

Proyecto_RNAseq

Análisis de Expresión Diferencial

Karla Ximena González Platas

2025-02-05

Contents

Introducción	1
Instalación y carga de paquetes	1
Selección de Proyecto	2
Preparación de los datos	6
Filtrar genes de baja expresión	8
Normalización de los datos	10
Determinar el modelo estadístico	11
Visualizar matriz	13
Expresión diferencial	14
Visualizar genes DE	18

Introducción

El conjunto de datos analizado en este proyecto se deriva del estudio de Muluhngwi y Klinge (2021), que explora las interacciones regulatorias entre los miembros de la familia miR-29 y los lncRNAs (ARN largos no codificantes) en el contexto de la resistencia a la terapia endocrina. En este estudio, se aplicó el análisis de RNA-seq para investigar la expresión de lncRNAs regulados por miR-29b-1-3p y miR-29a-3p en células de cáncer de mama sensibles a la terapia endocrina (MCF-7) y resistentes a la terapia endocrina (LCC9). Estas líneas celulares fueron empleadas para estudiar los efectos de la regulación de lncRNAs por los miR-29b-1-3p y miR-29a-3p en la progresión del cáncer de mama y la resistencia endocrina. Para ello, se realizaron transfecciones en las células MCF-7 y LCC9 con pre-miR-29b-1-3p y pre-miR-29a-3p para evaluar los efectos de su sobreexpresión, mientras que el uso de anti-miR-29 y un control negativo permitió analizar la inhibición de estos miRNAs. Asimismo, se llevaron a cabo co-transfecciones combinando pre-miR-29b-1-3p o pre-miR-29a-3p con anti-miR-29 para examinar posibles efectos compensatorios en la regulación de los lncRNAs.

Instalación y carga de paquetes

```

# Instalar BiocManager si no está instalado
#if (!requireNamespace("BiocManager", quietly = TRUE)) {
#   install.packages("BiocManager")
#}

# Instalar paquetes de Bioconductor
#BiocManager::install(
#  c(
#    "edgeR",
#    "ExploreModelMatrix",
#    "limma",
#    "recount3"
#  )
#)

# Instalar paquetes de CRAN
#install.packages(c(
#  "pheatmap",
#  "patchwork",
#  "RColorBrewer",
#  "cowplot"
#))

## Cargar los paquetes
library("recount3")
library("limma")
library("edgeR")
library("ExploreModelMatrix")
library("cowplot")
library("RColorBrewer")
library("pheatmap")

```

Selección de Proyecto

```

# Obtener la lista de proyectos disponibles
human_projects <- available_projects()

## 2025-02-06 18:32:11.702399 caching file sra.recount_project.MD.gz.

## 2025-02-06 18:32:12.442418 caching file gtex.recount_project.MD.gz.

## 2025-02-06 18:32:13.029812 caching file tcga.recount_project.MD.gz.

# Ver los proyectos disponibles
dim(human_projects)

## [1] 8742      6

```

```
# Esto nos indica cuántos proyectos están disponibles (número de filas)
# y cuántas columnas de información se proporcionan para cada proyecto.
```

```
# Mostrar las primeras filas para inspeccionar su estructura y contenido
head(human_projects)
```

```
##      project organism file_source      project_home project_type n_samples
## 1 SRP107565     human      sra data_sources/sra data_sources    216
## 2 SRP149665     human      sra data_sources/sra data_sources     4
## 3 SRP017465     human      sra data_sources/sra data_sources   23
## 4 SRP119165     human      sra data_sources/sra data_sources    6
## 5 SRP133965     human      sra data_sources/sra data_sources  12
## 6 SRP096765     human      sra data_sources/sra data_sources    7
```

```
# Seleccionar un estudio de interés
human_projects[709, ]
```

```
##      project organism file_source      project_home project_type n_samples
## 709 SRP075398     human      sra data_sources/sra data_sources    18
```

```
# Filtrar el dataframe para seleccionar un proyecto específico basado en su ID y tipo
project_info <- subset(
  human_projects,
  project == "SRP075398" & project_type == "data_sources"
)
```

```
# Mostrar la información del proyecto seleccionado para confirmar que se ha
# filtrado correctamente
project_info
```

```
##      project organism file_source      project_home project_type n_samples
## 709 SRP075398     human      sra data_sources/sra data_sources    18
```

```
# Crear un objeto de tipo RangedSummarizedExperiment (RSE) con la información a nivel de genes
rse_gene_SRP075398 <- create_rse(project_info)
```

```
## 2025-02-06 18:32:21.216004 downloading and reading the metadata.
```

```
## 2025-02-06 18:32:21.925683 caching file sra.sra.SRP075398.MD.gz.
```

```
## 2025-02-06 18:32:22.518392 caching file sra.recount_project.SRP075398.MD.gz.
```

```
## 2025-02-06 18:32:23.079376 caching file sra.recount_qc.SRP075398.MD.gz.
```

```
## 2025-02-06 18:32:23.790696 caching file sra.recount_seq_qc.SRP075398.MD.gz.
```

```
## 2025-02-06 18:32:24.482664 caching file sra.recount_pred.SRP075398.MD.gz.
```

```
## 2025-02-06 18:32:24.727526 downloading and reading the feature information.
```

```

## 2025-02-06 18:32:25.242203 caching file human.gene_sums.G026.gtf.gz.

## 2025-02-06 18:32:26.729366 downloading and reading the counts: 18 samples across 63856 features.

## 2025-02-06 18:32:27.353656 caching file sra.gene_sums.SRP075398.G026.gz.

## 2025-02-06 18:32:27.988332 constructing the RangedSummarizedExperiment (rse) object.

# Explorar el objeto RSE
rse_gene_SRP075398

## class: RangedSummarizedExperiment
## dim: 63856 18
## metadata(8): time_created recount3_version ... annotation recount3_url
## assays(1): raw_counts
## rownames(63856): ENSG00000278704.1 ENSG00000277400.1 ...
##   ENSG00000182484.15_PAR_Y ENSG00000227159.8_PAR_Y
## rowData names(10): source type ... havana_gene tag
## colnames(18): SRR3544525 SRR3544526 ... SRR3544537 SRR3544540
## colData names(175): rail_id external_id ...
##   recount_pred.curated.cell_line BigWigURL

## Información sobre el RSE creado
metadata(rse_gene_SRP075398)

## $time_created
## [1] "2025-02-06 18:32:28 CST"
##
## $recount3_version
##   package ondiskversion loadedversion
##   recount3 recount3      1.16.0      1.16.0
##   path
##   recount3 /usr/local/lib/R/site-library/recount3
##   loadedpath attached is_base      date
##   recount3 /usr/local/lib/R/site-library/recount3      TRUE    FALSE 2024-10-29
##   source md5ok          library
##   recount3 Bioconductor 3.20 (R 4.4.2)    NA /usr/local/lib/R/site-library
##
## $project
## [1] "SRP075398"
##
## $project_home
## [1] "data_sources/sra"
##
## $type
## [1] "gene"
##
## $organism
## [1] "human"
##
## $annotation
## [1] "gencode_v26"

