GCB/CIS/MTR 535: Introduction to Bioinformatics
(Spring 2017)

Course Director:

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
215-746-8083
bvoight@upenn.edu

Casey S. Greene, PhD
Department of Systems Pharmacology and Translational Therapeutics
10-131 Smilow Center for Translational Research (SCTR)
3400 Civic Center Blvd
215-573-2991
csgreene@upenn.edu

TAs:

Katie Siewert	ksiewert@mail.med.upenn.edu
Kelsey Johnson	kelsj@mail.med.upenn.edu

Guest Proctors:

Greg Grant	ggrant543@gmail.com

Location/Time:

MWF	10-11 AM	10-146 SCTR

Office Hours:

Ben	Wed	11-12PM	10-126 SCTR
	By appointment		10-126 SCTR
Casey	Fri	11-12PM	10-131 SCTR
	By appointment		10-131 SCTR
TAs	Tue	4-5 PM	10-100 SCTR
	Thu	4-5 PM	10-120 SCTR
	By appointment		

Course web site: We will use Canvas (https://upenn.instructure.com/) to disseminate and receive materials for the course. We will use Canvas to provide lecture material. We will also be using SageMathCloud (https://cloud.sagemath.com/) for learning and homework. Completed assignments will be uploaded there.

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 Python, R, and tools for reproducible research. This course emphasizes direct, hands-on experience with applications to current biological research problems. The course is **NOT INTENDED** for computer science students who want to learn about biologically motivated algorithmic problems; GCB/CIS/BIOL536 and GCB537 is more appropriate.

Areas include DNA sequence alignment, genetic variation and analysis, motif discovery, study design for high-throughput sequencing, RNA and gene expression, single gene and whole-genome analysis, machine learning, and topics in systems biology. The 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 BIOL421 or a graduate course in biology such as BIOL526 (Experimental Principles in Cell and Molecular Biology), BIOL527 (Advanced Molecular Biology and Genetics), BIOL528 (Advanced Molecular Genetics), BIOL540 (Genetic Systems), or equivalent, is a prerequisite.

Equipment prerequisite: IMPORTANT
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, and should be willing/capable to download and install free software from the Internet.

Schedule
The most up to date schedule for the course is available in the course calendar:
https://github.com/greenelab/GCB535/blob/master/Course_Calendar.md

Grading: Grades will be determined by: (i) the course project (50% total, 25% written proposal, 25% oral presentation), (ii) homework assignments (30%), (iii) in-class labs assignments (10%), and (iv) class participation (10%). For the homework grade, we will drop scores from the 3 lowest scoring assignments when determining the final grade.

Late grading policy: Our late grading policy is driven by fairness. This policy allows us to return assignments quickly while providing flexibility where it is required.

Pre-lab assignments are collected at the start of class. We are unable to accept late turn-ins. This material is required in order to fully participate in the in class exercises.

Homework is due at the start of class immediately following the class in which it was assigned (e.g. homework assigned Monday is due at the start of class on Wednesday). Homework will be accepted up to one class period after its due date (e.g. assigned Monday would be accepted at the start of class on Friday), but grades will be penalized (minus 30% 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
http://www.upenn.edu/academicintegrity/ai_violations.html

Date	Pre-class Type	Class Activity	Homework Due Dates
11-Jan	-	Lecture: Intro/Databases	-
13-Jan	Workbook + Lecture	Seeking Biological Info Online	SageMathCloud
16-Jan		MLK - No Class	
18-Jan	Workbook	UNIX Directories and Files	Seeking Bio Online
20-Jan	Workbook	UNIX Web Resources and Workflows	UNIX - I
23-Jan	Workbook + Lecture	Motif Discovery	UNIX – II
25-Jan	Workbook + Lecture	Cis-Regulatory Modules	Motifs – I
27-Jan	Workbook + Lecture	Regulatory Analysis	Motifs – II
30-Jan	Workbook + Lecture	Sequence Alignment	Motifs – III
1-Feb	Workbook	ChIP-Seq Primer – I	Alignment
3-Feb	Workbook	ChIP-Seq Primer – II	ChIP-Seq – I
6-Feb	Workbook	Phylogenies/Multi-Sequence Alignment	ChIP-Seq – II
8-Feb	Workbook	ENCODE Primer – I	Phylogenies
10-Feb	Workbook	More on ENCODE – II	ENCODE – I
13-Feb	Workbook	Intro to R – I	ENCODE - II
15-Feb	Workbook + Lecture	Sequencing Tech, Intro to R – II	R – I
17-Feb	Workbook	Tools for Reproducible Research	R – II
20-Feb	Workbook	Analysis of Gene Expression (Arrays) – I	Reproducibility
22-Feb	Workbook	Analysis of Gene Expression (Arrays) – II	Gene Expression – I
24-Feb	Workbook	Functional Enrichment Analysis	Gene Expression – II
27-Feb	Workbook + Lecture	RNA-Seq Primer – I	Gene Expression – III
1-Mar	Workbook	RNA-Seq Primer – II	RNA-Seq – I
3-Mar	Workbook	RNA-Seq Primer – III	RNA-Seq – II; Prelim Proposal
-		SPRING BREAK - No Class	
13-Mar	-	Project Meeting	RNA-Seq – III
15-Mar	Workbook + Lecture	Bioinformatics in Pharmacology	-
17-Mar	Workbook	Programming in Python - I	Pharmacology
20-Mar	Workbook	Programming in Python – II	Python – I
22-Mar	Workbook	Programming in Python – III	Python – II
24-Mar	Workbook	Programming in Python – IV	Python – III
27-Mar	Workbook	Programming in Python - V	Python – IV
29-Mar	Workbook	Machine Learning – I	Python – V
31-Mar	Workbook	Machine Learning – II	ML – I
3-Apr	Workbook	Machine Learning – III	ML – II
5-Apr	Workbook	Machine Learning – IV	ML – III
7-Apr	Workbook	Machine Learning – V	ML – IV
10-Apr	Workbook + Lecture	Intro to PLINK	ML – V
12-Apr	Workbook + Lecture	Analysis of Genetic Variation	Variation – I
14-Apr		"OPEN STUDIO"	Variation – II
17-Apr	-	"Demo Day" (Profs, TAs)	-
19-Apr	-	Student Presentations	-
21-Apr	-	Student Presentations	-
24-Apr	-	Student Presentations	-
26-Apr	-	Student Presentations	Project Report Due