

Documento técnico del TFM

Análisis comparativo de smallRNAs en síndrome de Rett mediado por MECP2 y síndrome atípico de Rett mediado por FOXP1.

Fernando Hernández Correa

Máster en bioinformática y bioestadística.
Bioinformática clínica.

Guerau Fernández Isern

Javier Luis Cánovas Izquierdo

10 de Junio de 2020

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN.....	3
2. DESARROLLO DE PIPELINE PARA ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.....	3
2.1 Datos de partida.....	3
2.2 Descarga de datos en crudo y adaptación al formato adecuado.....	3
2.3 Comprobación de la calidad de los reads.....	5
2.4 Anotación de miRNA.....	7
2.5 Control de calidad de datos y normalización.....	15
2.6 Análisis de expresión diferencial de miRNAs/isomiRs.....	20
2.7 Proceso de cruce de información de miRNAs/isomiRs DE en ambos experimentos.....	38
2.8 Análisis detallado de isomiRs DE.....	40
2.9 Clasificación supervisada.....	42
2.10 Integración gen-miRNA.....	43
2.11 Enriquecimiento GO.....	48
2.12 Enriquecimiento KEGG.....	61
3. RELACIÓN DE HERRAMIENTAS Y PROGRAMAS EMPLEADOS EN EL PROCESO DE ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.....	67
4. DIAGRAMA DE FLUJO DE PIPELINE PARA ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.....	69
ANEXO 1.....	71
ANEXO 2.....	78
ANEXO 3.....	85
REFERENCIAS.....	88

TFM_FHC_1_1_DOCUMENTO TÉCNICO

Fernando Hernández Correa

1. INTRODUCCIÓN

El presente documento desarrolla las tareas técnico-informáticas correspondientes al TFM que lleva por título **Análisis comparativo de smallRNAs en síndrome de Rett mediado por MECP2 y síndrome atípico de Rett mediado por FOXG1**.

En referencia al Plan de Trabajo, este documento recoge las tareas del grupo T1, encaminadas a la consecución del objetivo global G1, establecer un *pipeline* completo de todo el proceso de análisis que pueda ser factible de utilizar en comparación de experimentos similares sobre sRNAs. [1]

2. DESARROLLO DE PIPELINE PARA ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.

2.1 Datos de partida.

Se parte de 2 conjuntos de datos independientes y públicos referentes a 2 experimentos de smallRNA-seq.

- El estudio '*Quantitative analysis of wild type and MeCP2-/y mice hippocampus miRNAs using next generation sequencing*' cuenta con 2 muestras:

GSM1424567, secuenciación miRNA de hipocampo de ratones normales (wild tipe)
GSM1424566, secuenciación miRNA de hipocampo de ratones Mecp2-/Y

El estudio está publicado en:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59029>

- El estudio '*FOXG1 and DDX5 control PRKAR2B expression via miR200 in a mouse model of atypical Rett-syndrome*' cuenta con 18 muestras:

GSM2791062-70, 9 réplicas de smallRNA-seq de hipocampo de ratones normales
GSM2791053-61, 9 réplicas de smallRNA-seq de hipocampo de ratones Foxg1Cre/+

El estudio está publicado en:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE104169>

Ambos estudios se han realizado en la plataforma Illumina Hiseq 2000.

2.2 Descarga de datos en crudo y adaptación al formato adecuado.

Para el análisis de los datos de secuenciación de sRNA de los 2 experimentos, se parte de los datos en crudo, disponibles en *SRA (Sequence Read Archive)*. Para su descarga y adaptación de formato se dispone del conjunto de herramientas *SRA toolkit* [2]. Dentro de las herramientas disponibles, se escoge *fastq-dump*, que transforma los datos *sra* en crudo al formato *fastq*.

El comando que se utiliza para la descarga de los *datasets* en formato *fastq* es:

```
fastq-dump --gzip --skip-technical --readids -dumpbase --split-3 --clip  
--outdir dir_out SRRx
```

Opciones del comando:

--gzip	Salida comprimida
--skip-technical	Carga sólo las muestras biológicas
--readids	Añade el identificador de reads
-dumpbase	Formatea las secuencias usando bases (en SOLiD se usan colores)
--split	Divide la salida en uno (-spot), dos (-files) o 3 (-3) ficheros. Se escoge la opción -3 para diferenciar los reads alineados por la izquierda (3) y por la derecha (5), en caso de haber utilizado secuenciación paired-end, en el tercer fichero se guardan los reads no emparejados (singletone).
--clip	Elimina etiquetas utilizadas para la amplificación.
--outdir	Directorio de salida.
SRRx	Código SRR del dataset que se quiere procesar.

El código *SRR* hace referencia al *dataset SRA* que se quiere procesar. Para obtener los códigos de los *datasets* se descarga el fichero *SraAcclList.txt* (*Accession List*) de cada uno de los experimentos.

Para el primer experimento, se abre la página:

<https://www.ncbi.nlm.nih.gov/sra?term=SRP044018>

En la parte superior derecha, se pulsa *Send to:* y se escoge la opción *File* (*Format: Accession List*). El contenido del fichero *SraAcclList.txt* es:

```
SRR1504855  
SRR1504856
```

Del mismo modo se accede a la página del segundo experimento:

<https://www.ncbi.nlm.nih.gov/sra?term=SRP118722>

Y se descarga el fichero *SraAcclList.txt*, que tiene por contenido:

```
SRR6061838  
SRR6061839  
SRR6061840  
SRR6061841  
SRR6061842  
SRR6061843  
SRR6061844  
SRR6061845  
SRR6061846  
SRR6061847  
SRR6061848  
SRR6061849  
SRR6061850  
SRR6061851  
SRR6061852  
SRR6061853  
SRR6061854  
SRR6061855
```

Con el objeto de mecanizar el proceso, se ha diseñado el script *ProcesaSraAcclList2fastq.sh*

```

#!/bin/bash
#lee los codigos SRR pasados en el fichero SraAccList.txt como argumento
y ejecuta fastq-dump

gawk '{if($1 ~ /^SRR/){
    print "Procesando SRA to fastq "$1;
    system("fastq-dump --gzip --skip-technical --readids -dumpbase
--split-3 --clip --outdir ~/TFM/datos/fastq" $1);
    }
}' $1

#saca el informe de calidad de cada muestra con fastqc
gawk '{if($1 ~ /^SRR/){
    print "Informe de calidad "$1;
    system("fastqc ~/TFM/datos/fastq/" $1 ".fastq.gz");
    }
}' $1

```

Va leyendo secuencialmente los códigos *SRR* y ejecutando el comando *fastq-dump* para cada uno de ellos.

Se ejecuta el *script* para cada uno de los ficheros *SraAccList.txt*, del modo:

```

$ ProcesaSraAccList2fastq.sh SraAccList1.txt
$ ProcesaSraAccList2fastq.sh SraAccList2.txt

```

La salida del comando *fastq-dump* con la opción *--split-3* da como resultado un fichero único para cada muestra, en los 2 experimentos, por lo que se entiende que la secuenciación realizada es *single-end*.

Los ficheros en formato *fastq* comprimidos, se guardan en el directorio *~/TFM/datos/fastq*. Cada registro consta de 4 líneas, con el siguiente formato:

```

@SRR1504855.2.1 FC81C2CABXX:4:1101:1326:2088 length=49
AACAAATATCCTGGTGATGAGTATCTCGGATGCCGGCTTCTTCCTGGAAA
+SRR1504855.2.1 FC81C2CABXX:4:1101:1326:2088 length=49
FFFFFFFFDFFDDDC+C*->>,CBB6997>0BCDB5@D%%%%%%%%%%%%%

```

Las 2 primeras líneas indican el identificador de secuencia y la secuencia de nucleótidos. Las 2 líneas siguientes indican la calidad en la lectura de cada nucleótido.

2.3 Comprobación de la calidad de los reads.

Para la comprobación de la calidad de los *reads* se utiliza la herramienta *fastqc*. Es conveniente consultar el enlace anterior para la interpretación de resultados.

El *script* anterior, *ProcesaSraAccList2fastq.sh*, una vez que ha obtenido los ficheros *.fastq*, ejecuta el análisis de calidad *fastqc* en cada uno de ellos. Además de chequear la calidad de los *reads*, uno de los objetivos principales es localizar *adapters* en las secuencias, que es necesario eliminar antes de realizar el alineamiento de los *reads*.

Para ejecutar esta herramienta sólo es necesario pasar como parámetro el fichero *fastq* a analizar. Como ejemplo, se ve el proceso para un fichero del primer experimento, el *script* ha ejecutado:

```

$ fastqc SRR1504855.fastq.gz

```

Como resultado se obtienen 2 archivos:

SRR1504855_fastqc.zip Resultados del análisis.
SRR1504855_fastqc.html Informe en formato html.

En el informe se puede apreciar que la calidad de las secuencias por bases decrece por debajo de 28 a partir del nucleótido 40, todas las secuencias tienen una longitud de 49 nucleótidos y no se detecta la presencia de *adapters*.

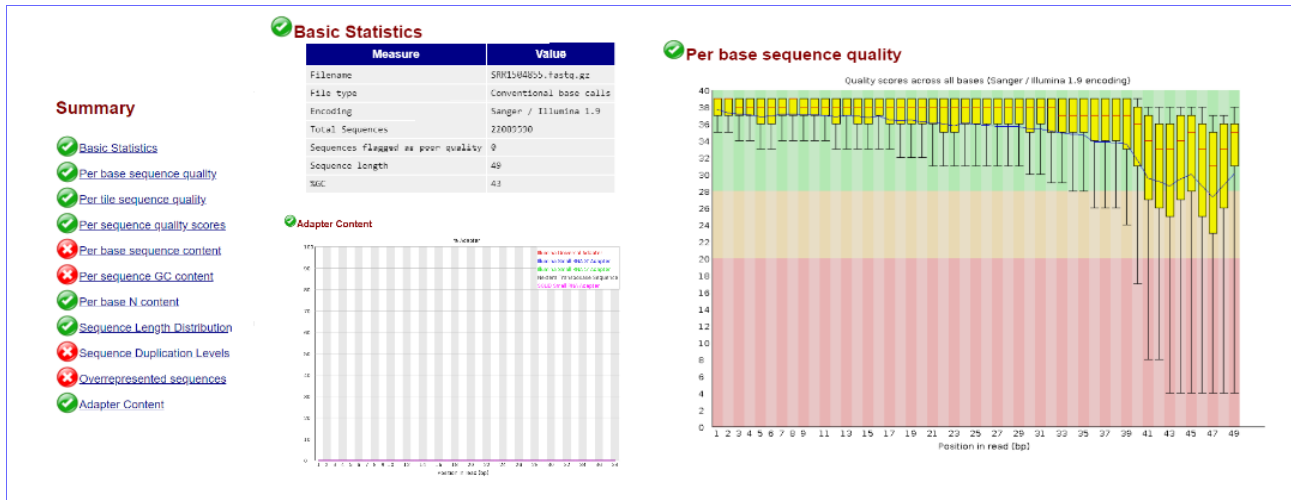


Figura 1. Resumen del informe de calidad *SRR1504855_fastqc.html*. Con estadísticas básicas, calidad de secuencia por bases y contenido de adapters.

Se comprueba lo mismo con un fichero del segundo experimento. En el informe *SRR6061841_fastqc.html* se puede ver que la calidad de las secuencias por bases es muy buena, todas las secuencias tienen una longitud de 50 nucleótidos y aquí sí se aprecia la presencia de *adapters*, concretamente *Illumina Small RNA 3 Adapter*.

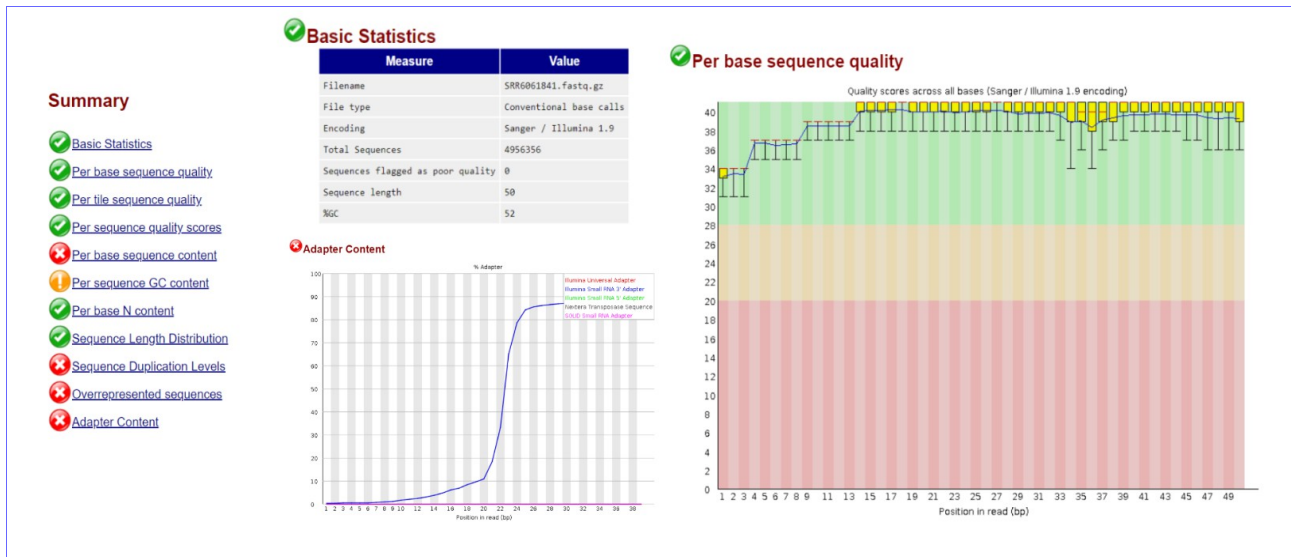


Figura 2. Resumen del informe de calidad *SRR6061841_fastqc.html*. Con estadísticas básicas, calidad de secuencia por bases y contenido de adapters.

Consultando la documentación de *Illumina*, el *adapter* tiene la secuencia:

5-TGGAATTCTCGGGTGCCAAGG-3

En principio no sería necesario eliminar *adapters* en las muestras del experimento 1 y sí en las muestras del experimento 2, antes de alinearlas frente a la base de datos de miRNA catalogados.

Sin embargo, cuando se realizó el procesamiento de los *reads* de esta forma, eliminando *aligners* de las muestras del experimento 2 y no de las del 1, los *reads* del experimento 2 se alinearon bien frente a la base de datos *miRBase* y, sin embargo, los *reads* del experimento 1 no se alinearon, lo que llevó a pensar que en las secuencias del experimento 1 existía algún *aligner* que había pasado desapercibido en el análisis de calidad.

Volviendo al paso anterior, se comprueba en la información de la muestra escogida en el experimento 1: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1424566> que se ha empleado *Qiagen miRNA Kit* para extraer y purificar el miRNA. Consultando la documentación de *Qiagen*, el *adapter* tiene la secuencia:

(5-3) AACTGTAGGCACCATCAAT

Sin embargo, no se encuentra esta secuencia repetida en los *reads* del experimento 1.

Volviendo al informe de calidad de la muestra *SRR1504855*, en la sección *Overrepresented sequences*, se observan los siguientes registros:

Sequence	Count	Percentage	Possible Source
CACCCGTAGAACCACCTTGCATCTCGTATGCCGCTTCTGCTTGAAAA	154364	0.6988088054146772	TruSeq Adapter, Index 8 (96% over 29bp)
TCTTTGGTTATCTAGCTGTATGTATCTCGTATGCCGCTTCTGCTTGAA	70815	0.3205808708989166	TruSeq Adapter, Index 2 (96% over 29bp)
TCCCTGAGACCCTAACTTGTGAATCTCGTATGCCGCTTCTGCTTGAAA	69026	0.31248203339219965	TruSeq Adapter, Index 8 (96% over 28bp)
AAGGGATTCTGATGTTGGTATCTCGTATGCCGCTTCTGCTTGAAAAA	32958	0.1492015017028383	TruSeq Adapter, Index 25 (96% over 28bp)
TGAGGTAGTAGGTTGTGGTTATATCTCGTATGCCGCTTCTGCTTGA	29199	0.1321844361982273	TruSeq Adapter, Index 25 (96% over 28bp)
AAGGGATTCTGATGTTGGTACATCTCGTATGCCGCTTCTGCTTGAAA	28838	0.13055018223516146	TruSeq Adapter, Index 9 (96% over 28bp)
TCCCTGAGACCCTTAACTGTATCTCGTATGCCGCTTCTGCTTGAAA	25240	0.1142619668359621	TruSeq Adapter, Index 2 (96% over 28bp)

Figura 3. Detalle del informe de calidad *SRR1504855_fastqc.html*. Con las filas de la tabla *Overrepresented sequences* donde aparecen *TruSeq Adapter*.

En la documentación de *Illumina*, para el *TruSeq Universal Adapter, Index 1 a 27*, todos terminan con la secuencia:

ATCTCGTATGCCGCTTCTGCTTG

Esta secuencia sí se encuentra repetida en los *reads* del experimento 1, por lo que se decide eliminarla como *adapter* antes de realizar el alineamiento de *reads*.

2.4 Anotación de miRNA.

Para la anotación de miRNA se ha escogido la herramienta *Miraligner* [3]. Para la definición del proceso se toma como referencia el tutorial *miRNA annotation* [4]. Lo primero que hay que hacer es instalar el conjunto de herramientas *seqcluster*. Se opta por la instalación completa de *bcbio*, para ello es necesario tener instalado previamente *python* y *git*.

Se realiza la instalación de *bcbio* sin genomas.

```
$ wget https://raw.githubusercontent.com/bcbio/bcbio-nextgen/master/scripts/bcbio_nextgen_install.py
$ python bcbio_nextgen_install.py ~/local/share/bcbio --tooldir=~/.local --aligners bwa --aligners bowtie2
```

2.4.1 Eliminar adapters.

Para eliminar los *adapters* localizados en la sección anterior, se utilizará la herramienta [cutadapt](#), dentro del kit de herramientas *bcbio*. El uso de *cutadapt* es el siguiente:

```
cutadapt -a ADAPTER --untrimmed-output fichero_notfound.fastq -o fichero_salida.fastq -m 17 -q 28 --overlap 8 fichero_entrada.fastq
```

Opciones del comando:

-a	Secuencia del ADAPTER regular 3 que se desea eliminar.
--untrimmed-output	Descarta los registros en los que no se encuentra el adapter y los guarda en un fichero de no encontrados.
-o	Path y fichero de salida.
-m	Eliminar reads con tamaño menor al indicado.
-q	Eliminar bases con calidad inferior a la indicada.
--overlap	Longitud mínima de solapamiento con el adapter.

A la vista de los informes de calidad obtenidos en el apartado anterior, se decide eliminar las secuencias en las que no se encuentra el adapter, considerando una coincidencia mínima de 8 pb y los *reads* con tamaños inferiores a 17 nucleótidos. En los *reads* del primer experimento las secuencias tienen una longitud de 49 nucleótidos y se está eliminando un *adapter* con una longitud de 24 nucleótidos; por lo que eliminar secuencias por debajo de los 17 nucleótidos parece una buena opción.

Lo mismo ocurre con los *reads* del segundo experimento; aunque aquí no se observó disminución de calidad en las bases, se aplica el mismo criterio y en cuanto a la longitud, los *reads* tienen 50 nucleótidos y se elimina un *adapter* de 19.

2.4.2 Unir secuencias iguales en un registro.

Se recomienda unir en un solo registro las secuencias iguales (*collapse*), para facilitar su procesamiento. Para ello se utiliza la opción *collapse* de *seqcluster*. El uso del comando es el siguiente:

```
seqcluster collapse -f path_fichero_entrada -o path_fichero_salida
```

Para facilitar el procesamiento de las muestras se diseña el *script EjecutaCutadaptCollapse.sh*

```
#!/bin/bash

#Para ficheros fastq en los que se ha detectado la presencia de adapters
#Se ejecuta primero cutadapt para eliminar adapters
#El parámetro 1 es el adapter detectado
#El parámetro 2 es el AccList con las muestras de reads
gawk -v ADAPTER=$1 '{
    if($1 ~ /^SRR/){
        print "Cutadapt " $1;
        system("cutadapt -a " ADAPTER " --untrimmed-output
```



```
~/TFM/datos/fastq_clean/" $1 "_notfound.fastq -o
/home/fernandohc/TFM/datos/fastq_clean/" $1 "_clean.fastq -m 17 --overlap
8 /home/fernandohc/TFM/datos/fastq/" $1 ".fastq.gz");
    }
}' $2

#se ejecuta de nuevo el informe de calidad, para compararlo con la
calidad antes de eliminar secuencias
gawk '{if($1 ~ /^SRR/){
    print "Informe de calidad "$1;
    system("fastqc ~/TFM/datos/fastq_clean/" $1 "_clean.fastq");
    }
}' $2

#Luego se ejecuta seqcluster collapse para dejar todos los reads iguales
en una sola linea
gawk '{if($1 ~ /^SRR/){
    print "Collapse " $1;
    system("seqcluster collapse -f
/home/fernandohc/TFM/datos/fastq_clean/" $1 "_clean.fastq -o
/home/fernandohc/TFM/datos/collapse");
    }
}' $2

#sobre los ficheros resultantes se eliminan las secuencias con 1 read
(x1)
gawk '{if($1 ~ /^SRR/){
    print "Quitando reads x1 " $1;
    system("quitaFreq1.sh ~/TFM/datos/collapse/" $1
"_clean_trimmed.fastq");
    }
}' $2
```

Se ejecuta el script para los 2 experimentos, cada uno con su secuencia *adater*:

```
$ EjecutaCutadaptCollapse.sh ATCTCGTATGCCGTCTTCTGCTTG SraAccList1.txt
$ EjecutaCutadaptCollapse.sh TGGAATTCTCGGGTGCCAAGG SraAccList2.txt
```

Los resultados finales se guardan en *~/TFM/datos/collapse/SRRx_clean_trimmed.fastq*

En primer lugar, lee de manera secuencial las muestras a procesar indicadas en el segundo parámetro (*SraAccList.txt*) y ejecuta *cutadapt* con el *ADAPTER* indicado en el primer parámetro. La salida se guarda como *~/TFM/datos/fastq_clean/SRRx_clean.fastq*
Una muestra de un registro de este tipo:

```
@SRR6061842.5.1 SN7001140:183:C65HPACXX:6:1102:10242:2247 length=50
TGGGGGGCCCAAGTCCTTCTGATCGAGGCC
+SRR6061842.5.1 SN7001140:183:C65HPACXX:6:1102:10242:2247 length=50
BCBFFFDCCCCDDDDACDDDDDDDEDDDDDBB
```

Se puede ver que a la secuencia se le ha recortado el *adapter 3*.

Sobre estos ficheros se ejecuta de nuevo *fastqc*, para obtener nuevos informes de calidad y poder compararlos con los realizados antes de la ejecución de *cutadapt*.

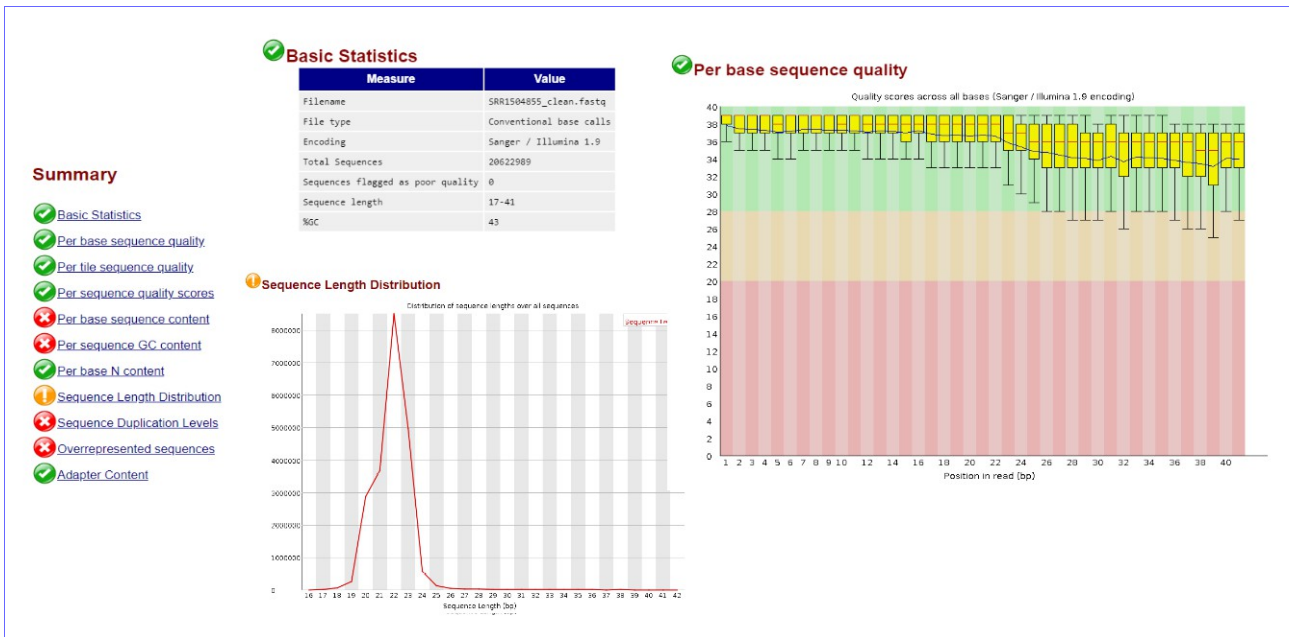


Figura 4. Resumen del informe de calidad SRR1504855_clean_fastqc.html. Con estadísticas básicas, calidad de secuencia por bases y distribución de secuencias por longitud.

Se analiza, en contraste con el anterior, el informe *SRR1504855_clean_fastqc.html*. Han pasado el proceso de corte el 93,4% de los reads y la curva de calidad por bases ha mejorado mucho. La mayoría de las secuencias están entre los 18 a 26 nucleótidos de longitud, siendo las de 22 las más representadas.

En el informe *SRR6061841_clean_fastqc.html* se comprueba que han pasado el proceso de corte el 91,2% de los reads y que la curva de calidad de las secuencias por bases sigue siendo muy buena, aunque ha empeorado un poco para reads por encima de 34 nucleótidos; las secuencias se distribuyen en su mayoría entre los 17 a 27 nucleótidos de longitud, aunque se ve un pequeño pico a los 33, siendo las de 22 las más representadas; se observa que se ha eliminado el adapter.

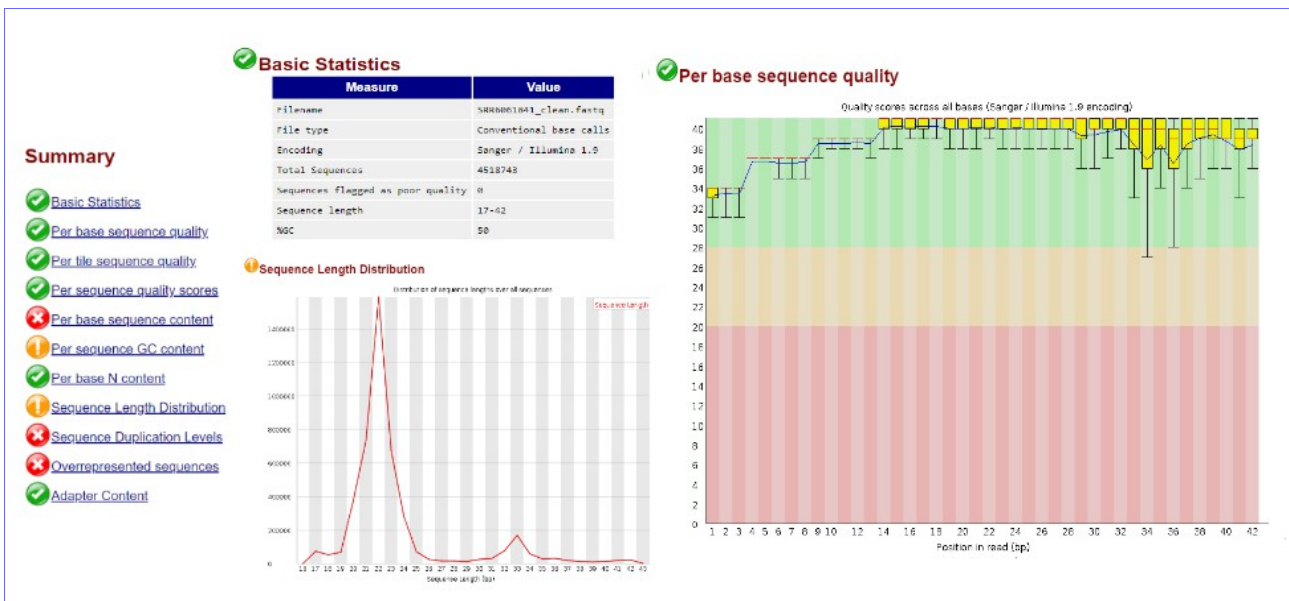


Figura 5. Resumen del informe de calidad SRR6061841_clean_fastqc.html. Con estadísticas básicas, calidad de secuencia por bases y distribución de secuencias por longitud.

Posteriormente, sobre este fichero, ejecuta *seqcluster collapse* y guarda la salida en:
~/TFM/datos/collapse/SRRx_clean_trimmed.fastq

Sobre estos resultados, se eliminan las secuencias que sólo aparecen 1 vez, para quitar ruido de fondo en la anotación de los *reads*. Para ello se ha diseñado el *script* *quitaFreq1.sh*, que se llama desde el *script* anterior (*EjecutaCutadaptCollapse.sh*).

```
#!/bin/bash
#elimina los reads con freq x1
#se pasa como argumento el fichero en el que se ha hecho collapse
#se salva con extension .tot
cp $1 $1.tot

#se recorre el fichero y se salva con los registros de freq > 1
gawk 'BEGIN{s=0}{
#si es fila de inicio de registro iniciamos s=0
    if($0 ~ /^@seq_/){
        s=0;
    }
#si es fila de inicio de registro con freq distinto de x1 imprimimos
registro
    if($0 ~ /^@seq_/ && $0 !~ /x1$/) {
        print $0;
        s=1;
    }
    if($0 !~ /^@seq_/ && s==1){
        print $0;
    }
}' $1 > $1.gt1
rm $1
mv $1.gt1 $1
```

Se dejan los ficheros, una vez eliminadas las secuencias *x1*, con el mismo nombre y se añade la extensión *.tot* a los ficheros originales, que contienen todas las secuencias. Una muestra de un registro de este tipo:

```
@seq_1_x12018
CATTGATCATCGACACTTCGAACGCACCTTGC GGCCCCGGGTT
+
BBBBEEEEGGGGGIIIIIIIIIIIIIIIIIIIIIIHIIGHHHIIGEF
```

Se puede ver que las secuencias se han numerado secuencialmente y se ha añadido un indicador del número de veces que se repite (*x12018*).

En el control de calidad de los *reads* se evaluaron distintas opciones de *cutadapt* hasta llegar a la aplicada y la combinación de los resultados de *cutadapt* con la herramienta [PRINSEQ](#). La herramienta *PRINSEQ*, además de obtener gráficas de calidad semejantes a las obtenidas con *fastqc*, permite, con la versión *standalone lite*, depurar los *reads* por distintos criterios de filtrado y recorte.

Sobre la salida de registros filtrados por *cutadapt* (*-q 28*), ya eliminado el *adapter*, se probaron las siguientes opciones de *prinseq-lite.pl*:

-trim_tail_left ;elimina colas de poli-A/T en el extremo 3' de las secuencias. A pesar de que se apreciaron posibles colas de poli-A/T, el programa apenas detecta secuencias con esta característica.

-min_qual_mean 28 ;elimina secuencias con calidad media por debajo de 28. No se consiguió una mejora importante en la curva de calidad por bases. Subiendo el parámetro a 32 se consigue bastante mejora, pero

se eliminan cerca del 5% de los *reads* (sumados al 4% eliminado por *cutadapt*) y se considera que se eliminan demasiados.

-lc_method entropy -lc_threshold 80 ;se utiliza el método de entropía para seleccionar las secuencias de mayor complejidad y eliminar las que sean más repetitivas. Con un umbral de 80, se eliminaron más del 4% de las secuencias y no se consiguió una mejora sustancial de la curva de calidad por bases.

Tras estas pruebas, se decide utilizar *cutadapt*, con los parámetros indicados al inicio de este apartado, en combinación con *fastqc* como herramienta de control de calidad, por ser su uso más sencillo que *PRINSEQ* y obtener unos buenos perfiles de calidad en las muestras analizadas.

2.4.3 Preparación de la base de datos.

En principio, no es necesario descargar toda la base de datos de *miRBase*; en general sólo son necesarios los ficheros [hairpin.fa](#) y [miRNA.str](#). Se descargan y se dejan descomprimidos en *~TFM/datos/database*

Es altamente recomendable filtrar *hairpin.fa* para que contenga sólo la especie recomendada. Para ello se desarrolla el *script* *filtraHairpin.sh*

```
#!/bin/bash

#script que selecciona una especie del fichero hairpin.fa
#el argumento 1 es la especie (por ejemplo: mmu, para raton; hsa, para hombre)
#el argumento 2 es el fichero hairpin.fa

#salvamos el fichero original comprimido con extension .tot
gzip -fc $2 > $2.tot.gz

#seleccionamos los registros de la especie indicada
gawk -v especie=$1 'BEGIN{s=0}{
#si es fila de inicio de registro iniciamos s=0
    if($0 ~ /^>/){
        s=0;
    }
#si es fila de inicio de registro de la especie seleccionada imprimimos registro
    if($0 ~ /^>/ && $0 ~ especie) {
        print $0;
        s=1;
    }
    if($0 !~ /^>/ && s==1){
        print $0;
    }
}' $2 > $2.sp #salvamos fichero con extension .sp

rm $2 #borramos fichero hairpin.fa original
mv $2.sp $2 #renombramos como hairpin.fa el fichero con solo la especie seleccionada
```

Se ejecuta el *script* para la especie a filtrar, en este caso ratón, *mmu*.

```
$ filtraHairpin.sh mmu ~/TFM/datos/database/hairpin.fa
```

El fichero original, con todas las especies, se salva como *hairpin.fa.tot.gz* y en *hairpin.fa* sólo quedan las secuencias de precursor miRNA de ratón (*mmu*).

2.4.4 Anotación de miRNA/isomiR con Miraligner.

Para mapear las secuencias de las muestras de los 2 experimentos frente a la base de datos miRBase se utilizará la herramienta [Miraligner](#). Es un programa Java; para ejecutarlo, se emplea el comando.

```
java -jar miraligner.jar -sub 1 -trim 3 -add 3 -s mmu -i fich_ent -db
database -o path_sal
```

Opciones del comando:

-sub	Missmatches permitidos (0/1)
-trim	Nucleótidos permitidos para recorte (máximo 3)
-add	Nucleótidos permitidos para adición (máximo 3)
-s	Especie, 3 letras (hsa, hombre; mmu, ratón)
-i	Fichero entrada
-o	Path salida

A la salida se generan 3 ficheros:

*.*mirna.opt* información del proceso.

*.*nomap* secuencias no mapeadas

*.*mirna* secuencias mapeadas

Ejemplo de un registro del fichero **.mirna*:

```
seq  name  freq  mir  start end  mism  add  t5  t3  s5  s3  DB
precursor  ambiguity
ATCGGATCTGTCTGAGCTTGGT  seq_100016_x3  3  mmu-miR-127-3p  42  62
9TC  T  A  ct  ATCATCGG  GGCTGGTC  miRNA mmu-mir-127  1
```

Los campos del fichero **.mirna* son:

seq	Secuencia
name	Nombre de la secuencia
freq	Frecuencia
mir	Nombre del miRNA
start	Inicio de la secuencia en el precursor
end	Fin de la secuencia en el precursor
mism	Posición nuc. sustituido nuc. en la secuencia nuc. en el precursor
add	Nucleótidos añadidos al extremo 3
t5	Nuc. en el extremo 5 diferentes de la secuencia anotada en miRBase
t3	Nuc. en el extremo 3 diferentes de la secuencia anotada en miRBase
s5	Nuc. offset al principio del miRNA anotado
s3	Nuc. offset al final del miRNA anotado

DB	Base de datos
precursor	Nombre del precursor
ambiguity	N.º de precursores detectados diferentes

2.4.5 Conversión a formato GFF3-srna

Se usa la herramienta *mirtop* para convertir a formato *GFF3-srna*, que es el formato adecuado para compartir información de isomiR. Se descarga el fichero [mmu.gff3](#) y se deja en *~TFM/datos/database*

El uso del comando es:

```
mirtop gff --format seqbuster --sps mmu --hairpin hairpin.fa --gtf
mmu.gff3 -o path_salida fich_ent.mirna
```

Para ejecutar los procesos *miraligner* y *mirtop* se diseña el *script EjecutaMiraligner.sh*

```
#!/bin/bash

#Ejecuta miraligner sobre los ficheros clean_trimmed.fastq
#despues de eliminar aligners con cutadapt
#y agrupar secuencias con collapse quitando reads x1
#según los prefijos del fichero AccList que se pasa como parametro
#se toma el numero de fichero AccList como numero de experimento para
separar la salida

gawk -v file=$1 'BEGIN{n=substr(file,11,1)}
  {if($1 ~ /^SRR/){
    print("Miraligner " $1);
    system("java -jar ~/miraligner/miraligner.jar -sub 1 -trim 3 -add 3
-s mmu -i ~/TFM/datos/collapse/" $1 "_clean_trimmed.fastq -db
~/TFM/datos/database -o ~/TFM/datos/miraligner/exp" n "/" $1 );
  }
}' $1

#ejecuta mirtop sobre la salida de miraligner para dar formato GFF3-srna
gawk -v file=$1 'BEGIN{n=substr(file,11,1)}
  {if($1 ~ /^SRR/){
    print("Mirtop " $1);
    system("mirtop gff --format seqbuster --sps mmu --hairpin
~/TFM/datos/database/hairpin.fa --gtf ~/TFM/datos/database/mmu.gff3 -o
~/TFM/datos/GFF3_srna/exp" n " ~/TFM/datos/miraligner/exp" n "/" $1
".mirna");
  }
}' $1
```

Se ejecuta el *script* para los 2 experimentos:

```
$ EjecutaMiraligner.sh SraAccList1.txt
$ EjecutaMiraligner.sh SraAccList2.txt
```

Ejecuta secuencialmente *miraligner* para todas las muestras del fichero *SraAccList.txt* pasado como argumento; toma como fichero de origen el procesado anteriormente con *cutadapt* (para eliminar *adapters*) y *collapse* (para unir todas las secuencias iguales en un solo registro con indicador de frecuencia) con nombre *SRRx_clean_trimmed.fastq*, la salida se reparte en 2 subdirectorios, por experimento, para tenerlos más ordenados.

```
~/TFM/datos/miraligner/exp1/SRRx.mirna
~/TFM/datos/miraligner/exp2/SRRx.mirna
```

Los ficheros *.mirna resultantes se procesan con *mirtop* para convertirlos a formato *GFF3-srna*. Los ficheros de salida se reparten en 2 subdirectorios.

```
~/TFM/datos/GFF3_srna/exp1/SRRx.gff
~/TFM/datos/GFF3_srna/exp2/SRRx.gff
```

Un ejemplo de registro en formato *.gff:

```
mmu-let-7b miRBasev22 isomiR      4      27      0      +      .      Read
GATGAGGTAGTAGGTTGTGTGGTT; UID iso-24-5URPV39W2E; Name mmu-let-7b-5p; Parent
mmu-let-7b; Variant iso_snv,iso_5p:-2; Cigar MA22M; Expression 1; Filter
Pass; Hits 1;
```

2.5 Control de calidad de datos y normalización.

Los ficheros de salida *.mirna de la herramienta *Miraligner* se utilizarán para el análisis de miRNAs/isomiR expresados diferencialmente. Para ello se empleará el entorno software R con paquetes propios del proyecto *Bioconductor* para análisis de experimentos de RNA-seq, como *DESeq2* [5] [6] y otros específicos del análisis de miRNAs/isomiRs, como *isomiRs* [7].

