

# 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**

**Marc Maceira Duch**

10 de Junio de 2020



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

**FICHA DEL TRABAJO FINAL**

<b>Título del trabajo:</b>	<i>Análisis comparativo de smallRNAs en síndrome de Rett mediado por MECP2 y síndrome atípico de Rett mediado por FOXC1</i>
<b>Nombre del autor:</b>	<i>Fernando Hernández Correa</i>
<b>Nombre del consultor/a:</b>	<i>Guerau Fernández Isern</i>
<b>Nombre del PRA:</b>	<i>Javier Luis Cánovas Izquierdo Marc Maceira Duch</i>
<b>Fecha de entrega (mm/aaaa):</b>	10/06/20
<b>Titulación::</b>	<i>Máster universitario en bioinformática y bioestadística UOC-UB</i>
<b>Área del Trabajo Final:</b>	<i>Bioinformática clínica</i>
<b>Idioma del trabajo:</b>	<i>Castellano</i>
<b>Palabras clave</b>	<i>MECP2, FOXC1, smallRNA.</i>
<b>Resumen del Trabajo (máximo 250 palabras):</b> <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i>	
<p>Los micro RNAs son un tipo de pequeños RNAs no codificantes que constituyen un mecanismo de control genético.</p> <p>En este trabajo se han analizado estudios de secuenciación de miRNAs en la investigación de dos síndromes similares. El síndrome típico de Rett producido por mutación del gen MECP2 y una variante producida por mutación del gen FOXC1. Ambos estudio se han centrado en el análisis de miRNAs. Los resultados iniciales no muestran miRNAs desregulados comunes. No se ha realizado un análisis completo de isomiRs, isoformas variantes de miRNAs, que podría ofrecer información adicional a la hora de buscar elementos de regulación genética comunes en los dos síndromes.</p> <p>El trabajo parte de los datos de secuenciación disponibles públicamente en la plataforma SRA y realiza un análisis completo de miRNAs/isomiRs diferencialmente expresados. Se han obtenido predicciones de genes target y se ha hecho un análisis de enriquecimiento GO y KEGG de rutas metabólicas que puedan estar desreguladas.</p> <p>Para ello, se ha desarrollado un pipeline completo que pueda ser exportable al análisis de otros experimentos similares.</p> <p>Se han encontrado miRNAs DE comunes en ambos experimentos, entre los que destacan miR200a y miR200b como infra-regulados y miR92a, miR148a y miR431 como sobre-regulados en ambas condiciones.</p> <p>El análisis de genes target de estos miRNAs ha arrojado conclusiones que pueden ser de interés en el estudio de ambos síndromes. Como algunas</p>	

metabólicas posiblemente desreguladas con elementos comunes y diferenciales.

El presente trabajo tiene un alcance puramente bioinformático y no se han realizado comprobaciones experimentales sobre los resultados.

**Abstract (in English, 250 words or less):**

miRNAs are a type of non-coding smallRNAs that make an important procedure on genetic control.

In this work, a full analysis of two smallRNA-seq experiments has been done. These two experiments consist on two similar disorders: the Rett syndrome as a result of MECP2 mutation and a atypical Rett syndrome as a result of FOXP1 mutation. Both experiments are focused in miRNA analysis. Initial results are not showing common miRNAs differentially expressed in both conditions, but a complete isomiRs analysis hasn't been done. IsomiRs are variants of the miRNA reference sequence, which could be significant in order to find common genetic elements in both disorders.

This work is based on smallRNA-seq datasets available on SRA platform and makes a full analysis of miRNAs/isomiRs DE. It has been done a prediction of targets and the analysis of GO terms, as well as a KEGG pathways enrichment over the results.

With this purpose, a full pipeline with bioinformatic tools has been performed, which can be used to analyze similar experiments.

Common miRNAs DE have been found, like miR200a and miR200b (down-regulated), as well as miR92a, miR148a and miR431 (up-regulated).

The analysis of targets of these miRNAs has shown interesting conclusions in the study of both syndromes. Like some destabilized pathways with common and differential elements.

This work has been made only in a bioinformatic way, without any experimental checking over the results.

# Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido.....	3
1.4 Planificación del Trabajo.....	3
1.5 Breve resumen de productos obtenidos.....	5
1.6 Breve descripción de los otros capítulos de la memoria.....	5
2. Panorámica del estudio.....	6
2.1 Antecedentes.....	6
2.2 Problemática abordada.....	8
2.3 Síndrome de Rett típico (tRTT) mediado por mutación del gen MECP2.....	10
2.4 Síndrome de Rett atípico (atRTT) mediado por mutación del gen FOXP1.....	13
2.5 Planteamiento inicial.....	15
3. Desarrollo del pipeline.....	16
3.1 Datos de partida.....	16
3.2 Descarga de datos en crudo y adaptación al formato adecuado.....	16
3.3 Comprobación de la calidad de los reads.....	17
3.4 Anotación de miRNAs.....	19
3.5 Control de calidad de datos y normalización.....	23
3.6 Análisis de expresión diferencial de miRNAs/isomiRs.....	26
3.7 Proceso de cruce de información de miRNAs/isomiRs DE en ambos experimentos. .....	34
3.8 Análisis detallado de isomiRs DE.....	35
3.9 Clasificación supervisada.....	36
3.10 Integración gen-miRNA.....	37
3.11 Enriquecimiento GO.....	41
3.12 Enriquecimiento KEGG.....	49
3.13 Relación de herramientas y programas empleados en el pipeline.....	52
3.14 Diagrama resumen del pipeline.....	54
4. Análisis de resultados.....	56
4.1 Análisis de resultados del experimento 1 (ratones Mecp2+/-).....	56
4.2 Análisis de resultados del experimento 2 (ratones Foxg1+/-).....	59
4.3 Análisis de resultados del cruce de los 2 experimentos.....	62
5. Conclusiones.....	66
6. Discusión.....	81
7. Glosario.....	85
8. Bibliografía.....	89

## Lista de figuras

Figura 1. Ejemplos de miRNAs en metazoos.....	6
Figura 2. Biogénesis de miRNA en mamíferos.....	7
Figura 3. Tipos de isomiR.....	8
Figura 4. Diagnóstico de variantes de RTT.....	10
Figura 5. Estructura del gen MECP2 y de la proteína MeCP2.....	11
Figura 6. Función de MeCP2.....	11
Figura 7. Interacción de MeCP2 con DGCR8.....	12
Figura 8. Defosforilación de MeCP2 S80 dependiente de calcio.....	13
Figura 9. Papel de FOXG1 en el mantenimiento de progenitor celular y en la supresión de diferenciación.....	14
Figura 10. Identificación de targets comunes en sobreexpresión de miR200 y deficiencia en Foxg1.....	14
Figura 11. En el citoplasma, FOXG1 coopera con DDX5.....	15
Figura 12. Resumen del informe de calidad SRR1504855_fastqc.html.....	18
Figura 13. Resumen del informe de calidad SRR6061841_fastqc.html.....	18
Figura 14. Detalle del informe de calidad SRR1504855_fastqc.html.....	19
Figura 15. Resumen del informe de calidad SRR1504855_clean_fastqc.html.....	20
Figura 16. Boxplot y diagrama de densidad de las muestras del experimento 1.....	23
Figura 17. Plot MA de las muestras del experimento 1. ....	24
Figura 18. Boxplot, diagrama de densidad y digramas MA antes de la normalización de las muestras del experimento 2.....	24
Figura 19. Boxplot, diagrama de densidad y digramas MA después de la normalización de las muestras del experimento 2.....	25
Figura 20. Representación MDS y dendograma de las muestras del experimento 2.....	25
Figura 21. Diagrama MDS de las muestras del experimento 2 una vez eliminadas las muestras mutante1 y control1.....	26
Figura 22. Gráfico isoPlot de diferenciación de isomiRs del experimento 1.....	26
Figura 23. Diagrama de Venn comparando los resultados del experimento 1.....	27
Figura 24. Gráfico isoPlot de diferenciación de isomiRs del experimento 2.....	31
Figura 25. Volcano plot de miRNAs DE en el experimento 2 .....	32
Figura 26. Digramas de Venn comparando los resultados del experimento 2.....	32
Figura 27. Representación PLS-DA.....	37
Figura 28. Interacción entre el miRNA y el mRNA target.....	37
Figura 29. Estructura secundaria de RNA en forma de bucle.....	38
Figura 30. Diagrama de categorías GO en el experimento 1 .....	41
Figura 31. Mapa de enriquecimiento de categorías GO del experimento 1 .....	42
Figura 32. Mapa de red bipartita de miRNAs DE y genes target del experimento 1.....	43
Figura 33. Mapa de enriquecimiento de categorías GO del experimento 2 .....	45
Figura 34. Mapa de red bipartita de miRNAs DE y genes target dell grupo 1 del experimento 2 .....	46
Figura 35. Mapa de red bipartita del grupo 2 del experimento 2 .....	47
Figura 36. Mapa de red bipartita del grupo 3 del experimento 2 .....	47
Figura 37. Mapa de enriquecimiento de categorías GO en el cruce de los 2 experimentos.....	48
Figura 38. Mapa de red bipartita de los miRNAs y genes target seleccionados en el grupo 1 del cruce de los 2 experimentos.....	50
Figura 39. Mapa de red bipartita del grupo 2 del cruce de los dos experimentos.....	51
Figura 40a. Primera parte del pipeline.....	54
Figura 40b. Segunda parte del pipeline,.....	55
Figura 41. Mapa de red bipartita de miRNAs y genes target seleccionados como más significativos para el experimento 1.....	57

