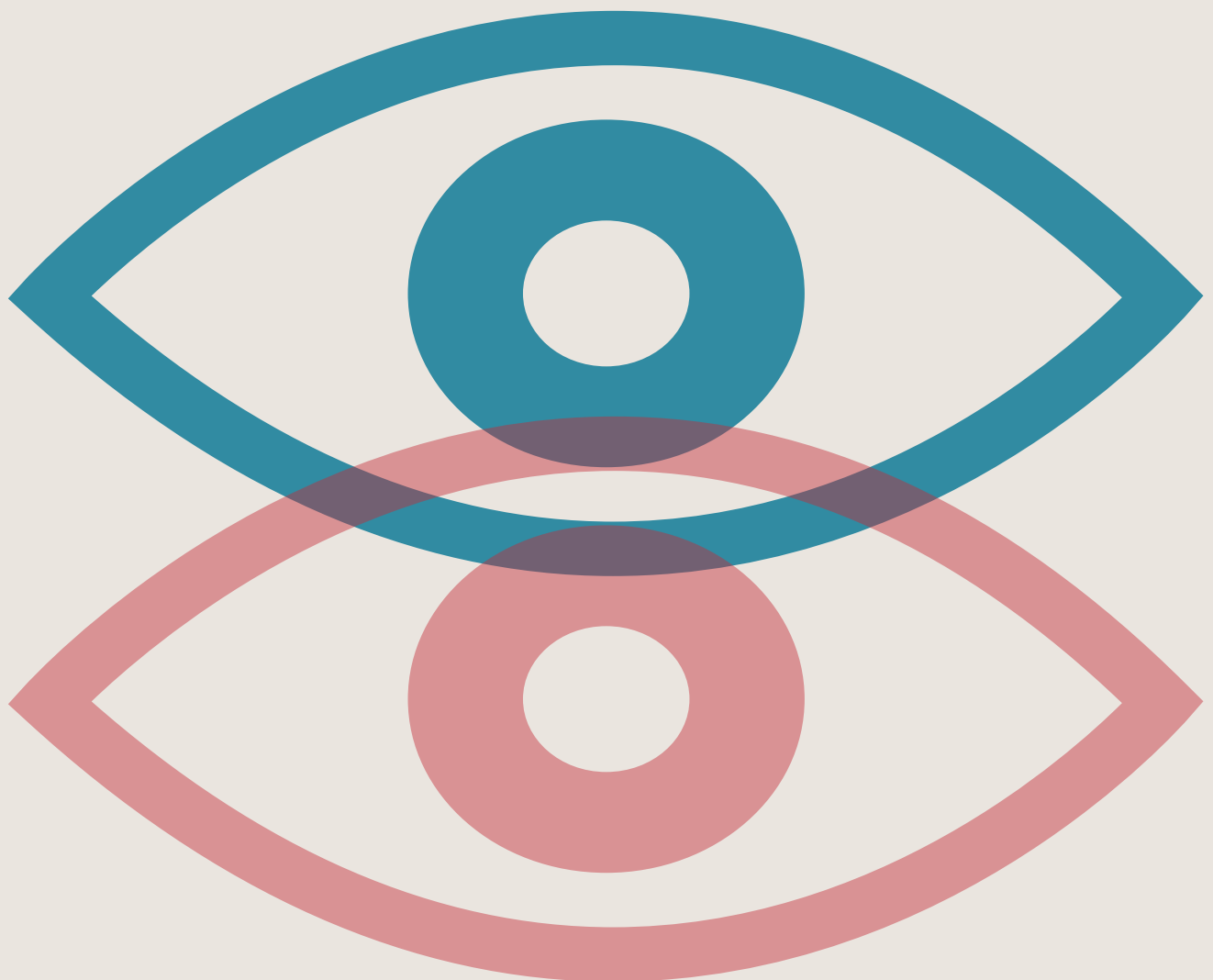


Estudios In Silico en Biomedicina

# **Revisión sistemática de la retinosis pigmentaria en humanos**

PILAR LÓPEZ FERNÁNDEZ



# Índice

1. Introducción.....	3
1.1 Contexto y justificación del trabajo.....	3
1.2 Objetivos del trabajo y planificación.....	4
2. Material y métodos.....	6
2.1 Literatura revisada y selección de los estudios.....	6
2.2 Estrategia de análisis bioinformático.....	6
2.2.1 Descarga de datos.....	8
2.2.2 Mapeo de las reads, cuantificación y control de calidad.....	9
2.2.3 Preprocesado de datos y análisis de la expresión diferencial.....	10
2.2.4 Anotación funcional.....	10
3. Resultados.....	12
3.1 Revisión de las bases de datos y análisis exploratorio.....	12
3.1.1 Calidad de las secuencias.....	14
3.1.2. Mapeo.....	15
3.1.3. Cuantificación y análisis exploratorio.....	16
3.2. Análisis de la expresión diferencial.....	18
3.3 Análisis funcional.....	20
4. Discusión.....	23
5. Conclusiones.....	25
6. Bibliografía.....	26
7. Material Suplementario.....	30

# 1. Introducción

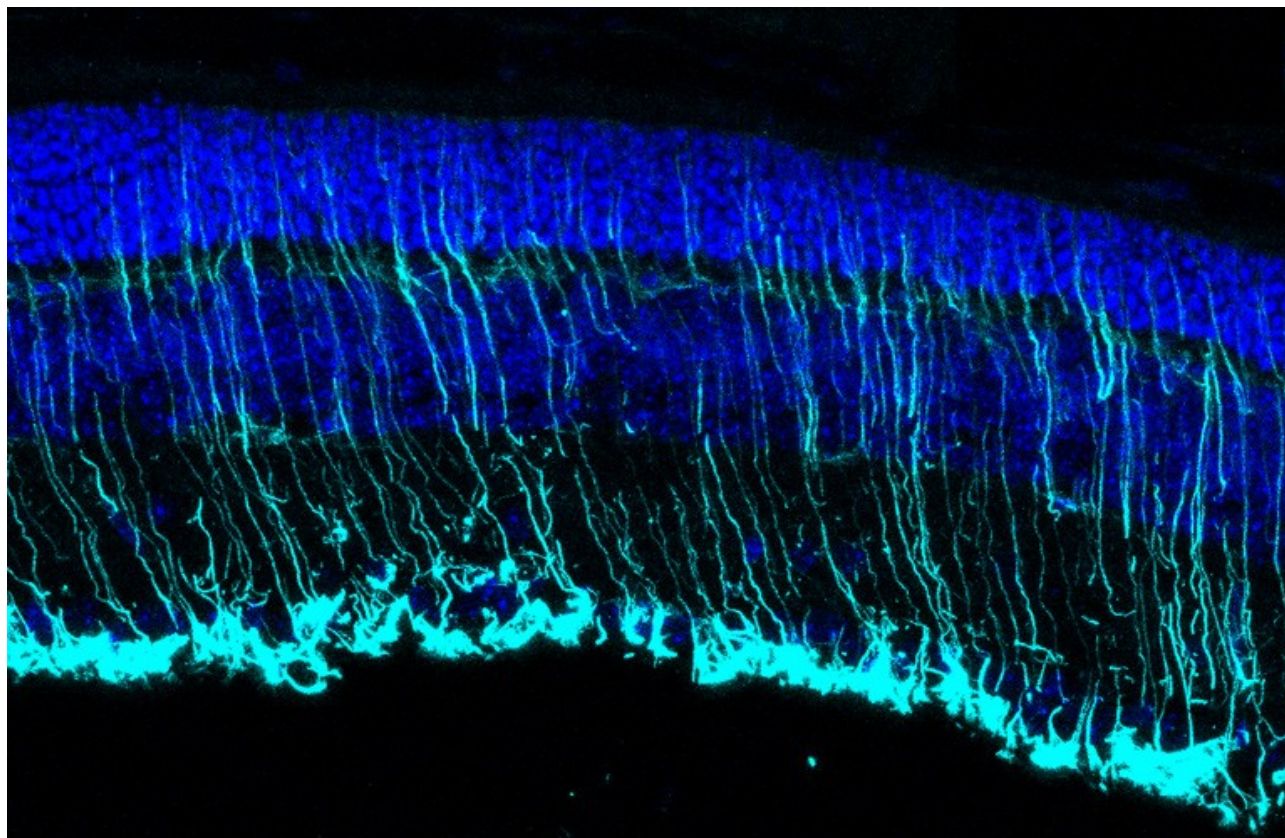
## 1.1 Contexto y justificación del trabajo

La **retinosis pigmentaria (RP)** es una rara enfermedad que se caracteriza por la degradación de los fotorreceptores y las células pigmentarias de la retina (RPE), que finalmente causa pérdida de la visión. RP es, a nivel genético, **altamente heterogénea** (con aproximadamente 80 genes relacionados con la enfermedad identificados hasta la fecha), donde además puede variar el tipo de herencia (ligada o no al sexo, dominante, recesivo, oligogénico o mitocondrial), y la mutación responsable (se han notificado unas 3000 distintas) (Newton & Megaw, n.d.).

Sin embargo, actualmente no hay tratamientos terapéuticos para curar esta patología o, al menos, reducir su progresión. La falta de conocimiento y el desarrollo de terapias se debe, en gran medida, a que los modelos animales no representan fielmente la enfermedad en humanos, que junto con la elevada heterogeneidad ocasiona que la patología se encuentre escasamente caracterizada (Arzalluz-Luque et al., 2021).

Una de las mutaciones con mayor prevalencia es la mutación del factor de procesamiento de pre-mRNA PRPF8, implicada en el *splicing* alternativo, que constituye la segunda causa más común de RP autosómica dominante (adRP) (Arzalluz-Luque et al., 2021). Recientemente, además, se ha relacionado la progresión de la enfermedad con la mutación m.3243A>G, una mutación en el genoma mitocondrial que, además de la degeneración de la retina, se ha asociado a otros trastornos clínicos, desde diabetes a encefalopatía (Chung et al., 2021).

A pesar de que varias terapias experimentales han avanzado a ensayos clínicos (<https://clinicaltrials.gov/ct2/results?term=retinitis+pigmentosa&Search=Search>), hoy en día no hay terapias preventivas ni tratamiento, dada su complicada etiología genética (Sánchez-Cruz et al., 2018). Por ello, es importante conocer e identificar aquellos procesos y mecanismos patológicos comunes, **independientemente de la mutación responsable**. Esto podría beneficiar, no solo a los a los pacientes diagnosticados con RP, sino con otras distrofias retinianas que engloban un espectro tan complejo de enfermedades.



**Figura 1:** Sección de la retina de un modelo de ratón para la Retinosis Pigmentaria. En cyan, las células de la glia de Müller y los astrocitos; en azul, los núcleos. Extraído de Sánchez-Cruz et al., 2018.

