Epilepsy\_var

Modibo K GOITA

2023-03-17

# Part 1 : Differential Genes Expression Analysis

##Install required packages

## Load required R packages

## Loading required package: S4Vectors

## Loading required package: stats4

## Loading required package: BiocGenerics

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, basename, cbind, colnames,  
## dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,  
## grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,  
## order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,  
## rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,  
## union, unique, unsplit, which.max, which.min

##   
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:base':  
##   
## expand.grid, I, unname

## Loading required package: IRanges

## Loading required package: GenomicRanges

## Loading required package: GenomeInfoDb

## Loading required package: SummarizedExperiment

## Loading required package: MatrixGenerics

## Loading required package: matrixStats

##   
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':  
##   
## colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,  
## colCounts, colCummaxs, colCummins, colCumprods, colCumsums,  
## colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,  
## colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,  
## colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,  
## colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,  
## colWeightedMeans, colWeightedMedians, colWeightedSds,  
## colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,  
## rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,  
## rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,  
## rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,  
## rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,  
## rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,  
## rowWeightedMads, rowWeightedMeans, rowWeightedMedians,  
## rowWeightedSds, rowWeightedVars

## Loading required package: Biobase

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

##   
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':  
##   
## rowMedians

## The following objects are masked from 'package:matrixStats':  
##   
## anyMissing, rowMedians

##   
## Attaching package: 'genefilter'

## The following objects are masked from 'package:MatrixGenerics':  
##   
## rowSds, rowVars

## The following objects are masked from 'package:matrixStats':  
##   
## rowSds, rowVars

##

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:AnnotationDbi':  
##   
## select

## The following object is masked from 'package:Biobase':  
##   
## combine

## The following object is masked from 'package:matrixStats':  
##   
## count

## The following objects are masked from 'package:GenomicRanges':  
##   
## intersect, setdiff, union

## The following object is masked from 'package:GenomeInfoDb':  
##   
## intersect

## The following objects are masked from 'package:IRanges':  
##   
## collapse, desc, intersect, setdiff, slice, union

## The following objects are masked from 'package:S4Vectors':  
##   
## first, intersect, rename, setdiff, setequal, union

## The following objects are masked from 'package:BiocGenerics':  
##   
## combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: knitr

##

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

##

## STEP 1: preparing reads or fragment counts

### Load features counts matrix dataset

### Counts col mames transformation

### Export row tab separated counts matrix

### loadind sample info

## 'data.frame': 36 obs. of 5 variables:  
## $ Accession : Factor w/ 36 levels "SRR9733947","SRR9733948",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ SampleNames: Factor w/ 36 levels "GSM3963731","GSM3963732",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ description: Factor w/ 36 levels "C1","C10","C11",..: 1 20 10 29 11 30 12 31 13 32 ...  
## $ Tissue : Factor w/ 2 levels "Hippocampus",..: 2 1 2 1 2 1 2 1 2 1 ...  
## $ Condition : Factor w/ 2 levels "mTLE","Unaffected": 1 1 1 1 1 1 1 1 1 1 ...

### Sample info row names asseignment

### Verification to sure the row names in colData matches to column names in counts data and in the same order

## [1] TRUE

## [1] TRUE

## STEP:2 Creating DESeq object from counts Matrix

### Pre-filtering the dataset

## [1] 47388

## [1] 33572

## STEP 3 using DESeq2 for differential expression analysis

### DESeq2 differential expression with DESeq function

## estimating size factors

## estimating dispersions

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

## -- replacing outliers and refitting for 144 genes  
## -- DESeq argument 'minReplicatesForReplace' = 7   
## -- original counts are preserved in counts(dds)

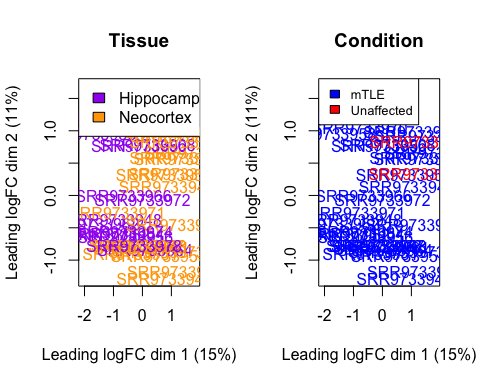
## estimating dispersions

## fitting model and testing

##   
## Attaching package: 'limma'

## The following object is masked from 'package:DESeq2':  
##   
## plotMA

## The following object is masked from 'package:BiocGenerics':  
##   
## plotMA



### Extract results from dds DESeq2 results function

### Summary of the results

##   
## out of 33572 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 4944, 15%  
## LFC < 0 (down) : 4626, 14%  
## outliers [1] : 0, 0%  
## low counts [2] : 3255, 9.7%  
## (mean count < 1)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

