

Análisis de datos con recount3

Ángel Román Zamora López

2025-02-12

Carga de paquetes

```
library("recount3")

## Cargando paquete requerido: SummarizedExperiment

## Cargando paquete requerido: MatrixGenerics

## Warning: package 'MatrixGenerics' was built under R version 4.4.2

## Cargando paquete requerido: matrixStats

## Warning: package 'matrixStats' was built under R version 4.4.2

##
## Adjuntando el paquete: 'MatrixGenerics'

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

## Cargando paquete requerido: GenomicRanges

## Cargando paquete requerido: stats4
```

```

## Cargando paquete requerido: BiocGenerics

##
## Adjuntando el paquete: 'BiocGenerics'

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

## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, 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, saveRDS, setdiff,
##     table, tapply, union, unique, unsplit, which.max, which.min

## Cargando paquete requerido: S4Vectors

##
## Adjuntando el paquete: 'S4Vectors'

## The following object is masked from 'package:utils':
##
##     findMatches

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

## Cargando paquete requerido: IRanges

## Warning: package 'IRanges' was built under R version 4.4.2

##
## Adjuntando el paquete: 'IRanges'

## The following object is masked from 'package:grDevices':
##
##     windows

## Cargando paquete requerido: GenomeInfoDb

## Warning: package 'GenomeInfoDb' was built under R version 4.4.2

## Cargando paquete requerido: Biobase

```

```

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

##
## Adjuntando el paquete: 'Biobase'

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

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

library("edgeR")

## Warning: package 'edgeR' was built under R version 4.4.2

## Cargando paquete requerido: limma

## Warning: package 'limma' was built under R version 4.4.2

##
## Adjuntando el paquete: 'limma'

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

library("ggplot2")

## Warning: package 'ggplot2' was built under R version 4.4.2

library("pheatmap")

## Warning: package 'pheatmap' was built under R version 4.4.2

library("limma")
library("RColorBrewer")
library("ComplexHeatmap")

## Cargando paquete requerido: grid

```

```

## =====
## ComplexHeatmap version 2.22.0
## Bioconductor page: http://bioconductor.org/packages/ComplexHeatmap/
## Github page: https://github.com/jokergoo/ComplexHeatmap
## Documentation: http://jokergoo.github.io/ComplexHeatmap-reference
##
## If you use it in published research, please cite either one:
## - Gu, Z. Complex Heatmap Visualization. iMeta 2022.
## - Gu, Z. Complex heatmaps reveal patterns and correlations in multidimensional
##   genomic data. Bioinformatics 2016.
##
##
## The new InteractiveComplexHeatmap package can directly export static
## complex heatmaps into an interactive Shiny app with zero effort. Have a try!
##
## This message can be suppressed by:
## suppressPackageStartupMessages(library(ComplexHeatmap))
## =====
## ! pheatmap() has been masked by ComplexHeatmap::pheatmap(). Most of the arguments
##   in the original pheatmap() are identically supported in the new function. You
##   can still use the original function by explicitly calling pheatmap::pheatmap().

##
## Adjuntando el paquete: 'ComplexHeatmap'

## The following object is masked from 'package:pheatmap':
##   pheatmap

```

Exploración de proyectos en recount3

```
mouse_projects <- available_projects(organism = "mouse")
```

```
## 2025-02-23 17:00:17.388363 caching file sra.recount_project.MD.gz.
```

Selección del proyecto de interés

```
proj_info <- subset(
  mouse_projects,
  project == "SRP151148" & project_type == "data_sources"
)
```

