Pre-Lab

Audience

This lab is appropriate for undergraduate biology students (200 level and up) and advanced high school students (completed AP Biology and advanced electives).

Important Terms

Here are some important terms you may already be familiar with as well as some resources to review before watching the presentation. If you are unfamiliar with any of these terms, please look them up before watching the presentation.

- Gene

- Expression (as in gene expression)

- Transcript (as in mRNA transcript)

- Splicing

- Read (as in high-throughput sequencing read)

- K-mer

- Graph (see: https://en.wikipedia.org/wiki/Graph_theory)

Bioinformatics Lab Computer Setup

To follow along with this laboratory you will need:

1. Access to an up-to-date web browser such as Chrome or Firefox.

2. (Required) A *DNA Subway* account through *CyVerse* to use *DNA Subway* Green Line. To get a free account, go to https://user.cyverse.org/

Resources

Scientific papers (open access)

> Why are de Bruijn graphs useful for genome assembly?
> https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5531759/
>
> Near-optimal probabilistic RNA-seq quantification (subscription required)
> https://www.nature.com/articles/nbt.3519

Ensembl database (human)

> https://useast.ensembl.org/Homo_sapiens/Info/Index

DNA Subway

> Green Line Walkthrough: https://learning.cyverse.org/projects/dnasubway_guide/en/latest/step7.html

Laboratory Questions

These questions can be completed by watching the video.

1. What is a sequencing read?

2. What is a transcriptome and where could you obtain a transcriptome for an organism?

3. What is the purpose of aligning reads to a reference genome or transcriptome?

4. What are some potential challenges with aligning reads?

5. What are the two major steps performed by the Kallisto software?