DESeq2 performs for each gene a hypothesis test to see whether evidence is sufficient to decide against the null hypothesis that there is zero effect of the tissue type and that the observed difference between epilepsy case and control was merely caused by experimental variability

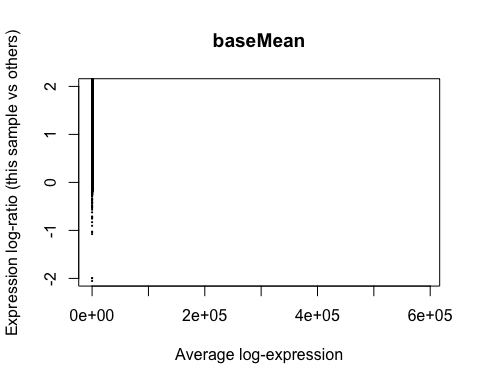
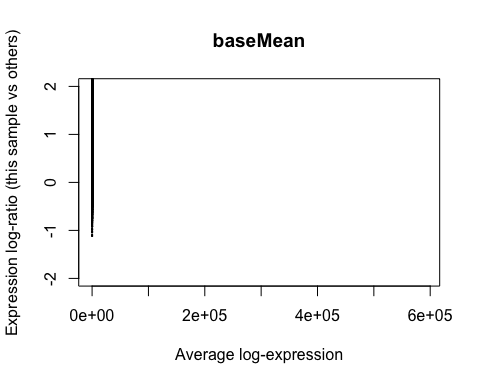
## [1] "Intercept" "Tissue\_Neocortex\_vs\_Hippocampus"  
## [3] "Condition\_Unaffected\_vs\_mTLE"

## using 'apeglm' for LFC shrinkage. If used in published research, please cite:  
## Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for  
## sequence count data: removing the noise and preserving large differences.  
## Bioinformatics. https://doi.org/10.1093/bioinformatics/bty895

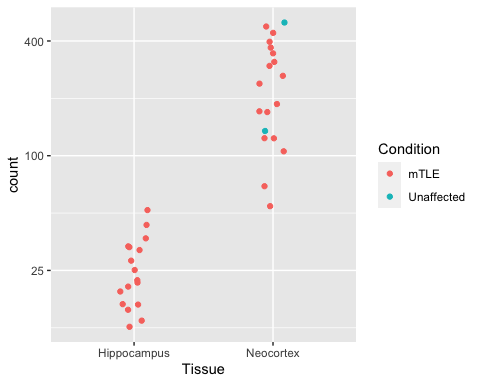
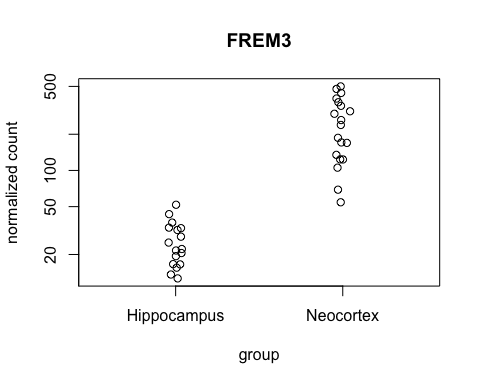
## computing FSOS 'false sign or small' s-values (T=1)

### MA plots (log ratio) and A (mean average)

Overview of the distribution of the estimated coefficients, or comparisons of interest, across all genes



