RNA sequence data and Differential Expression analysis in asthma treatment

**Introduction**

This research shows that RNA-Seq is used to characterize the human airway smooth muscle (HASM) transcriptome at baseline and under four asthma treatment conditions. Use R studio and statistical probability to analyze which treatment plan has the greatest impact on genes in several groups of controlled trials.

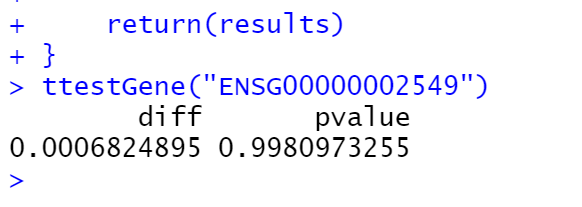
**Method**

We have adopted 1) T test, 2) Linear Regression and 3) Differential Expression Analysis. In this case we choose variables cell and dex because we care about the cell line and which samples are treated with dexamethasone versus which samples are untreated controls. The basis for this type of analysis is common when analyzing high-throughput data. It has the following steps: Extract the expression values for a single gene. Run compare the mean expression between two groups using a statistical test. Repeat former steps for every gene. Then, from dplyr package to filter out results based on padj < 0.01. 4) Gene set enrichment analysis (GSEA), is a method to identify classes of genes that are over-represented in a large set of genes.

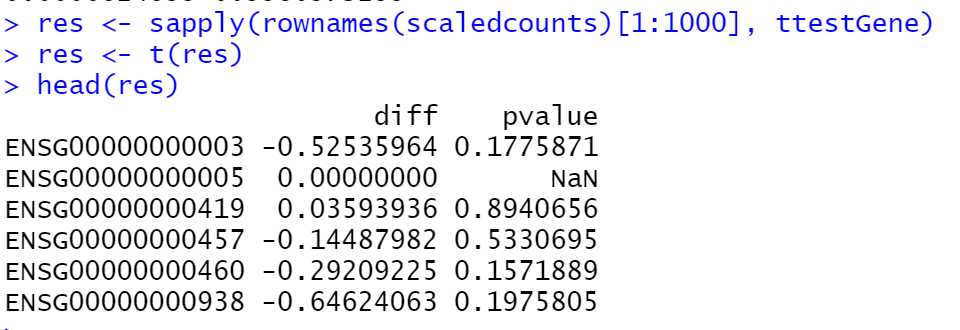
**Analysis**

From the previous section that our threshold is be a false discovery rate (FDR) of 0.1 (i.e. no more than 10% chance that the observed change in expression is due to chance). Not all the genes in the results from DESeq2 were assigned p-values so we'll start by filtering out the genes without p-values followed by storing the significant genes separately. 1 Filter out genes with no p-values 2 Keep significant genes only # 3 How many significant genes did we get? Last, we use the Gene Ontology (GO) to analysis.