Figura 42. Gráfico de enriquecimiento pathfindR en el experimento 1.....	58
Figura 43. Gráfico de enriquecimiento pathfindR en el experimento 2.....	61
Figura 44. Selección de la cadena de miRNA madura.....	63
Figura 45. Gráfico de enriquecimiento pathfindR en el cruce de ambos experimentos.....	64
Figura 46. Esquema del pathway KEGG mmu04010, MAPK signaling pathway.....	70
Figura 47. Esquema del pathway KEGG mmu04024, cAMP signaling pathway.....	71
Figura 48. Activación de la vía de las MAPK por la insulina .....	74
Figura 49. Activación de la vía de las PI3K por la insulina.....	75
Figura 50. Verificación experimental de targets con PAR-CLIP.....	81

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

El síndrome de Rett (*RTT*) es un desorden progresivo de neurodesarrollo que afecta a mujeres. Hay dos formas de *RTT*: típico (*tRTT*) y atípico (*atRTT*). Alrededor del 70-90% de los casos son *tRTT*, que proceden de mutaciones del gen *MECP2*. Una variante de *atRTT* es causada por mutaciones del gen *FOXG1*. Las mutaciones de pérdida de función en el gen *MECP2* son una causa primaria del síndrome de Rett. Las duplicaciones de loci que contienen *MECP2* pueden conducir a desórdenes del espectro del autismo en humanos. La dosis de proteína *MECP2* es crítica en el correcto desarrollo y función del sistema nervioso central (*CNS*). El tipo de mutación del gen *MECP2* puede determinar distintos fenotipos, pero la correlación genotipo-fenotipo en la caracterización de la enfermedad no está clara. Incluso la inactivación de uno de los cromosomas X (*XCI*) puede influir también en el fenotipo manifestado. En líneas generales, el síndrome de Rett se caracteriza por microcefalia, epilepsia y falta de habilidad para sentarse, caminar y hablar [1].

Las mutaciones en *FOXG1* causan una forma menos frecuente de síndrome de Rett atípica. *FOXG1* es un factor de transcripción clave para el desarrollo del cerebro anterior. Aunque muchas de las características del síndrome *atRTT* provocado por mutaciones del gen *FOXG1* se solapan con las del síndrome *tRTT*, algunas son diferenciales, como la disquinesia (exceso de movimientos o movimientos anormales e involuntarios) o la ausencia de arritmias respiratorias o regresión en el desarrollo, típicos del síndrome de Rett [2].

Tomando como base dos estudios independientes ya realizados, sobre análisis de *smallRNAs* en estos dos síndromes neurológicos similares producidos por alteraciones genéticas diferentes [3 y 4] y con datos de *smallRNA-seq* disponibles públicamente en *SRA* (*Sequence Read Archive*); el presente trabajo pretende hacer un análisis exhaustivo de estos datos para tratar de relacionar ambas casuísticas y, puesto que se trata de fenotipos similares, buscar las relaciones comunes de desregulación de *smallRNA* en los dos síndromes estudiados y sus puntos diferenciales.

En los dos experimentos de referencia, realizados en ratones, se analizan los *miRNAs* desregulados y sus implicaciones en dos situaciones diferentes. Por un lado, se identifican *miRNAs* desregulados en ratones con mutación en el gen *Mecp2*, que provoca síndrome de Rett típico (*tRTT*) en humanos [3]. Y por otro, se identifican *miRNAs* desregulados en ratones con mutación en el gen *Foxg1*, que provoca una clase de síndrome de Rett atípico (*atRTT*) [4]. En ambos estudios no se encuentran *miRNAs* desregulados comunes. Aunque las dos proteínas, *MeCP2* y *FOXG1* se asocian con el conjunto procesador de *miRNA*, parecen afectar a la maduración de diferentes *miRNAs*.

Entre los *miRNAs* reprimidos por *MeCP2* se encuentran miR-134, miR-383, miR-382 y miR-182. Entre los *miRNAs* regulados por *FOXG1* se encuentran los de la familia miR200.



Es habitual que se hagan análisis de *smallRNAs* desregulados en diferentes alteraciones genéticas sin tener en cuenta su distribución en sus distintas isoformas o *isomiRs*. En *locus miRNAs* específicos, distintos *isomiRs* pueden mostrar patrones de desregularización de expresión diferentes, con relaciones funcionales concretas, como la co-regulación de *mRNAs target*. Se piensa que estos *isomiRs* no son fruto de mutaciones al azar, sino que se producen deliberadamente como productos separados que tienen diferentes papeles en los procesos celulares.

Por eso se considera que un análisis más exhaustivo de distribución de los diferentes *isomiRs*, partiendo de los datos en crudo de ambos experimentos, puede indicar alguna relación de desregulación genética entre ambos síndromes. Es posible que en estos 2 estudios hayan pasado desapercibidos niveles desregulados de *isomiRs* en ambas anomalías, que puedan darnos idea de algún *pathway* común afectado por ambos genes, *MECP2* y *FOXG1*.

## 1.2 Objetivos del Trabajo

El objetivo principal del trabajo será hacer un análisis detallado sobre los datos en crudo disponibles de la caracterización de *isomiRs* de ambas anomalías genéticas, para encontrar elementos diferenciales y puntos en común entre ambos síndromes.

Para ello, se detallará un *pipeline*, utilizando herramientas de análisis de *miRNA/isomiR*, que nos permita cruzar datos de ambos orígenes de una manera sistematizada para comparar las dos situaciones y hallar vías genéticas y rutas metabólicas comunes.

Objetivos generales.

- G1. Establecer un *pipeline* completo de todo el proceso de análisis que pueda ser factible de utilizar en comparación de experimentos similares sobre *sRNAs*.
- G2. Realizar un análisis completo de *sRNAs (miRNA/isomiRs)* diferencialmente expresados en ambos experimentos y buscar posibles relaciones genéticas, funcionales o metabólicas entre los *sRNAs* comunes.

Objetivos específicos.

El objetivo general G1 de establecimiento de un *pipeline* se desglosa en los siguientes objetivos específicos:

- S1.1. Definir de forma detallada los procesos implicados en el análisis.
- S1.2. Definir las herramientas *software* empleadas.
- S1.3. En caso de que sea necesario, desarrollar pequeños *scripts* o subprocesos de conexión entre datos o herramientas dentro de *pipeline*.

El objetivo general G2 de análisis de *sRNAs* se desglosa en los siguientes objetivos específicos:

- S2.1. Obtener un conjunto de *sRNAs (miRNA/isomiRs)* diferencialmente expresados para ambos experimentos y completarlo con anotaciones de bases de datos de *sRNAs* que ayuden a identificarlos.
- S2.2. Cruzar las informaciones de los dos conjuntos de *sRNAs* para encontrar elementos coincidentes en ambos experimentos y buscar genes que se vean afectados por la desregulación de los *sRNAs* identificados.
- S2.3. Enriquecer la lista de *sRNAs (miRNA/isomiRs)* comunes anteriores y de genes regulados por ellos con categorías funcionales o rutas metabólicas

que ayuden a encontrar puntos en común entre los procesos biológicos afectados por ambos síndromes.

### 1.3 Enfoque y método seguido

La metodología en los dos estudios de referencia es la siguiente:

- El estudio '*Quantitative analysis of wild type and MeCP2-/- 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-/-*

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*.

