

GCB/CIS/MSTR 535 (Fall 2023): Introduction to Bioinformatics

Course Staff

Course Directors

Please use e-mail or the course slack to contact the course instructors. We do not reliably check the Canvas messaging functionality.

Benjamin F. Voight, PhD

Department of Systems Pharmacology and Translational Therapeutics
Department of Genetics
10-126 Smilow Center for Translational Research (SCTR)
3400 Civic Center Blvd

Ophir Shalem, PhD

Department of Genetics
Center of Cellular and Molecular Therapeutics
3501 Civic Center Boulevard
5th floor CTRB, room 5050

Teaching assistants

Mitch Conery

Guest Lecturers

If any, TBD.

Keywords

- Ben Voight: human genetics, statistics, human evolution, cardiometabolic disease
- Ophir Shalem: functional genomics, protein quality control, neurodegenerative disease
- Mitch Conery: computational genomics, gene regulation, and bone disease

Location and time

The course will be administered on campus for Fall 2023

Towne Building (<https://goo.gl/maps/uoVt4vjryf8ecv2Q9>); Access details are here (<https://facilities.upenn.edu/sites/default/files/pennaccess/PA0570-Towne.pdf>).

Classroom 217

MWF 1:45-2:45p

Office hours

Office hours are time that we (the instructors and TAs) set aside for you to drop by to talk about the course and ask any questions that you may have.

We are hosting office hours in or virtually, which will correspond to a designated time where a TA or instructor will be available to chat with you at the given time. Please feel free to attend at your convenience.

Office hours with Professors are generally open door. However, students may schedule meetings during this time with the course directors, and scheduled students will receive precedence. To schedule a meeting, send an e-mail or slack message to the course directors. There are some days where there are preexisting conflicts with regularly scheduled office hours. We've created a table with details (<https://github.com/bvoight/GCB535/blob/master/OfficeHours.md>), for your convenience.

- BFV: Friday 3-4p, 10-126 SCTR or by zoom
- TAs: TBD

Course web site

We will use CoCalc (<https://cocalc.com/>) for in class exercises and homework. Completed assignments will be collected through this system. Canvas (<https://upenn.instructure.com/>) is used to return grades to students and to provide additional materials for the course (recorded lectures, lecture slides, etc).

Course description

This course provides broad overview of bioinformatics and computational biology as applied to biomedical research. A primary objective of the course is to enable students to integrate modern bioinformatics tools into their research activities.

Course material is aimed to address biological questions using computational approaches and the analysis of data.

A basic primer in programming and operating in a UNIX environment will be presented, and students will also be introduced to R and tools for reproducible research. This course emphasizes direct, hands-on experience with applications to current biological research analysis problems. The course is **NOT INTENDED** for computer science students who want to learn about biologically motivated algorithmic problems; GCB/CIS/BIOL5360 and GCB5370 is more appropriate.

Areas multiple hypothesis testing, data processing and visualization, analysis pipeline development, sequence alignment, study design for high-throughput sequencing, RNA gene expression, single-cell RNA-Seq, single gene and whole-genome analysis, gene regulation and ENCODE, ChIP-Seq analysis, analysis of CRISPR screening data, and machine learning. Relevant principles underlying methods used for analysis in these areas will be introduced and discussed at a level appropriate for biologists without a background in computer science.

The course will assume a solid knowledge of modern biology. An advanced undergraduate course such as BIOL4210 or a graduate course in biology such as BIOL5260 (Experimental Principles in Cell and Molecular Biology), BIOL5270 (Advanced Molecular Biology and Genetics), BIOL5280 (Advanced Molecular Genetics), BIOL5400 (Genetic Systems), or equivalent, is a prerequisite.

Note on Reproducibility and Responsible Research

Scientific rigor and reproducibility (SRR) is essential for progress, and especially so in Bioinformatics. This course aims to address SRR by teaching students programming and computational pipeline building skills, so that students can generate results that are "technically reproducible" - that the same code run on the same input file generates the same output file or result each time. Moreover, the class will teach students technical and intuitional skills on how to look at, and evaluate, data sets they are provided that might be established or new. This helps to generate SRR because 'bugs' or technical artefacts in data files hinders the production of reproducible results. Students who leave the course will have developed their intuition and 'healthy skepticism' new data sets they encounter, and a framework to evaluate, visualize, and correct issues in data as they are encountered.

Equipment prerequisite

All students are required to bring a laptop to class for in class activities. TAs will provide help with the material, but students should be computer-capable with their own laptop.

Schedule

The most up to date schedule for the course is available in the [course calendar](https://github.com/bvoight/GCB535/blob/master/Course_Calendar.md) (https://github.com/bvoight/GCB535/blob/master/Course_Calendar.md).

Grading

Composition of grades

Grades are composed of:

- Homework Assignments (3 in total) - 50%
- Exam One - 25%
- Exam Two - 25%

Late grading policy:

- Homework will be accepted up to one week late, but grades will be penalized (minus 20% off). If you would like to have an assignment graded under this late policy, you must e-mail one of the course directors before the start of the class period in question.

Plagiarism policy

Consistent with the University of Pennsylvania's honor code and policies on academic integrity, we maintain a zero-tolerance policy on plagiarism. For assignments containing text that the instructors determine is plagiarized, students will receive a grade of zero for the assignment and the case will be referred to the Office of Student Conduct for disciplinary measures. Students may not always be aware of what constitutes plagiarism in their work. If you are unsure, please contact one of the course directors. Please see links below regarding the office of academic integrity for Penn's policy on plagiarism and our discretion on grading.

<https://provost.upenn.edu/policies/faculty-handbook/students/iv-d> (<https://provost.upenn.edu/policies/faculty-handbook/students/iv-d>)

http://www.upenn.edu/academicintegrity/ai_violations.html (http://www.upenn.edu/academicintegrity/ai_violations.html)

Regarding ChatGPT and other Generative AI / Large Language Models (LLMs) in the course

The recent release of powerful generative AI tools has unsettled higher education. Tools such as Chat-GPT, GPT-4, Bard, Dall-E 2, Llama 2, Claude, Pi and Bing are developing rapidly. These chat tools can write and revise text, code and images in response to prompts.

There may be cases where these tools can help you develop your understanding of the material and modules that we provide you for the class. If you are using these tools to help you work through 'in-class' materials, or in the course of helping you in your understanding of specific topics, we would ask that you indicate that by including a written note at the

top of the notebook by creating a 'new cell' (and adding a note).

Our rationale for providing (and grading) homework problems and exams is to give a 'checkpoint' that you are, in fact, working through the materials and are understanding them sufficiently that you can independently perform data analysis tasks put to you. As such using these tool to answer homework problems and/or exam questions is not consistent with this goal and as such is not allowed and constitutes plagiarism. Use without permission in these cases will be considered a violation of Penn's Code of Academic Integrity.