Como paso previo a este análisis se hará un control de calidad de las muestras y se estudiará la necesidad o no de normalización de las mismas. Se tratarán los datos de los dos experimentos por separado.

2.5.1 Experimento 1 (ratones mecp2+/-)

Para cargar los ficheros *.mirna de las muestras se emplea la función *IsomirDataSeqFromFiles* del paquete *isomiRs*. Se crea una *whitelist* con todas las secuencias presentes en los ficheros; se esta manera se asegura que se cargarán todas las secuencias mapeadas con *miraligner*.

```
f_list1<-
c("~/TFM/datos/miraligner/exp1/SRR1504855.mirna", "~/TFM/datos/miraligner/exp1/S
RR1504856.mirna")
de1 <- data.frame(row.names=c("mutante1" , "control1"), condition = c("mecp2kd",
"wt"))
sequences1<-read.csv("~/TFM/datos/miraligner/exp1/seq.txt", header = FALSE)
sequences1<-as.vector(sequences1$V1)
ids1 <- IsomirDataSeqFromFiles(f_list1, coldata=de1, whitelist=sequences1)
## Total samples filtered due to low number of hits: 0
head(counts(ids1))
##
##          mutante1 control1
## mmu-let-7a-1-3p      211      187
## mmu-let-7a-2-3p         5         4
## mmu-let-7a-5p    946176    773333
```

Se crea un objeto de la clase *IsomirDataSeq*, que almacena los datos en crudo de los ficheros de entrada y datos procesados para visualización y análisis. La función *counts* del paquete *DESeq2* obtiene una matriz de cuentas para cada miRNA en las distintas muestras.

Se representa un *boxplot* y un diagrama de densidad de los datos en crudo, utilizando *pseudocounts*.

```
pseudoCount1<-log2(counts(ids1)+1)
library(reshape2)
df1<-melt(pseudoCount1)
df1<-data.frame(df1,Condition=substr(df1$Var2,1,7))
colnames(df1)<-c("isomiR","sample","value","Condition")
library(ggplot2)
ggplot(df1,aes(x=sample,y=value,fill=Condition))+geom_boxplot()+xlab("")
+ylab(expression(log[2](count+1)))
ggplot(df1,aes(x=value,colour=sample,fill=sample))+ylim(c(0,0.15))
+geom_density(alpha=0.2,size=1.25)+
  facet_wrap(~Condition)+theme(legend.position="top")+xlab(expression(log[2]
(count+1)))
```

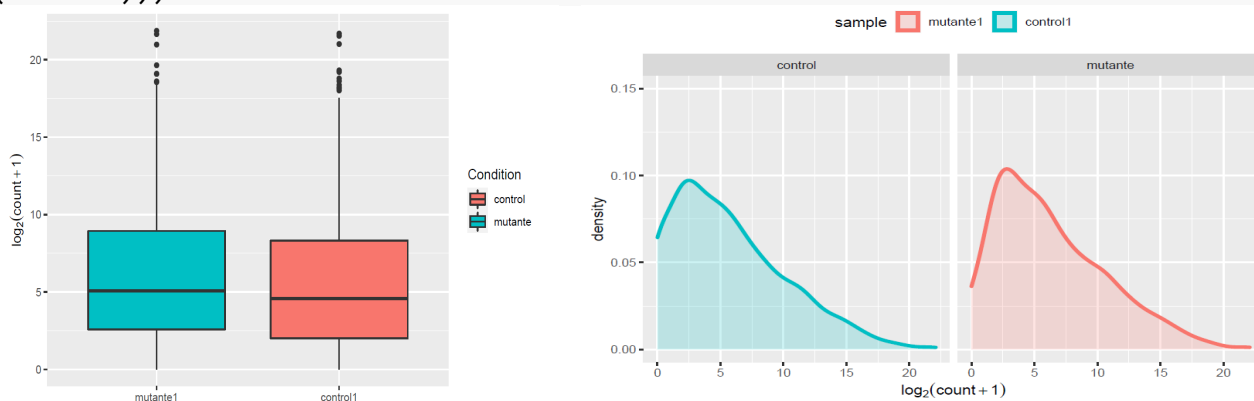


Figura 6. Boxplot y diagrama de densidad de las muestras del experimento 1. Mutante mecp2+/- frente a control (wt).

Se comparan las dos muestras con un diagrama MA. El diagrama MA es un *plot* de *log-fold change* (valores M. \log_2 de la ratio de cuentas para cada miRNA entre las dos muestras) frente a *log-average* (valores A. media de nivel de cuentas para cada miRNA).

```
x<-pseudoCount1[,1]
y<-pseudoCount1[,2]
M<-x-y
A<-(x+y)/2
dfMA<-data.frame(A,M)
ggplot(dfMA,aes(x=A,y=M))
+geom_point(size=1.5,alpha=1/5)+geom_hline(yintercept=0,color="bLue3")
+stat_smooth(se=FALSE,method="Loess",color="red3")
```

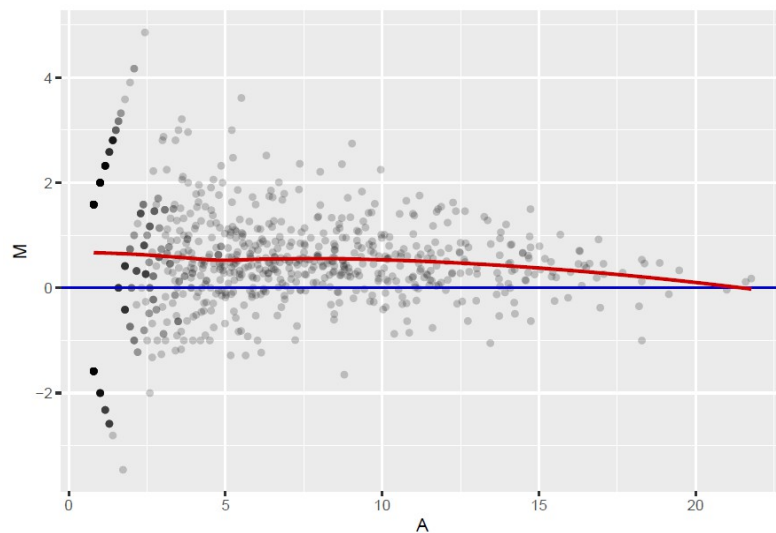



Figura 7. Plot MA de las muestras del experimento 1. Mutante frente a control.

La línea roja representa la desviación de la distribución entre las dos muestras. Las dos muestras están bastante equilibradas en cuanto a la distribución de las cuentas en los diferentes miRNAs. No tiene sentido hacer normalización porque hay una sola muestra por condición experimental.

2.5.2 Experimento 2 (ratones foxg1+/-)

Se cargan los ficheros de las muestras.

```
f_list2<-
c("~/TFM/datos/miraligner/exp2/SRR6061838.mirna", "~/TFM/datos/miraligner/exp2/SRR6061839.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061840.mirna", "~/TFM/datos/miraligner/exp2/SRR6061841.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061842.mirna", "~/TFM/datos/miraligner/exp2/SRR6061843.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061844.mirna", "~/TFM/datos/miraligner/exp2/SRR6061845.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061846.mirna", "~/TFM/datos/miraligner/exp2/SRR6061847.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061848.mirna", "~/TFM/datos/miraligner/exp2/SRR6061849.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061850.mirna", "~/TFM/datos/miraligner/exp2/SRR6061851.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061852.mirna", "~/TFM/datos/miraligner/exp2/SRR6061853.mirna",
  "~/TFM/datos/miraligner/exp2/SRR6061854.mirna", "~/TFM/datos/miraligner/exp2/SRR6061855.mirna")
de2 <-
data.frame(row.names=c("mutante1", "mutante2", "mutante3", "mutante4", "mutante5", "
mutante6", "mutante7", "mutante8", "mutante9", "control1", "control2", "control3",
"control4", "control5", "control6", "control7", "control8", "control9")
           , condition = c(rep("foxg1kd", 9), rep("wt", 9)))
sequences2<-as.vector(read.csv("~/TFM/datos/miraligner/exp2/seq.txt", header =
FALSE))
ids2 <- IsomirDataSeqFromFiles(f_list2, coldata=de2, whitelist=sequences2)
## Total samples filtered due to low number of hits: 0
head(counts(ids2))
##
##      mutante1 mutante2 mutante3 mutante4 mutante5 mutante6 mutante7
## mmu-let-7a-1-3p      3      96      125      190      166      93      184
## mmu-let-7a-2-3p      0       3       4       2       4       2       8
## mmu-let-7a-5p    14797   35802   29717   28325   24103   16278   22485
##
##      mutante8 mutante9 control1 control2 control3 control4 control5
## mmu-let-7a-1-3p      94      97      36      266      271      188      176
## mmu-let-7a-2-3p       6       0       0      11      12       5       6
## mmu-let-7a-5p    11870   21814   21339   42307   39615   38523   31971
```

```
##          control6 control7 control8 control9
## mmu-let-7a-1-3p      168      136      231      148
## mmu-let-7a-2-3p         6         2         11         5
## mmu-let-7a-5p     32536    14400    17486    25886
```

Se hace un *boxplot* y un diagrama de densidad de los datos en crudo, utilizando *pseudocounts*.

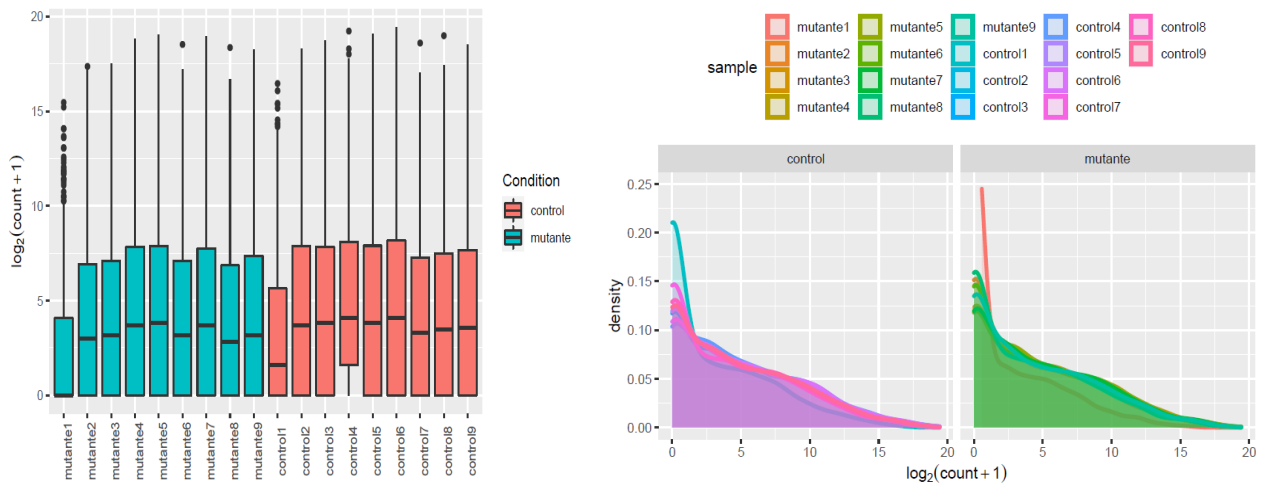


Figura 8. Boxplot y diagrama de densidad de las muestras del experimento 2. Ratones mutantes *foxg1+/-* frente a ratones control (*wt*).

Se hace un diagramas MA entre dos muestras mutantes, 2 y 3 (se descarta la 1 porque muestra un perfil muy diferente) y entre dos muestras de control, 1 y 2.

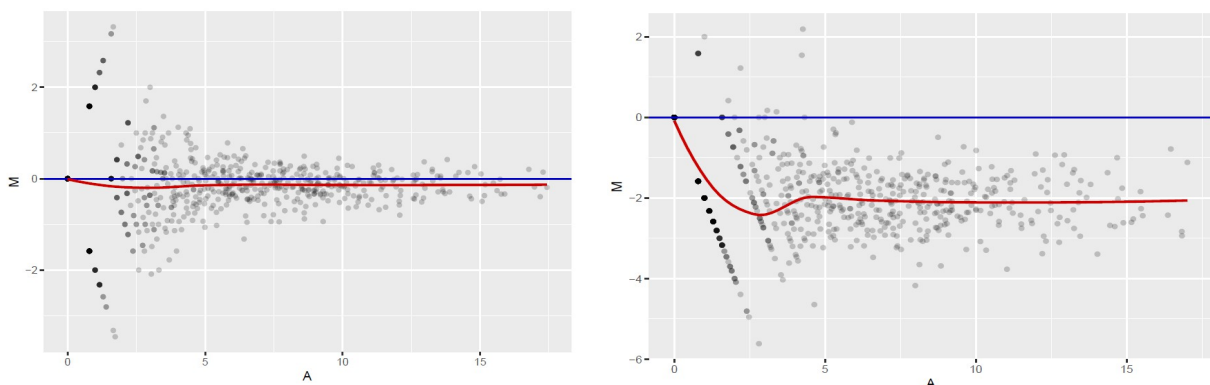


Figura 9. Diagramas MA de las muestras mutante2 frente a mutante3 y control1 frente a control2 de muestras del experimento 2 (ratones *foxg1+/-*).

En este caso sí que es necesaria la normalización, se observan distintos perfiles en muestras de la misma condición. Se hace la normalización y se muestra el *boxplot* y el diagrama de densidad. La función *isoNorm* realiza la normalización utilizando *rlog* del paquete *DESeq2*.

```
ids2<-isoNorm(ids2,formula = ~condition)
## converting counts to integer mode
pseudoCount2<-counts(ids2,norm=TRUE)
```

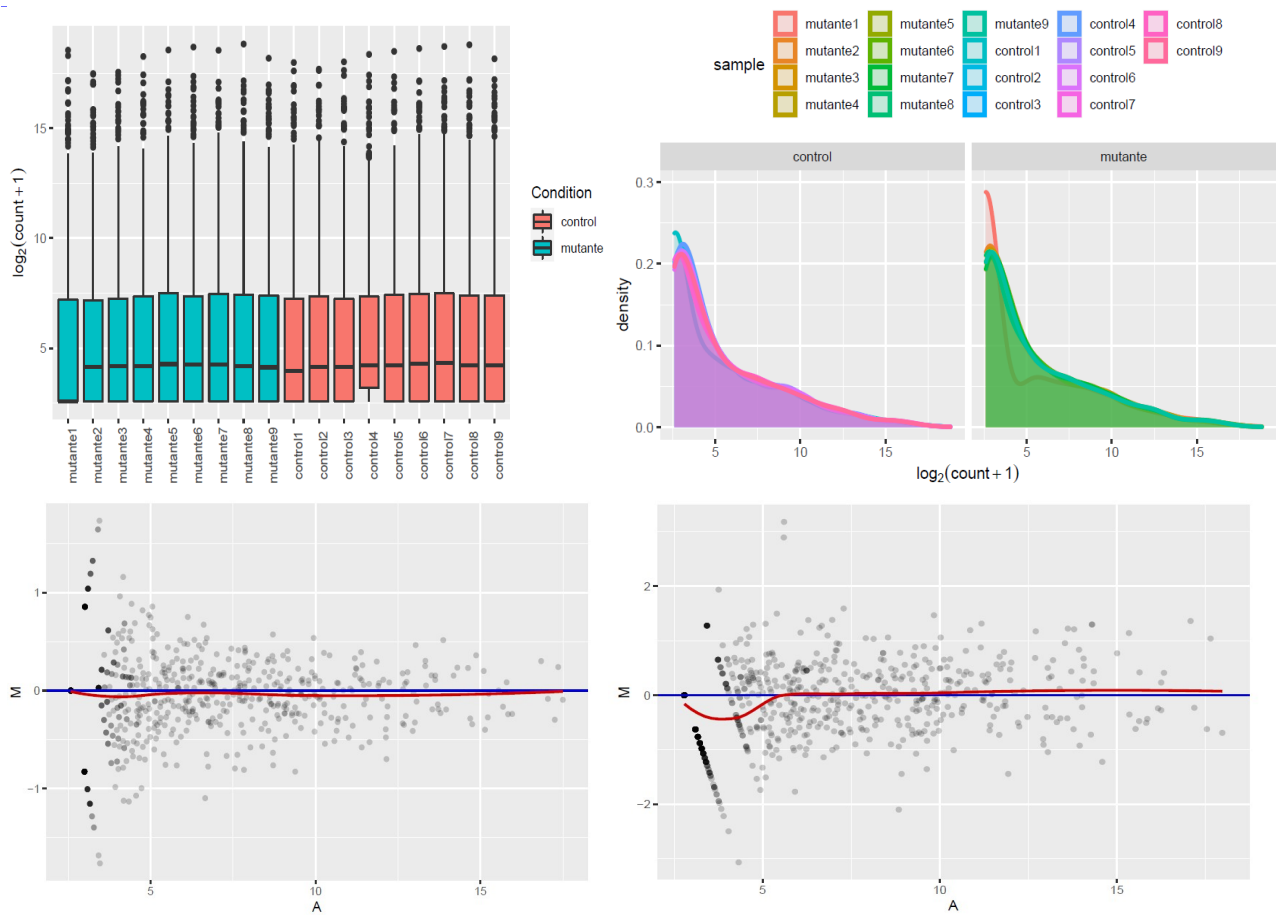


Figura 10. Boxplot, diagrama de densidad y diagramas MA de muestras del experimento 2 (ratones foxg1+/-) después del proceso de normalización. Los diagramas MA se corresponden a las muestras mutante2 frente a mutante3 y control1 frente a control2.

Se observa que las muestras se han normalizado, excepto la mutante1, que muestra un perfil de densidad muy diferente y la control1, que también se desvía de la distribución del resto de muestras de control. En ambos diagramas MA se aprecia un mejor ajuste de su distribución tras la normalización, aunque la muestra control1 sigue mostrando una desviación apreciable respecto a control2.

Se hace una representación MDS de las 2 primeras dimensiones y un dendograma de las muestras del experimento 2. El propósito del MDS (multidimensional scaling) es ofrecer una representación visual de los patrones de proximidad entre muestras. El dendograma es una representación de la distancia entre muestras que permite agruparlas en *clusters*.

```
#MDS
x<-t(as.matrix(counts(ids2,norm=TRUE)))
D<-dist(x)
mds2<-cmdscale(D,eig=TRUE)
X<-mds2$points[,1:2]
colnames(X)<-c("Coord.1","Coord.2")
rownames(X)<-rownames(x)
plot(X,type="n")
text(X[,1],X[,2],labels=rownames(X),cex=0.7)
#dendograma
c<-hclust(D)
plot(c)
```

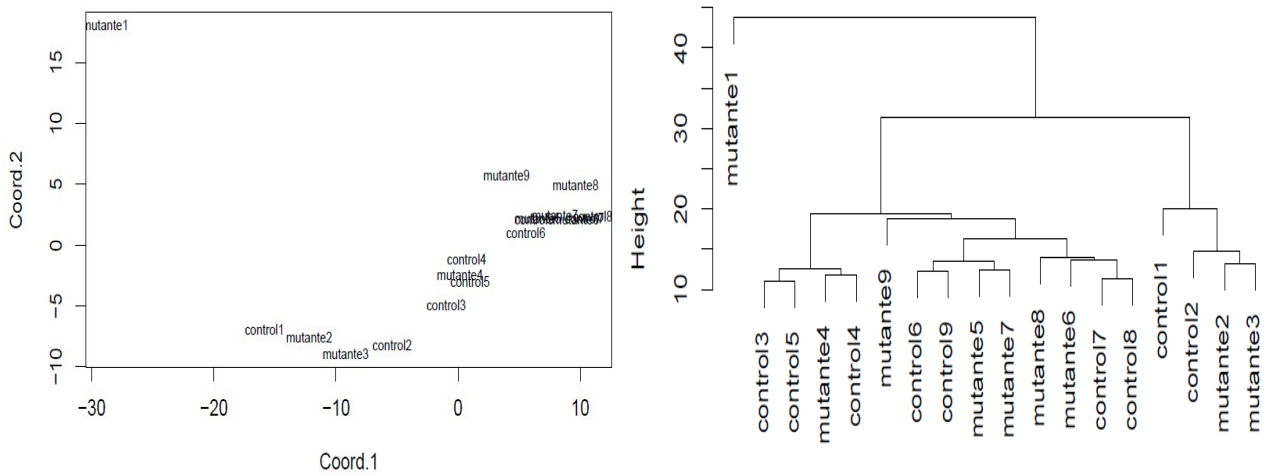
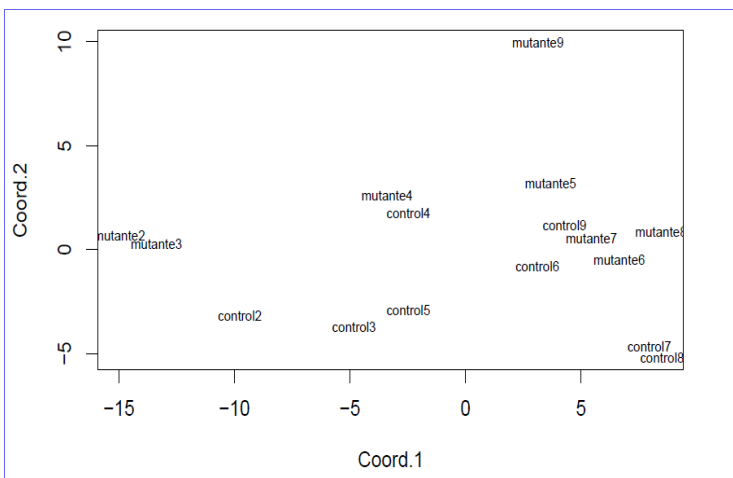


Figura 11. Representación MDS y dendrograma de las muestras del experimento 2 (ratones foxg1+/-).

Se aprecia claramente que la muestra mutante1 está muy alejada del resto y la control1 también muestra un perfil diferente al resto de muestras de control, por lo que se vuelven a cargar los datos sin estas dos muestras. No se aprecia una agrupación de las muestras mutantes frente a las de control.

```
head(counts(ids2,norm=TRUE))
##          mutante2  mutante3  mutante4  mutante5  mutante6  mutante7
## mmu-let-7a-1-3p  7.301568  7.480275  7.497241  7.342432  7.200009  7.557149
## mmu-let-7a-2-3p  4.387527  4.459664  4.071002  4.297185  4.215526  4.639647
## mmu-let-7a-5p   15.626349  15.182064  14.528583  14.313620  14.417722  14.309878
##          mutante8  mutante9  control12  control13  control14  control15
## mmu-let-7a-1-3p  7.516899  7.074673  7.869567  7.817304  7.213503  7.378615
## mmu-let-7a-2-3p  4.841502  3.525284  4.745711  4.762139  4.295301  4.451902
## mmu-let-7a-5p   14.311637  14.631419  15.038214  14.858848  14.660629  14.678051
##          control16  control17  control18  control19
## mmu-let-7a-1-3p  7.132599  7.610951  7.997761  7.332356
## mmu-let-7a-2-3p  4.386167  4.197624  4.890313  4.431522
## mmu-let-7a-5p   14.484449  14.163930  14.108814  14.570279
```

Se vuelve a repetir el diagrama MDS después de haber eliminado de la carga las 2 muestras.



Ahora sí se aprecia un cierto agrupamiento de muestras por condición con respecto a la segunda coordenada; las muestras de control se sitúan en la parte inferior del diagrama y las mutantes en la parte superior.

Figura 12. Diagrama MDS de las muestras del experimento 2 (ratones foxg1+/-) una vez eliminadas las muestras mutante1 y control1.

2.6 Análisis de expresión diferencial de miRNAs/isomiRs.

2.6.1 Experimento 1 (ratones *mecp2*+/-)

La función *isoCounts* permite obtener la matriz de cuentas por muestras diferenciadas por isomiRs, si se emplea sin parámetros se obtiene el mismo resultado que con la función *counts*; tan sólo se diferencian las variantes -3p y -5p de los miRNAs. Se realiza un gráfico de la diferenciación de tipo de isomiRs por condición con la función *isoPlot*.

```
#tipos de isomiRs
ids1c<-isoCounts(ids1)
isoPlot(ids1c,type="all")
## Ussing 12154 isomiRs.
```

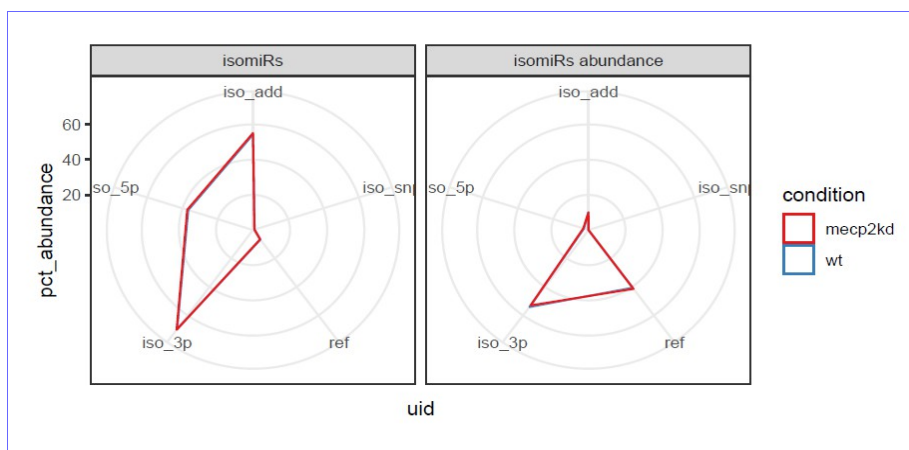


Figura 13. Gráfico isoPlot que muestra la diferenciación de isomiRs entre las 2 condiciones del experimento 1 (ratones *mecp2*+/-).

No se aprecian diferencias entre las 2 condiciones en la abundancia de isomiRs. En este experimento no hay réplicas biológicas. Al intentar hacer el análisis de expresión diferencial con métodos de los paquetes *isomiRs* y *DESeq2*, se produce el siguiente error:

"The design matrix has the same number of samples and coefficients to fit, so estimation of dispersion is not possible. Treating samples as replicates was deprecated in v1.20 and no longer supported since v1.22."

Análisis previos de miRNAs.

Como análisis previo, se comparten los datos de expresión diferencial por miRNAs del estudio de referencia con los obtenidos en este trabajo. Se descarga del artículo de referencia del experimento 1 [8] el fichero *mecpe2.xls*, que contiene las tablas:

Table S1A. RNA-seq Data of MeCP2 Knockout.

Table S1B. RNA-Seq of MeCP2 Overexpression.

Se copia la tabla *S1A* en el fichero *counts_exp1.csv* y se carga para comprobar los resultados de miRNAs DE que reporta el artículo. Según dice el artículo:

"Normalized with total reads, 314 of all 720 (43.6%) sequenced miRNAs from the hippocampus of mec2 Ko mice were found to be significantly upregulated for over 1.5-fold, as compared to those found in the wild-type (WT) littermates. In contrast, only 63 (8.7%) of them were downregulated."

En la tabla aparecen 355 miRNAs, todos *up-regulated*; aunque con *logFC* mayor que 1.5 sólo hay 66. Para hacer la comparación con los datos descargados y procesados en este

trabajo se tendrán en cuenta los miRNAs, sin diferenciar las variantes -3p y -5p; para ello se trunca el nombre del miRNA, dejando sólo el nombre principal.

```
#importamos la matriz de counts del paper del experimento 1
cexp1<-read.csv("~/TFM/R/counts_exp1.csv",header=TRUE,stringsAsFactors=FALSE )
cexp1<-data.frame(cexp1,log2(abs(cexp1$MeCP2.KO.count/cexp1$WT.count)))
colnames(cexp1)<-c(colnames(cexp1)[1:4],"logFC")
#truncamos el nombre del miR y agrupamos las cunetas por miR
r<-vector()
for (i in 1:nrow(cexp1)){
r<-c(r, paste(unlist(strsplit(cexp1[i,1],"-"))
[1],unlist(strsplit(cexp1[i,1],"-"))[2],sep="-"))
}
cexp1<-data.frame(r,cexp1)
r1<-unique(r)
cexp1a<-matrix(0,ncol=2)
for (i in 1:length(r1)) {
x<-subset(cexp1,cexp1$r==r1[i])
c<-apply(x[,3:4],2,sum)
f<-c(c[1:2])
cexp1a<-rbind(cexp1a,f)
}
cexp1a<-cexp1a[2:nrow(cexp1a),]
rownames(cexp1a)<-r1
#calculamos logFC para los miR agrupados y obtenemos DE por logFC > 1.5
cexp1a<-data.frame(cexp1a,log2(abs(cexp1a[,2]/cexp1a[,1])))
colnames(cexp1a)<-c(colnames(cexp1a)[1:2],"logFC")
cexp1aDE<-subset(cexp1a,cexp1a$logFC > 1.5)
rownames(cexp1aDE)
## [1] "miR-1249" "miR-129" "miR-5128" "miR-3059" "miR-3085" "miR-328"
## [7] "miR-7005p" "miR-874" "miR-376a" "miR-344c" "miR-677" "miR-1264"
## [13] "miR-410" "miR-496" "miR-879" "miR-7b" "miR-188" "miR-484"
```

Se obtienen 18 miRNAs *up-regulated* DE para los datos ofrecidos por el experimento 1, filtrando por $\logFC > 1.5$.

Se hace lo mismo con los datos descargados y procesados en este trabajo. Para hallar los miRNAs *up-reg* DE, se pone como condición que el número de *reads* en la muestra mutante1 sea mayor de 100 y se usa como criterio que $\logFC > 1.5$; como lo indicado en el artículo.

```
## [1] "miR-1249" "miR-300" "miR-3059" "miR-3078" "miR-3085" "miR-328"
## [7] "miR-3547" "miR-376a" "miR-410" "miR-484" "miR-496a" "miR-5099"
## [13] "miR-5128" "miR-758" "miR-7b" "miR-874" "miR-879"
```

Se han obtenido 17 miRNAs *up-regulated* DE para los datos procesados. Se representan las coincidencias de los miRNAs DE obtenidos de los datos del artículo y por el proceso de este trabajo en un diagrama de Venn.

```
#Library(Vennerable)
x1<-rownames(cexp1aDE)
x2<-rownames(c1aDEup)
lista_miR<-list(Art=x1,Pro=x2)
V<-Venn(lista_miR)
plot(V,type="circles")
```

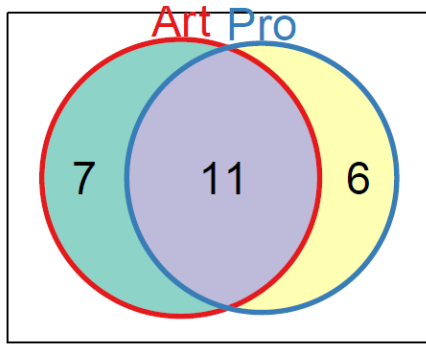


Figura 14. Diagrama de Venn comparando los resultados de miRNAs DE up-reg en Art) obtenidos del artículo de referencia y Pro) procesados y analizados en este trabajo.

Del conjunto de los 18 miRNAs identificados como DE *up-reg* con los datos del artículo del experimento 1, 11 coinciden con los identificados por el proceso y análisis de datos del presente trabajo. Aunque hay una coincidencia importante, también se muestran ciertas discrepancias. Para seleccionar los miRNAs DE no es suficiente utilizar el valor de *logFC* como único criterio; es conveniente

combinarlo con otro indicador, como el *p-valor*, obtenido por algún test de contraste de hipótesis. El hecho de que sólo se cuente con dos muestras, una para cada condición, también dificulta la determinación de los miRNAs diferencialmente expresados. Para solventar este problema de falta de réplicas se pueden utilizar distintos procedimientos de análisis.

Análisis con NOISEq.

El paquete *NOISEq* [10] utiliza un método no paramétrico para hacer análisis de expresión diferencial en experimentos NGS. Puede trabajar con réplicas técnicas reales, utilizando el método *NOISEq-real* o simular réplicas técnicas en experimentos sin réplicas con el método *NOISEq-sim*.

NOISEq calcula los estadísticos de expresión diferencial M (log2-ratio de las 2 condiciones) y D (diferencia entre condiciones). Para utilizar la función *noiseq* se ajustan los siguientes parámetros: *k=NULL*, los 0s se sustituyen por la mitad del siguiente valor no 0 en la matriz de expresión; *norm="n"*, indica el método de normalización ("n", no normalización); *pnr=0.2*, tamaño de las muestras simuladas como porcentaje de la profundidad de secuenciación; *nss=5*, número de réplicas simuladas; *v=0.02*, variabilidad de las muestras y *replicates="technical"* es el tipo de réplicas.

```
myfactors<-
data.frame(Condition=c("mutante","control"),ConditionRun=c("mutante1","control1"),Run=c("1","1"))
myfactors
##   Condition ConditionRun Run
## 1  mutante     mutante1   1
## 2  control     control1   1
mydata<-readData(data=counts(ids1),factors=myfactors)
rdds1<-
noiseq(mydata,factor="Condition",k=NULL,norm="n",pnr=0.2,nss=5,v=0.02,lc=1,replicates="technical")
## [1] "Computing (M,D) values..."
## [1] "Computing probability of differential expression..."
head(rdds1@results[[1]])
##           control_mean mutante_mean      M      D      prob
## mmu-let-7a-1-3p      187         211 -0.17420473    24 0.4358501
## mmu-let-7a-2-3p         4          5 -0.32192809     1 0.3455649
## mmu-let-7a-5p    773333    946176 -0.29101879 172843 0.6463041

rdds1DE<-subset(rdds1@results[[1]],rdds1@results[[1]]$prob>0.90)
rdds1DE
##           control_mean mutante_mean      M      D      prob
## mmu-miR-122-5p      946         2494 -1.398549   1548 0.9181626
```


## mmu-miR-1249-3p	17	173	-3.347165	156	0.9396515
## mmu-miR-126a-3p	23207	55781	-1.265214	32574	0.9241816
## mmu-miR-129-2-3p	232	1483	-2.676322	1251	0.9789335
## mmu-miR-136-5p	2640	7754	-1.554403	5114	0.9311510
## mmu-miR-137-3p	11501	24044	-1.063918	12543	0.9102429
## mmu-miR-153-3p	586	1524	-1.378890	938	0.9114572
## mmu-miR-187-3p	2556	5418	-1.083873	2862	0.9033791
## mmu-miR-203-3p	3699	8542	-1.207439	4843	0.9147307
## mmu-miR-206-3p	3661	9955	-1.443184	6294	0.9302006

Una característica (en este caso la expresión de un miRNA) se considera diferencialmente expresada si sus valores M y D son mayores que el ruido generado al comparar réplicas de la misma condición. Por esta comparación, *NOISeq* obtiene la probabilidad de expresión diferencial para esta característica.

Si el *odds ratio* $Pr(\text{expresión diferencial})/Pr(\text{no expresión diferencial})$ es mayor que un determinado umbral, la característica se considera diferencialmente expresada entre condiciones. Se ha establecido el umbral *prob* > 0.9 para obtener los miRNAs DE. Se seleccionan 43 miRNAs DE. Aquí se muestran los 10 primeros, el listado completo figura en el ANEXO 1.

Análisis con DESeq2.

Para poder hacer un análisis de expresión diferencial con *DESeq2* se necesitan, al menos, 3 réplicas de cada condición.

Para generar estas réplicas se usa el paquete *Polyester* [9]. Las muestras se generan con la función *simulate_experiment*. Esta función necesita como origen un fichero FASTA con la muestra de referencia, a partir de la cual se generan las réplicas.

Para obtener réplicas de la muestra SRR1504855 (ratones mecp2+/-) se parte del fichero *SRR1504855_clean_trimmed.fastq* (secuencias en las que se ha eliminado el *adapter* con *cutadapt* y se han unido en un solo registro las secuencias iguales con *collapse*). Se genera el fichero en formato FASTA, eliminando las 2 líneas de calidad de cada registro, se obtienen *SRR1504855_clean_trimmed.fasta* y un fichero con los *counts* de cada *read*, extraído del factor multiplicador de la cabecera de cada registro, *nt.txt*.

El código en R para generar las simulaciones es el siguiente:

```
#Library(polyester)
#Library(Biostrings)
numtx<-count_transcripts("~/TFM/pruebas/SRR1504855_clean_trimmed.fasta")
fold_change_values = sample(c(0.5, 1, 2), size=numtx*2,prob=c(0.05, 0.9, 0.05),
replace=TRUE)
fold_changes = matrix(fold_change_values, nrow=numtx)
nt<-read.csv("~/TFM/pruebas/nt.txt",header=FALSE)
nt<-as.vector(nt$V1)
basemeans<-nt*fold_changes
simulate_experiment(fasta("~/TFM/pruebas/SRR1504855_clean_trimmed.fasta",
outdir = "~/TFM/pruebas/", num_reps=2, reads_per_transcript = nt,
size=1,fold_changes = fold_changes)
```

En el parámetro *fold_changes* se genera una matriz de *fold_changes* aleatorios con número de filas igual al número de *reads* y número de columnas igual al número de muestras a simular, en este caso 2.

