



ANÁLISIS DE DATOS BULK RNA-seq

Carmen González Gálvez

CONTENIDO

1. Descripción del diseño experimental y metodología del artículo	3
1.1 Introducción	3
1.2 Diseño experimental	3
1.3 Generación de datos RNA-seq	3
2. Análisis de la matriz de expresión	4
2.1 Preprocesamiento de Datos	4
Instalación y Carga de Librerías	4
Selección y Filtrado de Archivos	4
Creación de Metadatos	4
2.2 Creación de la matriz de expresión	5
Lectura de los Archivos de Expresión	5
Estructura de la Matriz de Expresión	5
2.3 Filtrado de Genes	6
Cálculo de Genes sin Expresión	6
Filtrado de Genes con Baja Expresión	6
Visualización y transformación de los Datos	6
2.4 Normalización	7
Cálculo de CPM (Counts per Million)	8
Transformación Logarítmica	8
2.5 Análisis Multivariante (PCA)	9
Cálculo del PCA	9
Interpretación de los Resultados	11
Cálculo de Distancias y Matriz de Correlación	12
2.6 Análisis de Expresión Diferencial (EDA)	14
Preparación del Modelo	14
Definición de Contrastes	14
Prueba Estadística	14
Visualización de Resultados	15
2.7 Análisis de enriquecimiento	17
Análisis de sobrerrepresentación (ORA)	18
3. Conclusiones	21
4. Métodos – Tabla de recursos clave	22

1. Descripción del diseño experimental y metodología del artículo

El artículo es el siguiente: <https://www.pnas.org/doi/10.1073/pnas.2102698118>

1.1 Introducción

La **inmunidad entrenada** define adaptaciones duraderas de la inmunidad innata basadas en modificaciones transcripcionales y epigenéticas de las células mieloides y sus progenitores en la médula ósea. Este estudio analiza el **impacto del grupo hemo lábil** como un inductor de inmunidad entrenada previamente no reconocido, evaluando su efecto en la regulación a largo plazo de la diferenciación de células madre hematopoyéticas y progenitoras.

El grupo **hemo lábil** se refiere a una forma libre que puede ser liberada durante la hemólisis o infecciones y actuar como molécula de señalización inmunológica. El **β -glucano** es un polisacárido presente en la pared celular de hongos, levaduras y cereales, reconocido por su capacidad de modular el sistema inmunológico, activa vías de inmunidad entrenada que aumentan la respuesta inmune frente a futuras exposiciones a patógenos. El **LPS** es un lipopolisacárido presente en la membrana externa de bacterias Gram-negativas que actúa como un potente activador de la respuesta inmune, uniéndose al receptor **TLR4**, desencadenando una respuesta inflamatoria a través de la activación de NF- κ B y otras vías de señalización.

1.2 Diseño experimental

- **Aislar los monocitos humanos** primarios de donantes sanos
- Tratar los monocitos con **hemo** (50 μ M) o **β -glucano** durante 24 horas, cogiendo muestras a las 4 horas y a las 24 horas.
- Después de 6 días sin estímulos, estimular los macrófagos diferenciados con **LPS** (10 ng/ml)
- Evaluar los **cambios en la expresión génica** mediante la secuenciación de RNA y en la regulación epigenética mediante **ChIP-seq y ATAC-seq**.
- Además, administraron el grupo hemo lábil a ratones para evaluar su impacto en la hematopoyesis y en la resistencia a la sepsis (infección).

1.3 Generación de datos RNA-seq

- **Aislamiento y cultivo celular:** se obtuvieron monocitos de sangre periférica, se cultivaron en medio RPMI con suero humano, se expusieron a diferentes condiciones experimentales (hemo, β -glucano, LPS) y se diferenciaron a macrófagos en un modelo de inmunidad entrenada.
- **Extracción y procesamiento del RNA:** se extrajo el RNA con QIAzol, se evaluó la calidad del RNA y se generaron bibliotecas con Illumina.

- **Secuenciación y análisis:** la secuenciación se realizó con la plataforma de Illumina (lecturas emparejadas, paired-end), se alinearon las secuencias con el genoma de referencia humano hg19 usando Bowtie, se cuantificó la expresión génica con MMSEQ y se identificaron los genes diferencialmente expresados DESeq2.
-

2. Análisis de la matriz de expresión

2.1 Preprocesamiento de Datos

El análisis empieza con la configuración del entorno en R y la preparación de los datos para su procesamiento. Esto incluye la instalación y carga de librerías, la selección de archivos de expresión génica y la creación de metadatos.

Las muestras se obtienen de la base de datos de GEO, específicamente la serie GSE111003 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111003>)

Instalación y Carga de Librerías

Se utilizan diversos paquetes en R para el procesamiento y análisis de los datos:

- **Manipulación de datos:** dplyr, stringr
- **Visualización de datos:** ggplot2, ComplexHeatmap, ggpubr, RColorBrewer
- **Análisis de expresión génica:** edgeR, lima
- **Anotación de genes humanos:** org.Hs.eg.db
- **Análisis de enriquecimiento:** msigdb, clusterProfiler, fgsea

Estas librerías permiten filtrar, normalizar, analizar y visualizar los datos de expresión génica.

Selección y Filtrado de Archivos

Se observan los archivos disponibles en el directorio de trabajo y se seleccionan aquellos que contienen información relevante. Se filtran los archivos con los patrones de nombre específicos, asegurando que solo se analicen las muestras correctas. Se excluyen los archivos con formatos no deseados (.bw), manteniendo únicamente los archivos de expresión en formato .txt.

Creación de Metadatos

Se extrae información clave de los nombres de los archivos y se organiza en una tabla de metadatos (figura 1). Se generan las siguientes columnas:

- **Sample.ID:** Identificador de la muestra, combinando la condición y la muestra (T0_HD34, BG_4h_HD34, etc.).

- **Sample:** se asigna la muestra, por ejemplo, HD34.
- **Treatment:** se etiqueta el tratamiento como BG o LPS, dependiendo de si el nombre de la condición contiene LPS.
- **Time:** según el tiempo que se recogen las muestras tras el tratamiento (0h, 4h, 24h)
- **Condition:** se une la columna treatment con la columna time

Se verifica la consistencia de los metadatos para evitar errores en los análisis posteriores.

Description: df [45 x 5]

	Sample.ID <chr>	Sample <chr>	Treatment <chr>	Time <chr>	Condition <chr>
HD34_T0	HD34_T0	HD34	No treatment	0h	No treatment_0h
HD34_RPMI_4h	HD34_RPMI_4h	HD34	RPMI	4h	RPMI_4h
HD34_BG_4h	HD34_BG_4h	HD34	BG	4h	BG_4h
HD34_RPMI_24h	HD34_RPMI_24h	HD34	RPMI	24h	RPMI_24h
HD34_BG_24h	HD34_BG_24h	HD34	BG	24h	BG_24h
HD34_RPMI_d6	HD34_RPMI_d6	HD34	RPMI	d6	RPMI_d6
HD34_RPMI_d6_4h	HD34_RPMI_d6_4h	HD34	RPMI	d6_4h	RPMI_d6_4h
HD34_RPMI_d6_LPS	HD34_RPMI_d6_LPS	HD34	RPMI_LPS	d6_4h	RPMI_LPS_d6_4h
HD34_BG_d6_LPS	HD34_BG_d6_LPS	HD34	BG_LPS	d6_4h	BG_LPS_d6_4h
HD37_T0	HD37_T0	HD37	No treatment	0h	No treatment_0h

1-10 of 45 rows

Previous 1 2 3 4 5

Figura 1. Metadatos

2.2 Creación de la matriz de expresión

Una vez generados los metadatos, se crea la matriz de expresión, en la cual se organizan los valores de expresión génica para su análisis.

Lectura de los Archivos de Expresión

Se leen los archivos de expresión asegurando que contengan las columnas esperadas: **feature_id** (genes) y **unique_hits** (niveles de expresión) y se combinan los valores de expresión de todos los archivos en una única matriz.

Estructura de la Matriz de Expresión

Cada fila va a representar un gen (**feature_id**), cada columna representa una muestra, con un nombre único basado en **Sample.ID** y , con todo ello, se comprueba la correcta creación de la matriz, con sus respectivas expresiones génicas por cada gen y cada muestra (figura 2).

