Genomics of Human Populations

Syllabus and Schedule

INSTRUCTOR: Professor Jonathan M. Flowers,

OFFICE: Center for Genomics and Systems Biology, 12 Waverly Place, 6th Floor, Room 607

EMAIL: jmf11@nyu.edu

SESSIONS AND LOCATION: Two 75 min sessions per week. Location TBD.

OFFICE HOURS: 2 hours per week (TBD) or by appointment

COURSE OVERVIEW:

This course covers topics in the field of human population genetics including human ancestry and admixture, linkage disequilibrium, natural selection in the human genome, genome-wide association studies (GWAS), genetic architecture of human traits and diseases, and application of population genomics to the study of cancer and disease. The course includes lecture and recitation components with the latter geared to teaching students fundamental skills in population genomic data analysis.

By the end of the course students will be able to:

· Understand principles of population genetics theory and its application to human population genomic datasets

· Conduct analysis of human population genomic datasets using standard methodological approaches

· Draw biological inferences from human population genomic data

PREREQUISITES:

Undergraduate Molecular Biology and Genetics; Introductory Statistics with R (recommended)

COURSE DETAILS:

The course consists of 28 sessions. Each week will be roughly divided into i) a lecture with discussion of assigned readings and a ii) recitation/workshop. The assigned reading will draw from a variety of sources including textbooks and primary research papers. The recitation/workshop section will consist of a hands-on tutorial on how to conduct population genomic data analysis using either the R statistical programming language or using third party software tools. The latter tools are typically command-line driven software (e.g., ADMIXTURE, PLINK, or GCTA) that generate primary outputs for subsequent analysis.

The course requires introductory level coding experience in R. A laptop computer will be needed. If you do not have access to one let the instructor know so arrangements can be made.

ASSESSMENTS:

Assessments for the course include 8 quizzes which will assess student's understanding of key concepts from the lectures. The 4 homework assignments provide students with an opportunity to analyze real population genomic dataset and draw biological inferences from the data. These homeworks will allow students to apply methods introduced in the recitation sessions and conduct analysis and interpretation using concepts introduced in the lectures.

RECOMMENDED TEXTS:

An Introduction to Population Genetics: Theory and Applications (2013) by Rasmus Nielsen and Montgomery Slatkin

Population Genomics with R (2020) by Emmanuel Paradis, CRC Press

GRADING:

Quizzes	10 points each (x 8) = 80 points
Homework	30 points each (x 4) = 120 points
Participation	20 points

ASSESSMENT:

“Participation” is evaluated as follows:

- Preparation of questions/criticisms/ideas based on the assigned reading.
- Responding to other students’ questions/criticisms/ideas.
- Participation in discussions in the recitation

Students will be evaluated separately for first half and second half of course (10 points each half)

Four problem sets will be assigned bi- or tri-weekly

Eight Quizzes will serve as knowledge checks from readings and lectures from the previous week.

POLICIES:

Students are expected to know and understand the policies on academic integrity, including University, CAS, and GSAS policies:

https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-for-students-at-nyu.html ; https://cas.nyu.edu/content/nyu-as/cas/academic-integrity.html; Also see the GSAS Statement on Academic Integrity: http://gsas.nyu.edu/content/nyu-as/gsas/about-gsas/policies-and-procedures/gsas-statement-on-academic-integrity.html.

If a student is caught cheating or plagiarizing, the Instructor may, at her or his discretion, give the student an academic sanction, which may include a warning and/or reduction of the grade on an assessment item (e.g. exam) or even the final course grade (in consultation with the Director of the Master's Program, who may meet with the faculty and the student to discuss the nature of the offense). Depending on the severity of the infraction, this could even mean failure of the student in the course. The student may appeal any grade reduction to the Director of the Master's Program. The departmental decision is final. In addition, any substantial case brought to the Director of the Master's Program must be referred to the Dean's office for possible disciplinary action.

If you have any questions or uncertainties about these policies, please consult the Instructor, Director of the Master's Program, or Dean's office.