Con este propósito, la investigación se dirige en área de la bioinformática como una ciencia integradora en el ámbito de la biología molecular y genética con la ardua tarea de analizar gran cantidad de datos mediante las tecnologías y herramientas computacionales, que ha permitido el gran avance en el conocimiento de las ciencias ómicas.

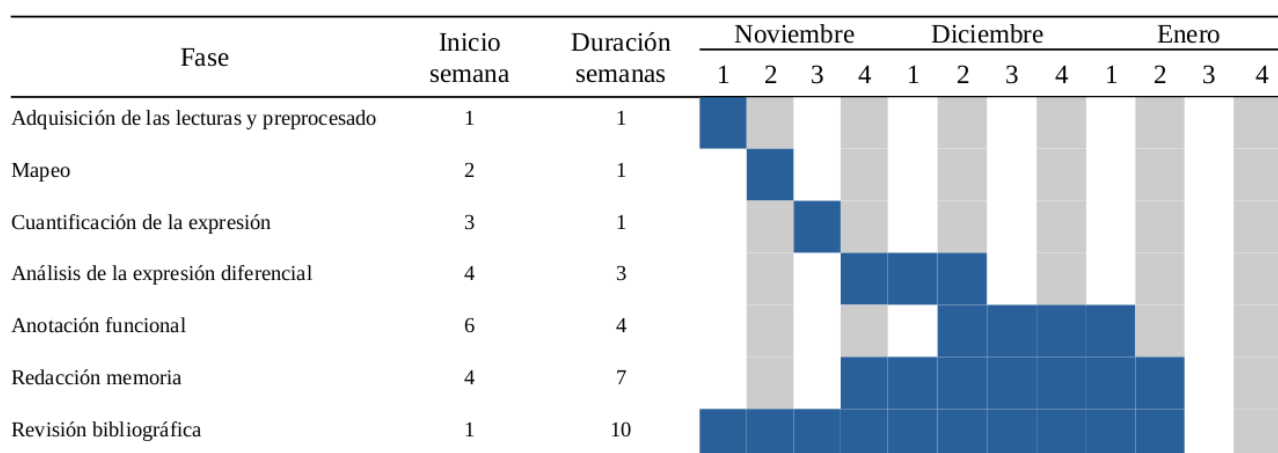
El objetivo de este análisis es, por tanto, estudiar aquellos genes diferencialmente expresados y estudiar el papel de estas mutaciones en la RE, como marco hacia el diseño de nuevas oportunidades terapéuticas para una patología tan compleja.

## 1.2 Objetivos del trabajo y planificación

El objetivo principal de este trabajo es identificar aquellos procesos biológicos y mecanismos moleculares asociados a la retinosis pigmentaria, mediante el estudio de aquellos genes que presentan una expresión diferencial significativa entre los grupos de individuos contrastados.

Los recursos computacionales utilizados, detallados en el apartado de Material y Métodos a continuación, se han utilizado para realizar las siguientes tareas:

1. Adquisición y preprocesado de las lecturas
2. Mapeo contra genoma de referencia
3. Cuantificación de la expresión génica
4. Análisis de la expresión diferencial
5. Anotación funcional



**Figura 2:** Cronograma de las tareas realizadas en el análisis.

## **2. Material y métodos**

### **2.1 Literatura revisada y selección de los estudios**

La búsqueda se inició introduciendo el término “*retinitis pigmentosa*” en la base de datos *Gene Expression Omnibus* (GEO) (Barrett et al., 2013). Estos estudios fueron filtrados por: i) organismo: “*Homo sapiens*”, ii) tipo de estudio: “*expression profiling by high throughput sequencing*” y iii) número de muestras: mínimo de 12.

Se descartaron aquellos repositorios de datos de *single cell*, diseños experimentales distintos a caso vs control (por ejemplo, series temporales) y estudios no basados en retinosis pigmentaria u otras alteraciones de la visión. No se tuvo en cuenta, entre los criterios de exclusión, la fecha del repositorio ni características de los individuos como el sexo o la edad.

### **2.2 Estrategia de análisis bioinformático**

La tecnología de secuenciación RNA-seq permite cuantificar y estimar los perfiles de expresión génica con el objetivo, usualmente, de detectar las diferencias significativas de expresión de los genes en distintos grupos. El uso de este tipo de tecnología está ampliamente extendido y, si bien no existe un único pipeline a seguir, hay una gran cantidad de aspectos a tener en cuenta y de los cuáles dependerá la elección de unas aplicaciones y unos algoritmos de análisis frente a otros. Entre estas consideraciones se encuentran: el diseño experimental de partida, objetivos principales del estudio, la existencia previa de un genoma de referencia o las propias capacidades del equipo.

Cada uno de los potenciales escenarios presenta métodos óptimos diferentes para la cuantificación, la normalización y el análisis de la expresión diferencial. En la Tabla 1 (página siguiente), se ha recogido una breve descripción de algunos de los software más usados en este trabajo:

## *Revisión sistemática de la retinosis pigmentaria en humanos*

**Tabla 1:** Soluciones bioinformáticas para el análisis de RNA-seq

<b>Pasos de análisis</b>	<b>Programa</b>	<b>Descripción</b>
<b>Adquisición de datos</b>	SRA tool kit ( <i>Toolkit Documentation : Software : Sequence Read Archive : NCBI/NLM/NIH</i> , n.d.)	Control de calidad de secuencias, búsqueda de anomalías en las lecturas de la librería
	Fastqc (Van Verk et al., 2013)	Creación de informes sobre el control de calidad
<b>Control de calidad</b>	MultiQC (Ewels et al., 2016)	Filtrado y eliminación de adaptadores de las lecturas
	Cutadapt (Saeidipour & Bakhshi, 2013)	Alineamiento de las lecturas obtenidas por RNA-seq sobre el genoma de referencia. Eficiente en coste de memoria y tiempo de cómputo.
<b>Filtrado y eliminación de adaptadores</b>	HISAT2 (Kim et al., 2015)	Control de calidad de secuencias, búsqueda de anomalías en las lecturas de la librería
	Sam Tools (Li et al., 2009)	Conjunto de herramientas para procesar alineamientos de lectura de secuencias de DNA
<b>Alineamiento</b>	HTSeq count (Anders et al., 2015)	Cuantificación de la expresión génica
	DeSeq2 (Love et al., 2014)	Método estadístico de evaluación de la expresión génica diferencial entre muestras. Distribución binomial negativa para controlar la sobre dispersión. Utiliza un Fold Change ajustado y estimaciones del error estándar, mejorando la detección de los DEGs.
<b>Cuantificación</b>		
<b>Detección DEGs</b>		

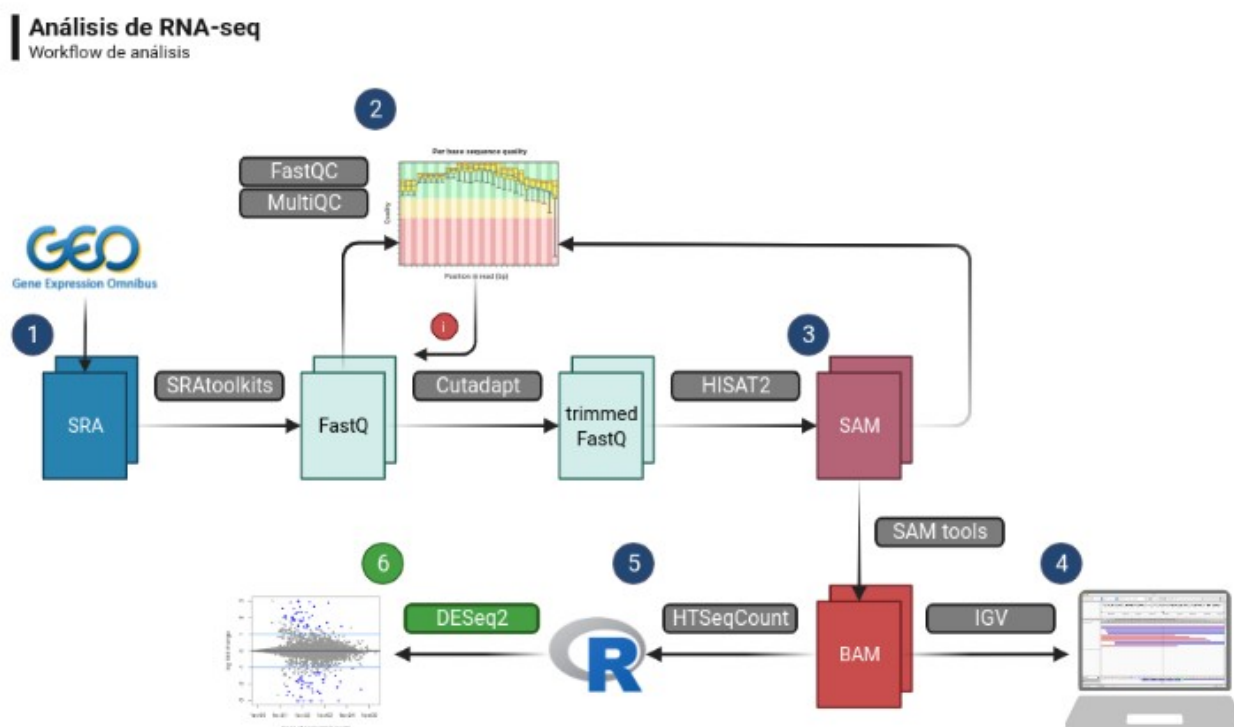
Para todos los estudios escogidos se ha seguido el mismo proceso de análisis: i) adquisición de los datos, control de calidad y normalización, ii) análisis exploratorio, iii) análisis de la expresión génica diferencial, iv) análisis del enriquecimiento funcional.

Se ha usado la versión 4.1.2 de RStudio (*R Core Team (2019) R A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.* -

References - Scientific Research Publishing, n.d.) en el análisis. Además, se ha recogido la totalidad de los paquetes y librerías usadas en la Tabla S1, en el apartado Material Suplementario.

### 2.2.1 Descarga de datos

Tras la elección de los estudios, la obtención de los ficheros fastq a partir de los proyectos SRA (Sequence Read Archive) desde el repositorio de datos GEO se realizó con *prefetch* y *fasterq-dump* del paquete *NCBI SRA Toolkit* (Toolkit Documentation : Software : Sequence Read Archive : NCBI/NLM/NIH, n.d.). Los scripts que contienen todos los parámetros de ejecución se encuentran disponibles en [https://github.com/PilarLF/Estudios\\_insilico](https://github.com/PilarLF/Estudios_insilico).



**Figura 3:** Esquema general de los pasos a seguir en el análisis de datos de RNA-seq. 1) Descarga de los datos crudos del repositorio GEO. 2) Control de calidad de las lecturas .fastq y creación de los informes con MultiQC. Si tras filtrar y recortar las lecturas con Cutadapt el control de calidad ofrece un balance negativo, se repite el control. 3) Alineamiento de las lecturas frente al genoma de referencia. 4) Visualización de las lecturas mapeadas contra el genoma. 5) Estimación de la abundancia de los conteos. 6) Análisis de la expresión diferencial a nivel génico mediante DESeq2.



