Introduction to RNA Seq Course

We will meet in PAR 206. You will need to bring your personal laptops. Please make sure you have an ssh client installed on your laptop. All macs come with terminal, so no installation is required. For windows laptops, install putty and winSCP. You will also need a TACC account with multi factor authentication set up.

Course Overview

This is a course designed to give you an overview of RNA-Sequencing in a hands-on manner. It will comprise of lectures and guided tutorials. For the tutorials, we will be used a canned dataset, but a large part of the last class day will be devoted to practicing the techniques learned on your own dataset. This course has the following objectives:

1. To teach you about the different options that are available to you when setting up a RNA-Seq study.
2. To teach you about the different options that are available to you when analyzing a RNA-Seq dataset.
3. To familiarize you with how the Texas Advanced Computing Center (TACC) can be used to simplify and speed up your data analysis.
4. To familiarize you with some of the typically used RNA-Seq analyses methods.
5. To provide a vocabulary to understand NGS and RNA-Seq terminology and to provide give you a starting point of where to begin you own data analysis, and enough experience that you can begin that analysis on your own.

Your Instructors

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Expertise</th>
<th>How to contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dhivya Arasappan</td>
<td>CBRS</td>
<td>Unix, TACC, RNA-Seq, hybrid genome assembly, pacbio</td>
<td><a href="mailto:dasappan@austin.utexas.edu">dasappan@austin.utexas.edu</a> or come to FNT 1.206D</td>
</tr>
<tr>
<td>Ben Jack</td>
<td>Wilke Lab</td>
<td>Python, R, RNA-seq, pseudoalignment tools</td>
<td><a href="mailto:b.jack@utexas.edu">b.jack@utexas.edu</a> or MBB 3.232</td>
</tr>
</tbody>
</table>

Day 1: Introduction to RNA-Seq

- **Logistics-Day 1**
- **Part 1: Lecture: Introduction to RNA-Seq**
- **Part 2: Unix and TACC Refresher**
  - The Quickest Unix Refresher ever
  - TACC Overview
    - Using The Lonestar5 cluster
    - Lonestar5 directory structure
    - Submitting jobs to Lonestar5
    - Lonestar5 TACC/Unix Exercise
- **Part 3: Getting NGS data ready**
  - The Big Picture
  - Evaluating your raw data
- **Day 1 take aways**
  - Overview of TACC Submission

Day 2: Read Mapping and Pseudomapping

- **Logistics-Day 2**
- **Read Mapping**
  - Lecture: Introduction to reading mapping
Day 2 take aways

Day 3: Testing for Differential Expression, Splice Variant Analysis and Novel Transcript assembly and Visualization

- Logistics-Day 3
- The Big Picture
- Part 1: Differential expression of annotated genes/transcripts
  - Lecture: Gene counting and differential expression
  - Gene counting
  - Testing for differential expression
  - Lecture: Visualization
  - Plotting DESEQ2 Results
  - Visualization using Integrated Genome Viewer (IGV)
- Part 2: The new tuxedo suite for finding novel transcripts
  - Running the new tuxedo suite and Results
- Day 3 take aways

Day 4: Downstream Analysis of Differentially Expressed Genes And Practice

- Logistics-Day 4
- Resources
- The Big Picture

- Part 1: GO Enrichment and Pathway Analysis
  - (optional) Advanced GO/KOG Analysis
- Part 2: Clustering using WGCNA
- Lecture: Example project downstream analysis
- Part 3: Practice time/Bring your own RNA-Seq Data

MANY THANKS

Many thanks to Dr. Scott Hunicke-Smith and Dr. Jeff Barrick who taught the first Introduction to NGS course at UT. Many slides and wiki pages have been borrowed and adapted from their initial course.

Also, thanks to Anna Battenhouse and Benni Goetz who have also provided some material for this course.