

Proyecto

Andrea Monzalvo

28/2/2021

Contents

Descarga de datos	1
Explorar los atributos	2
Algunos atributos de interés para el análisis	4
Resumen de las variables de interés	5
Filtrado	6
Mujeres	7
Normalización de datos	7
Modelo	8
Expresión diferencial	8
PDF de heatmap	11
Hombres	12
Normalización de datos	13
Expresión diferencial	13
Concordancia	18

Descarga de datos

El set de datos que seleccioné era un perfil de RNA-seq de pacientes con Parkinson

```
library("recount3")

human_projects <- available_projects()

rse_gene_Parkinson <- create_rse(
    subset(
        human_projects,
        project == "ERP014707" & project_type == "data_sources"
    )
)
```

```
)  
assay(rse_gene_Parkinson, "counts") <- compute_read_counts(rse_gene_Parkinson)
```

Explorar los atributos

```
rse_gene_Parkinson$sra.sample_attributes[1:3]
```

```
## [1] "Alias;;E-MTAB-4586:GIS_DaDisease1_RNARep1|Broker name;;ArrayExpress|cell_line;;ND38477|Description  
## [2] "Alias;;E-MTAB-4586:GIS_DaDisease1_RNARep2|Broker name;;ArrayExpress|cell_line;;ND38477|Description  
## [3] "Alias;;E-MTAB-4586:GIS_DaDisease2_RNARep1|Broker name;;ArrayExpress|cell_line;;ND35367|Description
```

Parece que hay un campo de descripción muy largo que sería mejor quitar:

```
rse_gene_Parkinson$sra.sample_attributes <- gsub("\\\\|Description;;Protocols: Total RNAs were extracted ..",  
rse_gene_Parkinson$sra.sample_attributes <- gsub (" From days 14, cells were fed with maturation medium ..",
```

```
rse_gene_Parkinson$sra.sample_attributes[1:2]
```

```
## [1] "Alias;;E-MTAB-4586:GIS_DaDisease1_RNARep1|Broker name;;ArrayExpress|cell_line;;ND38477|disease;  
## [2] "Alias;;E-MTAB-4586:GIS_DaDisease1_RNARep2|Broker name;;ArrayExpress|cell_line;;ND38477|disease;
```

Ver cómo quedan los atributos limpios

```
rse_gene_Parkinson <- expand_sra_attributes(rse_gene_Parkinson)  
  
colData(rse_gene_Parkinson) [  
 ,  
 grep("sra_attribute", colnames(colData(rse_gene_Parkinson)))  
 ]
```

```
## DataFrame with 32 rows and 15 columns  
##           sra_attribute.Alias sra_attribute.Broker_name  
##           <character>          <character>  
## ERR1338001 E-MTAB-4586:GIS_DaDi.. ArrayExpress  
## ERR1338002 E-MTAB-4586:GIS_DaDi.. ArrayExpress  
## ERR1338003 E-MTAB-4586:GIS_DaDi.. ArrayExpress  
## ERR1338004 E-MTAB-4586:GIS_DaDi.. ArrayExpress  
## ERR1338005 E-MTAB-4586:GIS_DaDi.. ArrayExpress  
## ...          ...  
## ERR1338028 E-MTAB-4586:GIS_Da_s.. ArrayExpress  
## ERR1338029 E-MTAB-4586:GIS_Da_s.. ArrayExpress  
## ERR1338030 E-MTAB-4586:GIS_Da_s.. ArrayExpress  
## ERR1338031 E-MTAB-4586:GIS_Da_s.. ArrayExpress  
## ERR1338032 E-MTAB-4586:GIS_Da_s.. ArrayExpress  
##           sra_attribute.cell_line sra_attribute.disease  
##           <character>          <character>
```

```

## ERR1338001      ND38477    Parkinson's disease
## ERR1338002      ND38477    Parkinson's disease
## ERR1338003      ND35367    Parkinson's disease
## ERR1338004      ND35367    Parkinson's disease
## ERR1338005      ND34391    Parkinson's disease
## ...
## ...
## ERR1338028      ND40018    Parkinson's disease
## ERR1338029      GM23280A   Healthy
## ERR1338030      GM23280A   Healthy
## ERR1338031      IR1.7     Healthy
## ERR1338032      IR1.7     Healthy
##           sra_attribute.ENA_checklist sra_attribute.genotype
##           <character>          <character>
## ERR1338001      ERC000011  PARK2_c.255delA
## ERR1338002      ERC000011  PARK2_c.255delA
## ERR1338003      ERC000011  LRRK2_G2019S_homo
## ERR1338004      ERC000011  LRRK2_G2019S_homo
## ERR1338005      ERC000011  SNCA_triplication
## ...
## ...
## ERR1338028      ERC000011  LRRK2_G2019S_homo
## ERR1338029      ERC000011  wild type genotype
## ERR1338030      ERC000011  wild type genotype
## ERR1338031      ERC000011  wild type genotype
## ERR1338032      ERC000011  wild type genotype
##           sra_attribute.INSDC_center_name sra_attribute.INSDC_first_public
##           <character>          <character>
## ERR1338001      GIS        2016-06-07T17:01:18Z
## ERR1338002      GIS        2016-06-07T17:01:18Z
## ERR1338003      GIS        2016-06-07T17:01:18Z
## ERR1338004      GIS        2016-06-07T17:01:18Z
## ERR1338005      GIS        2016-06-07T17:01:18Z
## ...
## ...
## ERR1338028      GIS        2016-06-07T17:01:18Z
## ERR1338029      GIS        2016-06-07T17:01:18Z
## ERR1338030      GIS        2016-06-07T17:01:18Z
## ERR1338031      GIS        2016-06-07T17:01:18Z
## ERR1338032      GIS        2016-06-07T17:01:18Z
##           sra_attribute.INSDC_last_update sra_attribute.INSDC_status
##           <character>          <character>
## ERR1338001      2016-03-24T16:17:20Z    public
## ERR1338002      2016-03-24T16:17:20Z    public
## ERR1338003      2016-03-24T16:17:20Z    public
## ERR1338004      2016-03-24T16:17:20Z    public
## ERR1338005      2016-03-24T16:17:20Z    public
## ...
## ...
## ERR1338028      2016-03-24T16:17:20Z    public
## ERR1338029      2016-03-24T16:17:20Z    public
## ERR1338030      2016-03-24T16:17:20Z    public
## ERR1338031      2016-03-24T16:17:20Z    public
## ERR1338032      2016-03-24T16:17:20Z    public
##           sra_attribute.obsolete_sex sra_attribute.organism
##           <character>          <character>
## ERR1338001      Male       Homo sapiens
## ERR1338002      Male       Homo sapiens