```

```

## 
## $recount3_url
## [1] "http://duffel.rail.bio/recount3"

```

```

## Número de genes y número de muestras
dim(rse_gene_SRP075398)

```

```

## [1] 63856    18

```

El estudio **SRP075398** se compuso de **18 muestras**, para las cuales tenemos **63,856 genes** en GENCODE v26. La información específica de la anotación está disponible rowRanges() como se muestra a continuación con la columna gene_id utilizada para identificar genes en cada una de las anotaciones.

```

# Información sobre los genes
rowRanges(rse_gene_SRP075398)

```

```

## GRanges object with 63856 ranges and 10 metadata columns:
##           seqnames      ranges strand | source
##           <Rle>      <IRanges> <Rle> | <factor>
## ENSG00000278704.1 GL000009.2 56140-58376 - | ENSEMBL
## ENSG00000277400.1 GL000194.1 53590-115018 - | ENSEMBL
## ENSG00000274847.1 GL000194.1 53594-115055 - | ENSEMBL
## ENSG00000277428.1 GL000195.1 37434-37534 - | ENSEMBL
## ENSG00000276256.1 GL000195.1 42939-49164 - | ENSEMBL
## ...
##   ...     ...     ...     ...     ...
## ENSG0000124334.17_PAR_Y chrY 57184101-57197337 + | HAVANA
## ENSG0000185203.12_PAR_Y chrY 57201143-57203357 - | HAVANA
## ENSG0000270726.6_PAR_Y chrY 57190738-57208756 + | HAVANA
## ENSG0000182484.15_PAR_Y chrY 57207346-57212230 + | HAVANA
## ENSG0000227159.8_PAR_Y chrY 57212184-57214397 - | HAVANA
##           type bp_length phase          gene_id
##           <factor> <numeric> <integer> <character>
## ENSG00000278704.1  gene     2237    <NA>  ENSG00000278704.1
## ENSG00000277400.1  gene     2179    <NA>  ENSG00000277400.1
## ENSG00000274847.1  gene     1599    <NA>  ENSG00000274847.1
## ENSG00000277428.1  gene      101    <NA>  ENSG00000277428.1
## ENSG00000276256.1  gene     2195    <NA>  ENSG00000276256.1
## ...
##   ...     ...     ...     ...
## ENSG0000124334.17_PAR_Y  gene     2504    <NA>  ENSG00000124334.17_PA..
## ENSG0000185203.12_PAR_Y  gene     1054    <NA>  ENSG00000185203.12_PA..
## ENSG0000270726.6_PAR_Y  gene      773    <NA>  ENSG00000270726.6_PA..
## ENSG0000182484.15_PAR_Y  gene     4618    <NA>  ENSG00000182484.15_PA..
## ENSG00000227159.8_PAR_Y  gene     1306    <NA>  ENSG00000227159.8_PA..
##           gene_type gene_name level
##           <character> <character> <character>
## ENSG00000278704.1  protein_coding BX004987.1 3
## ENSG00000277400.1  protein_coding AC145212.2 3
## ENSG00000274847.1  protein_coding AC145212.1 3
## ENSG00000277428.1  misc_RNA      Y_RNA 3
## ENSG00000276256.1  protein_coding AC011043.1 3
## ...
##   ...     ...     ...
## ENSG0000124334.17_PAR_Y  protein_coding      IL9R 2
## ENSG0000185203.12_PAR_Y  antisense        WASIR1 2

```

```

##      ENSG00000270726.6_PAR_Y processed_transcript AJ271736.10          2
##      ENSG00000182484.15_PAR_Y transcribed_unproces..          WASH6P          2
##      ENSG00000227159.8_PAR_Y unprocessed_pseudogene          DDX11L16          2
##                                havana_gene          tag
##                                <character> <character>
##      ENSG00000278704.1          <NA>          <NA>
##      ENSG00000277400.1          <NA>          <NA>
##      ENSG00000274847.1          <NA>          <NA>
##      ENSG00000277428.1          <NA>          <NA>
##      ENSG00000276256.1          <NA>          <NA>
##      ...          ...
##      ENSG00000124334.17_PAR_Y OTTHUMG00000022720.1          PAR
##      ENSG00000185203.12_PAR_Y OTTHUMG00000022676.3          PAR
##      ENSG00000270726.6_PAR_Y OTTHUMG00000184987.2          PAR
##      ENSG00000182484.15_PAR_Y OTTHUMG00000022677.5          PAR
##      ENSG00000227159.8_PAR_Y OTTHUMG00000022678.1          PAR
##      -----
##      seqinfo: 374 sequences from an unspecified genome; no seqlengths

```

Preparación de los datos

```
# Convertir las cuentas por nucleotido a cuentas por lectura usando compute_read_counts().
assay(rse_gene_SRP075398, "counts") <- compute_read_counts(rse_gene_SRP075398)
```

```
# Inspeccionar la información experimental de cada muestra
rse_gene_SRP075398$sra.sample_attributes[]
```

```

## [1] "cell line;;LCC9|source_name;;LCC9 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [2] "cell line;;LCC9|source_name;;LCC9 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [3] "cell line;;LCC9|source_name;;LCC9 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [4] "cell line;;LCC9|source_name;;LCC9 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [5] "cell line;;LCC9|source_name;;LCC9 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"
## [6] "cell line;;LCC9|source_name;;LCC9 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"
## [7] "cell line;;LCC9|source_name;;LCC9 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"
## [8] "cell line;;MCF-7|source_name;;MCF-7 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [9] "cell line;;MCF-7|source_name;;MCF-7 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [10] "cell line;;MCF-7|source_name;;MCF-7 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"
## [11] "cell line;;MCF-7|source_name;;MCF-7 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"
## [12] "cell line;;LCC9|source_name;;LCC9 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [13] "cell line;;LCC9|source_name;;LCC9 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [14] "cell line;;MCF-7|source_name;;MCF-7 cell line pre-miR-29b-1 transfected|transfection;;Pre-miR-29b-1"
## [15] "cell line;;MCF-7|source_name;;MCF-7 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [16] "cell line;;MCF-7|source_name;;MCF-7 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [17] "cell line;;MCF-7|source_name;;MCF-7 cell line Anti-miR-29a transfected|transfection;;Anti-miR-29a"
## [18] "cell line;;MCF-7|source_name;;MCF-7 cell line Pre-miR-29a transfected|transfection;;Pre-miR-29a"