Creación del objeto RangedSummarizedExperiment (RSE)

```
rse_gene_SRP151148 <- create_rse(proj_info)
```

```
## 2025-02-23 17:00:28.789716 downloading and reading the metadata.
```

```

## 2025-02-23 17:00:29.783549 caching file sra.sra.SRP151148.MD.gz.

## 2025-02-23 17:00:30.728428 caching file sra.recount_project.SRP151148.MD.gz.

## 2025-02-23 17:00:31.597146 caching file sra.recount_qc.SRP151148.MD.gz.

## 2025-02-23 17:00:32.622198 caching file sra.recount_seq_qc.SRP151148.MD.gz.

## 2025-02-23 17:00:33.602881 caching file sra.recount_pred.SRP151148.MD.gz.

## 2025-02-23 17:00:34.030008 downloading and reading the feature information.

## 2025-02-23 17:00:34.834475 caching file mouse.gene_sums.M023.gtf.gz.

## 2025-02-23 17:00:36.844122 downloading and reading the counts: 500 samples across 55421 features.

## 2025-02-23 17:00:37.536611 caching file sra.gene_sums.SRP151148.M023.gz.

## Warning in grep(pattern, bfr, value = TRUE): unable to translate ' El n<a3>mero
## de serie del volumen es: 2296-ODDF' to a wide string

## Warning in grep(pattern, bfr, value = TRUE): input string 2 is invalid

## Warning in grep(pattern, bfr, value = TRUE): unable to translate ' El n<a3>mero
## de serie del volumen es: 2296-ODDF' to a wide string

## Warning in grep(pattern, bfr, value = TRUE): input string 2 is invalid

## Warning in grep(pattern, bfr, value = TRUE): unable to translate ' El n<a3>mero
## de serie del volumen es: 2296-ODDF' to a wide string

## Warning in grep(pattern, bfr, value = TRUE): input string 2 is invalid

## 2025-02-23 17:00:47.862567 constructing the RangedSummarizedExperiment (rse) object.

assay(rse_gene_SRP151148, "counts") <- compute_read_counts(rse_gene_SRP151148)

```

Expansión de atributos y revisión de datos

```

rse_gene_SRP151148 <- expand_sra_attributes(rse_gene_SRP151148)
colData(rse_gene_SRP151148)[, grep("sra_attribute", colnames(colData(rse_gene_SRP151148)))]

```

```

## DataFrame with 500 rows and 8 columns
##           sra_attribute.antibodies sra_attribute.cell_type_short
##                           <character>                         <character>
## SRR7414204    SiglecF+IL5raIntSSCHi                      EoSSCHi
## SRR7414242    SiglecF+IL5raIntSSCLO                      EoSSCLO
## SRR7414251    SiglecF+IL5raIntSSCHi                      EoSSCHi
## SRR7414342    PI:- CD4:- CD8:- CD1..                      EoSSCLO
## SRR7413874    Lin:- cKit:Lo Sca1:L..                      CLP
## ...
## SRR7413973    Lin:- cKit:Lo Sca1:L..                      CLP
## SRR7413974    IL-5:Hi Ly6G:Hi Sigl..                     NeutPB
## SRR7413975    IL-5:Hi Ly6G:Hi Sigl..                     NeutPB
## SRR7413976    IL-5:Hi Ly6G:Hi Sigl..                     NeutPB
## SRR7413977    IL-5:Hi Ly6G:Hi Sigl..                     NeutPB
##           sra_attribute.cell_type sra_attribute.in_vitro_treatment
##                           <character>                         <character>
## SRR7414204    Eosinophil High Side..                   none
## SRR7414242    Eosinophil Low Side ..                  none
## SRR7414251    Eosinophil High Side..                   none
## SRR7414342    Eosinophil Low Side ..                  none
## SRR7413874    Common Lymphoid Prog..                 none
## ...
## SRR7413973    Common Lymphoid Prog..                 none
## SRR7413974    Neutrophil Periphera..                none
## SRR7413975    Neutrophil Periphera..                none
## SRR7413976    Neutrophil Periphera..                none
## SRR7413977    Neutrophil Periphera..                none
##           sra_attribute.lineage sra_attribute.source_name
##                           <character>                         <character>
## SRR7414204    Eosinophil Lineage      Eosinophil High Side..
## SRR7414242    Eosinophil Lineage      Eosinophil Low Side ..
## SRR7414251    Eosinophil Lineage      Eosinophil High Side..
## SRR7414342    Eosinophil Lineage      Eosinophil Low Side ..
## SRR7413874    Restricted Potential..  Common Lymphoid Prog..
## ...
## SRR7413973    Restricted Potential..  Common Lymphoid Prog..
## SRR7413974    Neutrophil Lineage     Neutrophil Periphera..
## SRR7413975    Neutrophil Lineage     Neutrophil Periphera..
## SRR7413976    Neutrophil Lineage     Neutrophil Periphera..
## SRR7413977    Neutrophil Lineage     Neutrophil Periphera..
##           sra_attribute.strain/background sra_attribute.tissue
##                           <character>                         <character>
## SRR7414204        C57BL/6          Bone Marrow
## SRR7414242        C57BL/6          Bone Marrow
## SRR7414251        C57BL/6          Bone Marrow
## SRR7414342        C57BL/6          Bone Marrow
## SRR7413874        C57BL/6          Bone Marrow
## ...
## SRR7413973        C57BL/6          Bone Marrow
## SRR7413974        C57BL/6          Peripheral Blood
## SRR7413975        C57BL/6          Peripheral Blood
## SRR7413976        C57BL/6          Peripheral Blood
## SRR7413977        C57BL/6          Peripheral Blood

