Lab Intro and Computer Setup

C. Ryan Campbell

Duke University

c.ryan.campbell@duke.edu

31 Aug 2017

Overview

Today's Goals

- Get familiarized with course software
- Install software
- Produce a graph and save the code

Computer Requirements

- Bring your computer to class
- Run some software locally
  - How much empty storage do your laptops have?
- Connect to cluster

Cluster Computing

- Duke Computing Cluster (DCC)
- SLURM workload manager
- Many cores, a lot of RAM, never shuts down

Local vs. Cluster

- Which tasks are good for each of the following?
  - RNAseq analysis

Local vs. Cluster

- Which tasks are good for each of the following?
  - RNAseq analysis
  - Statistical analysis

Local vs. Cluster

- Which tasks are good for each of the following?
  - RNAseq analysis
  - Statistical analysis
  - Sequence data 'cleaning'

Local vs. Cluster

- Which tasks are good for each of the following?
  - RNAseq analysis
  - Statistical analysis
  - Sequence data 'cleaning'
  - Producing plots

Local vs. Cluster

- Which tasks are good for each of the following?
  - RNAseq analysis (DCC)
  - Statistical analysis (DCC/Local)
  - Sequence data 'cleaning' (DCC)
  - Producing plots (Local)

Research Tools

- github
  - Website for version control as well as script and code storage

Research Tools

- github
  - Website for version control as well as script and code storage
- R
  - Statistical software, data visualization and analysis

Research Tools

- github
  - Website for version control as well as script and code storage
- R
  - Statistical software, data visualization and analysis
- kallisto/tophat/cufflinks
  - Various RNAseq analysis packages

Programming Languages

- R
  - free statistical software
- bin/bash
  - unix language, automate many tasks, interact with cluster computers

Software

- R - R-Studio

Software

- R - R-Studio
- github - SourceTree

Software

- R - R-Studio
- github - SourceTree
- data analysis software -
    - fastqc
    - trimmomatic
    - kallisto/tophat/cufflinks
        - we'll worry about these later

github

- Software version control
- Shared and collaborative script editing
- Free!
- SourceTree

github

- Create a github account
- github.com
- Create a repository for this course
- github Repo Tutorial
- Make a repo for this course instead of "hello-world"

SourceTree

- Download and install SourceTree
- SourceTree Webpage
- If you already have your own software, feel free to continue to use that
- Clone your repository onto your own computer

SourceTree - Log in

- Open SourceTree
- Under Preferences "Add Account" and log into your github

SourceTree - Clone a repo

- Click "+New Repository"
- Select "Clone from URL"
- On your github repo page click the green button labeled "Clone or download"
- Copy the link and clone

SourceTree - Test it out

- add a folder "LabIntro" maybe
- make a text file
  - WARNING - github handles .txt files well, .doc files not well, so use "notepad++" or "Sublime" or "TextWrangler" to edit the files
- Test SourceTree by committing the changes to the master branch, etc

R

- Statistical analysis and visualization
- Shared and collaborative script editing
- Free!

R and R-Studio

- Download and install R
- R download
- Download and install R-Studio
  - Desktop and Open Source Version
- R-Studio download

R-Studio Test

- Download my .Rmd script and datafile from github
  - hypoxia analysis Rmd
  - hsap hypoxia gene exp FORCLASS diff
- Save it to your repo
- Run it in R-Studio to produce the graph

R-Studio Test
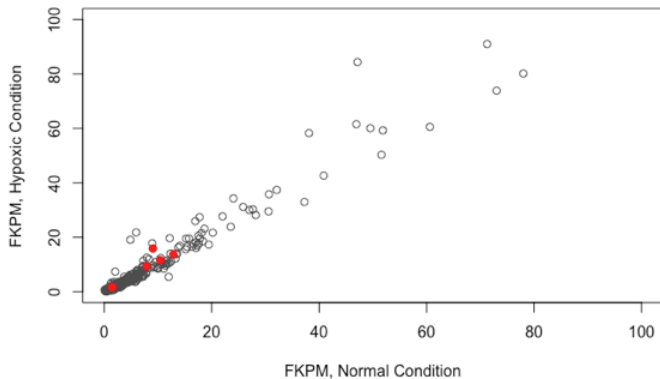
R-Studio Assignment

- Pick a different Gene Family
- Modify my code and plot the data
- Turn in -
  - Your Plot
  - A short paragraph about which gene family you picked an the results you saw

RNAseq Diff Files

- Output from a comparison across conditions
- Columns -
  - gene family
  - sample 1 and 2 - experimental condition of the sample
  - status - were there non-zero values on both samples
  - value 1 and 2 - relative amount of expression measured
  - log fold change - difference between conditions, roughly
    - negative numbers = decrease in expression from 1 to 2
  - test stat - test statistic
  - p-value
  - q-value
  - significant - less than .05 p-val

Hypoxia Gene Families

- TOMM - Translocase of outer mitochondrial membrane
- TIMM - Translocase of inner mitochondrial membrane
- ZNF - Zinc Finger
- RAS - involved in transmitting signals within cells
- OR - Olfactory Receptors
- PK - Protein Kinases
- HK - Hexokinase
- NFK - Nuclear Factor Kappa

The End