Description: df [56,625 x 45]

	HD34_T0 <Int>	HD34_RPMI_4h <Int>	HD34_BG_4h <Int>	HD34_RPMI_24h <Int>	HD34_BG_24h <Int>	HD34_RPMI_d6 <Int>
ENSG00000000003	3	0	1	2	0	2
ENSG00000000005	0	0	0	0	0	0
ENSG000000000419	334	436	476	381	133	70
ENSG000000000457	135	169	147	205	70	56
ENSG000000000460	62	80	85	161	61	35
ENSG000000000938	4995	4307	9446	2760	713	1216
ENSG000000000971	9	21	17	34	10	19
ENSG000000001036	145	284	279	673	295	334
ENSG000000001084	271	484	1619	933	431	625
ENSG000000001167	688	986	1041	791	296	239

1-10 of 56,625 rows | 1-10 of 45 columns

Figura 2. Matriz de expresión (filas = genes, columnas = muestras)

2.3 Filtrado de Genes

Para evitar ruido en el análisis, se eliminan genes con poca o ninguna expresión. Esto permite enfocarse en genes relevantes.

Desde el inicio hay 56625 genes, y a partir de estos genes se empieza a filtrar:

Cálculo de Genes sin Expresión

Se calcula el número de genes que tienen 0 en todas las muestras, se elimina este conjunto de genes y se mantienen 36547 genes

Filtrado de Genes con Baja Expresión

Se usa `rowSums()` para conservar únicamente los genes con al menos una cuenta de expresión en alguna muestra y, para que sea aún más preciso, se aplica `filterByExpr()`, basándose en la columna Condition, para eliminar genes con niveles de expresión muy bajos. Por lo que, nos quedamos con 17884 genes.

Visualización y transformación de los Datos

Se visualiza la distribución de los valores de expresión con histogramas antes (figura 3) y después (figura 4) de la transformación a escala logarítmica (`log2`), la cual mejora la distribución y reduce la variabilidad técnica.

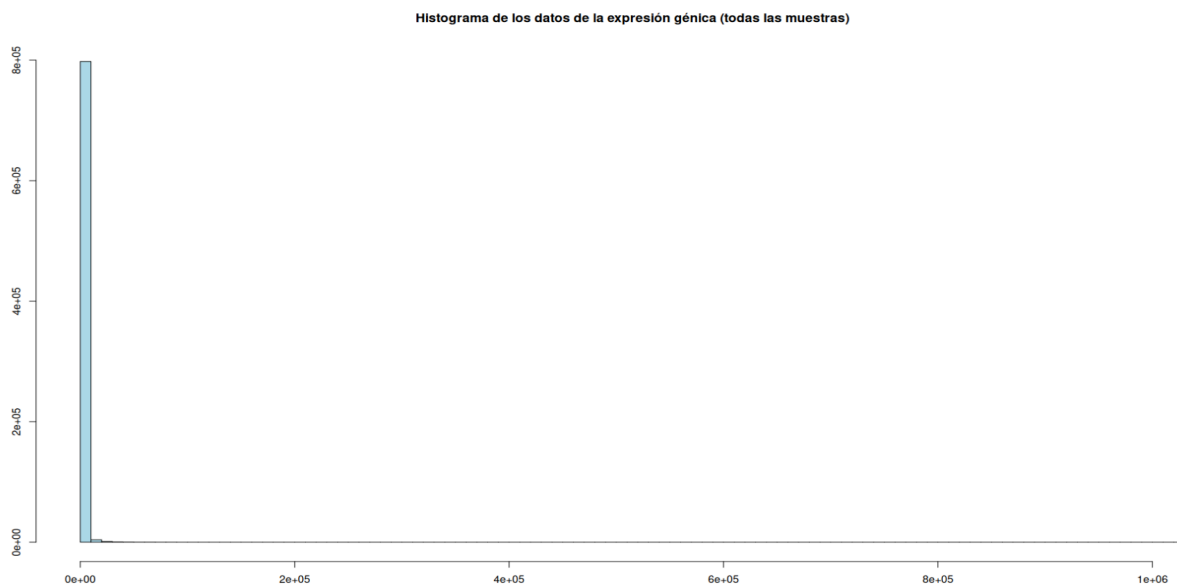


Figura 3. Histograma de los datos de la expresión génica (todas las muestras)

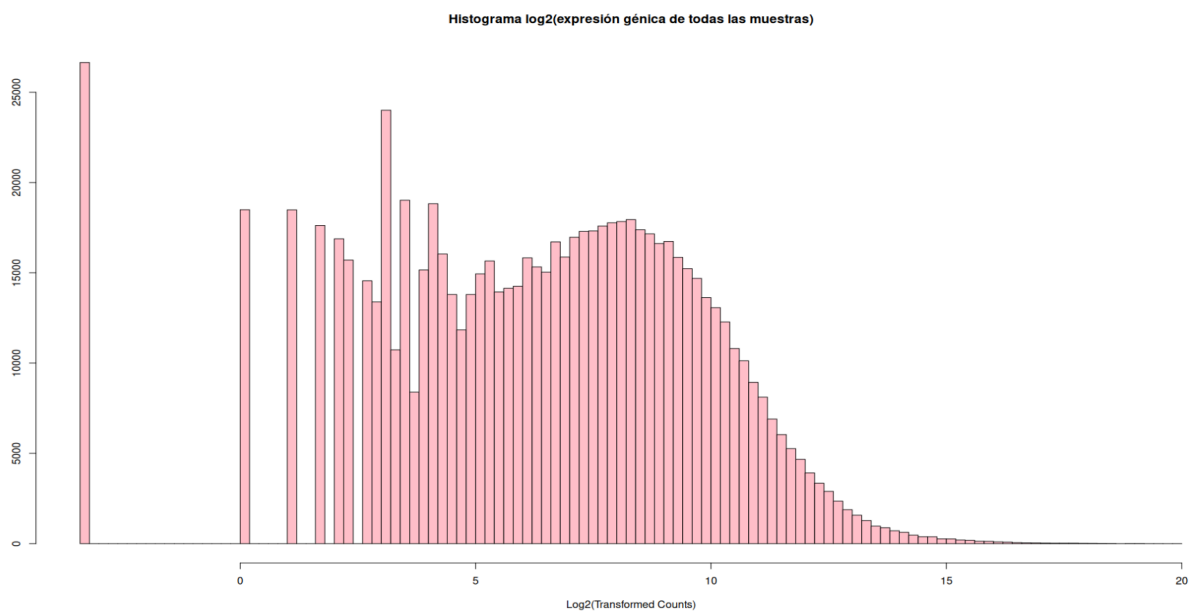


Figura 4. Histograma log2(expresión génica de todas las muestras)

2.4 Normalización

Para **eliminar los sesgos causados por las diferencias en la profundidad de secuenciación** entre las muestras, normalizamos los datos. Las cuentas crudas generalmente no se utilizan en los análisis; en su lugar, las transformamos para tener en cuenta las variaciones en el tamaño de la librería. Una transformación común es el counts-per-million (CPM), que estandariza los datos como si todas las muestras tuvieran 1 millón de cuentas. Esto ayuda a eliminar la variabilidad no deseada y garantiza una comparación justa entre las muestras.

Además, otra transformación necesaria para muchos análisis es la transformación logarítmica de los CPM. Este paso pretende hacer que los datos se parezcan más a los datos que siguen una distribución normal.

Cálculo de CPM (Counts per Million)

Se ajustan los valores de expresión para que todas las muestras tengan el equivalente a 1 millón de cuentas y se grafican las distribuciones de CPM antes y después de la normalización (figura 5).