```

```

## ERR1338003           Male      Homo sapiens
## ERR1338004           Male      Homo sapiens
## ERR1338005          Female    Homo sapiens
## ...                  ...
## ERR1338028          Female    Homo sapiens
## ERR1338029          Female    Homo sapiens
## ERR1338030          Female    Homo sapiens
## ERR1338031           Male     Homo sapiens
## ERR1338032           Male     Homo sapiens
##           sra_attribute.Sample_Name sra_attribute.SRA_accession
##                               <character>      <character>
## ERR1338001           ERS1094726   ERS1094726
## ERR1338002           ERS1094727   ERS1094727
## ERR1338003           ERS1094728   ERS1094728
## ERR1338004           ERS1094729   ERS1094729
## ERR1338005           ERS1094730   ERS1094730
## ...                  ...
## ERR1338028           ERS1094721   ERS1094721
## ERR1338029           ERS1094722   ERS1094722
## ERR1338030           ERS1094723   ERS1094723
## ERR1338031           ERS1094724   ERS1094724
## ERR1338032           ERS1094725   ERS1094725
##           sra_attribute.Title
##                               <character>
## ERR1338001 GIS_DaDisease1_RNARep1
## ERR1338002 GIS_DaDisease1_RNARep2
## ERR1338003 GIS_DaDisease2_RNARep1
## ERR1338004 GIS_DaDisease2_RNARep2
## ERR1338005 GIS_DaDisease3_RNARep1
## ...                  ...
## ERR1338028 GIS_Da_splicing_Dise..
## ERR1338029 GIS_Da_splicing_Wt1..
## ERR1338030 GIS_Da_splicing_Wt1..
## ERR1338031 GIS_Da_splicing_Wt2..
## ERR1338032 GIS_Da_splicing_Wt2..

```

Algunos atributos de interés para el análisis

```

sra_attribute.disease sra_attribute.obsolete_sex sra_attribute.genotype sra_attribute.cell_line

#Enfermedad
rse_gene_Parkinson$sra_attribute.disease <- as.factor(rse_gene_Parkinson$sra_attribute.disease)
#Sexo
rse_gene_Parkinson$sra_attribute.obsolete_sex <- factor(rse_gene_Parkinson$sra_attribute.obsolete_sex)
#Genotipo (mutación)
rse_gene_Parkinson$sra_attribute.genotype <- as.factor(rse_gene_Parkinson$sra_attribute.genotype)
#Linea celular
rse_gene_Parkinson$sra_attribute.cell_line <- factor(rse_gene_Parkinson$sra_attribute.cell_line)

#Una vista a los niveles en cada atributo
rse_gene_Parkinson$sra_attribute.disease

## [1] Parkinson's disease Parkinson's disease Parkinson's disease

```

```

## [4] Parkinson's disease Parkinson's disease Parkinson's disease
## [7] Healthy           Healthy           Healthy
## [10] Healthy          Healthy          Healthy
## [13] Parkinson's disease Parkinson's disease Parkinson's disease
## [16] Parkinson's disease Parkinson's disease Healthy
## [19] Healthy          Parkinson's disease Healthy
## [22] Healthy          Healthy          Healthy
## [25] Parkinson's disease Parkinson's disease Parkinson's disease
## [28] Parkinson's disease Healthy         Healthy
## [31] Healthy          Healthy
## Levels: Healthy Parkinson's disease

```

```
rse_gene_Parkinson$sra_attribute.obsolete_sex
```

```

## [1] Male   Male   Male   Male   Female Female Female Female Female
## [11] Male   Male   Male   Male   Male   Female Female Female Female Male
## [21] Female Female Male   Male   Male   Male   Female Female Female Female
## [31] Male   Male
## Levels: Female Male

```

```
rse_gene_Parkinson$sra_attribute.genotype
```

```

## [1] PARK2_c.255delA    PARK2_c.255delA    LRRK2_G2019S_homo  LRRK2_G2019S_homo
## [5] SNCA_triplication SNCA_triplication wild type genotype wild type genotype
## [9] wild type genotype wild type genotype wild type genotype wild type genotype
## [13] PARK2_c.255delA    PARK2_c.255delA    LRRK2_G2019S_homo  SNCA_triplication
## [17] SNCA_triplication wild type genotype wild type genotype LRRK2_G2019S_homo
## [21] wild type genotype wild type genotype wild type genotype wild type genotype
## [25] LRRK2_G2019S_homo  LRRK2_G2019S_homo  LRRK2_G2019S_homo  LRRK2_G2019S_homo
## [29] wild type genotype wild type genotype wild type genotype wild type genotype
## 4 Levels: LRRK2_G2019S_homo PARK2_c.255delA ... wild type genotype

```

```
rse_gene_Parkinson$sra_attribute.cell_line
```

```

## [1] ND38477  ND38477  ND35367  ND35367  ND34391  ND34391  GM23280A GM23280A
## [9] GM23279A GM23279A IR1.7   IR1.7   ND38477  ND38477  ND35367  ND34391
## [17] ND34391  GM23280A GM23280A ND35367  GM23279A GM23279A IR1.7   IR1.7
## [25] ND35367  ND35367  ND40018  ND40018  GM23280A GM23280A IR1.7   IR1.7
## Levels: GM23279A GM23280A IR1.7 ND34391 ND35367 ND38477 ND40018

```

Resumen de las variables de interés

```

# Resumen de las variables de interés
summary(as.data.frame(colData(rse_gene_Parkinson)[
  ,
  grepl("^sra_attribute.[disease|genotype|cell_line|obsolete_sex]", colnames(colData(rse_gene_Parkins
]))])

```

```
## sra_attribute.cell_line           sra_attribute.disease
```

```

## GM23279A:4           Healthy          :16
## GM23280A:6           Parkinson's disease:16
## IR1.7    :6
## ND34391   :4
## ND35367   :6
## ND38477   :4
## ND40018   :2
##           sra_attribute.genotype sra_attribute.obsolete_sex
## LRRK2_G2019S_homo : 8           Female:16
## PARK2_c.255delA   : 4           Male   :16
## SNCA_triplication : 4
## wild type genotype:16
##
##
##
## sra_attribute.organism
## Length:32
## Class :character
## Mode  :character
##
##
##
## $Female
##           Healthy Parkinson's disease
##           10             6
##
## $Male
##           Healthy Parkinson's disease
##           6              10

with(colData(rse_gene_Parkinson), tapply(sra_attribute.disease, sra_attribute.obsolete_sex, summary))

## $Female
##           Healthy Parkinson's disease
##           10             6
##
## $Male
##           Healthy Parkinson's disease
##           6              10

with(colData(rse_gene_Parkinson), tapply(sra_attribute.genotype, sra_attribute.obsolete_sex, summary))

## $Female
##   LRRK2_G2019S_homo  PARK2_c.255delA  SNCA_triplication wild type genotype
##           2             0                 4                  10
##
## $Male
##   LRRK2_G2019S_homo  PARK2_c.255delA  SNCA_triplication wild type genotype
##           6             4                 0                  6