```

```
# Expandir los atributos en columnas separadas para facilitar su uso
rse_gene_SRP075398 <- expand_sra_attributes(rse_gene_SRP075398)
```

```
# Extraer y mostrar las columnas que contienen atributos
colData(rse_gene_SRP075398) [
```

```

  ,
  grepl("^sra_attribute", colnames(colData(rse_gene_SRP075398)))
]

```

```

## DataFrame with 18 rows and 3 columns
##           sra_attribute.cell_line sra_attribute.source_name
##           <character>          <character>
## SRR3544525          LCC9    LCC9 cell line pre-m..
## SRR3544526          LCC9    LCC9 cell line Anti...
## SRR3544527          LCC9    LCC9 cell line Anti...
## SRR3544528          LCC9    LCC9 cell line Anti...
## SRR3544529          LCC9    LCC9 cell line Pre-m..
## ...
## ...
## SRR3544534          MCF-7   MCF-7 cell line pre...
## SRR3544535          MCF-7   MCF-7 cell line Anti...
## SRR3544536          MCF-7   MCF-7 cell line Anti...
## SRR3544537          MCF-7   MCF-7 cell line Anti...
## SRR3544540          MCF-7   MCF-7 cell line Pre...
##           sra_attribute.transfection
##           <character>
## SRR3544525          Pre-miR-29b-1
## SRR3544526          Anti-miR-29a
## SRR3544527          Anti-miR-29a
## SRR3544528          Anti-miR-29a
## SRR3544529          Pre-miR-29a
## ...
## ...
## SRR3544534          Pre-miR-29b-1
## SRR3544535          Anti-miR-29a
## SRR3544536          Anti-miR-29a
## SRR3544537          Anti-miR-29a
## SRR3544540          Pre-miR-29a

```

Ajustar el tipo de dato de las variables categóricas

```

rse_gene_SRP075398$sra_attribute.cell_line <- factor(rse_gene_SRP075398$sra_attribute.cell_line)

rse_gene_SRP075398$sra_attribute.source_name <- factor(tolower(rse_gene_SRP075398$sra_attribute.source_name))

rse_gene_SRP075398$sra_attribute.transfection <- factor(rse_gene_SRP075398$sra_attribute.transfection)

```

Resumen estadístico de las variables seleccionadas

```

summary(as.data.frame(colData(rse_gene_SRP075398)[

```

```

  ,
  grepl("^sra_attribute.[cell_line|source_name|transfection]", colnames(colData(rse_gene_SRP075398)))
])
```

## sra_attribute.cell_line	sra_attribute.source_name
## LCC9 :9	lcc9 cell line anti-mir-29a transfected :3
## MCF-7:9	lcc9 cell line pre-mir-29a transfected :3
##	lcc9 cell line pre-mir-29b-1 transfected :3
##	mcf-7 cell line anti-mir-29a transfected :3
##	mcf-7 cell line pre-mir-29a transfected :3
##	mcf-7 cell line pre-mir-29b-1 transfected:3

```

## sra_attribute.transfection
## Anti-miR-29a :6
## Pre-miR-29a  :6
## Pre-miR-29b-1:6
##
##
```

Calcular la proporción de lecturas asignadas a genes para evaluar la calidad de las muestras

```

rse_gene_SRP075398$assigned_gene_prop <-
  rse_gene_SRP075398$recount_qc.gene_fc_count_all.assigned /
  rse_gene_SRP075398$recount_qc.gene_fc_count_all.total
```

Resumen de la nueva variable para identificar si las muestras tienen una asignación adecuada de lecturas

```

summary(rse_gene_SRP075398$assigned_gene_prop)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	0.6076	0.6405	0.6603	0.6585	0.6696	0.7017

Filtrar genes de baja expresión

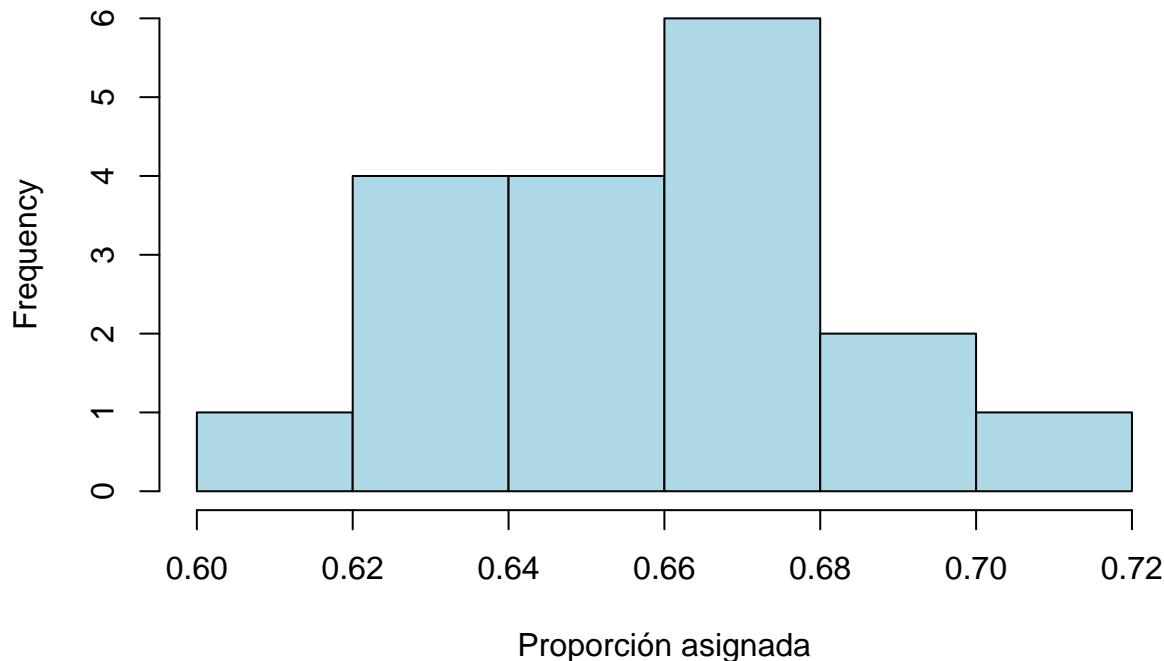
```