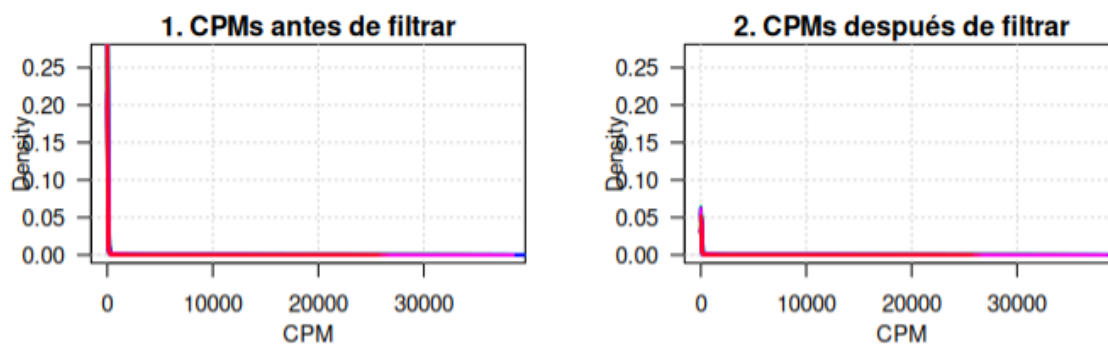


Figura 5. Normalización por CPM antes y después de filtrar los genes. Se observa una mejora en la normalización al filtrar los genes porque se están quitando genes que pueden causar ruido

Transformación Logarítmica

Se aplica $\log_2(\text{CPM} + 1)$ para estabilizar la varianza y mejorar la comparabilidad entre muestras y se grafican las densidades de expresión antes y después de la transformación (figura 6).

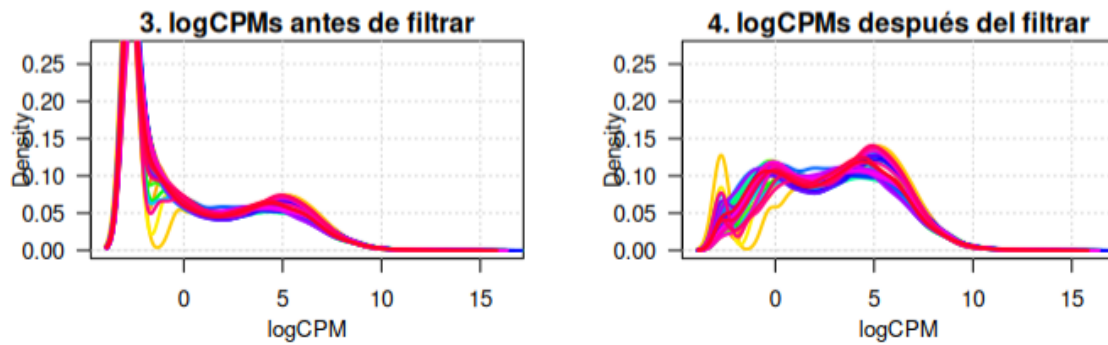


Figura 6. Normalización por $\log_2(\text{CPM})$ antes y después de filtrar los genes. Al aplicar la escala logarítmica para acercarlos a una distribución normal, se observa cómo mantener los genes de baja expresión afecta considerablemente la distribución general de los datos., aunque no siga una distribución perfecta, se parece bastante.

2.5 Análisis Multivariante (PCA)

Se utiliza el **Análisis de Componentes Principales (PCA)**, ya que sirve para la reducción de dimensiones. Su objetivo es transformar un conjunto de variables correlacionadas en un nuevo conjunto de variables llamadas **componentes principales**, que son linealmente independientes y capturan la mayor variabilidad posible de los datos.

Cálculo del PCA

Se aplica `prcomp()` a los datos normalizados para escalar los datos y que todos tengan la misma unidad de medida. Se grafican los dos primeros componentes principales (PC1 y PC2), gracias a que antes se ha creado una función que genera un gráfico PCA y se divide por muestra (figura 7), tratamiento (figura 8), tiempo (figura 9) y condición (figura 10).

