

## Assignment 1 - Akavia

Load up the CCLE expression. In order to get the files, please open up <https://portals.broadinstitute.org/ccle> and register. Once you've logged in, you can download the RPKM data, which is called **CCLE\_DepMap\_18q3\_RNAseq\_RPKM\_20180718.gct**.

I've provided a short matlab script to load the data, and I've provided an additional file sampleType.txt that has three columns - sample name, cell line name, tissue type.

The assignment includes

- 1) Loading the data
- 2) Identifying noise threshold by comparing potentially identical cell lines specified in a table below.
- 3) Identifying noise threshold by comparing 4 of the potentially identical cell lines to the dataset average
- 4) Running t-test between BREAST and PROSTATE samples - getting a list of genes, and their p-value
- 5) Optional- running t-test between BREAST and all other samples

Potentially identical cell lines

253JBV	253J
HCC1588	LS513
DLD1	HCT15

Matlab code for loading the data (also present as AkaviaAssingment1Skeleton.m)

```
%% Define files
ccleExpressionFile =
'~/Documents/Data/CCLE/expression/CCLE_DepMap_18q3_RNAseq_RPKM_20180718.gct';
ccleSampleTypeFile = '~/Documents/Data/CCLE/expression/sampleType.txt';
%% Read data
ccleExpression = readtable(ccleExpressionFile, 'FileType', 'text',
'Delimiter', '\t', 'HeaderLines', 2, 'ReadRowNames', 1, 'ReadVariableNames',
1);
genes = ccleExpression.Properties.RowNames;
geneSymbols = ccleExpression.Description;
expressionMatrix = ccleExpression{:, 2:end};
ccleSamples = readtable(ccleSampleTypeFile);
```