**Assignment #2: RNA-seq for Gene Expression Analyses P/BIO 381**

Your assignment for next week is to contrast differential expression between healthy (H) and sick (S) separately within each location, and then compare those results to running with location in the model as a control variable. Look at the degree that these two approaches converge, and then overall what we see in terms of number of significant differentially expressed genes going up vs. down.

Below are general guidelines that describe the elements of the assignment you should include in your short write-up. Please use 2 pages maximum to demonstrate your understanding of the conceptual background and technical details for using RNA sequencing for gene expression analyses. You should include relevant tables and figures (within the two-page limit) with legends. You may discuss the assignment with classmates, but the assignment should be prepared individually. Please provide your code for this assignment in your github lab notebook and note this in your assignment. Due Wednesday, March 8.

Use DESeq2\_SSW\_round2.1 RKT RMD file in Eco Geno folder for tutorial

Compare healthy vs sick individuals in entire data set:

1. H vs S within intertidal

2. H vs S within subtidal

3. between intertidal and subtidal (all H vs all S, with ‘location’ as a control variable)

* Clear statement of objective (1 sentence).

We will analyze RNA-seq count data for *Pisaster* ochraceus, using the R package DESeq2, to investigate differential gene expression patterns between healthy individuals and sick individuals showing signs of sea star wasting disease (SSWD) in two locations (intertidal and subtidal) of the Monterey, California coastline.

* Conceptual background on what the analysis does (2-3 sentences).

(Reference: Love, M.I., Huber, W., Anders, S., Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology* 2014, **15**:550. 10.1186/s13059-014-0550-8)

The package DESeq2 uses negative binomial generalized linear models to compare mRNA expression levels between sample groups and locations to infer genetic differences that may be relevant to our study (pathogen resistance or susceptibility). Our RNA-seq count data will be organized in a table to show the exons of each gene by row for each of our samples. By assigning a design formula, we can choose the explanatory and controlling variables that will be included in our model (with the last variable defaulting to represent the main effect of the model). After running the model, a results table can be easily generated and ordered by the smallest adjusted p-value to establish the significance of differentially expressed genes. Using the corresponding plots, we can visualize the log2 fold changes of expressed genes corresponding to a given variable.

* Verbal description of the mechanics of the pipeline (3-4 sentences).

mechanics (look at papers) = a short description of how you went from sequence data to analysis of expression codes - generally describe that we had the raw data, cleaned, visualized, mapped them to the reference transcriptome,

describe models = how you set them up in deseq2

Using the transcriptome

Looking at protein coding mRNA

The epidermal biopsy of sea stars was obtained and tissue prep was completed in the lab. RNA was extracted and a poly-a tail was used to sequence mRNA. and RNA library generation, high throughput sequencing using short sequence reads, transcriptome reconstruction, reads are mapped: aligning to a reference genome, read quantification: counts per transcript (or per exon),

R data analyses:

group analysis:You just need to add the "contrast" statement to the 'res2' results object. It's the contrast statement that let's you test for your custom hypotheses of interest. The generic form is:  
  
res2 <- results(dds2, contrast=list(c("int.H."), c("Int.S.")), listValues=c(1/2, -1/2))  
  
So, this custom contrast tests differences between H and S but ONLY for the intertidal group. You'd want to set up a similar contrast separately to test H vs. S for the subtidal group.

* Present results (3-5 sentences).
* Tables and figures with legends.
* Interpretation (3-5 sentences).
* Critical thinking (2-3 sentences). What would you do differently? What would you do next?