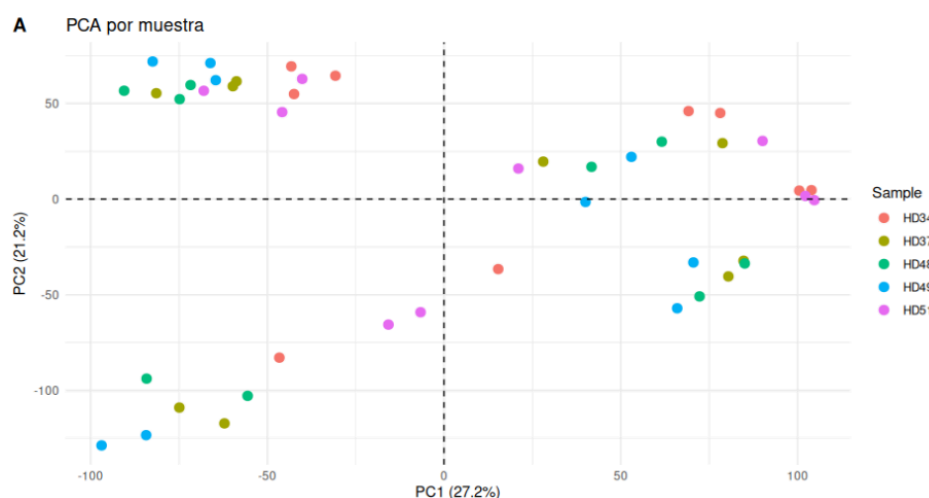


Figura 7. PCA por muestra

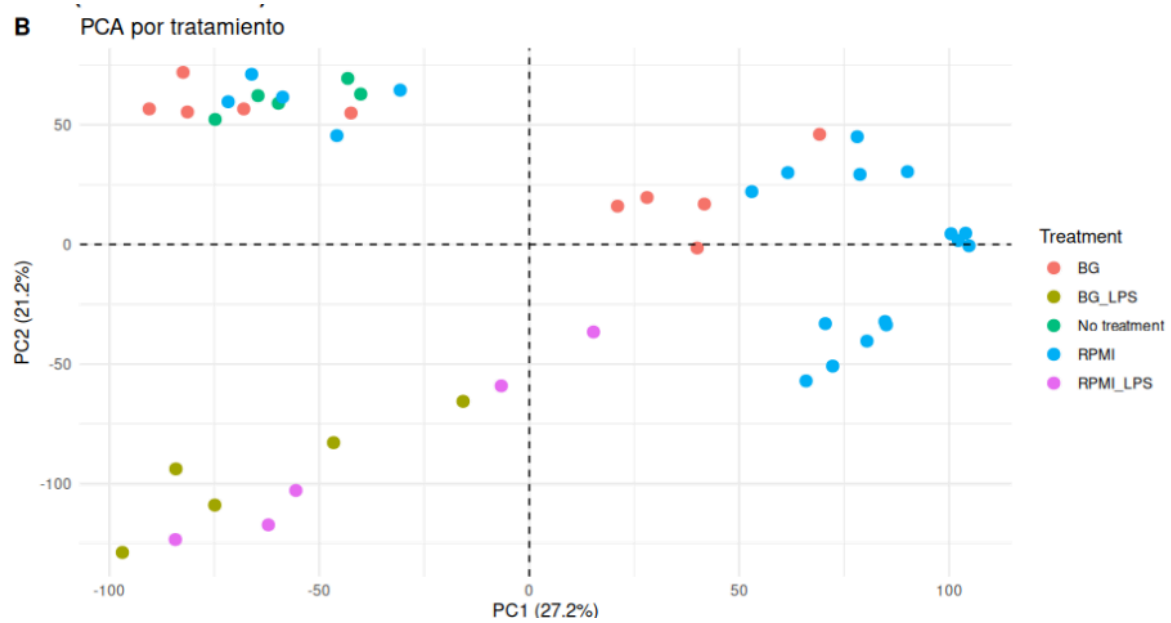


Figura 8. PCA por tratamiento

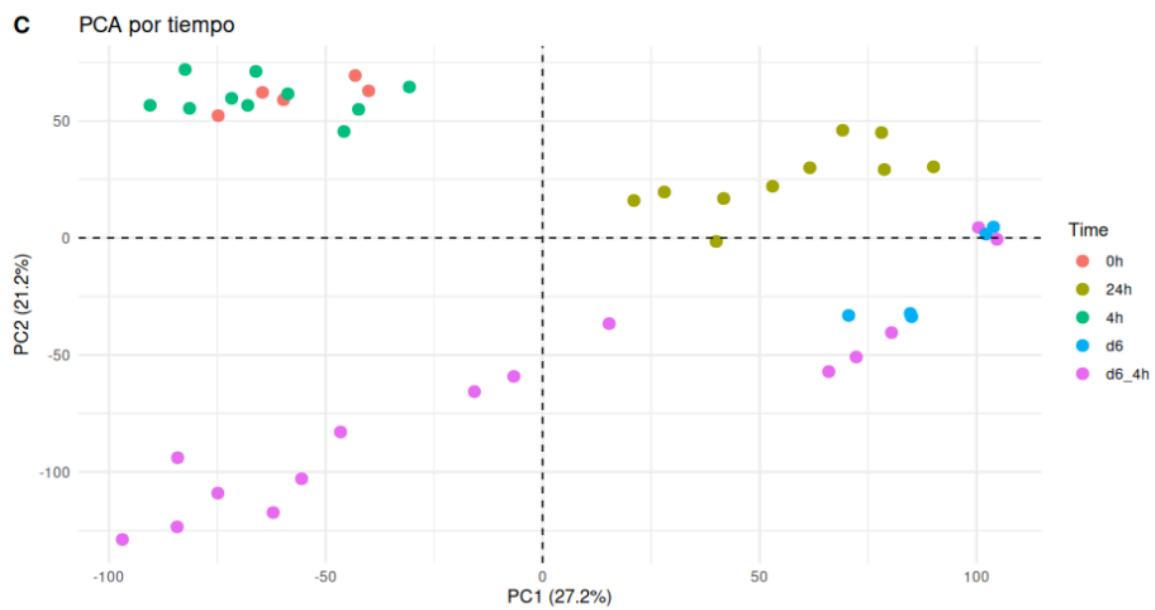


Figura 9. PCA por tiempo

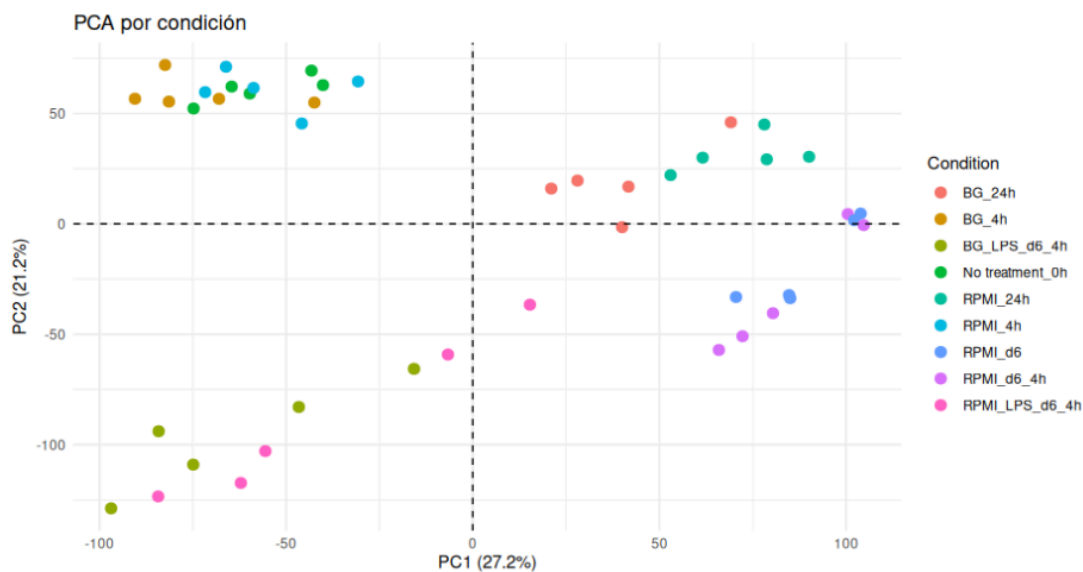


Figura 10. PCA por condición

Interpretación de los Resultados

- **Por Muestra** : no se observa un patrón claro en la distribución de los puntos en el espacio de los primeros dos componentes principales (PC1 y PC2). Esto indica que la variabilidad entre muestras individuales no es suficiente para agruparlas de manera diferenciada, sugiriendo que otros factores (como el tratamiento o el tiempo) podrían tener una mayor influencia en la variabilidad de los datos.
- **Por Tratamiento** : en el eje PC1 (27.2% de variabilidad explicada), se observa que las muestras tratadas con RPMI tienden a ubicarse en la parte positiva del eje x. Esto es lógico, ya que RPMI corresponde a los controles, mientras que las demás muestras han recibido algún tipo de tratamiento.

En el eje PC2 (21.2% de variabilidad explicada), la distribución de los tratamientos no es tan clara, por lo que no se puede extraer una conclusión sólida sobre cómo los tratamientos afectan este componente.

- **Por Tiempo** : en el eje PC1, se observa que las muestras tomadas a las 24 horas de aplicar BG tienden a ubicarse en la parte positiva del eje x. Esto sugiere que este grupo de muestras presenta características diferenciadas en comparación con las demás.

En el eje PC2, se nota que las muestras tomadas después de 6 días más 4 horas (independientemente de si fueron tratadas con Control + LPS o BG + LPS) se agrupan en la parte negativa del eje y. Esto indica que el tiempo transcurrido después de la exposición al tratamiento con LPS podría estar generando un efecto claro en la expresión génica.

- **Por condición:** al igual que por muestra, no se observa un patrón claro en ambas PCs.

Por último se **explora los PCs**, para ver cómo se distribuyen:

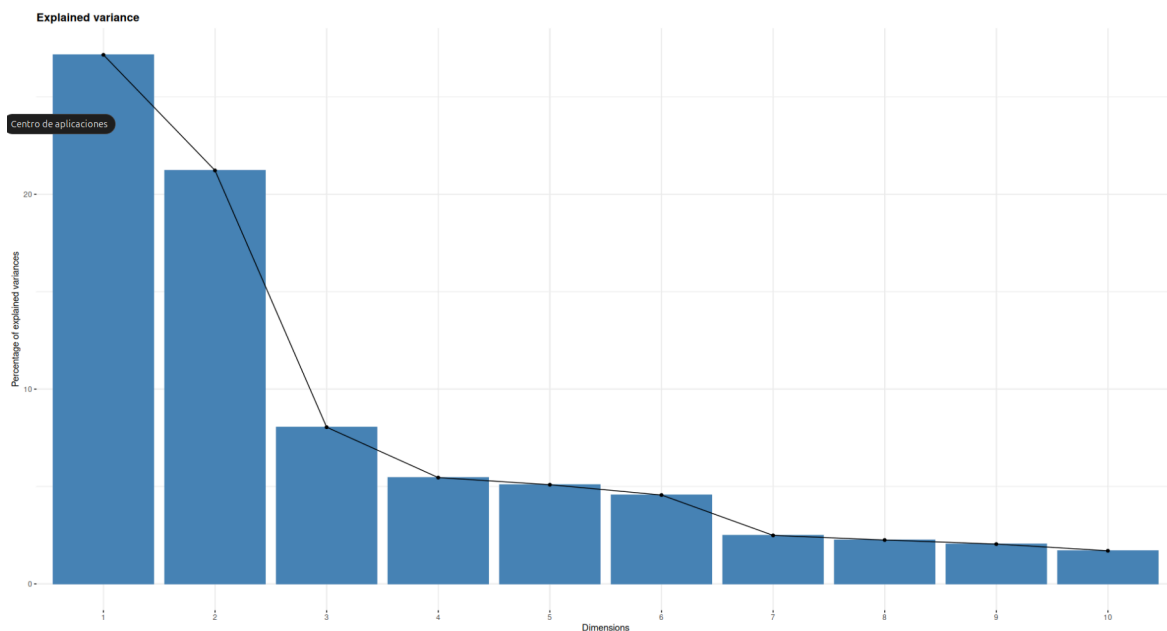


Figura 11. Porcentaje de variabilidad explicada por los PCS. El PC1 es el componente que más variabilidad explica, seguido del PC2 y, así, sucesivamente disminuyendo el porcentaje de variabilidad explicada

Cálculo de Distancias y Matriz de Correlación

Otra idea para ver la similitud entre muestras es creando matrices de distancias euclidianas y representarlas con mapas de calor o generando matrices de correlación. Por lo que, se debería de ver algo similar al PCA

Se puede calcular la distancia euclidiana en el espacio PCA y en el espacio transcripcional, pero he decidido hacerlo en el espacio PCA porque proporciona una versión más limpia de la distancia entre muestras, mientras que el espacio transcripcional refleja diferencias a nivel bruto.