El enfoque del presente trabajo es puramente bioinformático, se parte de datos *sRNA-seq* de dos experimentos publicados de libre acceso y se quiere cruzar ambas informaciones para obtener relaciones comunes entre ambos experimentos, uno sobre el síndrome típico de Rett (*tRTT*) mediado por mutación del gen *MECP2* y el otro sobre un síndrome atípico de Rett (*atRTT*) mediado por mutación del gen *FOXG1*. Las conclusiones que se obtengan de dicho análisis se contrastarán bioestadísticamente, pero no serán sometidas a verificación experimental.

Se partirá de los datos *sRNA-seq* en crudo, no de los ficheros de cuentas de *miRNA*, disponibles en ambos experimentos, puesto que se quiere hacer un análisis más detallado de *sRNA* para buscar aquellos que se encuentren desregulados en ambas situaciones y que puedan haber pasado desapercibidos por un simple análisis de *miRNA*.

A partir de los nuevos ficheros de cuentas de *miRNA/isomiR* obtenidos, se utilizará una metodología de análisis típica de experimentos de *sRNA-seq* adaptada a las circunstancias concretas del caso de estudio, pero que a su vez sirva como punto de partida para la creación de un *pipeline* genérico, con los ajustes de procesamiento y herramientas *software* necesarios, que pueda ser utilizado en el contraste de experimentos de *sRNA-seq*.

Se trabajará con herramientas *software* libres, sin que sea necesaria la adquisición de licencias de ningún producto.

### 1.4 Planificación del Trabajo

Se hace la siguiente distribución de tareas en base a los objetivos definidos en el punto 1.2. En la columna T se marca el tiempo previsto para cada tarea en horas; en total suman 300 horas.

TAREAS G1	Obj.	T horas	TAREAS G2	Obj.	T horas
T1.1. Definir y ajustar la herramienta software utilizada para descarga y adaptación de datos (p.e. fastq-dump).	S1.2	6	T2.1. Descargar datos en crudo y adaptarlos al formato adecuado.	S2.1	6
T1.2. Definir y ajustar la herramienta software de análisis de isomiR adecuada (p.e. isomir-SEA, isomiRID, Miraligner)		24	T2.2. Obtener los reads miRNA/isomiR válidos para cada experimento.		30
T1.3. Definir los procesos implicados en el análisis DE con R/Bioconductor.	S1.1	6	T2.3. Realizar control de calidad	S2.1	6
			T2.4. Valorar y realizar normalización		6
			T2.5. Identificar miRNA/isomiR diferencialmente expresados		24
T1.4. Definir el proceso de anotación con la BdD adecuada (p.e. MiRBase).		12	T2.6. Anotar los resultados		12
T1.5. Desarrollar subproceso de cruce de información.	S1.3	12	T2.7. Cruzar ambas informaciones de miRNA/isomiR DE		24
T1.6. Definir y ajustar el software de búsqueda de secuencias génicas target (p.e. TargetScan, miRanda, miRDB, Microcosm)	S1.2	18	T2.8. Buscar posibles genes objetivo (target) regulados por estos miRNA/isomiR	S2.2	30
T1.7. Definir el proceso de enriquecimiento de resultados con R/Bioconductor	S1.1	12	T2.9. Enriquecer la lista de posibles genes regulados con GO.	S2.3	24
			T2.10 Enriquecer la lista de posibles genes regulados con KEGG.		24
			T2.11. Obtener categorías funcionales y rutas metabólicas comunes.		24

Se distribuyen las fases principales del proyecto según el siguiente calendario.

Del 17/3 al 10/4	Del 11/4 al 30/4	Del 1/5 al 16/5
Análisis <i>sRNA-seq</i> de los dos experimentos con identificación de <i>miRNA/isomiR</i> diferencialmente expresados y anotación de resultados.	Cruce de ambos análisis con identificación de <i>miRNA/isomiR</i> DE comunes y búsqueda de posibles genes regulados por estos <i>miRNA/isomiR</i>	Enriquecimiento con categorías funcionales y rutas metabólicas de los genes objetivo ( <i>targets</i> ) identificados en la fase anterior.

La temporización propuesta para el desarrollo del proyecto tiene una primera fase (PEC 2) del 17 de marzo al 22 de abril y una segunda fase (PEC 3) del 23 de abril al 18 de mayo.

Según la planificación propuesta, en la entrega de la PEC 2 se habrá realizado el análisis *sRNA-seq* de los dos experimentos y se habrá avanzado bastante en el cruce de resultados, por lo que se estará en buena situación para decidir si se continúa el proyecto con las pautas previstas o es necesario cambiar el enfoque del mismo. En la entrega de la PEC 3 ya se habrán finalizado todas las tareas de análisis y durante la PEC 4 (Cierre de memoria; del 19 de mayo al 10 de junio) se redactará la memoria y se completarán las conclusiones.

### 1.5 Breve resumen de productos obtenidos

Este proyecto no tiene un producto propiamente dicho. Se ha obtenido como resultado 2 documentos detallados, uno de carácter técnico informático, con el detalle de procesos y *software* empleado, que constituya un *pipeline* válido en análisis comparativos de *sRNA-seq* y otro de carácter biológico, con las conclusiones y contrastes estadísticos de los análisis realizados. Los dos documentos, que se anexan junto a la memoria, son:

*fhernandezcor\_TFM202006\_1\_Documento técnico*  
*fhernandezcor\_TFM202006\_2\_Documento biológico*

### 1.6 Breve descripción de los otros capítulos de la memoria

Capítulo 2. Panorámica del estudio. Introduce la problemática a estudiar y presenta un panorama del conocimiento de partida.

Capítulo 3. Desarrollo del *pipeline*. Descripción de las tareas que conforman el *pipeline* para análisis de datos de experimentos de *smallRNA-seq*.

Capítulo 4. Análisis de resultados. Analiza los resultados obtenidos por el proceso descrito en el capítulo anterior desde el punto de vista de la problemática indicada en el capítulo 2.

Capítulo 5. Conclusiones. Elabora las conclusiones obtenidas tras el análisis de los resultados.

Capítulo 6. Discusión. Reflexión y análisis sobre los objetivos, planificación y resultados del trabajo. Descripción de líneas de trabajo futuras que no se han podido abordar.

## 2. Panorámica del estudio.

### 2.1 Antecedentes.

Los microRNA (*miRNA*), cadenas cortas de *RNA* no codificador que se doblan sobre sí mismas en forma de horquilla (*hairpin*), constituyen una herramienta de control genético. Son reguladores post-transcripcionales de la expresión génica. La mayoría proceden de regiones intergénicas, que a su vez son complementarias con secuencias de *RNA* mensajero (*mRNA*) de genes endógenos del organismo. Como consecuencia de dicha complementariedad, los *miRNA* se emparejan con sus *mRNA* diana (*targets*) e interfieren con su expresión.

Los *miRNA* son una clase de pequeños *RNAs* (*smallRNA* o *sRNA*), dentro de los cuales se cuentan también los pequeños *RNA* de interferencia (*siRNA*). Estructuralmente, los *miRNA* son muy similares a los *siRNA*, pero presentan ciertas peculiaridades distintivas. Los *miRNA* tienen un tamaño aproximado de 22 nucleótidos y se originan a partir de un precursor de *RNA* con estructura de horquilla (*hairpin*) de complementariedad imperfecta, que contiene la secuencia de *miRNA* en uno de sus brazos (Fig. 1).

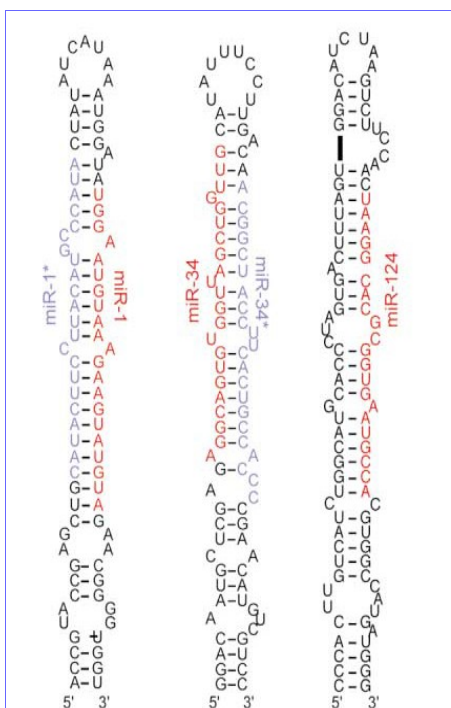


Figura 1. Ejemplos de *miRNAs* en metazoos. Las secuencias corresponden a los precursores (*hairpin*) *mir-1*, *mir-34* y *mir-124* de *C. elegans*, pero se han encontrado secuencias homólogas en mamíferos. (Lagos-Quintana et al., 2001,2002; Lau et al., 2001; Lee and Ambros, 2001) [5].

