# Statistical Genetics

## Syllabus

**Course description:** This course will cover basic concepts, methods, and software tools needed to critically evaluate and analyze genetics and genomics data. Specifically, we will cover the following topics: 1) basic concepts in molecular and population genetics, 2) genome-wide association studies, 3) gene-level association analysis, 4) polygenic risk score, and 5) Mendelian randomization.

**Course objectives:** While the field of statistics is at a crossroads, genetics and genomics data impose novel and unique challenges for statisticians. How to extract useful information from big genetics and genomics data (sometime more than 10 TB) remains an active research area. At the end of this course, students will grasp the necessary background knowledge to understand research advancements in statistical genetics. Students who are interested in methodology development will find interesting research topics to pursue further. Students who are interested in applied research will learn cutting-edge methods and tools.

**Prerequisite courses:** STA 5167 and STA 5326.

**Recommended prerequisite:** STA 5168.

Note: This course is computationally extensive. Familiarity with R, Python, and UNIX computing environment is highly recommended. Students are encouraged to discuss course prerequisites with the instructor.

**Lecture:** TBA

**Instructor:** Chong Wu

**Email:** cwu3@fsu.edu; **Office:** Love Building 309 ; **Office Hour:** TBA

**Teaching Assistant:** TBA

**Email:** TBA; **Office:** Love Building 430; **TA Office Hour:** 4:00–5:00 PM, Monday, or by appointment

**Class Website:** https://fsu.instructure.com/

**Texts and Reading Materials:** Due to the fast-growing nature of statistical genetics fields, we do not have a textbook that meets our needs. However, several Lecture notes and suggested readings will be available at the course website.

**Textbook (optional):** *Probability and Statistics with Applications: A Problem Solving Text*, by Leonard A. Asimow and Mark M. Maxwell, ACTEX Publications, 2010.

**Homework:** The homework assignments will be posted roughly biweekly. Students are encouraged to discuss with each other but should work independently and submit their own work.

**Final project:** Students will design and analyze real genetics and genomics datasets and summarize their methods and results in a brief report. Students need to find their teammates, propose a research question, analyze real datasets, and write a report. Novel methods developments are also encouraged. The final project details will be distributed at the course website.

**Presentation:** Formal presentation is really important for both academia and industry jobs. The final project needs to be presented and will be evaluated based on the quality of the presentation.

**Final grade:** It will be determined by a weighted average of the following items: (1) homework(40%), (2) final project (50%), and (3) presentation (10%).

**Course Policies**

· Classroom policies: The classroom environment is an important factor for effective learning. In order to not distract other students' attention please follow these classroom policies. The first one of these is the university policy. Remember that no food or drinks are allowed in the classroom. Turn off all audible alarms (cell phones, pagers, calculators, watches etc.) Do not use cell phones in the class. Come to the class on time. Opening and closing the classroom door in the middle of a class cause distraction to the students and the teacher. Do not talk to other students without permission while the professor is teaching. More than one conversation creates noise and makes it difficult for the students to pay attention to the lecture.

· Attendance: You are required to attend all classes. The class activities will help you assimilate the lessons more easily, giving you an opportunity for active learning. Do not let this opportunity slip away. Any foreseen absence must be cleared with the instructor. If the absence is due to emergencies, it is the student's responsibility to notify the instructor at the earliest opportunity of the emergency.

· **Collecting returned homework/exam:** It is the student's responsibility to retrieve his or her homework/exam whenever they are returned and to check grades on the class website. If you notice any mistake in recording grades, please inform the instructor about it as soon as possible but no late than one week from grades been posted online.

· Homework re-grade: You have one week to request a re-grade of a homework from the date on which the graded homework is available to the students of the class. Submit a written request detailing the nature of the grading error to the instructor along with the relevant homework.

· Contacting the instructor or TA outside the class: You are strongly encouraged to come to the instructor or TA during their office hours. If your schedule conflicts with the office hours, you can make an appointment. You may ask the instructor or TA brief questions by e-mail, but you may be asked to come to office hours if the instructor or TA thinks that the questions are better answered in person. When you send e-mails, remember the following: always send e-mails from your FSU accounts. The e-mails from non-FSU accounts may not reach me due to filters. Always write your full name at the end of each e-mail message you send.

· Academic honor policy: The Florida State University Academic Honor Policy outlines the University expectations for the integrity of students' academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process. Students are responsible for reading the Academic Honor Policy and for living up to their pledge to "… be honest and truthful and … [to] strive for personal and institutional integrity at Florida State University." (Florida State University Academic Honor Policy)

· Students with disabilities: Students with disabilities in need of academic accommodation should: 1. Register with and provide documentation to the Student Disability Resource Center; 2. Bring a letter to the instructor indicating the type of accommodation needed. This should be done during the first week of class. See https://dos.fsu.edu/sdrc/ for more information.

**Detailed teaching plan**

| Topics | Details |
|---|---|
| **Lecture 1** Syllabus; Important concepts in molecular and population genetics | Linkage disequilibrium, HWE, heritability, etc. |
| **Lecture 2** Genome-wide association studies | Plink, genotype imputation, population stratification, mixed-effect models, etc. |
| **Lecture 3** Gene-level association analysis | SKAT, aSPU, PrediXcan, GWAS summary data, TWAS, etc. |
| **Lecture 4** Polygenic risk score | P+T, LDpred, LassoSum, etc. |
| **Lecture 5** Two sample Mendelian randomization | IVW, robust MR, practical guidance, etc. |
| **Lecture 6** Advanced topics in statistical genetics (selective) | Whole genome-sequencing data analysis, eQTL, sQTL, and DNA methylation |
| **Lecture 7** Final project presentation | Each group will have ten minutes for presentation and five minutes for Questions |

# Email to be sent to students

Hi All,

This is Chong Wu, an Assistant Professor in Statistics. I am writing to recommend one new course (Statistical Genetics) that will be offered in 2020 FALL. In this course, we will learn basic concepts, methods, and software tools needed to critically evaluate and analyze genetics and genomics data. Specifically, we will cover the following topics: 1) basic concepts in molecular and population genetics, 2) genome-wide association studies, 3) gene-level association analysis, 4) polygenic risk score, and 5) Mendelian randomization.

While statistics is at a cross-road, training for big data and expertise of data science techniques (e.g., how to handle big data) goes a long way to help your career, no matter in academia or industry. This new course will not only pave the road to the statistical genetics but also provide opportunities to conducting some real research, which may have some impact and improve our understanding of the genetic basis of many complex diseases.

My teaching phoiolsy is to help students learn the beauty of statistics. I view statistics as a powerful tool for extracting information from messey real-world data. While theory can provide critical scaffolding for practice, this course will focus on solving real-world problems by the state-of-the-art methods. That's the reason why we do not have a textbook and need to read many pretty new papers.

Because of the nature of the course setting (involve research paper reading and real projects), I suggest only second year (or above) students think about taking this course. Please feel free to contact me if you have any questions or are not sure if you can take it.

Thanks,

Chong

# Email to be sent to Faculty

(If the enrolment is low, I will send out this email)

Hi All,

I will offer a new course (Statistical Genetics) in 2020 FALL. In this course, we will learn basic concepts, methods, and software tools needed to critically evaluate and analyze genetics and genomics data. Specifically, we will cover the following topics: 1) basic concepts in molecular and population genetics, 2) genome-wide association studies, 3) gene-level association analysis, 4) polygenic risk score, and 5) Mendelian randomization.

Currently, enrolment is low. I wonder if you can recommend this course to your students. Thank you for your help.

Thanks,