En el parámetro *reads_per_transcript* se pasa un vector con el número de *counts* por *read* (*nt*).

La función nos da el error: *no se encuentra objeto basemean*s; se declara como *basemean*s = *nt* * *fold_changes*. Parece un *bug* de la función, lo debería crear internamente.

Genera 4 muestras:

sample_01_1.fasta sample_01_2.fasta sample_02_1.fasta sample_02_2.fasta

Para el análisis se utilizan las muestras *01_1* y *02_1*.

Sobre estas muestras se ejecuta *collapse* y *miraligner*, para obtener los ficheros **.mirna* que se cargan en el objeto *IsomirDataSeq*.

Se hace lo mismo con la muestra SRR1504856 (ratones wt).

Se cargan los ficheros en un nuevo *IsomirDataSeq*.

```
f_list1<-c(
"~/TFM/datos/miraligner/exp1/SRR1504855.mirna", "~/TFM/datos/miraligner/exp1/SRR1504855_1.mirna",
"~/TFM/datos/miraligner/exp1/SRR1504855_2.mirna", "~/TFM/datos/miraligner/exp1/SRR1504856.mirna",
"~/TFM/datos/miraligner/exp1/SRR1504856_1.mirna", "~/TFM/datos/miraligner/exp1/SRR1504856_2.mirna"
)
de1 <- data.frame(row.names=c("mutante1", "mutante2", "mutante3", "control1",
"control2", "control3"),
condition = c(rep("mecp2kd",3), rep("wt",3)))
#al haberse generado nuevas secuencias con la simulación, no se emplea whitelist
ids1r <- IsomirDataSeqFromFiles(f_list1, coldata=de1)
## Total samples filtered due to low number of hits: 0
head(counts(ids1r))
##
##      mutante1 mutante2 mutante3 control1 control2 control3
## mmu-let-7a-2-3p      5      2      0      4      3      2
## mmu-let-7a-5p    796669  199994 1123133  633393  141346  82037

ids1r<-isoNorm(ids1r,formula = ~condition)
## converting counts to integer mode
head(counts(ids1r,norm=TRUE))
##
##      mutante1 mutante2 mutante3 control1 control2 control3
## mmu-let-7a-2-3p  2.799970  2.786777  1.608076  2.876860  3.285396  2.982888
## mmu-let-7a-5p  18.420368  17.714385  20.331468  18.602592  17.727525  16.901433
```

Se hace el análisis DE con *DESeq2*. El paquete *DESeq2* basa su análisis de expresión diferencial en la distribución binomial negativa. El objeto *IsomirDataSeq* se puede transformar en un objeto *DESeqDataSet* con la función *DESeqDataSetFromMatrix*

El análisis se hace en tres pasos:

- Normalización por estimación de los tamaños de las muestras y factores de normalización.
- Estimación de la dispersión.
- Ajuste de los datos en un modelo lineal general basado en la distribución binomial negativa y cálculo del test paramétrico de Wald.

```
#Library(DESeq2)
dds1r<-DESeqDataSetFromMatrix(counts(ids1r),colData(ids1r),design=~condition)
## converting counts to integer mode
dds1r<-estimateSizeFactors(dds1r)
dds1r<-estimateDispersions(dds1r)
## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
dds1r<-nbinomWaldTest(dds1r)
plotMA(dds1r)
```

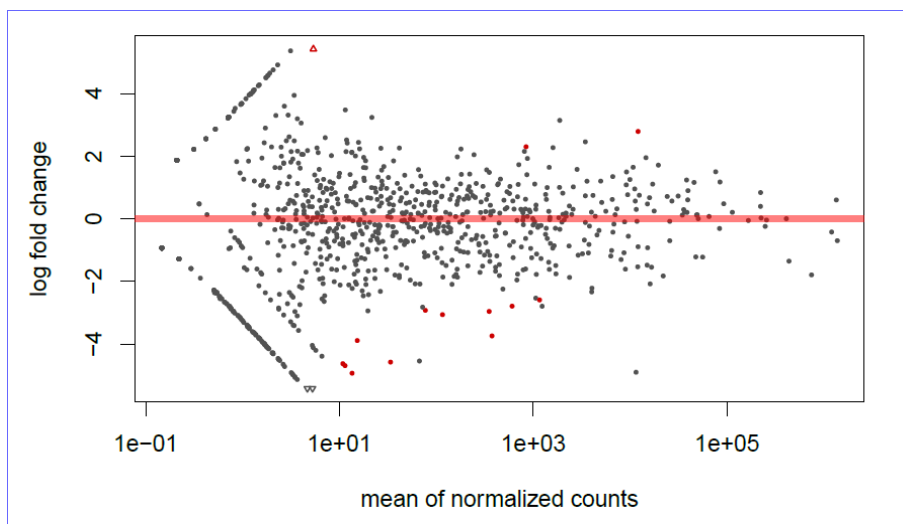


Figura 15. Gráfico MA entre las muestras mutantes y de control del experimento 1 (ratones *mecp2+/-*). Los puntos rojos indican miRNAs diferencialmente expresados con un p-valor ajust. < 0.1.

```
rdds1r<-results(dds1r,format="DataFrame")
rdds1rDE<-subset(rdds1r,rdds1r$padj<0.1)
rdds1rDE
## log2 fold change (MLE): condition wt vs mec2kd
## Wald test p-value: condition wt vs mec2kd
## DataFrame with 14 rows and 6 columns
##
```

	baseMean	log2FoldChange	lfcSE
	<numeric>	<numeric>	<numeric>
## mmu-miR-1247-5p	13.4930241134505	-4.93304692881724	1.41971855447812
## mmu-miR-1249-3p	33.6625638953492	-4.57820671824336	1.22454328266129
## mmu-miR-125b-1-3p	844.217503997863	2.29809578555483	0.713566461049782
## mmu-miR-129-2-3p	350.734310762634	-2.96062444395187	0.723462133449591
## mmu-miR-1298-5p	12096.1584694936	2.78735097437078	0.802361907189191
##
	stat	pvalue	padj
	<numeric>	<numeric>	<numeric>
## mmu-miR-1247-5p	-3.4746653928395	0.000511490937607038	0.0540000273995132
## mmu-miR-1249-3p	-3.73870551010136	0.000184970250512413	0.034080768656912
## mmu-miR-125b-1-3p	3.22057707445208	0.00127932787184828	0.0942864641552182
## mmu-miR-129-2-3p	-4.0923004910224	4.27114676501112e-05	0.015739175829066
## mmu-miR-1298-5p	3.47393233576522	0.000512890355219257	0.0540000273995132
##

Los resultados muestran los siguientes valores: *baseMean*, media normalizada de cuentas; *log2FoldChange*, log2 de la ratio de cuentas entre las 2 condiciones; *lfcse*, error estándar de log2FC; *stat*, estadístico del test; *pvalue*, p-valor de significación estadística de este cambio; *padj*, p-valor ajustado para pruebas múltiples por el método de Benjamini-Hochberg, con control FDR (false discovery rate).

Se han seleccionado 14 miRNAs DE, con un p-valor ajustado < 0.1. Aquí se muestran sólo los 4 primeros; el listado completo se puede consultar en el ANEXO 1.

Análisis con isomiRs.

La función *isoDE* del paquete *isomiRs* utiliza funciones del paquete *DESeq2* para el cálculo de expresión diferencial. Tiene parámetros para usar sólo un subconjunto de isomiRs; en esta caso se utiliza sin parámetros.

```
#analisis con isoDE
dds1ri<-isoDE(ids1r,formula=~condition)
## converting counts to integer mode
rdds1ri<-results(dds1ri,format="DataFrame")
rdds1riDE<-subset(rdds1ri,rdds1ri$padj<0.1)
rdds1riDE
## log2 fold change (MLE): condition wt vs mecp2kd
## Wald test p-value: condition wt vs mecp2kd
## DataFrame with 14 rows and 6 columns
##
##           baseMean  log2FoldChange      lfcSE
##           <numeric>      <numeric>      <numeric>
## mmu-miR-1247-5p  13.4930241134505 -4.93304692881724  1.41971855447812
## mmu-miR-1249-3p  33.6625638953492 -4.57820671824336  1.22454328266129
## mmu-miR-125b-1-3p 844.217503997863  2.29809578555483  0.713566461049782
## mmu-miR-129-2-3p 350.734310762634 -2.96062444395187  0.723462133449591
## mmu-miR-1298-5p 12096.1584694936  2.78735097437078  0.802361907189191
## ...
##           stat          pvalue      padj
##           <numeric>      <numeric>      <numeric>
## mmu-miR-1247-5p  -3.4746653928395  0.000511490937607038  0.0540000273995132
## mmu-miR-1249-3p  -3.73870551010136  0.000184970250512413  0.034080768656912
## mmu-miR-125b-1-3p 3.22057707445208  0.00127932787184828  0.0942864641552182
## mmu-miR-129-2-3p  -4.0923004910224  4.27114676501112e-05  0.015739175829066
## mmu-miR-1298-5p  3.47393233576522  0.000512890355219257  0.0540000273995132
## ...
##           ...
##           ...
##           ...

rownames(rdds1riDE)[order(rownames(rdds1riDE))]

## [1] "mmu-miR-1247-5p"  "mmu-miR-1249-3p"  "mmu-miR-125b-1-3p"
## [4] "mmu-miR-129-2-3p" "mmu-miR-1298-5p"  "mmu-miR-148a-5p"
## [7] "mmu-miR-200a-5p"  "mmu-miR-328-3p"   "mmu-miR-335-3p"
## [10] "mmu-miR-34b-5p"  "mmu-miR-376a-3p"  "mmu-miR-5128"
## [13] "mmu-miR-7090-3p"  "mmu-miR-92a-1-5p"
```

La tabla de resultados es idéntica a la obtenida en el análisis con *DESeq2*. Se han seleccionado 14 miRNAs DE, con un p-valor ajustado < 0.1.

Análisis con RankProd.

El paquete *RankProd* [11] utiliza el test no paramétrico Rank Product (RP) o el Rank Sum (RS). Ambos métodos se basan en suposiciones relativamente débiles:

- Sólo una minoría de todas las características medidas están desreguladas (up-reg/down-reg).
- Las medidas son independientes entre los experimentos replicados.
- La mayoría de los cambios son independientes unos de otros.
- La varianza de las medidas es aproximadamente igual en todas las medidas.

La función *RankProducts* tiene como parámetros: *data*, matriz de counts; *cl*, vector de clases y *gene.names*, vector con los nombres de los genes (en este caso miRNAs). Como

ya se ha hecho la normalización de los datos con \log_2 , se establece el parámetro `logged=TRUE`.

```
#Library(RankProd)
datos<-as.matrix(counts(ids1r,norm=TRUE))
clases<-c(rep(0,3),rep(1,3))
origen<-c(1,2,2,1,2,2) #si queremos considerar 2 orígenes, muestra real y simulada
miRnames<-rownames(datos)
#considerando sólo 1 origen
RP.out<-RankProducts(datos,clases,logged=TRUE,na.rm=FALSE,plot=FALSE,rand=123)
## Rank Product analysis for unpaired case
##
##
## done
```

Con la función `plotRP` se ve un gráfico del *pdf* (*percentage of false prediction*) para cada miRNA.

```
plotRP(RP.out,cutoff=0.1)
## No genes found using the input cutoff class 1 < class 2
## No genes found using the input cutoff: class 1 > class 2
```

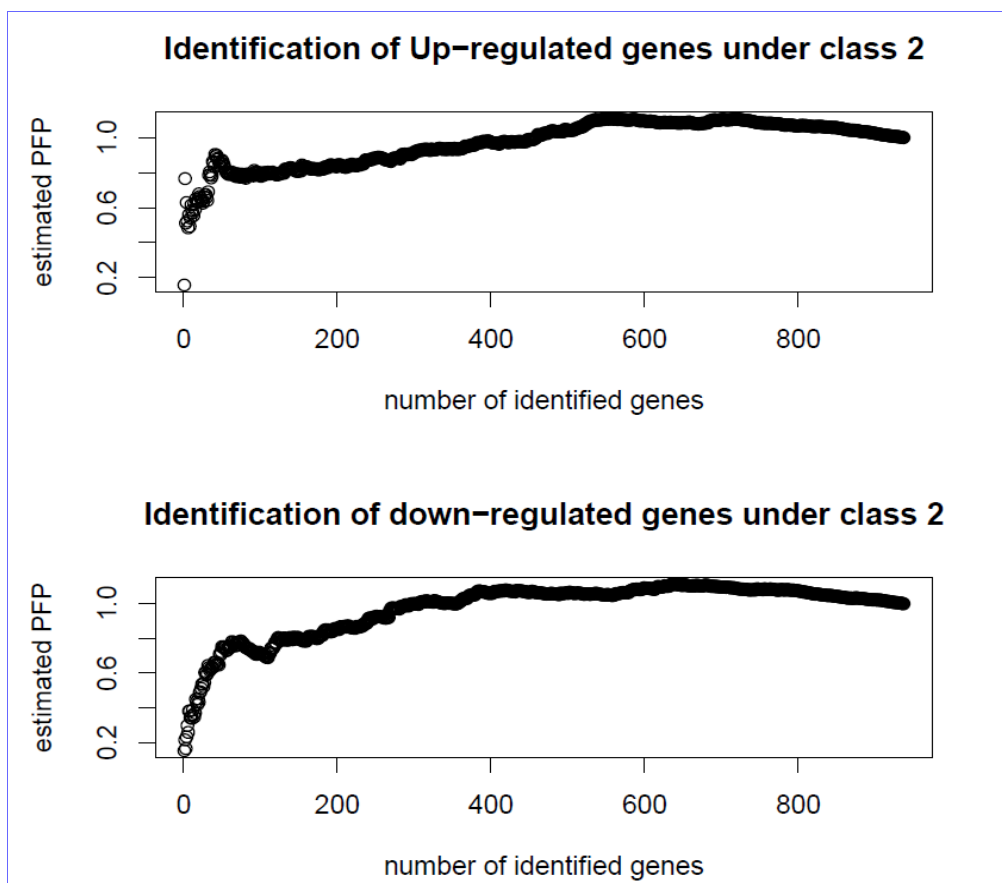


Figura 16. `plotRP`. Muestra los valores de *pdf* para los distintos miRNAs. Arriba, miRNAs up-reg bajo la clase 2 (control). Abajo, miRNAs down-reg bajo la clase 2 (control). En rojo se muestran los que tienen *pdf* < 0.1.

Con la función *topGene* se generan 2 tablas con los genes (en este caso miRNAs) identificados como DE según el criterio elegido, *pdf* o *p-value*. La tabla 1 nos da los miRNAs *up-reg* en la condición 2 (class 1 < class 2) y la tabla 2 los *down-reg* (class 1 > class 2).

El criterio *pdf* es más restrictivo, se generan las tablas de miRNAs DE con un umbral *pdf* < 0.1.

```
topGene(RP.out,cutoff = 0.1,method = "pdf", logged=TRUE, logbase = 2,gene.names
= miRnames)
## No genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
## NULL
##
## $Table2
## NULL
```

Por este criterio no se obtienen ningún miRNA DE. El criterio *pval* es menos restrictivo; el p-valor obtenido por la función es equivalente al p-valor ajustado por una corrección de prueba múltiple, como la de Benjamini-Hochberg; se establece un umbral *pval* < 0.05.

```
topGene(RP.out,cutoff = 0.05,method = "pval", logged=TRUE, logbase =
2,gene.names = miRnames)
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
##
##      gene.index  RP/Rsum  FC:(class1/class2)   pdf  P.value
## mmu-miR-672-5p      700    15.46      0.0639 0.1579 0.0001685
## mmu-miR-1298-3p     93     31.82      0.1608 0.7659 0.0016350
## mmu-miR-1968-5p    227    31.83      0.1691 0.5114 0.0016370
## mmu-miR-1298-5p     94    37.65      0.1829 0.6294 0.0026870
## mmu-miR-140-5p     123    38.17      0.1391 0.5241 0.0027970
## mmu-miR-574-5p     662    39.57      0.1893 0.4847 0.0031040
## mmu-miR-125b-1-3p   76    43.91      0.2097 0.5603 0.0041860
## mmu-miR-200a-5p    251    43.99      0.2106 0.4928 0.0042080
## mmu-miR-150-3p     142    47.38      0.2199 0.5403 0.0051900
## mmu-miR-186-3p     182    51.56      0.2251 0.6155 0.0065690
##
## $Table2
##
##      gene.index  RP/Rsum  FC:(class1/class2)   pdf  P.value
## mmu-miR-335-3p     451    15.33     13.3600 0.1533 0.0001637
## mmu-miR-300-3p     347    21.07     12.1000 0.2156 0.0004602
## mmu-miR-1249-3p    71     22.05     11.0900 0.1657 0.0005306
## mmu-miR-1b-5p     249    27.00      9.4130 0.2331 0.0009950
## mmu-miR-129-2-3p   91    31.54      7.5930 0.2985 0.0015930
## mmu-miR-1247-5p    70    31.97      7.3210 0.2590 0.0016590
## mmu-miR-376a-3p    518    38.42      6.5110 0.3815 0.0028500
## mmu-miR-3084-5p    397    40.24      6.4570 0.3817 0.0032590
## mmu-miR-7090-3p    841    40.37      6.3670 0.3425 0.0032900
## mmu-miR-296-5p    332    42.36      6.0340 0.3540 0.0037780
```

Se han seleccionado 59 miRNAs down-reg en la clase mutante y 62 up-reg. Aquí se muestran los 10 primeros, las tablas completas están en el ANEXO 1.

En ninguno de los procedimientos de análisis se identifica miR-134 como diferencialmente expresado. Este miRNA es uno de los que se indican como más relevantes en el estudio de referencia [8] por estar sobreexpresado en el síndrome de carencia de proteína MECP2. Los resultados de expresión diferencial para miR-134-5p encontrados en los distintos análisis son:

```
NOISeq:   prob = 0.696
DESeq2:   padj = 0.244
RankProd: p.value = 0.129
```

2.6.2 Experimento 2 (ratones *foxg1*+/-)

En este experimento, al disponer de suficientes réplicas por condición, se hace el análisis de expresión diferencial directamente con el paquete *isomiRs*.

Se hace un gráfico de la diferenciación de tipo de isomiRs por condición.

```
#tipos de isomiRs
ids2c<-isoCounts(ids2)
isoPlot(ids2c,type="all")
## Ussing 14684 isomiRs.
```

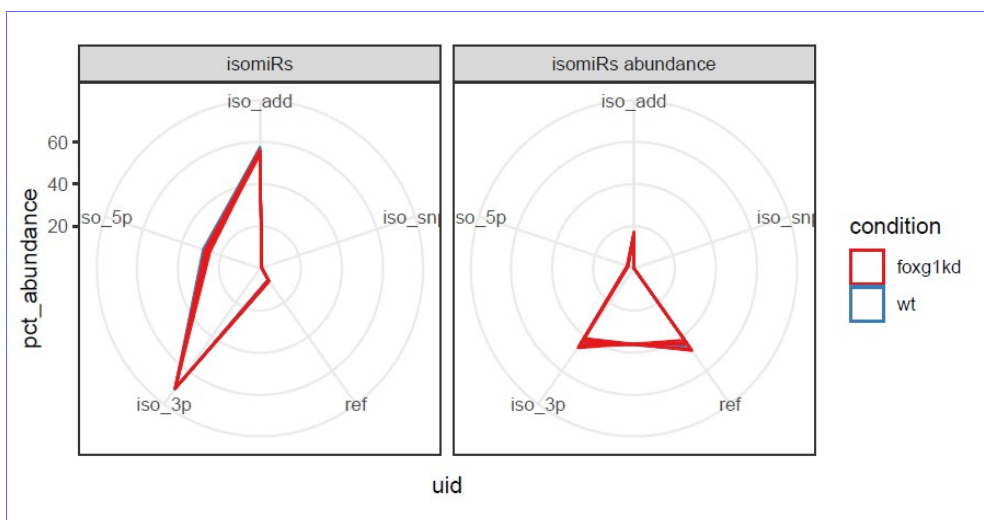


Figura 17. Gráfico isoPlot que muestra la diferenciación de isomiRs entre las 2 condiciones del experimento 2 (ratones *foxg1*+/-).

Muestra diferencias entre las 2 condiciones en isomiRs que difieren en el extremo 5'.

Análisis de expresión diferencial con isomiRs.

Se hace el análisis de expresión diferencial con la función *isoDE*, sin parámetros que seleccionen tipos de isomiRs.

```
#analisis DE
dds2<-isoDE(ids2,formula=~condition)
## converting counts to integer mode
plotMA(dds2)
```

```

rdds2<-results(dds2,format="DataFrame")
rdds2DE<-subset(rdds2,rdds2$padj<0.1) #padj < 0.10
rdds2DE
## log2 fold change (MLE): condition wt vs foxg1kd
## Wald test p-value: condition wt vs foxg1kd
## DataFrame with 19 rows and 6 columns
##
##           baseMean      log2FoldChange      lfcSE
##           <numeric>      <numeric>      <numeric>
## mmu-miR-1264-3p 116.944759377436 -1.18975388271475 0.259528642156348
## mmu-miR-1264-5p 10.1808050513326 -0.913339027758931 0.308967042459093
## mmu-miR-126a-5p 12473.4702568484 0.474339310031975 0.161183219783677
## mmu-miR-1298-5p 378.640663899583 -1.59465587299015 0.43128192790617
## mmu-miR-187-3p 1070.93563315195 0.493814244432877 0.0838708797725633
## ...
##
##           stat          pvalue          padj
##           <numeric>      <numeric>      <numeric>
## mmu-miR-1264-3p -4.58428739436784 4.55537325197631e-06 0.000661667964849559
## mmu-miR-1264-5p -2.95610502819198 0.00311550980082151 0.0994420212668773
## mmu-miR-126a-5p 2.94285788972688 0.00325197659908893 0.0994420212668773
## mmu-miR-1298-5p -3.69747900342601 0.00021775122375044 0.0140093658226092
## mmu-miR-187-3p 5.88779139758611 3.91390687722523e-09 2.27397989566786e-06
## ...

```

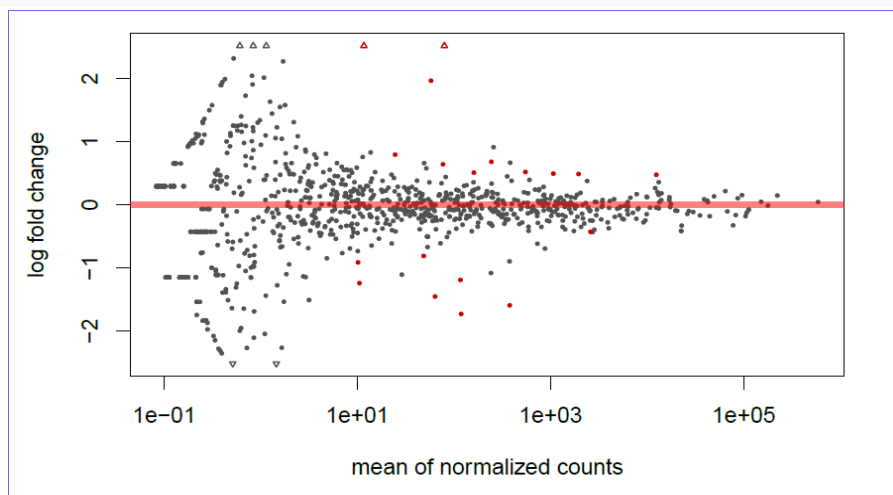


Figura 18. Gráfico MA entre las muestras mutantes y de control del experimento 1 (ratones mecp2+/-). Los puntos rojos indican miRNAs diferencialmente expresados con un p-valor ajust. < 0.1.

Se han seleccionado 19 miRNAs DE. Aquí se muestran sólo 5, para ver el listado completo consultar el ANEXO 1.

Comparación con los resultados del artículo.

Los resultados de miRNAs DE relacionados en el artículo de referencia [12] se muestran en la siguiente figura.

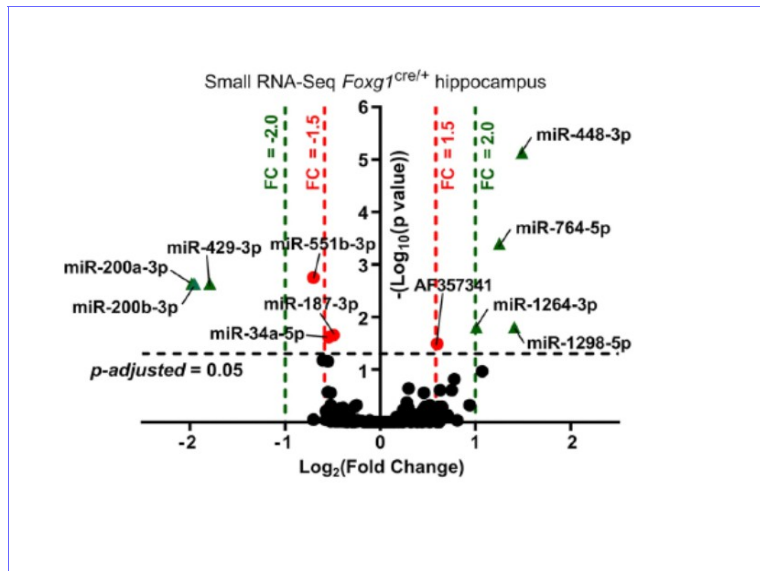


Figura 19. Volcano plot de miRNAs DE en el experimento 2 (ratones foxg1+/-) según el artículo de ref. [12].

Comparándolos con los obtenidos por el procesamiento de los datos en crudo, se obtienen los siguientes diagramas de Venn. Para los isomiRs *up-regulated* y *down-regulated*.

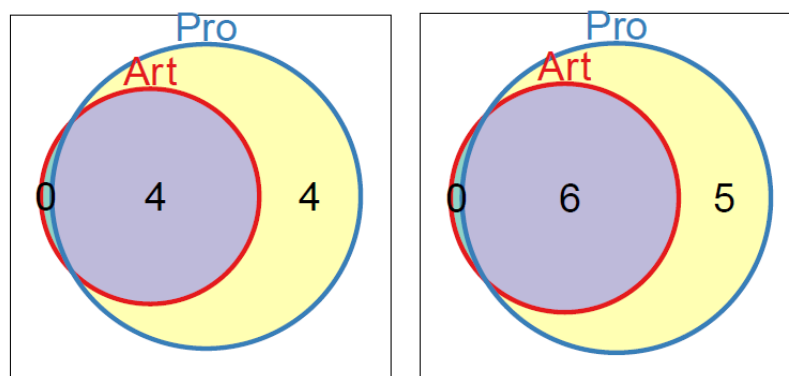


Figura 20. Diagramas de Venn comparando los miRNAs DE seleccionados en Art) Artículo de referencia y Pro) Procesados en este trabajo. A la izquierda los up-reg y a la derecha los down-reg.

Hay que tener en cuenta que los criterios de selección son algo diferentes, pero la coincidencia es notable.

Para comparar los miRNAs desregulados en ambos experimentos, mientras que en este segundo se comprueba que el procesamiento y análisis de datos en crudo de este trabajo obtiene datos muy similares a los ofrecidos por el artículo de referencia, en el primero se ha evidenciado la dificultad de obtener los mismos resultados y ante la ausencia de réplicas en el estudio se han valorado distintos análisis de expresión diferencial, observando que los resultados obtenidos con *RankProd* ofrecen un amplio espectro de miRNAs sobre el que trabajar.

Por ello, se decide hacer una análisis con *RankProd* de los datos del segundo experimento y hacer una comparación de miRNAs desregulados entre ambos experimentos utilizando como base los resultados obtenidos por este análisis.

Análisis con RankProd.

Se hace un análisis de miRNAs DE con *RankProd* para los datos del experimento 2.

```
datos2<-as.matrix(counts(ids2,norm=TRUE))
clases2<-c(rep(0,8),rep(1,8))
miRnames2<-rownames(datos2)
RP.out2<-
RankProducts(datos2,clases2,logged=TRUE,na.rm=FALSE,plot=FALSE,rand=123)
## Rank Product analysis for unpaired case
##
##
## done
plotRP(RP.out2,cutoff=0.1)
```

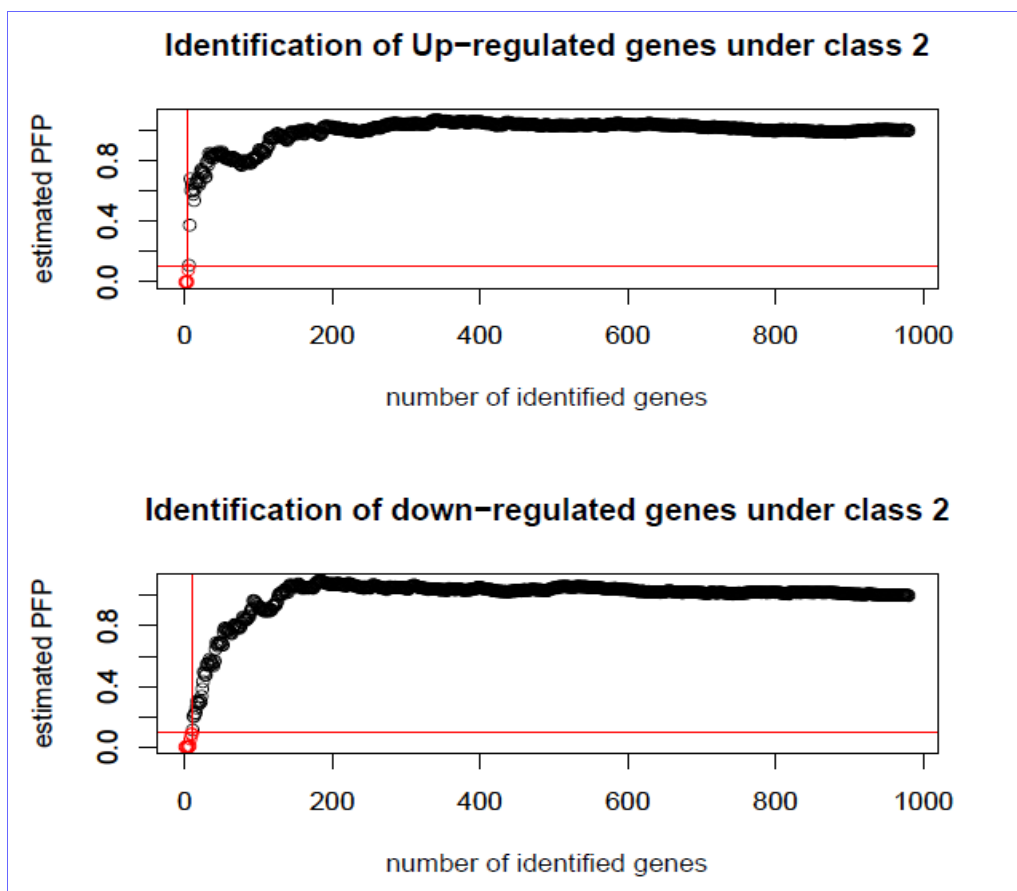


Figura 21. plotRP. Muestra los valores de pfp para los distintos miRNAs. Arriba, miRNAs up-reg bajo la clase 2 (control). Abajo, miRNAs down-reg bajo la clase 2 (control). En rojo se muestran los que tienen pfp < 0.1.

```
topGene(RP.out2,cutoff = 0.1,method = "pfp", logged=TRUE, logbase =
2,gene.names = miRnames2)
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
##          gene.index RP/Rsum FC:(class1/class2)      pfp  P.value
## mmu-miR-200b-3p      236    11.84           0.3072 1.447e-07 1.476e-10
```

```

## mmu-miR-200a-3p      235  18.03      0.4043 3.279e-06 6.691e-09
## mmu-miR-429-3p      527  20.09      0.4122 5.561e-06 1.702e-08
## mmu-miR-10b-5p      37   29.37      0.4969 9.419e-05 3.844e-07
## mmu-miR-383-5p     508  77.91      0.7029 7.648e-02 3.902e-04
##
## $Table2
##           gene.index RP/Rsum FC:(class1/class2)      pfp  P.value
## mmu-miR-448-3p      534  19.87      2.353 1.515e-05 1.546e-08
## mmu-miR-3552        473  30.29      1.993 2.400e-04 4.897e-07
## mmu-miR-1264-3p     77   36.46      1.921 6.672e-04 2.042e-06
## mmu-miR-1298-3p     90   36.56      1.783 5.107e-04 2.084e-06
## mmu-miR-1298-5p     91   38.00      2.094 5.471e-04 2.791e-06
## mmu-miR-582-3p     652  48.12      1.558 2.546e-03 1.559e-05
## mmu-miR-7080-3p     873  56.79      1.671 6.903e-03 4.931e-05
## mmu-miR-455-3p     541  78.89      1.507 5.167e-02 4.218e-04
## mmu-miR-532-5p     627  87.10      1.120 8.446e-02 7.757e-04
## mmu-miR-764-3p     908  88.08      1.515 8.131e-02 8.297e-04
topGene(RP.out2,cutoff = 0.05,method = "pval", logged=TRUE, logbase =
2, gene.names = miRnames2)
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
##           gene.index RP/Rsum FC:(class1/class2)      pfp  P.value
## mmu-miR-200b-3p     236  11.84      0.3072 1.447e-07 1.476e-10
## mmu-miR-200a-3p     235  18.03      0.4043 3.279e-06 6.691e-09
## mmu-miR-429-3p     527  20.09      0.4122 5.561e-06 1.702e-08
## mmu-miR-10b-5p      37   29.37      0.4969 9.419e-05 3.844e-07
## mmu-miR-383-5p     508  77.91      0.7029 7.648e-02 3.902e-04
.....
## $Table2
##           gene.index RP/Rsum FC:(class1/class2)      pfp  P.value
## mmu-miR-448-3p      534  19.87      2.3530 1.515e-05 1.546e-08
## mmu-miR-3552        473  30.29      1.9930 2.400e-04 4.897e-07
## mmu-miR-1264-3p     77   36.46      1.9210 6.672e-04 2.042e-06
## mmu-miR-1298-3p     90   36.56      1.7830 5.107e-04 2.084e-06
## mmu-miR-1298-5p     91   38.00      2.0940 5.471e-04 2.791e-06
.....

```

Con un umbral de $pval < 0.05$ se han obtenido 61 miRNAs down-reg en la clase mutante y 63 up-reg. Aquí se muestran los 5 primeros, para ver las tablas completas consultar el ANEXO 1.

2.6.3 Comparación entre ambos experimentos.

Al no disponer de réplicas en el experimento 1, se han hecho distintos análisis de expresión diferencial por distintos métodos.

NOISeq, método no paramétrico. Los miRNAs diferencialmente expresados con una probabilidad superior a 0.90 en el experimento 1 son:

```

## [1] "mmu-miR-122-5p" "mmu-miR-1249-3p" "mmu-miR-126a-3p" "mmu-miR-129-2-3p"
## [5] "mmu-miR-136-5p" "mmu-miR-137-3p" "mmu-miR-153-3p" "mmu-miR-187-3p"
## [9] "mmu-miR-203-3p" "mmu-miR-206-3p" "mmu-miR-296-5p" "mmu-miR-29b-3p"
## [13] "mmu-miR-300-3p" "mmu-miR-3059-5p" "mmu-miR-3085-3p" "mmu-miR-3086-5p"

```

```
## [17] "mmu-miR-323-3p" "mmu-miR-328-3p" "mmu-miR-335-3p" "mmu-miR-344-3p"
## [21] "mmu-miR-3547-3p" "mmu-miR-369-3p" "mmu-miR-370-3p" "mmu-miR-376a-3p"
## [25] "mmu-miR-376a-5p" "mmu-miR-380-3p" "mmu-miR-380-5p" "mmu-miR-410-3p"
## [29] "mmu-miR-412-5p" "mmu-miR-433-3p" "mmu-miR-433-5p" "mmu-miR-488-3p"
## [33] "mmu-miR-488-5p" "mmu-miR-496a-3p" "mmu-miR-5099" "mmu-miR-5128"
## [37] "mmu-miR-532-3p" "mmu-miR-770-5p" "mmu-miR-7a-5p" "mmu-miR-7b-5p"
## [41] "mmu-miR-874-3p" "mmu-miR-879-5p" "mmu-miR-92b-3p"
```

DESeq2, simulando réplicas con *polyester*. Los miRNAs diferencialmente expresados con un p-value ajustado inferior a 0.10 en el experimento 1 son:

```
## [1] "mmu-miR-1247-5p" "mmu-miR-1249-3p" "mmu-miR-125b-1-3p"
## [4] "mmu-miR-129-2-3p" "mmu-miR-1298-5p" "mmu-miR-148a-5p"
## [7] "mmu-miR-200a-5p" "mmu-miR-328-3p" "mmu-miR-335-3p"
## [10] "mmu-miR-34b-5p" "mmu-miR-376a-3p" "mmu-miR-5128"
## [13] "mmu-miR-7090-3p" "mmu-miR-92a-1-5p"
```

isomiRs. Las funciones que utiliza *isoDE* para calcular los miRNAs DE son las mismas que las del paquete *DESeq2*. Al haber empleado *isoDE* sin parámetros para la diferenciación de isomiRs, los resultados son los mismos que los anteriores.