```

Formateo de variables de interés

```
rse_gene_SRP151148$sra_attribute.in_vitro_treatment <- factor(tolower(rse_gene_SRP151148$sra_attribute.in_vitro_treatment))
rse_gene_SRP151148$sra_attribute.cell_type_short <- factor(tolower(rse_gene_SRP151148$sra_attribute.cell_type_short))
rse_gene_SRP151148$sra_attribute.tissue <- factor(tolower(rse_gene_SRP151148$sra_attribute.tissue))
```

Filtrado de muestras

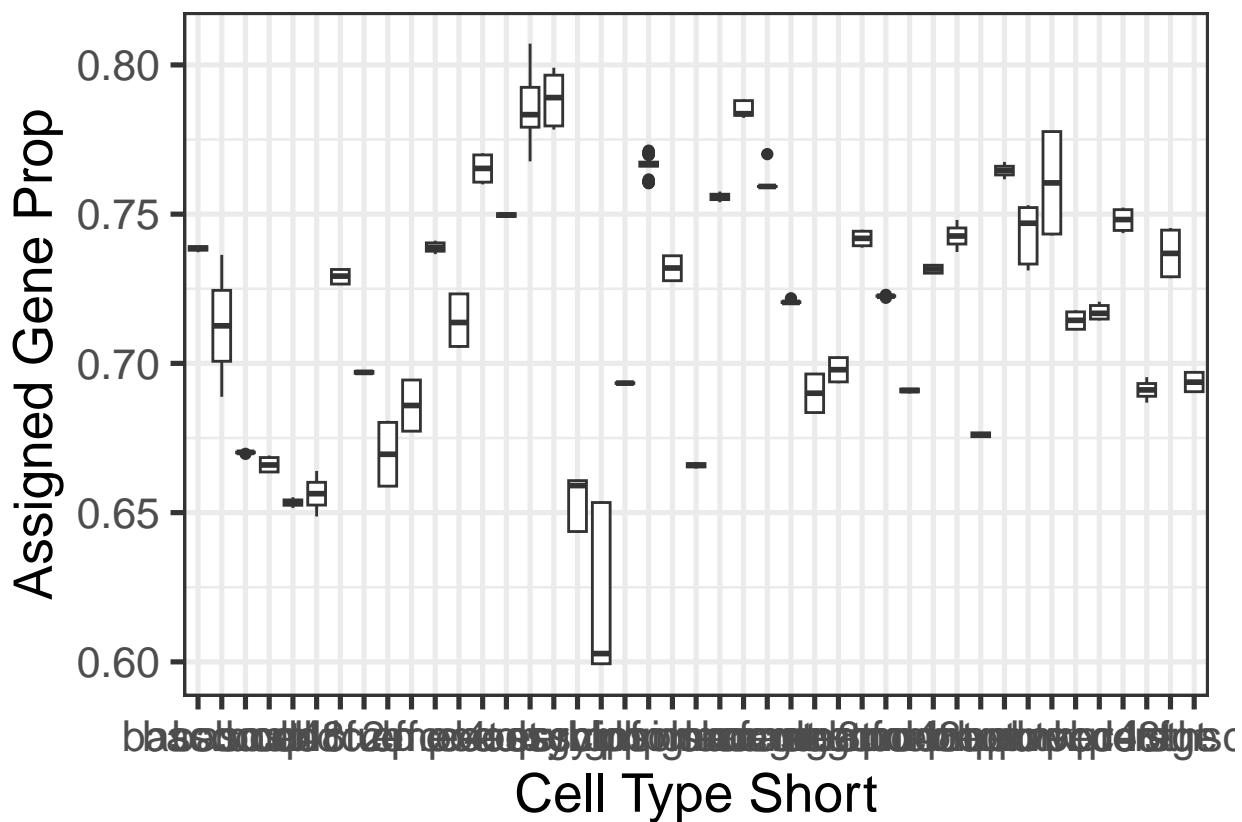
```
rse_gene_SRP151148$assigned_gene_prop <- rse_gene_SRP151148$recount_qc.gene_fc_count_all.assigned / rse_gene_SRP151148[, rse_gene_SRP151148$assigned_gene_prop > 0.5]
gene_means <- rowMeans(assay(rse_gene_SRP151148, "counts"))
rse_gene_SRP151148 <- rse_gene_SRP151148[gene_means > 0.1, ]
```

