

The University of North Carolina at Chapel Hill

Fall Semester 2016

BIOS 782 Statistical Methods in Genetic Studies

I. TIME AND PLACE MW 1:25–2:40 pm MC 2301

II. INSTRUCTORS

<p>Danyu Lin, Ph.D. Dennis Gillings Distinguished Professor Office: MC 3101E Phone: 919 843-5134 E-mail: lin@bios.unc.edu Office hours: M 3:30–4:30 pm or by appointments</p> <p>Yun Li, Ph.D. Associate Professor Office: MC 4115F Phone: 919 966-7255 E-mail: yunli@med.unc.edu Office hours: W 12:25–1:25 pm or by appointments</p>
--

III. TEACHING ASSISTANTS Alex Wong
Email: alexwky@live.unc.edu

IV. TEXTBOOK

(LL) Laird, N. M. and Lange, C. (2011). *The Fundamentals of Modern Statistical Genetics*. Springer.

V. COURSE DESCRIPTION

Recent scientific and technological advances have led to a proliferation of genetic studies, including genome-wide association studies, sequencing studies, and trans-omics studies. These studies present unique statistical challenges. There is a rich literature on statistical methods for the designs and analysis of modern genetic studies. It is important for statisticians to be exposed to these methods and to be able to apply them to biomedical studies. Furthermore, statistical genetics is a very active research area, presenting many exciting opportunities for the development of innovative and high-impact statistical methods.

This course provides a comprehensive survey of the statistical methods that have been recently developed for the designs and analysis of genetic studies. Specific topics include molecular and population genetics, association analysis, case-control sampling and secondary phenotypes, haplotypes

and untyped SNPs, copy number variation, population stratification, meta-analysis, multiple testing, next-generation sequencing studies, rare variants, trait-dependent sampling, integrative analysis, risk prediction, and RNA sequencing studies.

This course is targeted primarily at the Ph.D. students and will be taught at a rigorous statistical level. The students will learn the theoretical justifications for the methods as well as the skills to apply them to real studies. They will also be exposed to current research topics and open problems.

VI. COURSE FORMAT

- Some early lectures will draw materials from the textbook. Most lectures will be based primarily on recent journal articles.
- Lecture notes/slides will be posted on Sakai.
- Students are required to read appropriate textbook chapters and journal articles.
- Students are required to attend lectures and participate in discussions.
- There will be several homework assignments. The solutions are due one week after assignments.
- There will be an in-class mid-term exam.
- There will be a final data analysis project. The students are required to submit a detailed report.
- The grade will be determined by homework (30%), mid-term exam (20%), final project (40%), and class attendance/participation (10%).

VII. LIST OF LECTURES

Likelihood inference and EM algorithm
Molecular genetics (LL Ch. 1)
Principles of inheritance (LL Ch. 2)
Population genetics (LL Ch. 3)
Coalescent models
Gene mapping (LL Ch. 5)
Genetic association analysis (LL Ch. 7)
Genome-wide association studies and PLINK (LL Ch. 11)
Population stratification (LL Ch. 8)
Case-control designs and secondary phenotypes
Analysis of haplotypes-disease association
Meta-analysis
Adjustment for multiple testing (LL §10.1)
Risk prediction
Integrative analysis
Hidden-Markov models
Haplotype phasing algorithms
Imputation of untyped SNPs
Next-generation sequencing studies

Association analysis with rare variants
Meta-analysis of sequencing studies
Trait-dependent sampling
Family studies
Copy number variation
RNA sequencing studies

Note: Relevant background materials from the textbook (LL) are shown in parentheses. Most lectures will draw materials from journal articles.