**Case Study 6: Gene Expression of corals at natural CO2 seeps vs. non-seep sites**

In today’s activity, we will reproduce the results of Kenkel et al. " Functional genomic analysis of corals from natural CO2-seeps reveals core molecular responses involved in acclimatization to ocean acidification." *Global Change Biology* (2017):1-14.

The counts tables for the coral and symbionts were downloaded from the original study. We will replicate some of the analyses but with some updates to the code for today’s most current analysis tools.

1. How many transcripts are in the original counts table for the coral?
2. After we filtered the counts tables for low counts, how many remain for the coral data set?
3. Make a copy of the PCA from the filtered coral data set below. Which samples are potential outliers? Can you justify removing those samples, yes or no? Are there any commonalities between these outlier(s)?
4. After removing the outliers and re-plotting the PCA, how do you feel about the outliers you removed?
5. Which principal components show the best grouping of the data based on the location (origin) and treatment (seep or non-seep)? How much variation is explained by these axes? Does this make sense given the paper’s findings?
6. Fill out the table below with your results for the significant differential expression:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **HOST ONLY** | Origin | Origin - paper | Treatment | Treatment - paper |
| Upregulated |  | 320 |  | 31 |
| Downregulated |  | 183 |  | 30 |
| **TOTAL** |  | 503 |  | 61 |

1. What happens if you change the alpha level? What do you think could be causing differences between your estimates and the original paper?
2. We will attempt to pull out the top 50 significant genes to replicate Figure 1 PCA. How does your plot compare to panel Fig 1c?
3. Next, we will plot individual genes that are significantly different between origin, treatment or the interaction. Are these more or less useful than other figures presented in the paper? Why or why not?
4. We will replicate the Venn diagram in their Figure 1A panel. How does your plot compare?
5. I am missing the gene names and associate GO annotation for these genes, thus we cannot replicate the GO analysis without further analysis or emailing the corresponding author. What does this say about the reproducibility of this paper thus far? How confident are you in the results?
6. What does this say about working with non-model systems?