From 21/06/07 to 21/06/11

* Filtered the data using R, with a CPM of 10 on at least 36 sample.
* Normalized the counts using the TMM method.
* Determined and plotted:
  + The number of samples per tissue.
  + The number of counts and the mean count per sample.
  + The number of counts and the mean count per gene.
  + The number of counts and the mean count per tissue.
  + The number of counts and the mean count per gene per tissue type.
  + The number of counts and the mean count per sample per tissue type.
* Implemented PCA in python and applied it on the bladder, cervix uteri, fallopian tube and the bone tissues.
* Plotted the standard deviation in relation to the mean count per gene and the mean count per sample for the previously mentioned tissues.

21/06/14 to 21/06/18 goals

* Apply PCA on the complete dataset.
* Implement the mean-variance plot in R and compare it to the plot I made in python. Apply it then to the complete dataset.
* Select the top expressed genes, run them by gene ontology, perform hierarchical clustering on them and plot the heatmap. (Note: I should read about the different methods of correlation and how they would impact the results).
* Apply the previous work to the TCGA dataset.
* Compare the histone variants and the histone chaperones expression patterns across GTEx and TCGA.