Biophysics 206: Concepts in Cancer Genome Analysis
Fall 2021

Professor: Gad Getz (gadgetz@broadinstitute.org)

TFs:

- Nicholas Haradhvala (njharlen@g.harvard.edu)
- Kristy Schlueter-Kuck (kschluet@broadinstitute.org)

Time and Location: Classes will be held Tuesday and Thursdays 3:00-4:15pm at the Broad Institute board room. This can be reached by arriving at 415 Main Street, Cambridge, and proceeding up the steps in the lobby to the mezzanine level.

COVID Policies: All university COVID guidelines should be following, including weekly testing and wearing of masks in class. A zoom link is available on the Canvas page to attend remotely should a student experience any potential COVID symptoms, or for other COVID-related reasons feel uncomfortable to come to class in person.

Prerequisites: Proficient programming skills (python, R, Matlab, Julia), basic probability (STAT110 or equivalent) and biology fundamentals.

Example stats topics: Hypothesis testing and multiple hypothesis correction, Bayesian graphical hierarchical models, Conjugate priors, Expectation maximization, Mixture models, Regression, Principle Component Analysis, Clustering, Non-negative matrix factorization

Description: Cancer arises through a complex evolutionary process in which a series of genomic and epigenomic changes allow a cell lineage to gain fitness advantages such as increased growth rates, evasion of the immune system, and resistance to therapeutics. In this course, we will cover how analysis of next generation sequencing and other high throughput datasets can shed light on the different aspects of this process. Examples of topics include discovery of mutations that drive tumor development, mathematical modeling of clonal development, and use of single-cell sequencing techniques to survey the tumor immune microenvironment. We will focus on developing practical skills for applying statistical and computational methodologies to study these problems.

Grading: Homework (30%), Participation (30%), Final Projects (40%).

Assignments: The first half of the course will have problem sets that give a chance to get hands-on familiarity with the types of data surveyed in the lecture topics. Students are free to discuss assignments together, but all code and solutions submitted should be individually written. The course will culminate in a final project, which can be carried out in teams of two to three (see final project handout for additional details).

Coding may be done in the student's language of choice, however all code must be well documented and able to run with a simple commandline call or as a notebook. If students do not wish to perform compute on their laptops, bash, python and R environments are available on Harvard servers through FAS OnDemand, linked on the course Canvas page.

Week	Tuesday	Speaker	Thursday	Speaker
8/31	No class		Introduction	Gad Getz
9/7	Driver discovery	Gad Getz	Statistical modeling	Gad Getz
9/14	Driver statistics	Gad Getz	Next-gen sequencing data	Heng Li
9/21	Mutation calling	Gad Getz	Copy number alterations	Rameen Beroukhim
9/28	Purity and ploidy	Gad Getz	Tumor evolution	Gad Getz
10/5	DNA Damage and Repair	Nick Haradhvala	Mutational signatures	Nick Haradhvala
10/12	RNA-seq 1	François Aguet	RNA-seq 2	François Aguet
10/19	Dimensionality reduction	Gad Getz	scRNA technologies	Alex Shalek
10/26	Figure Design	Esther Rheinbay	Epigenetics	Esther Rheinbay
11/2	Tumor Immunology	Nir Hacohen	Perturbation screens	David Root
11/9	ML in Drug Discovery	Wengong Jin	Modeling cancer evolution	Franziska Michor
11/16	Proteomics	Steve Carr/D.R. Mani	Survival Analysis	Sam Freeman
11/23	TBD		Thanksgiving	No class
11/30	Final presentations		Final presentations	