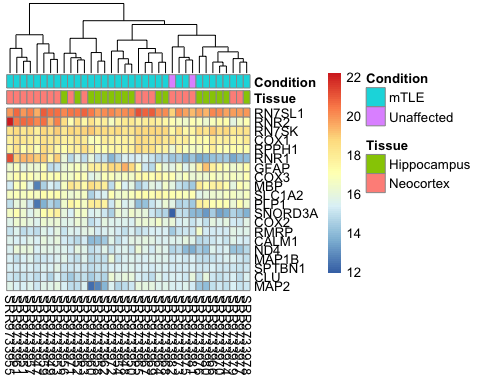
### Counts plot of individual genes



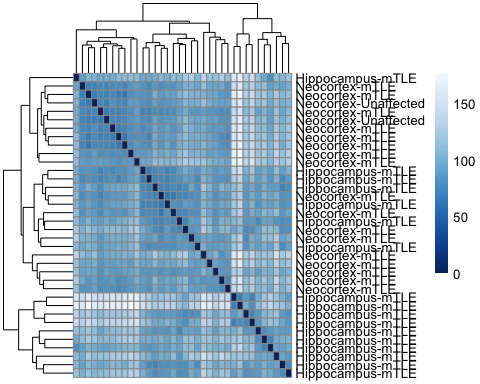
### Extract VST transformed value for all genes for each sample

### Generate heatmap

### Heatmap of top 20 significant genes



### Heatmap of the sample-to-sample distances



### Significance

subset results, consider a fraction of 10% false positives acceptable FDR alpha=0.1

##   
## out of 7617 with nonzero total read count  
## adjusted p-value < 0.05  
## LFC > 0 (up) : 3530, 46%  
## LFC < 0 (down) : 3074, 40%  
## outliers [1] : 0, 0%  
## low counts [2] : 0, 0%  
## (mean count < 1)  
## [1] see 'cooksCutoff' argument of ?results  
## [2] see 'independentFiltering' argument of ?results

### Extract up regulated genes

## log2 fold change (MLE): Tissue Hippocampus vs Neocortex   
## Wald test p-value: Tissue Hippocampus vs Neocortex   
## DataFrame with 371 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue  
## <numeric> <numeric> <numeric> <numeric> <numeric>  
## PLPP4 225.382 1.39279 0.139935 9.95316 2.44302e-23  
## SLC14A1 722.730 2.24567 0.226687 9.90647 3.90177e-23  
## CHST9 149.219 1.75749 0.189570 9.27091 1.84572e-20  
## CABP7 274.378 2.62004 0.298227 8.78539 1.55828e-18  
## SMOC1 745.662 1.20902 0.137839 8.77127 1.76665e-18  
## ... ... ... ... ... ...  
## GSN-AS1 192.5682 1.01998 0.437832 2.32963 0.0198260  
## HPN-AS1 58.6077 1.09381 0.476987 2.29316 0.0218389  
## LOC102723996 112.2532 1.03628 0.452179 2.29175 0.0219200  
## LOC124901117 76.9703 1.02423 0.455983 2.24621 0.0246909  
## LOC100507336 167.7216 1.17928 0.538155 2.19134 0.0284271  
## padj  
## <numeric>  
## PLPP4 1.64087e-19  
## SLC14A1 1.97150e-19  
## CHST9 3.29158e-17  
## CABP7 1.74972e-15  
## SMOC1 1.87731e-15  
## ... ...  
## GSN-AS1 0.0597895  
## HPN-AS1 0.0644182  
## LOC102723996 0.0645874  
## LOC124901117 0.0708522  
## LOC100507336 0.0791173

### Extract down regulated

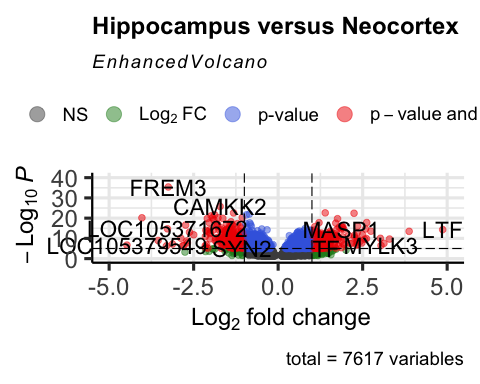
## log2 fold change (MLE): Tissue Hippocampus vs Neocortex   
## Wald test p-value: Tissue Hippocampus vs Neocortex   
## DataFrame with 6 rows and 6 columns  
## baseMean log2FoldChange lfcSE stat pvalue padj  
## <numeric> <numeric> <numeric> <numeric> <numeric> <numeric>  
## LINC01128 365.4474 -0.534180 0.1122264 -4.75984 1.93748e-06 3.23318e-05  
## SDF4 875.7912 0.147221 0.0562756 2.61607 8.89495e-03 3.16996e-02  
## DVL1 1575.9139 -0.299536 0.0891066 -3.36154 7.75080e-04 4.38070e-03  
## MRPL20-AS1 211.7329 0.262184 0.1077302 2.43370 1.49452e-02 4.78653e-02  
## TMEM240 99.9251 -0.441069 0.1526187 -2.89000 3.85239e-03 1.60827e-02  
## SSU72 669.7723 -0.188879 0.0860629 -2.19467 2.81875e-02 7.85586e-02