Normalización de datos

```
dge <- DGEList(counts = assay(rse_gene_SRP151148, "counts"), genes = rowData(rse_gene_SRP151148))
dge <- calcNormFactors(dge)
```

Exploración de la distribución de datos

```
ggplot(as.data.frame(colData(rse_gene_SRP151148)), aes(y = assigned_gene_prop, x = sra_attribute.cell_type_short))
  geom_boxplot() +
  theme_bw(base_size = 20) +
  ylab("Assigned Gene Prop") +
  xlab("Cell Type Short")
```



Modelado estadístico

```

mod <- model.matrix(~ sra_attribute.cell_type_short + sra_attribute.in_vitro_treatment +
                     rse_gene_SRP151148$sra_attribute.tissue + assigned_gene_prop,
                     data = colData(rse_gene_SRP151148))
colnames(mod)

## [1] "(Intercept)"
## [2] "sra_attribute.cell_type_shortbasocult"
## [3] "sra_attribute.cell_type_shortbasopl"
## [4] "sra_attribute.cell_type_shortbcell"
## [5] "sra_attribute.cell_type_shortcd4t"
## [6] "sra_attribute.cell_type_shortcd8t"
## [7] "sra_attribute.cell_type_shortcdc2"
## [8] "sra_attribute.cell_type_shortcfue"
## [9] "sra_attribute.cell_type_shortclp"
## [10] "sra_attribute.cell_type_shortcmp"
## [11] "sra_attribute.cell_type_shorteffcd4t"
## [12] "sra_attribute.cell_type_shorteo"
## [13] "sra_attribute.cell_type_shorteocult"
## [14] "sra_attribute.cell_type_shorteoop"
## [15] "sra_attribute.cell_type_shorteoosschi"
## [16] "sra_attribute.cell_type_shorteoossclo"
## [17] "sra_attribute.cell_type_shorteryblpb"

```

```

## [18] "sra_attribute.cell_type_shorteryblpo"
## [19] "sra_attribute.cell_type_shortgmp"
## [20] "sra_attribute.cell_type_shortgmmpsiclecf"
## [21] "sra_attribute.cell_type_shortinfmono"
## [22] "sra_attribute.cell_type_shortlsk"
## [23] "sra_attribute.cell_type_shortmac"
## [24] "sra_attribute.cell_type_shortmaccult"
## [25] "sra_attribute.cell_type_shortmast"
## [26] "sra_attribute.cell_type_shortmegcd8t"
## [27] "sra_attribute.cell_type_shortmegscf"
## [28] "sra_attribute.cell_type_shortmegtpo"
## [29] "sra_attribute.cell_type_shortmemcd4t"
## [30] "sra_attribute.cell_type_shortmemcd8t"
## [31] "sra_attribute.cell_type_shortmep"
## [32] "sra_attribute.cell_type_shortmonobm"
## [33] "sra_attribute.cell_type_shortmonopb"
## [34] "sra_attribute.cell_type_shortmpp"
## [35] "sra_attribute.cell_type_shortneutbm"
## [36] "sra_attribute.cell_type_shortneutpb"
## [37] "sra_attribute.cell_type_shortnk"
## [38] "sra_attribute.cell_type_shortnvecd4t"
## [39] "sra_attribute.cell_type_shortnvecd8t"
## [40] "sra_attribute.cell_type_shortpdc"
## [41] "sra_attribute.cell_type_shortprecfue"
## [42] "sra_attribute.cell_type_shortregt"
## [43] "sra_attribute.cell_type_shortretic"
## [44] "sra_attribute.cell_type_shortsthsc"
## [45] "sra_attribute.in_vitro_treatmentday 10 bm culture with scf + il-3"
## [46] "sra_attribute.in_vitro_treatmentday 3 bm culture with scf"
## [47] "sra_attribute.in_vitro_treatmentday 3 bm culture with tpo"
## [48] "sra_attribute.in_vitro_treatmentday5 cultured m-csf fetal liver"
## [49] "sra_attribute.in_vitro_treatmentnone"
## [50] "rse_gene_SRP151148$sra_attribute.tissuebone marrow (crush)"
## [51] "rse_gene_SRP151148$sra_attribute.tissueembryonic day 10.5 fetal liver"
## [52] "rse_gene_SRP151148$sra_attribute.tissueperipheral blood"
## [53] "rse_gene_SRP151148$sra_attribute.tissueperitoneal cavity lavage"
## [54] "rse_gene_SRP151148$sra_attribute.tissuespleen"
## [55] "rse_gene_SRP151148$sra_attribute.tissuespleen and lymph nodes"
## [56] "assigned_gene_prop"

