

HW1

Course: Advanced Biostatistics/Biochip Informatics

Solve the following problems with the given data

1. Filter the data by removing all rowSums equal to zero and hence show the number of genes that remain.
2. Compute the log transformed pseudo counts and visualize the histograms for 1 tumor sample and 1 normal sample using the given raw counts and pseudo counts.
3. Show the mean and variance relationship with plots for the normal and tumor samples separately.
4. Normalize the raw data by the following methods:
 - a) Total Count (TC)
 - b) Upper Quantile (UQ)
 - c) Trimmed Mean of M-values (TMM)
 - d) Relative Log Expression (RLE)
5. Make a comparison between different normalization methods (and original count data) by boxplots.
6. Perform differential expression analysis with DESeq2. Show the results for Wald test, and likelihood ratio test (LRT) and interpret the results. Give a list of top 5 differentially expressed significant genes based on adjusted p -values for both Wald and LRT.
7. Provide a principal component analysis (PCA) plot for the types of samples (normal and tumor).
8. Use edgeR to estimate the gene expression variance for all genes. Show the common dispersion, and tagwise dispersion.
9. Provide the plot for the tagwise biological coefficient of variation (square root of dispersion).