### Ordering significant by pvalue

### Annotating and exporting results

## 'select()' returned 1:many mapping between keys and columns

### Volcano plot



### Exporting results

# Part 2 : Functional enrichment analysis of the DE genes

### Import libraries

## Loading required package: BiasedUrn

## Loading required package: geneLenDataBase

##   
## Attaching package: 'geneLenDataBase'

## The following object is masked from 'package:S4Vectors':  
##   
## unfactor

## clusterProfiler v4.4.4 For help: https://yulab-smu.top/biomedical-knowledge-mining-book/  
##   
## If you use clusterProfiler in published research, please cite:  
## T Wu, E Hu, S Xu, M Chen, P Guo, Z Dai, T Feng, L Zhou, W Tang, L Zhan, X Fu, S Liu, X Bo, and G Yu. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021, 2(3):100141

##   
## Attaching package: 'clusterProfiler'

## The following object is masked from 'package:AnnotationDbi':  
##   
## select

## The following object is masked from 'package:IRanges':  
##   
## slice

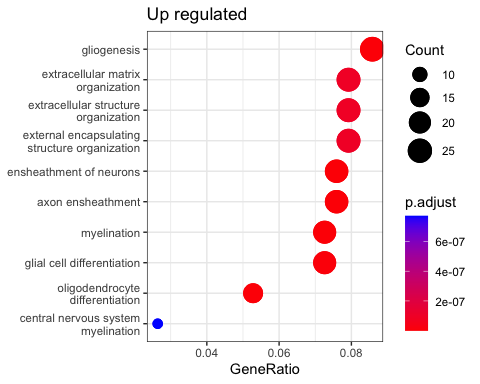
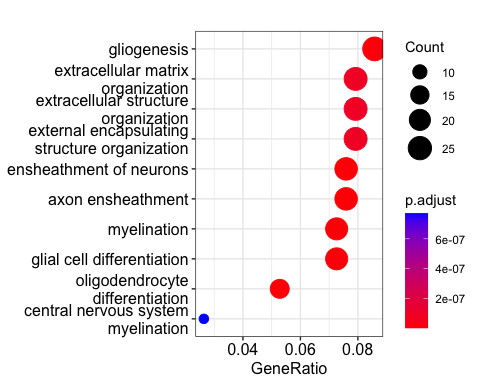
## The following object is masked from 'package:S4Vectors':  
##   
## rename

## The following object is masked from 'package:stats':  
##   
## filter

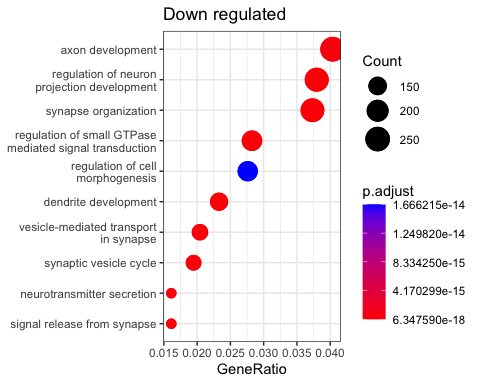
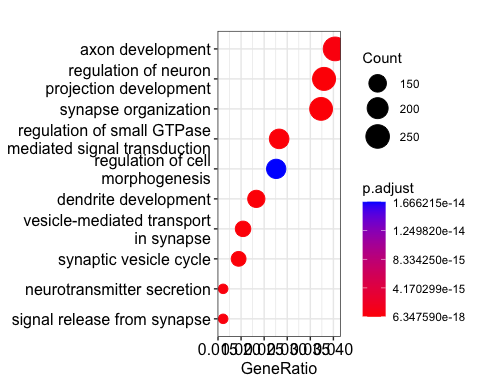
### Enrichment using GO Biological Process (BP)

### Up regulated and Down regulated GO plot

## Scale for colour is already present.  
## Adding another scale for colour, which will replace the existing scale.



## Scale for colour is already present.  
## Adding another scale for colour, which will replace the existing scale.



### Saving plot on png format

## quartz\_off\_screen   
## 2

## quartz\_off\_screen   
## 2

### Exporting GO results