# Guardar el objeto original
rse_gene_SRP075398_unfiltered <- rse_gene_SRP075398

# Visualizar la distribución de la proporción de lecturas asignadas a genes en cada muestra

hist(rse_gene_SRP075398$assigned_gene_prop,
      main = "Proporción de lecturas asignadas a genes",
      xlab = "Proporción asignada", col = "lightblue")
```

Proporción de lecturas asignadas a genes



```
# Verificar si existen muestras de baja calidad antes del filtrado
table(rse_gene_SRP075398$assigned_gene_prop < 0.3)

##
## FALSE
##     18

# Filtrar las muestras con proporción de lecturas asignadas superior a 0.3
rse_gene_SRP075398 <- rse_gene_SRP075398[, rse_gene_SRP075398$assigned_gene_prop > 0.3]

# Crear un objeto DGEList, para el análisis diferencial usando edgeR
dge <- DGEList(counts = assay(rse_gene_SRP075398, "counts"))

# Filtrar genes de baja expresión considerando combinaciones de transfección y línea celular
keep <- filterByExpr(dge, group = interaction(
  rse_gene_SRP075398$sra_attribute.transfection,
  rse_gene_SRP075398$sra_attribute.cell_line
))
rse_gene_SRP075398 <- rse_gene_SRP075398[keep, ]

# Dimensiones finales
dim(rse_gene_SRP075398)

## [1] 23741    18
```

```

# Porcentaje de genes retenidos
round(nrow(rse_gene_SRP075398) / nrow(rse_gene_SRP075398_unfiltered) * 100, 2)

## [1] 37.18

```

Se descartó los genes de baja expresión porque no contribuyen significativamente a las conclusiones biológicas. De modo que, después del filtrado se obtuvieron 23741 lo cual representa el 37.18% de genes retenidos.

Normalización de los datos

```

# Crear un objeto DGEList para normalización
dge <- DGEList(
  counts = assay(rse_gene_SRP075398, "counts"),
  genes = rowData(rse_gene_SRP075398)
)

# Normalización TMM
dge <- calcNormFactors(dge)

dge

## An object of class "DGEList"
## $counts
##          SRR3544525 SRR3544526 SRR3544527 SRR3544528 SRR3544529
## ENSG00000223972.5     44      31      54      44      62
## ENSG00000227232.5    297     264     405     352     242
## ENSG00000238009.6     32      27      19      32      20
## ENSG00000233750.3     10      15       6       1       9
## ENSG00000268903.1     17       7       9       6      17
##          SRR3544530 SRR3544531 SRR3544532 SRR3544533 SRR3544538
## ENSG00000223972.5     37      51      13      18      16
## ENSG00000227232.5    215     277     200     204     245
## ENSG00000238009.6     15      25      19      30      48
## ENSG00000233750.3     7       9      13       9      14
## ENSG00000268903.1     6      13       8       3       4
##          SRR3544539 SRR3544523 SRR3544524 SRR3544534 SRR3544535
## ENSG00000223972.5     13      71      52      10      11
## ENSG00000227232.5    105     509     353     217     123
## ENSG00000238009.6     23      45      28      26      31
## ENSG00000233750.3     3       16       7       8       3
## ENSG00000268903.1     1      35      24       0       4
##          SRR3544536 SRR3544537 SRR3544540
## ENSG00000223972.5     7       9      25
## ENSG00000227232.5    74     163     183
## ENSG00000238009.6     32      27      45
## ENSG00000233750.3     8       3       6
## ENSG00000268903.1     0       2       3
## 23736 more rows ...
##
## $samples

```

```

##          group lib.size norm.factors
## SRR3544525     1 40711468    1.0516143
## SRR3544526     1 44717701    1.0300052
## SRR3544527     1 55798559    0.9927744
## SRR3544528     1 61298574    0.9995922
## SRR3544529     1 34513836    1.0385196
## 13 more rows ...
##
## $genes
##           source type bp_length phase      gene_id
## ENSG00000223972.5 HAVANA gene      1735     NA ENSG00000223972.5
## ENSG00000227232.5 HAVANA gene      1351     NA ENSG00000227232.5
## ENSG00000238009.6 HAVANA gene      3726     NA ENSG00000238009.6
## ENSG00000233750.3 HAVANA gene      3812     NA ENSG00000233750.3
## ENSG00000268903.1 HAVANA gene      755      NA ENSG00000268903.1
##           gene_type      gene_name level
## ENSG00000223972.5 transcribed_unprocessed_pseudogene      DDX11L1    2
## ENSG00000227232.5 unprocessed_pseudogene      WASH7P    2
## ENSG00000238009.6 lincRNA      RP11-34P13.7    2
## ENSG00000233750.3 processed_pseudogene      CICP27    1
## ENSG00000268903.1 processed_pseudogene      RP11-34P13.15   2
##           havana_gene      tag
## ENSG00000223972.5 OTTHUMG00000000961.2      <NA>
## ENSG00000227232.5 OTTHUMG00000000958.1      <NA>
## ENSG00000238009.6 OTTHUMG00000001096.2 overlapping_locus
## ENSG00000233750.3 OTTHUMG00000001257.3 pseudo_consens
## ENSG00000268903.1 OTTHUMG00000182518.2      <NA>
## 23736 more rows ...

```

Determinar el modelo estadístico

```

# Construcción de la matriz de diseño para el modelo lineal.
mod <- model.matrix(
  ~ sra_attribute.cell_line + sra_attribute.transfection + assigned_gene_prop,
  data = colData(rse_gene_SRP075398)
)

# Cada columna representa un coeficiente del modelo
colnames(mod)

## [1] "(Intercept)"
## [2] "sra_attribute.cell_lineMCF-7"
## [3] "sra_attribute.transfectionPre-miR-29a"
## [4] "sra_attribute.transfectionPre-miR-29b-1"
## [5] "assigned_gene_prop"

# Visualizar la matriz de diseño completa
# Las filas representan muestras, mientras que las columnas son las variables del modelo
mod

