Progress:

We have created a "simple_teximport" package, which is a simplified, python version of the tximport library we used in lab4. Our simple_teximport is able to extract and combine raw counts from different datasets into a single dataframe suitable for further analysis. The package also includes a filtering function that allows users to filter out genes with low read counts based on the user's determined parameters.

We have also created a "simple_deseq" package, which is intended to be a simplified, python version of the DESeq2 library. We have finished four major functions: base mean calculation, log 2 fold change calculation, estimated dispersion calculation, and stats calculation. We have tried our test dataset with both packages.

Github Repository:

https://github.com/Kevinxsn/simple_deseq

Plans:
1. Timeline
   a. By Week 8, 24 May 2024:
      Finish Simple_Teximport package. Finish a portion of the simple_deseq functions
   b. By Week 9, 31 May 2024:
      Finish remaining simple_deseq functions: p-value, p-adjusted. Benchmark DESeq2 tool against our version and make adjustments according to benchmarking. Take care of the details in the code: e.g. error messages, parameters, helper methods, etc. Add possible additional features: e.g. data summary, differential expression visualization, etc.
      Finish and organize files on github
   c. By Week 10, 7 June 2024:
      Finish Presentation. Finish Report

2. Changes: New Dataset
   According to the instructor's feedback, we have changed our test dataset. The new dataset is a characterization of age-associated gene expression changes in mouse sweat glands. Cells were isolated from Footpad skin from fore paws of 3 months old and 28 months old male mice. Here is the link to the dataset: link