

Assignment W8: Statistical Analysis

H0 - Null hypothesis is that the relationship between sample type or age (fetal vs adult) and expression is exactly zero.

H1 - Alternative Hypothesis is that the relationship between sample type or age is not zero

Differential Gene Expression Analysis of RNASeq data using DESeq2

Filter genes with rowSums less than 10

Check factor levels:

fetal fetal fetal fetal fetal fetal adult adult adult adult adult adult

Levels: adult fetal

Create the DESeq Data object or container from the input data

Run the DESeq() function on this object brain_dds

The DESeq() will do the following in a single step

- * estimating size factors
- * estimating dispersions
- * gene-wise dispersion estimates
- * mean-dispersion relationship
- * final dispersion estimates
- * fitting model and testing

Correcting for Multiple Testing:

The DESeq() uses Benjamini Hochberg correction for the False Discovery Rate(padj-"BH adjusted p-values"). We run the results() on the above with an alpha parameter set to 0.05 to set the FDR at 0.05 level.

Next we reorder the result to order genes from lowest p-values to highest and save the table to tab delimited file differentialExp.txt (attached).

log2 fold change (MLE): type fetal vs adult

Wald test p-value: type fetal vs adult

DataFrame with 22872 rows and 6 columns

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>	<numeric>
LINC01103	1428.9972	3.05500	0.457759	6.67382	2.49234e-11	3.89328e-07
VPS28	14.3446	-4.66964	0.750377	-6.22305	4.87569e-10	3.80816e-06
PHLPP1	1285.1848	4.69888	0.763629	6.15336	7.58601e-10	3.95003e-06
IPO11	2016.3569	1.04338	0.171956	6.06774	1.29723e-09	5.06602e-06
RETSAT	16.1999	3.34740	0.605471	5.52860	3.22803e-08	1.00850e-04

Summary:

out of 22872 with nonzero total read count

adjusted p-value < 0.05

LFC > 0 (up) : 1209, 5.3%

LFC < 0 (down) : 1679, 7.3%

outliers [1] : 173, 0.76%

low counts [2] : 7078, 31%

(mean count < 4)

2888 genes have an adjusted p value below the given FDR cutoff, $\alpha = 0.05$ out of the total 22872, and are thus significant.

Volcano Plot: Plot the logFold change vs $-\log_{10}$ p-value

Log Fold Change Shrinkage for visualization using volcano plots and ranking genes

We can use the `lfcShrinkage()` with `coef="type_fetal_vs_adult"`, `type="apeglm"` to shrink the results data from above to use in volcano plot

Here we are going to use the `EnhancedVolcanoPlot()` with all default parameters