## (Intercept) sra_attribute.cell_lineMCF-7

```

```

## SRR3544525      1          0
## SRR3544526      1          0
## SRR3544527      1          0
## SRR3544528      1          0
## SRR3544529      1          0
## SRR3544530      1          0
## SRR3544531      1          0
## SRR3544532      1          1
## SRR3544533      1          1
## SRR3544538      1          1
## SRR3544539      1          1
## SRR3544523      1          0
## SRR3544524      1          0
## SRR3544534      1          1
## SRR3544535      1          1
## SRR3544536      1          1
## SRR3544537      1          1
## SRR3544540      1          1
##           sra_attribute.transfectionPre-miR-29a
## SRR3544525      0
## SRR3544526      0
## SRR3544527      0
## SRR3544528      0
## SRR3544529      1
## SRR3544530      1
## SRR3544531      1
## SRR3544532      0
## SRR3544533      0
## SRR3544538      1
## SRR3544539      1
## SRR3544523      0
## SRR3544524      0
## SRR3544534      0
## SRR3544535      0
## SRR3544536      0
## SRR3544537      0
## SRR3544540      1
##           sra_attribute.transfectionPre-miR-29b-1 assigned_gene_prop
## SRR3544525      1      0.6075824
## SRR3544526      0      0.6587443
## SRR3544527      0      0.6946650
## SRR3544528      0      0.7017052
## SRR3544529      0      0.6617664
## SRR3544530      0      0.6664334
## SRR3544531      0      0.6554926
## SRR3544532      1      0.6359671
## SRR3544533      1      0.6320064
## SRR3544538      0      0.6623243
## SRR3544539      0      0.6705981
## SRR3544523      1      0.6586547
## SRR3544524      1      0.6286454
## SRR3544534      1      0.6352978
## SRR3544535      0      0.6539348
## SRR3544536      0      0.6907278

```

```

## SRR3544537          0      0.6651556
## SRR3544540          0      0.6725733
## attr(,"assign")
## [1] 0 1 2 2 3
## attr(,"contrasts")
## attr(,"contrasts")$sra_attribute.cell_line
## [1] "contr.treatment"
##
## attr(,"contrasts")$sra_attribute.transfection
## [1] "contr.treatment"

```

Este modelo incluye:

- sra_attribute.cell_line: Efecto del tipo de línea celular (LCC9 o MCF-7).
- sra_attribute.transfection: Efecto del tratamiento por transfección (pre-miR-29a, pre-miR-29b-1, Anti-miR-29a).
- assigned_gene_prop: Proporción de lecturas asignadas a genes

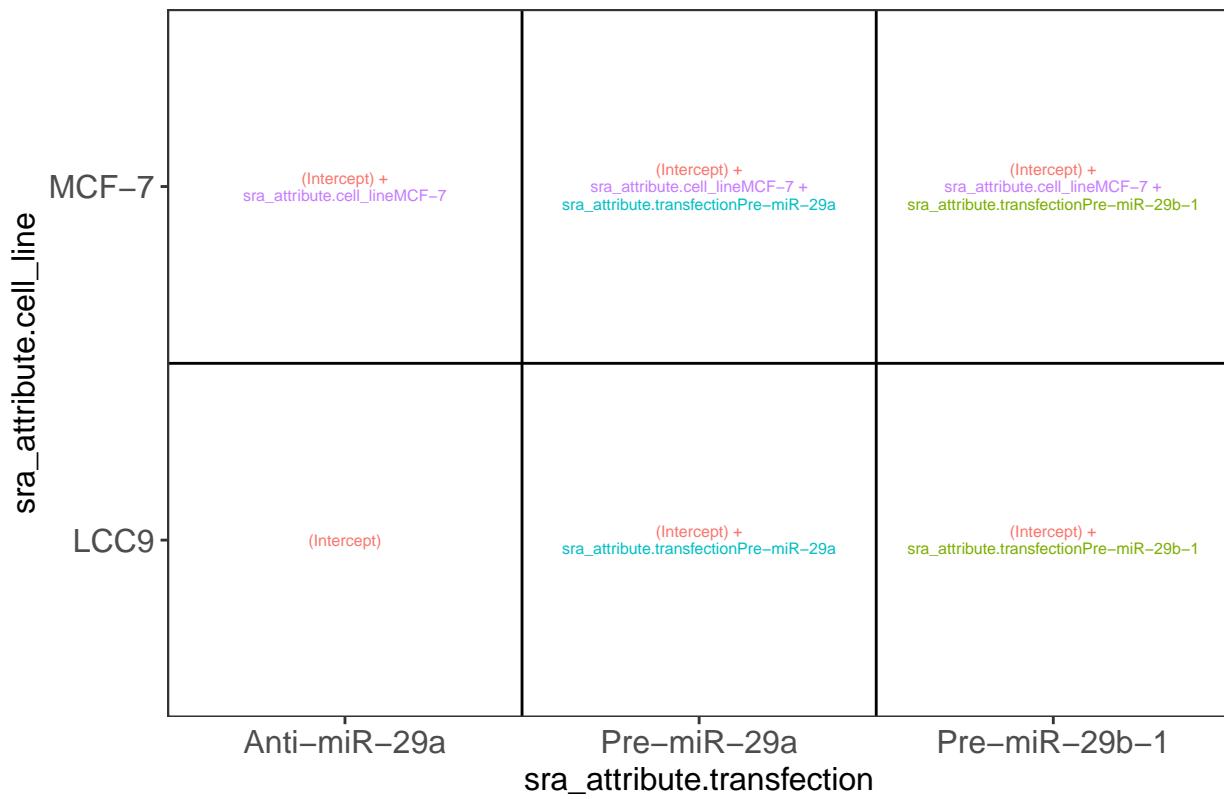
Visualizar matriz

```

## Crear las visualizaciones
vd <- ExploreModelMatrix::VisualizeDesign(
  sampleData = colData(rse_gene_SRP075398), # Metadatos de las muestras
  designFormula = ~ sra_attribute.cell_line + sra_attribute.transfection,
  textSizeFitted = 2
)

cowplot::plot_grid(plotlist = vd$plotlist)