RankProd, método no paramétrico con las mismas réplicas anteriores. Los miRNAs diferencialmente expresados con p-value inferior a 0.05 en el experimento 1 son:

```
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2

## [1] "mmu-miR-100-5p" "mmu-miR-1199-5p" "mmu-miR-12191-3p"
## [4] "mmu-miR-1224-5p" "mmu-miR-122b-3p" "mmu-miR-125b-1-3p"
## [7] "mmu-miR-126b-3p" "mmu-miR-129-5p" "mmu-miR-1298-3p"
## [10] "mmu-miR-1298-5p" "mmu-miR-129b-3p" "mmu-miR-140-5p"
## [13] "mmu-miR-150-3p" "mmu-miR-150-5p" "mmu-miR-181b-5p"
## [16] "mmu-miR-181c-5p" "mmu-miR-184-3p" "mmu-miR-1843a-3p"
## [19] "mmu-miR-186-3p" "mmu-miR-1912-3p" "mmu-miR-1912-5p"
## [22] "mmu-miR-192-5p" "mmu-miR-1931" "mmu-miR-1968-5p"
## [25] "mmu-miR-200a-5p" "mmu-miR-200b-5p" "mmu-miR-205-5p"
## [28] "mmu-miR-210-3p" "mmu-miR-212-5p" "mmu-miR-29b-1-5p"
## [31] "mmu-miR-3074-5p" "mmu-miR-3086-5p" "mmu-miR-30b-3p"
## [34] "mmu-miR-30e-5p" "mmu-miR-30f" "mmu-miR-3100-5p"
## [37] "mmu-miR-3102-3p" "mmu-miR-3470a" "mmu-miR-34b-3p"
## [40] "mmu-miR-34c-3p" "mmu-miR-3572-5p" "mmu-miR-362-5p"
## [43] "mmu-miR-382-5p" "mmu-miR-448-3p" "mmu-miR-450b-5p"
## [46] "mmu-miR-455-3p" "mmu-miR-539-3p" "mmu-miR-574-5p"
## [49] "mmu-miR-672-5p" "mmu-miR-673-5p" "mmu-miR-6970-5p"
## [52] "mmu-miR-6988-3p" "mmu-miR-7047-5p" "mmu-miR-7066-3p"
## [55] "mmu-miR-7668-5p" "mmu-miR-871-3p" "mmu-miR-9-3p"
## [58] "mmu-miR-99b-3p" "mmu-miR-9b-3p"

## [1] "mmu-miR-1247-5p" "mmu-miR-1249-3p" "mmu-miR-126a-3p" "mmu-miR-126b-3p"
## [5] "mmu-miR-129-2-3p" "mmu-miR-129b-5p" "mmu-miR-133a-3p" "mmu-miR-136-5p"
## [9] "mmu-miR-139-3p" "mmu-miR-141-3p" "mmu-miR-148a-5p" "mmu-miR-149-5p"
## [13] "mmu-miR-153-3p" "mmu-miR-154-3p" "mmu-miR-18a-5p" "mmu-miR-196b-5p"
## [17] "mmu-miR-19b-1-5p" "mmu-miR-1a-3p" "mmu-miR-1b-5p" "mmu-miR-210-5p"
## [21] "mmu-miR-219b-3p" "mmu-miR-223-3p" "mmu-miR-224-5p" "mmu-miR-296-5p"
## [25] "mmu-miR-29c-5p" "mmu-miR-300-3p" "mmu-miR-3059-5p" "mmu-miR-3061-3p"
```

```

## [29] "mmu-miR-3071-3p" "mmu-miR-3071-5p" "mmu-miR-3074-5p" "mmu-miR-3084-5p"
## [33] "mmu-miR-30f" "mmu-miR-323-3p" "mmu-miR-323-5p" "mmu-miR-328-3p"
## [37] "mmu-miR-33-5p" "mmu-miR-335-3p" "mmu-miR-34b-5p" "mmu-miR-350-5p"
## [41] "mmu-miR-3544-5p" "mmu-miR-376a-3p" "mmu-miR-410-3p" "mmu-miR-411-3p"
## [45] "mmu-miR-412-3p" "mmu-miR-423-3p" "mmu-miR-425-5p" "mmu-miR-431-5p"
## [49] "mmu-miR-488-5p" "mmu-miR-500-3p" "mmu-miR-505-5p" "mmu-miR-5128"
## [53] "mmu-miR-6928-3p" "mmu-miR-7010-5p" "mmu-miR-7070-3p" "mmu-miR-7090-3p"
## [57] "mmu-miR-758-3p" "mmu-miR-770-5p" "mmu-miR-92a-1-5p" "mmu-miR-92a-3p"
## [61] "mmu-miR-935" "mmu-miR-9b-3p"

```

En el experimento 2, como se dispone de réplicas biológicas, se ha hecho el análisis de miRNAs DE directamente con isomiRs. Los miRNAs diferencialmente expresados con una p-value ajustado inferior a 0.10 en el experimento 2 son:

```

## [1] "mmu-miR-1264-3p" "mmu-miR-1264-5p" "mmu-miR-126a-5p" "mmu-miR-1298-5p"
## [5] "mmu-miR-187-3p" "mmu-miR-193b-3p" "mmu-miR-195a-5p" "mmu-miR-200a-3p"
## [9] "mmu-miR-200b-3p" "mmu-miR-344c-3p" "mmu-miR-34a-5p" "mmu-miR-3552"
## [13] "mmu-miR-429-3p" "mmu-miR-448-3p" "mmu-miR-497a-5p" "mmu-miR-551b-3p"
## [17] "mmu-miR-6944-3p" "mmu-miR-764-3p" "mmu-miR-764-5p"

```

También se ha realizado el contraste con RankProd en el experimento 2.

```

## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
##
## [1] "mmu-let-7i-3p" "mmu-miR-10b-5p" "mmu-miR-12191-3p"
## [4] "mmu-miR-125b-2-3p" "mmu-miR-126a-5p" "mmu-miR-129-1-3p"
## [7] "mmu-miR-135a-1-3p" "mmu-miR-135a-5p" "mmu-miR-135b-5p"
## [10] "mmu-miR-143-3p" "mmu-miR-148a-3p" "mmu-miR-16-5p"
## [13] "mmu-miR-181c-3p" "mmu-miR-181d-5p" "mmu-miR-182-5p"
## [16] "mmu-miR-183-5p" "mmu-miR-1843b-3p" "mmu-miR-187-3p"
## [19] "mmu-miR-187-5p" "mmu-miR-188-5p" "mmu-miR-190b-5p"
## [22] "mmu-miR-193b-3p" "mmu-miR-195a-5p" "mmu-miR-1969"
## [25] "mmu-miR-200a-3p" "mmu-miR-200b-3p" "mmu-miR-30a-3p"
## [28] "mmu-miR-30d-3p" "mmu-miR-33-3p" "mmu-miR-344b-3p"
## [31] "mmu-miR-344c-3p" "mmu-miR-34a-5p" "mmu-miR-363-3p"
## [34] "mmu-miR-383-5p" "mmu-miR-409-5p" "mmu-miR-429-3p"
## [37] "mmu-miR-434-5p" "mmu-miR-466q" "mmu-miR-497a-5p"
## [40] "mmu-miR-501-3p" "mmu-miR-5099" "mmu-miR-5121"
## [43] "mmu-miR-532-5p" "mmu-miR-551b-3p" "mmu-miR-551b-5p"
## [46] "mmu-miR-5615-5p" "mmu-miR-582-3p" "mmu-miR-598-5p"
## [49] "mmu-miR-6926-3p" "mmu-miR-6941-3p" "mmu-miR-6952-3p"
## [52] "mmu-miR-6976-3p" "mmu-miR-6979-3p" "mmu-miR-7006-3p"
## [55] "mmu-miR-7093-3p" "mmu-miR-7117-3p" "mmu-miR-7651-5p"
## [58] "mmu-miR-8103" "mmu-miR-871-3p" "mmu-miR-92a-3p"
## [61] "mmu-miR-92b-5p"
## [1] "mmu-let-7f-5p" "mmu-miR-100-5p" "mmu-miR-10a-5p"
## [4] "mmu-miR-12193-3p" "mmu-miR-125b-2-3p" "mmu-miR-1264-3p"
## [7] "mmu-miR-1264-5p" "mmu-miR-1298-3p" "mmu-miR-1298-5p"
## [10] "mmu-miR-132-5p" "mmu-miR-143-3p" "mmu-miR-148a-3p"
## [13] "mmu-miR-1839-5p" "mmu-miR-1843b-3p" "mmu-miR-1912-3p"
## [16] "mmu-miR-212-3p" "mmu-miR-212-5p" "mmu-miR-221-5p"
## [19] "mmu-miR-3057-5p" "mmu-miR-3083-3p" "mmu-miR-31-5p"

```

```

## [22] "mmu-miR-3102-3p" "mmu-miR-340-3p" "mmu-miR-345-3p"
## [25] "mmu-miR-3475-3p" "mmu-miR-34b-3p" "mmu-miR-34b-5p"
## [28] "mmu-miR-34c-5p" "mmu-miR-3552" "mmu-miR-370-3p"
## [31] "mmu-miR-375-3p" "mmu-miR-409-5p" "mmu-miR-423-3p"
## [34] "mmu-miR-431-3p" "mmu-miR-434-5p" "mmu-miR-448-3p"
## [37] "mmu-miR-455-3p" "mmu-miR-466b-3p" "mmu-miR-466c-3p"
## [40] "mmu-miR-466e-3p" "mmu-miR-466p-3p" "mmu-miR-486a-5p"
## [43] "mmu-miR-486b-5p" "mmu-miR-5099" "mmu-miR-532-5p"
## [46] "mmu-miR-540-3p" "mmu-miR-543-3p" "mmu-miR-582-3p"
## [49] "mmu-miR-592-5p" "mmu-miR-666-5p" "mmu-miR-668-3p"
## [52] "mmu-miR-6944-3p" "mmu-miR-7080-3p" "mmu-miR-7080-5p"
## [55] "mmu-miR-7220-5p" "mmu-miR-7240-5p" "mmu-miR-764-3p"
## [58] "mmu-miR-764-5p" "mmu-miR-7653-3p" "mmu-miR-770-5p"
## [61] "mmu-miR-7b-5p" "mmu-miR-877-5p" "mmu-miR-92a-3p"

```

En ambos experimentos aparecen los siguientes miRNAs DE comunes, según los distintos análisis realizados:

Tabla 1. miRNAs DE comunes en los 2 experimentos por los distintos análisis realizados.

Exp1 NOISEq	Exp1 DESeq2/isomiRs	Exp1 RankProduct	Exp2 isomiRs	Exp2 RankProduct
miR-7a-5p				miR-7b-5p
miR-187-3p			miR-187-3p	miR-187-3p
miR-344-3p			miR-344c-3p	miR-344c-3p
		miR-100-5p		miR-100-5p
		miR-12191-3p		miR-12191-3p
		miR-125b-1-3p		miR-125b-2-3p
miR-126a-5p		miR-126b-5p	miR-126a-5p	miR-126a-5p
miR-129-2-3p	miR-129-2-3p	miR-129-2-3p		miR-129-1-3p
		miR-1298-3p		miR-1298-5p
	miR-148a-5p	miR-148a-5p		miR-148a-3p
		miR-181c-5p		miR-181c-3p
		miR-1843b-3p		miR-1843b-3p
		miR-1912-3p		miR-1912-3p
	miR-200a-5p	miR-200a-5p	miR-200a-3p	miR-200a-3p
		miR-200b-5p	miR-200b-3p	miR-200b-3p
		miR-212-5p		miR-212-5p
		miR-30(b/e)-5p		miR-30a-3p
		miR-3102-3p		miR-3102-3p
		miR-33-5p		miR-33-3p
	miR-34b-3p	miR-34b-3p	miR-34a-5p	miR-34b-3p
		miR-423-3p		miR-423-3p
		miR-431-5p		miR-431-3p
		miR-448-3p	miR-448-3p	miR-448-3p
		miR-455-3p		miR-455-3p
miR-770-5p		miR-770-3p		miR-770-5p
		miR-871-3p		miR-871-3p
		miR-92a-3p		miR-92a-3p

2.7 Proceso de cruce de información de miRNAs/isomiRs DE en ambos experimentos.

Se elabora un proceso de cruce de las tablas de ambos experimentos. En primer lugar, a las tablas generadas con *RankProd* en los 2 experimentos se les añaden las siguientes columnas:

base; miRNA base utilizado para cruzar los miRNAs/isomiRs, se selecciona el número de miR. Por ejemplo al miRNA "mmu-miR-125b-1-3p" le corresponde el miR.Base "miR-125"

min.pfp.base; mínimo valor de pfp entre los miRNAs correspondientes al mismo miRNA base.

min.pval.base; mínimo valor de pval entre los miRNAs correspondientes al mismo miRNA base.

Estos 2 últimos campos servirán para filtrar los miRNAs/isomiRs por distinto grado de significación estadística. Se recuerda que para generar las tablas de miRNAs/isomiRs DE en *RankProd* se utilizó el umbral $pval < 0,05$, por lo que filtrar posteriormente por un valor de min.pval.base igual o superior a 0,05 no tendría sentido.

Sobre las tablas de miRNAs/isomiRs resultantes se realiza el proceso de cruce de ambas informaciones, buscando miRNAs/isomiRs DE en ambos experimentos que tengan un miRNA base común.

El proceso se detalla en el ANEXO 2.

Se hacen 4 cruces de información entre los 2 experimentos:

- Tabla 1 del experimento 1 con tabla 1 del experimento 2.
- Tabla 2 del experimento 1 con tabla 2 del experimento 2.
- Tabla 1 del experimento 1 con tabla 2 del experimento 2.
- Tabla 2 del experimento 1 con tabla 1 del experimento 2.

Las tablas resultantes completas se muestran en el ANEXO 3.

Para seleccionar los miRNAs/isomiRs DE comunes más significativos, se filtran sólo aquellos que tengan un pvalue mínimo de miR.base menor que 0.025. Los resultados de los 4 cruces son los siguientes:

Tabla 1 del experimento 1 con tabla 1 del experimento 2. Son miRNAs/isomiRs down-reg en la clase mutante en ambos experimentos.

miRDE11.025

##	miR.base	miR	FC.exp1	pfp.exp1	pval.exp1	FC.exp2	
##	mmu-miR-12191-3p	miR-12191	mmu-miR-12191-3p	0.3148	0.6412	0.021210	0.7733
##	mmu-miR-200a-5p	miR-200	mmu-miR-200a-5p	0.2106	0.4928	0.004208	NA
##	mmu-miR-200b-5p	miR-200	mmu-miR-200b-5p	0.3449	0.8049	0.029210	NA
##	mmu-miR-200b-3p	miR-200	mmu-miR-200b-3p	NA	NA	NA	0.3072
##	mmu-miR-200a-3p	miR-200	mmu-miR-200a-3p	NA	NA	NA	0.4043
##	mmu-miR-34c-3p	miR-34	mmu-miR-34c-3p	0.2585	0.6310	0.012790	NA
##	mmu-miR-34b-3p	miR-34	mmu-miR-34b-3p	0.4838	0.8359	0.047280	NA
##	mmu-miR-34a-5p	miR-34	mmu-miR-34a-5p	NA	NA	NA	0.7396
##	mmu-miR-871-3p	miR-871	mmu-miR-871-3p	0.2700	0.6374	0.012250	0.7447
##		pfp.exp2	pval.exp2				
##	mmu-miR-12191-3p	6.732e-01	1.236e-02				
##	mmu-miR-200a-5p	NA	NA				
##	mmu-miR-200b-5p	NA	NA				
##	mmu-miR-200b-3p	1.447e-07	1.476e-10				
##	mmu-miR-200a-3p	3.279e-06	6.691e-09				
##	mmu-miR-34c-3p	NA	NA				

```
## mmu-miR-34b-3p      NA      NA
## mmu-miR-34a-5p    7.086e-01 1.808e-02
## mmu-miR-871-3p    6.135e-01 6.887e-03
```

Tabla 2 del experimento 1 con tabla 2 del experimento 2. Son miRNAs/isomiRs up-reg en la clase mutante en ambos experimentos.

miRDE22.025

```
##          miR.base          miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-148a-5p miR-148 mmu-miR-148a-5p 5.897 0.3704 0.005929 NA
## mmu-miR-148a-3p miR-148 mmu-miR-148a-3p NA NA NA 1.0600
## mmu-miR-423-3p  miR-423 mmu-miR-423-3p 4.223 0.5900 0.018890 1.1640
## mmu-miR-431-5p  miR-431 mmu-miR-431-5p 5.043 0.4181 0.008032 NA
## mmu-miR-431-3p  miR-431 mmu-miR-431-3p NA NA NA 1.0940
## mmu-miR-92a-1-5p miR-92 mmu-miR-92a-1-5p 5.024 0.4517 0.009159 NA
## mmu-miR-92a-3p  miR-92 mmu-miR-92a-3p 3.426 0.7526 0.040160 0.9537
##          pfp.exp2 pval.exp2
## mmu-miR-148a-5p NA NA
## mmu-miR-148a-3p 0.2025 0.00248
## mmu-miR-423-3p 0.5676 0.02027
## mmu-miR-431-5p NA NA
## mmu-miR-431-3p 0.5437 0.01664
## mmu-miR-92a-1-5p NA NA
## mmu-miR-92a-3p 0.5762 0.02116
```

Tabla 1 del experimento 1 con tabla 2 del experimento 2

miRDE12.025

```
##          miR.base          miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-1298-3p miR-1298 mmu-miR-1298-3p 0.1608 0.7659 0.001635 1.783
## mmu-miR-1298-5p miR-1298 mmu-miR-1298-5p 0.1829 0.6294 0.002687 2.094
## mmu-miR-212-5p  miR-212 mmu-miR-212-5p 0.2379 0.5850 0.006867 1.337
## mmu-miR-212-3p  miR-212 mmu-miR-212-3p NA NA NA 1.278
## mmu-miR-34b-3p  miR-34 mmu-miR-34b-3p 0.4838 0.8359 0.047280 1.363
## mmu-miR-34b-5p  miR-34 mmu-miR-34b-5p NA NA NA 1.309
## mmu-miR-34c-5p  miR-34 mmu-miR-34c-5p NA NA NA 1.261
## mmu-miR-34c-3p  miR-34 mmu-miR-34c-3p 0.2585 0.6310 0.012790 NA
##          pfp.exp2 pval.exp2
## mmu-miR-1298-3p 0.0005107 2.084e-06
## mmu-miR-1298-5p 0.0005471 2.791e-06
## mmu-miR-212-5p 0.4697000 1.342e-02
## mmu-miR-212-3p 0.5906000 2.411e-02
## mmu-miR-34b-3p 0.2970000 5.455e-03
## mmu-miR-34b-5p 0.3765000 9.220e-03
## mmu-miR-34c-5p 0.6733000 3.229e-02
## mmu-miR-34c-3p NA NA
```

Tabla 2 del experimento 1 con tabla 1 del experimento 2

miRDE21.025

```
##          miR.base          miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-92a-1-5p miR-92 mmu-miR-92a-1-5p 5.024 0.4517 0.009159 NA
## mmu-miR-92a-3p  miR-92 mmu-miR-92a-3p 3.426 0.7526 0.040160 0.9537
## mmu-miR-92b-5p  miR-92 mmu-miR-92b-5p NA NA NA 0.8468
##          pfp.exp2 pval.exp2
## mmu-miR-92a-1-5p NA NA
## mmu-miR-92a-3p 0.7487 0.02216
## mmu-miR-92b-5p 0.8366 0.03756
```

2.8 Análisis detallado de isomiRs DE.

2.8.1 Experimento 1 (ratones mutantes *mecp2*+/-)

Se hace un análisis de isomiRs DE en el experimento 1, con todas las variantes de isomiRs indicadas en la función *isoDE*.

```
dds1ri_iso<-isoDE(ids1r,formula=~condition,ref=TRUE,iso5=TRUE,iso3=TRUE,
add=TRUE,snv=TRUE,seed=TRUE)
## converting counts to integer mode
rdds1ri_iso<-results(dds1ri_iso,format="DataFrame")
rdds1riDE_iso<-subset(rdds1ri_iso,rdds1ri_iso$padj<0.1)
rownames(rdds1riDE_iso)
## [1] "mmu-miR-1247-5p;iso_3p:a"
## [2] "mmu-miR-1249-3p;ref"
## [3] "mmu-miR-1981-3p;iso_3p:c"
## [4] "mmu-miR-30f;iso_snp:5CT;iso_5p:GCT;iso_3p:ctc"
## [5] "mmu-miR-335-3p;ref"
## [6] "mmu-miR-376c-3p;iso_snp:6GA"
## [7] "mmu-miR-7090-3p;iso_add:TT"
#aplicamos isoAnnotate a Los isomiRs DE
fi1<-vector()
for (i in 1:nrow(rdds1riDE_iso)){
fi1<-c(fi1,unlist(strsplit(rownames(rdds1riDE_iso)[i],";"))[1])
}
fn<-vector()
for (i in 1:length(fi1)){
fn<-c(fn,grep(fi1[i],isoAnnotate(ids1r)$uid))
}
iso1DEa<-isoAnnotate(ids1r)[fn,]
#isomiRs más representados en clase mutante de cada miRNA
seqs<-data.frame()
for (i in 1:length(isom1)){
f<-subset(iso1DEa[,c("seq","uid")],iso1DEa$uid==isom1[i])
seqs<-rbind(seqs,f)
}
seqs
##          seq          uid
## 150  ACCCGTCCCGTTCGTCCCGG  mmu-miR-1247-5p;iso_3p:a
## 162  ACGCCCTTCCCCCTTCTCA  mmu-miR-1249-3p;ref
## 536  CATCTAACCTGGCCTTGA  mmu-miR-1981-3p;iso_3p:c
## 792  GCTGCAAACATCCGACTGAAAG mmu-miR-30f;iso_snp:5CT;iso_5p:GCT;iso_3p:ctc
## 1580 TTTTCATTATTGCTCCTGACC  mmu-miR-335-3p;ref
## 41   AACATGGAGGAAATTCACGT  mmu-miR-376c-3p;iso_snp:6GA
## 1559 TTTGATGTCTGACTTGCAGTT  mmu-miR-7090-3p;iso_add:TT
```

Se han seleccionado 7 isomiRs DE, de los que se muestra el detalle de anotación de la secuencia y el identificador.

Se consultan estas secuencias en la plataforma *Isomir Bank* (<https://mcg.ustc.edu.cn/bsc/isomir/>) [13], una base de datos de isomiRs para varias especies y tejidos. Filtrando los tejidos por: *brain*, *hippocampus* y *cortex*, se obtienen evidencias experimentales de que estos isomiRs se manifiestan en genotipos *Dgcr8* (heterocigoto y ko) y *Dicer* (heterocigoto y ko).

2.8.2 Experimento 2 (ratones mutantes foxg1+/-)

Se hace un análisis de isomiRs DE en el experimento 2, con todas las variantes de isomiRs indicadas en la función *isoDE*.

Lista de isomiRs DE

```
## [1] "mmu-miR-10a-5p;iso_5p:t" "mmu-miR-10a-5p;iso_add:A;iso_5p:t"
## [3] "mmu-miR-125a-5p;iso_add:A" "mmu-miR-1264-3p;iso_3p:t"
## [5] "mmu-miR-1264-3p;ref" "mmu-miR-129-2-3p;ref"
## [7] "mmu-miR-1298-5p;iso_add:A" "mmu-miR-1298-5p;ref"
## [9] "mmu-miR-181c-3p;iso_add:A;iso_3p:c" "mmu-miR-181d-5p;iso_add:A;iso_3p:t"
## [11] "mmu-miR-181d-5p;ref" "mmu-miR-187-3p;iso_3p:g"
## [13] "mmu-miR-187-3p;iso_add:A" "mmu-miR-187-3p;iso_add:T"
## [15] "mmu-miR-187-3p;iso_add:T;iso_3p:g" "mmu-miR-187-3p;ref"
## [17] "mmu-miR-195a-5p;iso_3p:c" "mmu-miR-195a-5p;ref"
## [19] "mmu-miR-200b-3p;iso_3p:C" "mmu-miR-34a-5p;iso_3p:T"
## [21] "mmu-miR-3552;ref" "mmu-miR-429-3p;ref"
## [23] "mmu-miR-448-3p;iso_3p:t" "mmu-miR-448-3p;iso_add:T;iso_3p:at"
## [25] "mmu-miR-497a-5p;iso_3p:a" "mmu-miR-551b-3p;iso_3p:T"
## [27] "mmu-miR-652-3p;iso_add:T" "mmu-miR-6944-3p;iso_add:T"
```

Hay 28 isomiRs DE en el experimento 2, de los que se ve en detalle la anotación de *seq* y *uid*.

##	seq	uid
## 1680	ACCCTGTAGATCCGAATTTGTG	mmu-miR-10a-5p;iso_5p:t
## 1681	ACCCTGTAGATCCGAATTTGTGA	mmu-miR-10a-5p;iso_add:A;iso_5p:t
## 10930	TCCCTGAGACCCCTTTAACCTGTGAA	mmu-miR-125a-5p;iso_add:A
## 4339	CAAATCTTATTTGAGCACCTG	mmu-miR-1264-3p;iso_3p:t
## 4341	CAAATCTTATTTGAGCACCTGT	mmu-miR-1264-3p;ref
## 710	AAGCCCTTACCCCAAAAAGCAT	mmu-miR-129-2-3p;ref
## 14011	TTCATTCGGCTGTCCAGATGTAA	mmu-miR-1298-5p;iso_add:A
## 14010	TTCATTCGGCTGTCCAGATGTA	mmu-miR-1298-5p;ref
## 1590	ACCATCGACCGTTGAGTGGACA	mmu-miR-181c-3p;iso_add:A;iso_3p:c
## 457	AACATTTCATTGTTGTCGGTGGGA	mmu-miR-181d-5p;iso_add:A;iso_3p:t
## 463	AACATTTCATTGTTGTCGGTGGGT	mmu-miR-181d-5p;ref
## 11424	TCGTGTCTTGTGTTGCAGCCG	mmu-miR-187-3p;iso_3p:g
## 11430	TCGTGTCTTGTGTTGCAGCCGGA	mmu-miR-187-3p;iso_add:A
## 11437	TCGTGTCTTGTGTTGCAGCCGGT	mmu-miR-187-3p;iso_add:T
## 11444	TCGTGTCTTGTGTTGCAGCCGT	mmu-miR-187-3p;iso_add:T;iso_3p:g
## 11429	TCGTGTCTTGTGTTGCAGCCGG	mmu-miR-187-3p;ref
## 9615	TAGCAGCACAGAAATATTGG	mmu-miR-195a-5p;iso_3p:c
## 9620	TAGCAGCACAGAAATATTGGC	mmu-miR-195a-5p;ref
## 9251	TAATACTGCCTGGTAATGATGAC	mmu-miR-200b-3p;iso_3p:C
## 13011	TGGCAGTGTCTTAGCTGGTTGTT	mmu-miR-34a-5p;iso_3p:T
## 2941	AGGCTGCAGGCCCACTTCCT	mmu-miR-3552;ref
## 9260	TAATACTGTCTGGTAATGCCGT	mmu-miR-429-3p;ref
## 14233	TTGCATATGTAGGATGTCCCA	mmu-miR-448-3p;iso_3p:t
## 14244	TTGCATATGTAGGATGTCCCT	mmu-miR-448-3p;iso_add:T;iso_3p:at
## 4893	CAGCAGCACACTGTGGTTTGT	mmu-miR-497a-5p;iso_3p:a
## 8072	GCGACCCATACTTGGTTTCAGT	mmu-miR-551b-3p;iso_3p:T
## 1189	AATGGCGCCACTAGGTTGTGT	mmu-miR-652-3p;iso_add:T
## 9121	TAACCTTCCCTTGTGCCCTCAGT	mmu-miR-6944-3p;iso_add:T

Su consulta en *Isomir Bank* ofrece unos resultados similares a los del experimento 1: se manifiestan en genotipos Dgcr8 (heterocigoto y ko) y Dicer (heterocigoto y ko).

2.9 Clasificación supervisada.

PLS-DA (Partial Least Squares Discriminant Analysis) es una técnica apropiada para análisis de conjuntos de datos de alta dimensionalidad y multicolinealidad. Es un método supervisado con la intención de conseguir una reducción de dimensiones en una situación donde se quiere relacionar una variable respuesta binaria (en este caso genotipo mutante o normal y con ello, manifestación del síndrome o no) con un conjunto de variables predictoras (expresión de isomiRs). El procedimiento de reducción de dimensionalidad se basa en la transformación ortogonal de las variables originales (isomiRs) en un conjunto de variables latentes no correladas linealmente (llamadas componentes) tales que maximizan la separación entre las diferentes clases en los primeros componentes (en este caso se usarán los 2 primeros).

2.9.1 Experimento 1 (ratones *mecp2* +/-)

Se implementa este método usando *Discriminer* dentro de la función *isoPLSDA*.

```
pls.ids1<-isoPLSDA(ids1r,"condition",nperm=2)
isoPLSDAplot(pls.ids1)
```

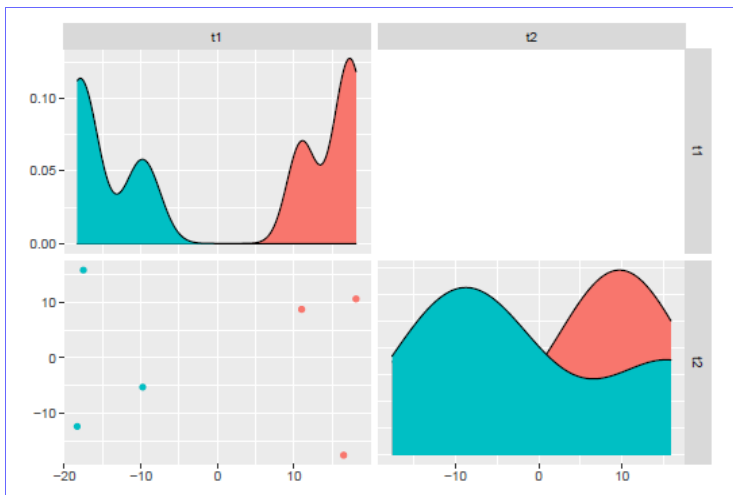


Figura 22. Representación PLS-DA de las muestras del experimento 1 (ratones *mecp2* +/-)

Este método utiliza la suma de cuadrados capturada por el modelo como una medida de bondad de ajuste. El p-valor nos dice la significación estadística de la separación de grupos usando la expresión de miRNAs; en este modelo el p-valor=0,6275.

En el gráfico se representa la distribución de muestras a lo largo de las 2 primeras componentes (t1 y t2). El color rojo corresponde a las muestras mutantes y el azul a las de control.

2.9.2 Experimento 2 (ratones *foxg1* +/-)

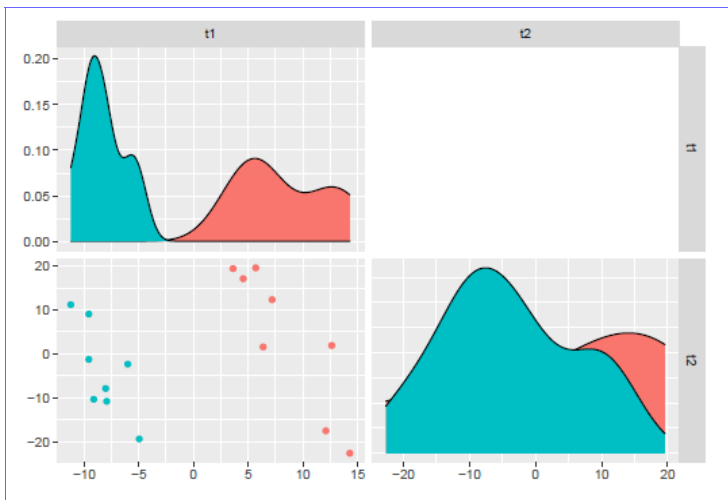


Figura 23. Representación PLS-DA de las muestras del experimento 2 (ratones *foxg1* +/-)

Para el modelo PLS-DA del experimento 2 el p-valor = 0,5. La interpretación del gráfico es la misma que para el experimento 1

2.10 Integración gen-miRNA

2.10.1 Herramientas de predicción de genes target.

Para la descripción y selección de las herramientas de predicción de *targets* se ha tomado como referencia el artículo [14]. Hay 4 características comunes para las herramientas de predicción de *targets* (genes diana) de miRNAs:

- Coincidencia en la región germen (*seed match*).

La secuencia germen (*seed*) de un miRNA se define como los primeros 2-8 nucleótidos, empezando desde el extremo 5' y continuando hacia el extremo 3'.

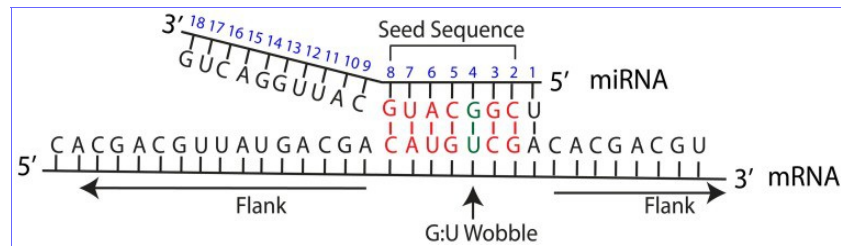


Figura 24. Interacción entre el miRNA y el mRNA target. Se representa la secuencia seed con los emparejamientos de nucleótidos WC (Watson Crick) marcados en rojo y un ejemplo de G-U wobble marcado en verde. Los flancos se refieren a las secuencias del mRNA a cada lado de la región target que hibrida con el miRNA [14].

Los principales tipos de *seed match*, por orden de menor a mayor precisión, son:

- 6mer. Coincidencia perfecta de 6 nucleótidos del *seed* miRNA con el mRNA.
- 7mer-m8. Coincidencia perfecta de los nucleótidos 2-8 del *seed* miRNA.
- 7mer-A1. Coincidencia perfecta de los nucleótidos 2-7 del *seed* miRNA, con adición de una A en el nucleótido 1.
- 8mer. Coincidencia perfecta de los nucleótidos 2-8 del *seed* miRNA, con adición de una A en el nucleótido 1.

- Conservación.

La conservación se refiere al mantenimiento de una secuencia a lo largo de la especiación. El análisis de conservación se puede centrar en las regiones 3' UTR o 5'UTR del mRNA, el miRNA, o una combinación de las 3. En general hay mayor conservación en la región *seed* del miRNA.

- Energía libre (free energy).

La energía libre de Gibbs se puede usar como medida de estabilidad de un sistema biológico. Si la unión del miRNA con un mRNA target candidato se predice como estable, se considerará más probable que sea un verdadero *target*. Una región con alta energía libre (ΔG positivo) es más inestable que una con baja energía libre (ΔG negativo).

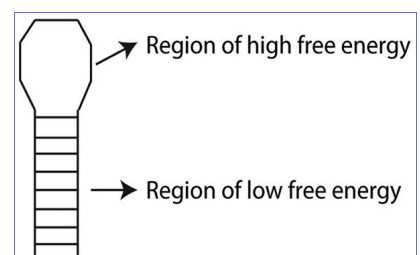


Figura 25. Estructura secundaria de RNA en forma de bucle (hairpin loop) como ejemplo de zonas de alta energía libre (menos estable) y baja energía libre (más estable) [14].

- Accesibilidad del sitio (*site accessibility*).

La accesibilidad del sitio es una medida de la facilidad con que un miRNA puede localizar e hibridar un mRNA *target*. La hibridación miRNA:mRNA se realiza en dos pasos; primero el miRNA se una a una región corta accesible del mRNA, posteriormente la estructura secundaria del mRNA se despliega para que el miRNA complete su unión.

Se puede evaluar la cantidad de energía necesaria para hacer un sitio accesible al miRNA como medida de probabilidad de que un mRNA sea *target* del miRNA.

Otras características menos comunes que tienen en cuenta las herramientas de predicción de *targets* son:

- Abundancia de sitios *target*. Medida de cuántos sitios *target* se localizan en la región 3' UTR.
- Contenido local de AU. Concentración de nucleótidos AU que flanquean la región *seed* del miRNA.
- Presencia de GU *wobble* en el *seed match*. Posibilidad de que G se empareje con U.
- Emparejamiento 3' compensatorio. Es una medida de conservación. En una pequeña proporción de interacciones miRNA:mRNA, hay pares conservados en el extremo 3' del miRNA que pueden compensar *seed mismatches*.
- Contribución de la posición. Analiza la posición del sitio *target* dentro del mRNA.
- SVM (*Support Vector Machine*). Es una medida de predicción que utiliza técnicas de Machine Learning.

2.10.2 Predicción de genes *target* en el experimento 1 (ratones *mecp2+/-*).

Se emplearán distintas herramientas para la predicción de *targets*.

TargetScan.

Herramienta disponible en <http://www.targetscan.org/> [15]. La salida de esta herramienta proporciona una clasificación de *targets* predichos por la eficacia de *targeting* (*context + scores*) o por la probabilidad de *target* conservado (PCT, *probabilities of conserved targeting*).

La conservación de 3' UTR se determina siguiendo un análisis de un k-mer específico (8mer, 7mer-m8 o 7mer-1A) y se combinan las consideraciones de conservación por tipo de k-mer.

La combinación *context + scores* muestra la probabilidad de que un *target* predicho sea efectivamente un *target*. La puntuación se deriva de resultados experimentales. En el *score* se consideran características adicionales como el emparejamiento 3' compensatorio, contenido local de AU y la contribución de la posición.

Aquí se utilizará la función *mirna2targetscan* del paquete *targetscan.Hs.eg.db* [16] de Bioconductor, que da la clasificación por PCT. Para la obtención de genes *target* de los miRNAs DE se pone un umbral de significación en las tablas de RP de $pfp < 0.5$

```
#Obtención de miRNA-target con mirna2targetscan
#Library(AnnotationDbi)
#Library(targetscan.Mm.eg.db)
#miRs DE con pfp<0.5
mirDEe1_05<-subset(rownames(t1e1DE),t1e1DE$pfp<0.5)
```

```

mirDEe1_05<-c(mirDEe1_05,subset(rownames(t2e1DE),t2e1DE$pfpc<0.5))
mirDEe1_05<-unique(mirDEe1_05)
mirna_ma1<-mirna2targetscan(mirDEe1_05, species = "mmu")
## Missing miRNAs: mmu-miR-129b-5pmmu-miR-7090-3p
#Descargamos fichero Mus_musculus.gene.info de ftp.ncbi.nlm.nih.gov
#anotacion
anno<-read.delim(file=~ /TFM/R/Mus_musculus.gene_info.gz",header=FALSE,skip=1)
genes<-mirna_ma1$gene
nf<-match(genes,anno$V2)
mirna_ma1a<-cbind(mirna_ma1,anno[nf,3],
  substr(anno[nf,6],nchar(as.vector(anno[nf,6])))-
  17,nchar(as.vector(anno[nf,6]))),
  anno[nf,9])
colnames(mirna_ma1a)<-c(colnames(mirna_ma1),c("gene.name","Ensembl","description"))

#Length(unique(mirna_ma1a$gene))
#Length(unique(subset(mirna_ma1a$gene,mirna_ma1a$PCT>0.5 &
  mirna_ma1a$Seedmatch=="8mer")))
#se escriben los genes target seleccionados con PCT>0.5 y Seedmatch=8mer
genes1ts<-unique(subset(mirna_ma1a[,c("gene","gene.name","Ensembl")],
  mirna_ma1a$PCT>0.5 & mirna_ma1a$Seedmatch=="8mer"))

```

Como umbral de filtrado de targets a la salida de TargetScan se utiliza el criterio de que el PCT sea mayor que 0,5 y que el *seed match* sea igual a 8mer.

miRanda

La herramienta está disponible en <http://www.microrna.org/> [17]. miRanda utiliza un análisis en tres pasos:

- Las secuencias miRNA se escanean frente a las secuencias 3' UTR de mRNA (ambos ficheros proporcionados por el usuario, en formato FASTA) para chequear coincidencias.
- Se filtran los *targets* candidatos por un umbral de energía libre (establecido por el usuario).
- Se utiliza la medida de conservación como filtro final.