```

Filtrado

```

# Guardar copia de seguridad
rse_gene_Parkinson_unfiltered <- rse_gene_Parkinson

```

En un análisis donde se ovularon todas las muestras independientemente del sexo, se notó que muchos genes con expresión diferencial estaban relacionados con las diferencias sexuales, como espermatogénesis. Por ello se decidió hacer dos análisis independientes, uno con mujeres y otro con hombres.

Mujeres

```
#filtrado muestras
rse_gene_Parkinson_female <- rse_gene_Parkinson[,rse_gene_Parkinson$sra_attribute.obsolete_sex == 'Female']
rse_gene_Parkinson_female$sra_attribute.obsolete_sex <- droplevels(rse_gene_Parkinson_female$sra_attribute)
rse_gene_Parkinson_female$sra_attribute.obsolete_sex

## [1] Female Female Female Female Female Female Female Female Female
## [11] Female Female Female Female Female Female
## Levels: Female

# Calculo de los niveles medios de expresión de los genes en las muestras.

gene_means <- rowMeans(assay(rse_gene_Parkinson_female, "counts"))
summary(gene_means)

##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
##      0.0      0.2      3.2     634.3    155.6 1033111.8

## Eliminamos genes con un promedio de expresión menor que la mediana
rse_gene_Parkinson_female <- rse_gene_Parkinson_female[gene_means > 3.1, ]

## Dimensiones finales
dim(rse_gene_Parkinson_female)

## [1] 32054     16

## Porcentaje de genes que retuvimos
round(nrow(rse_gene_Parkinson_female) / nrow(rse_gene_Parkinson_unfiltered) * 100, 2)

## [1] 50.2
```

Normalización de datos

```
library("edgeR") # BiocManager::install("edgeR", update = FALSE)

## Loading required package: limma

##
## Attaching package: 'limma'

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

```

dge <- DGEList(
  counts = assay(rse_gene_Parkinson_female, "counts"),
  genes = rowData(rse_gene_Parkinson_female)
)
dge <- calcNormFactors(dge)

```

Modelo

```

mod <- model.matrix(~ sra_attribute.disease,
  data = colData(rse_gene_Parkinson_female)
)
colnames(mod)

## [1] "(Intercept)"
## [2] "sra_attribute.diseaseParkinson's disease"

```

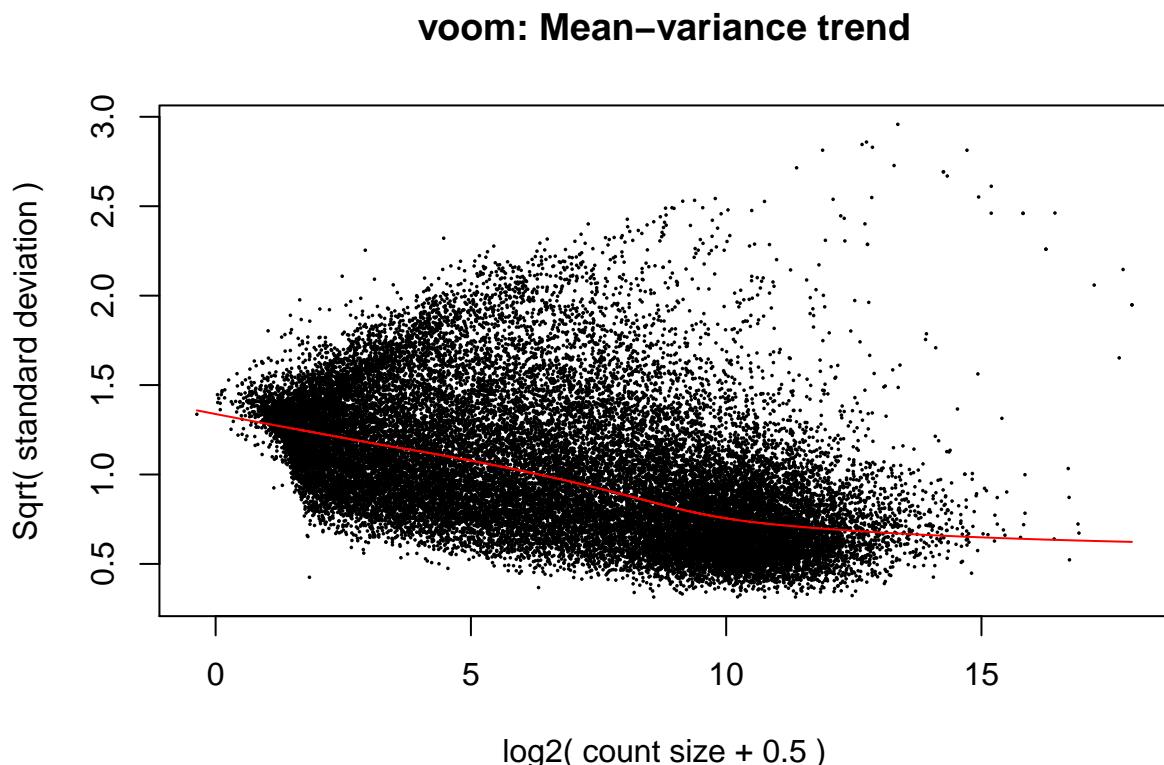
Expresión diferencial

Análisis de expresión diferencial con modelo estadístico.