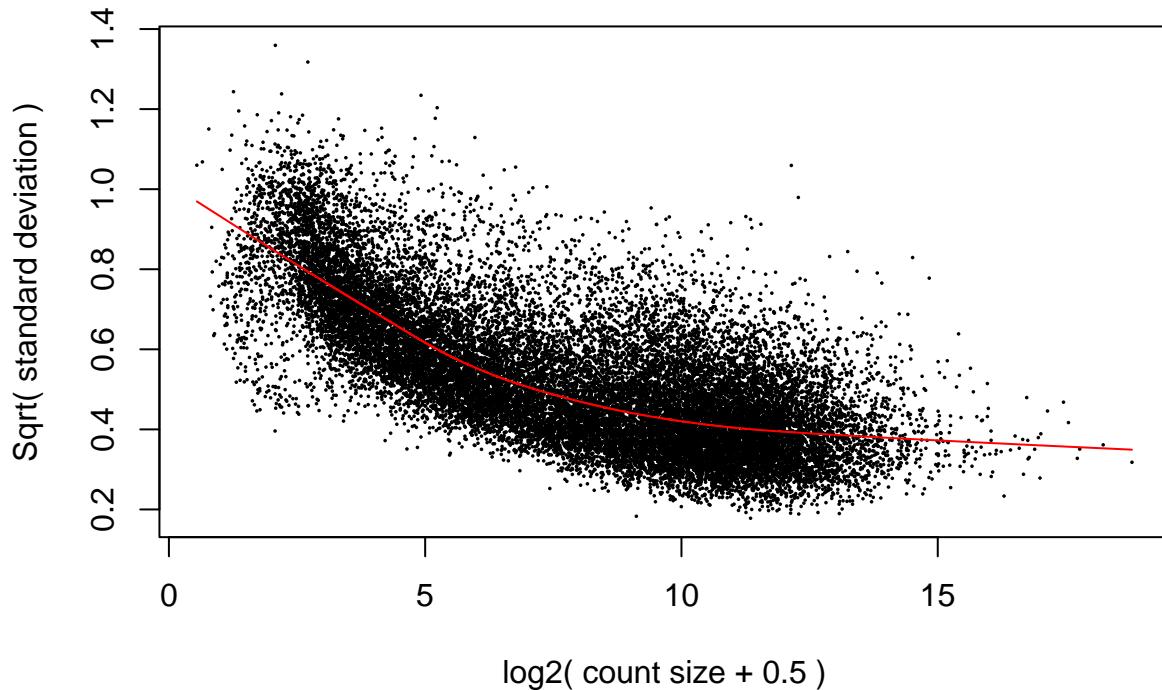
```



Expresión diferencial

```
# Convertir los datos de conteo a valores log2 y ajusta las varianzas para hacerlos aptos para un análisis
vGene <- voom(dge, mod, plot = TRUE)
```

voom: Mean-variance trend



```
# Ajuste del modelo lineal y cálculo de estadísticas empíricas de Bayes
eb_results <- eBayes(lmFit(vGene))

# Extraer la tabla de genes diferencialmente expresados.
de_results <- topTable(
  eb_results,
  coef = 2, # Se refiere al coeficiente del segundo término en el modelo
  number = nrow(rse_gene_SRP075398),
  sort.by = "none"
)

# Dimensiones y vista preliminar de los resultados
dim(de_results)

## [1] 23741     16

head(de_results)

##           source type bp_length phase      gene_id
## ENSG00000223972.5 HAVANA gene       1735      NA ENSG00000223972.5
## ENSG00000227232.5 HAVANA gene       1351      NA ENSG00000227232.5
## ENSG00000238009.6 HAVANA gene       3726      NA ENSG00000238009.6
## ENSG00000233750.3 HAVANA gene       3812      NA ENSG00000233750.3
## ENSG00000268903.1 HAVANA gene        755      NA ENSG00000268903.1
```

```

## ENSG00000269981.1 HAVANA gene      284    NA ENSG00000269981.1
##                                     gene_type   gene_name level
## ENSG00000223972.5 transcribed_unprocessed_pseudogene DDX11L1    2
## ENSG00000227232.5          unprocessed_pseudogene WASH7P    2
## ENSG00000238009.6           lincRNA    RP11-34P13.7    2
## ENSG00000233750.3 processed_pseudogene    CICP27    1
## ENSG00000268903.1 processed_pseudogene RP11-34P13.15    2
## ENSG00000269981.1 processed_pseudogene RP11-34P13.16    2
##                               havana_gene      tag      logFC AveExpr
## ENSG00000223972.5 OTTHUMG00000000961.2 <NA> -1.57964773 -0.7349755
## ENSG00000227232.5 OTTHUMG00000000958.1 <NA> -0.65730720  2.4021094
## ENSG00000238009.6 OTTHUMG00000001096.2 overlapping_locus 0.50892797 -0.5758187
## ENSG00000233750.3 OTTHUMG00000001257.3 pseudo_consens 0.07505824 -2.5024602
## ENSG00000268903.1 OTTHUMG00000182518.2 <NA> -2.26677866 -2.9392186
## ENSG00000269981.1 OTTHUMG00000182738.2 <NA> -0.20655991 -1.1726322
##                               t      P.Value adj.P.Val     B
## ENSG00000223972.5 -9.0583130 5.502604e-08 1.724585e-07 8.293369
## ENSG00000227232.5 -4.2492998 5.203366e-04 8.224575e-04 -1.699547
## ENSG00000238009.6  3.2989311 4.153389e-03 5.833271e-03 -3.230385
## ENSG00000233750.3  0.2097501 8.363087e-01 8.516989e-01 -7.250773
## ENSG00000268903.1 -4.8120051 1.545050e-04 2.624197e-04  0.559485
## ENSG00000269981.1 -0.8707765 3.957782e-01 4.289509e-01 -7.156075

```