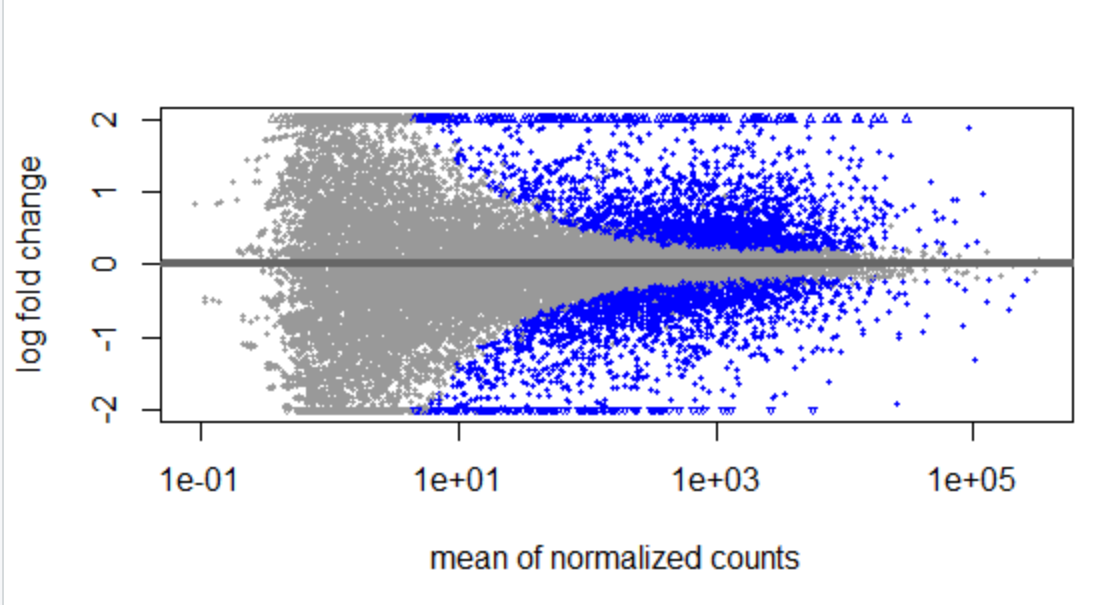
**Result**



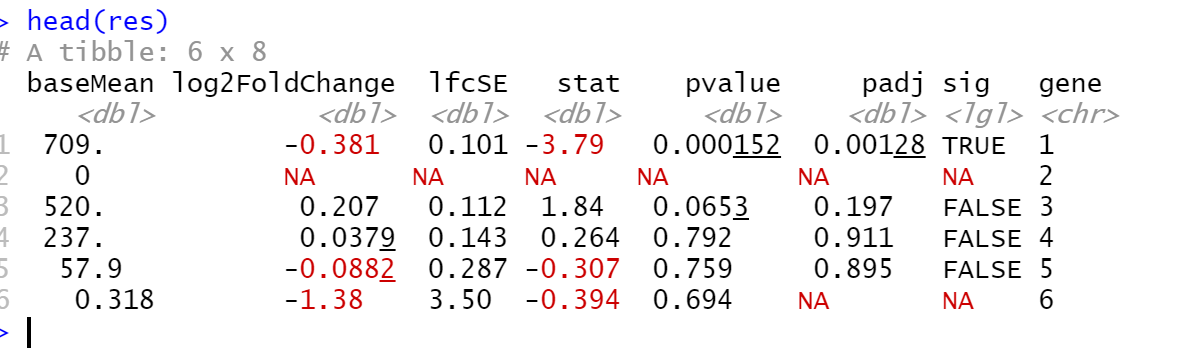
the expression values for a single gene.