```

library("limma")
vGene <- voom(dge, mod, plot = TRUE)

```



```

eb_results <- eBayes(lmFit(vGene))

de_results <- topTable(
  eb_results,
  coef = 2,
  number = nrow(rse_gene_Parkinson_female),
  sort.by = 'none'
)
dim(de_results)

## [1] 32054     16

head(de_results)

##           source type score phase      gene_id
## ENSG00000223972.5 HAVANA gene  1735     NA ENSG00000223972.5
## ENSG00000278267.1 ENSEMBL gene    68     NA ENSG00000278267.1
## ENSG00000227232.5 HAVANA gene  1351     NA ENSG00000227232.5
## ENSG00000243485.5 HAVANA gene  1021     NA ENSG00000243485.5
## ENSG00000239945.1 HAVANA gene  1319     NA ENSG00000239945.1
## ENSG00000238009.6 HAVANA gene  3726     NA ENSG00000238009.6
##                      gene_type   gene_name level
## ENSG00000223972.5 transcribed_unprocessed_pseudogene DDX11L1    2
## ENSG00000278267.1                         miRNA MIR6859-1    3
## ENSG00000227232.5 unprocessed_pseudogene WASH7P    2
## ENSG00000243485.5                      lincRNA MIR1302-2HG    2
## ENSG00000239945.1                      lincRNA RP11-34P13.8    2
## ENSG00000238009.6                      lincRNA RP11-34P13.7    2
##           havana_gene      tag      logFC AveExpr
## ENSG00000223972.5 OTTHUMG00000000961.2 <NA> -0.30274947 -1.3546351
## ENSG00000278267.1 <NA> <NA>  0.31891164 -3.7441047
## ENSG00000227232.5 OTTHUMG00000000958.1 <NA>  0.30909153  3.2555417
## ENSG00000243485.5 OTTHUMG00000000959.2 ncRNA_host -0.06299235 -0.8853096
## ENSG00000239945.1 OTTHUMG00000001097.2 overlapping_locus -1.05854989 -3.3959671
## ENSG00000238009.6 OTTHUMG00000001096.2 overlapping_locus -1.06140212 -2.1129587
##           t      P.Value adj.P.Val      B
## ENSG00000223972.5 -0.4579156  0.653058005 0.9174395 -5.775443
## ENSG00000278267.1  0.3671316  0.718243417 0.9329793 -5.637864
## ENSG00000227232.5  1.1806463  0.254684546 0.8985234 -5.760595
## ENSG00000243485.5 -0.1519260  0.881109743 0.9724578 -5.908379
## ENSG00000239945.1 -2.0610624  0.055622494 0.8078493 -3.940885
## ENSG00000238009.6 -3.5529828  0.002584878 0.2613744 -1.421028

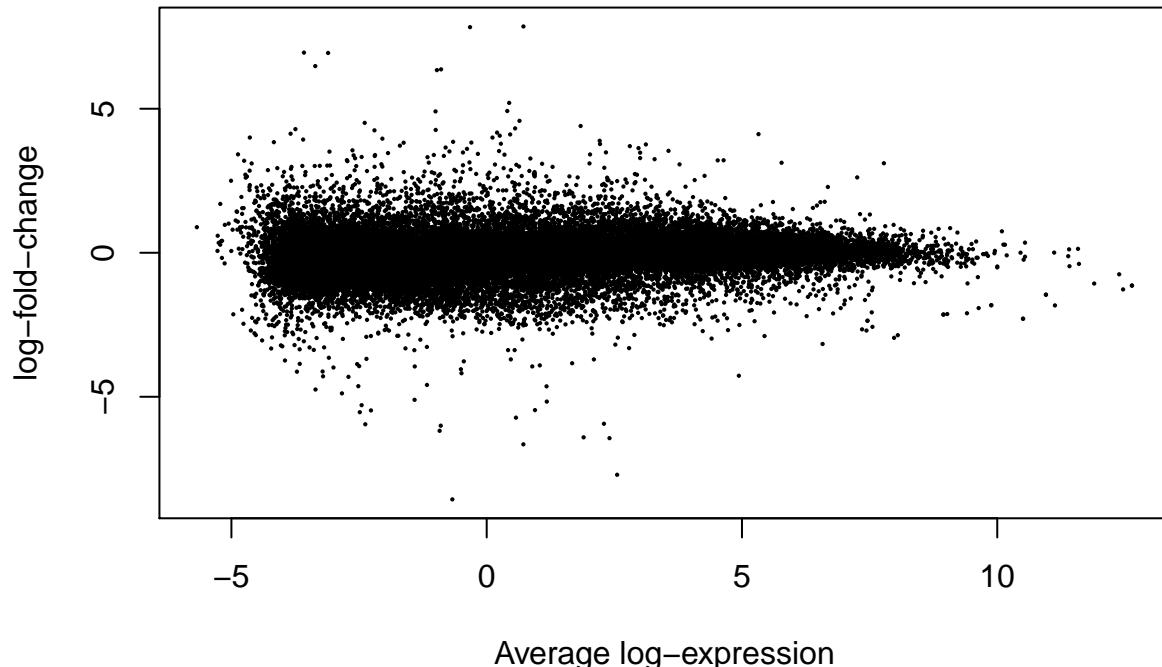
## Genes diferencialmente expresados entre pre y post natal con FDR < 5%
table(de_results$adj.P.Val < 0.05)

## 
## FALSE  TRUE
## 31977    77

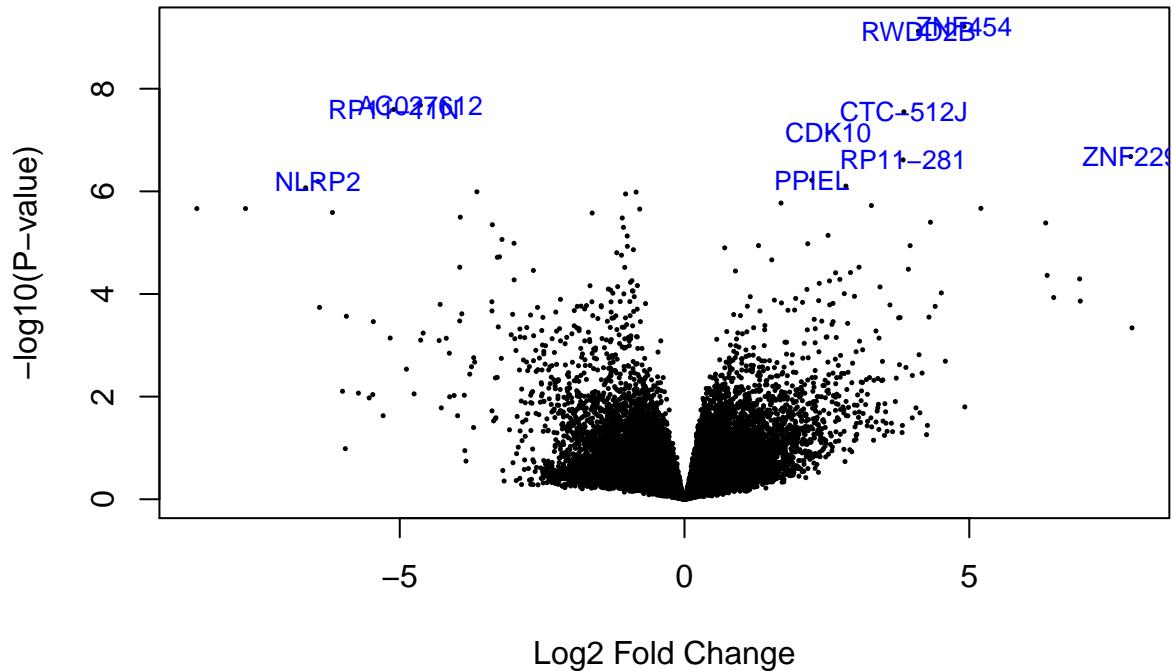
```

```
## Visualicemos los resultados estadísticos
limma::plotMA(eb_results, coef = 2)
```

sra_attribute.diseaseParkinson's disease