### ***2.2.2 Mapeo de las reads, cuantificación y control de calidad***

El control de calidad de las *reads* obtenidas se ha realizado con FastQC (Van Verk et al., 2013) (ya que las lecturas proceden de la plataforma Illumina), que genera un informe *html* que informa de las principales estadísticas sobre la calidad de las lecturas de secuenciación de cada muestra. Este paso nos permite evaluar si nuestros datos presentan algún problema, ocasionado durante la preparación de la librería o en la reacción de secuenciación, antes de proseguir con el análisis, y aporta información sobre el contenido de G-C de la muestra, cantidad de bases catalogadas como “N” o el número de secuencias sobre representadas, entre otros parámetros. A continuación, los informes *html* obtenidos para cada muestra del análisis se han reunido en uno solo mediante MultiQC (Ewels et al., 2016), generando las estadísticas, en conjunto, de las muestras de cada estudio, lo que nos permite obtener una visión global de los datos de análisis.

Tras evaluar si los parámetros se encuentran dentro de los estándares esperados, se ha eliminado los adaptadores de las secuencias mediante Cutadapt v3.2 (Martin, 2011). En este punto, es opcional volver a realizar el análisis de calidad para evaluar de nuevo las estadísticas de los datos.

El alineamiento frente al genoma de referencia humano se ha realizado mediante HISAT2 v2.2.1 (Kim et al., 2015). Si bien hasta la fecha algunos de los alineadores más habituales eran TopHat (Trapnell et al., 2009) y Bowtie2 (Langmead et al., 2009), hoy se encuentran en bajo mantenimiento, ya que han sido reemplazados en gran medida por HISAT2 que, con la misma funcionalidad, proporciona mayor precisión y es mucho más eficaz (Kim et al., 2015). Los archivos de entrada fueron en formato fastq, filtrados y recortados previamente con Cutadapt.

A continuación, se realiza la conversión de los ficheros SAM, obtenidos con HISAT2, a BAM, para poder ser usados como archivos de entrada de HTSeqCount v0.13.5-1 (Anders et al., 2015). Para cumplir con los argumentos necesarios para correr HTSeqCount, los ficheros BAM se ordenaron por nombre de lectura (por defecto se ordenan por coordenada cromosómica) y se indexaron, mediante las herramientas *sort* e *index*, respectivamente, del paquete SamTools (Li et al., 2009). Por último, se realiza la cuantificación de las lecturas mediante HTSeqCount.

### *2.2.3 Preprocesado de datos y análisis de la expresión diferencial*

El análisis de la expresión diferencial se realizó, en Rstudio, con el paquete DESeq2 (Love et al., 2014), a partir de los ficheros de cuantificación de las lecturas obtenidas por HTSeqCount y se generó la matriz de conteo a nivel de gen para cada estudio.

La variables, de tipo factor, que se han tenido en cuenta en todos los estudios es el grupo experimental al que pertenece el individuo (caso o control), sin considerar otros niveles.

Se han descartado los genes de menor expresión detectada mediante *rowSums* del paquete edgeR. Para ello, se mantienen los genes que presenten, como mínimo, 1 CPM (conteo por millón) en al menos cuatro muestras (ya que en ambos estudios trabajamos con 12 réplicas). Tras filtrar se han normalizado los datos de conteo mediante VST (*variance stabilizing transformation*) ofrecida por DESeq2. El análisis exploratorio de los datos (visualización de los niveles de expresión, análisis de los componentes principales y análisis por clustering) se realizó con el fin de identificar aquellas muestras que presentasen anomalías y posibles efecto batch.

Los contrastes se han realizado con DESeq2. Se ha usado el método Test de Wald (Basu et al., 2017) con la función *nbinomWaldTest()*, usado comúnmente en DESeq2, para comparar ambos grupos experimentales. Los p-valores se han ajustado mediante el método Benjamini & Hochberg (BH).

### *2.2.4 Anotación funcional*

Para detectar las funciones asociadas al fenotipo de los casos, se ha realizado un *gen set enrichment analysis* (GSEA), en Rstudio, mediante el paquete clusterProfiler (Yu et al., 2012). En un primer paso, se ordenan los genes según el valor del estadístico  $\log_2\text{FoldChange}$  obtenido en el análisis de la expresión diferencial. Después, se realiza el GSEA mediante la implementación de clusterProfiler en Rstudio. Dado que estamos analizando (Caso - Controles), un valor de  $\log_2\text{FC} < 0$  indicaría que esa función o ruta se encuentra infrarepresentado en casos, pero sobreexpresado en controles, y viceversa. Cuanto mayor sea su valor absoluto, mayor es el nivel de sobre o infrarepresentación (Català-senent et al., 2021). Dado que se trata de un análisis de la expresión a nivel de gen, GSEA es un buen método de integración, ya que usa todos los genes del estudio, que son ordenados por su nivel de expresión y no solo por su significatividad, como es el caso de mgdsa (Mullin et al., 2021; Ferrer et al., 2020). Las funciones con un p-valor ajustado  $> 0.05$  fueron consideradas significativas.

### ***Revisión sistemática de la retinosis pigmentaria en humanos***

La anotación funcional se obtuvo de las bases de datos *Gene Ontology Biological Processes* (GO-BP) (Ashburner et al., 2000) y *Kyoto Encyclopedia of Genes and Genomes* (KEGG) (Kanehisa & Goto, 2000).

### 3. Resultados

#### 3.1 Revisión de las bases de datos y análisis exploratorio

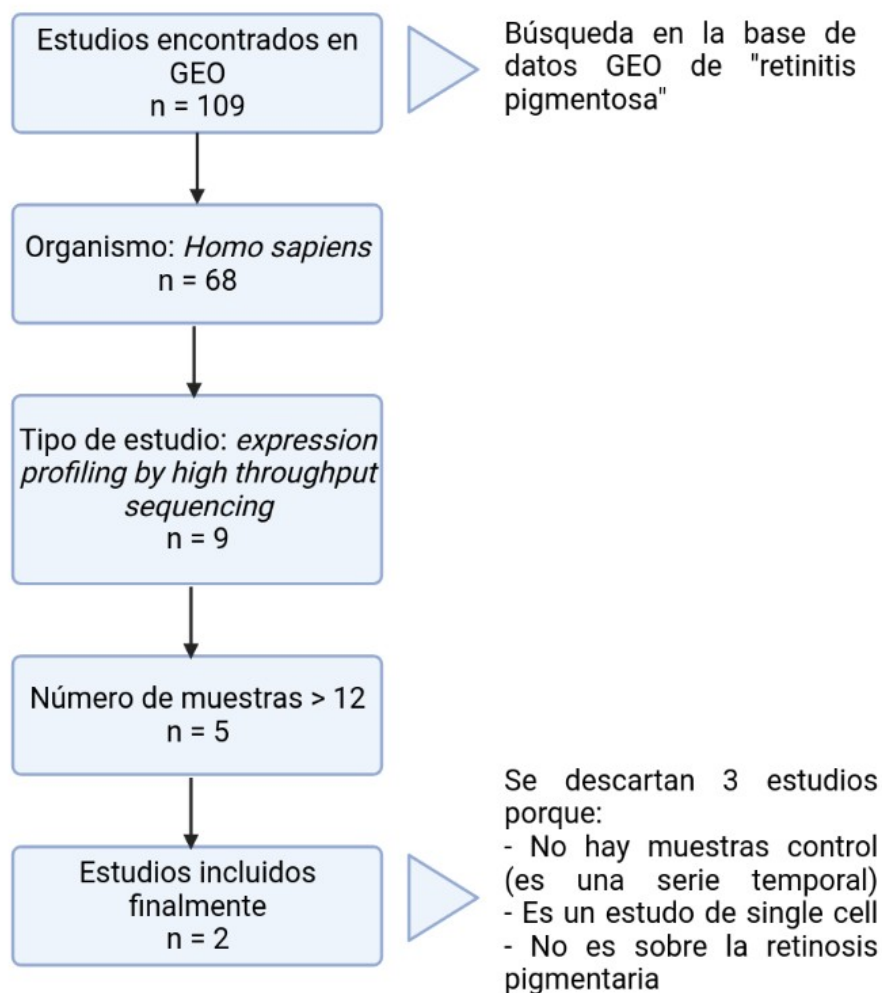
Tras incorporar los filtros ya mencionados en la base de datos GEO y, tras revisar los 9 repositorios resultantes, solo 2 han sido seleccionados para el análisis: un estudio de fibroblastos y otro estudio de fibroblastos y células madre pluripotenciales inducidas (iPSC) a RPE (células pigmentarias de la retina).

En total, ambos estudios contienen 6 pacientes (en ningún caso se tuvo en consideración la edad o el sexo de los individuos): 3 individuos sanos como control y 3 individuos diagnosticados con retinosis pigmentaria o con patologías relacionadas con la pérdida de la visión. Cabe destacar que los sujetos del estudio GSE175477 fueron escogidos porque presentaban, además de retinosis pigmentaria, otros desórdenes como diabetes, pérdida de oído, pérdida de memoria y bipolaridad. La distribución de las muestras se recoge en la Tabla 2. De cada muestra se tomaron tres réplicas (ver Figura 5).