Se calcula la distancia euclidiana entre muestras y se representa la matriz de correlación con un heatmap.

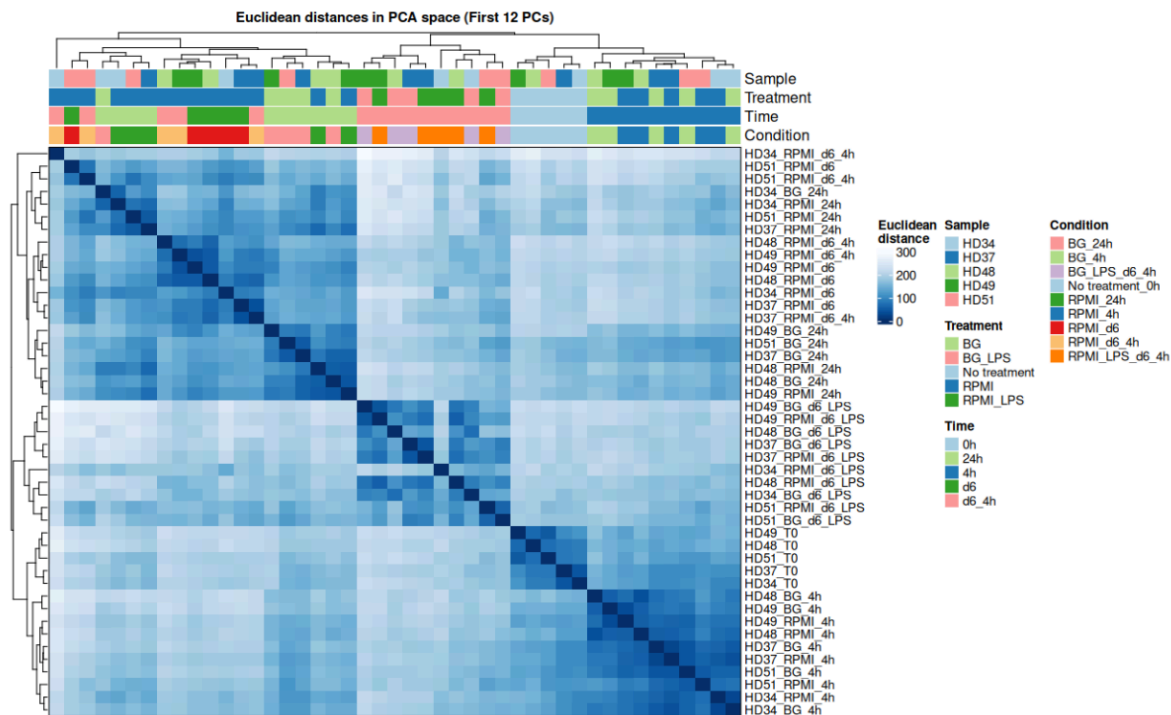


Figura 12. Matriz de distancias euclídeas

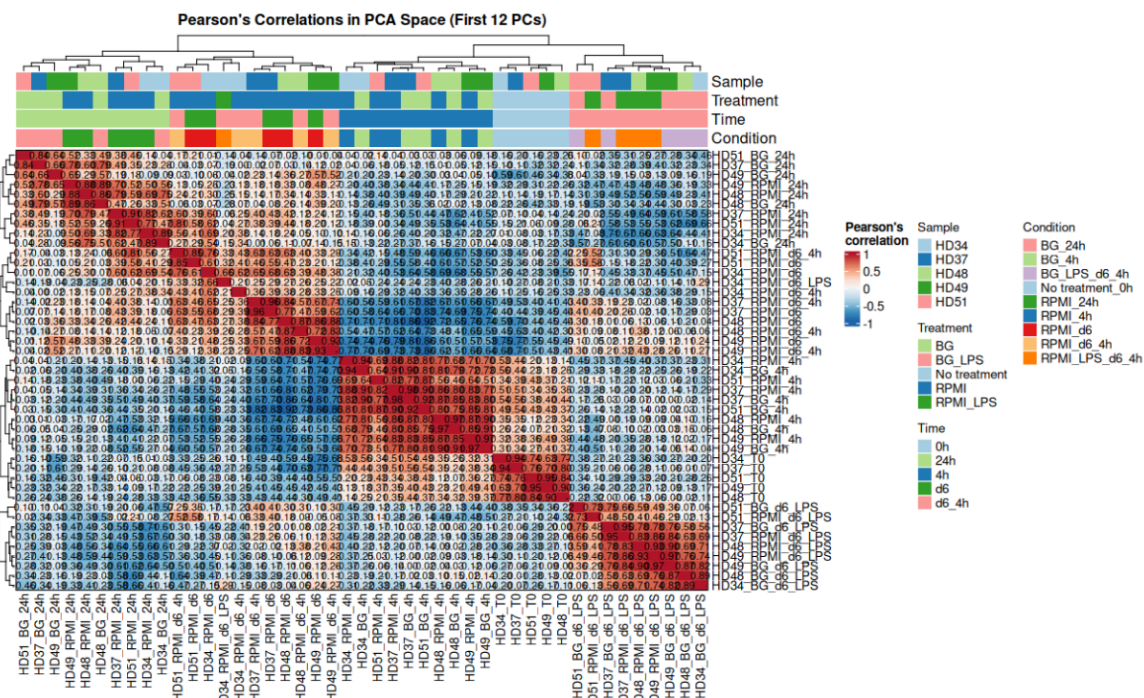


Figura 13. Matriz de correlación

2.6 Análisis de Expresión Diferencial (EDA)

Para identificar genes diferencialmente expresados (**DEGs**), se emplea **limma** por varias razones:

- Implementa un conjunto de funciones diseñadas para realizar análisis en datos transcriptómicos.
- Los **datos no cumplen con la homocedasticidad** (varianza igual), es decir, los la varianza aumentan con la media. **Para corregirlo, limma usa la función 'voom()',** siendo necesario para que los modelos lineales generen p-valores confiables.
- Utiliza un **método para estimar la varianza**, optimizado para estudios transcriptómicos donde hay muchas características (genes), pero un número reducido de réplicas.

Preparación del Modelo

Se crea una matriz de diseño o que genera una matriz sin intercepto, donde cada nivel de Condition se representa como una columna con valores binarios (1 si la muestra pertenece a esa condición, 0 en caso contrario) y se ajusta un modelo lineal para cada gen con **lmFit()**.

Definición de Contrastes

Se establecen comparaciones clave como:

```
# Definir los contrastes basados en Condition
contr <- makeContrasts(
  ConditionNo.treatment_0h - ConditionBG_4h ,
  ConditionNo.treatment_0h - ConditionBG_24h ,
  ConditionRPMI_4h - ConditionBG_4h,
  ConditionRPMI_24h - ConditionBG_24h,
  ConditionRPMI_d6_4h - ConditionRPMI_LPS_d6_4h,
  ConditionRPMI_LPS_d6_4h - ConditionBG_LPS_d6_4h,
  levels = colnames(design)
)
```

Luego se aplican los contrastes al modelo ajustado, se ajusta la varianza con **eBayes()**

Prueba Estadística

Se realizan varios data frames de cada contraste para obtener los logFC (diferencias entre ambas condiciones) y se extraen genes diferencialmente expresados con **logFC** significativo y **p-value** ajustado < 0.05, obteniendo:

- ConditionNo.treatment_0h - ConditionBG_4h: 4389 genes
- ConditionNo.treatment_0h - ConditionBG_24h: 7874
- ConditionRPMI_4h - ConditionBG_4h: 12
- ConditionRPMI_24h - ConditionBG_24h: 541
- ConditionRPMI_d6_4h - ConditionRPMI_LPS_d6_4h: 8540
- ConditionRPMI_LPS_d6_4h - ConditionBG_LPS_d6_4h: 2