Cada *target* candidato obtiene una puntuación individual que se combina con la abundancia de sitios predichos. Se consideran coincidencias a lo largo de toda la secuencia del miRNA, dando más peso a la región *seed*. Se permiten un número limitado de GU *wobbles*, inserciones y deleciones. También calcula la energía libre.

En la versión miRanda-mirSVR se utiliza miRanda para localizar candidatos y mirSVR para puntuarlos. Incorpora otras características, como accesibilidad del sitio, contenido AU en los flancos, posición del *target* dentro de 3' UTR y longitud de UTR. La puntuación mirSVR proporciona una indicación del efecto regulatorio del miRNA sobre el mRNA.

Para utilizar el programa interactivo, sería necesario disponer de las secuencias mRNA donde se quieran buscar los *targets*. En lugar de esto, se descarga del servidor el fichero de predicciones para ratón con alta conservación y SVR *mouse_predictions_S_C_aug2010.txt*. Se quita el símbolo "#" de la cabecera y se guarda el fichero como *mouse_predictions_S_C.txt* para importarlo a un data.frame.

```

pred_miranda<-read.table("~/TFM/datos/mouse_predictions_S_C.txt",sep="\t",header =
TRUE)
head(pred_miranda)

```

```

##   mirbase_acc mirna_name gene_id gene_symbol transcript_id ext_transcript_id
## 1 MIMAT000121 mmu-let-7g 109019 Obfc2a uc007axn.1 NM_028696
## 2 MIMAT000121 mmu-let-7g 52150 Kcnk6 uc009gbg.1 NM_001033525
## 3 MIMAT000121 mmu-let-7g 18514 Pbx1 uc007dlf.1 L27453
##   mirna_alignment alignment
## 1 uuGACA-UGUU--UG-AU-GAUGGAgU |||| |::| || || |||||
## 2 uugACAUGUUUGAUG-AUGGAGU | |::||| |:| |||||
## 3 uuGACAUGUUUGA-UGAUGGAGU :||| :| || |||||
##   gene_alignment mirna_start mirna_end gene_start gene_end
## 1 ugCUGUAAUGAUGACAUAUCUACCUac 3 21 1276 1302
## 2 ugaUAUGUAAAAUGCAUACCUca 2 20 2612 2634
## 3 ccUUGU-UACUCUCACUACCUcu 2 21 615 636
##   genome_coordinates conservation align_score seed_cat energy
## 1 [mm9:1:51526787-51526813:-] 0.6211 132 0 -16.75
## 2 [mm9:7:30007421-30007443:-] 0.5016 131 6 -15.96
## 3 [mm9:1:170088363-170088384:-] 0.6311 153 7 -17.32
##   mirsvr_score
## 1 -0.1295
## 2 -0.2266
## 3 -0.1566

```

Para la obtención de genes *target* de los miRNAs DE se pone un umbral de significación en las tablas de RP de $pfp < 0.5$. Para cruzarlos con las predicciones *target* de miRanda disponibles, hay que recortar el nombre del miRNA, pues en el fichero (campo *mirna_name*) no se distinguen las variantes -5p y -3p.

```

#recortamos el nombre del miRNA, en miRanda no diferencia variantes -3p-5p
mirDEe1_05t<-vector()
for (i in 1:length(mirDEe1_05)){
mirDEe1_05t[i]<-paste(unlist(strsplit(mirDEe1_05[i],"-"))[1],
                    unlist(strsplit(mirDEe1_05[i],"-"))[2],
                    unlist(strsplit(mirDEe1_05[i],"-"))[3],sep="-")
}
mirDEe1_05t<-unique(mirDEe1_05t)
#obtenemos los genes de predicción de target de la tabla pred_miranda
nf<-vector()
for (i in 1:length(mirDEe1_05t)){
nf<-c(nf,grep(mirDEe1_05t[i],pred_miranda$mirna_name))
}
targets_e1<-pred_miranda[nf,c(2:4,15:19)]
genes1mr<-unique(subset(targets_e1[,c("gene_id","gene_symbol")],
                        targets_e1$mirsvr_score <= -1.2))

```

Como umbral de filtrado de targets obtenidos de miRanda se utiliza el criterio de que el mirSVR sea menor que -1,2 (a menor puntuación se considera mejor candidato).

DIANA MicroT-CDS

La herramienta está disponible en <http://diana.imis.athena-innovation.gr/DianaTools> [18, 19]. DIANA MicroT-CDS incorpora una aproximación Machine Learning para identificar las características más relevantes extraídas de datos PAR-CLIP (*photoactivable-ribonucleoside-enhanced crosslinking and immunoprecipitation*).

Se utilizan secuencias 3' UTR y CDS. Las características más importantes para ambas regiones son el peso de la categoría de unión (estimación de la eficiencia de unión basada en la coincidencia primaria en una región *seed* extendida), distancia al extremo más

cercano de la región (3' UTR o CDS), distancia al sitio de unión adyacente, energía libre predicha para el híbrido, conservación y contenido en AU.

Se realizan dos modelos separados de unión del miRNA al CDS y al 3' UTR, con dos puntuaciones separadas que se combinan en una sola.

Se seleccionan los *targets* de la lista de miRNAS DE anterior directamente en el Web Server. Se importa el fichero resultante.

```
#importamos el fichero
diana1<-read.csv("~/TFM/R/Diana1s.csv",header=TRUE,stringsAsFactors = FALSE)
#formateamos las columnas
diana1b<-data.frame(substr(diana1[,2],1,18),
                    substr(diana1[,2],21,nchar(as.vector(diana1[,2]))-1),
                    diana1$Mirna.Name,diana1$miTG.score)
colnames(diana1b)<-c("Ensembl","gene.symbol","miRNA","miTG.score")
#seleccionamos los targets con miTG.score > 0.98
genes1di<-unique(subset(diana1b[,c("Ensembl","gene.symbol")],diana1b$miTG.score>0.97))
```

Como umbral de filtrado de targets obtenidos de DIANA MicroT-CDS se utiliza el criterio de que la puntuación *miTG* sea mayor que 0,97.

DIANA TarBase

Otra herramienta de predicción de targets disponible en DIANA es TarBase, que a diferencia de las anteriores basa sus resultados en *targets* miRNA validados experimentalmente en lugar de predicciones. La herramienta está disponible en http://carolina.imis.athena-innovation.gr/diana_tools [20].

Se consultan directamente en el Web server los miRNAs DE seleccionados anteriormente, pero los resultados no parecen muy prometedores para el presente estudio. Se contabilizan 31 genes target, implicados en los siguientes pathways:

- *Intestinal immune network for IgA production*
- *Inflammatory bowel disease (IBD)*
- *Maturity onset diabetes of the young*
- *HTLV-I infection*

Comparación de resultados.

Con la herramienta TargetScan se han seleccionado 898 genes *target*; considerando un umbral de selección de PCT > 0.5 y Seedmatch = 8mer.

Con la herramienta miRanda se han seleccionado 589 genes *target*; considerando un umbral de selección de mirsvr_score < -1.2.

Con la herramienta MicroT-CDS se han seleccionado 632 genes *target*; considerando un umbral de selección de miTG.score > 0.98.

Targets que coinciden en las predicciones con las 3 herramientas

Coincidencias	TargetScan	miRanda	MicroT-CDS
TargetScan	-		
miRanda	117	-	
MicroT-CDS	152	66	-

En vista de los resultados, se escogen como candidatos los genes *target* obtenidos con TargetScan para la realización de análisis posteriores.

2.10.3 Predicción de genes target en el experimento 2 (ratones foxg1+/-).

TargetScan

Con la herramienta TargetScan se han seleccionado 835 genes *target*; considerando un umbral de selección de PCT > 0.5 y Seedmatch = 8mer.

DIANA TarBase

En cuanto a *targets* validados experimentalmente se utiliza la herramienta DIANA TarBase y se localizan 40 genes *target*, uno de ellos implicado en el *pathway*: *Lysine degradation*.

2.11 Enriquecimiento GO.

2.11.1 Enriquecimiento GO Experimento 1 (ratones mecp2+/-)

Para hacer un enriquecimiento de categorías GO (Gene Ontology) se emplea la lista de genes target obtenida por la herramienta TargetScan. Para el proceso se utiliza la librería *clusterProfile* de Bioconductor [21].

```
#cogemos los targets de TargetScan
gene1<-as.vector(genes1ts$gene.name)
ego1 <- enrichGO(gene1,org.Mm.eg.db, keyType = "SYMBOL", ont = "BP")
dotplot(ego1, showCategory=30)
emapplot(ego1)
cnetplot(ego1)
```

Se comprueba que entre las 30 categorías GO más representadas en los genes target se encuentran bastantes referentes a la diferenciación neuronal.

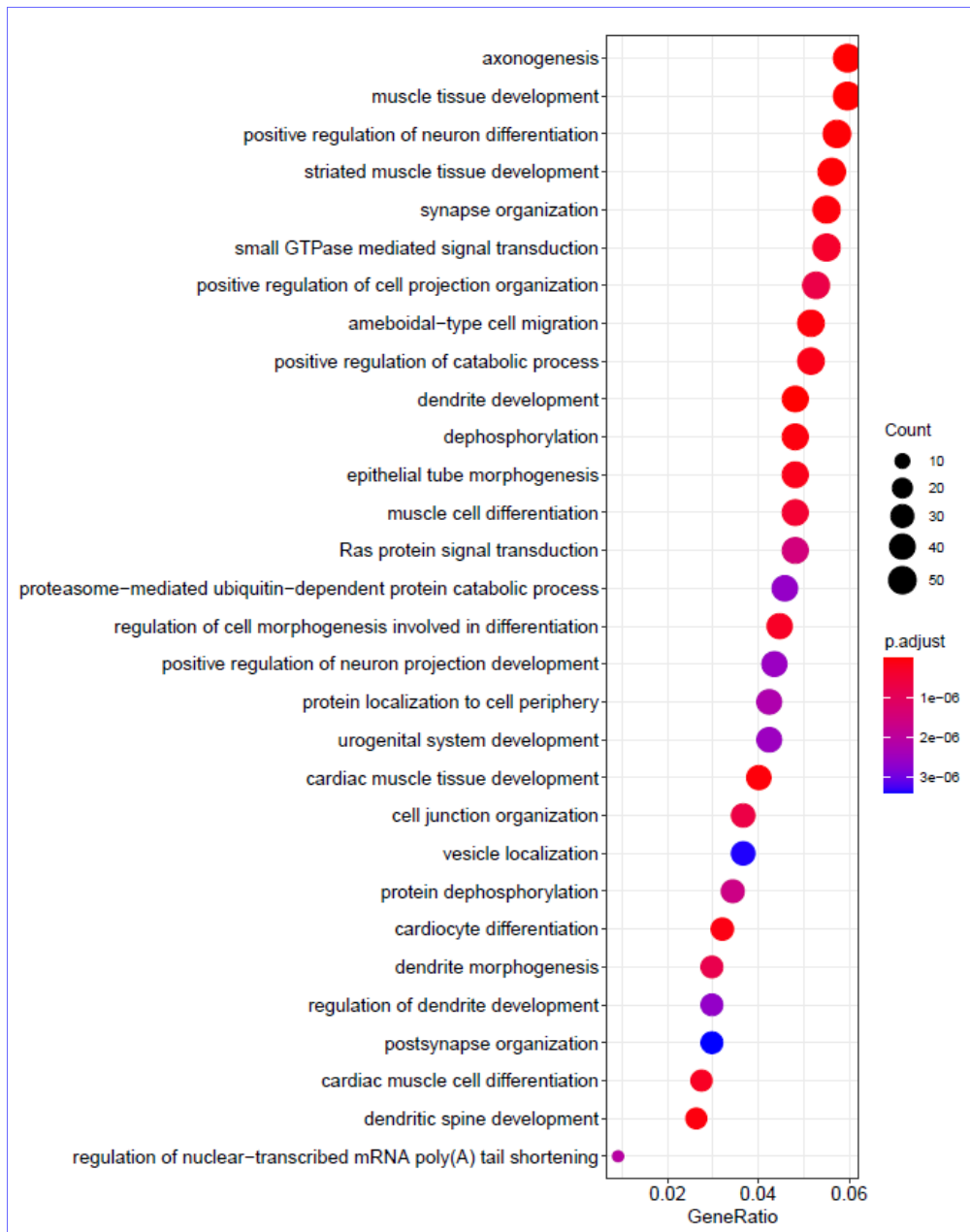


Figura 26. Diagrama de puntos de categorías GO más representativas entre los genes target seleccionados del experimento 1 (ratones mecp2+/-).

En el mapa de enriquecimiento se ve que hay 2 redes importantes de categorías relacionadas, la referente a diferenciación neuronal y otra referente al desarrollo muscular.

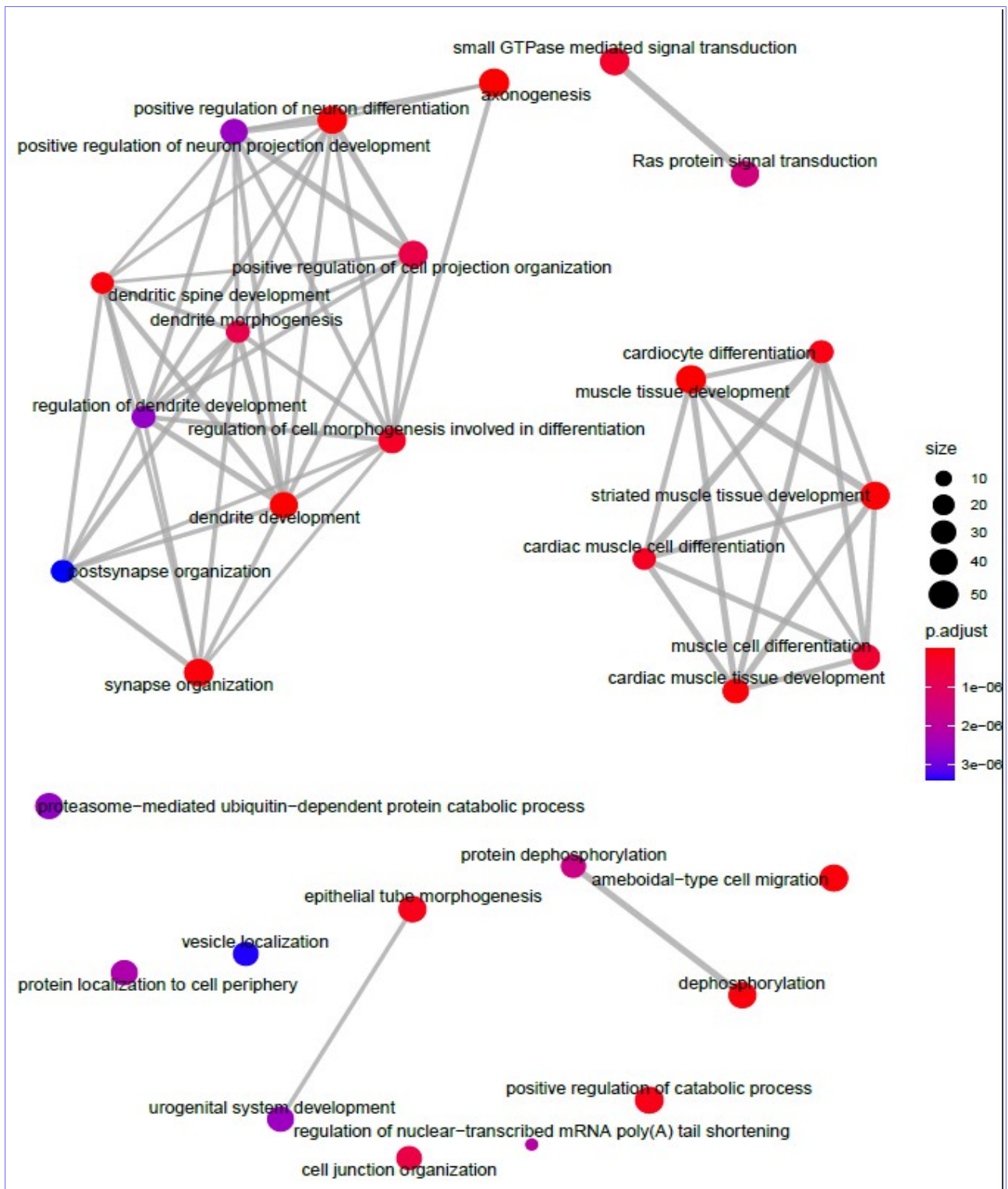


Figura 27. Mapa de enriquecimiento de categorías GO más representativas entre los genes target seleccionados del experimento 1 (ratones *mecp2+/-*)

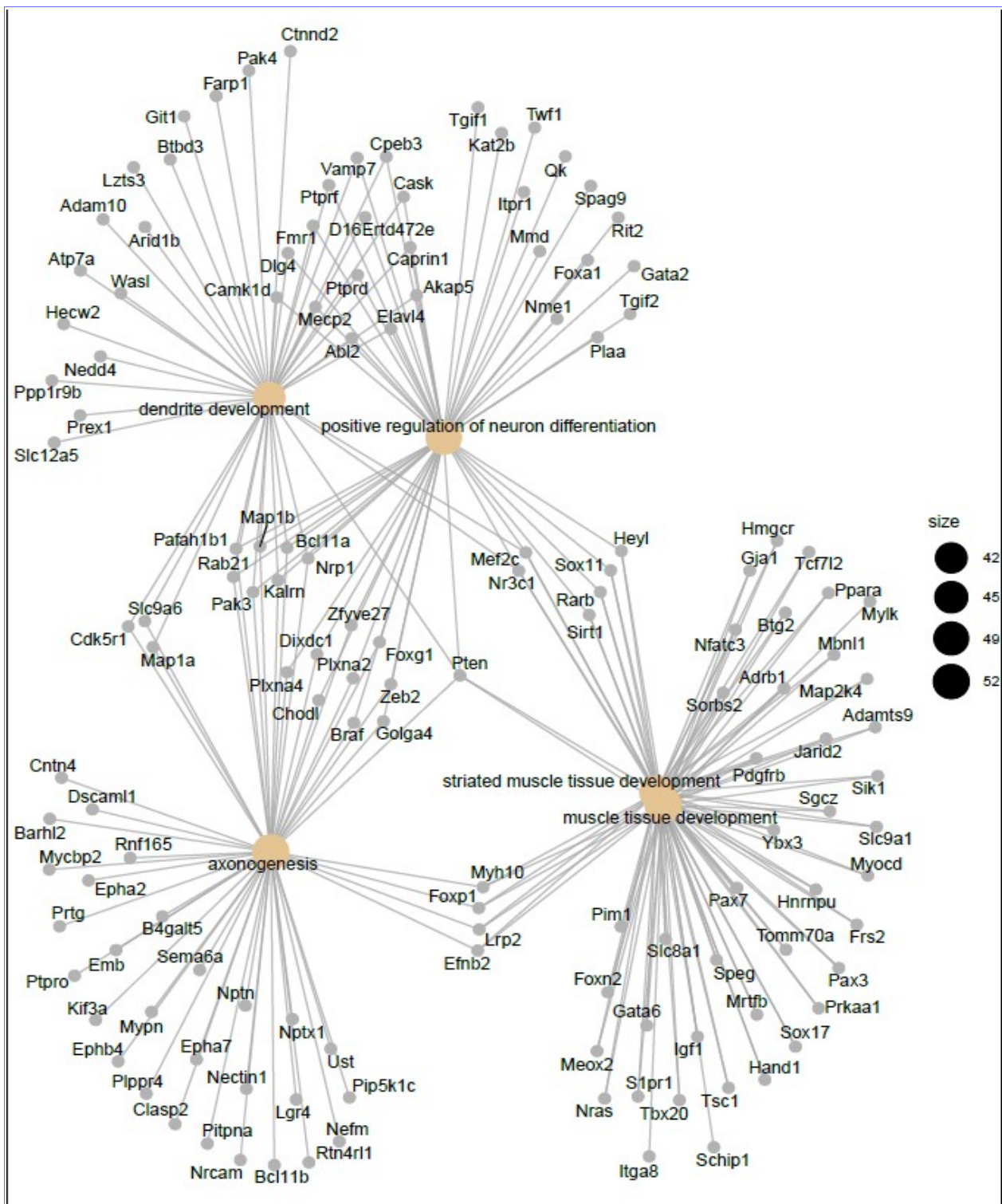


Figura 28. cnetplot de los genes *target* seleccionados en el experimento 1 (ratones *mecp2*^{+/-}), en el que se puede ver la relación entre posibles genes *target* y procesos biológicos.

Se quiere hacer una representación de red bipartita en la que se vea la interrelación entre los 2 tipos de nodos, por un lado los miRNAs DE y por otro los posibles genes *target* de estos miRNAs. Se hace una selección de los genes *target* que según el mapa de enriquecimiento GO tienen relación con categorías implicadas en la diferenciación neuronal. Éstas son:

Categoría	Descripción	Nº genes
GO:0007409	<i>Axonogenesis</i>	52
GO:0045666	<i>Positive regulation of neuron differentiation</i>	52
GO:0010976	<i>Positive regulation of neuron projection development</i>	38
GO:0031346	<i>Positive regulation of cell projection organization</i>	46
GO:0060996	<i>Dendritic spine development</i>	23
GO:0048813	<i>Dendrite morphogenesis</i>	26
GO:0050773	<i>Regulation of dendrite development</i>	26
GO:0010769	<i>Regulation of cell morphogenesis involved in differentiation</i>	39
GO:0016358	<i>Dendrite development</i>	42
GO:0099173	<i>Postsynapse organization</i>	26
GO:0050808	<i>Synapse organization</i>	48

Para la creación de la red bipartita se utilizará el paquete *igraph* [22] y el modelo se representará con *cytoscape* [23].

```
#grafico bipartito
library(igraph)
#seleccionamos genes para las categorias GO que tienen que ver con neurogénesis
got<-c("GO:0007409", "GO:0045666", "GO:0010976", "GO:0031346", "GO:0060996",
       "GO:0048813", "GO:0050773", "GO:0010769", "GO:0016358", "GO:0099173", "GO:0050808")
nf<-match(got, ego1$ID)
genes1s<-vector()
for (i in 1:length(nf)){
genes1s<-c(genes1s, unlist(strsplit(ego1[nf[i], "geneID"], "/")))
}
genes1s<-unique(genes1s)
#seleccionamos los emparejamientos miRNA-gene
pares<-mirna_mala[, c(10,6)]
nf<-vector()
for (i in 1:length(genes1s)){
nf<-c(nf, grep(genes1s[i], pares$gene.name))
}
pares_gos<-pares[nf,]
#hacemos el grafico de red bipartita
#library(igraph)
g <- graph.data.frame(pares_gos, directed = FALSE)
#bipartite.mapping(g)
V(g)$type <- bipartite_mapping(g)$type
#plot(g)
#exportamos el grafico a cytoscape
write_graph(g, file="~/TFM/R/bpn1.graphml", format = "graphml")
```

El mapa de red bipartita resultante se muestra a continuación. Hay 126 genes *target* diferentes implicados en estas funciones biológicas.

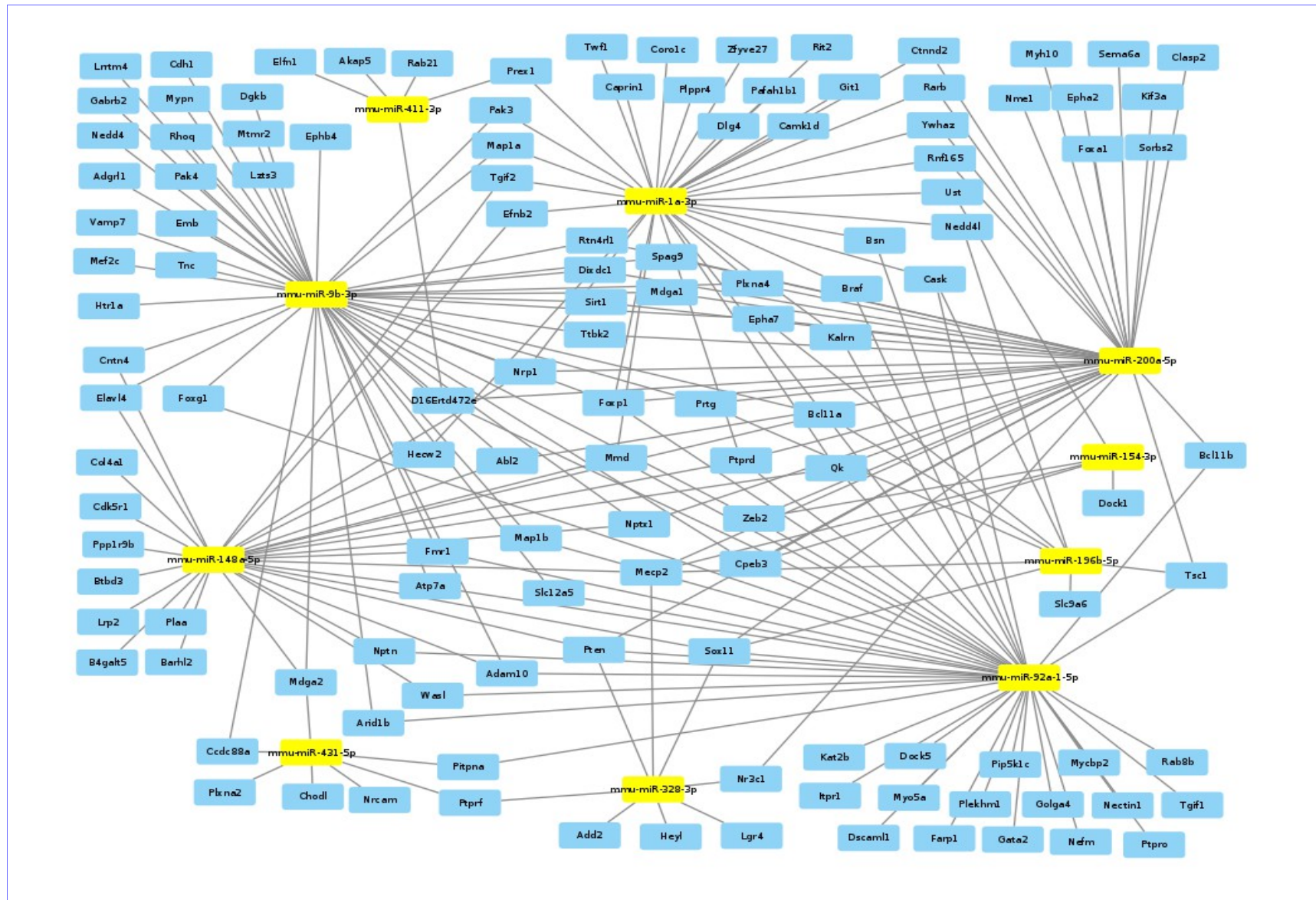


Figura 29. Mapa de red bipartita de miRNAs DE y genes target seleccionados del experimento 1 (ratones mecp2+/-).

2.11.2 Enriquecimiento GO Experimento 2 (ratones *foxg1+/-*)

Se procede de igual manera que en el experimento 1, el mapa resultante del enriquecimiento GO de los genes *target* seleccionados para este experimento es el siguiente:

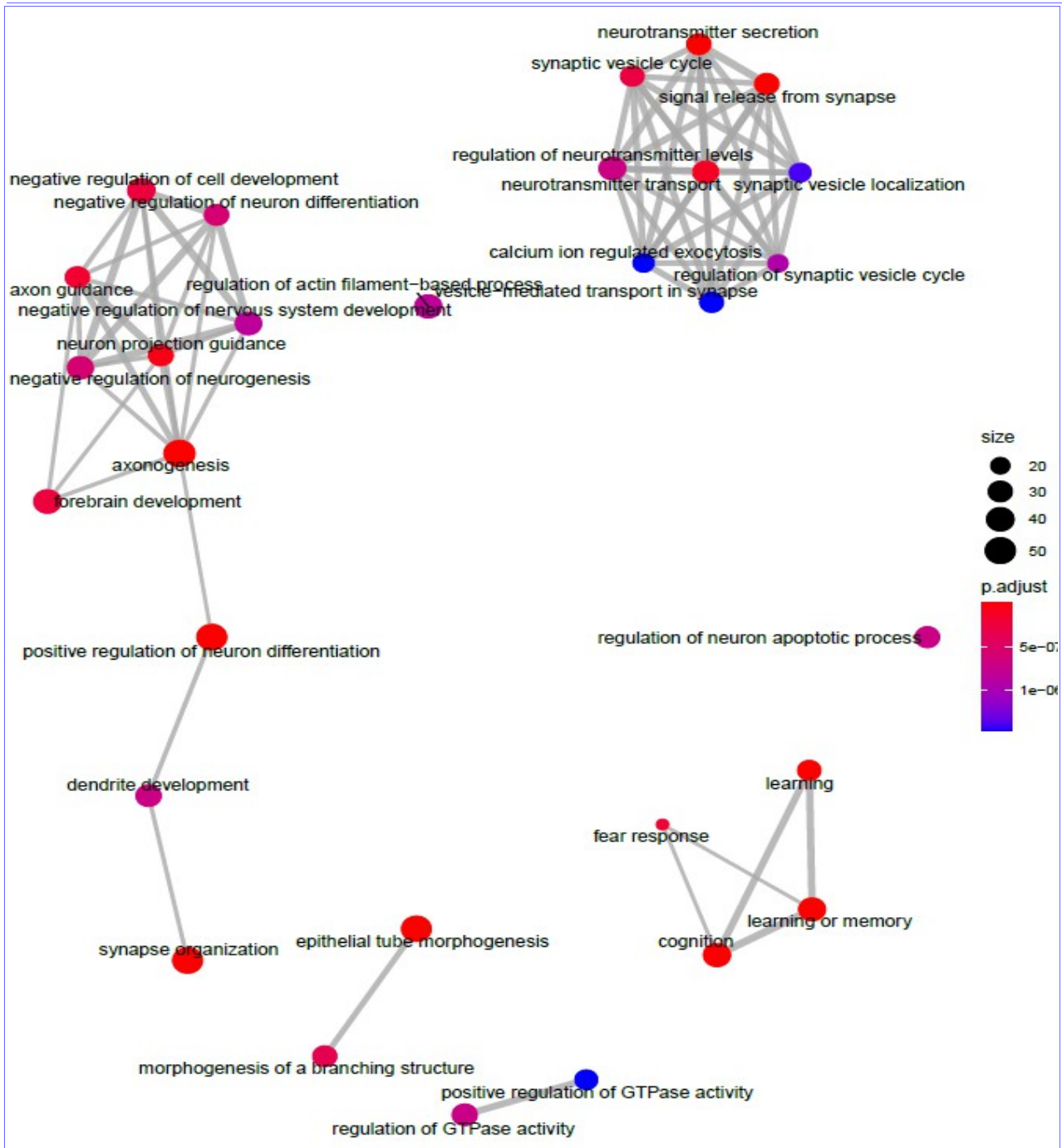


Figura 30. Mapa de enriquecimiento de categorías GO más representativas entre los genes *target* seleccionados del experimento 2 (ratones *foxg1+/-*)

Aquí se pueden apreciar 3 redes de características relacionadas con el sistema nervioso. En la primera de ellas se encuentran las categorías:

Categoría	Descripción	Nº genes
GO:0050808	<i>Synapse organization</i>	48
GO:0016358	<i>Dendrite development</i>	42
GO:0045666	<i>Positive regulation of neuron differentiation</i>	49
GO:0030900	<i>Forebrain development</i>	40
GO:0007409	<i>Axonogenesis</i>	52
GO:0050768	<i>Negative regulation of neurogenesis</i>	36
GO:0097485	<i>Neuron projection guidance</i>	32
GO:0051961	<i>Negative regulation of nervous system development</i>	37
GO:0007411	<i>Axon guidance</i>	31
GO:0045665	<i>Negative regulation of neuron differentiation</i>	31
GO:0010721	<i>Negative regulation of cell development</i>	40

Hay 147 genes *target* diferentes implicados en estas funciones biológicas.

En la segunda se encuentran las categorías:

Categoría	Descripción	Nº genes
GO:0007269	<i>Neurotransmitter secretion</i>	31
GO:0099504	<i>Synaptic vesicle cycle</i>	29
GO:0099643	<i>Signal release from synapse</i>	31
GO:0001505	<i>Regulation of neurotransmitter levels</i>	38
GO:0006836	<i>Neurotransmitter transport</i>	35
GO:0097479	<i>Synaptic vesicle localization</i>	25
GO:0017156	<i>Calcium ion regulated exocytosis</i>	24
GO:0098693	<i>Regulation of synaptic vesicle cycle</i>	22
GO:0099003	<i>Vesicle-mediated transport in synapse</i>	29

Hay 51 genes *target* diferentes implicados en estas funciones biológicas.

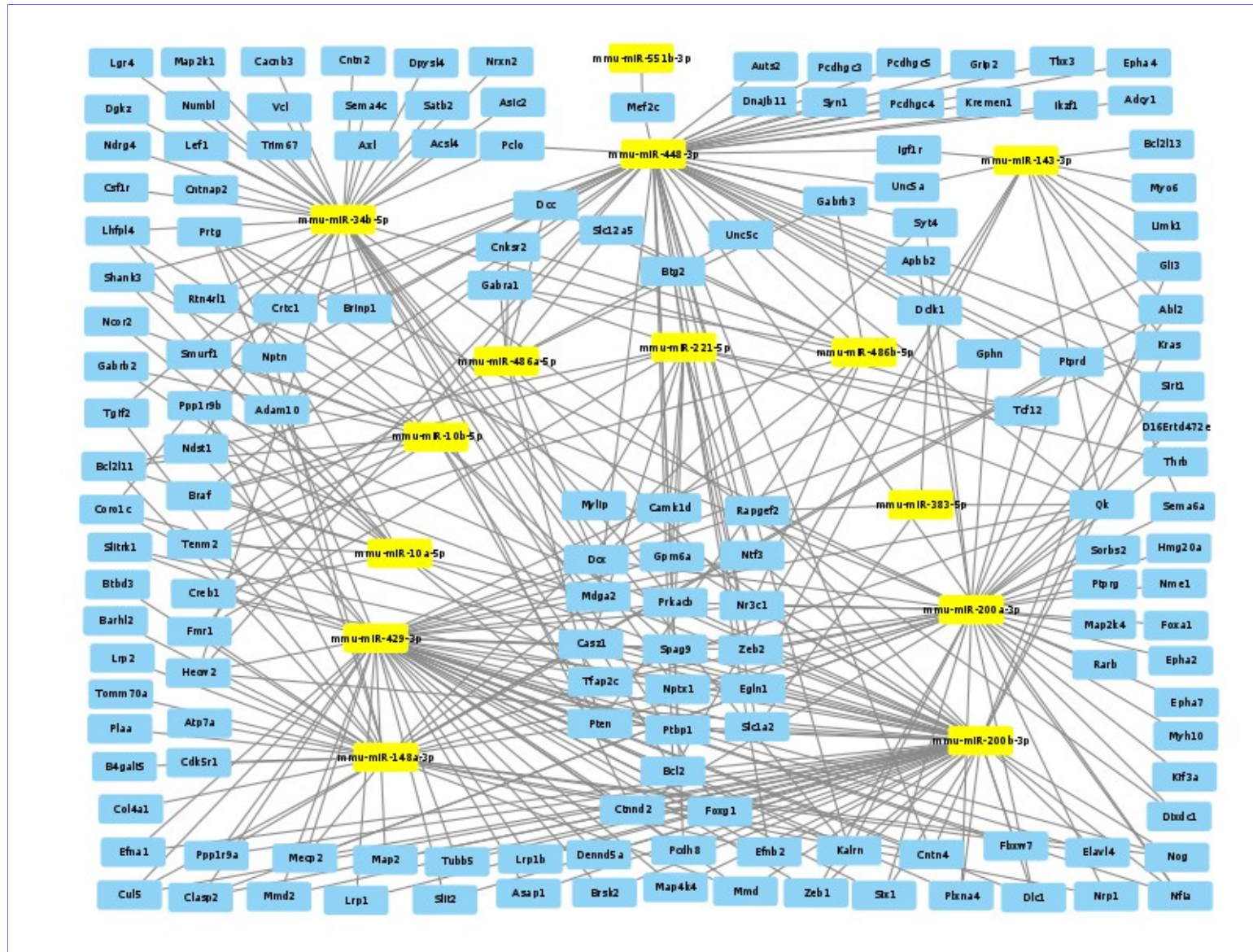


Figura 31. Mapa de red bipartita de miRNAs DE y genes target seleccionados en el grupo 1 del experimento 2 (ratones foxg1+/-).