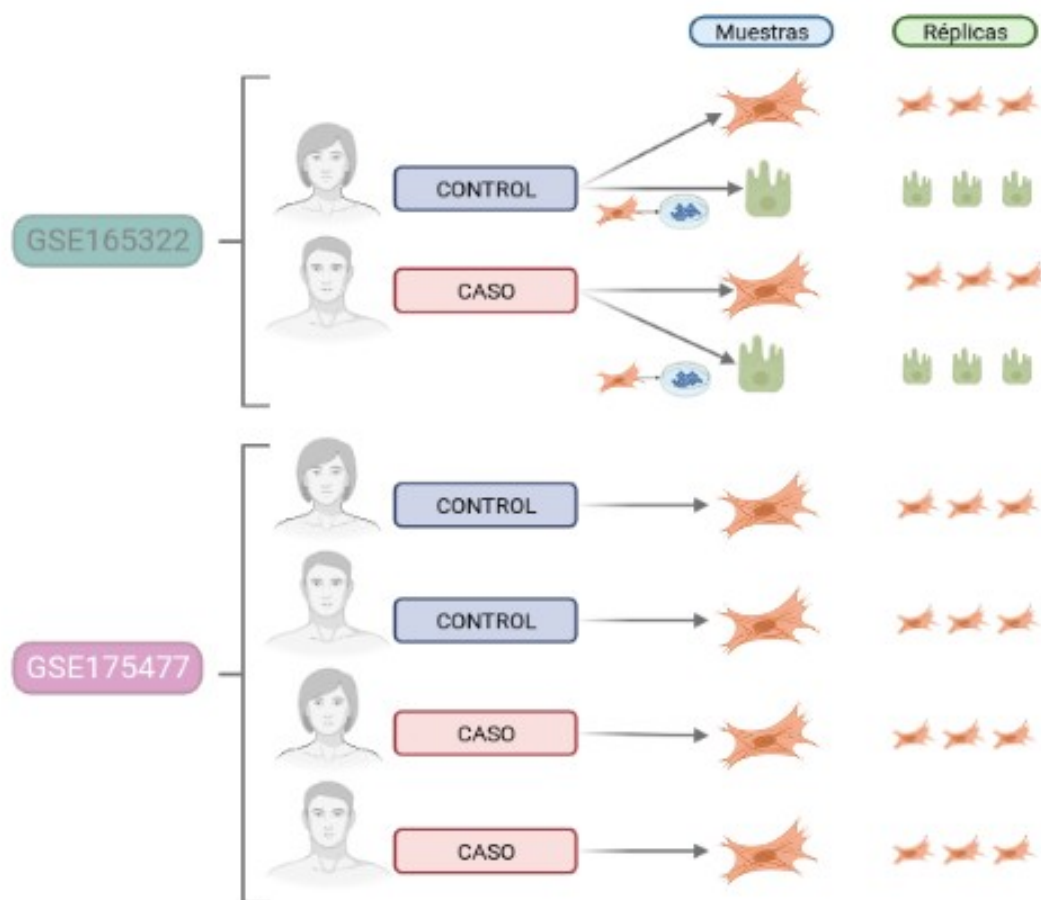
**Tabla 2:** Descripción de los estudios seleccionados. Se recoge la plataforma de secuenciación usada y el tipo celular usado para el análisis.

GEO acceso	Plataforma	Tejido/tipo celular	Referencia
GSE165322	Illumina HiSeq2500	Fibroblastos de la piel y RPE procedentes de iPSC	Arzalluz-Luque et al., 2021
GSE175477	Illumina MiSeq	Fibroblastos	

## Revisión sistemática de la retinosis pigmentaria en humanos



**Figura 4:** Diagrama del proceso de selección y cribado de los estudios revisados en la base de datos GEO.



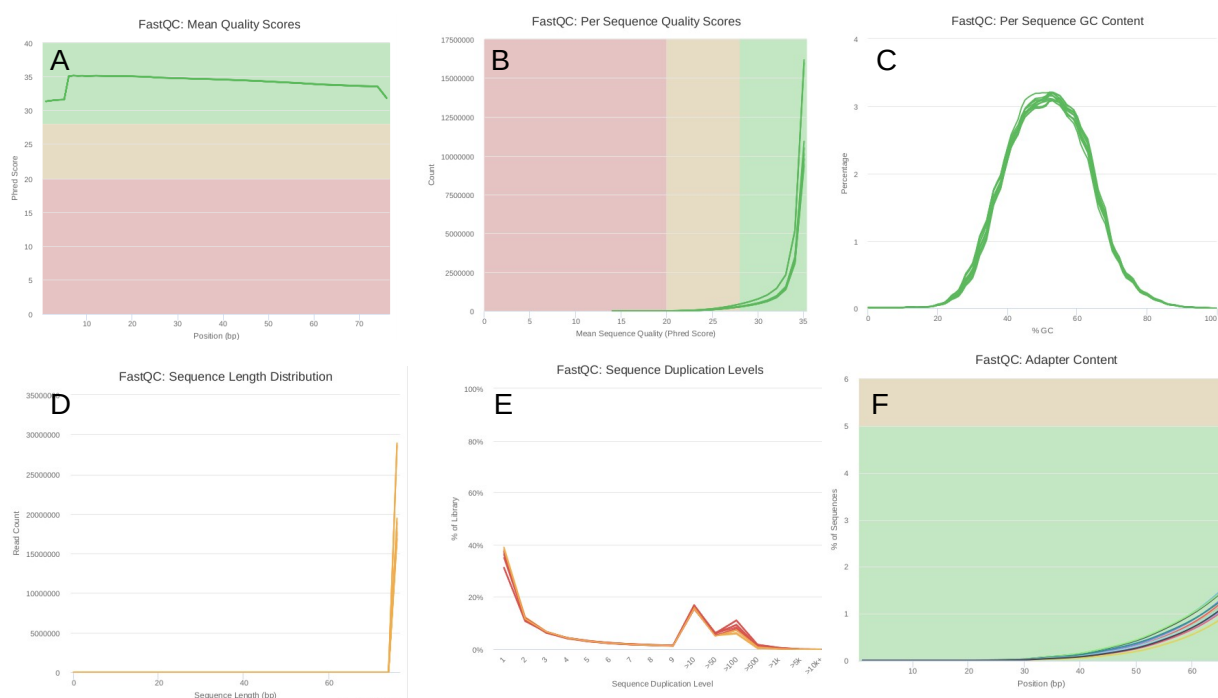
**Figura 5:** Visualización de los individuos de estudio, las muestras y las réplicas en cada grupo de estudio. Elaboración propia.

### 3.1.1 Calidad de las secuencias

En general, las lecturas muestran una buena calidad de secuencia (*Phred Score* por encima de 30 en todos los casos), una distribución normal del contenido en GC y no hay detección de secuencias sobrerrepresentadas. No obstante, el nivel de duplicación relativo en cada secuencia se ha determinado elevado, el 15,4% de las secuencias presenta 10 veces el mismo fragmento, y el 9,5% presentan el mismo fragmento más de 100 veces. Si bien es cierto que es inevitable cierto nivel de duplicación como consecuencia del propio análisis de RNA-seq, en términos generales, los principales módulos para evaluar la calidad de la secuencia presentan el símbolo de aprobación.

## Revisión sistemática de la retinosis pigmentaria en humanos

Por otra parte, el análisis mediante FastQC determinó la presencia de adaptadores Truseq de Illumina (3' AGATCGGAAGAG CACACGTCTGAACTCAGTCA), que son eliminados mediante Cutadapt. En la Figura 6 se recoge los resultados de algunos de los parámetros evaluados.



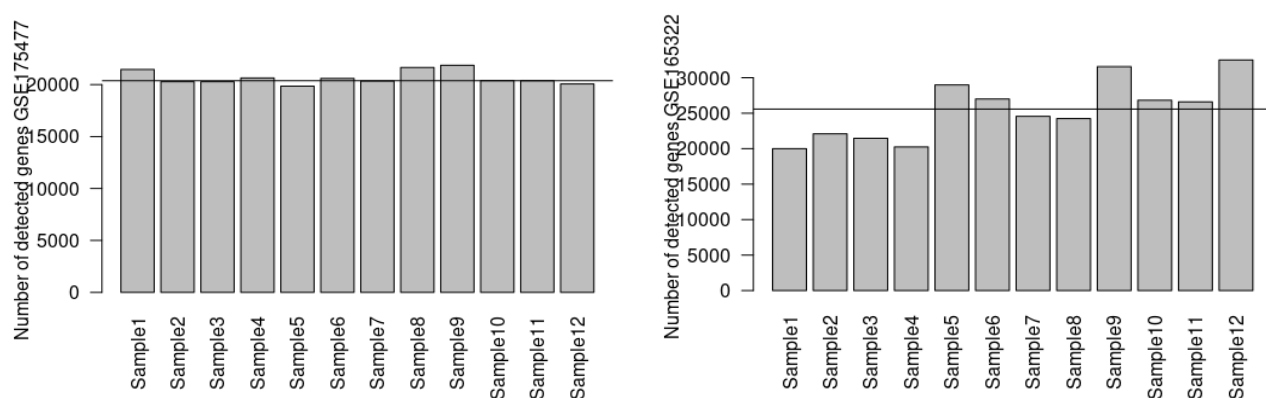
**Figura 6:** Evaluación de la calidad de las secuencias de RNA mediante FastQC y MultiQC para las secuencias del estudio GSE175477. A) Calidad media de las secuencias. B) Puntuación de calidad de cada secuencia. C) Contenido GC de cada secuencia. D) Distribución de la longitud de cada secuencia. E) Niveles de duplicación de las secuencias. F) Presencia de adaptadores.

### 3.1.2. Mapeo

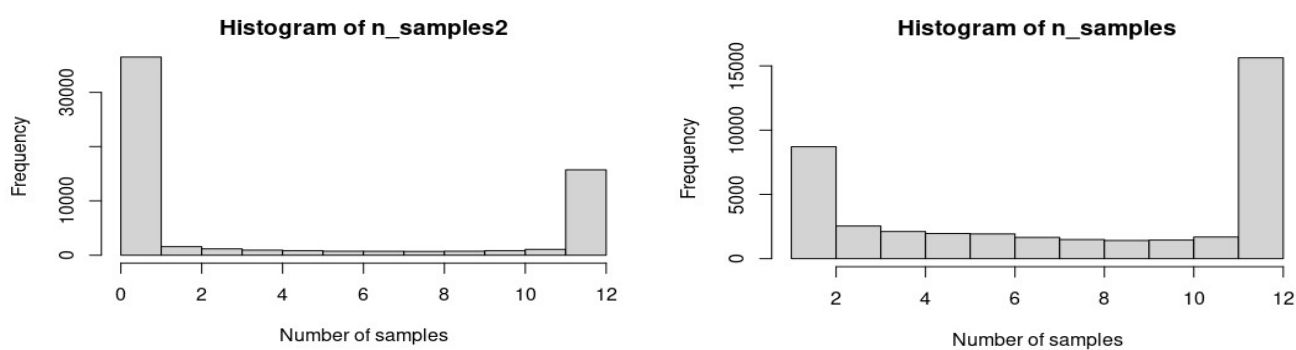
El mapeo de las lecturas se ha realizado con el genoma de referencia humano GRCh38.p13 mediante el programa HISAT2, y ha presentado una tasa de alineamiento del 97% para las lecturas de los estudios analizados.

### 3.1.3. Cuantificación y análisis exploratorio

Tras la cuantificación, llevada a cabo con HTSeqCount, se obtuvieron los archivos de cuantificación de expresión génica para cada una de las muestras. Estos datos fueron evaluados para asegurarnos de que no era necesario descartar ninguna muestra:



**Figura 7:** Número de genes detectados en cada una de las muestras de estudio. Con una línea horizontal se ha señalado la media de genes detectados en cada uno de los dos estudios. Estos diagramas de barras han sido generados mediante el paquete de R ggplot2.

