

Cancer Genomic Workshop

Fall 2023
Course Syllabus

Course Overview:

As we step into the era of precision medicine, the need for extensive genomic data analysis is more pressing than ever. With the advent of high-throughput technologies, a plethora of genomic, transcriptomic, and epigenomic datasets from cancer patients have been sequenced, providing a wealth of data for systematic bioinformatics analysis. This three-day workshop, comprising three-hour daily sessions, is tailored to assist cancer bioinformaticians, researchers, and graduate students who are interested in incorporating expansive cancer genomic datasets from an open-source platform - The Cancer Genome Atlas (TCGA) - into their research.

Enrolled participants will gain insights into the fundamental principles of TCGA, including the different data categories and structures, methods for downloading data, and basic pipelines for data analysis within the TCGA platform. The workshop will also provide a versatile R script scaffold for mining TCGA datasets, empowering attendees to analyze their own unique cohorts of cancer genomic datasets.

Prerequisites:

Participants are expected to have a fundamental understanding of genomics. It is necessary to bring a laptop equipped with R (version >4.1) and R-Studio. Experience with R programming will be advantageous for workshop attendees.

Instructor Information:

Xianglong Tan

Postdoc Researcher, Biological Chemistry Department, David Geffen
School of Medicine, UCLA

xianglongtan@mednet.ucla.edu. Cell: 310-210-7456

Office Hours:

I am open to meeting with students outside of class hours, by appointment. The most efficient methods to contact me are via email or the cell phone number provided above. In the event of an urgent matter, you can usually find me in the lab during typical workdays, located in Room 357 of the Biomedical Sciences Research Building (BSRB).

Course Materials:

There is no mandatory textbook for this course. As we will be collectively coding in R during the sessions, please ensure that your laptop has both R and R-Studio pre-installed prior to the commencement of the course.

Course Outline:

Day I. Introduction and Guide to Data Downloading

- A. Background introduction of TCGA program
- B. Exploring the divergent datasets from GDC Portal
- C. Build a cohort and analyze with GDC Portal
- D. Understanding TCGA's barcode system
- E. Download datasets from that cohort with GDC Portal/GDC-Client

Day II. R Package TCGAbiolinks

- A. Introducing TCGAbiolinks package
- B. Three main function of TCGAbiolinks package – data query, download and prepare
- C. Practice with TCGA-READ data via TCGAbiolinks.

Day III. Advancing to Downstream Data Analysis via Pseudo-Single Cell Analysis

- A. Conducting differential expression analysis using DESeq2
- B. Dimensionality reduction and clustering using Uniform Manifold Approximation and Projection (UMAP)
- C. Survival analysis by Kaplan-Meier Curve

Exam (for students who enroll for academic credit only):

On Day III, students will receive an exam and homework. These exams will consist of 10 quick questions covering the material addressed in 3 days' unit. Exams will be distributed at the conclusion Day III.

Specific Homework assignment (for students who enroll for academic credit only):

At the conclusion of Day III, students will receive a specific homework assignment. Students will be expected to use the provided R script scaffold, given during the course, to analyze a provided cohort dataset and construct a UMAP as their final result.

Grading (for students who enroll for academic credit only):

Considering that this workshop accommodates both online and in-person attendees, grades will be determined based on the end-of-unit exams and the Specific Aims assignment. The weight distribution is as follows:

Participation	25 %
Day III Exam	25%
Specific Homework Assignment	25%