```

# Genes diferencialmente expresados con FDR < 5%
table(de_results$adj.P.Val < 0.05)

```

```

##
## FALSE TRUE
## 4642 19099

```

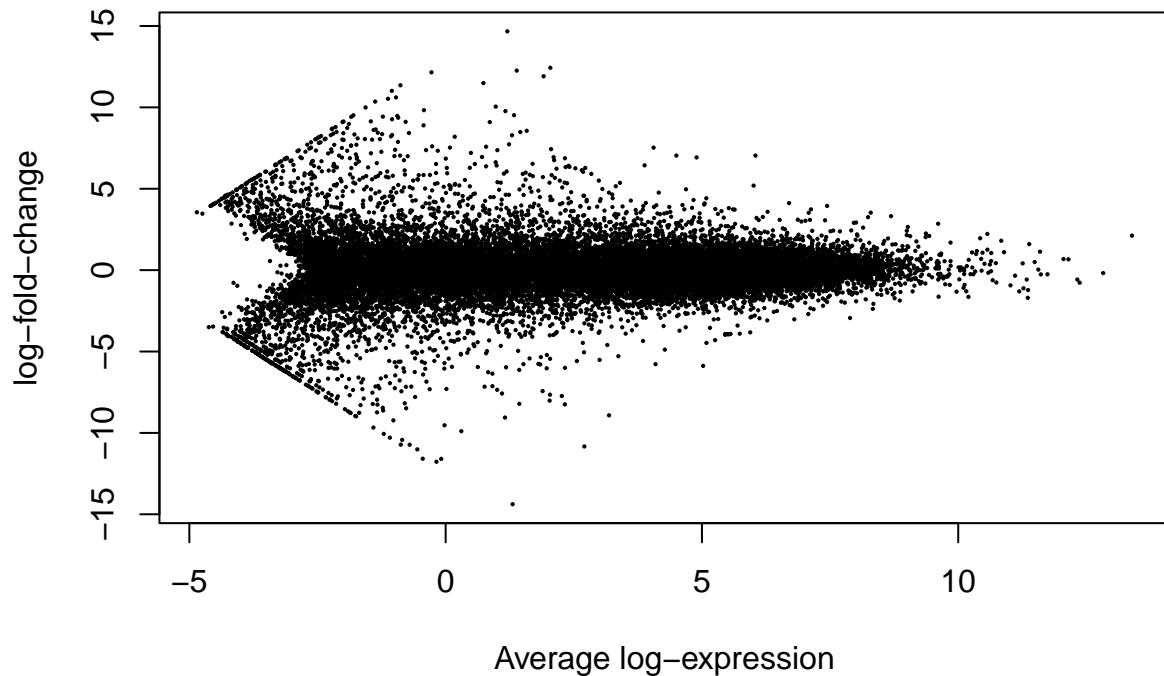
```

# Visualizar los resultados estadísticos

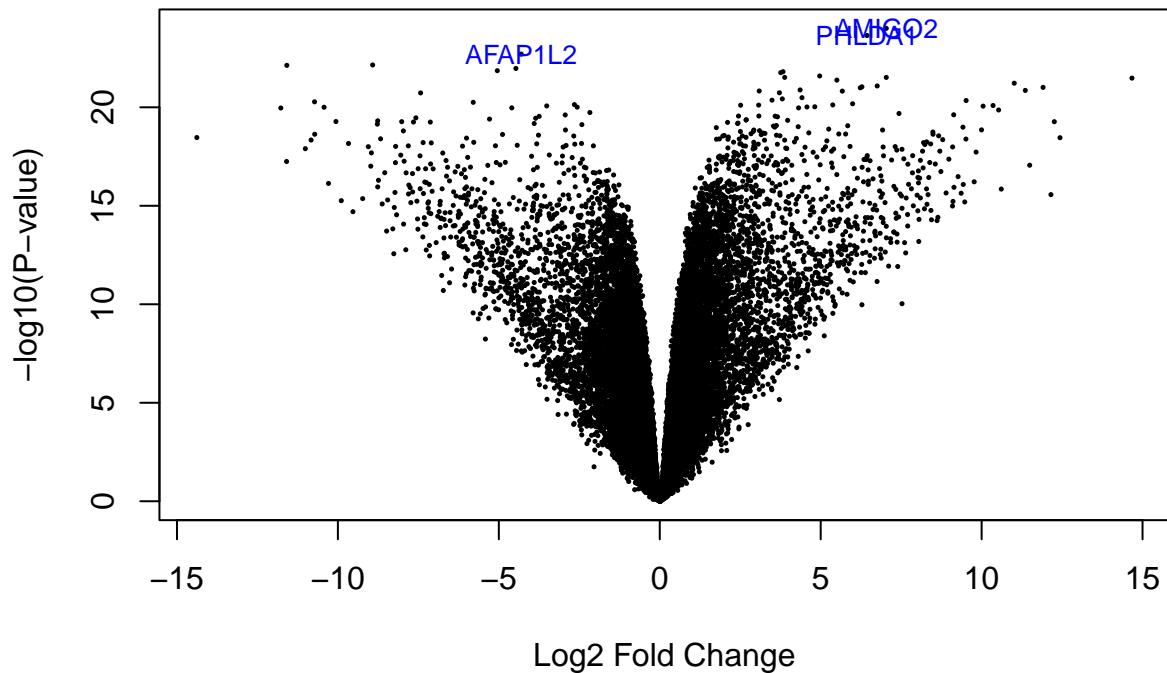
plotMA(eb_results, coef = 2)

```

sra_attribute.cell_lineMCF-7



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



```
# Información de los 3 genes más significativos
de_results[de_results$gene_name %in% c("AMIGO2", "AFAP1L2", "PHLDA1"), ]
```

```
##           source type bp_length phase      gene_id
## ENSG00000169129.14 HAVANA gene     5789    NA ENSG00000169129.14
## ENSG00000139211.6 HAVANA gene     3956    NA ENSG00000139211.6
## ENSG00000139289.13 HAVANA gene     8069    NA ENSG00000139289.13
##           gene_type gene_name level      havana_gene tag
## ENSG00000169129.14 protein_coding AFAP1L2      2 OTTHUMG00000019086.3 <NA>
## ENSG00000139211.6 protein_coding AMIGO2      2 OTTHUMG00000169616.1 <NA>
## ENSG00000139289.13 protein_coding PHLDA1      2 OTTHUMG00000169783.2 <NA>
##           logFC AveExpr      t   P.Value adj.P.Val
## ENSG00000169129.14 -4.302232 5.257918 -77.07393 1.991303e-23 1.575851e-19
## ENSG00000139211.6    7.038259 6.044177  91.73069 9.814316e-25 2.330017e-20
## ENSG00000139289.13   6.438081 3.881338  87.38040 2.274059e-24 2.699422e-20
##           B
## ENSG00000169129.14 43.77924
## ENSG00000139211.6  46.45332
## ENSG00000139289.13 45.06009
```

Visualizar genes DE