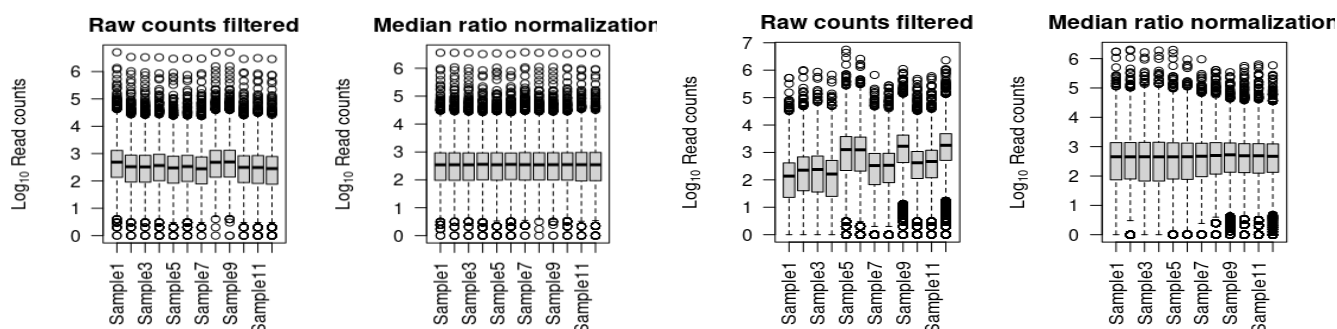


**Figura 8:** Distribución de la expresión de los genes de cada muestra. Como vemos, en GSE175477 (izquierda) la mayoría de genes no se ha detectado en ninguna de las 12 muestras. EN GSE165322 (derecha), ocurre el caso contrario.



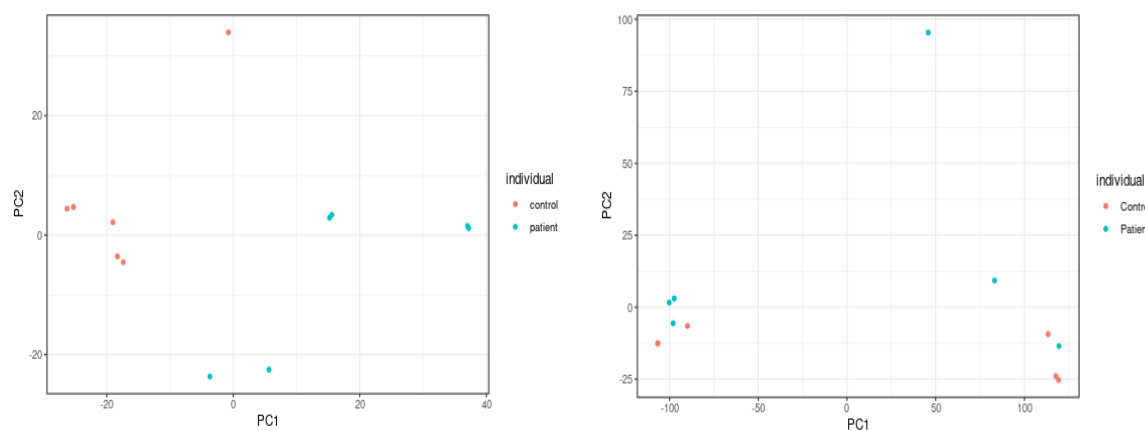
## Revisión sistemática de la retinosis pigmentaria en humanos

Con el objetivo de minimizar las diferencias entre muestras y normalizar con respecto al tamaño de la librería, la normalización de los datos de expresión se ha realizado con VST, ofrecida por DESeq2. Podemos ver los diagrama de cajas antes y después de normalizar:



**Figura 9:** Visualización mediante diagrama de cajas de los daatos de expresión génica para cada una de las 12 muestras de los dos estudios. A izquierda, GSE175477; a la derecha, GSE165322.

El análisis exploratorio mostró un buen comportamiento de las muestras en el estudio GSE175477. No obstante, detectó un efecto batch en GSE165322, que fue considerado y corregido mediante ComBat-seq (Zhang et al., 2020), del paquete *Surrogate Variable Analysis* (sva).



**Figura 10:** Análisis de los componentes principales como paso del análisis exploratorio. Se ha marcado por colores a qué grupo experimental pertenece cada muestra (salmón, controles; azul, pacientes). A la izquierda, el análisis de las muestras correspondientes al estudio GSE174577. A la derecha, GSE165322.

### 3.2. Análisis de la expresión diferencial

Como se ha descrito en el apartado de Material y métodos, para cada estudio se ha evaluado el contraste (Casos - Controles). El número de genes diferencialmente expresados se ha recogido en la siguiente tabla:

**Tabla 3: Número de genes diferencialmente expresados en cada estudio.** Se ha mostrado el total de genes analizados inicialmente, cuántos han pasado el filtro de expresión mínima y, por último, el número de genes considerados significativos en función del p-valor ajustado y log2FoldChange.

Paso	GSE175477 genes	GSE165322 genes
Crudos	61546	40543
Filtrar >1CPM	12651	15578
DGE genes p.ajustado < 0.05	3016	300
DGE genes p.ajustado < 0.05 y log2FC > 2 (en valor absoluto)	197	84

Podemos apreciar una gran diferencia entre ambos estudios. Los genes significativos y con un valor absoluto de log2FC superior a 2 se recogen en la tabla siguiente:

**Tabla 4: Genes significativos en cada estudio, tras los contrastes (Casos - Controles) respectivos.** En rojo, los genes sobreexpresados (log2FC > 2); en azul, infraexpresados (log2FC < -2).

Estudio	Genes significativos
<b>GSE175477</b>	<p> SYN1, SH2D2A, IBSP, APBA2, RAB27B, PTGER3, COL11A1, LIMCH1, ABCC9, MRV11, NTN4, SORBS1, IL11, MMP11, SUSP2, GGT5, RASD2, APOL1, CSF2RB, JAG1, PCSK1N, WFDC1, SFRP1, STMN2, TMEM176B, LGALS3BP, DTX4, KRT18, ADGRD1, BACH2, CDH6, PDE4D, IGFBP2, IGFBP5, PRG4, TNFSF4, TREH, GALNT12, TEK, KCNJ8, INHBA, SRGN, LYPD3, POM121L9P, MTUS1, BST2, GFPT2, KRT34, MYBPH, POSTN, BEX1, CD36, LMO2, IL6, ERMN, CCN3, IL33, SULF1, STRA6, BMPR1B, LGR5, HAPLN3, MFGE8, NKD1, ABCA8, GATA6, TINAGL1, MALL, CORIN, SFRP2, HAPLN1, SCUBE3, SLC22A3, CSGALNACT1, ANKRD1, LYPD6B, KCNE4, BMP6, PTPRD, DEPTOR, TIAM1, SST, NMNAT2, ACAN, ADAMTS4, CLIC6, IGF2BP1, SLC37A1, INHBB, IGFN1, IGFBP7, TMEM155, GASK1B, EDIL3, RNF180, SAMD3, PI16, FNDC1, TMEM130, CCDC68, PLEKHA7, SPINT2, PTGDR, PCDH7, TM4SF1, KCNK3, KRT19, SCG2, ISG20, PDE3A, C15orf53, A2M, PRR15, FOXL1, LMO7DN, HTR1F, C3orf80, GALNT9, CSMD1, ALDH1A3, KCND2, DRD1, SLC24A3, KRT14, LUZP2, PLA2G2A, HES4, RELN, CACNA1H, MFAP5, NMB, MBP, AKR1B10, ITGBL1, HIST2H2BF, MT-TL1, DIO2, KRTAP2-3, EBF2, KRTAP1-5, TEX26-AS1, AC073621.1, LINC01436, HLA-DPA1, AL357060.1, IL12A-AS1, LINC00942, AC091563.1, PKD1P5, </p>

## Revisión sistemática de la retinosis pigmentaria en humanos

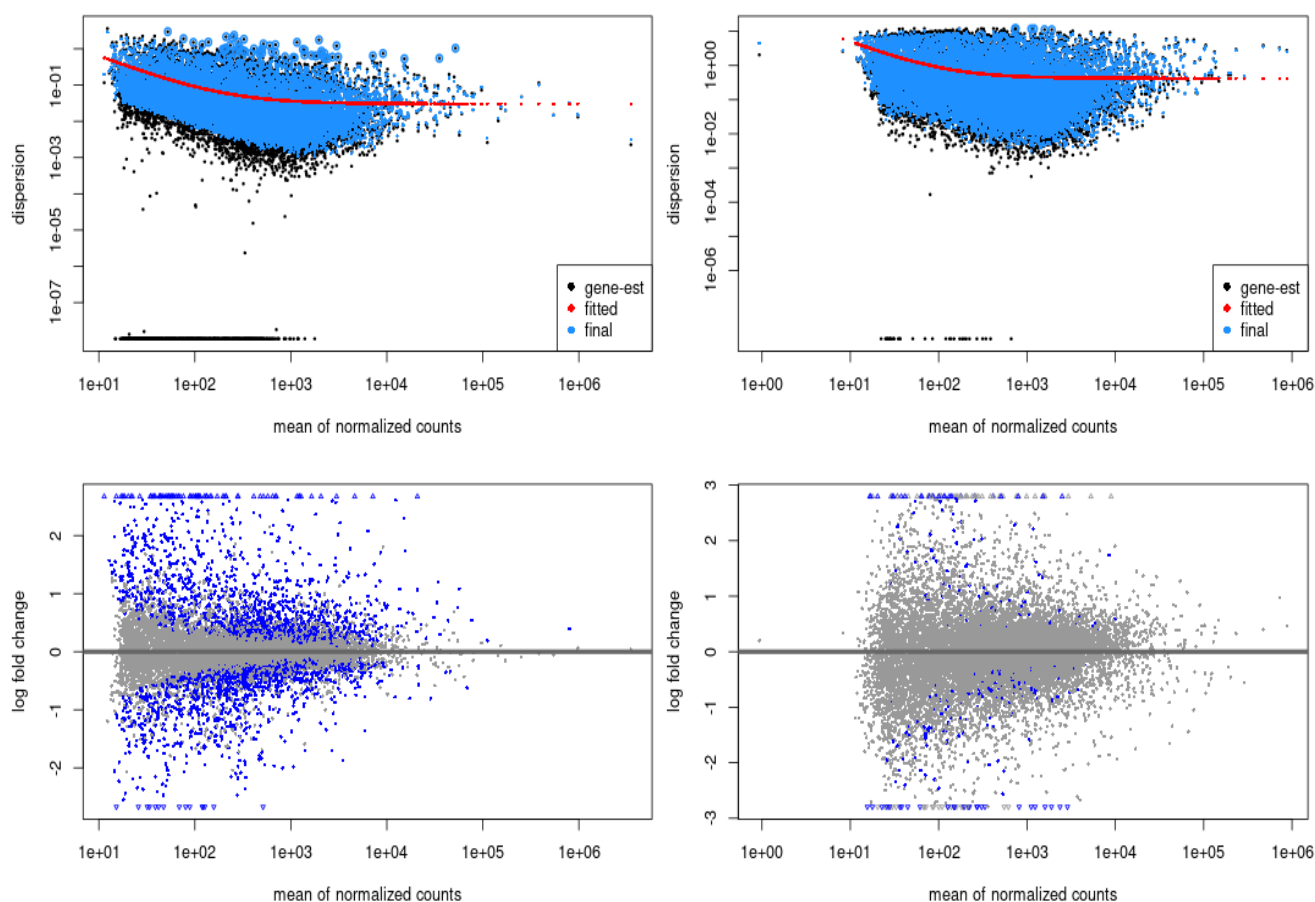
LINC02361, AL355607.2, AC134312.5, AC005837.1, LINC01050,

