Critical and quantitative analyses of next generation sequencing data

Instructor: Donald Vander Griend, Dept. of Surgery and Ben May; prostate@uchicago.edu

Teaching Assistant: Hannah Brechak, Committee on Cancer Biology; hbrechka@uchicago.edu

BSD Bootcamp on Quantitative Biology @ MBL

Goal:

Next generation sequencing is becoming mainstream and learning how to analyze and critically evaluate sequencing data will be crucial for nearly every field of study. Also, it's a quick way to obtain data for your project with no bench time! In this 3hr workshop we will cover the fundamentals of analyzing next generation sequencing (NGS) data, from raw Fastq sequencing files to quantative data analyses. The ***overall objective*** of this workshop is to provide a basic understanding of NGS data analysis, experimental considerations for both sequencing and data analyses, and how to access public data for use in their research. First, we will use the Galaxy platform and some simple datasets to cover key concepts which include experimental design, quality control analyses, read trimming, utilization and considerations of reference genomes, alignment processes, file types (i.e. BAM and GTF), FPKM and RPKM values, and false discovery rates. Second, we'll utilize the UCSC visualizing browser to introduce students to visualizing their sequencing data, and overlaying their data with publically-available resources such as ENCODE. Third and finally, we will cover major sequencing initiatives (TCGA, Broad, ENCODE, 1000 genomes) and how to access and analyze publically-available datasets for use in their own research.

Audience:

This workshop is intended for all biologists interested in understanding, and more so for using, next-generation sequencing approaches for their research. Given how commonplace sequencing has become, and understanding of the fundamentals will enhance any research project and provide a foundation for additional learning if desired.

Installation:

No software is necessary for this course, as we will be using a cloud computing platform. Please go to https://usegalaxy.org/ and set up an account.

Workshop Resources:

1. My powerpoint slides
2. Two simple datasets to play with
3. Understanding of how to access many more dataset

Readings and Additional Resources:

These readings are a good starting point after the workshop is completed to begin your NGS analyses journey.

1. Rodriguez-Ezpeleta et al. (Editors). Bioinformatics for High Throughput Sequencing. Springer 2012. Download for free from here: http://link.springer.com/book/10.1007%2F978-1-4614-0782-9
2. Garber et al. Computational methods for transcriptome annotation and quantification using RNA-seq. Nature Methods 2011, PMID 21623353.
3. Trapnell et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology, PMID 23222703.
4. Sims et al. Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews Genetics, PMID 24434847.