```

Análisis diferencial de expresión con limma

```
vGene <- voom(dge, mod, plot = TRUE)
```

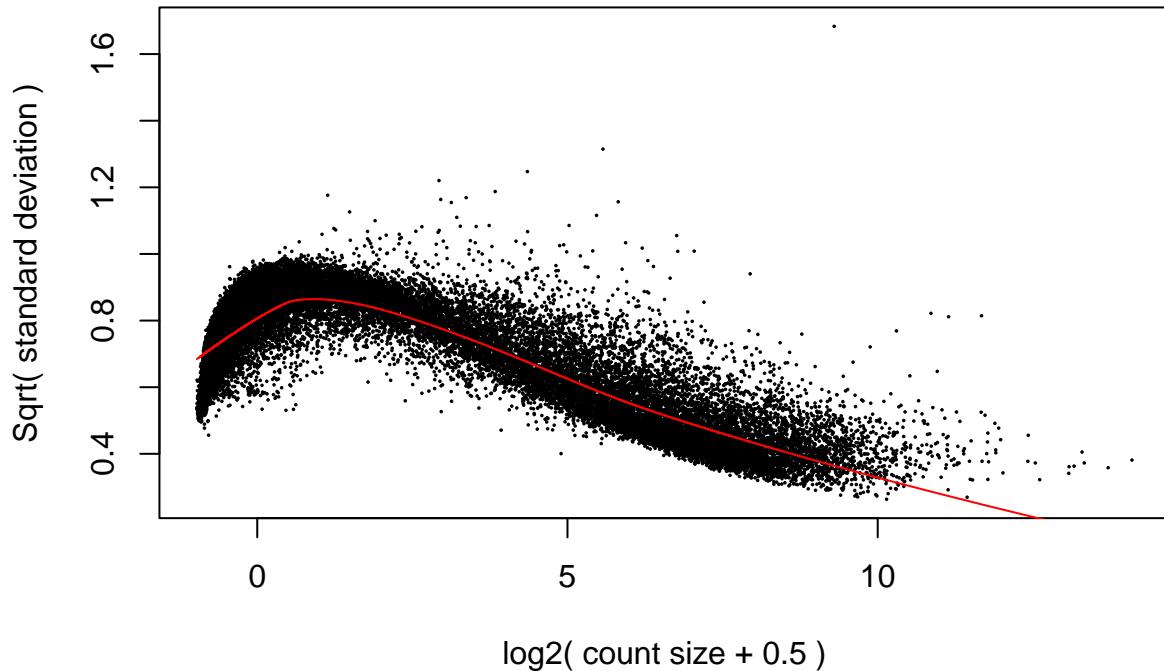
```

## Coefficients not estimable: sra_attribute.cell_type_shortretic sra_attribute.in_vitro_treatmentday 10.5 fetal liver

## Warning: Partial NA coefficients for 27462 probe(s)