CACNA1G, SDK2, IGSF9B, PITPNM3, BEX4, PTN, KIF20A, THBS4, NACA3P, NDP, HOXD10, HOXD11, FAIM2, INHBE, ABCA12, TMEM132D, ASTN1, GRIA4, MARVELD2, ZSWIM5, LHX8, HAND2, DCHS1, ZNF608, NPIPA5, DLK1, GPAT2, CPNE4, NPIPB15, STMN3, EGFL6, RASA4DP, HAND2-AS1, SHISA9, PWP2, PSG4, MIR4300HG, PRSS51

**GSE165322**

CDH10, CCDC85A, PITX1, LXN, DRP2, CACNG7, TMEM59L, SFRP4, ADGB, PREX1, PAX8, KRT17, LRRC17, LSP1, H19, HSPA12B, EPHB2, GSTM1, ESPNL, GTF2H2, SHC3, NCAM1, CAMK4, ERG, SCRG1, GUCY1A1, PCDH19, LDLRAD4, NSG1, AFAP1L2, RASGRP1, A2M, OLFML2A, C3orf70, RPL13P12, HLA-DPB1, AC138969.2, XIST, HLA-K, KANTR, ACTG1P20, LINC00960, PSG4, CEBPA, STK19B, GAPDHP35, PKD1P5, ABCC6P2

KDM5D, DDX3Y, ZFY, TRHDE, PRKY, USP9Y, GBP3, E2F8, USP6, RPS4Y1, TXLNGY, ACP6, NLGN4Y, KRT4, TTTY14, HERC2P3, CHRM2, UTY, SRY, EEF1GP1, NPIPB15, EIF1AY, AGAP9, AL365357.1, AC023157.1, MTND2P28, HDGFP1, RPS28, AC012146.1, NPIPB2, A SMTLAS1, MBL1P, AC004923.4, AL590560.2, AC011043.1



**Figura 11:** Dispersión (arriba) y log2FoldChange (abajo) para cada uno de los genes evaluados en DESeq2. A la izquierda, GSE175477; a la derecha, GSE165322.

## Revisión sistemática de la retinosis pigmentaria en humanos

La dispersión estimada de los datos se ha representado en la Figura 10, donde los puntos negros son las estimaciones de dispersión de cada gen por separado, los puntos azules representan las estimaciones finales tras reducir la estimación de cada gen hacia la línea de tendencia, marcada por la línea roja. También puede verse, en la parte inferior de la figura, la distribución de genes que presenta un 2logFC positivo (clasificado como *up-regulated*) o negativo (*down-regulated*).

### 3.3 Análisis funcional

Tras llevar a cabo el análisis de la expresión diferencial se ha realizado el análisis de enriquecimiento funcional para los términos GO y rutas KEGG. En la Tabla 5 puede comprobarse la diversidad de los resultados significativos obtenidos entre ambos estudios.

**Tabla 5:** Número de términos GO y rutas KEGG significativas en cada estudio tras implementar GSEA. Valores positivos de enrichment score indican sobrerrepresentación en casos; negativo, infrarepresentación (y, por tanto, sobreexpresión en controles).

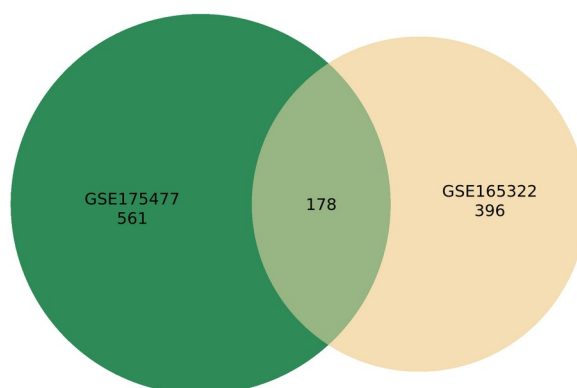
Estudio	Términos GO		Rutas KEGG	
	Enrichment Score	Enrichment Score	Enrichment Score	Enrichment Score
	Positivo	Negativo	Positivo	Negativo
GSE175477	372	189	9	11
GSE165322	267	129	27	5

En los contrastes individuales realizados encontramos una gran cantidad de términos GO significativos en ambos estudios. Además, se han identificado un elevado número de funciones biológicas significativas en ambos (178 términos GO, de los cuales 68 se encuentran sobreexpresados y 110 infraexpresados). Los procesos biológicos con los valores más extremos de enriquecimiento (*enrichment score*) se encuentran recogidos en la Figura 10. Entre los procesos más significativos que se hallan **infraexpresados** en **pacientes**, se encuentran multitud de términos que implican alteraciones en el ciclo celular (meiosis y mitosis), varios términos relacionados con el spliceosoma y el splicing de RNA, y alteraciones en la gliogénesis, reguladores sinápticos y neurotransmisores y otras anomalías relacionadas con el sistema nervioso, como el comportamiento (GO:0007610), el aprendizaje o la memoria (GO:0007611). En cuanto a los términos GO más

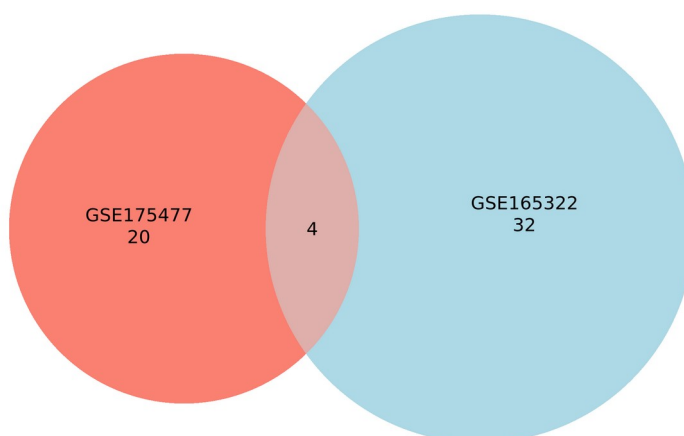
## Revisión sistemática de la retinosis pigmentaria en humanos

significativos a nivel de **sobreexpresión** en **casos**, se encuentra la ruta de señalización receptores acoplados a proteínas G, y otros tan interesantes relacionados con el desarrollo de órganos y sistemas sensoriales. En cualquier caso, la expresión diferencial de alguno de estos genes puede contribuir a crear un desequilibrio en las funciones de pacientes con RP. Los resultados significativos se han representado gráficamente mediante *dotplots* (Figura 14 y Figura 15), y con *treemaps* (en el apartado de Material Suplementario).

Por el contrario, solo encontramos 4 rutas KEGG significativas común para ambos grupos, y todas se corresponden a rutas infraexpresadas: hsa03440 (recombinación homóloga), hsa03010 (ribosomas), hsa04110 (ciclo celular) y hsa03460 (ruta de anemia Fanconi). Es decir, vemos un perfil de funciones alteradas que concuerda con los resultados obtenidos en términos GO.

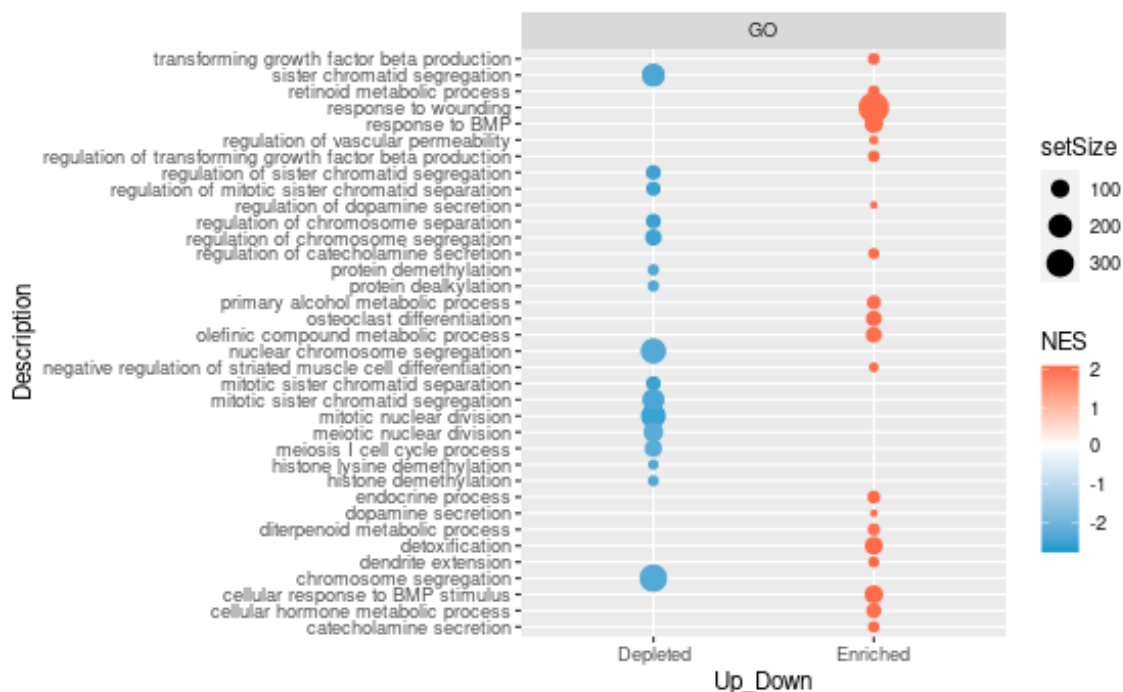


**Figura 12:** Visualización de la distribución del número de procesos biológicos significativos entre ambos estudios mediante un diagrama de Venn.

