**Group 1 & 2 Exercise instructions**

You will be analyzing a subset of the data looking at **treatment effects of EVPA compared to EV** on either CF primary monocyte-derived macrophages (Group 1) or WT primary monocyte-derived macrophages (Group 2).

Setup

1. To get started, read the raw count csv file for your group into R using the read.csv() function. **Make sure to set the row.names argument to 1** so that the ENSEMBL gene IDs from the first column of the csv file get read in as row names of your count data.
2. Inspect the column names of your data which contain the sample names. The first element of a sample name indicates whether the cells were CF or WT, the second element is the unique donor ID (letter+number combination) and the last element indicates the treatment (either EV or EVPA). Because your samples are already informatively named, you will not need to rename your counts table columns.
3. **You need to create two grouping factors, one for the treatment groups and one for the donors.**
4. Use cbind() to double-check that your new grouping factors agree with the sample names in the columns.

Part 1 – Exploratory Analysis

Perform exploratory analysis of the data using the script from the Exploratory Analysis module as a template.

Here are some ideas of what you may want to try:

1. Visualize your library sizes.
2. Are there any systematic differences in total counts based on your factors?
3. Determine how many genes had zero counts and remove these genes (rows) from the data set. What does your count distribution look like before and after this?
4. Make a boxplot of log2 raw counts after removing all genes that had zero counts.
5. Make a Cluster Dendrogram of the raw data.
6. Do a simple normalization of your data and visualize your normalized data with a boxplot of log2 raw counts and a cluster dendrogram. Did normalization change anything?
7. Make a PCA plot and use color to highlight groups.
8. Is more of the variability due to donor or treatment?

Part 2 – Differential Gene Expression Analysis

Perform differential gene expression analysis with edgeR using the script from the edgeR module as a template.

1. When creating the DGEList object, include the gene annotation information from the "Annotations.csv" file as your argument in “genes =”.
2. Include both treatment and donor in your experimental design matrix.
3. **When running the glmLRT function, be careful to use the correct name to specify the comparison of interest. For the purpose of this exercise, we are interested in the treatment effects.**
4. After executing all the code of the edgeR workflow, answer the following questions:
5. How many genes are differentially expressed (p<0.05 & log2 FC > 1) between the two genotypes?

b) Make an MA plot of edgeR results with DE genes (p<0.05 & log2 FC > 1) highlighted in red.

c) Generate a Volcano plot of edgeR results.

d) Is there an uneven number of genes that are up- or down-regulated when comparing the two genotypes?

Optional Part 3 – Pathway Analysis (Bonus)

If there is time left, you may try to perform pathway analysis (using R or an online tool) to see if any pathways are enriched for genes that are differentially expressed between EVPA and EV treatment.