

**BST 283 Cancer Genome Analysis**  
**Harvard T. H. Chan School of Public Health**  
**Spring 2017**

**Instructor:** Scott L. Carter  
**TA's:** TBN  
**Lectures:** M, W 1:30-3:20 PM

**Contact information:**

**Office:** Dana-Farber Center for Life Sciences Building 11026  
**Phone:** 617. 632.3038

**E-mail:** [scarter@jimmy.harvard.edu](mailto:scarter@jimmy.harvard.edu)

**Learning Objectives**

Achieve a basic working knowledge for modern statistical computing techniques used to characterize and interpret cancer genome sequencing datasets. By the end of this course students will be able to apply state-of-the art analysis to cancer genome datasets and to critically evaluate papers employing cancer genome data.

**Overview**

Classes will be organized into 1 2-hour lecture / week. Homeworks will involve application of methods presented in lectures to real cancer genome datasets.

This course will begin with a basic introduction to DNA, genes, and genomes for students with no biology background. We will then introduce cancer as an evolutionary process and review landmarks in the history of cancer genetics.

We will then discuss the basics of sequencing technology and modern NGS. We will cover the main steps involved in turning billions of short sequencing reads into a representation of the somatic genetic alterations characterizing an individual patient's cancer. We will build on this foundation to study topics related to identifying mutations under positive selection from multiple tumors sampled in a population.

**Outcome measures**

***Class Participation***

Active learning through class participation and discussion are an important component of the course. Students are expected to attend and participate in all classes.

## ***Assignments***

There will be five (5) homework assignments. These will require you to apply methods discussed in class to real cancer genome datasets. In some cases, key statistical aspects of a method must be filled in by the students. Each assignment will be graded on a scale of up to 10 points each. The assignment is due before the start of class on the date due. Projects, including code, graphics, and a written summary, will be submitted electronically.

## ***Project***

You will complete a final project for the course that will explore an area of cancer genome analysis in detail. Students will propose projects to the instructor and incorporate feedback into their project research plan, which must be approved by the instructor. Examples of student projects include: detailed evaluation of competing methods or approaches on a defined dataset, or extending an existing algorithm or method. The project includes an in-class presentation. In addition, it is expected that after the presentation, there will be in-depth in-class discussion on key aspects of the project.

The class project is worth 50 points. The grade is based on the group presentation, the post-presentation discussion (i.e., your ability to ask and respond to relevant questions).

## **Grading Criteria**

Grades are based on a total of 100 points. There are a total of the five (5) homework assignments (up to 50 points total) plus the grade on the class project (up to 50 points).

## **Schedule**

Weekly Th. TBD

## **Prerequisites**

To accommodate both statistics students with little biology background as well as biology students with little data analysis background, we will allow biology and statistics students to work as pairs to complete the homeworks and final project. These assignments will require basic programming and data manipulation ability in a statistical scripting language such as R (<http://projects.iq.harvard.edu/rtc/event/introduction-r>). In addition, basic knowledge of probability and statistical inference are encouraged.

## **Tools**

R, Unix

## **Recommended readings**

Information Theory, Inference and Learning Algorithms

By David J. C. MacKay

## **Course outline**

Date	Topic	Reading
23-Jan	Introduction	Weinberg textbook/ Garraway, Levi A., and Eric S. Lander. "Lessons from the cancer genome." Cell 153.1 (2013): 17-37.
25-Jan	DNA, genes, and genomes	Weinberg textbook/ Garraway, Levi A., and Eric S. Lander. "Lessons from the cancer genome." Cell 153.1 (2013): 17-37.
30-Jan	Cancer as an evolutionary process	Nowell 1976
1-Feb	Cancer as an evolutionary process	Nowell 1976
6-Feb	Landmarks in cancer genetics	Bishop/Varmus/Weinberg
8-Feb	DNA sequencing	Illumina / Gnirke Nat. Biotech 2009
13-Feb	Alignment	Li and Durbin 2009 Bioinformatics 25 (14), 1754-1760
14-Feb	Point-mutation calling	Cibulskis 2013 Nat. Biotech / Strelka
22-Feb	Point-mutation calling	Cibulskis 2013 Nat. Biotech / Strelka
27-Feb	Copy-number alteration calling	Chaing Nat. Methods 2009 / Olshen 2007
1-Mar	Copy-number alteration calling	Chaing Nat. Methods 2009 / Olshen 2007
6-Mar	Structural alterations	Campbell / Berger
8-Mar	Data quality control	Stewart
20-Mar	Absolute copy number	Carter Nat. Biotech 2012
22-Mar	Tumor clonal substructure	Nik-Zainal 2012
27-Mar	Tumor clonal substructure	Nik-Zainal 2012
29-Mar	Multi-sample analysis	Gerlinger NEJM 2012 / Landau Cell 2013
3-Apr	Phylogenetic inference	Stachler Nat. Genetics 2015
5-Apr	DNA damage mechanisms	Lawrence 2013
10-Apr	Mutational processes	Alexandrov 2013
12-Apr	Recurrent somatic copy number alterations	Beroukhim/Getz PNAS 2008, Mermel Genome Biol. 2010
17-Apr	Recurrent somatic point mutations	Lawrence Nature 2013 / Lawrence Nature 2014
19-Apr	RNA sequencing	Wang 2008 Cell
24-Apr	DNA methylation	Bernstein 2007 Cell
26-Apr	Chip-seq	Bernstein 2007 Cell
1-May	Clinical cancer genomics	Van Allen Nat. Medicine 2014
5-May	Final project presentations	
8-May	Final project presentations	
10-May	Final project presentations	