El *miRNA* primario se procesa con dos enzimas de tipo *Rnasa III*, *Drosha* en el núcleo celular y *Dicer* en el citoplasma; el segmento resultante da origen a la molécula de *miRNA*, que se asocia con un complejo ribonucleoproteico, *RISC* (*RNA-induced silencing complex*) (Fig. 2).

El *RISC* se une al *mRNA* diana (*target*) de dos formas distintas, o bien en la región 3' *UTR* (*untranslated region*), si la complementariedad entre ambas cadenas no es muy precisa, o bien en la secuencia codificadora del *mRNA* vía emparejamiento de bases, si las regiones son perfectamente complementarias [5].

La unión del *RISC* al *mRNA* inhibe su procesamiento posterior, bloqueando su traducción o promoviendo la degradación.

La biogénesis de *miRNAs* puede implicar la producción de isoformas, conocidas como *isomiRs*. Se piensa que se producen deliberadamente como productos separados que definen papeles diferentes en la célula y que no representan errores de secuenciación. Los *isomiRs* pueden estar extendidos o truncados en cualquier extremo comparados con el *miRNA* maduro, presumiblemente por una escisión imperfecta por *Drosha* o *Dicer*.

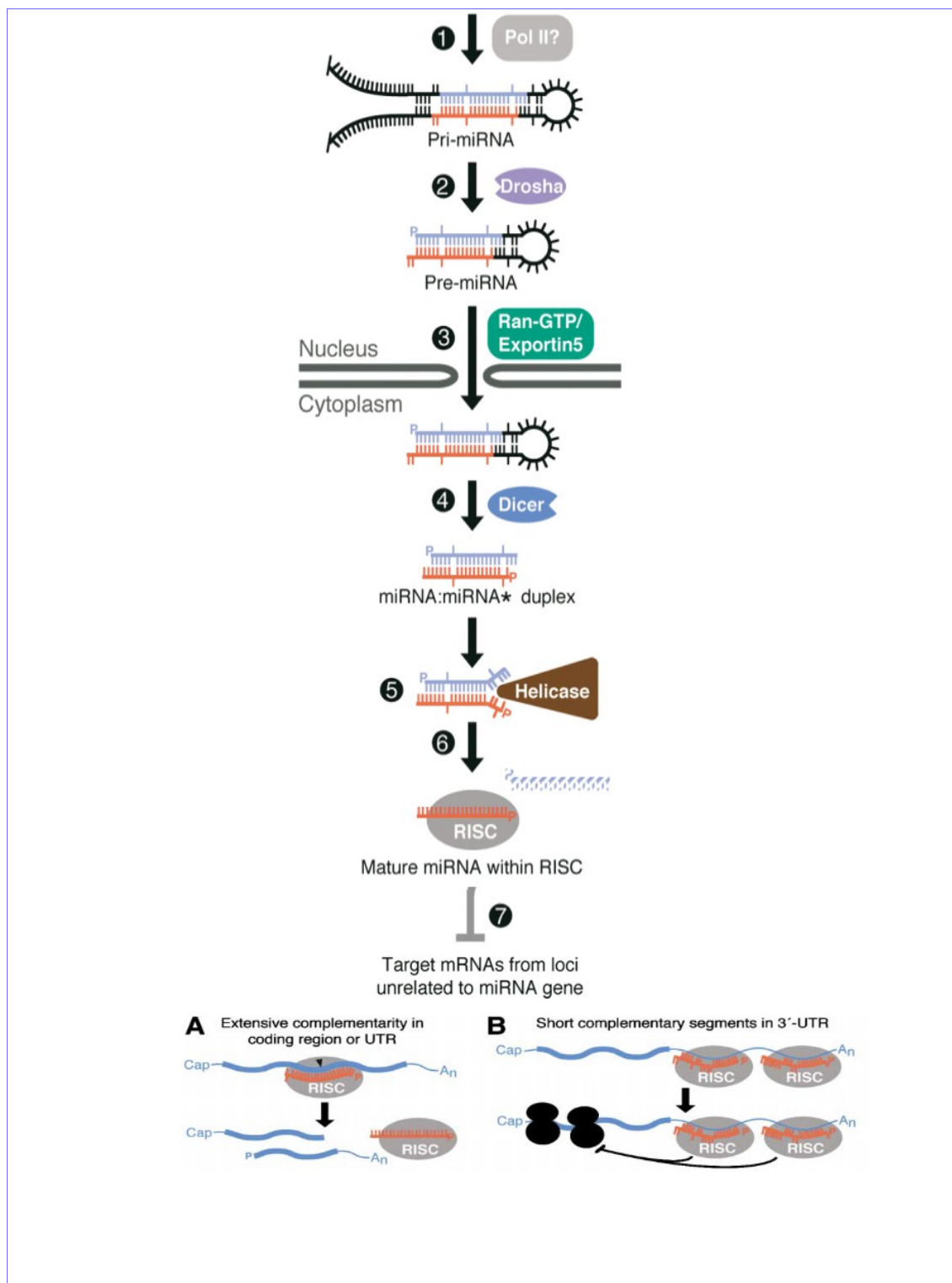


Figura 2. Biogénesis de *miRNA* en mamíferos. 1) División del *pri-miRNA*; se libera un bucle (*stem loop*) de ~60-70 nuc. 2) La endonucleasa *RnaseIII Drosha* corta las hebras del bucle en lugares próximos al *pre-miRNA*. 3) El *pre-miRNA* se transporta al citoplasma. 4) El corte nuclear de *Drosha* define uno de los extremos del *miRNA* maduro, el otro extremo se procesa en el citoplasma por la enzima *Dicer*. 5) El duplex *miRNA* se divide en dos hebras por la acción de la helicasa. 6) Una de las hebras, el *miRNA* maduro, se une al complejo *RISC*. 7) El complejo *RISC* se puede unir al *mRNA target* de 2 modos: A) a la región codificadora, con una amplia complementariedad o B) a la región no codificadora 3' UTR, si la complementariedad es menor. [5]

Se pueden considerar las siguientes variantes de *isomiRs* sobre el *miRNA* maduro (Fig. 3):

- *5' template addition*. Adición de nucleótidos en el extremo 5' siguiendo la plantilla del *pre-miRNA*.
- *5' template deletion*. Supresión de nucleótidos en el extremo 5'.
- *3' template addition*. Adición de nucleótidos en el extremo 3' siguiendo la plantilla del *pre-miRNA*.
- *3' template deletion*. Supresión de nucleótidos en el extremo 3'.
- *3' non-template addition*. Adición de nucleótidos en el extremo 3' sin seguir la plantilla del *pre-miRNA*.
- *Seed SNP*. Sustitución de 1 nucleótido en el germen (*seed*) del *miRNA* maduro.
- *Tail SNP*. Sustitución de 1 nucleótido en la cola (*tail*) del *miRNA* maduro.

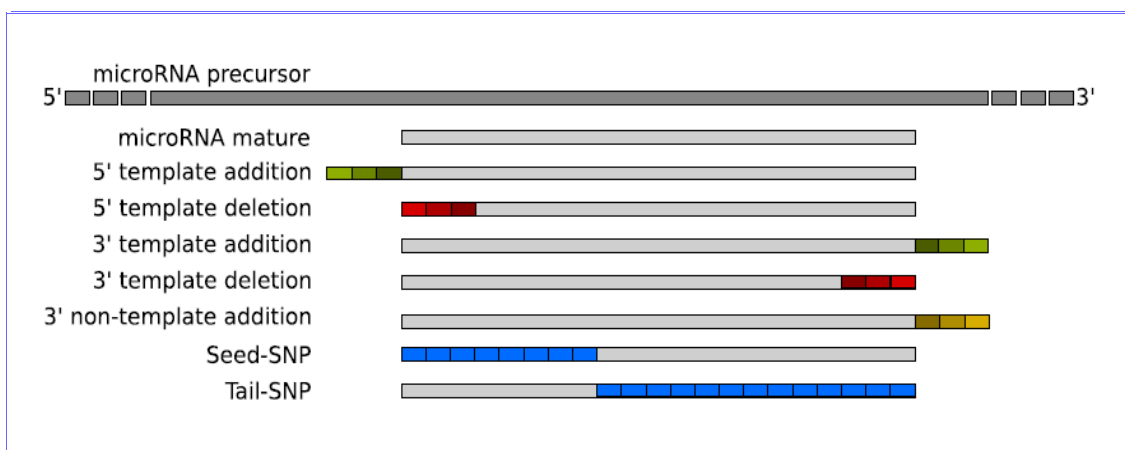


Figura 3. Tipos de *isomiR*. Las cajas verdes representan adición de nucleótidos, las amarillas adiciones *non-template*, las rojas deleciones y las azules *SNPs* (*single nucleotide polymorphism*) [6].

