# Peer-graded Assignment: Exploratory Analysis

[**Instructions**](https://www.coursera.org/learn/genomic-data-science-project/peer/YQkBz/exploratory-analysis)

After summarizing your genomic data the next step is to load the data into R for analysis with Bioconductor. Load the phenotype table and the gene expression matrix created in the previous steps into R (it may be a good idea to create a SummarizedExperiment object). The next step is to explore the data for important features.

Hint: In this, and the next questions, it will be important to account for differences in sequencing depth between samples. The classic way to do this is to compute the number of reads mapped to a feature per million reads mapped.

## **Review criteria**

Upload a pdf file with the main results of your exploratory analysis (at most 3 pages). At minimum do the following. Make a boxplot of the expression levels for each sample and assess whether any transformation should be applied. Then perform a principal components analysis on the data and plot the top two principal components in a scatter plot. Color the points by variables you collected in your phenotype table object and see if there are any correlations between principal components and known variables, especially variables that are not the phenotype of interest. Keep these in mind in the next section.

* Did they upload a document with an appropriate exploratory analysis?
* Did they make boxplots by sample and pick a transformation?
* Did they do a principal components analysis with the points colored by phenotype and other variables?

**PROMPT**

Upload a pdf file with the main results of your exploratory analysis (at most 3 pages). At minimum do the following. Make a boxplot of the expression levels for each sample and assess whether any transformation should be applied. Then perform a principal components analysis on the data and plot the top two principal components in a scatter plot. Color the points by variables you collected in your phenotype table object and see if there are any correlations between principal components and known variables, especially variables that are not the phenotype of interest. Keep these in mind in the next section.

Exploratory Data Analysis

[Exploratory Data Analysis](https://coursera-assessments.s3.amazonaws.com/assessments/1632243158002/9a26ddd7-537d-4356-d5ca-ea3007a5b618/Assignment%20W7_%20Exploratory%20Analysis.pdf)