

Course Outline

STAT 540: Statistical Methods for High Dimensional Biology

Contents

Overview	1
Teaching Team	1
Course-level learning objectives	1
Selected topics	2
Course communication	2
Schedule	2
Prerequisites and Resources	2
Evaluation	3
Copyright	4

Overview

Winter Term 2 2020 (January 11, 2021 - April 14, 2021)

STAT 540 is a 3 credit course with a mandatory computing seminar

This course is cross-listed as STAT 540, BIOF 540, and GSAT 540

Teaching Team

Virtual office hours listed below are on a drop-in basis (no appointment necessary) and Zoom links will be posted in Canvas. For more info on the Teaching team, including brief bios, see the [People](#) page.

Instructor

Keegan Korthauer, Ph.D (She/Her/Hers) - keegan@stat.ubc.ca

Teaching Assistants

Jenkin Tsui (He/Him/His) - jenkin.tsui@aya.yale.edu

Marco Tello Palencia (He/Him/His) - Marco.TelloPalencia@bcchr.ca

Sina Jafarzadeh (He/Him/His) - jafarzadeh91@gmail.com

Guest Lecturers

We are also fortunate to have several guest lectures throughout the course. See the [Lecture](#) schedule and [People](#) page for more info.

Course-level learning objectives

- Perform exploratory data analysis and visualize genomics data
- Apply tailored statistical methods to answer questions using high dimensional biological data
- Make your work reproducible, reusable, and shareable
- Work with real data in a collaborative model

Selected topics

- Basics of molecular genetics/genomics and data collection assays (methods)
- Basic probability and math foundations
- Exploratory data analysis and data quality control
- Normalization, batch correction
- Basic statistical inference (“one gene at a time”) – linear models
- Large-scale inference (“genome-wide”) – multiple testing
- Analysis of microarray, RNASeq, and epigenetics data
- Principal Component Analysis and Clustering (unsupervised machine learning)
- Classification and cross validation (supervised machine learning)
- Gene set analysis and gene networks
- Genome-wide association analysis (GWAS)
- Guest lectures on special topics: Single-cell RNA-seq, Causal Inference in Genomics, & Polygenic Risk Scores and Phenome-Wide Association Studies

Course communication

In the web-oriented setting, it is vital that we have a dedicated plan for how we will communicate throughout the semester.

General questions: We encourage you to use the [course Github Discussion repository](#) (you will be granted access after we collect your Github user IDs) for posting questions, so that the message can be seen by the entire teaching team, and so that others in the class who might have the same question can learn from responses. You are also welcome to share your input on questions posted by others. Course announcements will also be posted in this repository, so please check it regularly (and/or subscribe to notifications).

Private matters: For private matters, please contact the Teaching team by email (listed above).

Group work: In your final project teams, we encourage you to make use of (1) the discussion feature in your GitHub Teams group, and (2) the issues feature in your group’s project repo. In addition, you are encouraged to meet regularly via a platform of your choice (e.g. Microsoft Teams). Please reach out if you have any questions or challenges in this space.

Schedule

Lectures (Sec 201)

- Time : Mon Wed 9:30 - 11am*
- Location: Online (Zoom) - Meeting links and recordings posted in Canvas
- *Note that some lectures will be asynchronous/prerecorded - see [Lectures](#) for lecture materials and schedule

Seminars (Sec S2B)

- Time : Wed 12pm - 1pm
- Location: Online (Zoom) - Meeting links posted in Canvas
- See [Seminars](#) for schedule and seminar materials
- We strongly recommend reading the seminar materials prior to attending each seminar.

Prerequisites and Resources

This interdisciplinary course requires a broad range of skills at the interface of statistics, molecular biology / genomics, and computing. While you may have some background in one or more of the following areas, you are not expected to be an expert in all. That said, to be successful in the course, you may need to spend a bit more time in the areas that you have less experience in. Although there are no official prerequisites for the course, here is a list of useful skills to bring into the course and/or learn along the way.

Statistics:

- You should have already taken a university level introductory statistics course.
- [This free online book “Modern Statistics for Modern Biology”](#) by Susan Holmes and Wolfgang Huber is a great reference for introduction or review of many of the statistical concepts that are relevant for this course.
- [This free online book “Data Analysis for the Life Sciences”](#) by Rafael Irizarry and Michael Love is another great resource for introduction or review of many of the statistical concepts relevant in this course, with an emphasis on implementation in R.