Estudios recientes indican que los *isomiRs* 5' experimentan un desplazamiento de la región germen (*seed*) que cambia el conjunto de *mRNAs* diana (*target*) comparados con el *mRNA* original. El conjunto de *mRNAs target* también puede variar por cambios simples de nucleótido en la región *seed* (*SNP*; *single nucleotide polymorphism*). Los *miRNAs* maduros también pueden adquirir una cola 3' de polinucleótidos *non-template* [6].

Muchos estudios se focalizan en estas múltiples variantes de *miRNA* debido al papel versátil que juegan los *isomiRs* en la regulación post-transcripcional de *mRNA*. Para estudiar una expresión detallada de niveles de *miRNA/isomiRs* es necesario el análisis de interacciones *miRNA-miRNA* o *isomiR-isomiR*, combinado con homología y/o agrupamiento (*clustered*) de *miRNAs*. Como ejemplo, se ha comprobado que la expresión de muchos *isomiRs* está diferenciada por género en humanos [7].

## 2.2 Problemática abordada.

Muchos estudios se focalizan únicamente en el análisis de la expresión diferenciada de *miRNAs*. Se ha comprobado la conveniencia de extender este análisis a la expresión de *miRNAs/isomiRs* para entender en más profundidad la complejidad del problema. Dentro de esta problemática, se ha escogido el análisis de *isomiRs* diferenciales en dos síndromes similares.

El síndrome de Rett (*RTT*) es un desorden progresivo de neurodesarrollo que afecta principalmente a mujeres. Hay dos formas de *RTT*: típico (*tRTT*) y atípico

(*atRTT*). Alrededor del 70-90% de los casos son *tRTT*, que proceden de mutaciones del gen *MECP2*.

Una variante de *atRTT* es causada por mutaciones del gen *FOXG1*, aunque se han descrito otras variantes del síndrome que se mencionarán a continuación.

*RTT* se caracteriza por un desarrollo aparentemente normal en los primeros 6 a 18 meses de vida, seguido de un periodo de regresión en el lenguaje y en la habilidades motoras. Los pacientes pierden el uso normal de las manos y lo sustituyen por movimientos repetidos y estereotipados. Al nacer tienen una medida de circunferencia de la cabeza normal, pero muestran un crecimiento craneal post-natal bajo, dando lugar a microcefalia. También son características propias de pacientes *RTT* la privación social, disfunción en la comunicación, pérdida de frases adquiridas y discapacidad cognitiva. La discapacidad locomotora también es común. Incluye otras características adicionales del espectro autista, como ataques de pánico, disfunciones respiratorias, bruxismo, disfunción en patrones de sueño, cifosis o escoliosis, bajo crecimiento somático, pies y/o manos hipotrofiados y fríos [1].

La dificultad de un diagnóstico temprano del síndrome de Rett radica en que el desarrollo de las niñas con este síndrome es normal durante los primeros meses de vida. A partir del sexto mes pueden aparecer signos sutiles, como pérdida de interés, menor contacto visual y dificultad para sentarse o gatear, que pueden pasar desapercibidos. El diagnóstico normalmente se produce cuando se aprecia un bajo crecimiento de la cabeza (microcefalia) o se produce pérdida de habilidades o de etapas fundamentales de desarrollo, en un deterioro rápido que puede tener lugar entre los 1 a 4 años de edad.

La variante de *RTT* llamada congénita está provocada por mutaciones en el gen *FOXG1*. Las características del síndrome *FOXG1* incluyen deficiencia moderada del crecimiento post-natal, microcefalia post-natal severa, retardo mental severo con ausencia o mínimo desarrollo del lenguaje, deficiente interacción social, discinesia (movimientos anormales e involuntarios, sobre todo de la muscularidad orofacial) mezclada con características de atetosis (trastorno neurológico caracterizado por movimientos involuntarios, lentos y ondulatorios, predominantemente en pies, manos o cabeza), corea y distonía, epilepsia, patrones de sueño alterado, irritabilidad especialmente en la infancia, excesivos episodios de llanto, aspiración recurrente y frecuente reflujo gastrointestinal [2].

El número de mutaciones registradas no es suficientemente alto para tener una correlación clara genotipo-fenotipo. Pero parece que el fenotipo característico se da con mutaciones que causan la rotura de la proteína y en algunas mutaciones de pérdida de sentido (*missense*) o de otro tipo. Grandes deleciones en la región 14q12, que incluye *FOXG1* y otros genes, pueden causar un fenotipo más severo. Sin embargo, algunas mutaciones *missense* o de posibles regiones reguladoras parecen tener fenotipos menos severos.

Este fenotipo se interpreta como parecido al síndrome de Rett, excepto porque no hay una etapa temprana de desarrollo normal. Aunque muchas características se solapan en ambos síndromes, otras como la verdadera discinesia y anomalías en las imágenes cerebrales (giros simplificados, circunvoluciones aplanadas y reducidas, reducción del volumen de materia blanca en lóbulos frontales, hipogénesis del cuerpo calloso) son propias de alteraciones de *FOXG1*, mientras que otras como presencia de regresión y arritmias respiratorias son propias del síndrome de Rett. Por estos motivos, muchos autores prefieren hablar de síndrome *FOXG1* con un reconocimiento clínico específico.

El fenotipo del síndrome *FOXG1*, más allá de su similitud con el síndrome de Rett típico (mediado por *MECP2*-) ya mencionada, se solapa con otros



síndromes complejos de desarrollo como el síndrome de Angelman (asociado a *UBE3A*-), el de Pitt-Hopkins (asociado a *TCF4*-) y síndromes asociados a *ARX*-, *CDKL5*-, *CNTNAP2*-, *NRXN1*-, *SLC9A6*- y síndrome materno de cromosoma 15 isodiccéntrico. La mayoría están asociados a un solo gen y muestran un elevado espectro de deficiencias cognitivas, sociales, afectivas, de autonomía y motoras, además de epilepsia. Se ha comenzado a emplear el término encefalopatías del desarrollo para referirse a este conjunto de síndromes [2]. El síndrome *CDKL5*, asociado al cromosoma X, se caracteriza por aparición de crisis temprana. También se ha diagnosticado una variante ligada al gen *MECP2* que se caracteriza por el mantenimiento del habla (variante *PSD* o Zappella) (Fig. 4).

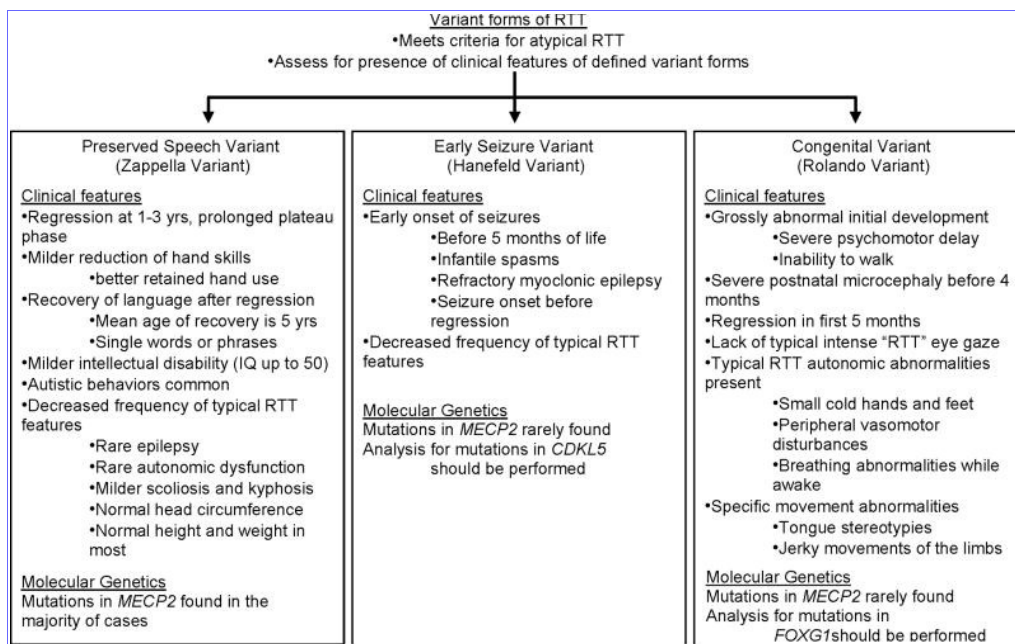


Figura 4. Diagnóstico de variantes de RTT [8].

