BST281: Genomic Data Manipulation, Spring 2019

Monday 01: Course overview

Convey the conceptual understanding and technical tools to do high-quality quantitative biology research.

<http://huttenhower.sph.harvard.edu/bst281>

## Location

Lecture: MW 3:45-5:15 FXB G11

Lab: R 3:45-5:15 FXB G03 (optional, problem set focus)

## Instructors and teaching assistant

Curtis Huttenhower, [chuttenh@hsph.harvard.edu](mailto:chuttenh@hsph.harvard.edu) Office hours: W 10:00-11:00, SPH1 413

Eric Franzosa, [franzosa@hsph.harvard.edu](mailto:franzosa@hsph.harvard.edu) Office hours: W 10:00-11:00, SPH1 413

Mike MacArthur, [macarthur@g.harvard.edu](mailto:macarthur@g.harvard.edu) Office hours: F 9:30-10:30, FXB second floor atrium

## Course structure

6 biweekly problem during the first 3/4 of the course.

Assigned during class on Monday, due by end-of-day (midnight) the following week's Friday (~2 weeks).

Group midterm journal club, plus final project (including 2-4 page individual writeups).

## Collaboration policy

Assignments and code may be discussed but not copied; all problem sets must be submitted individually and will be checked for duplications.

Questions but not specific code should be posted publicly on the course discussion board; please email questions about specific code privately to TAs.

*Quantitative biology*

Quantitative biology: umbrella term for biological research that requires data nuance.

Computational biology and bioinformatics often used interchangeably.

Sometimes the former refers more to biological applications and the latter to algorithms / biophysics.

Also includes biostatistics and general quantitative thinking.

Computational experiments should be handled much like laboratory experiments.

Express hypotheses, design first, and be constructively skeptical.

Be constructively skeptical: if a result looks too good to be true, it probably is.

Modularize computational processes.

Look at the data going in and coming out of every step.

Include positive and negative control inputs with known and null outputs, respectively.

Document everything: electronic lab notebook, data, and code.

Can be assisted by literate programming tools, simple readme files, and comments.

Keep your workspace clean: use consistent file layouts and naming conventions.

# Textbooks

Bioinformatics overview: Pevsner, Chapter 1 p3-15

'Omics overview: Lesk, Chapter 1 p3-9, Chapter 2 p41-50

# Literature

[A systems approach to mapping DNA damage response pathways. Workman et al, Science 2006.](https://www.ncbi.nlm.nih.gov/pubmed/16709784)

[A quick guide to organizing computational biology projects. Noble, PLoS Comp. Bio. 2009.](https://www.ncbi.nlm.nih.gov/pubmed/19649301)

[Genomic signatures to guide the use of chemotherapeutics. Potti et al, Nature Medicine 2006.](https://www.ncbi.nlm.nih.gov/pubmed/17057710)

[Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. Baggerly and Coombes, Annals of Applied Statistics 2010.](https://arxiv.org/abs/1010.1092)