Bio Data Science

19 Feb 2016
Project Proposal
Meredith Bache-Wiig & Lauren Saunders

1) The data will be coming from an RNASeq experiment that we performed this year. It is low input RNASeq data from melanophores isolated from juvenile thyroid ablated and thyroid intact zebrafish.
2) What is the effect of thyroid hormone on zebrafish melanophore development at a transcriptional level? We know there are differences in cell number and phenotype, but we want to know how these changes are brought about transcriptionally. Subquestions include: What are relevant differentially expressed genes? Are there ways that we can improve the experimental design and data collection to gain more information from future cell type specific studies?
3) We will map the reads and produce SAM files to begin the MATLAB analysis, there will likely be some data wrangling needed to get the data into a readable format. WE will look at correlations of fpkm values between samples we know to be related, to assess sample homogeny. We will also be mapping reads to chromosomes to visualize distribution of reads, this will help us develop benchmarking tools to assess library quality and determine optimization methods for sample collection. We will determine which genes are differentially expressed and we plan to do some higher level pathway analysis using this published Matlab method (PMID: 25367050).
4) Correlations of fpkms will be analyzed with PCA and likely scatter plotted. We will use volcano plots to visualize differentially expressed genes, and perhaps isolate several genes to visualize using a heat map. We will be able to view our reads mapped onto chromosomes to assess their distribution. We will also generate pathway maps using our new Matlab tool.
5) Lauren and Meredith work in the same lab, so can communicate in person readily. We will also utilize GitHub for more remote collaboration.