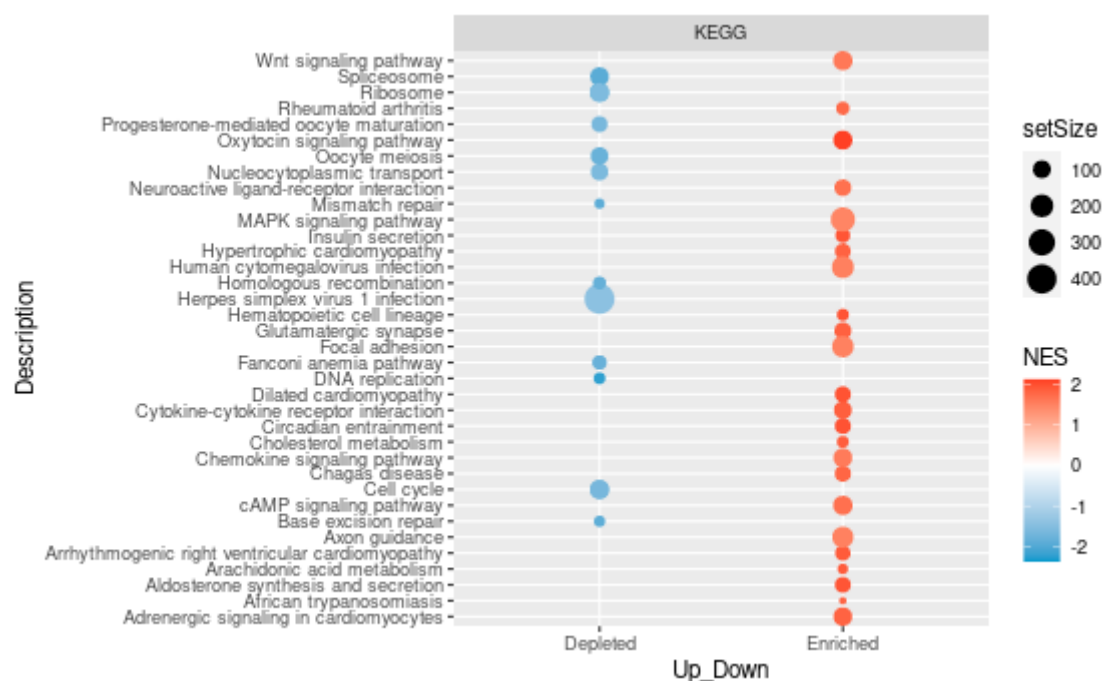


**Figura 13:** Visualización de la distribución del número de rutas KEGG significativas entre ambos estudios mediante un diagrama de Venn.

## Revisión sistemática de la retinosis pigmentaria en humanos



**Figura 14:** Procesos biológicos significativos en GSE175477 y GSE165322. Se han separado en función de su valor NES (enrichment score). El tamaño del punto indica su valor NES absoluto.



**Figura 15:** Rutas KEGG significativas en GSE175477 y GSE165322. Se han separado en función de su valor NES (enrichment score). El tamaño del punto indica su valor NES absoluto.

## **4. Discusión**

### **Estrategia y limitaciones.**

En esta revisión se ha seleccionado un bajo número de estudios, dado que se ha preferido mantener un alto criterio de exclusividad, a fin de asegurar su comparabilidad. Además, los dos estudios seleccionados cuentan con el mismo número de réplicas biológicas y coinciden en el tejido utilizado, lo que puede favorecer al poder estadístico del análisis.

### **Resultados obtenidos.**

Como se había mencionado, la retinosis pigmentaria (RP), es una distrofia de la retina caracterizada por la degeneración progresiva y la muerte de los fotorreceptores rod, lo que deriva en una pérdida de la visión. Dada la heterogeneidad génica que ocasiona esta patología, es difícil, pero primordial, conocer y comprender los mecanismos que ocasionan su muerte celular, con el objetivo principal de obtener dianas terapéuticas sobre los que centrar el tratamiento. Focalizarse en encontrar los procesos comunes que subyacen en los casos de RP, en lugar de tratamientos específicos de mutación, ha demostrado ser una estrategia innovadora para combatir la degeneración de la retina.

Las múltiples rutas que pueden ocasionar la muerte de los fotorreceptores rod, se encuentran: la acumulación de proteínas mal plegadas o mal localizadas, lo que deriva en un estrés del retículo endoplasmático (ER), desregulación de cGMP, acumulación de ión Calcio, estrés oxidativo y respuestas inflamatorias (Newton & Megaw, n.d.).

Por lo que respecta al transporte de calcio, se ha encontrado multitud de rutas significativas asociadas al transporte de este ión (*GO:0017156*, regulación de la exocitosis del ión calcio; *GO:0006816*, transporte del ión calcio; *GO:0070588*, transporte transmembrana del ión calcio).

A pesar de que no haya una evidencia clara, se ha hipotetizado sobre el papel de la deficiencia en el transporte de  $\text{Ca}^{2+}$  y su acumulación intracelular como un causante del estrés en ER, ocasionando la activación de enzimas proteolíticas y la posible muerte celular del fotorreceptor (Cheng et al., 2018; van Vliet et al., 2017). De hecho, la acumulación de calcio también se ha relacionado con la apoptosis, en rutas independientes del estrés en ER, vía una apoptosis dependiente de caspasas (Comitato et al., 2014).

## ***Revisión sistemática de la retinosis pigmentaria en humanos***

Por otra parte, también podemos ver una sobreexpresión en pacientes en las rutas de producción de citocinas y producción de factor de necrosis tumoral, cuya inhibición farmacológica se ha demostrado que reduce la muerte celular de fotorreceptores (Xu et al., 2017).

No podemos pasar por alto el elevado número de procesos biológicos infraexpresados en pacientes relacionados con el ciclo celular y la replicación del DNA. Es interesante ver la ruta de preprocesado de RNAm, ya que hay una extensa bibliografía al respecto como factor de la aparición de la retinosis pigmentaria (Arzalluz-Luque et al., 2021; Tanackovic et al., 2011; Yang et al., 2021). Mutaciones en los factores de procesado del pre-RNAm ocasionan una desregulación del espliceosoma, afectando a multitud de genes relacionados con funciones específicas de retina y otros procesos biológicos más generales como la apoptosis o la inflamación, también vistos como procesos significativos del análisis.

El análisis del enriquecimiento funcional también ha presentado la ruta de señalización acoplada a proteína G y varios procesos relacionados con el mantenimiento de la matriz extracelular, sobreexpresados en pacientes. Los receptores acoplados a proteínas G constituyen la mayor familia de proteínas de membrana y tienen un rol fundamental en multitud de procesos fisiológicos. Además, se ha visto que las proteínas de la matriz extracelular juegan un papel clave en el soporte de las células pigmentarias de la retina, tanto *in vivo* como *ex vivo* (Arzalluz-Luque et al., 2021). Por lo tanto, podemos hipotetizar que estos genes se encuentran estrechamente relacionados con el mecanismo subyacente de la enfermedad y pueden ser buenos candidatos como dianas terapéuticas.

Debemos tener en cuenta que la retinosis pigmentaria se expresa de múltiples formas. Si bien hay muy pocos procesos moleculares y mecanismos celulares que a día de hoy se consideren comunes a todas (como la inflamación, la activación de la microglía o la pérdida de células fotorreceptoras) (Sánchez-Cruz et al., 2018), el desarrollo de tratamientos cuya diana sean estos mecanismos patológicos es clave, dada la gran complejidad genética de RP, que impide que la terapia génica (aunque ideal) sea factible.

Tras los resultados obtenidos en este pequeño análisis, se pone en relieve la necesidad de conocer, aún más en profundidad, cómo estos mecanismos impactan en una enfermedad tan heterogénea como la retinosis pigmentaria, y así poder mejorar nuestra comprensión de la enfermedad, mejorar su tratamiento y su diagnóstico.



## 5. Conclusiones

En conclusión, este trabajo aborda la complejidad de la retinosis pigmentaria, bajo la estrategia de analizar los mecanismos celulares y moleculares comunes que subyacen en pacientes diagnosticados con RP, pero bajo mutaciones totalmente distintas (portadores de la mutación m.3243A>G, en el gen mitocondrial MT-TL1; y portadores de una mutación del factor de procesamiento de pre-mRNA PRPF8, implicada en el *splicing* alternativo). El objetivo principal es arrojar luz sobre posibles tratamientos, cuyas dianas sean estos mecanismos. Esto supondría una alternativa a las terapias génicas que, si bien se están poniendo en práctica con buenos resultados, **no son capaces de abordar** la totalidad de los casos diagnosticados con la enfermedad.

## 6. Bibliografía

- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166. <https://doi.org/10.1093/BIOINFORMATICS/BTU638>
- Arzalluz-Luque, Á., Cabrera, J. L., Skottman, H., Benguria, A., Bolinches-Amorós, A., Cuenca, N., Lupo, V., Dopazo, A., Tarazona, S., Delás, B., Carballo, M., Pascual, B., Hernan, I., Erceg, S., & Lukovic, D. (2021). Mutant PRPF8 Causes Widespread Splicing Changes in Spliceosome Components in Retinitis Pigmentosa Patient iPSC-Derived RPE Cells. *Frontiers in Neuroscience*, 15(April). <https://doi.org/10.3389/fnins.2021.636969>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics* 2000 25:1, 25(1), 25–29. <https://doi.org/10.1038/75556>
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*, 41(D1), 991–995. <https://doi.org/10.1093/nar/gks1193>
- Basu, A., Ghosh, A., Mandal, A., Martín, N., & Pardo, L. (2017). A wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator. *Electronic Journal of Statistics*, 11(2), 2741–2772. <https://doi.org/10.1214/17-EJS1295>
- Català-senent, J. F., Andreu, Z., Roig, F. J., & Marta, R. (2021). A deep transcriptome meta-analysis reveals sex-based molecular differences in Multiple Sclerosis. 1–29.
- Cheng, S. Y., Wang, S. C., Lei, M., Wang, Z., & Xiong, K. (2018). Regulatory role of calpain in neuronal death. *Neural Regeneration Research*, 13(3), 556. <https://doi.org/10.4103/1673-5374.228762>

## ***Revisión sistemática de la retinosis pigmentaria en humanos***

- Chung, C. Y., Singh, K., Kotiadis, V. N., Ahn, J. H., Kabir, L., Topley, E., Tan, J., Andrews, W. D., Plotegher, N., Guella, G., Pitceathly, R. D. S., Szabadkai, G., Yuneva, M., & Duchon, M. R. (2021). Constitutive activation of the PI3K-Akt-mTORC1 pathway sustains the m.3243A>G mtDNA mutation. *BioRxiv*, 2020.06.18.159103. <https://doi.org/10.1101/2020.06.18.159103>
- Comitato, A., Sanges, D., Rossi, A., Humphries, M. M., & Marigo, V. (2014). Activation of Bax in three models of retinitis pigmentosa. *Investigative Ophthalmology & Visual Science*, 55(6), 3555–3562. <https://doi.org/10.1167/IOVS.14-13917>
- Ewels, P., Magnusson, M., Lundin, S., & Källér, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047. <https://doi.org/10.1093/BIOINFORMATICS/BTW354>
- Ferrer, F. C., Pascual, M., Hidalgo, M. R., Malmierca-Merlo, P., Guerri, C., & García-García, F. (2020). Unveiling Sex-Based Differences in the Effects of Alcohol Abuse: A Comprehensive Functional Meta-Analysis of Transcriptomic Studies. *Genes* 2020, Vol. 11, Page 1106, 11(9), 1106. <https://doi.org/10.3390/GENES11091106>
- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30. <https://doi.org/10.1093/NAR/28.1.27>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements Daehwan HHS Public Access. *Nature Methods*, 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), 1–10. <https://doi.org/10.1186/GB-2009-10-3-R25>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078. <https://doi.org/10.1093/BIOINFORMATICS/BTP352>