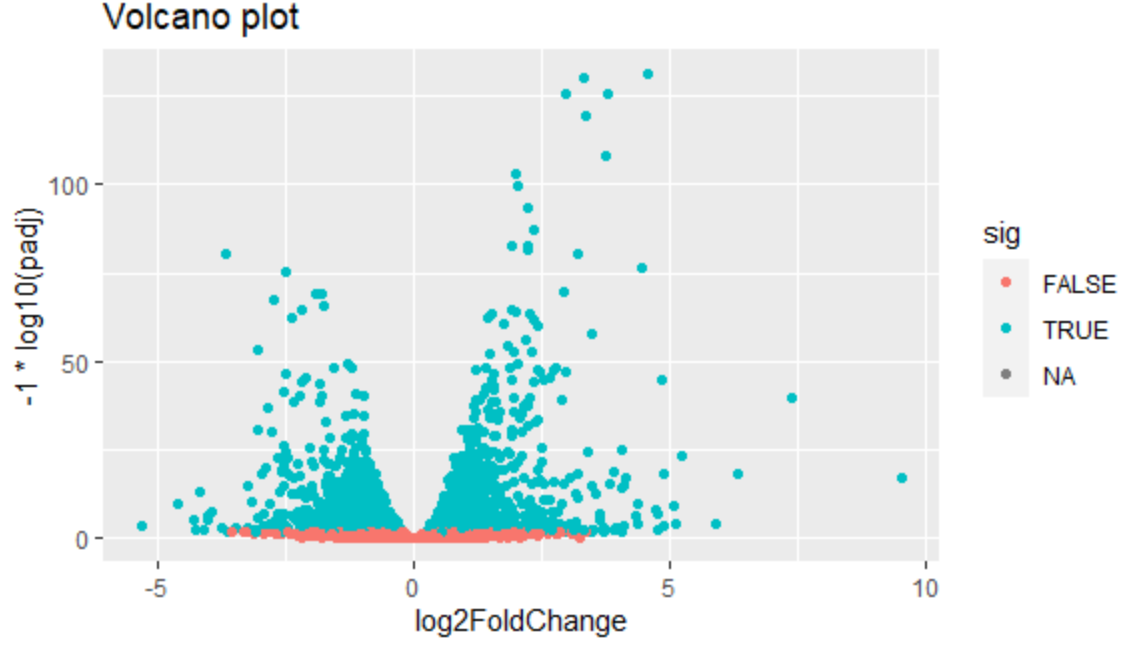
The expression values for 1000 genes



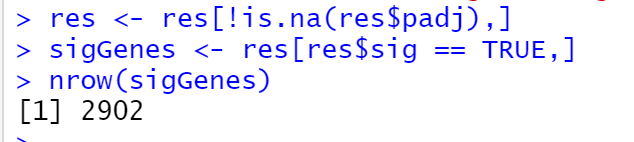
The plot shows the log2 fold changes attributable to a given variable over the mean of normalized counts for all the samples in the DESeqDataSet. Points represent genes and will be colored red if the adjusted p value is less than 0.1. Points which fall out of the window are plotted as open triangles pointing either up or down.



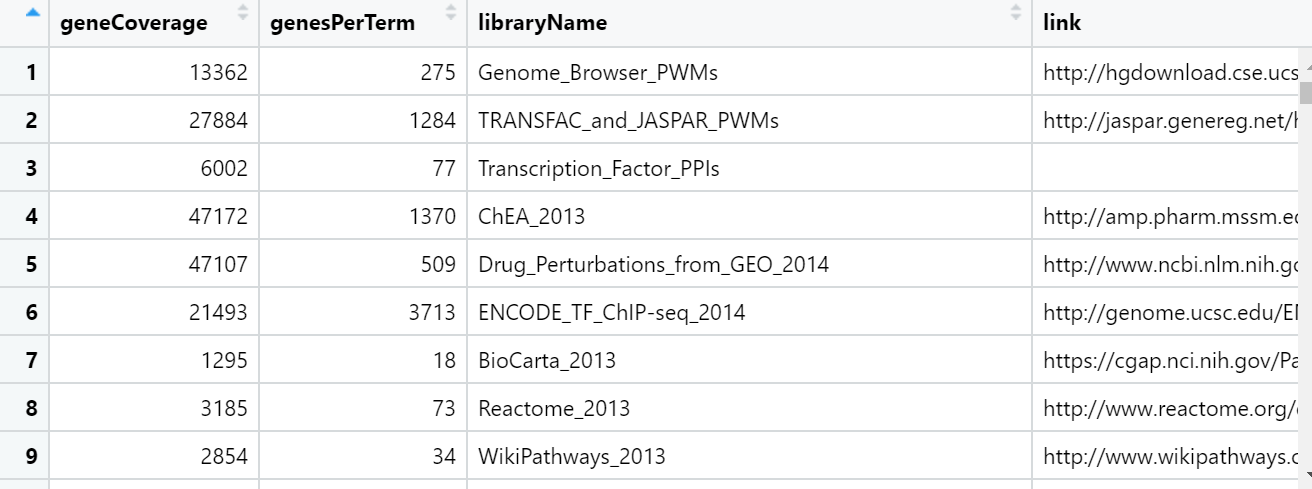
 Add a column that tell us whether each gene is significant.



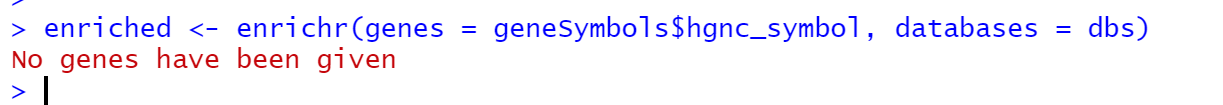
Gene sig plot



Keep sig gene



 The genes in this dataset use Ensembl indentifiers, while enrichR expects gene symbols. We'll use the biomaRt package to map our Ensembl IDs to gene symbols.



No genes can be compared

Conclusion

Through the T test of a single gene, the difference in mean is very small relative to the variance, hence the large p-value. With control experiments and filtering non-significant cells, it is found that the significant cells are low. There is no significant relationship between it and dexamethasone. Gene set enrichment analysis (GSEA) is a method to identify classes of genes that are over-represented in a large set of genes. After database comparison, there are no similar genes.

Overall, this analysis has no data showing that Dex can relate to the RNA sequence of asthma patients