Visualización de Resultados

- **Histogramas de p-valores:** Se grafican las distribuciones de significancia, de los cuales hay tres contrastes que tienen la mayoría de p-valores menores a 0.05 y otros tres que no se les ve mucha significancia estadística. Los que tienen mayor significancia coinciden con los que tienen más genes diferenciados:
 - ConditionNo.treatment_0h - ConditionBG_4h
 - ConditionNo.treatment_0h - ConditionBG_24h
 - ConditionNo.treatment_0h - ConditionBG_24h
 - ConditionRPMI_24h - ConditionBG_24h (este más bien no, pero esta más o menos)

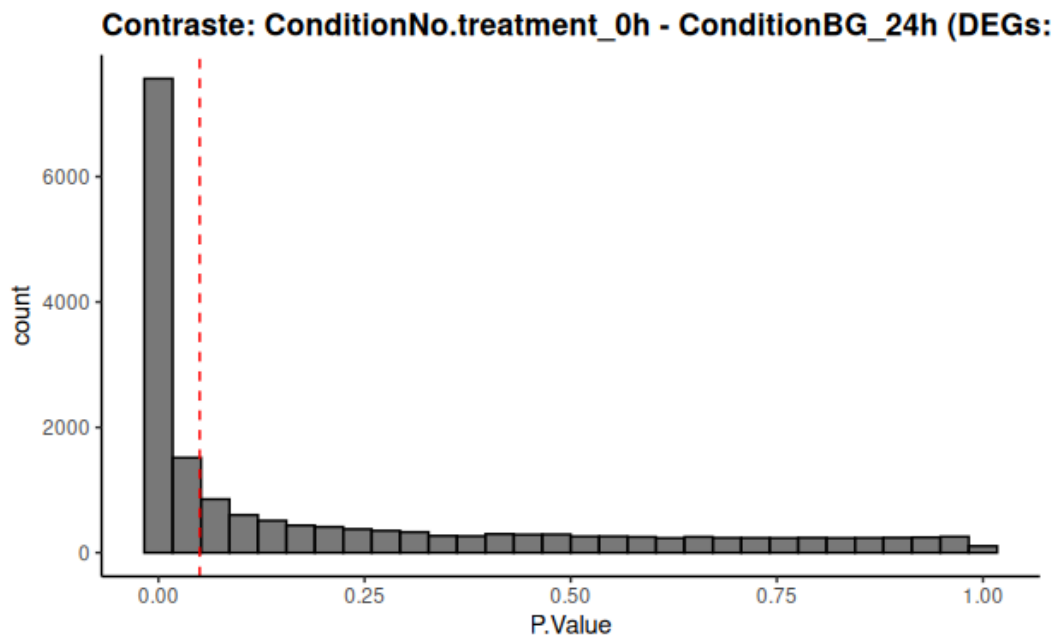


Figura 14. Distribución de p-valores para el contraste de ConditionNO.treatment_0h - ConditionBG_24h. Los que son significativos tienen una distribución parecida.

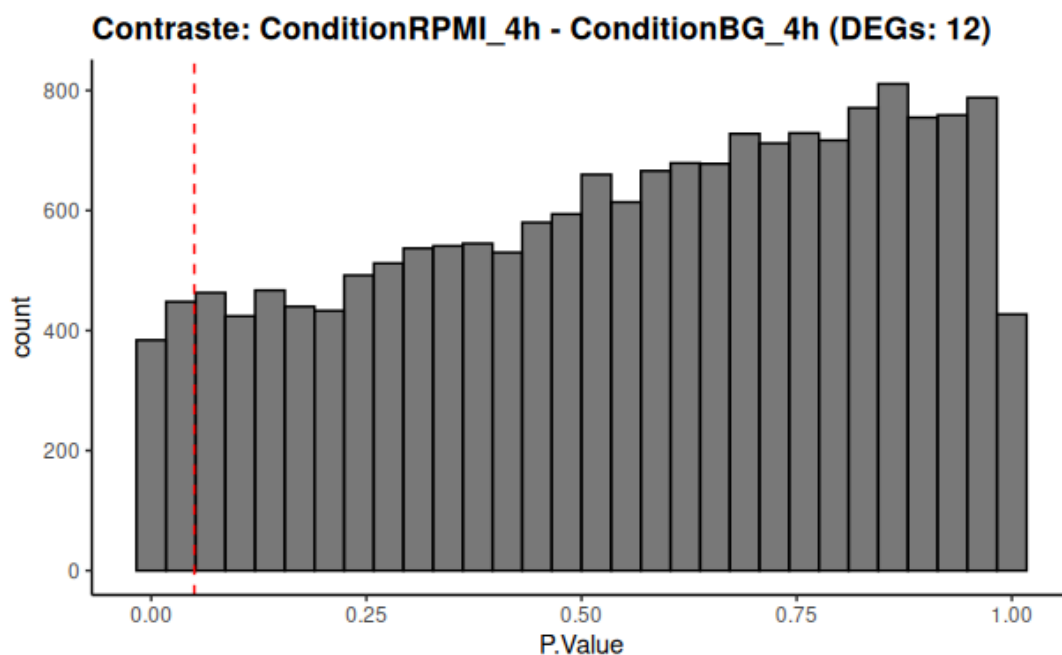


Figura 15. Distribución de p-valores para el contraste de ConditionRPMI_4h - ConditionBG_4h. Los que son no significativos tienen una distribución parecida

Gráficos de Volcán: Es una herramienta visual clave para evaluar los genes diferencialmente expresados (DEGs) en función de dos factores: la magnitud del cambio en la expresión génica (logFC) y su significancia estadística (adj.P.Val). Uno de los gráficos de los contrastes con p-valores más significativos es este:

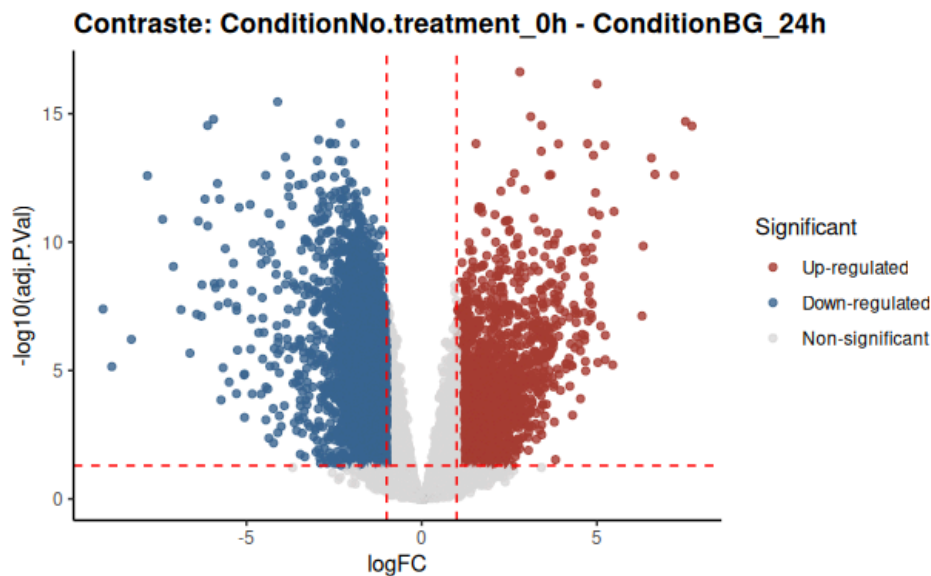


Figura 16. Gráfico de volcano con el contraste de ConditionNO.treatment_0h - ConditionBG_24h. En la leyenda se observa: up-regulated que significa genes con $\log FC \geq 1$ y p-valor ajustado ≤ 0.05 (más expresados en la condición experimental; down-regulated, son genes con $\log FC \leq -1$ y p-valor ajustado ≤ 0.05 (menos expresados en la condición experimental); y non-significant, genes que no cumplen con los criterios anteriores, es decir, $\log FC$ entre -1 y 1 o p-valor ajustado > 0.05 .

