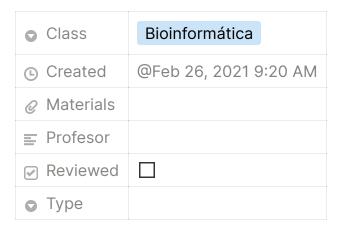
Revisión y SpatialLIBD



- ¿Debemos explorar las relaciones entre nuestras variables?
 Sí, es muy importante
- ¿Por qué usamos el paquete edger?

Toma en cuenta la suma de los niveles de expresión de todos los genes de todas las muestras y las compara. Normalizando de acuerdo a dicha comparación.

- ¿Por qué es importante el argumento sort.by en topTable()?

 Permite reordenar los datos de acuerdo a cierto parámetro
- ¿Por qué es importante el argumento coef en toptable()?

 Es el índice de la categoría que queremos evaluar. Este se encuentra en model.matrix

Ejercicio

```
speaqeasy_data <- file.path(tempdir(), "rse_speaqeasy.RData")
download.file("https://github.com/LieberInstitute/SPEAQeasy-example/blob/master/rse_s
peaqeasy.RData?raw=true", speaqeasy_data, mode = "wb")
library("SummarizedExperiment")</pre>
```

Revisión y SpatialLIBD 1

```
load(speaqeasy_data, verbose = TRUE)
rse_gene
```

```
## Loading objects:
## rse_gene

## class: RangedSummarizedExperiment
## dim: 60609 40
## metadata(0):
## assays(1): counts
## rownames(60609): ENSG000000223972.5 ENSG000000227232.5 ... ENSG000000210195.2 ENSG000
00210196.2
## rowData names(10): Length gencodeID ... NumTx gencodeTx
## colnames(40): R13896_H7JKMBBXX R13903_HCTYLBBXX ... R15120_HFY2MBBXX R15134_HFFGHB
BXX
## colData names(67): SAMPLE_ID FQCbasicStats ... AgeDeath BrNum
```

- ¿Cuántos genes y muestras tenemos en estos datos?
- ¿Hay diferencias en totalAssignedGene O mitoRate entre los grupos de diagnosis (PrimaryDx)?
- Grafica la expresión de SNAP25 para cada grupo de diagnosis.
- Sugiere un modelo estadistico que podríamos usar en una análisis de expresión diferencial. Verifica que si sea un modelo full rank. ¿Cúal sería el o los coeficientes de interés?

full rank es que no todas dependen linealmente

```
### ¿Hay diferencias en totalAssignedGene o mitoRate entre los grupos de diagnosis (P
rimaryDx)?
```{r}
summary(rse_gene)
rse_gene
.``
```{r}
assay(rse_gene, "counts") <- compute_read_counts(rse_gene)</pre>
```

Revisión y SpatialLIBD 2

```
rse_gene$sra.sample_attributes
 ```{r}
\tt df_rse_gene < -data.frame (rse_gene \$total Assigned Gene, rse_gene \$mitoRate, rse_gene \$prima) = (rse_gene \$prima) = (rse_
df_rse_gene
 ```{r}
 library(ggplot2)
 ggplot(df_rse_gene, aes(y=rse_gene.totalAssignedGene,x=rse_gene.PrimaryDx,group=rse_g
ene.PrimaryDx)) + geom_boxplot()
Sí existe expresión
Grafica la expresión de SNAP25 para cada grupo de diagnosis.
geneIndex <- which(rowRanges(rse_gene)$Symbol=="SNAP25")</pre>
boxplot(assay(rse_gene)[geneIndex,]~as.vector(colData(rse_gene)$PrimaryDx), xlab = "D
x", ylab = "Expresión", main = "Expresion de SNAP25 según Dx")
Sugiere un modelo estadistico que podríamos usar en una análisis de expresión diferen
cial. Verifica que si sea un modelo full rank. ¿Cúal sería el o los coeficientes de i
nterés?
```{r}
df2 <- data.frame(sex = colData(rse_gene)$Sex, prim = colData(rse_gene)$PrimaryDx)</pre>
vd <- ExploreModelMatrix::VisualizeDesign(</pre>
 sampleData = df2,
 designFormula = ~ sex + prim,
 textSizeFitted = 4
```

Revisión y SpatialLIBD 3