We explored the original Toden et al. dataset by investigating the distribution of inherent data along with using Principal Component Analysis (PCA) to investigate genes that may affect Alzheimer’s diagnosis.

- Genes accounting for the highest variation of subjects are related to brain function, inflammation, cell signaling.
- The approximations of normal distributions are probably adequate for most of the genes, though less so for genes with lower average counts.
- While the genes with the most explained variance have biological significance in Alzheimer’s disease, the reverse is not true.

We benchmarked Toden et al.’s data and classifier by investigating the effect of the trained coefficients, normalization, and consistency of data.

- Z-score normalization of data was used to develop classifier and performance decreases if TPM are used.
- Variation in data along each gene are consistent regardless of subset.
- Classifier is accurate in its predictions aside from a few patients near cut-off score.

References