.edu, 206 Blockley Hall
Mingyao Li, PhD, mingyao@penncmedicine.upenn.edu, 213 Blockley Hall

Lecture Time: 10:15 -11:45 am Tuesday and Thursday

Location: 418 Blockley Hall

Office Hour: by appointment

Evaluation:

1. Two homework assignments (20% of the final grade). Submit your homework on Canvas in word or pdf. Late homework is NOT accepted.
2. One paper presentation (25% of the final grade). Students will form 2-person teams and each team will select and present one paper from the list provided by the instructors, either in the 1st half or 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 write down their questions and submit the questions to the instructor 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 (~15 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 12, 2024**. The final project is due on **May 10, 2024**.
4. Class participation and paper presentation discussion (15% 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 (2021). <https://www.biorxiv.org/content/10.1101/2021.05.11.443152v2>

BSTA 787 2024 Spring Course Schedule

#	Date	Content	Instructor
		1st Half focusing on statistical genetics	
1	1/18	Introduction	Xiao
2	1/23	Linkage analysis	Xiao
3	1/25	Basics of genetic association studies + Haplotype analysis	Xiao
4	1/30	GWAS	Xiao
5	2/1	Genotype imputation	Xiao
6	2/6	Paper presentation (1): Genomewide Association Study of Severe Covid-19 with Respiratory Failure. <i>New England Journal of Medicine</i> 383(16):1522-1534 (2020)	Students
7	2/8	eQTL + allele-specific expression (ASE)	Xiao
8	2/13	Paper presentation (2): Fast and efficient QTL mapper for thousands of molecular phenotypes. <i>Bioinformatics.</i> 32(10): 1479-1485 (2016)	Students
9	2/15	Polygenic Risk Scores	Guest Lecture
10	2/20	Paper presentation (3): A robust and efficient method for Mendelian randomization with hundreds of genetic variants. <i>Nature Communications</i> 11, 376 (2020)	Students
11	2/22	Paper presentation (4): STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data. <i>Nature Methods</i> 19: 1532–1533 (2022)	Students
12	2/27	Paper presentation (5): OTTERS: a powerful TWAS framework leveraging summary-level reference data. <i>Nature Communications</i> 14(1):1271 (2023)	Students
		2nd Half focusing on statistical genomics	
13	2/29	Intro of RNA-seq and single-cell RNA-seq	Li
14	3/12	Long read sequencing	Guest Lecture
15	3/14	Single-cell RNA-seq clustering and annotation	Li
16	3/19	Paper presentation (6): Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. <i>Genome Biology</i> 15:550 (2014)	Students
17	3/21	Gene expression imputation in single-cell RNA-seq	Li
18	3/26	Paper presentation (7): Dictionary learning for integrative, multimodal and scalable single-cell analysis. <i>Nature Biotech</i> (2023)	Students
19	3/28	Cell type deconvolution analysis	Li
20	4/2	Intro to spatial transcriptomics	Li
21	4/4	Paper presentation (8): Robust alignment of single-cell and spatial transcriptomes with CytoSPACE. <i>Nature Biotech</i> (2023)	Students
22	4/9	How to model histology image information in spatial transcriptomics	Li
23	4/11	Paper presentation (9): Integration of spatial and single-cell data across modalities with weakly linked features. <i>Nature Biotech</i> (2023)	Students
24	4/16	Paper presentation (10): CellCharter reveals spatial cell niches associated with tissue remodeling and cell plasticity. <i>Nature Genetics</i> (2024)	Students
25	4/18	Final Project Presentation	Students
26	4/23	Final Project Presentation	Students
27	4/25	Final Project Presentation	Students
28	4/30	Final Project Presentation	Students