**Hackdash #3, 2018**

[Abundance files and study design.](https://drive.google.com/drive/folders/1scfKwSRCpwp-2OM0xZVDpu2moeCKIGwo?usp=sharing)

Because everyone now thinks you are an expert at RNAseq data analysis, a collaborator asks for help in identifying the major transcriptional changes that occur in skin following infection with the protozoan parasite *Leishmania braziliensis*. You have been given RNAseq data from the skin of 5 healthy volunteers and 5 patients with cutaneous leishmaniasis. The data has already been aligned using Kallisto, but you will need to carry this data through exploratory analysis, differential expression, and functional enrichment analysis. Your goal in the next two hours is to answer the following questions:

**Questions:**

1. What is the gene name of the 12391th row of the output of tximport? (**2 points**)

2. What are the total number of counts for each sample within Txi\_gene$counts (**2 points**)

3. How many genes were filtered out for having cpm>1 in at least 5 samples? **(2 points)**

3. Create a dendrogram and include an image here. Describe how your samples are clustering. (**2 points**)

4. Complete a PCA and include an image here. What do PC1 correspond to? Are there any outliers in the dataset? Which sample was this? (**4 points**)

5. How many genes are differentially expressed (FDR < 0.05, abs LFC > 1) when you compare disease vs. healthy skin? How many of them were up and downregulated in disease relative to healthy skin? (**2 points**)

6. Create a heatmap of all differentially expressed genes (FDR < 0.05, abs LFC > 1). Label the groups with colors. (**4 points**)

7. Comparing cutaneous leishmaniasis to healthy skin, and using GO BP terms and the DAVID web resource, what are the two most enriched Annotation Clusters upregulated in disease? **(5 points)**

8. Using GSEA software to investigate only the Biocarta portion of the C2 collection from MSigDB, answer the following questions:

a) What are the top 3 enriched pathways in disease condition? Indicate normalized enrichment score and FDR. **(3 points).**

b) What is the 11th enriched pathway in disease condition relative to healthy skin? Indicate NES and FDR. **(3 points)**

c) What is the 4th enriched pathway in disease condition relative to healthy skin? How many genes from the dataset were observed in this pathway? How many genes had significant sore enrichment in this pathway? How many genes this pathway has originally from the database? **(4 points).**

d) Give a brief description and interpretation of enriched signatures in healthy skin relative to disease. **(2 points).**