

Assignment W7: Exploratory Analysis

Load and Preprocess Data:

Load the expression data and phenotype data into R and preprocess, create a feature table with gene annotations.

SummarizedExperiment: Combine all the above into a SummarizedExperiment object to use for further analysis

class: SummarizedExperiment

dim: 78764 12

metadata(0):

assays(1): counts

rownames(78764): NM_001376549 NM_001376535 ... NR_144628 NR_003682

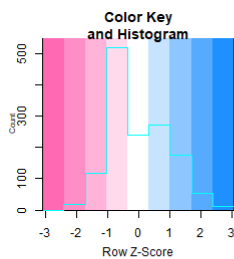
rowData names(4): rownames.gene_counts. GENENAME ENSEMBL SYMBOL

colnames(12): SRR1554537 SRR1554538 ... SRR1554561 SRR1554534

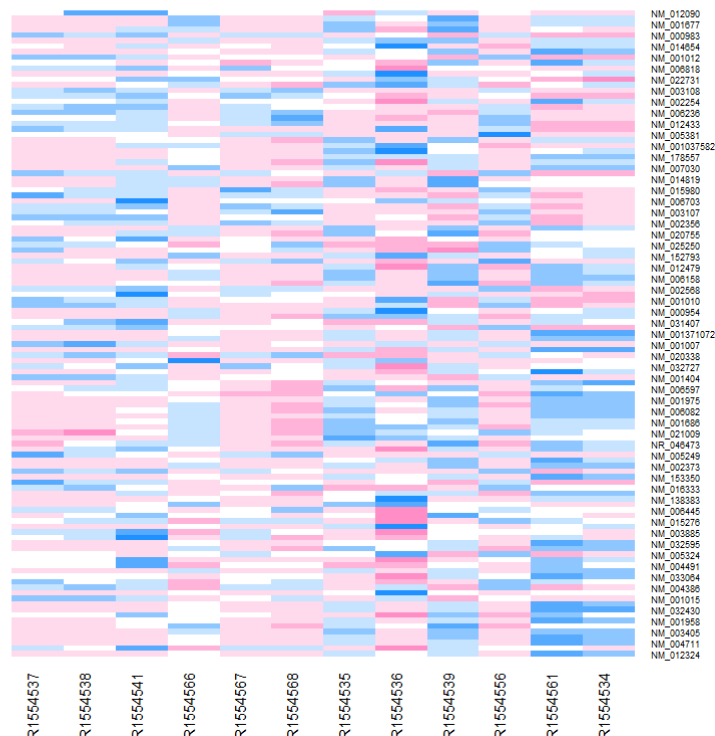
colData names(13): sample.id type ... BioSample.id avg_per_base_seq_qual

Explore Data using edgeR:

- Create a DGEList Object brain_dge and add to it samples and genes data.
- Filter gene expression data to remove lowly expressed genes.
- Calculate the normFactors to determine effective library size for further analysis (lib.size * norm.factors)
- Compute Counts per Million (CPM) and plot heatmap



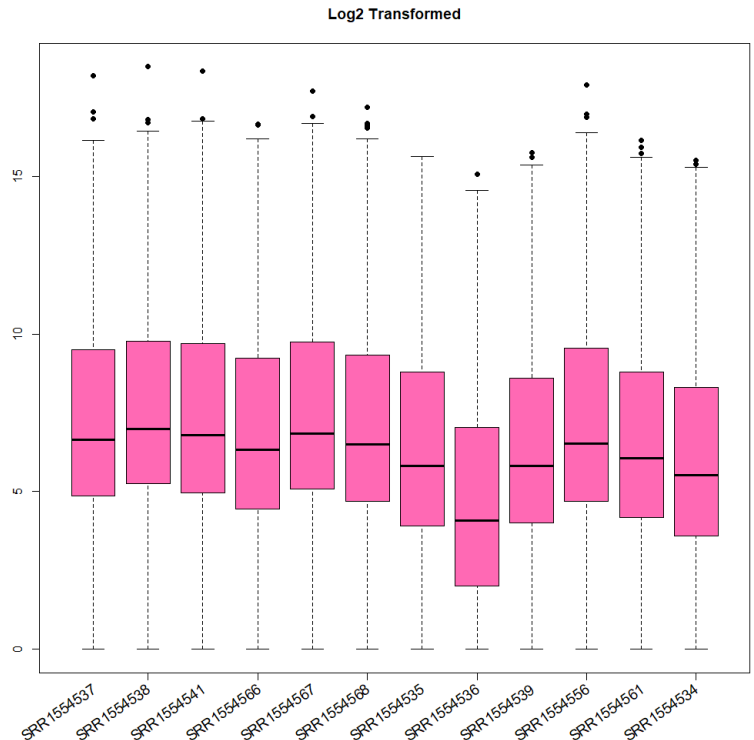
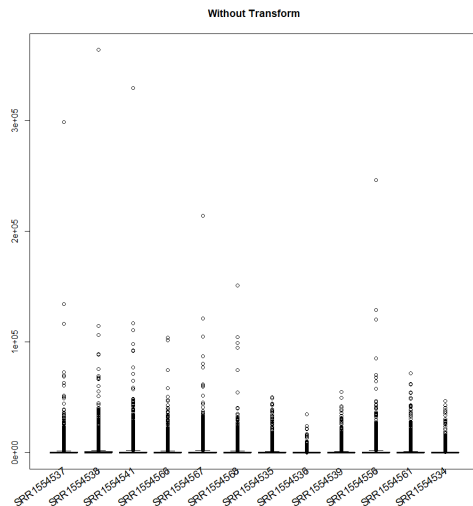
CPM