COURSE SCHEDULE

WEEK 1: Introduction to Population Genomics I

Lecture: Mutation, Genetic Drift, and Hardy-Weinberg

Reading: Nielsen and Slatkin Chapters 1 and 2

Recitation: TBD

Homework: Problem set 1 (*due end of Week 3)

WEEK 2: Introduction to population genomics II

Lecture: Genomic approaches to studies of kinship and inbreeding

Reading: Nielsen and Slatkin Chapter 1 “Deviations from Hardy-Weinberg”

Recitation: Calculation of kinship, relatedness and inbreeding coefficients

Homework: Problem set 1 due

WEEK 3: Introduction to population genomics III

Lecture: Mutation, population genetic diversity and the site frequency spectrum

Reading: Nielsen and Slatkin Chapter 3

Recitation: Calculating population genetic summary statistics in R

Homework: Problem set 1 (*due end of Week 3)

WEEK 4: Admixture and Genomic Ancestry I

Lecture: Wahlund Effect, population subdivision, and population genomic approaches to inferring population structure.

Reading: Nielsen and Slatkin Chapter 4 and 5

Recitation: Population structure inferred from model-based clustering

Homework: Problem Set 2 (*due Week 6)

WEEK 5: Admixture and Genomic Ancestry II

Lecture: Distinguishing lineage sorting and admixture using population genomic data

Readings:

Green et al. 2010. A Draft Sequence of the Neanderthal Genome. Science. 328: 710-722

Reich D et al. 2009. Reconstructing Indian Population History. Nature. 461:489-494

Recitation: Application of admixture statistics and the inference of introgression using ADMIXTOOLS 2 R package

Homework: Problem set 2 (*due Week 6)

WEEK 6: Recombination in the human genome I: Linkage disequilibrium

Lecture: Linkage disequilibrium concepts and inference

Reading: Nielsen and Slatkin Chapter 6

Recitation: Calculation of linkage disequilibrium in R

Homework: problem set 2 due

WEEK 7: Recombination in the Human Genome II: Genetic Map Reconstruction with Genomic Data

Lecture: Genetic map reconstruction from trios and other mapping resources

Reading: Kong et al. 2010. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 467:1099-1103.

Recitation: Genetic map reconstruction from pedigree data

Homework: none

WEEK 8: GWAS I: Mapping disease with case-control studies

Lecture: Experimental design of case control studies, statistical analysis of case control data

Reading: Pearson TA, Manolio TA. 2008. How to interpret a genomewide association study. JAMA. 299:1335-1344.

Recitation: Analysis of case-control data to map human disease

Homework: Problem Set 3 (*due Week 11)

WEEK 9: GWAS II: Mapping quantitative traits with population-based designs

Lecture: Approaches to controlling for population structure and kinship with mixed linear models

Reading: TBD

Recitation: GWAS mapping with the Gapit package in R

Homework: Problem set 3 (*due Week 11)

WEEK 10: GWAS III: Advanced methods and applications

Lecture: Mass-scale applications of GWAS (e.g., Biobank)

Reading: Bycroft C. et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature. 562:203-209.

Recitation: GWAS mapping with the Gapit package in R II

Homework: Problem set 3 (*due Week 11)

WEEK 11: The missing heritability problem in human genetics

Lecture: The genetic architecture of phenotypic traits and the causes of missing heritability

Reading: Génin, E. 2020. Missing heritability of complex diseases: case solved? Human Genetics. 139: 103-113.

Recitation: Estimation of SNP-based heritability with GCTA

Homework: Problem Set 3 due

WEEK 12: Adaptation in the Human Genome

Lecture: Inferring positive and balancing selection using population genomic methods

Reading: Nielsen and Slatkin Chapters 7 and 8

Recitation: TBD

Homework: Problem set 4 (*due Week 14)

WEEK 13: Population genomics of disease I

Lecture: Background selection and the dynamics of deleterious mutations in human populations

Reading: TBD

Recitation: TBD

Homework: Problem set 4 (*due Week 14)

WEEK 14: Population genomics of disease II

Lecture: Applications of population genomics in the study of cancer and disease

Reading: Gurdasani D. et al. 2019. Genomics of disease risk in globally diverse populations. Nature Reviews Genetics. 20:520-535.

Recitation: TBD

Homework: Problem set 4 due