BSTA 787: Methods for Statistical Genetics and Genomics in Complex Human Diseases
Spring 2023

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), expression quantitative trait locus analysis (eQTL) and allele-specific expression (ASE) analyses, analysis of copy number 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 Xiao, PhD, rxiao@pennmedicine.upenn.edu, xiaor@chop.edu, 206 Blockley Hall
Mingyao Li, PhD, mingyao@pennmedicine.upenn.edu, 213 Blockley Hall

Lecture Time: 12:00pm-1:30pm Tuesday and Thursday

Location: 701 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 present one paper 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 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 (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) 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 7, 2023**. The final project is due on **May 5, 2023**. Each student will do a 10-minute presentation about their final project at the end of the semester.
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 2023 Spring Course Schedule

#	Date	Content	Instructor
		1st Half focusing on statistical genetics	
1	1/12	Introduction	Xiao
2	1/17	Linkage analysis	Xiao
3	1/19	Basics of genetic association studies + population stratification	Xiao
4	1/24	Haplotype analysis	Xiao
5	1/26	GWAS	Xiao
6	1/31	eQTL + allele-specific expression (ASE)	Xiao
7	2/2	Paper presentation: **Genomewide Association Study of Severe Covid-19 with Respiratory Failure.** *New England Journal of Medicine* 383(16):1522-1534 (2020)	Students
8	2/7	Imputation	Xiao
9	2/9	Paper presentation: **Fast and efficient QTL mapper for thousands of molecular phenotypes**. *Bioinformatics*. 32(10): 1479-1485 (2016)	Students
10	2/14	Paper presentation: **The personal and clinical utility of polygenic risk scores.** *Nat Rev Genet* **19:** 581–590 (2018)	Students
11	2/16	Copy number variations (CNVs)	Guest Lecture
12	2/21	Paper presentation: **STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data.** *Nat Met* 19**:** 1532–1533 (2022)	Students
13	2/23	Paper presentation: **Integrative approaches for large-scale transcriptome-wide association studies.** *Nat Genet* 48(3):245-52 (2016)	Students
		2nd Half focusing on statistical genomics	
14	2/28	Intro of RNA-seq and single-cell RNA-seq	Li
15	3/2	Single-cell RNA-seq clustering and cell type annotation	Li
16	3/14	Paper presentation: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biology 15:550* (2014)	Students
17	3/16	Denoising in single-cell RNA-seq	Li
	3/21	No class	
18	3/23	Paper presentation: **Joint probabilistic modeling of single-cell multi-omic data with totalVI.** *Nature Methods 18:272-282* (2021)	Students
19	3/28	Cell-type deconvolution in RNA-seq	Li
20	3/30	Intro of spatial transcriptomics -- clustering and deconvolution	Li
21	4/4	Paper presentation: **Cell type-specific inference of differential expression in spatial transcriptomics.** *Nature Methods* 19:1076-1087 (2022)	Students
22	4/6	How to model histology information in spatial transcriptomics	Li
23	4/11	Super-resolution gene expression in spatial transcriptomics	Li
24	4/13	Paper presentation: **Robust alignment of single-cell and spatial transcriptomes with CytoSPACE.** *BioRxiv* **doi:** https://doi.org/10.1101/2022.05.20.488356 (2022) (accepted by *Nature Biotechnology*)	Students
25	4/18	Integration of single-cell RNA-seq and spatial transcriptomics	Li
26	4/20	Paper presentation: **Graph deep learning for the characterization of tumor microenvironments from spatial protein profiles in tissue specimens.** *Nature Biomedical Engineering, in press* (2022)	Students
27	4/25	Final Project Presentation	Students
28	4/27	Final Project Presentation	Students