### 2.3 Síndrome de Rett típico (tRTT) mediado por mutación del gen *MECP2*.

El gen *MECP2* se sitúa en el cromosoma X, de modo que su mutación afecta principalmente a mujeres. Sus mutaciones en hombres suelen ser letales y raramente sobreviven; los casos en que sobreviven suelen estar relacionados con diversas formas de mutaciones leves.

Al estar en el cromosoma X, el gen *MECP2* está sometido al proceso de silenciamiento de uno de los cromosomas X (*XCI*, *X chromosome inactivation*) que ocurre en mujeres. El gen tiene 4 exones y genera 2 isoformas de la proteína *MECP2* (*methyl-CpG binding protein*), que están implicadas en el silenciamiento transcripcional de la cromatina. La forma *MECP2* es la más abundante y en su transcripción se emplean los 4 exones; la forma *MECP2B* emplea sólo los exones 1,3 y 4. El cuarto exón es el más largo [1].

La proteína *MECP2* tiene 4 dominios funcionales (Fig. 5):

- *MBD* (*methyl-CpG-binding domain*). Responsable de la metilación de dinucleótidos CpG en regiones promotoras.
- *TRD* (*transcriptional repression domain*).
- Localizador de la señal nuclear. Responsable del transporte de *MECP2* al núcleo.
- Segmento C-terminal. Facilita la unión al nucleosoma.

Las interacciones entre el complejo de represión transcripcional y la cromatina lleva a la de-acetilación de las histonas H3 y H4, lo que da como resultado una compactación de la cromatina, haciéndola inaccesible a los componentes de la maquinaria transcripcional.

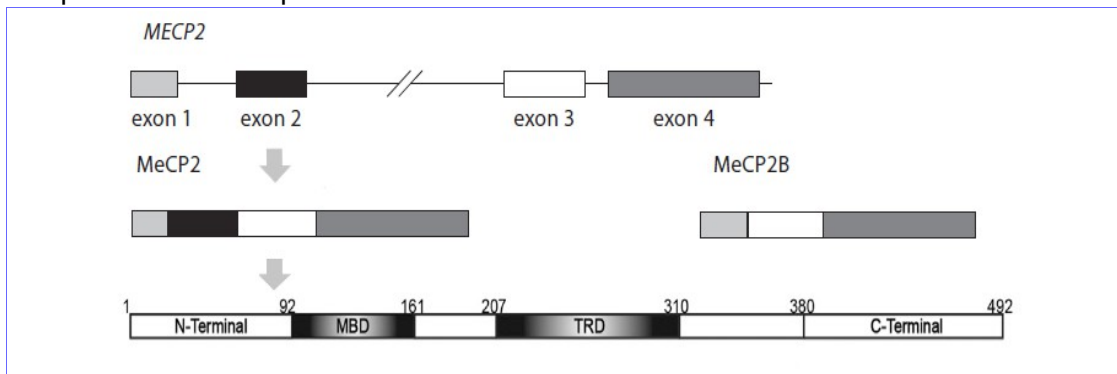


Figura 5. Estructura del gen *MECP2* y de la proteína MeCP2 [1, 3].

Las mutaciones del gen *MECP2* pueden causar decrecimiento del tamaño del cuerpo neuronal y árboles dendríticos y provocan niveles anormales en la expresión de genes implicados en la génesis sináptica. Esto sugiere que *MECP2* juega un papel importante en la maduración neuronal.

El 67% de todas las mutaciones de *MECP2* están causadas por transiciones C>T en 8 dinucleótidos CpG, localizados en los exones 3 y 4. Un conjunto de mutaciones de pérdida de sentido (*missense*) ocurre en el extremo 5' de *MBD*, la mayoría en *TRD* e implican aminoácidos conservados evolutivamente en dominios funcionales de la proteína. En la región distal de *MBD* aparecen mutaciones sin sentido (*nonsense*), de desplazamiento de trama (*frameshift*) y corte (*splicing*) que dan como resultado la terminación prematura de la proteína. En el dominio C-terminal ocurren grandes deleciones de nucleótidos.

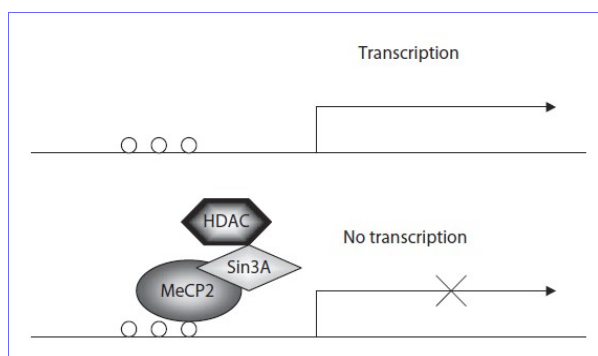


Figura 6. Función de *MECP2*. *MECP2* actúa como represor transcripcional. Se une a CpG metilado y a través del corepresor Sin3A recluta a la histona de-acetilasa (HDAC), que provoca la de-acetilación de las histonas H3 y H4. El resultado es una compactación de la cromatina, a la que la maquinaria transcripcional no puede acceder [1].

No hay una evidencia clara en la correlación genotipo-fenotipo. Las mismas mutaciones pueden provocar fenotipos diferentes, lo que sugiere que no sólo el tipo de mutación influye en la severidad del fenotipo; la inactivación de uno de los cromosomas X (*XCI*) puede ser un factor adicional. En algunos pacientes *RTT* no se han encontrado mutaciones en *MECP2*, lo que supone que puede haber otros *locus* responsables de *RTT*. Se han relacionado con este síndrome los genes *CDKL5* (también en el cromosoma X), que provoca una variante con manifestación más temprana de los síntomas y *Netrin G1*.

Sin embargo, se han encontrado ciertas correlaciones entre el tipo de mutación del gen *MECP2* y el fenotipo *RTT*. Las mutaciones que provocan formas más leves de la enfermedad corresponden a mutaciones de cambio de sentido (*missense*) en un sólo nucleótido, mutaciones de pérdida de sentido (*nonsense*) o deleciones en una región de alta concentración de mutaciones (*hotspot*) entre

las pb 1030-1207. Y las que provocan formas más severas de la enfermedad son las de terminación prematura y las que conducen a una rotura completa o parcial de la región codificante de la señal de localización nuclear.

*MECP2* silencia distintos genes [1] (Fig. 6):

- *LINE1*. Retrotransposón.
- *BDNF* (*brain-derived neurotrophic factor*). Implicado en la plasticidad neuronal, aprendizaje y memoria. La patología *RTT* puede ser debida a la desregulación de este proceso.
- *DLX5*. Implicado en la producción de enzimas que sintetizan *GABA*, un importante neurotransmisor. En pacientes *RTT* podría suponer cambios de actividad neuronal.
- *Crh*. Producción de corticotropina.
- *FXYP domain*. Controla la excitabilidad celular por modulación de la actividad  $\text{Na}^+/\text{K}^+/\text{ATPasa}$ .

En el trabajo de referencia [3] se ha evaluado la posibilidad alternativa de que *MECP2* controle reguladores post-transcripcionales, como *miRNA*, que actúan como supresores de muchas proteínas importantes en proliferación celular, desarrollo y tumorigénesis. Para ello se ha trabajado con ratones con una mutación que anula el gen *Mecp2* (*null*) y ratones transgénicos con *Mecp2* sobreexpresado (*TG*), comparados con ratones normales (*wt*). En mutaciones de pérdida de función del gen *MECP2*, como se ha indicado anteriormente, se manifiesta *RTT* y con la ganancia de función del gen se desarrolla en humanos un desorden severo de neurodesarrollo del espectro autista.