```
volcanoplot(eb_results, coef = 2, highlight = 10, names = de_results$gene_name)
```



PDF de heatmap

```

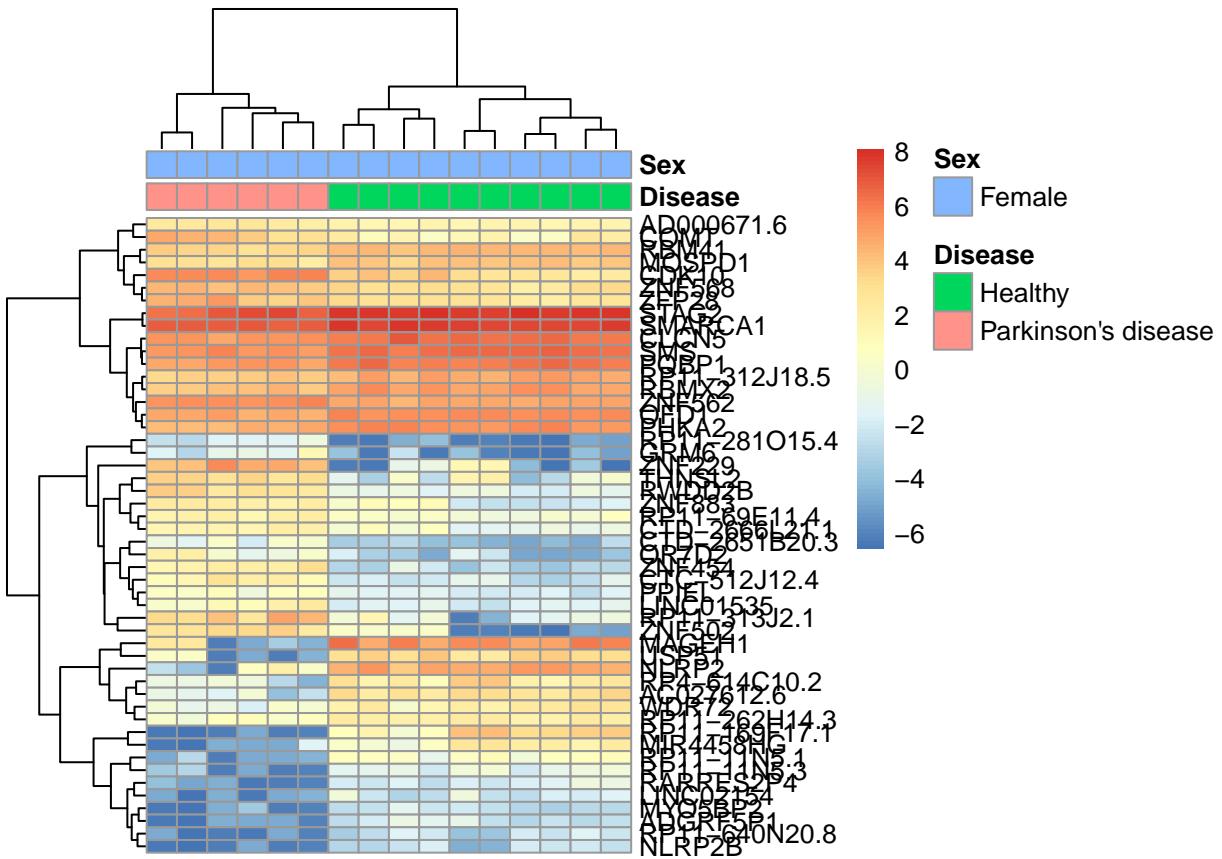
## Extraer valores de los genes de interés
exprs_heatmap <- vGene$E[rank(de_results$adj.P.Val) <= 50, ]

## Creamos una tabla con información de las muestras
## y con nombres de columnas más amigables
df <- as.data.frame(colData(rse_gene_Parkinson_female)[, c("sra_attribute.disease", "sra_attribute.obso")]
colnames(df) <- c("Disease", "Sex")

rownames(exprs_heatmap) <- rowRanges(rse_gene_Parkinson_female)$gene_name[
  match(rownames(exprs_heatmap), rowRanges(rse_gene_Parkinson_female)$gene_id)
]
library("pheatmap")

pheatmap(
  exprs_heatmap,
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  show_rownames = TRUE,
  show_colnames = FALSE,
  annotation_col = df
)

```



```
pdf("pheatmap_Parkinson_female.pdf", height = 14, useDingbats = FALSE)
pheatmap(
```

```
    exprs_heatmap,
    cluster_rows = TRUE,
    cluster_cols = TRUE,
    show_rownames = TRUE,
    show_colnames = FALSE,
    annotation_col = df
)
dev.off()
```

```
## pdf
## 3
```

Hombres

```
#filtrado muestras
rse_gene_Parkinson_male <- rse_gene_Parkinson[,rse_gene_Parkinson$sra_attribute.obsolete_sex == 'Male']
rse_gene_Parkinson_male$sra_attribute.obsolete_sex <- droplevels(rse_gene_Parkinson_male$sra_attribute.
rse_gene_Parkinson_male$sra_attribute.obsolete_sex
```

```
## [1] Male Male
```

```

## [16] Male
## Levels: Male

# Calculo de los niveles medios de expresión de los genes en las muestras.

gene_means <- rowMeans(assay(rse_gene_Parkinson_male, "counts"))
summary(gene_means)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##      0.0     0.2     3.0    632.8   150.6 988953.8

## Eliminamos genes con un promedio de expresión menor que la mediana
rse_gene_Parkinson_male <- rse_gene_Parkinson_male[gene_means > 3.1, ]

