Bahaa Noori DSP 539 Dr. Rachel Schwartz

**Differential expression of genes after in the presence and absence of FANCD2**

This experiment was done on Fanconi Anemia ear fibroblast cell line. This cell line (+G) missing FANCD2 gene which encode for FANCD2 protein that play a key role in the DNA repair. We made a complement cell line to the (G) and called it (D2). We set RNAseq experiment in which we treated those cell lines with Aphidicolin (induces double strand breaks in DNA). In this experiment we had 4 replicates of each cell line. We got the data, and it was analyzed.

In this project, I did a principal component analysis (PCA) on a .csv file containing all the differentially expressed genes. This table include the all the replicates, p-values, log2fold change and gene names. I exported this table to RStudio and did a PCA on this table and then I graphed the data.

The data was analyzed using DESeq2, which is a count-based statistical method that require input data obtained from RNA-seq or other high-throughput sequencing experiment in the form of a matrix of un-normalized counts. After analysis with DESeq2, a PCA can be performed then visualize the sample-to-sample distance. In this method, the data points are projected on a 2D plane where they spread out into two directions that explain most of the differences. The x-axis is the direction that separate the data points the most (written as PC1), and the y-axis is the direction that separate the data the second most (written as PC2).