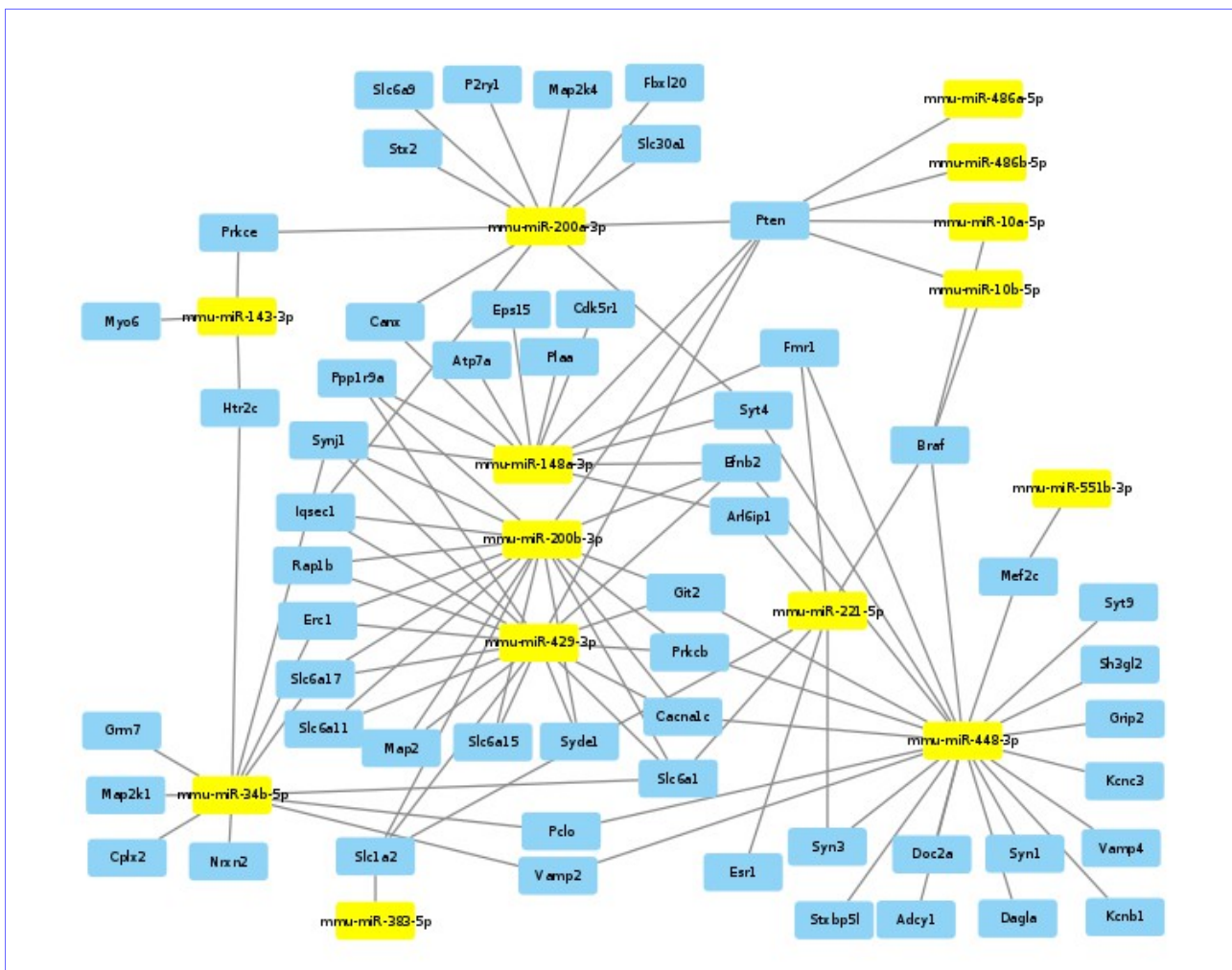


Figura 31. Mapa de red bipartita del grupo 2 del experimento 2 .

En la tercera se encuentran:

Categoría	Descripción	Nº genes
GO:0007612	<i>Learning</i>	28
GO:0042596	<i>Fear response</i>	15
GO:0007611	<i>Learning or memory</i>	38
GO:0050890	<i>Cognition</i>	39

Hay 45 genes *target* diferentes implicados en estas funciones biológicas.

2.11.4 Cruce de ambos experimentos.

Con el proceso de cruce de miRNAs/isomiRs DE de los experimentos 1 y 2 se obtuvo la tabla de miRNAs DE comunes. Posteriormente, a partir de esta tabla de cruce se obtuvo una predicción de genes *target*. Sobre esta lista de genes se hace un enriquecimiento GO. El mapa de enriquecimiento resultante es el siguiente.

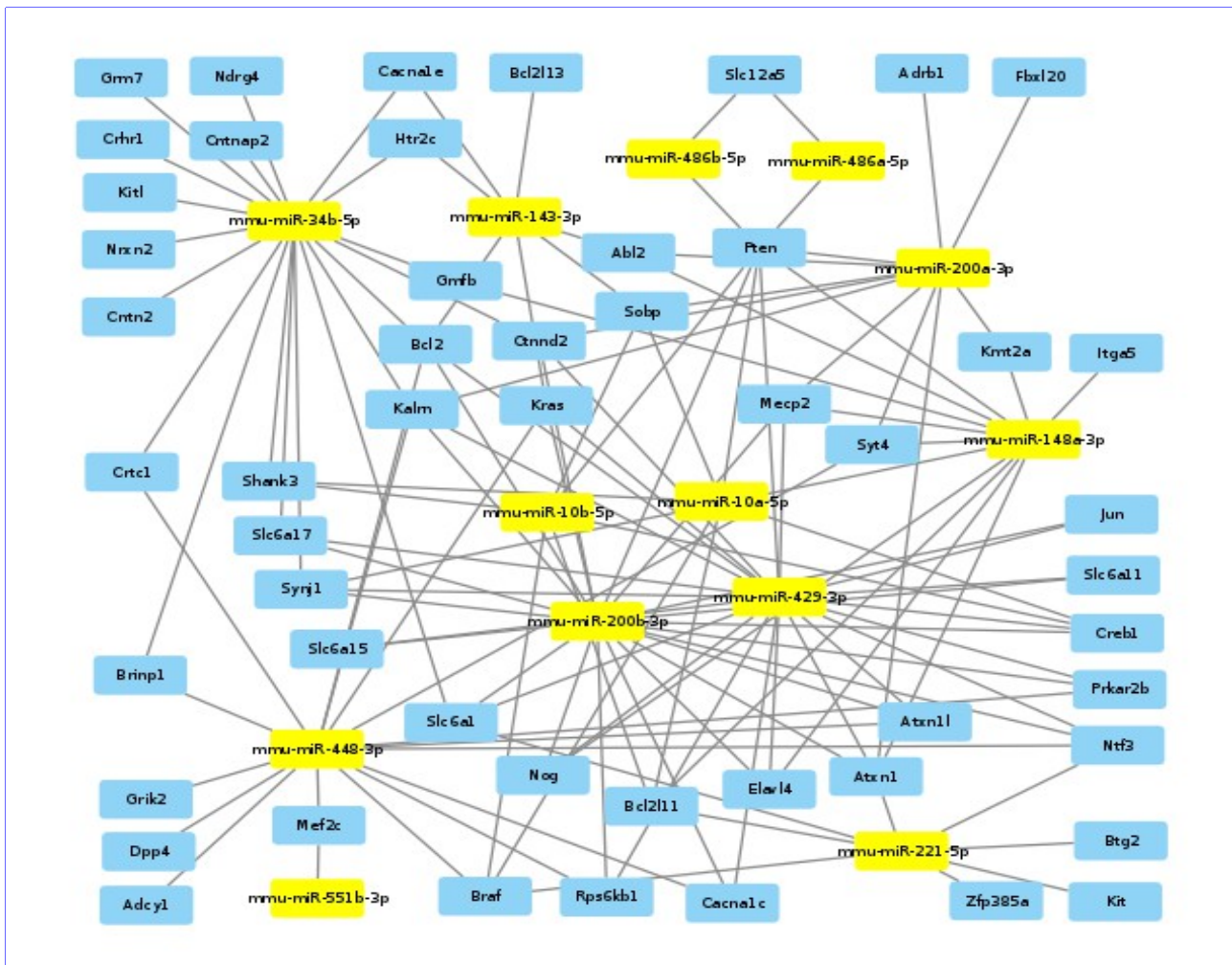


Figura 32. Mapa de red bipartita del grupo 3 del experimento 2 .

Aquí se distinguen 2 agrupaciones de categorías funcionales con referencias al desarrollo del sistema nervioso. En el primer grupo se encuentran las categorías:

Categoría	Descripción	Nº genes
GO:0016358	Dendrite development	32
GO:0050808	Synapse organization	48
GO:0045665	Negative regulation of neuron differentiation	28
GO:0010769	Regulation of cell morphogenesis involved in differentiation	38
GO:0097485	Neuron projection guidance	32
GO:0045666	Positive regulation of neuron differentiation	49
GO:0007409	Axonogenesis	51
GO:0050770	Regulation of axonogenesis	27
GO:0007411	Axon guidance	31
GO:0030900	Forebrain development	39

Hay 127 genes *target* diferentes implicados en estas funciones biológicas.

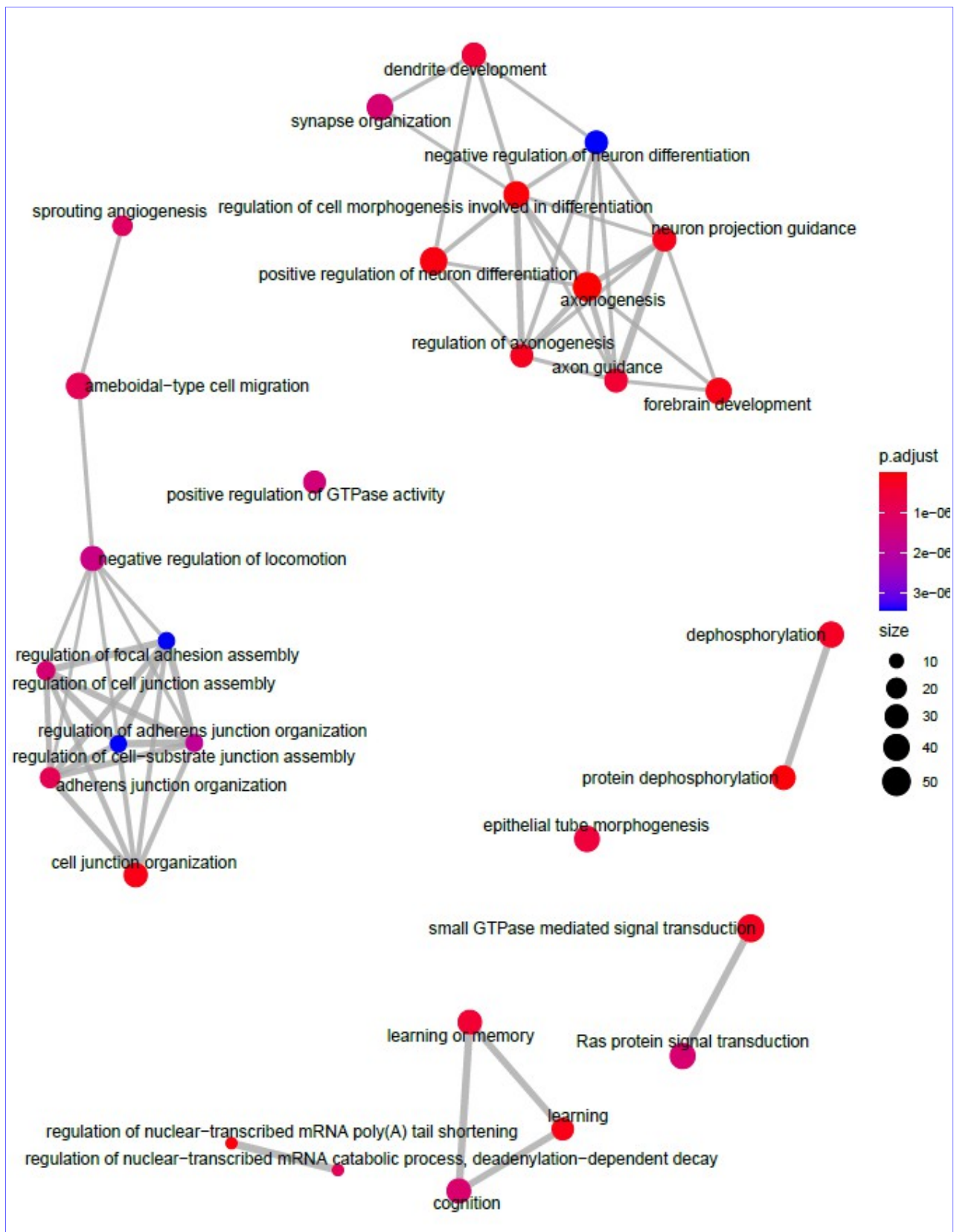


Figura 33. Mapa de enriquecimiento de los genes target seleccionados en el cruce de los 2 experimentos.

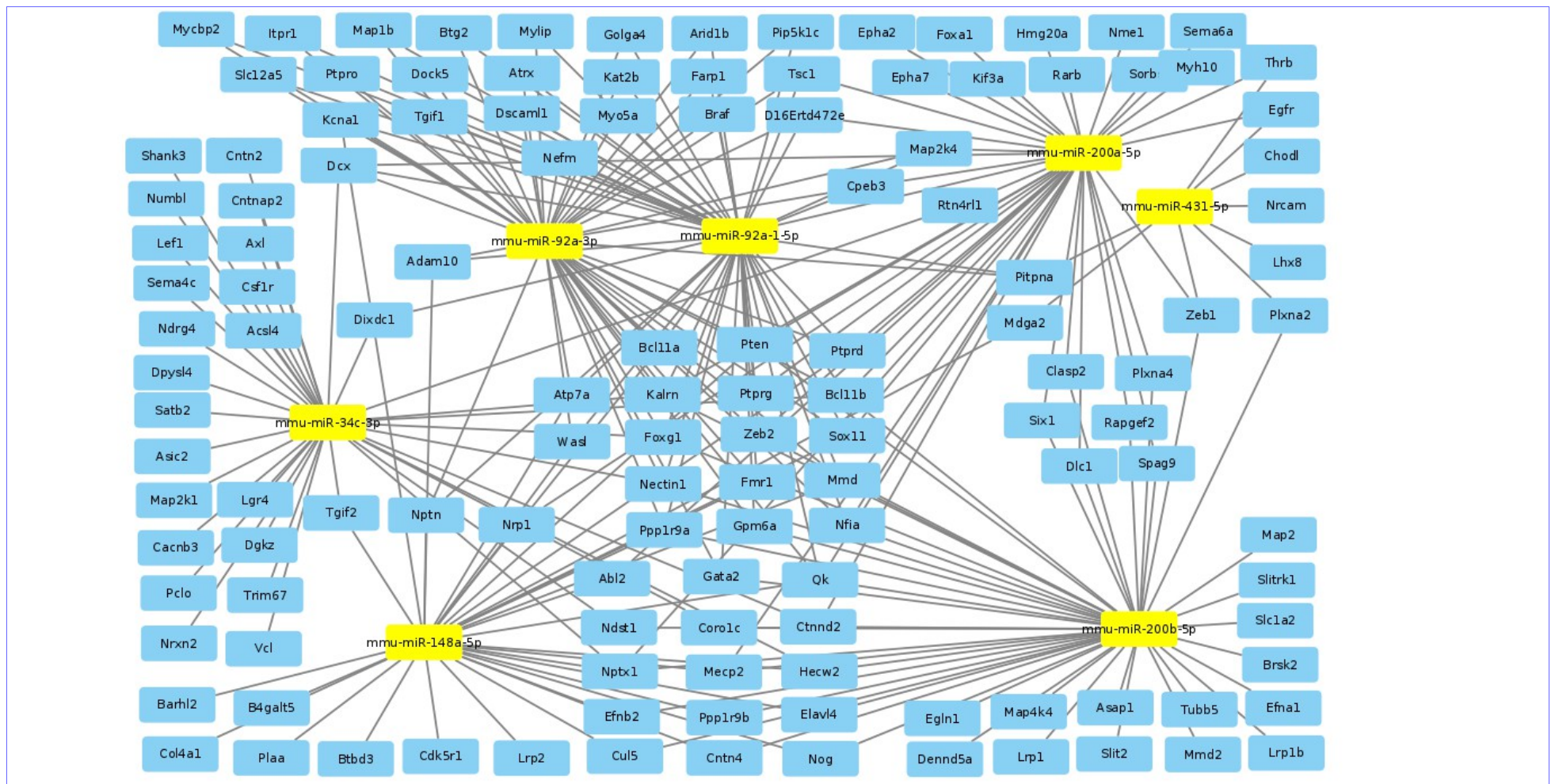


Figura 34. Mapa de red bipartita de los miRNAs y genes target seleccionados en el grupo 1 de enriquecimiento GO del cruce de los 2 experimentos.

En el segundo agrupamiento se encuentran las siguientes categorías:

Categoría	Descripción	Nº genes
GO:0007611	Learning or memory	32
GO:0007612	Learning	25
GO:0050890	Cognition	33

Hay 33 genes *target* diferentes implicados en estas funciones biológicas.

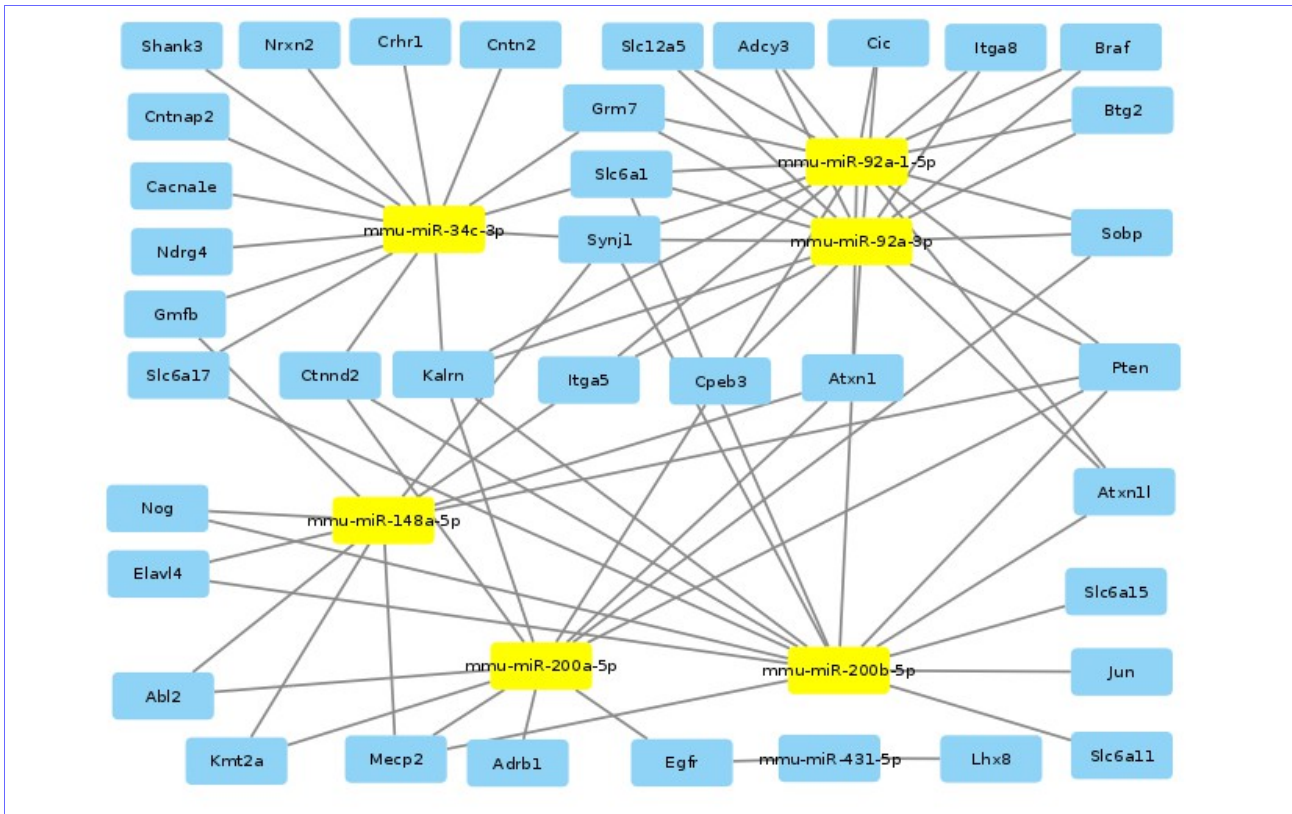


Figura 35. Mapa de red bipartita del grupo 2 de categorías GO en el cruce de los dos experimentos.

2.12 Enriquecimiento KEGG.

El enriquecimiento KEGG (Kyoto Encyclopedia of Genes and Genomes) también se realiza con el paquete *clusterProfile* de Bioconductor [21].

2.12.1 Enriquecimiento KEGG en el experimento 1 (ratones *mecp2*^{+/-}).

Se enriquece la información de genes *target* seleccionados por TargetScan en el experimento 1 y se filtran los *pathways* con más presencia de genes seleccionados en las categorías GO destacadas en el apartado anterior.

```
kk1 <- enrichKEGG(gene = genes1ts$gene, organism = 'mmu', pvalueCutoff = 0.05)
#buscamos en los pathways el numero de genes implicados de los seleccionados en GO
#con caracter neurologico
kk1r <- kk1[, c("ID", "geneID")]
for (i in 1:nrow(kk1r)) { #bucle por pathway
  g <- unlist(strsplit(kk1r$geneID[i], "/"))
```

```

id<-kk1r$ID[i]
nf<-vector()
for(j in 1:length(g)){
  nf<-c(nf,match(g[j],genes1ts$gene))
}
gn<-genes1ts$gene.name[nf] #genes por symbol en el pathway
nf<-vector()
for(i in 1:length(gn)){ #busqueda en genes seleccionados por GO
  nf<-c(nf,match(gn[i],genes1s,nomatch = 0))
}
nf<-subset(nf,nf>0)
if (length(nf)>5){
  print(subset(kk1[,c("ID", "Description", "GeneRatio", "p.adjust")],kk1$ID==id))
}
}
##          ID          Description GeneRatio    p.adjust
## mmu04810 mmu04810 Regulation of actin cytoskeleton    27/359 4.828233e-05
## mmu04510 mmu04510 Focal adhesion                    24/359 0.0001505022
## mmu04151 mmu04151 PI3K-Akt signaling pathway         32/359 0.001025688
## mmu05206 mmu05206 MicroRNAs in cancer               26/359 0.004910252
## mmu04360 mmu04360 Axon guidance                    17/359 0.009229576

```

2.12.2 Enriquecimiento KEGG en el experimento 2 (ratones foxg1+/-).

Se realiza el mismo proceso de enriquecimiento KEGG para los genes target seleccionados en el experimento 2, filtrando los pathways por mayor presencia de los genes seleccionados por los 3 grupos categorías GO señaladas anteriormente. Para el primer grupo se seleccionan los pathways:

##	ID	Description	GeneRatio	p.adjust
##	mmu04727	mmu04727 GABAergic synapse	16/312	1.411396e-05
##	mmu04015	mmu04015 Rap1 signaling pathway	25/312	1.411396e-05
##	mmu04010	mmu04010 MAPK signaling pathway	30/312	1.411396e-05
##	mmu04211	mmu04211 Longevity regulating pathway	15/312	3.458688e-05
##	mmu01522	mmu01522 Endocrine resistance	15/312	4.285126e-05
##	mmu04151	mmu04151 PI3K-Akt signaling pathway	32/312	4.50833e-05
##	mmu04360	mmu04360 Axon guidance	21/312	4.804338e-05
##	mmu01521	mmu01521 EGFR tyrosine kinase inhibitor resistance	13/312	0.0001029755
##	mmu04014	mmu04014 Ras signaling pathway	22/312	0.0004958467
##	mmu05215	mmu05215 Prostate cancer	13/312	0.0007639823
##	mmu04068	mmu04068 FoxO signaling pathway	15/312	0.0009080047
##	mmu04340	mmu04340 Hedgehog signaling pathway	8/312	0.004438143
##	mmu05206	mmu05206 MicroRNAs in cancer	23/312	0.004586942
##	mmu05166	mmu05166 Human T-cell leukemia virus 1 infection	19/312	0.007043665

##	mmu04510	mmu04510	Focal adhesion	16/312	0.01028232
##	mmu04928	mmu04928	Parathyroid hormone synthesis, sec. and act.	10/312	0.01897064

Para el segundo se seleccionan:

##	ID	Description	GeneRatio	p.adjust
##	mmu04721	mmu04721	Synaptic vesicle cycle	8/312 0.02296373

Y para el tercero:

##	ID	Description	GeneRatio	p.adjust
##	mmu04010	mmu04010	MAPK signaling pathway	30/312 1.411396e-05
##	mmu01522	mmu01522	Endocrine resistance	15/312 4.285126e-05
##	mmu04151	mmu04151	PI3K-Akt signaling pathway	32/312 4.50833e-05
##	mmu05224	mmu05224	Breast cancer	17/312 0.000365198
##	mmu04934	mmu04934	Cushing syndrome	17/312 0.000825063
##	mmu05166	mmu05166	Human T-cell leukemia virus 1 infection	19/312 0.007043665

2.12.3 Enriquecimiento KEGG cruce de los 2 experimentos.

Mismo proceso para el primer grupo de genes *target* seleccionado:

##	ID	Description	GeneRatio	p.adjust
##	mmu04151	mmu04151	PI3K-Akt signaling pathway	31/293 0.0002060361
##	mmu04010	mmu04010	MAPK signaling pathway	26/293 0.0006653252
##	mmu04015	mmu04015	Rap1 signaling pathway	20/293 0.001505787
##	mmu04810	mmu04810	Regulation of actin cytoskeleton	20/293 0.001505787
##	mmu04510	mmu04510	Focal adhesion	19/293 0.001505787
##	mmu04360	mmu04360	Axon guidance	17/293 0.003284403

Y para el segundo grupo:

##	ID	Description	GeneRatio	p.adjust
##	mmu04510	mmu04510	Focal adhesion	19/293 0.001505787

2.12.4 Enriquecimiento KEGG por subredes activas.

Sobre los resultados de miRNAs DE y sus genes targets se hace una selección de los más significativos, con el objeto de acotarlos y obtener conclusiones. Los criterios empleados se indican en el documento biológico.

Sobre los genes target seleccionados, se ha hecho un análisis de pathways con pathfindR [24], que tiene en cuenta información de interacciones entre complejos proteína-proteína, identificando primero subredes activas para representar un análisis de enriquecimiento usando éstas. Una subred activa se define como un conjunto de genes interconectados en

una red de interacción proteína-proteína (PIN, protein-protein interaction network), que consiste esencialmente en un conjunto de genes alterados significativamente.

Experimento 1 (ratones Mecp2+/-)

Los miRNAs sobre los que se centra el análisis son:

Down-reg: miR-200a-5p.

Up-reg: miR-328-3p, miR-1a-3p, miR-92a1-5p, miR-411-3p, miR-431-5p, miR-9b-3p y miR-148a-5p.

Para preparar el dataset de entrada, se ha considerado que los miRNAs *down-regulated* (en este caso miR-200a-5p), provocarán que sus genes target estén sobre-expresados y por el contrario, los genes target de los miRNAs *up-reg* (el resto de los seleccionados en este experimento) estarán infra-expresados.

```
#instrucciones para analizar pathways con pathfindR
#experimento 1
#gs1 es el listado de genes seleccionados del experimento 1
#los 2 primeros son target de miR-200a-5p (down-reg)
#el resto son target de miRs up-reg
#preparamos el df de entrada
set.seed(123456)
lfcu<-sample(c(0.25, 0.5, 1), size=2,prob=c(0.2, 0.6, 0.2), replace=TRUE)
lfc<-sample(c(-0.25, -0.5, -1), size=length(gs1)-2,prob=c(0.2, 0.6, 0.2), replace=TRUE)
lfc<-c(lfcu,lfc)
rand_p_vals <- sample(seq(0.001, 0.05, length.out = 5),size = length(gs1),replace = TRUE)
df1s <- data.frame(Gene_symbol = gs1,logFC = lfc, p_val = rand_p_vals)
#ejecutamos pathfindR
pf1s <- run_pathfindR(df1s,gene_sets = "mmu_KEGG",pin_name_path = "mmu_STRING",
                      convert2alias = FALSE,output_dir = "~/TFM/R/pathfindR_1s")
```

Los pathways seleccionados, con sus genes target implicados son:

	Term_Description	Up_regulated	Down_regulated
1	Axon guidance	NRP1 PAK3, EFNB2,	EPHA7, EPHB4
2	Inositol phosphate metabolism		MTMR2, PTEN
3	Cocaine addiction		CDK5R1, DLG4
4	Phosphatidylinositol signaling system		PTEN, MTMR2
5	Breast cancer		PTEN, HEYL
6	TGF-beta signaling pathway		TGIF1, TGIF2
7	FoxO signaling pathway		PTEN, FOXG1
8	Focal adhesion		TNC, PTEN, PAK3
9	Notch signaling pathway		HEYL
10	Endometrial cancer		PTEN
11	Renal cell carcinoma		PAK3
12	ErbB signaling pathway		PAK3
13	Regulation of actin cytoskeleton		PAK3, GIT1
14	Small cell lung cancer		PTEN
15	Central carbon metabolism in cancer		PTEN

Experimento 2 (ratones Foxg1+/-)

Se identifican los siguientes miRNAs desregulados como más significativos en el segundo experimento:

Down-reg: miR-429, miR-200b y miR-200a (de la misma familia).

Up-reg: miR-448.

Con estos miRNAs se localizan los siguientes genes *target* con predicciones más ajustadas, asociados a posibles *pathways* desregulados:

	Term_Description	Up_regulated	Down_regulated
1	Colorectal cancer	JUN, RPS6KB1	BCL2
2	Rap1 signaling pathway	RAP1B, P2RY1, RAPGEF2	IGF1R
3	Synaptic vesicle cycle	STX2, SLC6A1, SLC6A11, SLC1A2	
4	HIF-1 signaling pathway	RPS6KB1, EGLN1	IGF1R, BCL2
5	Breast cancer	JUN, RPS6KB1	IGF1R
6	Neurotrophin signaling pathway	RAP1B, JUN	BCL2
7	MAPK signaling pathway	RAPGEF2, RAP1B, MAP4K4, JUN	CACNA1C, IGF1R, MEF2C
8	GABAergic synapse	SLC6A1, SLC6A11	GPHN, CACNA1C
9	TGF-beta signaling pathway	NOG, RPS6KB1	SMURF1
10	Choline metabolism in cancer	RPS6KB1, JUN	
11	PD-L1 expression and PD-1 checkpoint pathway in cancer	JUN, RPS6KB1	
12	Prostate cancer	ZEB1	IGF1R, BCL2
13	Focal adhesion	RAP1B, JUN	IGF1R, BCL2
14	Human immunodeficiency virus 1 infection	RPS6KB1, JUN	BCL2
15	Apoptosis	JUN	BCL2
16	Renal cell carcinoma	EGLN1, RAP1B, JUN	
17	Autophagy - animal	RPS6KB1	IGF1R, BCL2
18	Measles	JUN	BCL2
19	Fluid shear stress and atherosclerosis	JUN	MEF2C, BCL2
20	Hepatitis B	JUN	BCL2
21	cAMP signaling pathway	ADRB1, RAP1B, JUN	CACNA1C
22	ErbB signaling pathway	ABL2, JUN, RPS6KB1	
23	Longevity regulating pathway - multiple species	RPS6KB1	IGF1R
24	Long-term potentiation	RAP1B	CACNA1C
25	Amphetamine addiction	JUN	CACNA1C
26	AGE-RAGE signaling pathway in diabetic complications	JUN	BCL2
27	Epstein-Barr virus infection	JUN	BCL2
28	Human T-cell leukemia virus 1 infection	JUN	CRTC1
29	Ras signaling pathway	RAP1B, ABL2	IGF1R
30	Axon guidance	EFNB2, PLXNA4, SEMA6A	UNC5A, EPHA4
31	Longevity regulating pathway	RPS6KB1	IGF1R
32	GnRH signaling pathway	JUN	CACNA1C
33	mTOR signaling pathway	RPS6KB1	IGF1R
34	Estrogen signaling pathway	JUN	BCL2
35	SNARE interactions in vesicular transport	STX2	
36	Hepatocellular carcinoma	RPS6KB1	IGF1R
37	Platelet activation	P2RY1, RAP1B	
38	Gastric cancer	RPS6KB1	BCL2
39	Ovarian steroidogenesis		IGF1R
40	NOD-like receptor signaling pathway	JUN	BCL2
41	Cocaine addiction	JUN	
42	Insulin signaling pathway	RPS6KB1	PRKAR2B
43	Melanoma		IGF1R
44	Inositol phosphate metabolism	SYNJ1	
45	Amyotrophic lateral sclerosis (ALS)	SLC1A2	BCL2
46	Regulation of lipolysis in adipocytes	ADRB1	
47	Oxytocin signaling pathway	JUN	MEF2C, CACNA1C
48	Long-term depression		IGF1R
49	Protein digestion and absorption		DPP4
50	Cushing syndrome	KMT2A, RAP1B	CACNA1C
51	Mitophagy - animal	JUN	
52	Progesterone-mediated oocyte maturation		IGF1R
53	p53 signaling pathway		BCL2
54	Phosphatidylinositol signaling system	SYNJ1	
55	Leishmaniasis	JUN	
56	Viral myocarditis	ABL2	

Cruce de los 2 experimentos

En el cruce de datos de ambos experimentos se identifican los siguientes miRNAs desregulados como más significativos:

Down-reg: miR-200b y miR-200a.

Up-reg: miR-92a, miR-148a y miR-431.

Con estos miRNAs se identifican como posiblemente desregulados los siguientes pathways y sus genes *target* asociados:

	Term Description	Up_regulated	Down_regulated
1	ErbB signaling pathway	ABL2	EGFR, MAP2K4, BRAF
2	Hepatocellular carcinoma		EGFR, BRAF, PTEN, ARID1B
3	Endocytosis	ASAP1	PIP5K1C, EGFR, WASL
4	Axon guidance	EFNB2, SLIT2, PLXNA4, SEMA6A	PLXNA2, NRP1
5	Endometrial cancer		EGFR, PTEN, BRAF
6	Central carbon metabolism in cancer		EGFR, PTEN
7	HIF-1 signaling pathway	EGLN1	EGFR
8	Melanoma		EGFR, BRAF, PTEN
9	Inositol phosphate metabolism		PIP5K1C, PTEN
10	Glioma		EGFR, BRAF, PTEN
11	Focal adhesion		EGFR, PTEN, BRAF, PIP5K1C
12	Fc gamma R-mediated phagocytosis	ASAP1	PIP5K1C
13	Phosphatidylinositol signaling system		PIP5K1C, PTEN
14	Prostate cancer	ZEB1	EGFR, PTEN, BRAF
15	Bladder cancer		BRAF, EGFR
16	Yersinia infection		PIP5K1C, WASL, MAP2K4
17	MAPK signaling pathway	RAPGEF2, CSF1R, MAP4K4	EGFR, BRAF, MAP2K4
18	FoxO signaling pathway	FOXG1	PTEN, EGFR, BRAF
19	Regulation of actin cytoskeleton		EGFR, BRAF, PIP5K1C, WASL
20	Choline metabolism in cancer		EGFR, TSC1, PIP5K1C
21	Breast cancer		EGFR, BRAF, PTEN
22	PD-L1 expression and PD-1 checkpoint pathway in cancer		EGFR, PTEN
23	Hepatitis C		EGFR, BRAF
24	Hepatitis B		MAP2K4, BRAF
25	Adherens junction		WASL, EGFR
26	Relaxin signaling pathway		MAP2K4, EGFR
27	Parathyroid hormone synthesis, secretion and action		BRAF, EGFR
28	GnRH signaling pathway		MAP2K4, EGFR
29	Insulin signaling pathway		TSC1, BRAF
30	mTOR signaling pathway		TSC1, BRAF, PTEN
31	Rap1 signaling pathway	CSF1R, RAPGEF2	EGFR, BRAF
32	Ras signaling pathway	CSF1R, ABL2	EGFR
33	Phospholipase D signaling pathway		EGFR, TSC1, PIP5K1C
34	Autophagy - animal		PTEN, TSC1
35	Cholesterol metabolism	LRP1	MYLIP
36	Notch signaling pathway	NUMBL	
37	Human T-cell leukemia virus 1 infection		NRP1, PTEN, MAP2K4
38	Cellular senescence		TSC1, PTEN
39	Acute myeloid leukemia	CSF1R	BRAF
40	Renal cell carcinoma	EGLN1	BRAF
41	Thyroid cancer		BRAF
42	Malaria	LRP1	
43	RNA degradation		BTG2
44	Hedgehog signaling pathway	KIF3A	
45	PPAR signaling pathway	ACSL4	
46	Gap junction	TUBB5	EGFR
47	Cell adhesion molecules (CAMs)	CNTN2, CNTNAP2	NRCAM
48	Long-term depression		BRAF
49	Progesterone-mediated oocyte maturation		CPEB3, BRAF
50	Long-term potentiation		BRAF
51	p53 signaling pathway		PTEN
52	TGF-beta signaling pathway	NOG, TGIF2	
53	Non-small cell lung cancer		EGFR, BRAF
54	Fc epsilon RI signaling pathway		MAP2K4
55	Viral myocarditis	ABL2	

3. RELACIÓN DE HERRAMIENTAS Y PROGRAMAS EMPLEADOS EN EL PROCESO DE ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.