```

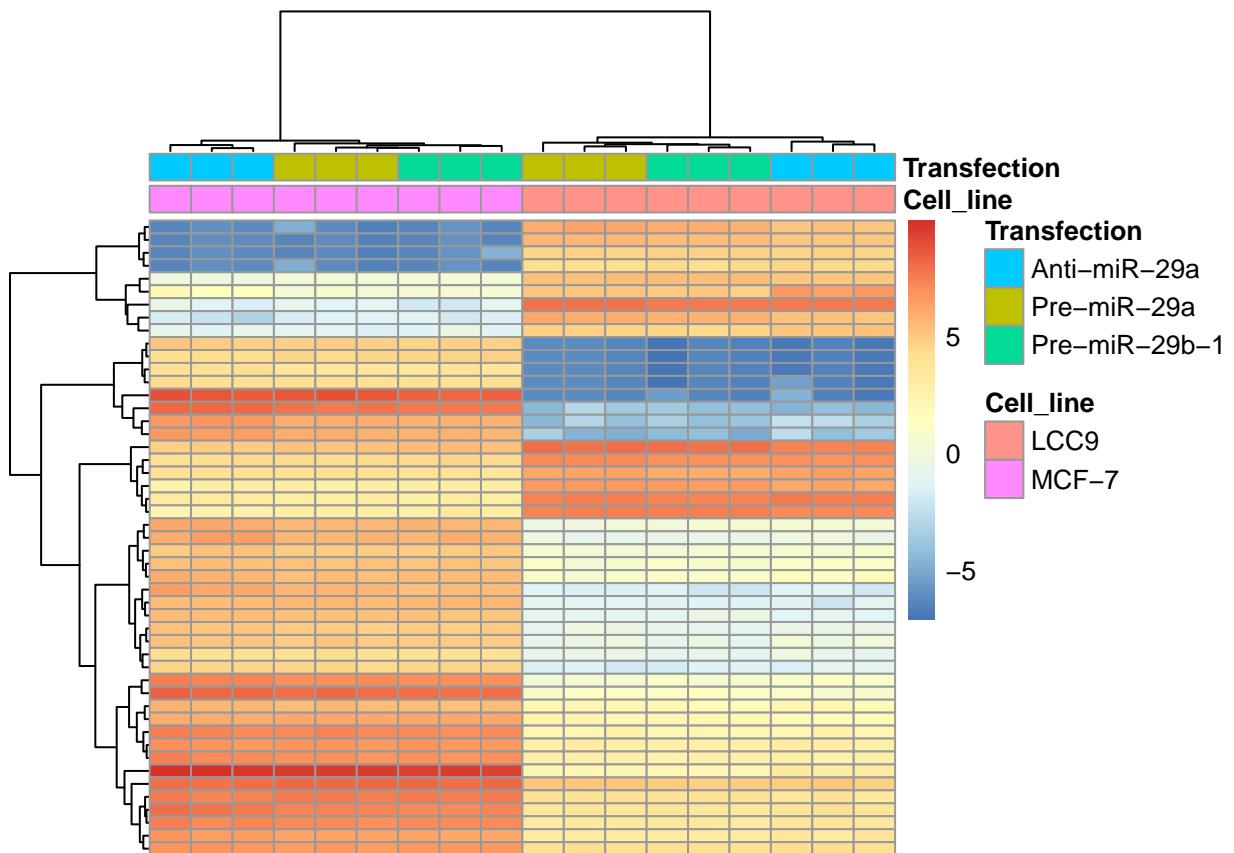
# Revisar los top 50 genes diferencialmente expresados

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

# Crear una tabla con información de las muestras y con nombres de columnas más amigables
df <- as.data.frame(colData(rse_gene_SRP075398)[, c("sra_attribute.cell_line", "sra_attribute.transfection")]
colnames(df) <- c("Cell_line", "Transfection")

# Hacer un heatmap
pheatmap(
  exprs_heatmap,
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  show_rownames = FALSE,
  show_colnames = FALSE,
  annotation_col = df
)

```



```

# MDS (multidimensional scaling)

## Convertir los grupos de Cell_line a colores
col.group <- df$Cell_line
levels(col.group) <- brewer.pal(nlevels(col.group), "Set2")

```

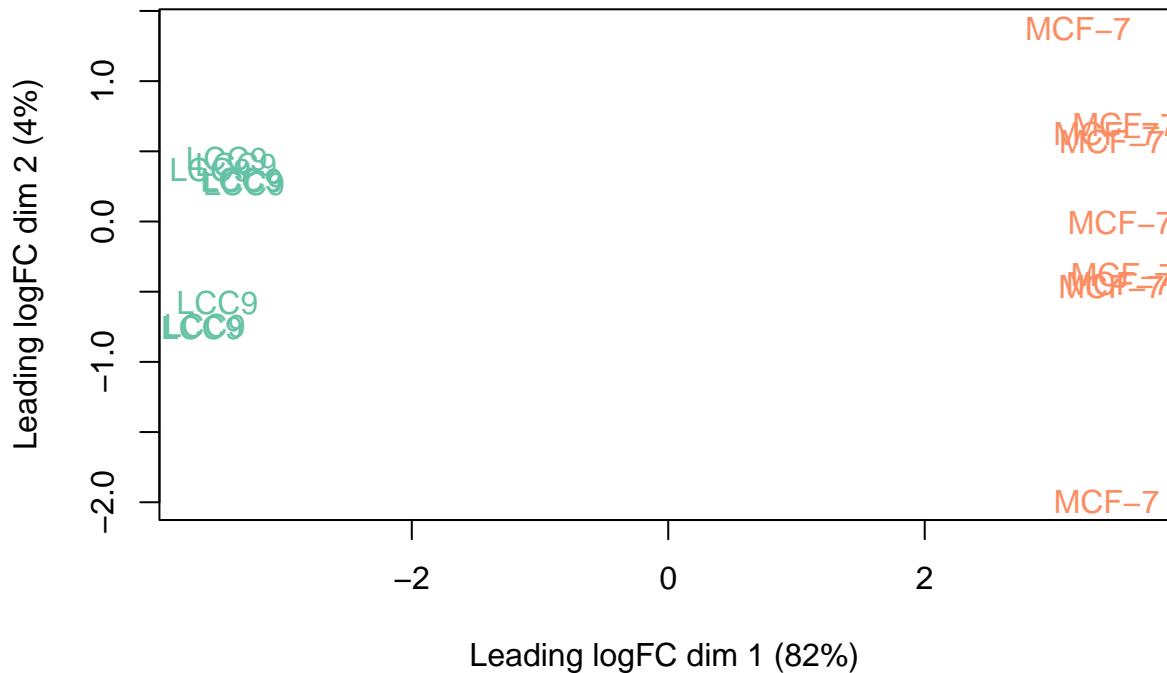
```

## Warning in brewer.pal(nlevels(col.group), "Set2"): minimal value for n is 3, returning requested palette

col.group <- as.character(col.group)

## MDS por grupos de Cell_line
plotMDS(vGene$E, labels = df$Cell_line, col = col.group)

```



```

## Convertir Transfection a colores
col.group <- df$Transfection
df$Transfection <- as.factor(df$Transfection) # Asegúrate de que sea un factor
colors <- brewer.pal(nlevels(df$Transfection), "Set2") # Generar paleta de colores
levels(col.group) <- colors # Asignar colores a los niveles
col.group <- as.character(col.group) # Convertir a vector de caracteres

## MDS por grupos de Transfection
plotMDS(vGene$E, labels = df$Transfection, col = col.group,
         main = "MDS Plot by Transfection", pch = 16)

## Agregar leyenda
legend("topright", legend = levels(df$Transfection), fill = unique(col.group),
       title = "Transfection Groups")

```

MDS Plot by Transfection