```

voom: Mean–variance trend



```
eb_results <- eBayes(lmFit(vGene))

## Coefficients not estimable: sra_attribute.cell_type_shortretic sra_attribute.in_vitro_treatmentday 10

## Warning: Partial NA coefficients for 27462 probe(s)

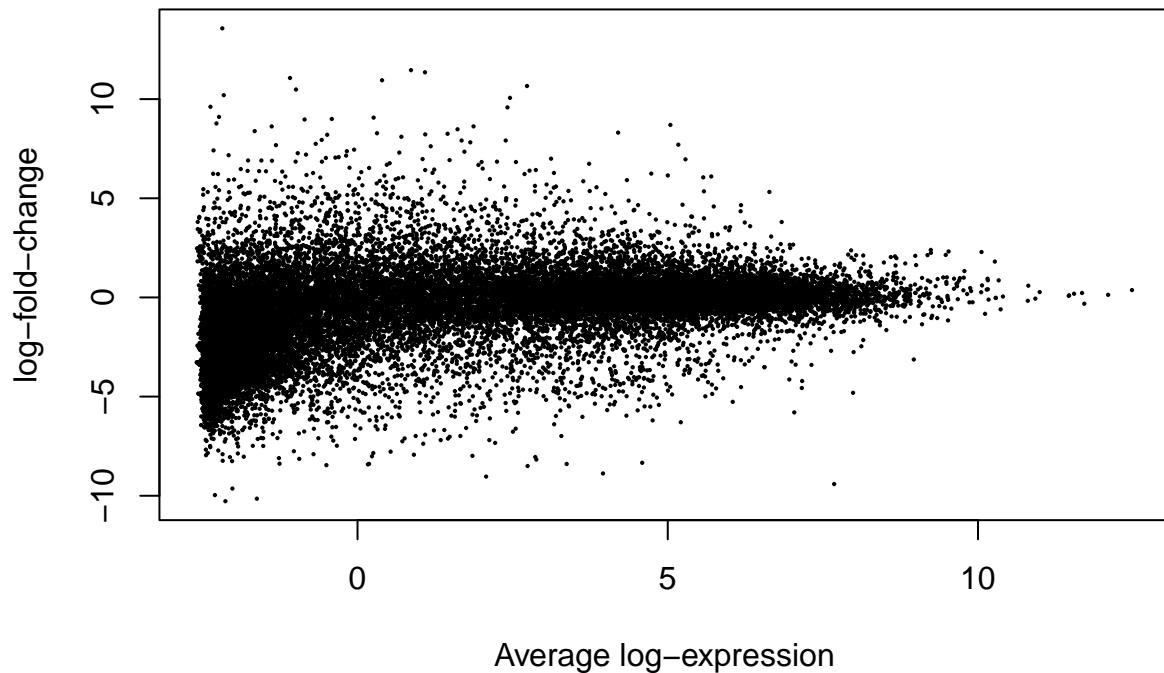
## Warning in .ebayes(fit = fit, proportion = proportion, stdev.coef.lim =
## stdev.coef.lim, : Estimation of var.prior failed - set to default value

de_results <- topTable(eb_results, coef = 2, number = nrow(rse_gene_SRP151148), sort.by = "none")
```

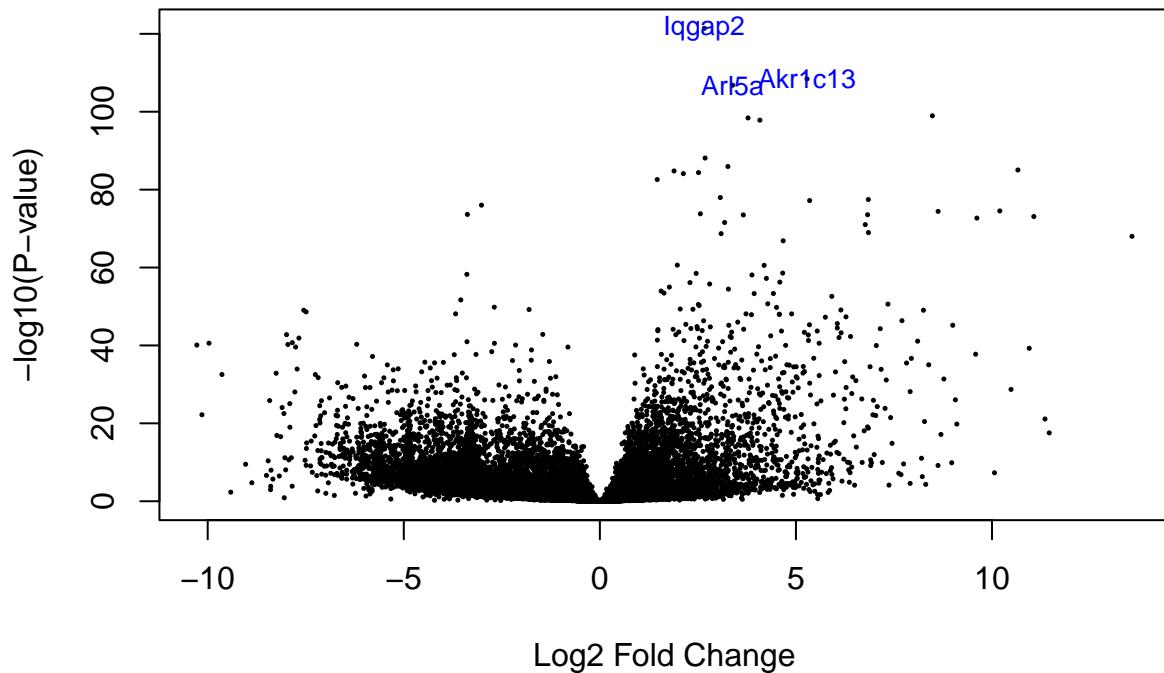
Visualización de resultados

```
plotMA(eb_results, coef = 2)
```