## Dimensiones finales
dim(rse_gene_Parkinson_male)

## [1] 31789     16

## Porcentaje de genes que retuvimos
round(nrow(rse_gene_Parkinson_male) / nrow(rse_gene_Parkinson_unfiltered) * 100, 2)

## [1] 49.78

```

Normalización de datos

```

library("edgeR") # BiocManager::install("edgeR", update = FALSE)
dge <- DGEList(
  counts = assay(rse_gene_Parkinson_male, "counts"),
  genes = rowData(rse_gene_Parkinson_male)
)
dge <- calcNormFactors(dge)

mod <- model.matrix(~ sra_attribute.disease,
  data = colData(rse_gene_Parkinson_male)
)
colnames(mod)

## [1] "(Intercept)"
## [2] "sra_attribute.diseaseParkinson's disease"

```

Expresión diferencial

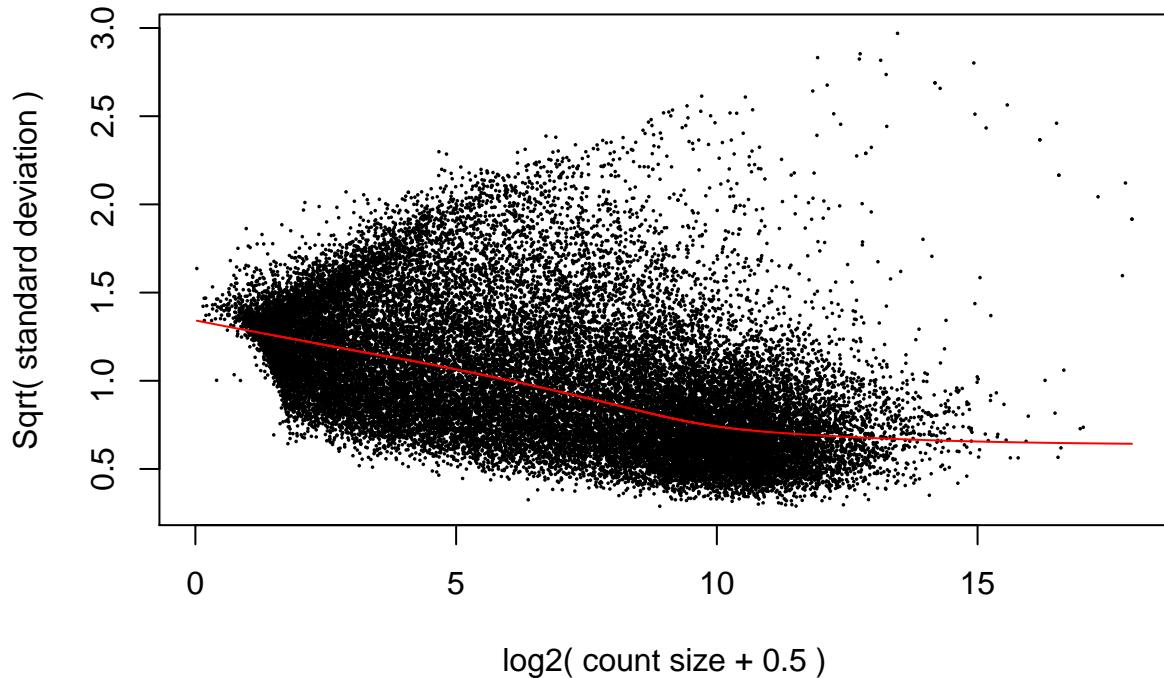
Análisis de expresión diferencial con modelo estadístico.

```

#library("limma")
vGene <- voom(dge, mod, plot = TRUE)

```

voom: Mean–variance trend



```

eb_results <- eBayes(lmFit(vGene))

de_results <- topTable(
  eb_results,
  coef = 2,
  number = nrow(rse_gene_Parkinson_male),
  sort.by = 'none'
)
dim(de_results)

## [1] 31789     16

head(de_results)

##           source type score phase      gene_id
## ENSG00000223972.5 HAVANA gene  1735     NA ENSG00000223972.5
## ENSG00000278267.1 ENSEMBL gene    68     NA ENSG00000278267.1
## ENSG00000227232.5 HAVANA gene  1351     NA ENSG00000227232.5
## ENSG00000243485.5 HAVANA gene  1021     NA ENSG00000243485.5
## ENSG00000238009.6 HAVANA gene  3726     NA ENSG00000238009.6
## ENSG00000233750.3 HAVANA gene  3812     NA ENSG00000233750.3
##                               gene_type   gene_name level
## ENSG00000223972.5 transcribed_unprocessed_pseudogene      DDX11L1     2
## ENSG00000278267.1                               miRNA      MIR6859-1     3
## ENSG00000227232.5 unprocessed_pseudogene      WASH7P     2

```

```

## ENSG00000243485.5          lincRNA  MIR1302-2HG      2
## ENSG00000238009.6          lincRNA  RP11-34P13.7    2
## ENSG00000233750.3          processed_pseudogene CICP27    1
##                               havana_gene      tag      logFC   AveExpr
## ENSG00000223972.5 OTTHUMG00000000961.2 <NA>  0.4586263 -1.4145458
## ENSG00000278267.1           <NA>      <NA> -1.0250940 -4.3294491
## ENSG00000227232.5 OTTHUMG00000000958.1 <NA> -0.1509094  3.1550342
## ENSG00000243485.5 OTTHUMG00000000959.2 ncRNA_host -0.3950490 -0.6766208
## ENSG00000238009.6 OTTHUMG00000001096.2 overlapping_locus -0.6121261 -3.1231353
## ENSG00000233750.3 OTTHUMG00000001257.3 pseudo_consens -0.1982091 -3.4795204
##                               t  P.Value adj.P.Val      B
## ENSG00000223972.5  0.6933553 0.4979414 0.9999047 -5.590545
## ENSG00000278267.1 -1.1786740 0.2556088 0.9999047 -4.999955
## ENSG00000227232.5 -0.6521205 0.5235002 0.9999047 -6.188367
## ENSG00000243485.5 -0.8408133 0.4127330 0.9999047 -5.567253
## ENSG00000238009.6 -0.7856754 0.4434365 0.9999047 -5.413840
## ENSG00000233750.3 -0.2866016 0.7780628 0.9999047 -5.624439

```

