

COURSE DESCRIPTION

Genome-sequencing is quickly becoming a commodity, and more than a million people have already analyzed their own genomes through direct-to-consumer companies. This course provides an introduction to current bioinformatics techniques for analyzing and interpreting human genomes. We will learn how to interpret a single genome in the context of an entire population, based on the often quoted concept: *interpreting one genome requires tens of thousands of genomes*. Topics covered include an introduction to human medical and population genetics, human ancestry, finding and interpreting disease-causing variants, genome-wide association studies, genetic risk prediction, analyzing next generation sequencing data, and how to scale current genomics techniques to analyze hundreds of thousands of genomes. We will also discuss the social impact of the personal genomics revolution.

Prerequisites

There are no official prerequisites for this course, but it is assumed students have some programming experience and are familiar with using the UNIX command line.

Course objectives

Though this course students will:

- Gain basic bioinformatics skills needed to analyze a personal genome using the UNIX command line.
- Gain the ability to critically read and interpret basic science and translational literature relevant to personal genomics.
- Demonstrate knowledge and understanding of the social impacts of the personal genomics revolution.
- Gain skills and experience necessary to carry out original research related to personal genomics.

Course structure

The course will consist of 5 modules: Intro to personal genomics, What can I do with my genome?, Complex traits, Next-generation sequencing, and mutation hunting. Each module will have a corresponding problem set due at the end of the module. In addition, each student will work in a group to complete a final project, described below.

Course sessions

Class meetings will consist of two sections separated by a short 5 minute break. The first section will usually consist of a lecture (45 minutes). The second session (30 minutes) will consist either of a journal club discussion or hands on experience using and developing bioinformatics methods. All details of the course will be posted at the course website:

https://gymreklab.github.io/teaching/personal_genomics/personal_genomics_2017.html

GENOTYPING FAQ

The goal of this course is to teach you how to analyze your genome. At this time, we are not able to provide genotyping or sequencing services to students, and instead homeworks will be performed on publicly available human genomes. You are welcome to analyze your own genome using a direct-to-consumer service such as

23andme or Ancestry.com. I recommend 23andme, which performs SNP genotyping arrays for \$200. You will also get access to a range of fun tools on their website that will complement some of the analyses we do in class. Homeworks for modules 1-3 are built to analyze the type of data available from 23andme.

Please keep in mind the following:

- It is **NOT** required that you analyze your own genome. Your grade is in no way influenced by whether or not you participate in 23andme.
- To ensure the previous point, **you do not need to tell me if you will be analyzing your own genome.**
- If you do 23andme, it is at your own risk and at your expense. We do not have funding to pay for the test.
- While you are free to analyze your genome, you must still complete the homework assignments with the provided genomes as well.
- 23andme can take up to six weeks, so sign up early if you want to use it during class.

RESOURCES

All resources including readings, problem sets, the final project, and how to use computational resources on XSEDE can be found at the course website (see link above).

GRADING

Your grade will be based on participation (10%), attendance (10%), problem sets (45% total), and a final project (35%). Attendance is mandatory and attendance will be taken.

There will be one problem set for each module that will mostly involve performing analyses on genomes using existing tools. PS1 is worth 5%. The other problem sets are worth 10% each. Late problem sets will lose one point for each day past the due date.

The final project will consist of a proposal (5%), paper (20%), and presentation (10%). Students will work in pairs to either design a new bioinformatics tool for analyzing human genomes, perform an analysis, or explore a theoretical problem. Projects and problem sets are described in detail on the course website.

Final grades will be out of 100 points. ≥ 70 points is passing. If you are taking the course for a letter grade, we will use the following scale: A+:97-100, A:93-96, A-:90-92, B+:87-89, B:83-86, B-:80-82, C+:77-79, C:73-76, C-:70-72, F:0-69.

POLICIES

Academic integrity

Collaboration on problem sets is encouraged. However each student must turn in a separate problem set and perform each exercise. The final project must consist of original work, and all sources must be properly cited.

Accommodations for Students with Disabilities

If you have a disability for which you are or may be requesting accommodations, please contact Office for Students with Disabilities. You must have documentation from the the Office before accommodations can be granted.

Disclaimer

While we have every intention of following this syllabus, any information here is subject to change.

SCHEDULE

Module	Date	Lecture	Homework
Intro to personal genomics	01-10 (T)	Introduction to your genome	PS1 out
What can I do with my genome?	01-12 (Th)	Basic population genetics	PS1 due, PS2 out
	01-17 (T)	Determining ancestry	
	01-19 (Th)	Phasing and imputation	
	01-24 (T)	Web 2.0 Genomics	
	01-26 (Th)	Social impact of personal genomics	PS2 due
Complex traits	01-31 (T)	Introduction to GWAS	PS3 out
	02-02 (Th)	Guest lecture TBD	
	02-07 (T)	Advanced GWAS topics	Project out
	02-09 (Th)	Scaling GWAS to millions of genomes	PS3 due
Next-gen sequencing (NGS)	02-14 (T)	Introduction to NGS	PS4 out
	02-16 (Th)	Short read alignment and variant calling	Proposal due
	02-21 (T)	Functional genomics - RNAseq, ChIPseq	
	02-23 (Th)	Storing, querying, visualization	
	02-28 (T)	Long read sequencing technologies	PS4 due
Mutation hunting	03-02 (Th)	Introduction to genetic mapping	PS5 out
	03-07 (T)	Filtering and prioritizing variants	
	03-09 (Th)	<i>De novo</i> mutations and constraint	PS5 due
	03-14 (T)	Interpreting non-coding variants	
Final Project	03-16 (Th)	Project presentations	Project due