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

R Kirsten Tyler

**Introduction:**

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

**Methods:**

The epidermal tissue of *P. ochraceus* sea stars was obtained and mRNA was extracted. A library was prepared, then sequenced using Illumina. The raw data was trimmed and cleaned using FastQC, then the reads were mapped to a reference transcriptome and loci were predicted using Trinity.

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. Our RNA-seq count data was organized in a table to show the exons of each gene by row for each sample. By assigning design formulas, we chose the explanatory and controlling variables that were included in our models (with the last variable defaulting to represent the main effect of the model). We then visualized the log2 fold changes of expressed genes corresponding to given variables.

**Analysis:**

The models were set up in DESeq2 (see <https://github.com/rkirstentyler/Ecological-Genomics-PBIO381/blob/master/DESeq_HW2_RKT.Rmd> for Rmd file with code) to compare gene expression in two ways: first, we looked at the effect of health, controlling for location. Then, we chose to compare health status between the two locations using a “group design” in DESeq2. To help focus on a subset of the data, we pasted together the different variables of interest into a group. Then, after running the model, we used the contrast statement to parse out and compare specific variables in the results. Summary tables and plots were created to visualize the results.

**Results:**

After running the models to explore differentially expressed genes in S and H individuals, some very surprising and interesting findings resulted. Model 1 focused on the contrast in gene expression between H and S in the entire data set, controlling for location. We found more genes differentially expressed in S than H, overall. There were 209 genes more highly expressed in S and there are 65 genes that are more highly expressed in H. In figure 1, we highlight an example of a gene showing this trend. Figure 2 is a summary of the results from this model.

Model 2 (summaries in figures 3 and 4) explored differential expression of S and H between locations. We found a large difference between intertidal and subtidal expression levels. In the intertidal location, we found 237 genes that were differentially expressed in S and 58 genes that were differentially expressed in H. This is a stark difference from what was found in the subtidal location. Only 7 genes were differentially expressed in S and 31 genes were differentially expressed in H in the subtidal location.

Between the two models, a trend is visible in that S had more differentially expressed genes than healthy individuals (see figure 1 for an example of this trend). The most interesting finding, however, was in the number of differentially expressed genes in S between locations. There were far more genes differentially expressed in S of the intertidal location than in the S of the subtidal location.