En dicho experimento, se encuentra que *MeCP2* se une directamente a la proteína *DGCR8* (región crítica 8 del síndrome de DiGeorge), un componente fundamental en la maquinaria de procesamiento del *miRNA* nuclear (Fig. 7). Esta interacción depende del extremo C-terminal de *MeCP2*, pero no del dominio *MDB* (*methyl-DNA-binding*). *DGCR8* contiene 2 dominios de unión a *RNA* (*RBD1* y *RBD2*) y una cola C-terminal que es necesaria para la unión con *Drosha* en el procesamiento de *miRNA*. Estos dominios *RNA-binding* pueden estar implicados en la unión con *MeCP2* y están cerca del extremo C-terminal, por lo que se produciría una interferencia con la asociación *DGCR8-RNA* y también con la interacción *DGCR8-Drosha* y de esta manera suprime el procesamiento de *miRNA* mediado por *DGCR8/Drosha*.

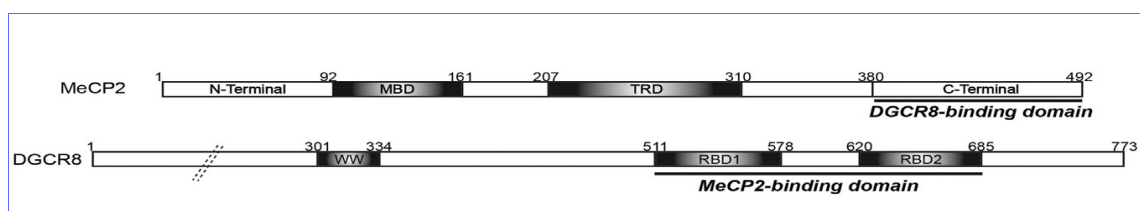


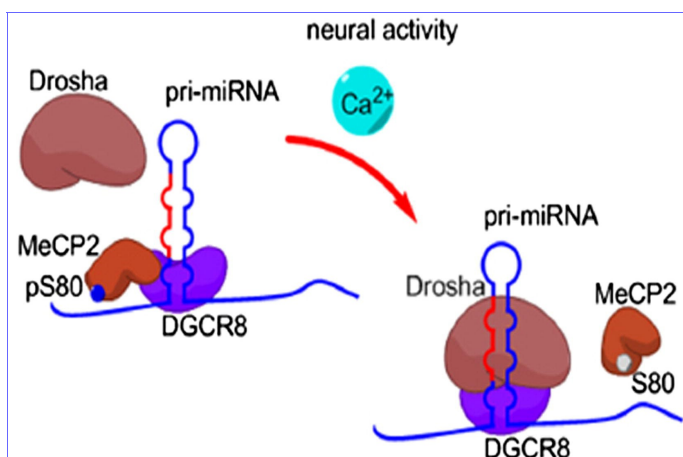
Figura 7. Interacción de *MeCP2* con *DGCR8*. La unión de ambas moléculas depende del extremo C-terminal de *MeCP2* y de los dominios de unión a *RNA* (*RBD1* y *RBD2*) de *DGCR8* [3].

Se propone también que la interacción entre *MeCP2* y *DGCR8* está regulada por el estado de fosforilación del residuo Ser80 de *MeCP2* (Fig. 8). La proteína *MeCP2* en estado fosforilado se une a *DGCR8* durante el estado de reposo en neuronas. Cuando se activa la señal de calcio neuronal, la defosforilación de *MeCP2* Ser80 lleva a una liberación de *DGCR8* y permite que la actividad de procesamiento dependiente de *miRNA* tenga lugar.

De modo que con baja expresión de *MeCP2* habrá una sobre-expresión de *miRNA*, que a su vez actuará como represor post-transcripcional de ciertos genes diana (*target*). En ratones *Mecp2-null* se ha estudiado sobre todo la

acción del *miRNA* sobre-expresado mir-134, un *miRNA* crítico para el desarrollo neuronal.

Figura 8. Defosforilación de MeCP2 S80 dependiente de calcio. MeCP2 con Ser80 en estado fosforilado (pS80) se une a DGCR8, impidiendo que interactúe con *Drosha* en el proceso de maduración de *miRNA*. La defosforilación de S80 mediada por la señal de calcio propicia que MeCP2 libere a DGCR8 y ésta pueda unirse a *Drosha* y llevar a cabo el proceso de maduración de *miRNA* [3].



En ratones con *Mecp2* sobre-expresado (*TG*), el nivel de miR-134 primario no se ve alterado, pero los niveles de miR-134 precursor y maduro están significativamente por debajo de lo normal, lo que sugiere que *MeCP2* regula la expresión de estos *miRNAs* a través de un mecanismo independiente de la transcripción, como el visto anteriormente.

*Mecp2* puede regular los niveles de expresión de genes diana (*target*) por ajuste fino de niveles de *miRNAs*, como miR-134, que tiene por objetivo 3 reguladores críticos del desarrollo y la plasticidad neuronal: *CREB*, *LIMK1* y *Pumilio2* [3].

Por tanto, *Mecp2* actúa como represor transcripcional por unión al *DNA* metilado y reclutamiento del complejo histona de-acetilasa (*HDAC*) y también como regulador post-transcripcional por supresión del procesamiento nuclear de *miRNAs*.

#### 2.4 Síndrome de Rett atípico (atRTT) mediado por mutación del gen *FOXG1*.

Una serie de estudios han implicado distintas deleciones o alteraciones de secuencia del gen *FOXG1* como responsables de la aparición de un desorden de desarrollo mental, descrito como una variable congénita del síndrome de Rett. Al situarse en el cromosoma 14, su alteración afecta tanto a mujeres como a hombres; siendo la mutación homocigótica normalmente letal al nacer y la heterocigótica causante de distintos grados de severidad del síndrome, dependiendo de la naturaleza de dicha mutación.

*FOXG1* (*Forkhead Box G1*) es un gen de la familia *forkhead*; factores de transcripción que juegan un papel importante en la regulación de genes que intervienen en el crecimiento, proliferación y diferenciación celular. Concretamente, *FOXG1* es un importante represor transcripcional que regula el desarrollo del telencéfalo desde el embrión hasta el estado adulto por múltiples mecanismos y cuya alteración en la expresión afecta a la formación y funcionalidad del córtex frontal en mamíferos (Fig. 9).

La regulación del ciclo celular por *FOXG1* está mediada por su unión al complejo *FOXO/SMAD*, que antagoniza los efectos de la señalización *TGFβ* (*Transforming Growth Factor β*), con lo que reduce la expresión de *Cdkn1a/p21* (*cyclin-dependent kinase inhibitor 1A*). Bajos niveles de *CDKN1A* impiden la salida del ciclo celular de las células madre neuronales y promueven la expansión del *pool* de células madre.

La actividad de *FOXG1* está mediada en parte por proteínas *TLE* (*Transducin-like Enhancer split*), un correpresor que se une a *FOXG1* para formar un

complejo represor transcripcional que conduce a una sobre-expresión de células madre en el estado de proliferación.

El mantenimiento del ciclo celular por *FOXG1* permite una proliferación celular prolongada y una expansión de las células madre neuronales.

Las familias génicas *FOX* y *SOX* (*SRY-box transcription factor*) son reguladores críticos de la diferenciación de células madre neuronales. *FOXG1* afecta a reguladores clave de metilación de *DNA* que facilitan la diferenciación de células madre neuronales. En enfermedades como el glioblastoma, el enriquecimiento *FOX/SOX2* abre el perfil de la cromatina y desregula la expresión de reguladores epigenéticos como *Foxo3*, *Plk1*, *Mycn*, *Dnmt1*, *Dnmt3b* y *Tet3* [9].

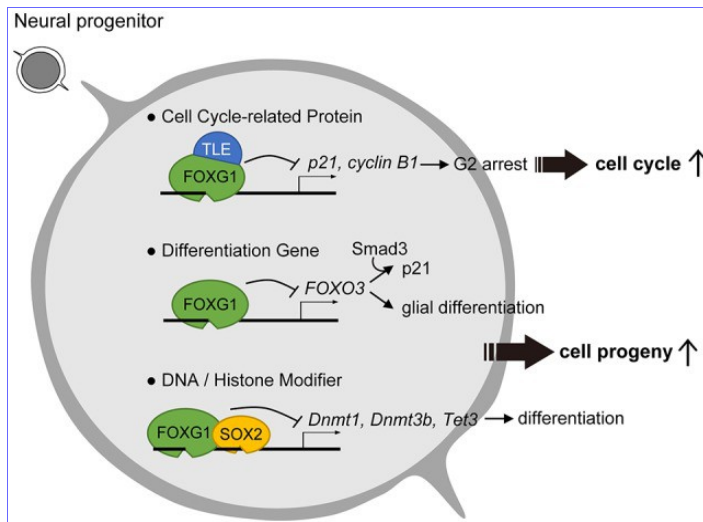


Figura 9. Papel de *FOXG1* en el mantenimiento de progenitor celular y en la supresión de diferenciación. Para impedir la salida del ciclo celular, *FOXG1* coopera con TLE para inhibir la expresión de proteínas clave en el ciclo celular. *FOXG1* también inhibe la diferenciación génica y la modificación *DNA*/Histona, impidiendo la diferenciación neuronal [9].