Boxplots: representan los genes más significativos en cada contrastes, uno de los gráfico es este:

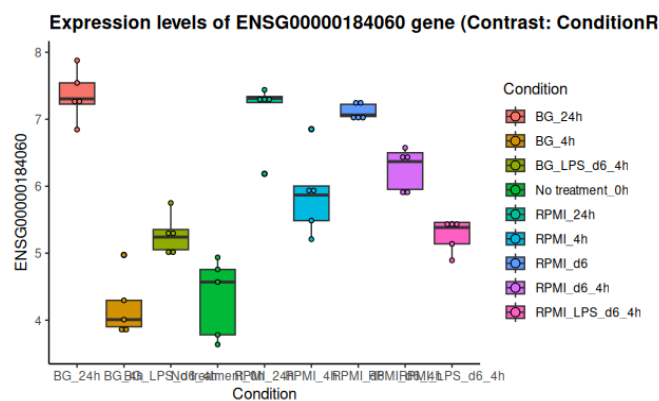


Figura 17. Boxplot del gen ENSG00000125740 para el contraste de ConditionRPMI_4h - ConditionBG_4h. Se observa que se encuentra muy diferenciado para la condición de BG_24h

2.7 Análisis de enriquecimiento

Este análisis permite agrupar genes en función de su función biológica, ayudando a interpretar los genes diferencialmente expresados (DEGs). Se van a utilizar las siguientes bases de datos:

- **KEGG**: información sobre rutas metabólicas y de señalización celular
- **Go terms**: funciones biológicas, moleculares y celulares.
- **MSigDB**: agrupa genes en conjunto funcionales clave para procesos biológicos

Primero se realiza una anotación de genes para que los identificadores de los genes desde ENSEMBL a SYMBOL y, seguido, se ejecuta el enriquecimiento funcional con las tres bases de datos mencionadas

Análisis de sobrerrepresentación (ORA)

He escogido el ORA mejor que GSEA (análisis de enriquecimiento de conjuntos de genes) porque:

- **Más fácil de usar**: necesitas un umbral de significancia para identificar los genes relevantes
- **Menos datos** necesarios: funciona bien con conjuntos de datos pequeños o cuando solo unos pocos genes son significativo
- **Más rápido**: menor demanda computacional y es más rápido en comparación con GSE
- **Más fácil de interpretar**: resultados directos y fáciles de entender si estás buscando cambios claros en la expresión génica
-

Este análisis identifica temas biológicos o vías significativamente enriquecidas en un conjunto de genes. Se enfoca en conjuntos de genes de bases de datos y compara las frecuencias observadas de los genes con lo que se esperaría por azar. Se realiza gracias al paquete [clusterProfiler](#). Se va crear un data frame por cada contraste en el que viene de qué proceso biológico vienen los genes diferenciados y un barplot para visualizarlo mejor.

Según los contrastes y las bases de datos se diferencian los siguientes genes:

- **GO terms**
 - ConditionNo.treatment_0h - ConditionBG_24h: genes responsables de las anfetaminas
 - ConditionRPMI_4h - ConditionBG_4h: genes que intervienen en la regulación positiva de la vía de señalización mediada por lipopolisacáridos
 - ConditionRPMI_d6_4h - ConditionRPMI_LPS_d6_4h: genes que intervienen en la regulación de la diferenciación de las células receptoras auditivas del oído interno
 - ConditionRPMI_LPS_d6_4h - ConditionBG_LPS_d6_4h: genes que tienen que ver con procesos metabólicos del urato
 -

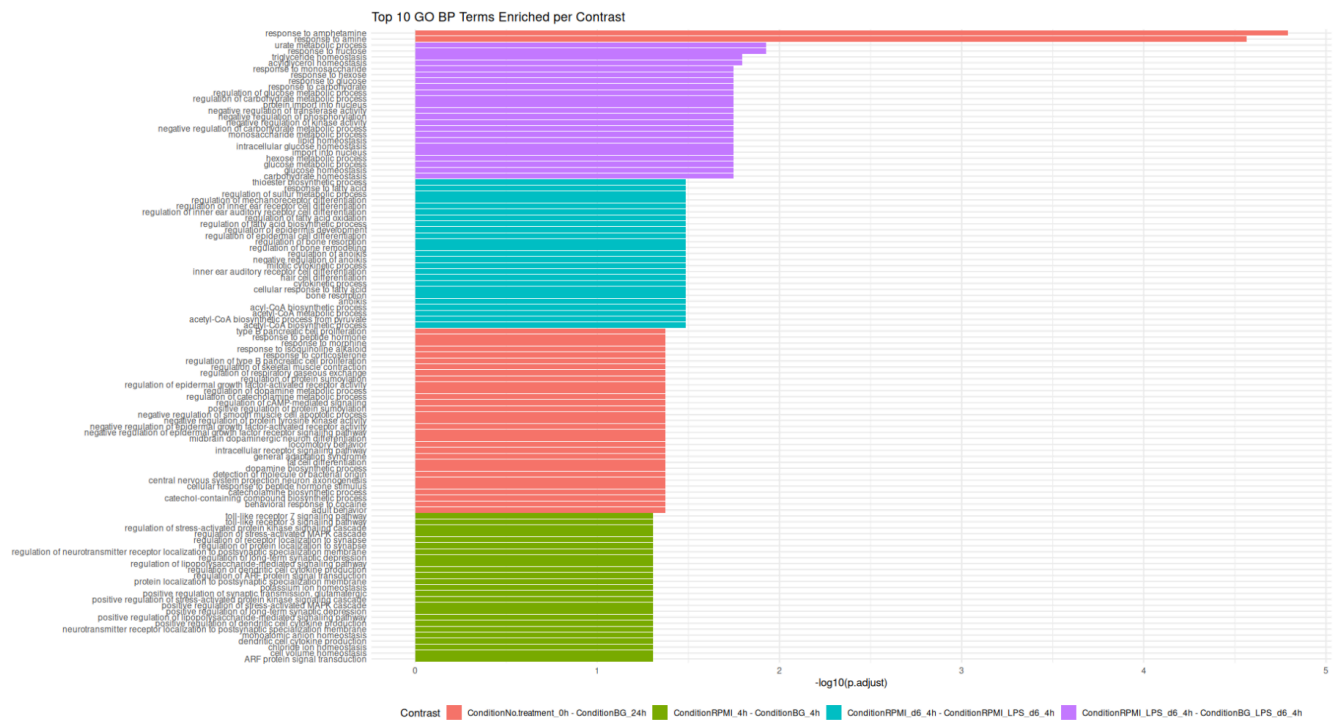


Figura 18. Barplot de los genes diferencialmente expresados según sus procesos biológicos en diferentes contrastes con la base de datos de GO terms

- **KEGG**

- ConditionNo.treatment_0h - ConditionBG_24h: genes responsables de la síntesis y secreción de la aldosterona
- ConditionRPMI_4h - ConditionBG_4h: genes que intervienen en la biosíntesis de estroides
- ConditionRPMI_d6_4h - ConditionRPMI_LPS_d6_4h: genes que tienen que ver con la endocitosis
- ConditionRPMI_LPS_d6_4h - ConditionBG_LPS_d6_4h: genes responsables en la cardiomiopatía diabética