Biology:

- No requirements, but you are expected to learn things like, e.g. the difference between DNA and RNA, and the difference between a gene and a genome.
- See [Seminar-0c](#) for some basic introductory material.
- [This free online book “Concepts of Biology”](#) by Fowler, Roush & Wise is a great resource for biological concepts, in particular chapters 6 and 9
- [This free online book “Biology”](#) by Clark, Douglas & Choi goes more in-depth, see Chapters 14, 15, and 16 for material on genetics that is particularly relevant for this course.
- No matter your prior experience, when you come across a new biological concept during the course or in your research, you might need to spend a bit of time ‘learning as you go’ - this is expected and something I still do regularly in my day-to-day research!

R:

- No experience required but be prepared to do a lot of self-guided learning if you haven’t taken other courses on R or used it in your research.
- Start now by installing [R](#) and the HIGHLY RECOMMENDED “integrated development environment” (IDE) [RStudio](#) - both are free and open source.
- You should be able to run R on your own computer during each seminar session.
- If you are new to R, check out [this blog post on getting started with R](#).
- [This free online book “Introduction to Data Science”](#) by Rafael Irizarry is also a great resource for getting more in-depth with R, programming basics, and the tidyverse. In particular see Chapters 1-5:
 - [Chapter 1: Getting Started with R and R Studio](#)
 - [Chapter 2: R Basics](#)
 - [Chapter 3: Programming Basics](#)
 - [Chapter 4: The tidyverse](#)
 - [Chapter 5: Importing data](#)

Other computing tools:

- In this course we’ll be using the version control software [Git](#) and its web-based hosting and collaborative platform [GitHub](#).
- [The online resource “Happy Git and GitHub for the useR”](#) from Jenny Bryan is a great reference for these tools as we learn them.
- Another helpful git resource is [Hadley Wickham’s webinar “Collaboration and time travel- version control with git, github and RStudio”](#)
- We’ll learn about using [R markdown](#) to generate readable and reproducible reports with code and text, and you’ll be using that a lot in this course - see [Chapter 18 of the ‘Happy Git’ resource: “Test drive R markdown”](#).

Evaluation

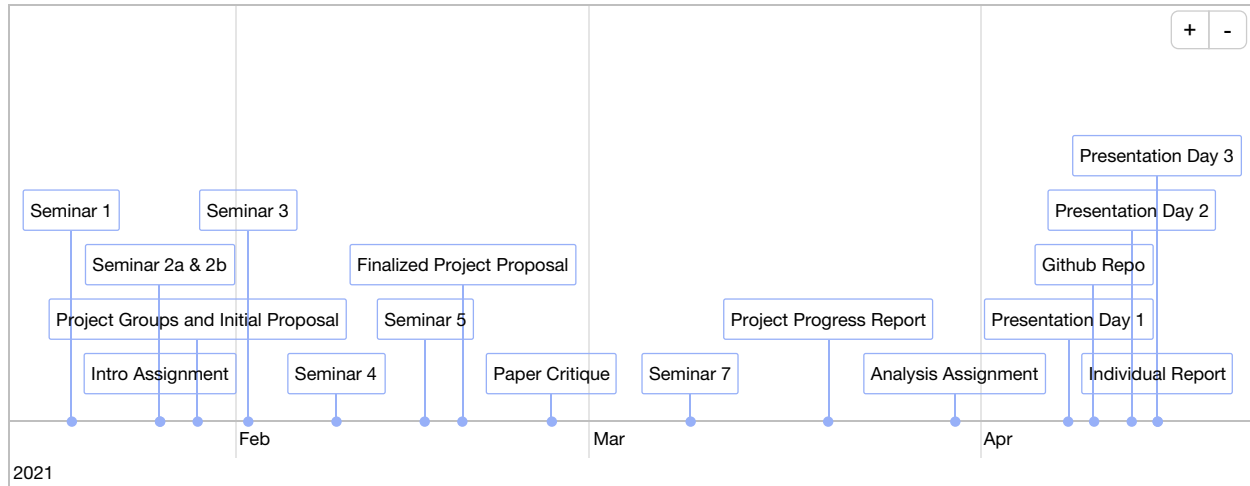
You will have three individual assignments, six seminar submissions (one divided into two parts), and one group project. Deadlines are all by 11:59 pm (Pacific time) on the due date, and late assignments are penalized 10% per day/partial day. If you anticipate having trouble meeting a deadline, or need to request an extension please reach out via email.

For more detail on each of these assignments, see the [course work page](#) (the header of each assignment on this page points to the relevant section of the course work page).

For help on your homework mechanics, please see this list of [assignment tips](#). For detailed instructions on how to turn in your assignments, see the [submission instructions on the course work page](#).

Timeline overview

Current date is indicated by the vertical red bar. Click the zoom buttons to zoom in on a date. Click and drag horizontally to slide date range and vertically to view all items. Reload the page to reset the view to include all due dates.



Copyright

All materials of the course are licensed under the [Attribution-NonCommercial 4.0 International](#) license.