

Project 2 – Signature creation and iLINCS API

This assignment builds upon the R/shiny class and expands the API example.

1. For the assignment use the dataset `TCGA_breast_cancer_ERstatus_allGenes.tsv`. You may choose to “debug” your code with a smaller dataset `TCGA_breast_cancer_ERpositive_vs_ERnegative_PAM50.tsv` or `TCGA_breast_cancer_LumA_vs_Basal_PAM50.tsv`.
 - Not a question
2. Your assignment is to develop code using R Shiny to present user interface that allows a user to upload a tsv file, presents an option to split the samples into two groups based on the available metadata and create a signature. The signature is then submitted to the iLincs API which retrieves concordant signatures.
 - Template already does this
4. Template is provided in the Project 2 github project (<https://github.uc.edu/uc-datascience/Project2.git>).
 - not a question
5. The template is missing the calculation of the differential expression – please use t-test to calculate t-statistic / p-value.
 - switched the output variable to `t.test()statisticfromt.test()`p.value
6. Expand the template to allow users to filter the input file to L1000 genes only (See the include L1000.txt file).
 - added a boolean checkbox that allows the user to select whether they want to filter for L1000 genes only
7. Further expand the template to allow users to submit only top 100 differentially expressed genes.
 - added a boolean checkbox that allows the user to select whether they want to filter for only the top 100 differentially expressed
8. Compare results with iLincs
 - The top similarity score was 0.5352301897 and the compound was WZ3146 when no filtering was used
 - The top similarity score was 0.4838501194 and the compound was ZINC01050121 when both filtering options (L1000 and top 100 differentially expressed genes) were used.
9. Extra credit for a heatmap or other visualizations.
 - Created a violin plot showing distributions of similarity
 - added some code for a heatmap, but was unable to use it to plot entire big dataset (not performant enough)

The assignment is due on – March 11, 2025 midnight.

The submission should be zip compressed file named “project2-[*your UC username*].zip” (e.g. “project2-lastnfi.zip”) which includes any supporting R files. The zip file should be uploaded canopy. The assignment entry in Canopy will be created shortly.