sra_attribute.cell_type_shortbasocult



```
volcanoplot(eb_results, coef = 2, highlight = 3, names = de_results$gene_name)
```

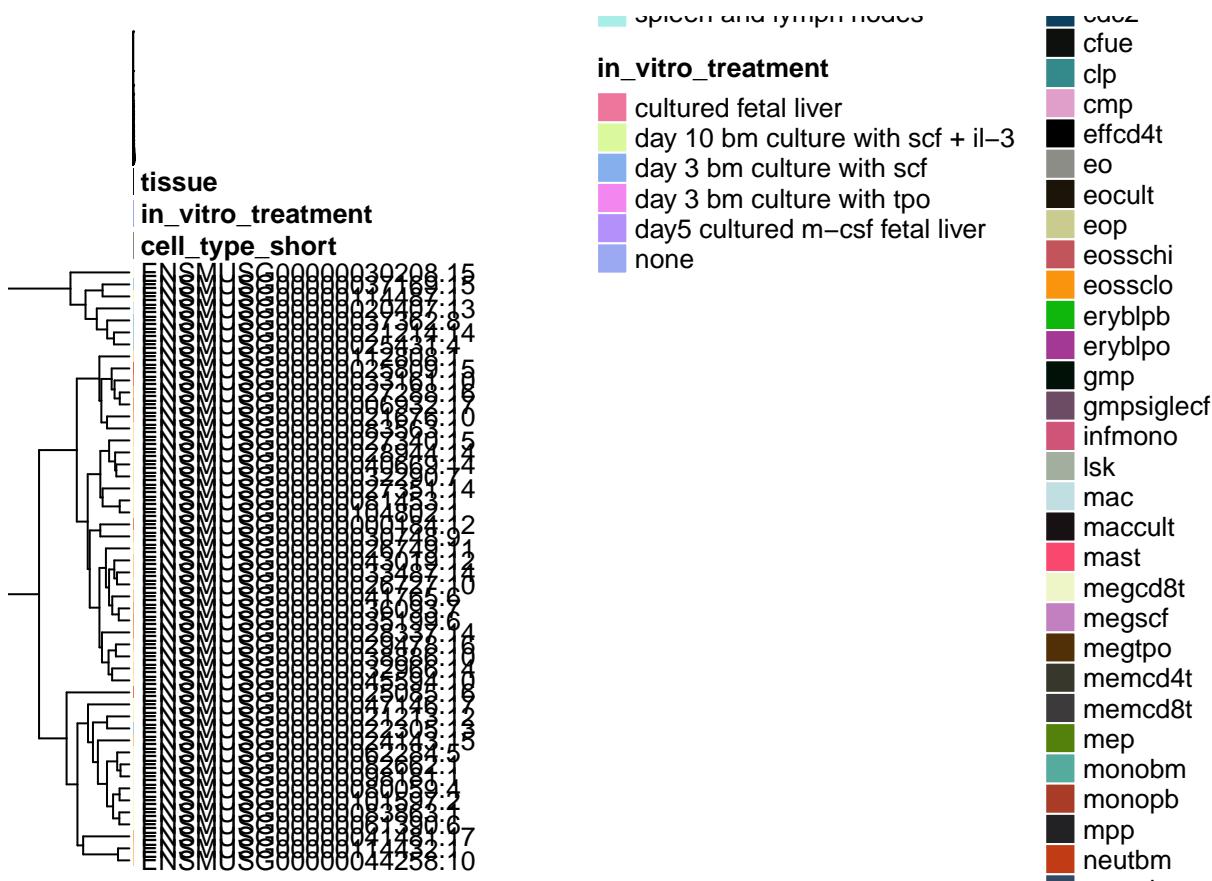


Heatmap de genes más significativos

```

exprs_heatmap <- vGene$E[rank(de_results$adj.P.Val) <= 50, ]
df <- as.data.frame(colData(rse_gene_SRP151148)[, c("sra_attribute.cell_type_short", "sra_attribute.in_")]
colnames(df) <- c("cell_type_short", "in_vitro_treatment", "tissue")
pheatmap(exprs_heatmap, cluster_rows = TRUE, cluster_cols = TRUE, show_rownames = TRUE, show_colnames =