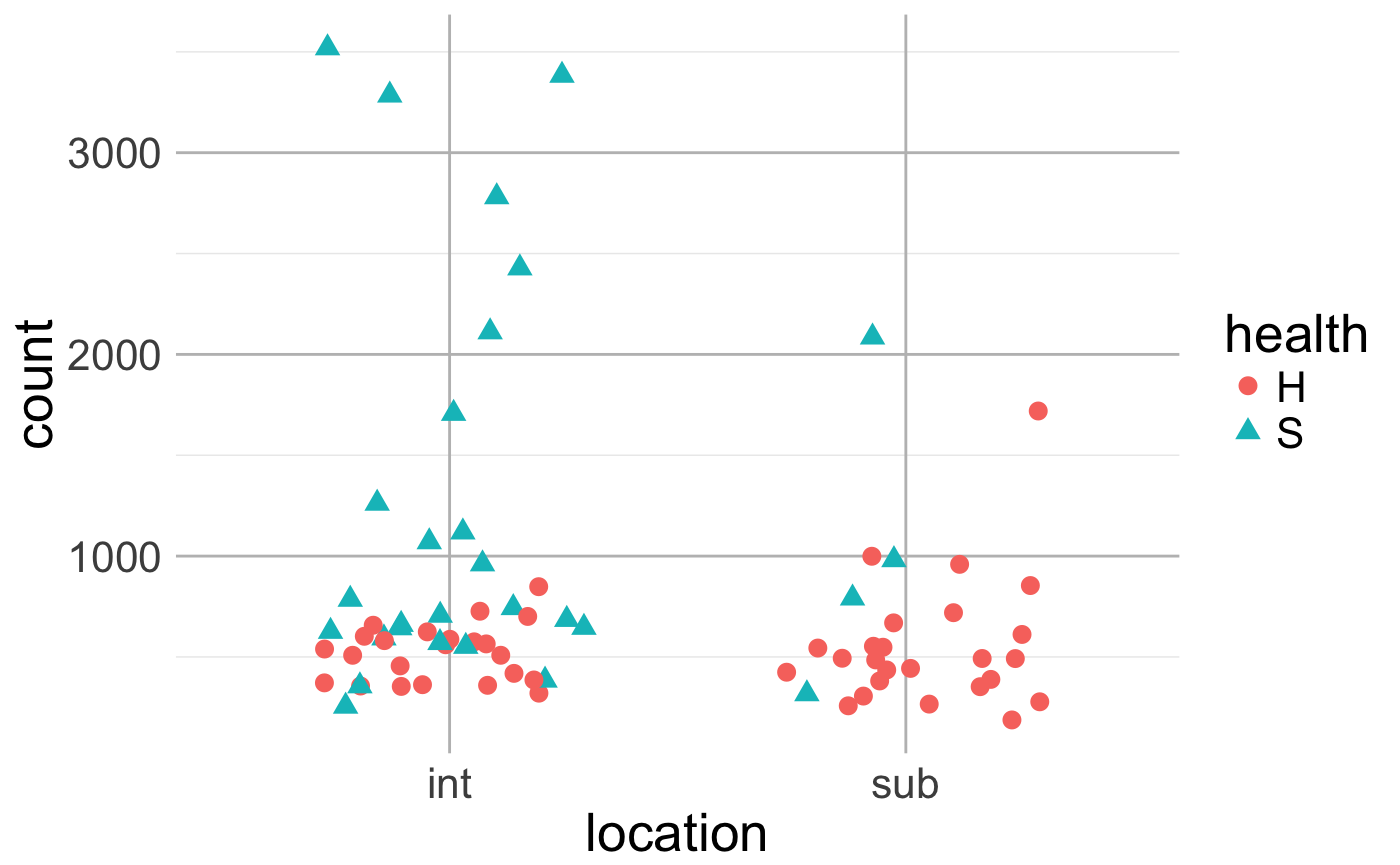
**Figures:**

Figure : Taking a closer look at a differentially expressed gene between locations.

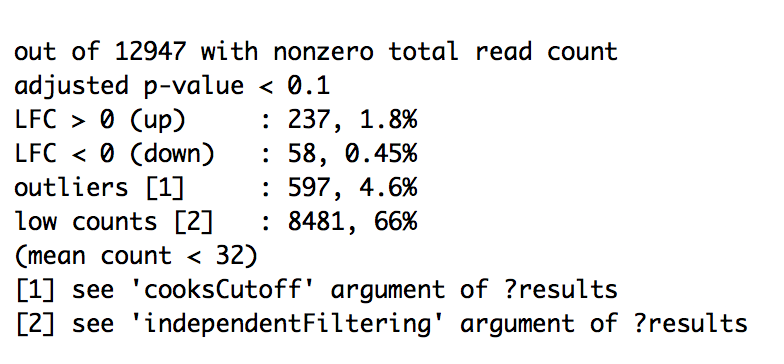
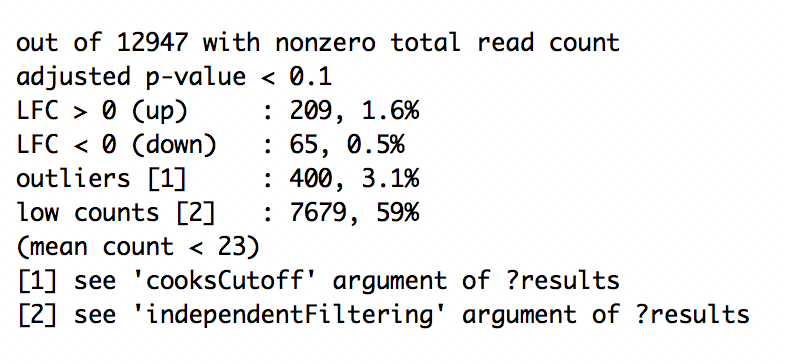


Figure : Summary of model 1. Figure 3: Summary of model 2, intertidal comparison.

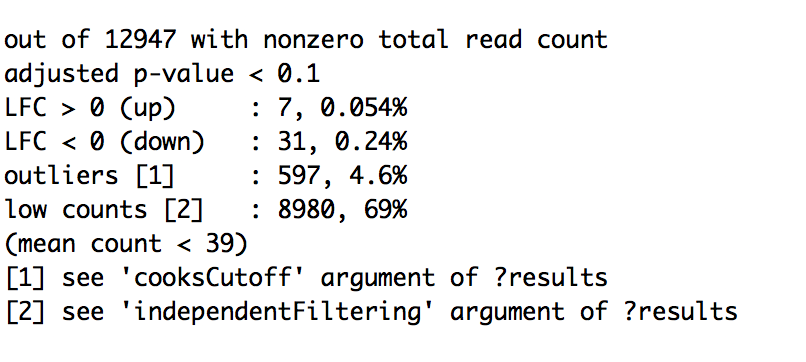


Figure : Summary of model 2, subtidal comparison

**Discussion:**

We found increased expression of genes in S in the intertidal location but this result does not tell us why. Perhaps there are genes associated with immune response that are expressed when a sea star is sick. This would explain the overall difference of S vs. H but it doesn’t explain the intertidal finding. Are there just more sick sea stars (or sicker?) in the intertidal zone? There may be positive selection for immune-related genes in the intertidal zone, causing sea stars located there (and their offspring) to have more genes that code for immune response.

Future studies should investigate the physical functions that these differentially expressed genes provide for the organism. Also, are sea stars found in the intertidal location genetically different than those found in the subtidal location? This research has provided us with new tools to further explore SSWD – a pathogen that is destroying sea star populations globally.