## ***Revisión sistemática de la retinosis pigmentaria en humanos***

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), 10–12. <https://doi.org/10.14806/EJ.17.1.200>
- Mullin, N. K., Anfinson, K. R., Riker, M. J., Wieland, K. L., Tatro, N. J., Scheetz, T. E., Mullins, R. F., Stone, E. M., & Tucker, B. A. (2021). Sensitive quantification of m.3243A>G mutational proportion in non-retinal tissues and its relationship with visual symptoms. *Human Molecular Genetics*, 00(00), 1–8. <https://doi.org/10.1093/hmg/ddab289>
- Newton, F., & Megaw, R. (n.d.). *Mechanisms of Photoreceptor Death in Retinitis Pigmentosa*. <https://doi.org/10.3390/genes11101120>
- R Core Team (2019) *R A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. - References - Scientific Research Publishing. (n.d.). Retrieved January 11, 2022, from [https://www.scirp.org/\(S\(lz5mqp453edsnp55rrgjt55\)\)/reference/ReferencesPapers.aspx?ReferenceID=2631126](https://www.scirp.org/(S(lz5mqp453edsnp55rrgjt55))/reference/ReferencesPapers.aspx?ReferenceID=2631126)
- Saeidipour, B., & Bakhshi, S. (2013). The relationship between organizational culture and knowledge management,& their simultaneous effects on customer relation management. *Advances in Environmental Biology*, 7(10), 2803–2809.
- Sánchez-Cruz, A., Villarejo-Zori, B., Marchena, M., Zaldivar-Díez, J., Palomo, V., Gil, C., Lizasoain, I., De La Villa, P., Martínez, A., De La Rosa, E. J., & Hernández-Sánchez, C. (2018). Modulation of GSK-3 provides cellular and functional neuroprotection in the rd10 mouse model of retinitis pigmentosa. *Molecular Neurodegeneration*, 13(1). <https://doi.org/10.1186/S13024-018-0251-Y>
- Tanackovic, G., Ransijn, A., Thibault, P., Elela, S. A., Klinck, R., Berson, E. L., Chabot, B., & Rivolta, C. (2011). PRPF mutations are associated with generalized defects in spliceosome formation and pre-mRNA splicing in patients with retinitis pigmentosa. *Human Molecular Genetics*, 20(11), 2116–2130. <https://doi.org/10.1093/HMG/DDR094>

## ***Revisión sistemática de la retinosis pigmentaria en humanos***

- Toolkit Documentation : Software : Sequence Read Archive : NCBI/NLM/NIH.* (n.d.). Retrieved January 11, 2022, from [https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit\\_doc](https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=toolkit_doc)
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* (Oxford, England), 25(9), 1105–1111. <https://doi.org/10.1093/BIOINFORMATICS/BTP120>
- Van Verk, M. C., Hickman, R., Pieterse, C. M. J., & Van Wees, S. C. M. (2013). RNA-Seq: Revelation of the messengers. *Trends in Plant Science*, 18(4), 175–179. <https://doi.org/10.1016/j.tplants.2013.02.001>
- van Vliet, A. R., Giordano, F., Gerlo, S., Segura, I., Van Eygen, S., Molenberghs, G., Rocha, S., Houcine, A., Derua, R., Verfaillie, T., Vangindertael, J., De Keersmaecker, H., Waelkens, E., Tavernier, J., Hofkens, J., Annaert, W., Carmeliet, P., Samali, A., Mizuno, H., & Agostinis, P. (2017). The ER Stress Sensor PERK Coordinates ER-Plasma Membrane Contact Site Formation through Interaction with Filamin-A and F-Actin Remodeling. *Molecular Cell*, 65(5), 885-899.e6. <https://doi.org/10.1016/J.MOLCEL.2017.01.020>
- Xu, X. J., Wang, S. M., Jin, Y., Hu, Y. T., Feng, K., & Ma, Z. Z. (2017). Melatonin delays photoreceptor degeneration in a mouse model of autosomal recessive retinitis pigmentosa. *Journal of Pineal Research*, 63(3). <https://doi.org/10.1111/JPI.12428>
- Yang, C., Georgiou, M., Atkinson, R., Collin, J., Al-Aama, J., Nagaraja-Grellscheid, S., Johnson, C., Ali, R., Armstrong, L., Mozaffari-Jovin, S., & Lako, M. (2021). Pre-mRNA Processing Factors and Retinitis Pigmentosa: RNA Splicing and Beyond. *Frontiers in Cell and Developmental Biology*, 9, 2062. <https://doi.org/10.3389/FCELL.2021.700276/BIBTEX>
- Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, 16(5), 284–287. <https://doi.org/10.1089/OMI.2011.0118/ASSET/IMAGES/LARGE/FIGURE1.JPEG>
- Zhang, Y., Parmigiani, G., & Johnson, W. E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics*, 2(3). <https://doi.org/10.1093/NARGAB/LQAA078>

## 7. Material Suplementario

**Disponibilidad de datos:** los datos usados para este estudio se encuentran disponibles, públicamente, en GEO (<https://www.ncbi.nlm.nih.gov/geo/>). Los nombres de acceso a la base de datos son:

- GSE175477: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse175477>
- GSE 165322: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse165322>

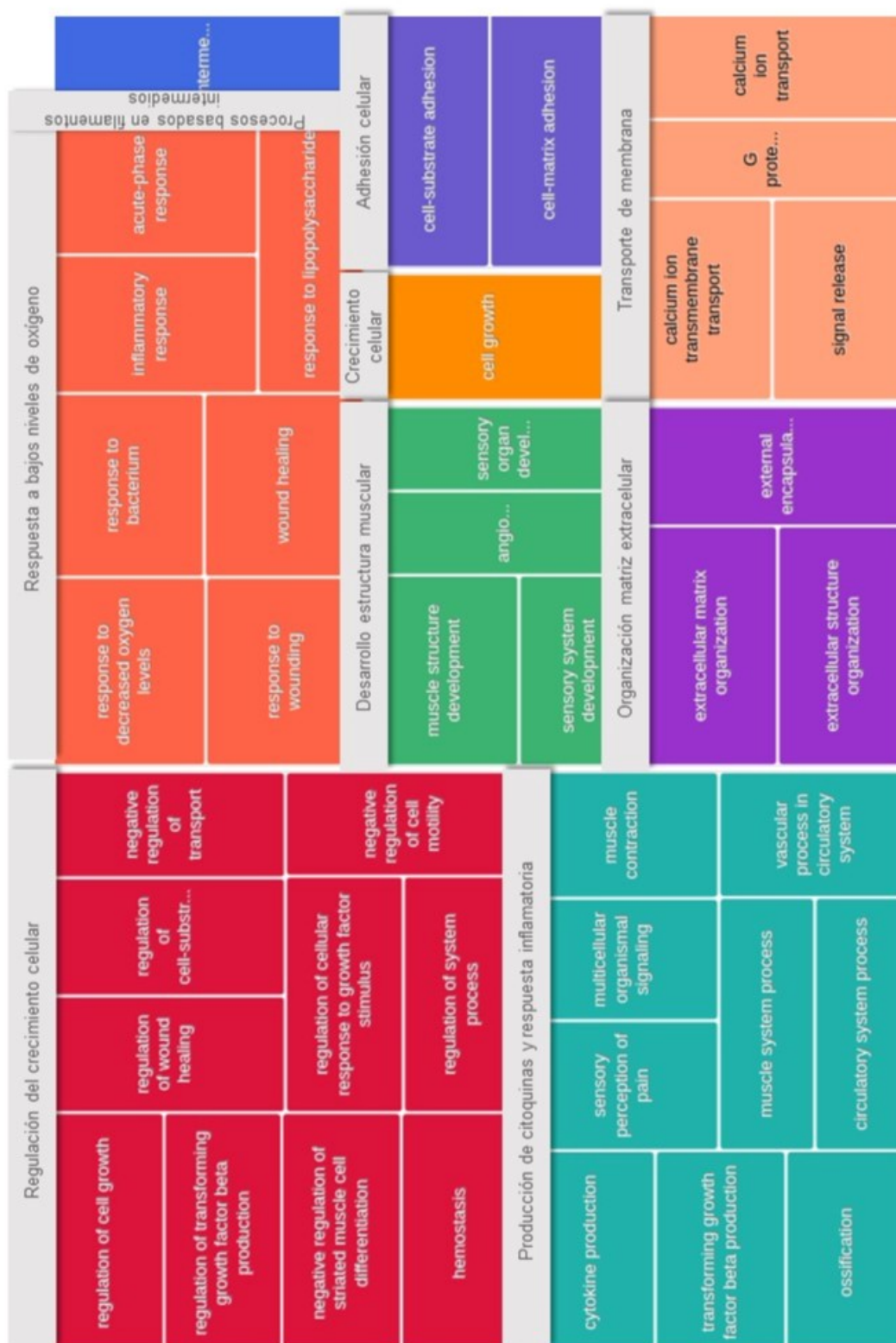
Los scripts que contienen todos los parámetros de ejecución están disponibles en [https://github.com/PilarLF/Estudios\\_insilico](https://github.com/PilarLF/Estudios_insilico).

**Tabla S1:** Software y las versiones respectivas usadas en este trabajo.

Software/Paquetes de R	Version
AnnotationDbi	1.56.5
clusterProfile	4.2.1
DESeq2	1.34.0
DOSE	3.20.1
edgeR	3.36.0
eulerr	6.1.1
genefilter	1.76.0
ggplot2	3.3.5
GO.db	3.14.0
KEGG.db	3.2.3
org.Hs.eg.db	3.14.0
pathview	1.34.0
PCAtools	2.6.0
R	4.1.2
ReportingTools	2.34.0
sva	3.42.0
VennDiagram	1.7.1

## Revisión sistemática de la retinosis pigmentaria en humanos

**Figura S1.** Treemap de los procesos biológicos sobreexpresados en pacientes, agrupados por el nivel jerárquico superior en que se incluyen. Realizado mediante Revigo (<http://revigo.irb.hr/>)



## Revisión sistemática de la retinosis pigmentaria en humanos

**Figura S2.** Treemap de los procesos biológicos infraexpresados en pacientes, agrupados por el nivel jerárquico superior en que se incluyen. Realizado mediante Revigo (<http://revigo.irb.hr/>).

