Quantitative and computational biology

GSK Graduate School, Spring 2018
Location:
Zuckerman Research Center (**ZRC-138**)
Meetings:
Class will meet from **1:00-3:00 pm** on the following days: February 5, 12, 26;
March 5, 12, 26; April 9, 16, 23, 30

Summary:
Students in this course will learn to apply quantitative exploratory data analysis techniques to different forms of experimental data. The course will begin with an introduction for students to computing via the UNIX shell, and to computing in the R programming language. The remaining lessons will be a blend of practical skills and theoretical concepts. Students will become comfortable performing exploratory data analysis, and will understand how concepts from statistics underlie the tools they use. Overall the goal of this course is to serve as a practical primer for various bioinformatic analyses, and should provide students with the foundation for future self-guided learning and skill acquisition in this discipline. These skills will enable them both to collaborate effectively with computational biologists, as well as begin to carry out their own computational experiments.

Specific topics include: practical aspects of data formatting and management; visualization of data; an introduction to probability, elementary statistics and hypothesis testing; experimental design and tools for differential expression analysis in RNA-seq and ChIP-seq; common data structures for working with biological sequences; enrichment testing for ranked gene sets; common bioinformatics tools and data quality assessment tools; introduction to structural biology and tools for visualization.

Towards the end of the course, students will be assigned a guided problem set to utilize the concepts and skills described in the course.

Course Instructors:
- Levi Naden (levi.naden@choderalab.org)
- Yuri Pritykin (*pritykin@cbio.mskcc.org*)
- Bobby Bowman (*bowmanr@mskcc.org*)

Assessment:
This class will not be graded. However, periodic student surveys in class and online are planned to assess if the presented material is properly learnt. All students will also be asked to complete a survey at the end of the course, to obtain feedback on the course and assess if this experimental format is successful and what changes need to be done when the course is offered in future years.

Class Sessions: All class sessions will meet from 1:00-3:00 pm
Lab 1 February 5, 2018   (Levi Naden)
       Introduction and getting started on the terminal
Lab 2 February 12, 2018 (Levi Naden)
       Basic mechanics of R
Lab 3 February 26, 2018 (Levi Naden)
       Basic mechanics of R: loading data and reading in files and plotting
Lab 4 March 5, 2018   (Bobby Bowman)
       Plotting data in R: TCGA data, boxplots, scatterplots, etc.
Lab 5 March 12, 2018   (Bobby Bowman)
       RNA-sequencing analysis
Lab 6 March 26, 2018   (Yuri Pritykin)
       Genomic data structures in R and Bioconductor: GenomicRanges,
       GenomicFeatures, Biostrings, BSgenome, AnnotationDbi.
Lab 7 April 9, 2018   (Bobby Bowman / Levi Naden)
       Bioinformatic tools including IGV, cBIO portal, file formats
Lab 8 April 16, 2018   (Yuri Pritykin)
       Practical ChIP-seq
Lab 9 April 23, 2018   (Bobby Bowman)
       Pathway analysis, GO enrichment, gene set enrichment analysis.
Lab 10 April 30, 2018   (Yuri Pritykin)
       Exploratory data analysis by PCA, clustering.

Books and Materials:
       Students will need a laptop computer on which they can install software (R
       and R Studio), and bring to class. All software used in the course is free,
       and run on recent versions of Linux, Mac OS X, and Microsoft Windows.
       R: *http://www.r-project.org*
       Rstudio: *http://www.rstudio.com*
       Gitshell (Windows only): *https://git-for-windows.github.io/*
       The course does not require the use of a specific textbook, but instead
       makes use of the following online resources:

Software Capentry
  - *Introduction to the Shell*
  - *R for Reproducible Scientific Analysis*

qBio-I course materials (courtesy of Jason Banfelder)
       http://physiology.med.cornell.edu/people/banfelder/qbio/schedule_2017/
Biomedical Data Science (courtesy of Raphael Irizarry and Michael Love)
  - *http://genomicsclass.github.io/book/*