```

## Genes diferencialmente expresados entre pre y post natal con FDR < 5%
table(de_results$adj.P.Val < 0.05)

```

```

##
## FALSE  TRUE
## 31740   49

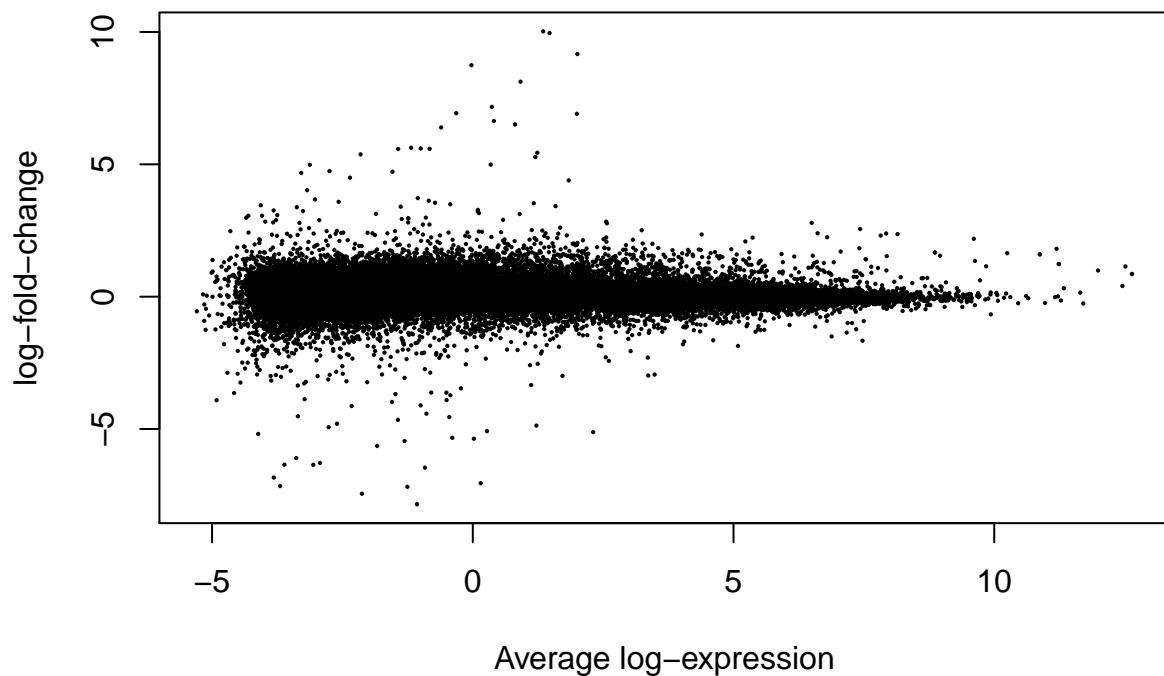
```

```

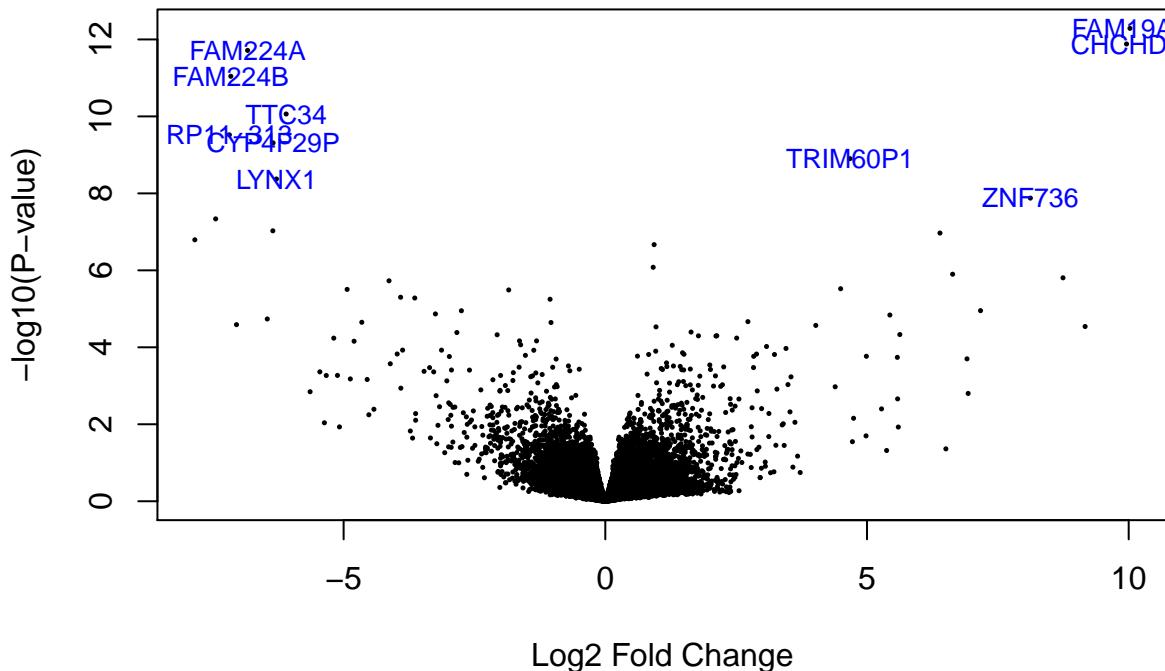
## Visualicemos los resultados estadísticos
limma:::plotMA(eb_results, coef = 2)

```

sra_attribute.diseaseParkinson's disease



