Project Title: Simplified DESeq

Project member:
Longmei Zhang A17012012
Sunan Xu A16906965
Zhirui Zhu A16804513

Description of Tool

We are aiming to create a simplified version of the DESeq2 tool in python. This tool would compare to DESeq2 we learned in the lab. Our goal is to normalize and analyze RNA-sequencing data reads and get statistics including fold change, p-value, adjusted p-value. The preprocessing steps (like filter out low reads counts) are included in the package. We are planning to use python as the coding language so the analytical process can be done in the python interface without support from extra packages such as R packages or tximport. We also provide an option to filter the significantly differentially expressed genes based on the p-value or the adjusted p-value.

Benchmark Process:

We are planning to benchmark against the RNA-sequencing data comparing the gene expression of mice fed with Chow and with High fat diet. We used DESeq2 to analyze this dataset. We would compare the result from our tool with the result we got from the lab, particularly the fold-change, p-values, adjusted p-values, and the differentially expressed genes.

Public Database:

We are planning to use single cell RNA-sequence results from the website GEO. The link gives a comparative gene expression RNA-sequencing data.

The study of this dataset investigates how the RNA MALAT1 affects lung cancer progression by using RNA-seq on mouse models and human cancer cells. It shows that increasing MALAT1 levels in cancer cells boosts their movement and changes the environment to favor tumor growth, mainly through increasing the recruitment of helpful immune cells and the release of a specific protein, CCL2. Blocking these changes reduces the cancer-promoting effects of MALAT1. The result suggests targeting MALAT1 could be a new way to treat lung cancer.