CS 461: Project 2 Write Up

Student: Cameron M.
Instructor: Phi Hung Nguyen / Tin Nguyen

Introduction:

The data set is acquired from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19804 . Contains data from genome wide-screening of transcriptional modulation in non-smoking females was performed. RNA is extracted from tumor and normal lung tissue for gene expression analysis.

Analysis:

Statistics Acquired:

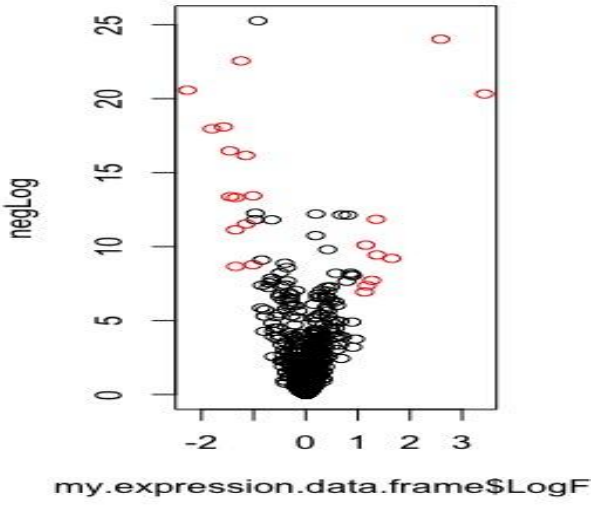
- T-Score and P-value [t.test(control vs condition)]
- LogFC [mean difference]
- Empirical T-test [e.T-score, e.P-value, e.E-score]
- Euclidean Distance between group means [E-score]

A t-test was utilized to compare control and condition groups

T-score, P.values, across the matrix were obtained and used to graph a volcano plot with

- LogFC vs (-logFC)

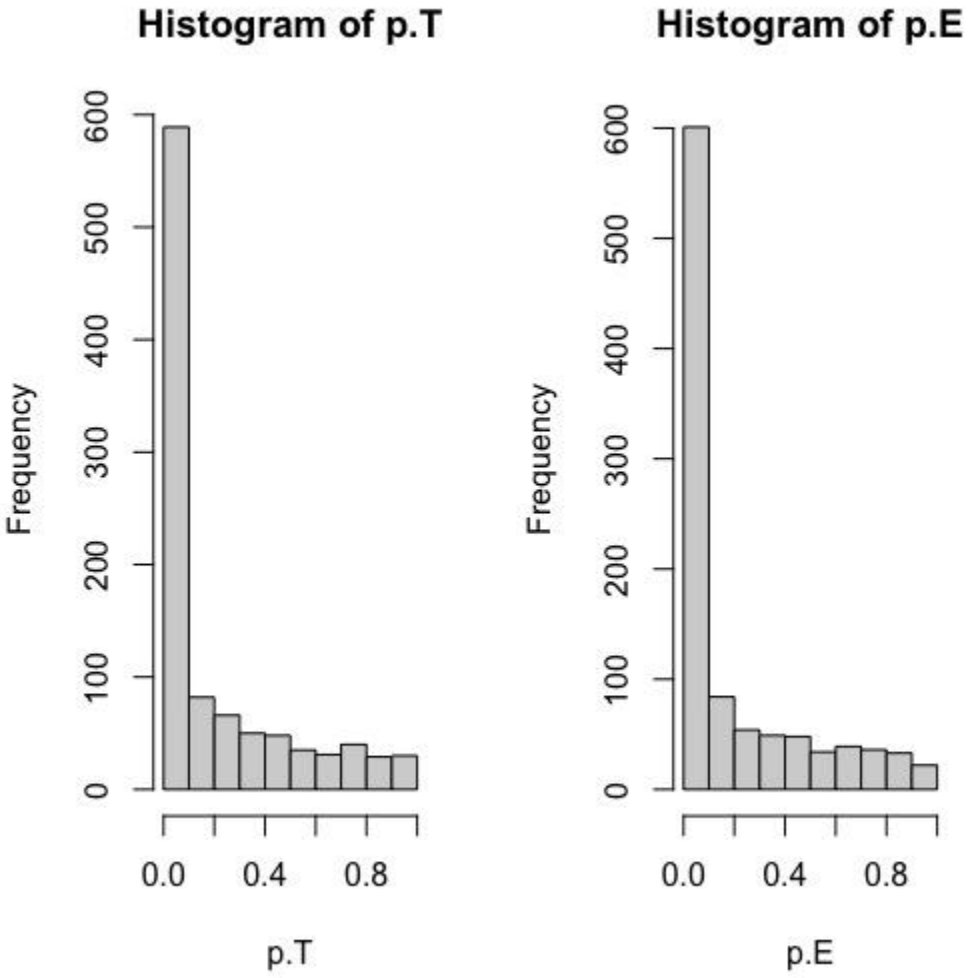
Red points indicating to differentially expressed genes



Due to the nature of the volcano plot, we see there is potential support to say that gene expression is not identical between groups but does not appear to have too much variance from the most dense point of distribution.

Randomly Scrambled DATA VALUES:

p.E by through permutation analysis with euclidean distance between group means
p.T ——————————————————————————> p.values for t.test between control and cancer groups



Permutation Analysis was utilized to find empirical p-values for T-score and E-score which are displayed above:

- High correlation (0.9906919)

This high correlation score reveals a relationship between the two variables. The observed data may not have occurred by chance and different from an expected from a null hypothesis. If the null hypothesis is “no gene expression occurred”, then we see some change but cannot conclude that it is positive or negative.

Further questions for future experiments:

- Which genes may be “overdriven” resulting in a negative impact on a cell’s contribution to human health within a lung cancer
- If the test subjects do not have a history of smoking, what stressor could have an influence on a cell’s ability to perform / come into existence