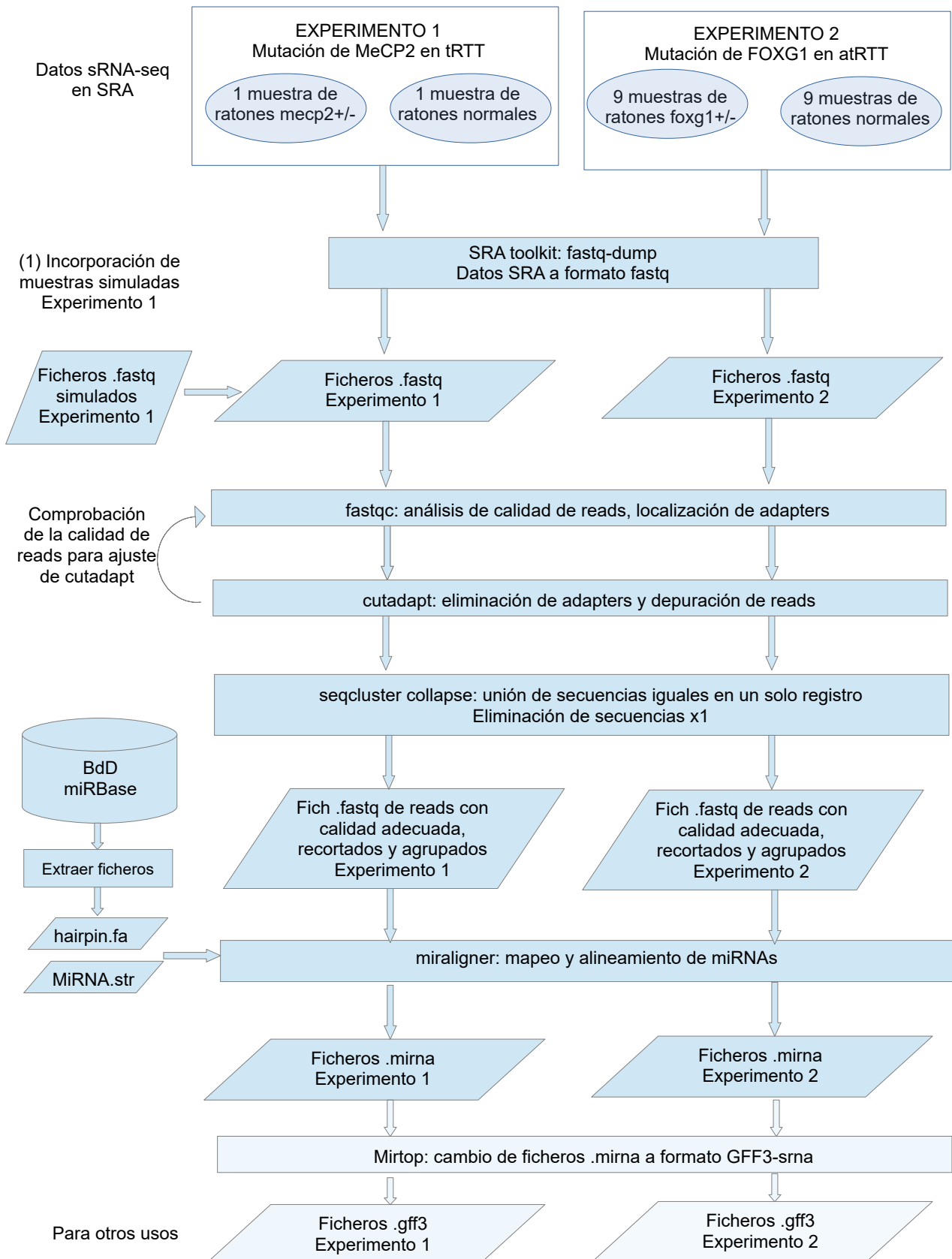
En el *pipeline* para el análisis y comparación de experimentos de smallRNA-seq se han empleado los siguientes programas y herramientas informáticos, todos ellos de libre acceso y sin necesidad de licencia.

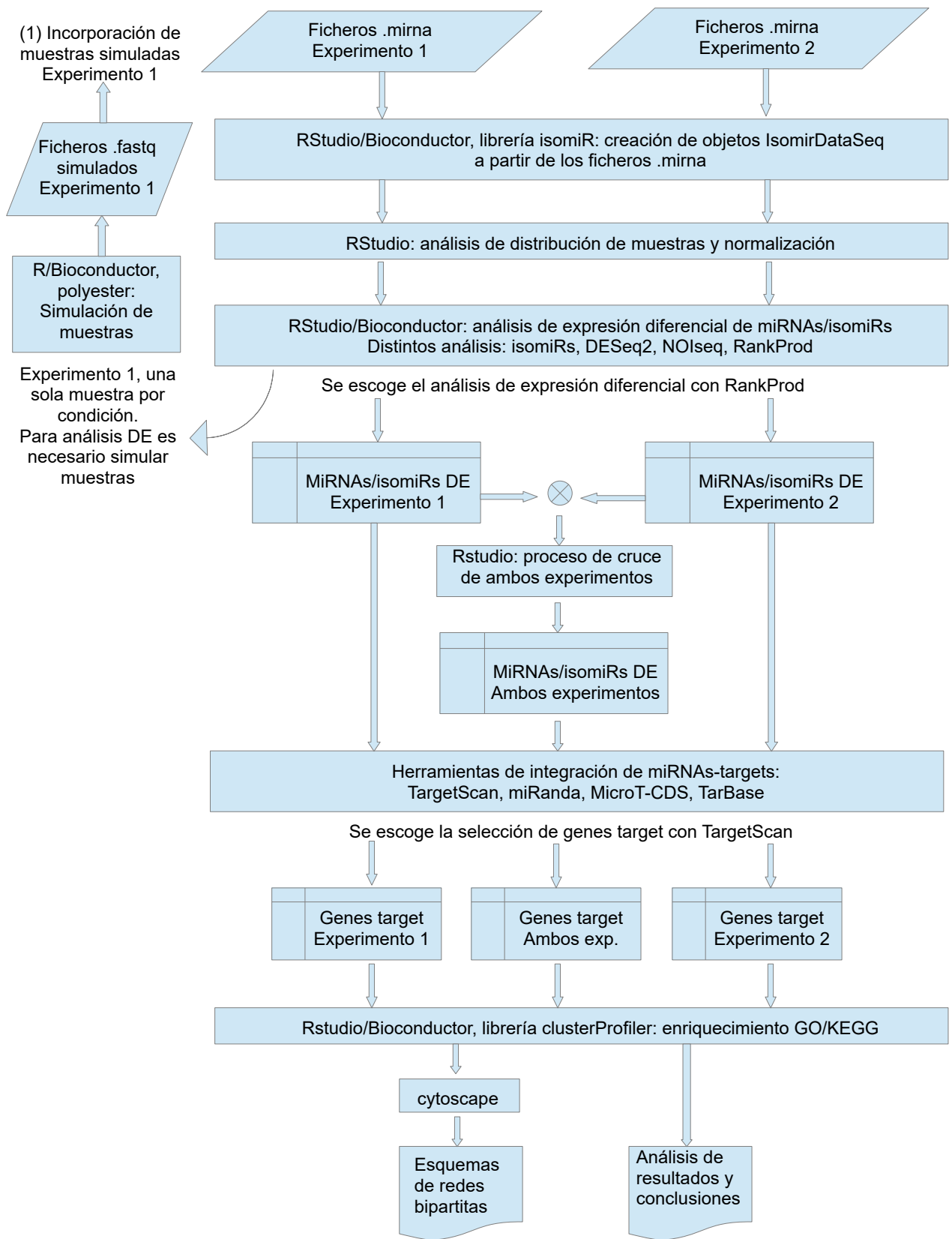
SRA toolkit. Conjunto de herramientas para la descarga de ficheros SRA (<i>Sequence Read Archive</i>) en diferentes formatos (FASTA, FASTQ, SAM y otros)	
fastq-dump	Convierte datos SRA en formato fastq
Herramientas de análisis de calidad de <i>reads</i>	
fastqc	Herramienta de control de calidad para datos de secuenciación de alto rendimiento.
PRINSEQ	Herramienta que genera estadísticas de calidad de datos de secuenciación y que también permite su filtrado, formateo y recorte.
bcbio. Herramienta que proporciona <i>pipelines</i> completos para un análisis automático de datos de secuenciación de alto rendimiento.	
cutadapt	Herramienta que permite eliminar los adapters de las secuencias, a parte de otras modificaciones y filtrados de los <i>reads</i> .
seqcluster	Herramienta para el procesamiento y análisis de datos de smallRNAs.
miraligner	Herramienta de mapeo y anotación de smallRNAs frente al repositorio miRBase.
mirtop	Herramienta para convertir datos .mirna al formato GFF3-srna.
RStudio. Conjunto de herramientas diseñadas para obtener una mayor versatilidad y productividad en el manejo del lenguaje de programación de análisis estadístico R. Además de sus funciones básicas, se han utilizado los siguientes paquetes adicionales.	
reshape2	Paquete de R que permite transformar datos en formatos <i>Wide & Long</i> .
ggplot2	Paquete de R para visualización de datos.
Vennerable	Paquete de R para representación de diagramas de Venn.
igraph	Paquete de R para la creación de redes.
pathfindR	Paquete de R para análisis de enriquecimiento por subredes activas.
Bioconductor. Plataforma basada en el lenguaje de programación R para el análisis de datos genómicos. Además de sus funciones básicas, se han utilizado los siguientes paquetes adicionales.	
isomiRs	Analiza isomiRs y miRNAs de smallRNA-seq.
DESeq2	Realiza análisis de expresión diferencial basados en la distribución binomial negativa.
NOIseq	Método no paramétrico de análisis exploratorio y expresión diferencial para datos RNA-seq.
polyester	Método de simulación de reads.
Biostrings	Manipulación eficiente de secuencias biológicas.
RankProd	Método no paramétrico de identificación de genes diferencialmente

	expresados.
Targetscan.Hs. eg.db	Predicción de targets de miRNA con TargetScan.
clusterProfiler	Análisis estadístico y visualización de perfiles funcionales de genes y agrupaciones de genes.
Herramientas on-line para análisis minuciosos de miRNAs/isomiRs	
IsomirBank	Base de datos de isomiRs para diferentes especies y tejidos.
TargetScan	Servidor web de predicción de targets de miRNAs.
miRanda	Algoritmo de búsqueda de targets de miRNAs.
DIANA MicroT- CDS	Algoritmo de detección de elementos de reconocimiento de miRNAs (MREs) en regiones genómicas.
DIANA TarBase	Base de datos de interacciones miRNA-genes target validadas experimentalmente
Herramientas para representación de redes.	
cytoscape	Plataforma para visualización de redes de interacción molecular e integración con perfiles de expresión genética.

4. DIAGRAMA DE FLUJO DE PIPELINE PARA ANÁLISIS COMPARATIVO DE EXPERIMENTOS DE smallRNA-seq.

El pipeline desarrollado anteriormente se resume en el siguiente diagrama de flujo.





ANEXO 1.

Identificación de miRNAs/isomiRs DE en los 2 experimentos por los distintos métodos de análisis empleados.

Experimento 1. Ratones mecp2+/-

NOISeq

miRNAs DE experimento 1 por análisis NOISeq.

##	control_mean	mutante_mean	M	D	prob
## mmu-miR-122-5p	946	2494	-1.398549	1548	0.9181626
## mmu-miR-1249-3p	17	173	-3.347165	156	0.9396515
## mmu-miR-126a-3p	23207	55781	-1.265214	32574	0.9241816
## mmu-miR-129-2-3p	232	1483	-2.676322	1251	0.9789335
## mmu-miR-129-5p	2640	7754	-1.554403	5114	0.9311510
## mmu-miR-137-3p	11501	24044	-1.063918	12543	0.9102429
## mmu-miR-153-3p	586	1524	-1.378890	938	0.9114572
## mmu-miR-187-3p	2556	5418	-1.083873	2862	0.9033791
## mmu-miR-203-3p	3699	8542	-1.207439	4843	0.9147307
## mmu-miR-206-3p	3661	9955	-1.443184	6294	0.9302006
## mmu-miR-296-5p	68	233	-1.776723	165	0.9154699
## mmu-miR-29b-3p	2570	7196	-1.485427	4626	0.9306230
## mmu-miR-300-3p	2147	6184	-1.526218	4037	0.9299366
## mmu-miR-3059-5p	22	142	-2.690316	120	0.9277719
## mmu-miR-3085-3p	50	269	-2.427606	219	0.9428194
## mmu-miR-3086-5p	1364	519	1.394037	845	0.9102429
## mmu-miR-323-3p	495	1222	-1.303744	727	0.9004752
## mmu-miR-328-3p	641	2812	-2.133200	2171	0.9741816
## mmu-miR-335-3p	214	1098	-2.359195	884	0.9713305
## mmu-miR-344-3p	1279	3437	-1.426134	2158	0.9230201
## mmu-miR-3547-3p	14	116	-3.050626	102	0.9241816
## mmu-miR-369-3p	2654	5729	-1.110115	3075	0.9056494
## mmu-miR-370-3p	7852	18260	-1.217555	10408	0.9203273
## mmu-miR-376a-3p	477	1685	-1.820687	1208	0.9577086
## mmu-miR-376a-5p	434	1094	-1.333846	660	0.9049630
## mmu-miR-380-3p	279	901	-1.691262	622	0.9445618
## mmu-miR-380-5p	11317	25602	-1.177765	14285	0.9169483
## mmu-miR-410-3p	1185	3682	-1.635603	2497	0.9611932
## mmu-miR-412-5p	581	1493	-1.361604	912	0.9107709
## mmu-miR-433-3p	2517	6813	-1.436585	4296	0.9291447
## mmu-miR-433-5p	52	229	-2.138764	177	0.9299366
## mmu-miR-488-3p	5284	14104	-1.416402	8820	0.9313622
## mmu-miR-488-5p	143	540	-1.916944	397	0.9394931
## mmu-miR-496a-3p	84	252	-1.584963	168	0.9145723
## mmu-miR-5099	519	1574	-1.600629	1055	0.9524287
## mmu-miR-5128	18	115	-2.675565	97	0.9194826
## mmu-miR-532-3p	152	469	-1.625517	317	0.9305174
## mmu-miR-770-5p	146	661	-2.178682	515	0.9553326
## mmu-miR-7a-5p	6688	16694	-1.319683	10006	0.9282999
## mmu-miR-7b-5p	3208	9409	-1.552367	6201	0.9315206
## mmu-miR-874-3p	97	359	-1.887927	262	0.9307286
## mmu-miR-879-5p	62	197	-1.667856	135	0.9076558
## mmu-miR-92b-3p	6649	13347	-1.005306	6698	0.9071806

DESeq2

miRNAs DE experimento 1 por análisis DESeq2.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
mmu-miR-335-3p	375.410577	-3.741966	0.8494551	-4.405137	1.057171e-05	0.007791353
mmu-miR-129-2-3p	350.734311	-2.960624	0.7234621	-4.092300	4.271147e-05	0.015739176
mmu-miR-376a-3p	605.799994	-2.791773	0.7305614	-3.821407	1.326924e-04	0.032598101
mmu-miR-1249-3p	33.662564	-4.578207	1.2245433	-3.738706	1.849703e-04	0.034080769
mmu-miR-1247-5p	13.493024	-4.933047	1.4197186	-3.474665	5.114909e-04	0.054000027
mmu-miR-1298-5p	12096.158469	2.787351	0.8023619	3.473932	5.128904e-04	0.054000027
mmu-miR-328-3p	1163.231994	-2.593303	0.7419693	-3.495163	4.737729e-04	0.054000027
mmu-miR-200a-5p	5.376047	6.138508	1.8215297	3.369974	7.517523e-04	0.069255184
mmu-miR-125b-1-3p	844.217504	2.298096	0.7135665	3.220577	1.279328e-03	0.094286464
mmu-miR-7090-3p	11.413508	-4.692451	1.4516095	-3.232585	1.226757e-03	0.094286464
mmu-miR-148a-5p	77.040842	-2.924563	0.9414312	-3.106508	1.893113e-03	0.099658865
mmu-miR-34b-5p	115.834367	-3.063645	0.9770392	-3.135642	1.714783e-03	0.099658865
mmu-miR-5128	10.843421	-4.627962	1.4840799	-3.118405	1.818329e-03	0.099658865
mmu-miR-92a-1-5p	15.278050	-3.890243	1.2276183	-3.168935	1.529985e-03	0.099658865

RankProd

miRNAs DE experimento 1 por análisis RankProd.

```
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
##
##      gene.index RP/Rsum FC:(class1/class2) pfp P.value
## mmu-miR-672-5p      700    15.46          0.0639 0.1579 0.0001685
## mmu-miR-1298-3p     93    31.82          0.1608 0.7659 0.0016350
## mmu-miR-1968-5p    227    31.83          0.1691 0.5114 0.0016370
## mmu-miR-1298-5p     94    37.65          0.1829 0.6294 0.0026870
## mmu-miR-140-5p     123    38.17          0.1391 0.5241 0.0027970
## mmu-miR-574-5p     662    39.57          0.1893 0.4847 0.0031040
## mmu-miR-125b-1-3p   76    43.91          0.2097 0.5603 0.0041860
## mmu-miR-200a-5p    251    43.99          0.2106 0.4928 0.0042080
## mmu-miR-150-3p     142    47.38          0.2199 0.5403 0.0051900
## mmu-miR-186-3p     182    51.56          0.2251 0.6155 0.0065690
## mmu-miR-212-5p     270    52.40          0.2379 0.5850 0.0068670
## mmu-miR-30f        421    53.56          0.6198 0.5697 0.0072960
## mmu-miR-129b-3p     95    54.61          0.3803 0.5545 0.0076940
## mmu-miR-7047-5p    809    58.57          0.2665 0.6228 0.0093060
## mmu-miR-362-5p     505    58.86          0.2170 0.5890 0.0094290
## mmu-miR-1843a-3p   176    62.54          0.2640 0.6497 0.0110900
## mmu-miR-382-5p     536    63.44          0.2316 0.6351 0.0115200
## mmu-miR-871-3p     903    64.92          0.2700 0.6374 0.0122500
## mmu-miR-34c-3p     490    66.01          0.2585 0.6310 0.0127900
## mmu-miR-205-5p     261    69.23          0.2783 0.6788 0.0144900
## mmu-miR-6970-5p    758    69.96          0.2766 0.6641 0.0148800
## mmu-miR-1224-5p     63    70.88          0.2860 0.6559 0.0154000
## mmu-miR-184-3p     175    71.64          0.2818 0.6447 0.0158300
## mmu-miR-99b-3p     934    72.74          0.2855 0.6426 0.0164600
```


## mmu-miR-129-5p	92	73.12	0.2928	0.6252	0.0166800
## mmu-miR-3572-5p	501	74.73	0.3004	0.6354	0.0176300
## mmu-miR-6988-3p	765	77.02	0.3045	0.6604	0.0190300
## mmu-miR-1912-5p	197	78.71	0.3022	0.6724	0.0200900
## mmu-miR-539-3p	638	79.95	0.3183	0.6751	0.0208900
## mmu-miR-150-5p	143	80.26	0.3102	0.6589	0.0211000
## mmu-miR-12191-3p	51	80.44	0.3148	0.6412	0.0212100
## mmu-miR-29b-1-5p	342	83.97	0.3228	0.6906	0.0235900
## mmu-miR-30b-3p	412	89.74	0.3465	0.7866	0.0277000
## mmu-miR-200b-5p	253	91.76	0.3449	0.8049	0.0292100
## mmu-miR-126b-3p	83	92.05	1.3560	0.7878	0.0294300
## mmu-miR-9b-3p	936	92.26	2.2920	0.7700	0.0295800
## mmu-miR-1199-5p	43	94.01	0.3455	0.7831	0.0309200
## mmu-miR-30e-5p	420	99.20	0.3585	0.8638	0.0350300
## mmu-miR-3470a	478	99.95	0.3456	0.8564	0.0356400
## mmu-miR-7668-5p	878	99.99	0.3475	0.8357	0.0356800
## mmu-miR-9-3p	918	104.60	0.3677	0.9043	0.0395700
## mmu-miR-1931	202	105.10	0.3628	0.8920	0.0399800
## mmu-miR-122b-3p	64	106.60	0.8461	0.9004	0.0413200
## mmu-miR-181c-5p	168	106.70	0.4207	0.8820	0.0414200
## mmu-miR-3086-5p	400	107.50	0.3778	0.8767	0.0421100
## mmu-miR-673-5p	702	107.70	0.3738	0.8607	0.0422500
## mmu-miR-7066-3p	822	108.20	0.3633	0.8512	0.0426900
## mmu-miR-181b-5p	166	108.40	0.6182	0.8370	0.0428800
## mmu-miR-455-3p	572	110.50	0.3755	0.8564	0.0447900
## mmu-miR-450b-5p	569	112.30	0.4007	0.8713	0.0464900
## mmu-miR-1912-3p	196	112.80	0.4306	0.8631	0.0469800
## mmu-miR-3102-3p	425	113.00	0.3835	0.8500	0.0471700
## mmu-miR-34b-3p	488	113.20	0.4838	0.8359	0.0472800
## mmu-miR-3074-5p	381	113.20	1.0070	0.8211	0.0473200
## mmu-miR-100-5p	21	113.30	0.3968	0.8073	0.0473900
## mmu-miR-3100-5p	424	113.90	0.4036	0.8032	0.0480000
## mmu-miR-192-5p	199	114.30	0.5121	0.7946	0.0483400
## mmu-miR-448-3p	563	115.30	0.4072	0.7961	0.0492800
## mmu-miR-210-3p	266	115.70	0.3931	0.7892	0.0496900

##

\$Table2

##	gene.index	RP/Rsum	FC:(class1/class2)	pfp	P.value
## mmu-miR-335-3p	451	15.33	13.3600	0.1533	0.0001637
## mmu-miR-300-3p	347	21.07	12.1000	0.2156	0.0004602
## mmu-miR-1249-3p	71	22.05	11.0900	0.1657	0.0005306
## mmu-miR-1b-5p	249	27.00	9.4130	0.2331	0.0009950
## mmu-miR-129-2-3p	91	31.54	7.5930	0.2985	0.0015930
## mmu-miR-1247-5p	70	31.97	7.3210	0.2590	0.0016590
## mmu-miR-376a-3p	518	38.42	6.5110	0.3815	0.0028500
## mmu-miR-3084-5p	397	40.24	6.4570	0.3817	0.0032590
## mmu-miR-7090-3p	841	40.37	6.3670	0.3425	0.0032900
## mmu-miR-296-5p	332	42.36	6.0340	0.3540	0.0037780
## mmu-miR-129b-5p	96	43.14	5.8010	0.3390	0.0039790
## mmu-miR-5128	629	46.61	5.8530	0.3869	0.0049550
## mmu-miR-9b-3p	936	46.74	2.2920	0.3601	0.0049960
## mmu-miR-3071-3p	375	47.41	3.6680	0.3480	0.0052000
## mmu-miR-148a-5p	138	49.69	5.8970	0.3704	0.0059290
## mmu-miR-328-3p	442	54.58	5.3820	0.4498	0.0076810
## mmu-miR-1a-3p	247	55.12	5.0430	0.4349	0.0078910
## mmu-miR-431-5p	558	55.48	5.0430	0.4181	0.0080320
## mmu-miR-92a-1-5p	920	58.23	5.0240	0.4517	0.0091590
## mmu-miR-411-3p	547	58.40	5.4250	0.4326	0.0092330
## mmu-miR-196b-5p	231	62.26	4.7910	0.4890	0.0109600

## mmu-miR-154-3p	150	63.54	4.5900	0.4927	0.0115700
## mmu-miR-153-3p	148	65.57	4.5950	0.5122	0.0125700
## mmu-miR-410-3p	545	67.83	4.7890	0.5364	0.0137400
## mmu-miR-149-5p	141	68.87	4.7400	0.5357	0.0142900
## mmu-miR-3071-5p	376	69.22	1.7350	0.5220	0.0144800
## mmu-miR-223-3p	294	71.47	4.3170	0.5459	0.0157300
## mmu-miR-3061-3p	361	75.41	4.1600	0.6037	0.0180400
## mmu-miR-33-5p	446	76.42	4.2940	0.6029	0.0186600
## mmu-miR-423-3p	552	76.80	4.2230	0.5900	0.0188900
## mmu-miR-505-5p	621	78.19	4.1320	0.5975	0.0197700
## mmu-miR-210-5p	267	81.64	4.0260	0.6445	0.0220100
## mmu-miR-18a-5p	188	82.05	4.0880	0.6327	0.0222800
## mmu-miR-219b-3p	283	82.11	1.7990	0.6152	0.0223200
## mmu-miR-139-3p	120	83.48	4.0120	0.6225	0.0232500
## mmu-miR-224-5p	295	84.83	3.8390	0.6294	0.0241800
## mmu-miR-3059-5p	360	86.17	3.5650	0.6363	0.0251200
## mmu-miR-3544-5p	498	86.87	3.6640	0.6317	0.0256200
## mmu-miR-34b-5p	489	88.53	3.8080	0.6443	0.0268200
## mmu-miR-758-3p	865	90.54	3.5590	0.6629	0.0283000
## mmu-miR-3074-5p	381	91.38	1.0070	0.6609	0.0289200
## mmu-miR-488-5p	597	92.45	3.6460	0.6633	0.0297300
## mmu-miR-7070-3p	828	92.52	3.7780	0.6490	0.0297800
## mmu-miR-350-5p	493	93.74	3.7980	0.6541	0.0307100
## mmu-miR-935	927	94.69	3.0180	0.6549	0.0314500
## mmu-miR-323-5p	436	95.07	3.4140	0.6466	0.0317400
## mmu-miR-136-5p	114	99.93	3.5380	0.7103	0.0356300
## mmu-miR-425-5p	555	100.50	3.5550	0.7043	0.0360800
## mmu-miR-7010-5p	776	102.60	3.1440	0.7232	0.0378200
## mmu-miR-92a-3p	922	105.30	3.4260	0.7526	0.0401600
## mmu-miR-19b-1-5p	243	106.00	3.3660	0.7495	0.0407900
## mmu-miR-30f	421	106.70	0.6198	0.7463	0.0414200
## mmu-miR-6928-3p	729	108.00	3.3930	0.7519	0.0425300
## mmu-miR-126b-3p	83	108.70	1.3560	0.7494	0.0431900
## mmu-miR-133a-3p	104	109.00	3.3180	0.7403	0.0434500
## mmu-miR-126a-3p	81	109.10	3.3020	0.7292	0.0435800
## mmu-miR-412-3p	549	110.20	3.3230	0.7324	0.0445500
## mmu-miR-141-3p	124	112.00	3.2860	0.7457	0.0461600
## mmu-miR-29c-5p	346	113.50	3.1720	0.7554	0.0475700
## mmu-miR-770-5p	889	113.60	3.2380	0.7443	0.0476600
## mmu-miR-500-3p	613	114.70	3.2640	0.7493	0.0487800
## mmu-miR-323-3p	435	115.80	2.8740	0.7519	0.0497500

Experimento 2. Ratones foxg1+/-

DESeq2

miRNAs DE experimento 2 por análisis DESeq2.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
mmu-miR-187-3p	1070.93563	0.4938136	0.08391247	5.884866	3.983775e-09	2.314573e-06
mmu-miR-448-3p	119.09585	-1.7305818	0.31285191	-5.531633	3.172637e-08	9.216512e-06
mmu-miR-3552	63.64475	-1.4516347	0.30278831	-4.794223	1.633063e-06	3.162699e-04
mmu-miR-1264-3p	116.94476	-1.1897546	0.25951780	-4.584482	4.551142e-06	6.610534e-04
mmu-miR-497a-5p	549.82976	0.5187534	0.11874562	4.368611	1.250392e-05	1.452955e-03
mmu-miR-551b-3p	244.02136	0.6817082	0.16736342	4.073221	4.636736e-05	4.489906e-03
mmu-miR-200a-3p	11.71034	3.1527803	0.81333753	3.876349	1.060355e-04	8.800944e-03

mmu-miR-429-3p	57.95360	1.9648623	0.52132302	3.768992	1.639081e-04	1.190383e-02
mmu-miR-1298-5p	378.64066	-1.5946559	0.43119912	-3.698189	2.171430e-04	1.401779e-02
mmu-miR-34a-5p	161.08898	0.5108312	0.13916866	3.670591	2.419901e-04	1.405963e-02
mmu-miR-764-5p	10.53341	-1.2415711	0.34213631	-3.628879	2.846551e-04	1.503497e-02
mmu-miR-195a-5p	1948.41025	0.4871754	0.13771216	3.537635	4.037273e-04	1.954713e-02
mmu-miR-764-3p	48.75197	-0.8147314	0.24149228	-3.373737	7.415519e-04	3.314167e-02
mmu-miR-200b-3p	79.65996	2.5301099	0.75456967	3.353050	7.992624e-04	3.316939e-02
mmu-miR-6944-3p	2599.23969	-0.4274745	0.12983220	-3.292515	9.929543e-04	3.846043e-02
mmu-miR-193b-3p	77.12036	0.6388193	0.20428844	3.127046	1.765724e-03	6.411784e-02
mmu-miR-1264-5p	10.18081	-0.9133592	0.30904984	-2.955378	3.122859e-03	9.955039e-02
mmu-miR-126a-5p	12473.47026	0.4743393	0.16120169	2.942521	3.255520e-03	9.955039e-02
mmu-miR-344c-3p	24.65975	0.7937112	0.26702286	2.972447	2.954366e-03	9.955039e-02

RankProd

miRNAs DE experimento 2 por análisis RankProd.

```
## Table1: Genes called significant under class1 < class2
##
## Table2: Genes called significant under class1 > class2
## $Table1
##
## gene.index RP/Rsum FC:(class1/class2) pfp P.value
## mmu-miR-200b-3p 236 11.84 0.3072 1.447e-07 1.476e-10
## mmu-miR-200a-3p 235 18.03 0.4043 3.279e-06 6.691e-09
## mmu-miR-429-3p 527 20.09 0.4122 5.561e-06 1.702e-08
## mmu-miR-10b-5p 37 29.37 0.4969 9.419e-05 3.844e-07
## mmu-miR-383-5p 508 77.91 0.7029 7.648e-02 3.902e-04
## mmu-miR-551b-3p 641 84.90 0.6414 1.084e-01 6.636e-04
## mmu-miR-187-5p 183 105.50 0.6804 3.346e-01 2.390e-03
## mmu-miR-551b-5p 642 120.00 0.6984 6.010e-01 4.906e-03
## mmu-miR-193b-3p 202 123.90 0.7088 6.348e-01 5.830e-03
## mmu-miR-497a-5p 601 127.50 0.7027 6.656e-01 6.792e-03
## mmu-miR-871-3p 950 127.80 0.7447 6.135e-01 6.887e-03
## mmu-miR-344c-3p 441 128.20 0.7069 5.717e-01 7.001e-03
## mmu-miR-5121 618 128.80 0.7210 5.400e-01 7.164e-03
## mmu-miR-126a-5p 80 134.40 0.7243 6.268e-01 8.954e-03
## mmu-miR-6976-3p 782 137.80 0.7232 6.664e-01 1.020e-02
## mmu-miR-195a-5p 214 138.90 0.7257 6.513e-01 1.063e-02
## mmu-miR-187-3p 182 140.70 0.7139 6.543e-01 1.135e-02
## mmu-miR-12191-3p 53 143.10 0.7733 6.732e-01 1.236e-02
## mmu-miR-181c-3p 165 143.60 0.7497 6.499e-01 1.260e-02
## mmu-miR-5099 613 144.00 1.0100 6.258e-01 1.277e-02
## mmu-miR-5615-5p 644 145.00 0.7418 6.167e-01 1.322e-02
## mmu-miR-143-3p 124 147.10 0.9906 6.326e-01 1.420e-02
## mmu-miR-135a-5p 107 152.00 0.7977 7.127e-01 1.673e-02
## mmu-miR-7093-3p 885 153.40 0.7568 7.136e-01 1.748e-02
## mmu-miR-34a-5p 461 154.40 0.7396 7.086e-01 1.808e-02
## mmu-miR-6941-3p 754 156.50 0.7395 7.275e-01 1.930e-02
## mmu-miR-532-5p 627 157.00 1.1200 7.121e-01 1.962e-02
## mmu-miR-135b-5p 109 158.80 0.7579 7.256e-01 2.073e-02
## mmu-miR-92a-3p 965 161.00 0.9537 7.487e-01 2.216e-02
## mmu-miR-190b-5p 190 164.20 0.7659 7.947e-01 2.433e-02
## mmu-miR-16-5p 156 164.50 0.7623 7.745e-01 2.450e-02
## mmu-miR-409-5p 514 165.50 0.9637 7.736e-01 2.526e-02
## mmu-miR-188-5p 185 169.50 0.7899 8.378e-01 2.821e-02
## mmu-miR-30a-3p 379 170.00 0.8649 8.259e-01 2.865e-02
## mmu-miR-582-3p 652 170.20 1.5580 8.061e-01 2.879e-02
```

```

## mmu-miR-6952-3p      763  170.50      0.7877  7.910e-01  2.906e-02
## mmu-miR-501-3p      605  172.90      0.9197  8.198e-01  3.095e-02
## mmu-miR-434-5p      533  174.40      1.0560  8.326e-01  3.229e-02
## mmu-miR-129-1-3p     86   174.50      0.7747  8.135e-01  3.237e-02
## mmu-miR-125b-2-3p   75   174.90      0.9207  7.997e-01  3.264e-02
## mmu-miR-1969        220  177.90      0.8322  8.446e-01  3.534e-02
## mmu-miR-6979-3p     784  179.50      0.7928  8.598e-01  3.685e-02
## mmu-miR-181d-5p     168  179.80      0.7722  8.451e-01  3.708e-02
## mmu-miR-92b-5p      967  180.30      0.8468  8.366e-01  3.756e-02
## mmu-miR-135a-1-3p   105  181.80      0.7812  8.491e-01  3.899e-02
## mmu-miR-7117-3p     888  182.90      0.7594  8.543e-01  4.010e-02
## mmu-miR-30d-3p      386  183.10      0.7924  8.391e-01  4.024e-02
## mmu-miR-148a-3p     134  184.50      1.0600  8.507e-01  4.167e-02
## mmu-miR-344b-3p     440  184.90      0.9267  8.421e-01  4.210e-02
## mmu-miR-1843b-3p    176  186.30      1.0310  8.537e-01  4.355e-02
## mmu-miR-598-5p      656  187.00      0.7784  8.493e-01  4.420e-02
## mmu-miR-466q        565  187.70      0.7821  8.472e-01  4.495e-02
## mmu-miR-363-3p      479  188.10      0.8025  8.395e-01  4.540e-02
## mmu-let-7i-3p       18   188.10      0.8107  8.244e-01  4.542e-02
## mmu-miR-6926-3p     737  188.50      0.7878  8.164e-01  4.582e-02
## mmu-miR-183-5p      170  188.70      0.9129  8.069e-01  4.611e-02
## mmu-miR-8103        944  189.20      0.8048  8.010e-01  4.659e-02
## mmu-miR-7651-5p     912  191.80      0.7700  8.370e-01  4.954e-02
## mmu-miR-7006-3p     814  191.90      0.8269  8.239e-01  4.960e-02
## mmu-miR-182-5p      169  192.10      0.8732  8.131e-01  4.978e-02
## mmu-miR-33-3p       418  192.10      0.8050  8.014e-01  4.988e-02
##
## $Table2
##
##      gene.index  RP/Rsum  FC:(class1/class2)      pfp      P.value
## mmu-miR-448-3p      534    19.87      2.3530  1.515e-05  1.546e-08
## mmu-miR-3552        473    30.29      1.9930  2.400e-04  4.897e-07
## mmu-miR-1264-3p      77     36.46      1.9210  6.672e-04  2.042e-06
## mmu-miR-1298-3p      90     36.56      1.7830  5.107e-04  2.084e-06
## mmu-miR-1298-5p      91     38.00      2.0940  5.471e-04  2.791e-06
## mmu-miR-582-3p      652    48.12      1.5580  2.546e-03  1.559e-05
## mmu-miR-7080-3p     873    56.79      1.6710  6.903e-03  4.931e-05
## mmu-miR-455-3p      541    78.89      1.5070  5.167e-02  4.218e-04
## mmu-miR-532-5p      627    87.10      1.1200  8.446e-02  7.757e-04
## mmu-miR-764-3p      908    88.08      1.5150  8.131e-02  8.297e-04
## mmu-miR-10a-5p       36    94.57      1.4820  1.131e-01  1.269e-03
## mmu-miR-148a-3p     134   106.20      1.0600  2.025e-01  2.480e-03
## mmu-miR-7240-5p     900   106.80      1.2540  1.935e-01  2.566e-03
## mmu-miR-221-5p      275   109.70      1.1480  2.088e-01  2.983e-03
## mmu-miR-5099        613   111.90      1.0100  2.180e-01  3.336e-03
## mmu-miR-764-5p      909   116.70      1.4180  2.572e-01  4.200e-03
## mmu-miR-12193-3p     54   121.40      1.3340  3.014e-01  5.229e-03
## mmu-miR-34b-3p      462   122.40      1.3630  2.970e-01  5.455e-03
## mmu-miR-143-3p      124   122.70      0.9906  2.846e-01  5.517e-03
## mmu-miR-434-5p      533   124.00      1.0560  2.872e-01  5.861e-03
## mmu-miR-540-3p      630   126.40      1.2630  3.022e-01  6.476e-03
## mmu-miR-1264-5p      78   126.50      1.3850  2.905e-01  6.521e-03
## mmu-miR-345-3p      449   130.90      1.0420  3.323e-01  7.799e-03
## mmu-miR-34b-5p      463   135.10      1.3090  3.765e-01  9.220e-03
## mmu-miR-3057-5p     323   140.80      1.2910  4.464e-01  1.139e-02
## mmu-miR-486b-5p     584   143.10      1.3120  4.656e-01  1.235e-02
## mmu-miR-486a-5p     582   144.80      1.3110  4.757e-01  1.311e-02
## mmu-miR-212-5p      255   145.40      1.3370  4.697e-01  1.342e-02
## mmu-miR-409-5p      514   146.90      0.9637  4.772e-01  1.412e-02
## mmu-miR-431-3p     528   151.90      1.0940  5.437e-01  1.664e-02

```

## mmu-miR-6944-3p	756	151.90	1.3280	5.268e-01	1.666e-02
## mmu-miR-132-5p	99	154.30	1.3150	5.516e-01	1.801e-02
## mmu-miR-668-3p	677	154.50	1.2550	5.381e-01	1.812e-02
## mmu-miR-31-5p	391	157.80	1.2610	5.794e-01	2.010e-02
## mmu-miR-423-3p	523	158.10	1.1640	5.676e-01	2.027e-02
## mmu-miR-92a-3p	965	159.50	0.9537	5.762e-01	2.116e-02
## mmu-miR-466c-3p	550	162.40	1.2830	6.115e-01	2.309e-02
## mmu-miR-466p-3p	564	163.10	1.2830	6.065e-01	2.352e-02
## mmu-miR-466b-3p	549	163.40	1.2830	5.967e-01	2.375e-02
## mmu-miR-212-3p	254	163.90	1.2780	5.906e-01	2.411e-02
## mmu-miR-466e-3p	554	166.20	1.2830	6.145e-01	2.571e-02
## mmu-miR-1839-5p	172	168.60	1.1850	6.432e-01	2.757e-02
## mmu-miR-7220-5p	892	171.70	1.2810	6.843e-01	3.002e-02
## mmu-miR-666-5p	674	171.80	1.1370	6.695e-01	3.006e-02
## mmu-miR-7b-5p	938	172.10	1.1800	6.608e-01	3.034e-02
## mmu-miR-375-3p	487	174.30	1.2140	6.848e-01	3.215e-02
## mmu-miR-34c-5p	465	174.50	1.2610	6.733e-01	3.229e-02
## mmu-let-7f-5p	15	176.50	1.1470	6.952e-01	3.405e-02
## mmu-miR-100-5p	23	176.70	1.0130	6.847e-01	3.423e-02
## mmu-miR-3083-3p	353	177.20	1.2600	6.806e-01	3.473e-02
## mmu-miR-1843b-3p	176	177.70	1.0310	6.747e-01	3.511e-02
## mmu-miR-7653-3p	913	177.90	1.2570	6.660e-01	3.534e-02
## mmu-miR-877-5p	959	183.80	1.2480	7.574e-01	4.096e-02
## mmu-miR-125b-2-3p	75	186.40	0.9207	7.922e-01	4.365e-02
## mmu-miR-370-3p	484	186.50	1.1710	7.798e-01	4.376e-02
## mmu-miR-7080-5p	874	187.60	1.1610	7.862e-01	4.492e-02
## mmu-miR-340-3p	432	187.80	1.2300	7.752e-01	4.509e-02
## mmu-miR-770-5p	933	189.00	1.2000	7.847e-01	4.644e-02
## mmu-miR-592-5p	654	189.40	1.2190	7.782e-01	4.685e-02
## mmu-miR-3102-3p	393	189.60	1.2370	7.682e-01	4.703e-02
## mmu-miR-543-3p	636	189.80	1.2150	7.596e-01	4.728e-02
## mmu-miR-1912-3p	193	190.30	1.1730	7.566e-01	4.787e-02
## mmu-miR-3475-3p	459	190.50	1.2130	7.466e-01	4.799e-02

ANEXO 2.