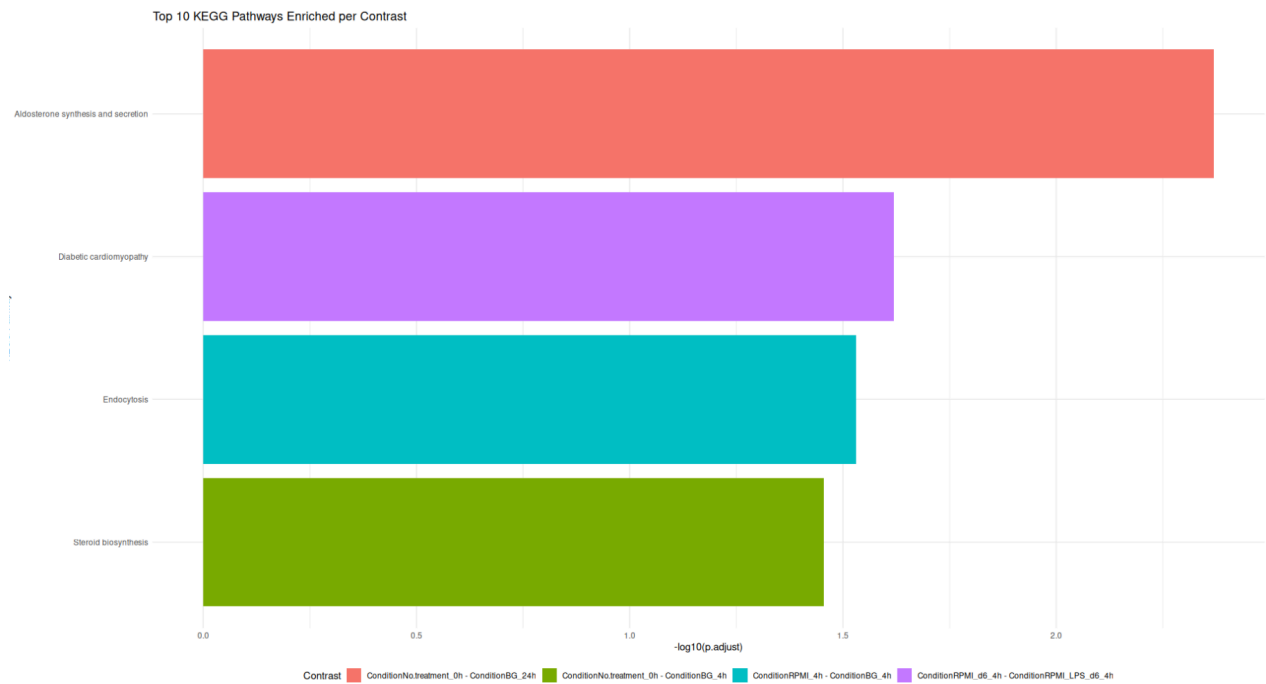


Figura 19. Barplot de los genes diferencialmente expresados según sus procesos biológicos en diferentes contrastes con la base de datos de KEGG

- **MSigDB**

- ConditionNo.treatment_0h - ConditionBG_24h: genes que intervienen en la señalización TNF-alfa a través de NFkB
- ConditionNo.treatment_0h - ConditionBG_4h: genes que intervienen en la señalización de IL2 STAT5
- ConditionRPMI_4h - ConditionBG_4h: genes del metabolismo xenobiótico
- ConditionRPMI_d6_4h - ConditionRPMI_LPS_d6_4h: genes que regulan la fosforilación oxidativa
- ConditionRPMI_LPS_d6_4h - ConditionBG_LPS_d6_4h: genes del metabolismo xenobiótico

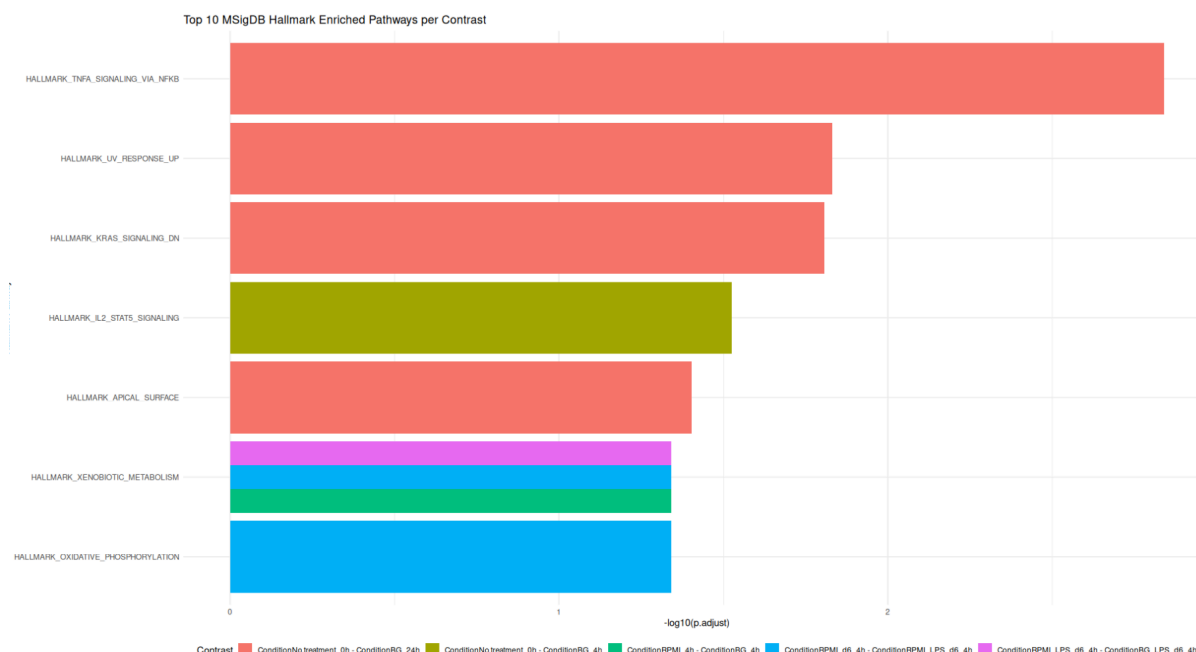


Figura 20. Barplot de los genes diferencialmente expresados según sus procesos biológicos en diferentes contrastes con la base de datos de MSigDB

3. Conclusiones

El análisis demuestra que la **condición biológica con mayor impacto** en la expresión génica es la exposición a **BG** a lo largo del tiempo, especialmente a las **24 horas**. Esta condición genera la mayor cantidad de genes diferencialmente expresados y muestra patrones claros en el análisis de PCA y enriquecimiento funcional.

Los **principales cambios inducidos por la estimulación** incluyen alteraciones en vías metabólicas, respuestas inmunológicas y procesos de señalización celular. Se observa que **BG** activa la respuesta inflamatoria y afecta rutas relacionadas con la regulación del metabolismo y la señalización inmune. Además, la exposición a **LPS** genera cambios específicos en la expresión de genes involucrados en la homeostasis celular y la respuesta al estrés.

A **nivel funcional**, los genes diferencialmente expresados están implicados en procesos clave como la biosíntesis de esteroides, la señalización de citoquinas (IL2-STAT5, TNF-alfa/NFkB), y el metabolismo de compuestos esenciales como el urato. En resumen, estos resultados sugieren que la **respuesta a los estímulos varía según el tipo de tratamiento y el tiempo de exposición**, con efectos más marcados en rutas de señalización inflamatoria y regulación metabólica.

4. Métodos – Tabla de recursos clave

Para ver mejor la tabla, adjunto el link:

<https://drive.google.com/file/d/11n7TJ24phNJBt13cJIAB-w5ML1xqeqBy/view?usp=sharing>

SOFTWARE AND ALGORITHMS				
Nº	Librerías	Descripción	Versión	URL
1	dplyr	Manipulación eficiente de datos	1.0.10	dplyr.tidyverse.org
2	stringr	Manipulación de cadenas de texto	1.4.0	stringr.tidyverse.org
3	ggplot2	Creación de gráficos visuales	3.4.0	ggplot2.tidyverse.org
4	ComplexHeatmap	Creación de mapas de calor complejos	2.10.0	jokergoo.github.io
5	ggpubr	Facilita la creación de gráficos con ggplot2	0.4.0	ggpubr.rstudio.com
6	RColorBrewer	Proporciona paletas de colores	1.1-2	cran.r-project.org
7	edgeR	Análisis de datos RNA-seq, especialmente para datos de conteo	3.38.4	bioconductor.org
8	limma	Análisis de expresión diferencial para datos transcriptómicos	3.54.1	bioconductor.org
9	org.Hs.eg.db	Anotación de genes humanos	3.17.0	bioconductor.org
10	msigdb	Conjuntos de genes y anotaciones para análisis de enriquecimiento	7.4	gsea-msigdb.org
11	clusterProfiler	Herramientas para análisis de enriquecimiento y visualización	4.4.4	yulab-smu.top
12	fgsea	Análisis de enriquecimiento de conjuntos basado en estadística acumulada	1.20.0	bioconductor.org
13	biomaRt	Acceso a bases de datos biológicas para anotación de genes	2.50.0	bioconductor.org
14	AnnotationDbi	Herramientas para interactuar con bases de datos de anotaciones	1.58.0	bioconductor.org
15	grid	Paquete base para la creación y manipulación de gráficos en R	4.4.2	rdocumentation.org
16	gridExtra	Extensión del paquete grid, mejora la personalización de gráficos	2.3	cran.r-project.org