```
volcanoplot(eb_results, coef = 2, highlight = 10, names = de_results$gene_name)
```



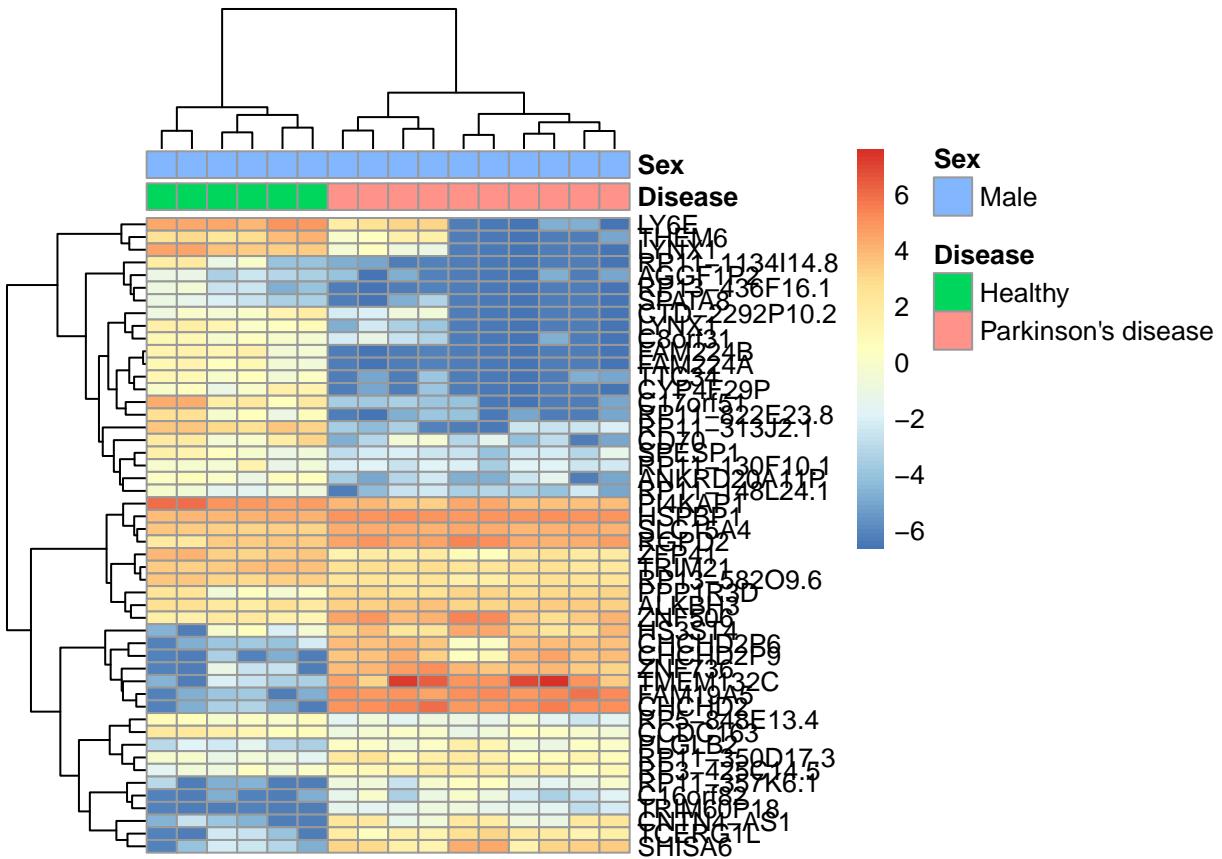
```
### PDF de heatmap
```

```
## Extraer valores de los genes de interés
exprs_heatmap <- vGene$E[rank(de_results$adj.P.Val) <= 50, ]

## Creamos una tabla con información de las muestras
## y con nombres de columnas más amigables
df <- as.data.frame(colData(rse_gene_Parkinson_male)[, c("sra_attribute.disease", "sra_attribute.obsolete")]
colnames(df) <- c("Disease", "Sex")

rownames(exprs_heatmap) <- rowRanges(rse_gene_Parkinson_male)$gene_name[
  match(rownames(exprs_heatmap), rowRanges(rse_gene_Parkinson_male)$gene_id)
]

pheatmap(
  exprs_heatmap,
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  show_rownames = TRUE,
  show_colnames = FALSE,
  annotation_col = df
)
```



```
pdf("pheatmap_Parkinson_Male.pdf", height = 14, useDingbats = FALSE)
pheatmap(
  exprs_heatmap,
```

```
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  show_rownames = TRUE,
  show_colnames = FALSE,
  annotation_col = df
)
dev.off()
```

```
## pdf
## 3
```

Concordancia

Top 50 de los genes con expresión diferencial en mujeres y hombres

```
genes_male<-rownames(exprs_heatmap)
genes_female <- rownames(exprs_heatmap)
genes_female
```

```
## [1] "RP13-436F16.1"   "TTC34"          "CHCHD2P6"        "RP5-848E13.4"
```

```

## [5] "CCDC163"          "RP11-313J2.1"      "TCERG1L"        "AGGF1P2"
## [9] "TRIM21"            "RP11-350D17.3"    "ALKBH3"         "RP11-357K6.1"
## [13] "TMEM132C"          "SLC15A4"          "SPESP1"         "SPATA8"
## [17] "HS3ST4"             "C16orf82"         "SHISA6"         "C17orf51"
## [21] "RP11-822E23.8"     "CD70"              "ZNF506"         "HSPBP1"
## [25] "PLGLB2"             "RGPD2"            "PPP1R3D"        "CYP4F29P"
## [29] "ANKRD20A11P"        "PI4KAP1"          "FAM19A5"        "CNTN4-AS1"
## [33] "RP11-148L24.1"      "RP11-130F10.1"    "RP3-425C14.5"   "CHCHD2"
## [37] "ZNF736"              "TRIM60P18"        "RP11-1134I14.8" "CTD-2292P10.2"
## [41] "THEM6"               "LYNX1"            "LYNX1"          "LY6E"
## [45] "C8orf31"             "ZFP41"            "RP13-58209.6"   "CHCHD2P9"
## [49] "FAM224B"             "FAM224A"          ""               ""

```

genes_male

```

## [1] "RP13-436F16.1"      "TTC34"           "CHCHD2P6"       "RP5-848E13.4"
## [5] "CCDC163"            "RP11-313J2.1"    "TCERG1L"        "AGGF1P2"
## [9] "TRIM21"              "RP11-350D17.3"   "ALKBH3"         "RP11-357K6.1"
## [13] "TMEM132C"           "SLC15A4"          "SPESP1"         "SPATA8"
## [17] "HS3ST4"              "C16orf82"         "SHISA6"         "C17orf51"
## [21] "RP11-822E23.8"       "CD70"              "ZNF506"         "HSPBP1"
## [25] "PLGLB2"              "RGPD2"            "PPP1R3D"        "CYP4F29P"
## [29] "ANKRD20A11P"         "PI4KAP1"          "FAM19A5"        "CNTN4-AS1"
## [33] "RP11-148L24.1"       "RP11-130F10.1"   "RP3-425C14.5"   "CHCHD2"
## [37] "ZNF736"              "TRIM60P18"        "RP11-1134I14.8" "CTD-2292P10.2"
## [41] "THEM6"                "LYNX1"            "LYNX1"          "LY6E"
## [45] "C8orf31"              "ZFP41"            "RP13-58209.6"   "CHCHD2P9"
## [49] "FAM224B"              "FAM224A"          ""               ""

```

Al parecer el gen RP11-313J2.1 se comparte entre ambos grupos

```
gene_pos<- match("RP11-313J2.1",rowRanges(rse_gene_Parkinson)$gene_name)
```

```
boxplot(with(colData(rse_gene_Parkinson),tapply(assay(rse_gene_Parkinson[gene_pos,], 'counts'),sra_attri
```

Expresión de RP11-313J2.1