Proceso de cruce de miRNAs/isomiRs comunes en los dos experimentos.

Primera parte. Enriquecimiento de las tablas procedentes del análisis RankProd de ambos experimentos con los campos *base*, *min.pfp.base* y *min.pval.base*.

```
#proceso de cruce de tablas de ambos experimentos

#tabla1 experimento1; añadimos columnas base, min.pfp.base, min.pval.base
m11<-vector()
#obtenemos Los miR.base
for (i in 1:nrow(t1e1DE)){
m11<-c(m11, paste(unlist(strsplit(rownames(t1e1DE)[i],"-"))[2],
                  gsub("[a-z]","",unlist(strsplit(rownames(t1e1DE)[i],"-"))
[3])),sep="-"))
}
#añadimos La columna miR.base a las tablas
t1e1DE<-data.frame(t1e1DE,m11)
t1e1DE<-t1e1DE[order(t1e1DE$m11),]
m11u<-unique(m11[order(m11)])
c1a<-matrix(0,ncol=2)
#vamos recorriendo las tablas y calculando min.pfp y min.pval por miR.base
for (i in 1:length(m11u)) {
x<-subset(t1e1DE,t1e1DE$m11==m11u[i])
r<-apply(x[,4:5],2,min)
  for(c in 1:nrow(x)){
    f<-c(r[1:2])
    c1a<-rbind(c1a,f)
  }
}
c1a<-c1a[2:nrow(c1a),]
#guardamos La nueva tabla con las nuevas columnas base, min.pfp.base y min.pval.base
t1e1DE<-data.frame(t1e1DE,c1a[,1],c1a[,2])
colnames(t1e1DE)<-c(colnames(t1e1DE)[1:5],"base","min.pfp.base","min.pval.base")

#tabla2 experimento1; añadimos columnas base, min.pfp.base, min.pval.base
m21<-vector()
for (i in 1:nrow(t2e1DE)){
m21<-c(m21, paste(unlist(strsplit(rownames(t2e1DE)[i],"-"))[2],
                  gsub("[a-z]","",unlist(strsplit(rownames(t2e1DE)[i],"-"))
[3])),sep="-"))
}
t2e1DE<-data.frame(t2e1DE,m21)
t2e1DE<-t2e1DE[order(t2e1DE$m21),]
m21u<-unique(m21[order(m21)])
c1a<-matrix(0,ncol=2)
for (i in 1:length(m21u)) {
x<-subset(t2e1DE,t2e1DE$m21==m21u[i])
r<-apply(x[,4:5],2,min)
  for(c in 1:nrow(x)){
    f<-c(r[1:2])
    c1a<-rbind(c1a,f)
  }
}
c1a<-c1a[2:nrow(c1a),]
t2e1DE<-data.frame(t2e1DE,c1a[,1],c1a[,2])
colnames(t2e1DE)<-c(colnames(t2e1DE)[1:5],"base","min.pfp.base","min.pval.base")

#tabla1 experimento2; añadimos columnas base, min.pfp.base, min.pval.base
```

```

m12<-vector()
for (i in 1:nrow(t1e2DE)){
m12<-c(m12, paste(unlist(strsplit(rownames(t1e2DE)[i], "-"))[2],
                    gsub("[a-z]", "", unlist(strsplit(rownames(t1e2DE)[i], "-"))
[3])), sep="-"))
}
t1e2DE<-data.frame(t1e2DE, m12)
t1e2DE<-t1e2DE[order(t1e2DE$m12),]
m12u<-unique(m12[order(m12)])
c1a<-matrix(0, ncol=2)
for (i in 1:length(m12u)) {
x<-subset(t1e2DE, t1e2DE$m12==m12u[i])
r<-apply(x[, 4:5], 2, min)
  for(c in 1:nrow(x)){
    f<-c(r[1:2])
    c1a<-rbind(c1a, f)
  }
}
c1a<-c1a[2:nrow(c1a),]
t1e2DE<-data.frame(t1e2DE, c1a[, 1], c1a[, 2])
colnames(t1e2DE)<-c(colnames(t1e2DE)[1:5], "base", "min.pfp.base", "min.pval.base")

#tabla2 experimento2; añadimos columnas base, min.pfp.base, min.pval.base
m22<-vector()
for (i in 1:nrow(t2e2DE)){
m22<-c(m22, paste(unlist(strsplit(rownames(t2e2DE)[i], "-"))[2],
                    gsub("[a-z]", "", unlist(strsplit(rownames(t2e2DE)[i], "-"))
[3])), sep="-"))
}
t2e2DE<-data.frame(t2e2DE, m22)
t2e2DE<-t2e2DE[order(t2e2DE$m22),]
m22u<-unique(m22[order(m22)])
c1a<-matrix(0, ncol=2)
for (i in 1:length(m22u)) {
x<-subset(t2e2DE, t2e2DE$m22==m22u[i])
r<-apply(x[, 4:5], 2, min)
  for(c in 1:nrow(x)){
    f<-c(r[1:2])
    c1a<-rbind(c1a, f)
  }
}
c1a<-c1a[2:nrow(c1a),]
t2e2DE<-data.frame(t2e2DE, c1a[, 1], c1a[, 2])
colnames(t2e2DE)<-c(colnames(t2e2DE)[1:5], "base", "min.pfp.base", "min.pval.base")
#en las 4 tablas hemos añadido las columnas base, min.pfp.base, min.pval.base

```

Segunda parte. Cruce de las tablas de ambos experimentos por miRNA base común.

```

#ponemos un umbral min.pval.base , para acotar la selección
#con umbral=0.05 seleccionamos todos los DE seleccionados anteriormente
#no tiene sentido un umbral > 0.05
umbral=0.025

#cruzamos las 2 tablas 1 de los 2 experimentos
#miRNAs down-reg en la clase mutante
#creamos los dataset con el umbral de corte
t1e1DEp<-subset(t1e1DE, t1e1DE$min.pval.base < umbral)
t1e2DEp<-subset(t1e2DE, t1e2DE$min.pval.base < umbral)
#obtenemos los miR.base

```



```

i11<-as.vector(unique(t1e1DEp$base))
d1a<-data.frame() #dataframe donde guardamos Los resultados del cruce
#vamos recorriendo Las 2 tablas por miR.base, buscando coincidencias de isomiR
for (i in 1:length(i11)){
t11<-subset(t1e1DEp,t1e1DEp$base==i11[i])
t12<-subset(t1e2DEp,t1e2DEp$base==i11[i])
nom1<-rownames(t11)
nom2<-rownames(t12)
#si se encuentra el miR.base en ambas tablas hacemos el cruce
if(length(nom1) > 0 && length(nom2) > 0){
f<-vector() #vector para filas
#si hay mas registros o Los mismos en tabla1 que en tabla 2 hacemos el cruce
#tomando La tabla 1 como referencia
if(length(nom1) >= length(nom2)){
n1<-vector() #vector donde se guardan Las coincidencias de isomiR
for (l in 1:length(nom1)){
#creamos La fila. Si el isomiR no se encuentra en La tabla 2 devuelve NA
f<-cbind(t11[nom1[l],3:5], t12[nom1[l],3:5])
d1a<-rbind(d1a,f) #añadimos La fila al dataframe
#guardamos el número de fila de La tabla 2 donde se ha encontrado el isomiR
#si no Lo encontramos ponemos un índice que no se encuentra en el dataset
n1<-c(n1,match(nom1[l],rownames(t12),nomatch=length(nom1)+1))
}
n1a<-subset(n1,n1<length(nom1)+1) #num. de coincidencias
#si hay menor número de coincidencias que de isomiRs en tabla 2
if(length(n1a)<length(nom2)){
#se selecciona un dataset donde no aparezcan Los índices matcheados -n1
t12n<-t12[-n1,]
#se recorre La tabla 2 para obtener Los datos de isomiRs no coincidentes
for (a in 1:length(rownames(t12n))){
#creamos La fila con Los resultados de La tabla 2 primero
#para poder registrar el nombre del isomiR como nombre de fila
#evidentemente, el resultado de tabla 1 será NA
f<-cbind(t12n[rownames(t12n)[a],3:5], t11[rownames(t12n)[a],3:5])
f<-cbind(f[4:6],f[1:3]) #Le damos La vuelta a La fila (primero tabla 1)
d1a<-rbind(d1a,f) #añadimos La fila al dataframe
}
}
}else{ #si hay más registros en tabla 2, La tomamos como referencia
n1<-vector()
for (l in 1:length(nom2)){
f<-cbind(t12[nom2[l],3:5], t11[nom2[l],3:5])
f<-cbind(f[4:6],f[1:3])
d1a<-rbind(d1a,f)
n1<-c(n1,match(nom2[l],rownames(t11),nomatch=length(nom2)+1))
}
n1a<-subset(n1,n1<length(nom2)+1)
if(length(n1a)<length(nom1)){
t11n<-t11[-n1,]
for (a in 1:length(rownames(t11n))){
f<-cbind(t11n[rownames(t11n)[a],3:5], t12[rownames(t11n)[a],3:5])
d1a<-rbind(d1a,f)
}
}
}
}
}
#obtenemos Los miR.base donde hay coincidencias entre Las 2 tablas
m11<-vector()

```

```

for (i in 1:nrow(d1a)){
m11<-c(m11, paste(unlist(strsplit(rownames(d1a)[i], "-"))[2],
gsub("[a-z]", "", unlist(strsplit(rownames(d1a)[i], "-"))[3]), sep="-"))
}
#guardamos Los resultados en un dataframe
miRDE11<-data.frame(m11, rownames(d1a), d1a[,1:3], d1a[,4:6])
colnames(miRDE11)<-
c("miR.base", "miR", "FC.exp1", "pfp.exp1", "pval.exp1", "FC.exp2", "pfp.exp2", "pval.exp2")

#cruzamos Las 2 tablas 2 de Los 2 experimentos
#miRNAs up-reg en La clase mutante
t2e1DEp<-subset(t2e1DE, t2e1DE$min.pval.base < umbral)
t2e2DEp<-subset(t2e2DE, t2e2DE$min.pval.base < umbral)
i22<-as.vector(unique(t2e1DEp$base))
d1a<-data.frame()
for (i in 1:length(i22)){
t21<-subset(t2e1DEp, t2e1DEp$base==i22[i])
t22<-subset(t2e2DEp, t2e2DEp$base==i22[i])
nom1<-rownames(t21)
nom2<-rownames(t22)
if(length(nom1) > 0 && length(nom2) > 0){
f<-vector()
if(length(nom1) >= length(nom2)){
n1<-vector()
for (l in 1:length(nom1)){
f<-cbind(t21[nom1[l], 3:5], t22[nom1[l], 3:5])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom1[l], rownames(t22), nomatch=length(nom1)+1))
#si nolo encontremos ponemos un índice que no se encuentra en el dataset
}
n1a<-subset(n1, n1<length(nom1)+1)
if(length(n1a)<length(nom2)){
#se selecciona un dataset donde no aparezcan Los indices matcheados -nl
t22n<-t22[-n1, ]
for (a in 1:length(rownames(t22n))){
f<-cbind(t22n[rownames(t22n)[a], 3:5], t21[rownames(t22n)[a], 3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
}
}
}else{
n1<-vector()
for (l in 1:length(nom2)){
f<-cbind(t22[nom2[l], 3:5], t21[nom2[l], 3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom2[l], rownames(t21), nomatch=length(nom2)+1))
}
n1a<-subset(n1, n1<length(nom2)+1)
if(length(n1a)<length(nom1)){
#se selecciona un dataset donde no aparezcan Los indices matcheados -nl
t21n<-t21[-n1, ]
for (a in 1:length(rownames(t21n))){
f<-cbind(t21n[rownames(t21n)[a], 3:5], t22[rownames(t21n)[a], 3:5])
d1a<-rbind(d1a, f)
}
}
}
}
}
}
}

```

```

}
#guardamos los resultados en un dataframe
m22<-vector()
for (i in 1:nrow(d1a)){
m22<-c(m22, paste(unlist(strsplit(rownames(d1a)[i], "-"))[2],
gsub("[a-z]", "", unlist(strsplit(rownames(d1a)[i], "-"))[3]), sep="-"))
}
miRDE22<-data.frame(m22, rownames(d1a), d1a[,1:3], d1a[,4:6])
colnames(miRDE22)<-
c("miR.base", "miR", "FC.exp1", "pfp.exp1", "pval.exp1", "FC.exp2", "pfp.exp2", "pval.exp2")

#cruzamos la tabla 1 del experimento 1 con la tabla 2 del experimento 2
#miRNAs down-reg en la clase mutante del experimento 1
#miRNAs up-reg en la clase mutante del experimento 2
t1e1DEp<-subset(t1e1DE, t1e1DE$min.pval.base < umbral)
t2e2DEp<-subset(t2e2DE, t2e2DE$min.pval.base < umbral)
i12<-as.vector(unique(t1e1DEp$base))
d1a<-data.frame()
for (i in 1:length(i12)){
t11<-subset(t1e1DEp, t1e1DEp$base==i12[i])
t22<-subset(t2e2DEp, t2e2DEp$base==i12[i])
nom1<-rownames(t11)
nom2<-rownames(t22)
if(length(nom1) > 0 && length(nom2) > 0){
f<-vector()
if(length(nom1) >= length(nom2)){
n1<-vector()
for (l in 1:length(nom1)){
f<-cbind(t11[nom1[l],3:5], t22[nom1[l],3:5])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom1[l], rownames(t22), nomatch=length(nom1)+1))
#si no lo encontramos ponemos un índice que no se encuentra en el dataset
}
n1a<-subset(n1, n1<length(nom1)+1)
if(length(n1a)<length(nom2)){
#se selecciona un dataset donde no aparezcan los índices matcheados -nl
t22n<-t22[-n1,]
for (a in 1:length(rownames(t22n))){
f<-cbind(t22n[rownames(t22n)[a],3:5], t11[rownames(t22n)[a],3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
}
}
}else{
n1<-vector()
for (l in 1:length(nom2)){
f<-cbind(t22[nom2[l],3:5], t11[nom2[l],3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom2[l], rownames(t11), nomatch=length(nom2)+1))
}
n1a<-subset(n1, n1<length(nom2)+1)
if(length(n1a)<length(nom1)){
#se selecciona un dataset donde no aparezcan los índices matcheados -nl
t11n<-t11[-n1,]
for (a in 1:length(rownames(t11n))){
f<-cbind(t11n[rownames(t11n)[a],3:5], t22[rownames(t11n)[a],3:5])
d1a<-rbind(d1a, f)
}
}
}
}

```

```

}
}
}
}
#guardamos Los resultados en un dataframe
m12<-vector()
for (i in 1:nrow(d1a)){
m12<-c(m12, paste(unlist(strsplit(rownames(d1a)[i],"-"))[2],
gsub("[a-z]", "", unlist(strsplit(rownames(d1a)[i],"-"))[3]), sep="-"))
}
miRDE12<-data.frame(m12, rownames(d1a), d1a[,1:3], d1a[,4:6])
colnames(miRDE12)<-
c("miR.base", "miR", "FC.exp1", "pfp.exp1", "pval.exp1", "FC.exp2", "pfp.exp2", "pval.exp2")

#cruzamos las tabla 2 del experimento 1 con la tabla 1 del experimento 2
#miRNAs up-reg en la clase mutante del experimento 1
#miRNAs down-reg en la clase mutante del experimento 2
t2e1DEp<-subset(t2e1DE, t2e1DE$min.pval.base < umbral)
t1e2DEp<-subset(t1e2DE, t1e2DE$min.pval.base < umbral)
i21<-as.vector(unique(t2e1DEp$base))
d1a<-data.frame()
for (i in 1:length(i21)){
t21<-subset(t2e1DEp, t2e1DEp$base==i21[i])
t12<-subset(t1e2DEp, t1e2DEp$base==i21[i])
nom1<-rownames(t21)
nom2<-rownames(t12)
if(length(nom1) > 0 && length(nom2) > 0){
f<-vector()
if(length(nom1) >= length(nom2)){
n1<-vector()
for (l in 1:length(nom1)){
f<-cbind(t21[nom1[l],3:5], t12[nom1[l],3:5])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom1[l], rownames(t12), nomatch=length(nom1)+1))
#si nolo encontremos ponemos un índice que no se encuentra en el dataset
}
n1a<-subset(n1, n1<length(nom1)+1)
if(length(n1a)<length(nom2)){
#se selecciona un dataset donde no aparezcan los índices matcheados -nl
t12n<-t12[-n1,]
for (a in 1:length(rownames(t12n))){
f<-cbind(t12n[rownames(t12n)[a],3:5], t21[rownames(t12n)[a],3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
}
}
}else{
n1<-vector()
for (l in 1:length(nom2)){
f<-cbind(t12[nom2[l],3:5], t21[nom2[l],3:5])
f<-cbind(f[4:6], f[1:3])
d1a<-rbind(d1a, f)
n1<-c(n1, match(nom2[l], rownames(t21), nomatch=length(nom2)+1))
}
n1a<-subset(n1, n1<length(nom2)+1)
if(length(n1a)<length(nom1)){
#se selecciona un dataset donde no aparezcan los índices matcheados -nl
t21n<-t21[-n1,]
for (a in 1:length(rownames(t21n))){

```

```

    f<-cbind(t21n[rownames(t21n)[a],3:5], t12[rownames(t21n)[a],3:5])
    d1a<-rbind(d1a,f)
  }
}
}
}
}
#guardamos los resultados en un dataframe
m21<-vector()
for (i in 1:nrow(d1a)){
m21<-c(m21, paste(unlist(strsplit(rownames(d1a)[i],"-"))[2],
gsub("[a-z]","",unlist(strsplit(rownames(d1a)[i],"-"))[3]),sep="-"))
}
miRDE21<-data.frame(m21,rownames(d1a),d1a[,1:3],d1a[,4:6])
colnames(miRDE21)<-
c("miR.base","miR","FC.exp1","pfp.exp1","pval.exp1","FC.exp2","pfp.exp2","pval.exp2")

#renombramos las tablas
miRDE11.025<-miRDE11
miRDE22.025<-miRDE22
miRDE12.025<-miRDE12
miRDE21.025<-miRDE21

```

ANEXO 3.

Tablas resultantes del proceso de cruce de miRNAs/isomiRs comunes en los dos experimentos.

miRDE11.05. Cruce de las 2 Tabla1 de los 2 experimentos. miRNAs down-reg en clase mutante.

miRDE11.05

##	miR.base	miR	FC.exp1	pdf.exp1	pval.exp1	
##	mmu-miR-12191-3p	miR-12191	mmu-miR-12191-3p	0.3148	0.6412	0.021210
##	mmu-miR-125b-1-3p	miR-125	mmu-miR-125b-1-3p	0.2097	0.5603	0.004186
##	mmu-miR-125b-2-3p	miR-125	mmu-miR-125b-2-3p	NA	NA	NA
##	mmu-miR-126b-3p	miR-126	mmu-miR-126b-3p	1.3560	0.7878	0.029430
##	mmu-miR-126a-5p	miR-126	mmu-miR-126a-5p	NA	NA	NA
##	mmu-miR-129b-3p	miR-129	mmu-miR-129b-3p	0.3803	0.5545	0.007694
##	mmu-miR-129-5p	miR-129	mmu-miR-129-5p	0.2928	0.6252	0.016680
##	mmu-miR-129-1-3p	miR-129	mmu-miR-129-1-3p	NA	NA	NA
##	mmu-miR-181c-5p	miR-181	mmu-miR-181c-5p	0.4207	0.8820	0.041420
##	mmu-miR-181b-5p	miR-181	mmu-miR-181b-5p	0.6182	0.8370	0.042880
##	mmu-miR-181c-3p	miR-181	mmu-miR-181c-3p	NA	NA	NA
##	mmu-miR-181d-5p	miR-181	mmu-miR-181d-5p	NA	NA	NA
##	mmu-miR-1843a-3p	miR-1843	mmu-miR-1843a-3p	0.2640	0.6497	0.011090
##	mmu-miR-1843b-3p	miR-1843	mmu-miR-1843b-3p	NA	NA	NA
##	mmu-miR-200a-5p	miR-200	mmu-miR-200a-5p	0.2106	0.4928	0.004208
##	mmu-miR-200b-5p	miR-200	mmu-miR-200b-5p	0.3449	0.8049	0.029210
##	mmu-miR-200b-3p	miR-200	mmu-miR-200b-3p	NA	NA	NA
##	mmu-miR-200a-3p	miR-200	mmu-miR-200a-3p	NA	NA	NA
##	mmu-miR-30f	miR-30	mmu-miR-30f	0.6198	0.5697	0.007296
##	mmu-miR-30b-3p	miR-30	mmu-miR-30b-3p	0.3465	0.7866	0.027700
##	mmu-miR-30e-5p	miR-30	mmu-miR-30e-5p	0.3585	0.8638	0.035030
##	mmu-miR-30a-3p	miR-30	mmu-miR-30a-3p	NA	NA	NA
##	mmu-miR-30d-3p	miR-30	mmu-miR-30d-3p	NA	NA	NA
##	mmu-miR-34c-3p	miR-34	mmu-miR-34c-3p	0.2585	0.6310	0.012790
##	mmu-miR-34b-3p	miR-34	mmu-miR-34b-3p	0.4838	0.8359	0.047280
##	mmu-miR-34a-5p	miR-34	mmu-miR-34a-5p	NA	NA	NA
##	mmu-miR-871-3p	miR-871	mmu-miR-871-3p	0.2700	0.6374	0.012250
##		FC.exp2	pdf.exp2	pval.exp2		
##	mmu-miR-12191-3p	0.7733	6.732e-01	1.236e-02		
##	mmu-miR-125b-1-3p	NA	NA	NA		
##	mmu-miR-125b-2-3p	0.9207	7.997e-01	3.264e-02		
##	mmu-miR-126b-3p	NA	NA	NA		
##	mmu-miR-126a-5p	0.7243	6.268e-01	8.954e-03		
##	mmu-miR-129b-3p	NA	NA	NA		
##	mmu-miR-129-5p	NA	NA	NA		
##	mmu-miR-129-1-3p	0.7747	8.135e-01	3.237e-02		
##	mmu-miR-181c-5p	NA	NA	NA		
##	mmu-miR-181b-5p	NA	NA	NA		
##	mmu-miR-181c-3p	0.7497	6.499e-01	1.260e-02		
##	mmu-miR-181d-5p	0.7722	8.451e-01	3.708e-02		
##	mmu-miR-1843a-3p	NA	NA	NA		
##	mmu-miR-1843b-3p	1.0310	8.537e-01	4.355e-02		
##	mmu-miR-200a-5p	NA	NA	NA		
##	mmu-miR-200b-5p	NA	NA	NA		
##	mmu-miR-200b-3p	0.3072	1.447e-07	1.476e-10		
##	mmu-miR-200a-3p	0.4043	3.279e-06	6.691e-09		
##	mmu-miR-30f	NA	NA	NA		
##	mmu-miR-30b-3p	NA	NA	NA		
##	mmu-miR-30e-5p	NA	NA	NA		
##	mmu-miR-30a-3p	0.8649	8.259e-01	2.865e-02		
##	mmu-miR-30d-3p	0.7924	8.391e-01	4.024e-02		
##	mmu-miR-34c-3p	NA	NA	NA		
##	mmu-miR-34b-3p	NA	NA	NA		

```
## mmu-miR-34a-5p 0.7396 7.086e-01 1.808e-02
## mmu-miR-871-3p 0.7447 6.135e-01 6.887e-03
```

miRDE22.05. Cruce de las 2 Tabla2 de los 2 experimentos. miRNAs up-reg en clase mutante.

miRDE22.05

```
##          miR.base          miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-148a-5p miR-148 mmu-miR-148a-5p 5.897 0.3704 0.005929 NA
## mmu-miR-148a-3p miR-148 mmu-miR-148a-3p NA NA NA 1.0600
## mmu-miR-34b-3p miR-34 mmu-miR-34b-3p NA NA NA 1.3630
## mmu-miR-34b-5p miR-34 mmu-miR-34b-5p 3.808 0.6443 0.026820 1.3090
## mmu-miR-34c-5p miR-34 mmu-miR-34c-5p NA NA NA 1.2610
## mmu-miR-423-3p miR-423 mmu-miR-423-3p 4.223 0.5900 0.018890 1.1640
## mmu-miR-431-5p miR-431 mmu-miR-431-5p 5.043 0.4181 0.008032 NA
## mmu-miR-431-3p miR-431 mmu-miR-431-3p NA NA NA 1.0940
## mmu-miR-770-5p miR-770 mmu-miR-770-5p 3.238 0.7443 0.047660 1.2000
## mmu-miR-92a-1-5p miR-92 mmu-miR-92a-1-5p 5.024 0.4517 0.009159 NA
## mmu-miR-92a-3p miR-92 mmu-miR-92a-3p 3.426 0.7526 0.040160 0.9537
##          pfp.exp2 pval.exp2
## mmu-miR-148a-5p NA NA
## mmu-miR-148a-3p 0.2025 0.002480
## mmu-miR-34b-3p 0.2970 0.005455
## mmu-miR-34b-5p 0.3765 0.009220
## mmu-miR-34c-5p 0.6733 0.032290
## mmu-miR-423-3p 0.5676 0.020270
## mmu-miR-431-5p NA NA
## mmu-miR-431-3p 0.5437 0.016640
## mmu-miR-770-5p 0.7847 0.046440
## mmu-miR-92a-1-5p NA NA
## mmu-miR-92a-3p 0.5762 0.021160
```

miRDE12.05. Cruce de Tabla1 experimento 1 con Tabla2 experimento 2.

miRDE12.05

```
##          miR.base          miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-100-5p miR-100 mmu-miR-100-5p 0.3968 0.8073 0.047390 1.0130
## mmu-miR-125b-1-3p miR-125 mmu-miR-125b-1-3p 0.2097 0.5603 0.004186 NA
## mmu-miR-125b-2-3p miR-125 mmu-miR-125b-2-3p NA NA NA 0.9207
## mmu-miR-1298-3p miR-1298 mmu-miR-1298-3p 0.1608 0.7659 0.001635 1.7830
## mmu-miR-1298-5p miR-1298 mmu-miR-1298-5p 0.1829 0.6294 0.002687 2.0940
## mmu-miR-1843a-3p miR-1843 mmu-miR-1843a-3p 0.2640 0.6497 0.011090 NA
## mmu-miR-1843b-3p miR-1843 mmu-miR-1843b-3p NA NA NA 1.0310
## mmu-miR-1912-5p miR-1912 mmu-miR-1912-5p 0.3022 0.6724 0.020090 NA
## mmu-miR-1912-3p miR-1912 mmu-miR-1912-3p 0.4306 0.8631 0.046980 1.1730
## mmu-miR-212-5p miR-212 mmu-miR-212-5p 0.2379 0.5850 0.006867 1.3370
## mmu-miR-212-3p miR-212 mmu-miR-212-3p NA NA NA 1.2780
## mmu-miR-3102-3p miR-3102 mmu-miR-3102-3p 0.3835 0.8500 0.047170 1.2370
## mmu-miR-34b-3p miR-34 mmu-miR-34b-3p 0.4838 0.8359 0.047280 1.3630
## mmu-miR-34b-5p miR-34 mmu-miR-34b-5p NA NA NA 1.3090
## mmu-miR-34c-5p miR-34 mmu-miR-34c-5p NA NA NA 1.2610
## mmu-miR-34c-3p miR-34 mmu-miR-34c-3p 0.2585 0.6310 0.012790 NA
## mmu-miR-448-3p miR-448 mmu-miR-448-3p 0.4072 0.7961 0.049280 2.3530
## mmu-miR-455-3p miR-455 mmu-miR-455-3p 0.3755 0.8564 0.044790 1.5070
##          pfp.exp2 pval.exp2
## mmu-miR-100-5p 6.847e-01 3.423e-02
## mmu-miR-125b-1-3p NA NA
## mmu-miR-125b-2-3p 7.922e-01 4.365e-02
## mmu-miR-1298-3p 5.107e-04 2.084e-06
## mmu-miR-1298-5p 5.471e-04 2.791e-06
## mmu-miR-1843a-3p NA NA
## mmu-miR-1843b-3p 6.747e-01 3.511e-02
## mmu-miR-1912-5p NA NA
## mmu-miR-1912-3p 7.566e-01 4.787e-02
## mmu-miR-212-5p 4.697e-01 1.342e-02
```



```

## mmu-miR-212-3p 5.906e-01 2.411e-02
## mmu-miR-3102-3p 7.682e-01 4.703e-02
## mmu-miR-34b-3p 2.970e-01 5.455e-03
## mmu-miR-34b-5p 3.765e-01 9.220e-03
## mmu-miR-34c-5p 6.733e-01 3.229e-02
## mmu-miR-34c-3p NA NA
## mmu-miR-448-3p 1.515e-05 1.546e-08
## mmu-miR-455-3p 5.167e-02 4.218e-04

```

miRDE21.05, Cruce de Tabla2 experimento 1 con Tabla1 experimento 2.

miRDE21.05

```

##
## miR.base miR FC.exp1 pfp.exp1 pval.exp1 FC.exp2
## mmu-miR-126b-3p miR-126 mmu-miR-126b-3p 1.3560 0.7494 0.043190 NA
## mmu-miR-126a-3p miR-126 mmu-miR-126a-3p 3.3020 0.7292 0.043580 NA
## mmu-miR-126a-5p miR-126 mmu-miR-126a-5p NA NA NA 0.7243
## mmu-miR-129-2-3p miR-129 mmu-miR-129-2-3p 7.5930 0.2985 0.001593 NA
## mmu-miR-129b-5p miR-129 mmu-miR-129b-5p 5.8010 0.3390 0.003979 NA
## mmu-miR-129-1-3p miR-129 mmu-miR-129-1-3p NA NA NA 0.7747
## mmu-miR-148a-5p miR-148 mmu-miR-148a-5p 5.8970 0.3704 0.005929 NA
## mmu-miR-148a-3p miR-148 mmu-miR-148a-3p NA NA NA 1.0600
## mmu-miR-30a-3p miR-30 mmu-miR-30a-3p NA NA NA 0.8649
## mmu-miR-30d-3p miR-30 mmu-miR-30d-3p NA NA NA 0.7924
## mmu-miR-30f miR-30 mmu-miR-30f 0.6198 0.7463 0.041420 NA
## mmu-miR-33-5p miR-33 mmu-miR-33-5p 4.2940 0.6029 0.018660 NA
## mmu-miR-33-3p miR-33 mmu-miR-33-3p NA NA NA 0.8050
## mmu-miR-34b-5p miR-34 mmu-miR-34b-5p 3.8080 0.6443 0.026820 NA
## mmu-miR-34a-5p miR-34 mmu-miR-34a-5p NA NA NA 0.7396
## mmu-miR-92a-1-5p miR-92 mmu-miR-92a-1-5p 5.0240 0.4517 0.009159 NA
## mmu-miR-92a-3p miR-92 mmu-miR-92a-3p 3.4260 0.7526 0.040160 0.9537
## mmu-miR-92b-5p miR-92 mmu-miR-92b-5p NA NA NA 0.8468
##
## pfp.exp2 pval.exp2
## mmu-miR-126b-3p NA NA
## mmu-miR-126a-3p NA NA
## mmu-miR-126a-5p 0.6268 0.008954
## mmu-miR-129-2-3p NA NA
## mmu-miR-129b-5p NA NA
## mmu-miR-129-1-3p 0.8135 0.032370
## mmu-miR-148a-5p NA NA
## mmu-miR-148a-3p 0.8507 0.041670
## mmu-miR-30a-3p 0.8259 0.028650
## mmu-miR-30d-3p 0.8391 0.040240
## mmu-miR-30f NA NA
## mmu-miR-33-5p NA NA
## mmu-miR-33-3p 0.8014 0.049880
## mmu-miR-34b-5p NA NA
## mmu-miR-34a-5p 0.7086 0.018080
## mmu-miR-92a-1-5p NA NA
## mmu-miR-92a-3p 0.7487 0.022160
## mmu-miR-92b-5p 0.8366 0.037560

```

REFERENCIAS

- [1] Lai, Mong-Hsun Tsai. "Common applications of next-generation sequencing technologies in genomic research." (2013). <http://tcr.amegroups.com/article/view/962/html>
- [2] SRA handbook. <https://www.ncbi.nlm.nih.gov/books/NBK242621/>
- [3] Amsel, Daniel, Andreas Vilcinskas, and André Billion. "Evaluation of high-throughput isomir identification tools: illuminating the early isomirome of *tribolium castaneum*." *Bmc Bioinformatics* 18.1 (2017): 359. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5543545/>
- [4] Pantano, L. "miRNA annotation". https://seqcluster.readthedocs.io/mirna_annotation.html
- [5] González, I. "Statistical analysis of RNA-Seq data". Universidad de Toulouse (2014)
- [6] Chamorro Poyo, Cristina. "Análisis de datos de RNA-Seq empleando diferentes paquetes desarrollados dentro del proyecto Bioconductor para estudios de expresión génica diferencial." TFM Máster de Bioinformática y Bioestadística UOC (2019). <http://openaccess.uoc.edu/webapps/o2/handle/10609/96466>
- [7] Pantano, Lorena. "Characterization of miRNA and isomiR molecules." <https://www.bioconductor.org/packages/devel/bioc/vignettes/isomiRs/inst/doc/isomiRs.html>
- [8] Cheng, Tian-Lin, et al. "MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex." *Developmental cell* 28.5 (2014): 547-560. [https://www.cell.com/developmental-cell/fulltext/S1534-5807\(14\)00073-2](https://www.cell.com/developmental-cell/fulltext/S1534-5807(14)00073-2)
- [9] Polyester. Documento de Bioconductor: <https://www.bioconductor.org/packages/release/bioc/vignettes/polyester/inst/doc/polyester.html>
- [10] Tarazona, S., Furió-Tarí, P., Ferrer, A., & Conesa, A. (2013). NOISeq: Differential Expression in RNA-seq. <https://www.bioconductor.org/packages/release/bioc/vignettes/NOISeq/inst/doc/NOISeq.pdf>
- [11] Hong, F.(2010). Bioconductor RankProd Package Vignette. <http://bioconductor.statistik.tu-dortmund.de/packages/3.6/bioc/vignettes/RankProd/inst/doc/RankProd.pdf>
- [12] Weise, Stefan C., et al. "FOXG1 regulates PRKAR2B transcriptionally and posttranscriptionally via miR200 in the adult hippocampus." *Molecular neurobiology* 56.7 (2019): 5188-5201. <https://link.springer.com/article/10.1007%2Fs12035-018-1444-7>
- [13] Zhang, Y., Zang, Q., Xu, B., Zheng, W., Ban, R., Zhang, H., ... & Shi, Q. (2016). IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics*, 32(13), 2069-2071. <https://www.ncbi.nlm.nih.gov/pubmed/27153728>
- [14] Peterson, S. M., Thompson, J. A., Ufkin, M. L., Sathyanarayana, P., Liaw, L., & Congdon, C. B. (2014). Common features of microRNA target prediction tools. *Frontiers in genetics*, 5, 23. <https://www.ncbi.nlm.nih.gov/pubmed/24600468>
- [15] Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *elife*, 4, e05005. <https://elifesciences.org/articles/05005>
- [16] Csardi G (2013). *targetscaan.Hs.eg.db: TargetScan miRNA target predictions for human*. R package version 0.6.1.

- [17] Doron Betel, Manda Wilson, Aaron Gabow, Debora S. Marks, Chris Sander, The microRNA.org resource: targets and expression, *Nucleic Acids Research*, Volume 36, Issue suppl_1, 1 January 2008, Pages D149–D153, <https://doi.org/10.1093/nar/gkm995>
- [18] Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG., "DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows.", *Nucleic Acids Res.* 2013 Jul;41(Web Server issue):W169-73. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3692048/>
- [19] Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG., "Functional microRNA targets in protein coding sequences.", *Bioinformatics.* , 2012 Jan 27. https://www.researchgate.net/publication/221785142_Functional_microRNA_targets_in_protein_coding_sequences
- [20] Dimitra Karagkouni, Maria D. Paraskevopoulou, Serafeim Chatzopoulos, Ioannis S. Vlachos, Spyros Tastsoglou, Ilias Kanellos, Dimitris Papadimitriou, Ioannis Kavakiotis, Sofia Maniou, Giorgos Skoufos, Thanasis Vergoulis, Theodore Dalamagas, Artemis G. Hatzigeorgiou; DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions, *Nucleic Acids Research*, gkx1141, <https://doi.org/10.1093/nar/gkx1141>
- [21] Yu G, Wang L, Han Y, He Q (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters." *OMICS: A Journal of Integrative Biology*, **16**(5), 284-287. doi: . [10.1089/omi.2011.0118](https://doi.org/10.1089/omi.2011.0118)
- [22] Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *InterJournal*, Complex Systems, 1695. <http://igraph.org>.
- [23] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov; 13(11):2498-504. <https://www.ncbi.nlm.nih.gov/pubmed/14597658>
- [24] Ulgen E, Ozisik O, Sezerman OU. 2019. pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. *Front. Genet.* <https://doi.org/10.3389/fgene.2019.00858>