**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. (done)

|  |  |  |  |
| --- | --- | --- | --- |
| Regulation | Intertidal | Subtidal | All (location control) |
| Up | 205 | 20 | 209 |
| Down | 37 | 113 | 65 |

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.

Subtidal has more down regulation of genes while Intertidal has more up regulated genes. There are 48 individuals from the Intertidal and 29 from the subtidal. When looking at all the individuals while controlling for location, there are more genes up regulated than down regulated, similar to the subset of just Intertidal.

Which gene most regulated.

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.

We use RNA-Seq to analyze the transcriptome at a snapshot in time. Expression measured with number of reads for each locus, these are called counts after assembly.

* Clear statement of objective (1 sentence).

Analysis of differential expression for healthy and sick individuals of P.--- for the Intertidal and Subtidal subsets and all data analyzed controlling for location.

* Conceptual background on what the analysis does (2-3 sentences). for using RNA sequencing for gene expression analyses

DESeq2 analyzes count data from RNA-Seq, the number of sequence fragments assigned to a gene for each sample.

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

Raw reads were assessed for quality and trimmed (adapters removed). Using FastQC, we decided on sequences with Phred Quality scores of 30 or greater.

* Present results (3-5 sentences).

PCA: “those linear combinations explain most of the variation to explain the data”

* Tables and figures with legends.
* Interpretation (3-5 sentences).

Dif between Log2 fold change.

Int: 2.5 (4x counts upregulated in sick individuals)

Sub: -5.7 (significantly down-regulated)

* Critical thinking (2-3 sentences). What would you do differently? What would you do next?

Intertidal

out of 12399 with nonzero total read count

adjusted p-value < 0.1

LFC > 0 (up) : 205, 1.7%

LFC < 0 (down) : 37, 0.3%

outliers [1] : 0, 0%

low counts [2] : 9381, 76%

(mean count < 41)

Subtidal

out of 12392 with nonzero total read count

adjusted p-value < 0.1

LFC > 0 (up) : 20, 0.16%

LFC < 0 (down) : 113, 0.91%

outliers [1] : 647, 5.2%

low counts [2] : 4289, 35%

(mean count < 13)

All (location)

out of 12947 with nonzero total read count

adjusted p-value < 0.1

LFC > 0 (up) : 209, 1.6%

LFC < 0 (down) : 65, 0.5%

outliers [1] : 400, 3.1%

low counts [2] : 7679, 59%

(mean count < 23)

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.

VSD= VST (search on R)