```

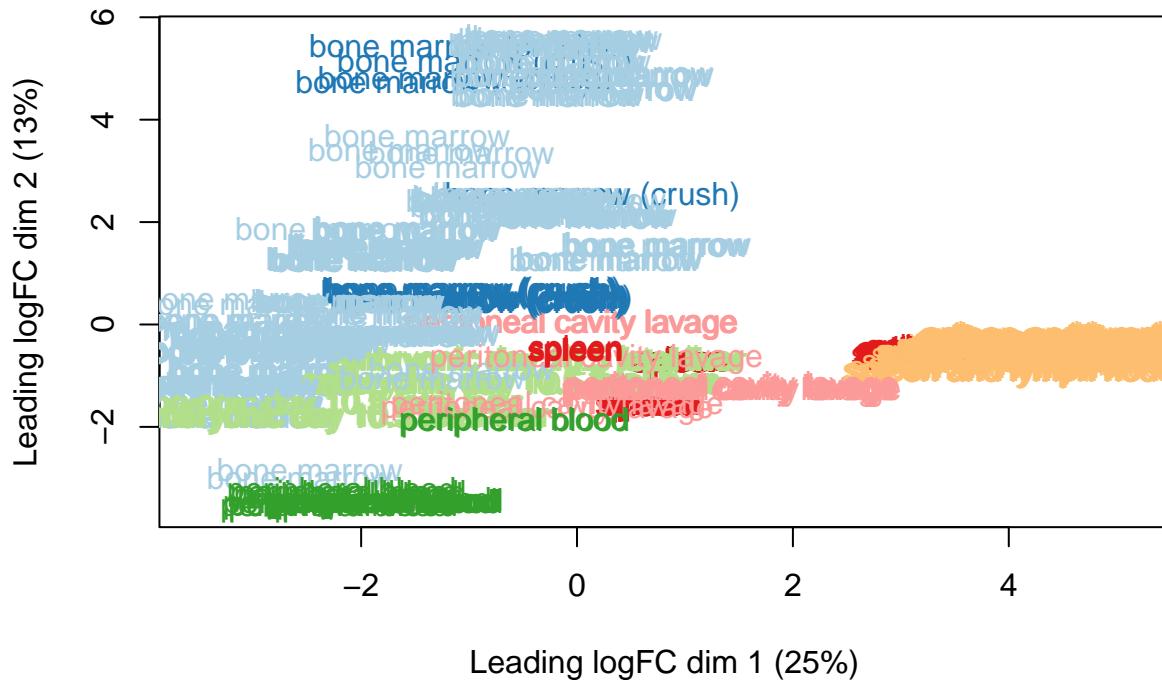


Análisis MDS

```

col.tissue <- df$tissue
levels(col.tissue) <- brewer.pal(nlevels(col.tissue), "Paired")
col.tissue <- as.character(col.tissue)
plotMDS(vGene$E, labels = df$tissue, col = col.tissue)

```



Guardado de resultados

```
pdf("pheatmap_con_nombres.pdf", height = 14, useDingbats = FALSE)
pheatmap(exprs_heatmap, cluster_rows = TRUE, cluster_cols = TRUE, show_rownames = TRUE, show_colnames = TRUE,
         dev.off()
```

```
## pdf
## 2
```

Interpretación Biológica de los Resultados

Diversidad Celular y Funcional:

El extenso número de niveles en la variable "tipo celular" indica que el estudio abarca una amplia gama

Impacto de Tratamientos In Vitro:

La inclusión de variables referentes a tratamientos in vitro sugiere que se evaluó el efecto de condiciones

Relación entre Tejido y Expresión Génica:

Al comparar muestras de distintos tejidos (como médula ósea vs. sangre periférica o hígado fetal), el análisis

Conclusiones sobre Regulación Genética:

La identificación de genes diferencialmente expresados mediante análisis estadístico (usando voom/limma)