Boxplots

Make boxplots for gene expression levels for each sample. We see that most values are close to zero and we need to do a log transform to view the data.

We can see that the data is quite skewed with large number of zero values

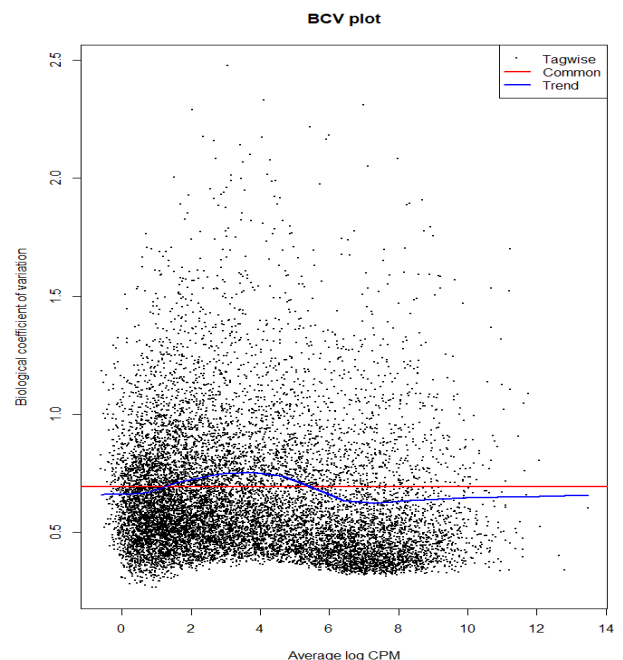
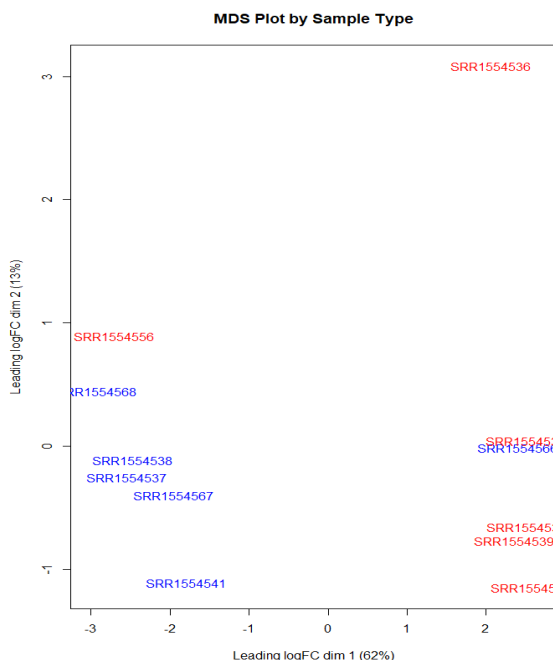


With data log transformed we can see the distribution much better and we see the gene expression is mostly similar with some outliers, with the exception of SR1554536 whose expression is much lower than that of the remaining samples.

MDS and BCV Plots

MDS plot produces a plot in which distances between samples correspond to leading biological coefficients of variation (BCV) between those samples.

The BCV plot shows dispersion estimates.



Principal Component Analysis: Compute PC1 and PC2 and plot. Color the plot based on different variables. We see that there does not seem to be a correlation between PCs and variables except the age(fetal vs adult), sample type (which is the variable we are investigating) and PC1.