Se han detectado también niveles de *FOXG1* en el citoplasma. Distintos niveles de *FOXG1* en la mitocondria de células neurales provocan o bien su división, o bien su fusión.

En el estudio de referencia [4], se ha realizado *sRNA-seq* en células del hipocampo de ratones heterocigotos *Foxg1<sup>cre/+</sup>* y se ha comparado con ratones normales (*wildtype*), lo que revela un grupo de *miRNAs* alterados significativamente en sus niveles de expresión. Dentro de los *miRNAs* cuya expresión se ve reducida por un bajo nivel de *FOXG1*, el estudio se centra en la familia *miR200*, conocida por su intervención en un proceso de control similar al de *FOXG1* en el desarrollo del cerebro frontal.

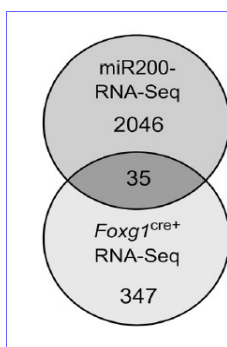


Figura 10. Identificación de *targets* comunes en sobreexpresión de *miR200* y deficiencia en *Foxg1*. Se identificaron 2081 genes diferencialmente expresados en condiciones de sobre-expresión de miembros de ella familia *miR200* y 382 genes en *Foxg1<sup>cre/+</sup>*, de los cuales 35 son comunes [4].

Para identificar dianas (*targets*) de *miR200* con un papel potencial en el síndrome *FOXG1*, se realizó *RNA-seq* en células con sobre-expresión de miembros de la familia *miR200* y se comparó con los resultados obtenidos en

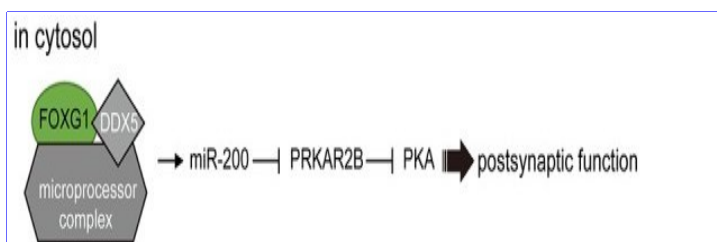
*Foxg1<sup>cre/+</sup>*, el resultado fue la identificación de 35 genes diferencialmente expresados comunes (Fig. 10); entre los que se encuentra *PRKAR2B*, cuyo producto (*cAMP-dependent protein kinase type II-beta regulatory subunit*) inhibe funciones post-sinápticas por atenuación de la actividad de *PKA* (proteína quinasa A).

Así, niveles incrementados de *PRKAR2B* pueden contribuir a disfunción neuronal en el síndrome *FOXG1*.

Esto indica que *FOXG1* suprime la transcripción de *Prkar2b* directamente, además de su regulación post-transcripcional mediada por miR200 (Fig. 11). En células N2a, *Drosha* no se localiza exclusivamente en el núcleo, también se han detectado niveles significativos en el citoplasma, donde interactúa con *FOXG1* y *DGCR8*, sugiriendo que *FOXG1* influye en la maduración de *miRNA*. *FOXG1* interacciona con *DDX5* (*ATP-dependent RNA helicase*) y se asocia con el complejo microprocesador, implicado en el proceso de maduración de miR200.

Figura 11.

En el citoplasma, *FOXG1* coopera con *DDX5* y el complejo microprocesador para inducir la biogénesis de miR200 y promover la función post-sináptica [9].



Se sugiere que *FOXG1* y la familia miR200 son ambos parte de una red multinivel que equilibra la expresión de la subunidad reguladora de *PKA*, *PRKAR2B*. De este modo, *FOXG1* y *MECP2* pueden afectar un *pathway* común dependiente de *PKA* para ajustar la función neuronal en el hipocampo.

## 2.5 Planteamiento inicial.

Se han escogido dos estudios con datos *GEO* (*Gene Expression Omnibus*) públicos sobre el estudio de estos 2 casos en ratones y su diferenciación en la expresión de *miRNAs*.

El estudio '*Quantitative analysis of wild type and Mecp2-/- mice hippocampus miRNAs using next generation sequencing*' analiza la expresión diferenciada de *miRNA* en ratones con mutaciones en el gen *Mecp2*:

<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*' analiza la expresión diferenciada de *miRNA* en ratones con mutaciones en el gen *Foxg1*:

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

Se identifican *miRNAs* desregulados no comunes entre las dos formas de *RTT*. Aunque ambos, *Mecp2* y *Foxg1* se asocian con el conjunto procesador de *miRNA*, parecen afectar a la maduración de diferentes *miRNAs*. Entre los *miRNAs* reprimidos por *Mecp2* se ha encontrado, entre otros, miR-134. Entre los *miRNAs* regulados por *Foxg1* se encuentran los de la familia miR200. No se ha realizado un análisis de *isomiRs*. Es posible que en estos 2 estudios hayan pasado desapercibidos niveles desregulados de *isomiRs* en ambas anomalías, que puedan darnos idea de algún *pathway* común afectado por ambos genes, *Mecp2* y *Foxg1*.

El objetivo principal del trabajo será hacer un análisis detallado sobre los datos en crudo disponibles de la caracterización de *isomiRs* de ambas anomalías genéticas, para encontrar puntos en común entre ambos síndromes. Para ello, se detallará un *pipeline*, utilizando herramientas de análisis de *miRNA/isomiR*, que permita cruzar datos de ambos orígenes de una manera sistematizada para poder comparar las dos situaciones y hallar vías genéticas y rutas metabólicas comunes.

### 3. Desarrollo del *pipeline*.

Uno de los objetivos del trabajo es establecer un *pipeline* completo de todo el proceso de análisis que pueda ser factible de utilizar en comparación de experimentos similares sobre *sRNAs* [10]. Para mayor detalle sobre los *scripts* en *bash* desarrollados en esta sección, se recomienda consultar el documento técnico *TFM\_FCH\_1\_1\_Documento tecnico*, anexo.

#### 3.1 Datos de partida.

Se parte de 2 conjuntos de datos independientes y públicos referentes a los 2 experimentos de *smallRNA-seq* mencionados anteriormente. La identificación de los *datasets SRA* es la siguiente:

Experimento 1		Experimento 2	
Ratones <i>Mecp2</i> +/-	Ratones <i>wt</i>	Ratones <i>Foxg1</i> +/-	Ratones <i>wt</i>
<i>SRR1504855</i>	<i>SRR1504856</i>	<i>SRR6061838</i>	<i>SRR6061847</i>
		<i>SRR6061839</i>	<i>SRR6061848</i>
		<i>SRR6061840</i>	<i>SRR6061849</i>
		<i>SRR6061841</i>	<i>SRR6061850</i>
		<i>SRR6061842</i>	<i>SRR6061851</i>
		<i>SRR6061843</i>	<i>SRR6061852</i>
		<i>SRR6061844</i>	<i>SRR6061853</i>
		<i>SRR6061845</i>	<i>SRR6061854</i>
		<i>SRR6061846</i>	<i>SRR6061855</i>

#### 3.2 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](#) [11], de las que 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:

<code>--gzip</code>	Salida comprimida
<code>--skip-technical</code>	Carga sólo las muestras biológicas
<code>--readids</code>	Añade el identificador de reads
<code>-dumpbase</code>	Formatea las secuencias usando bases (en SOLiD se usan colores)
<code>--split</code>	Divide la salida en uno ( <code>-spot</code> ), dos ( <code>-files</code> ) o 3 ( <code>-3</code> ) ficheros. Se escoge la opción <code>-3</code> para diferenciar los reads alineados por la izquierda (3) y por la derecha (5), en caso de haber utilizado secuenciación <i>paired-end</i> , en el tercer fichero se guardan los reads no emparejados ( <i>singletons</i> ).

--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.

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
FFFFFFFFDFDDDC+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.

### 3.3 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. 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:

```
$ 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* (Fig. 12).

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* (Fig. 13).

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.



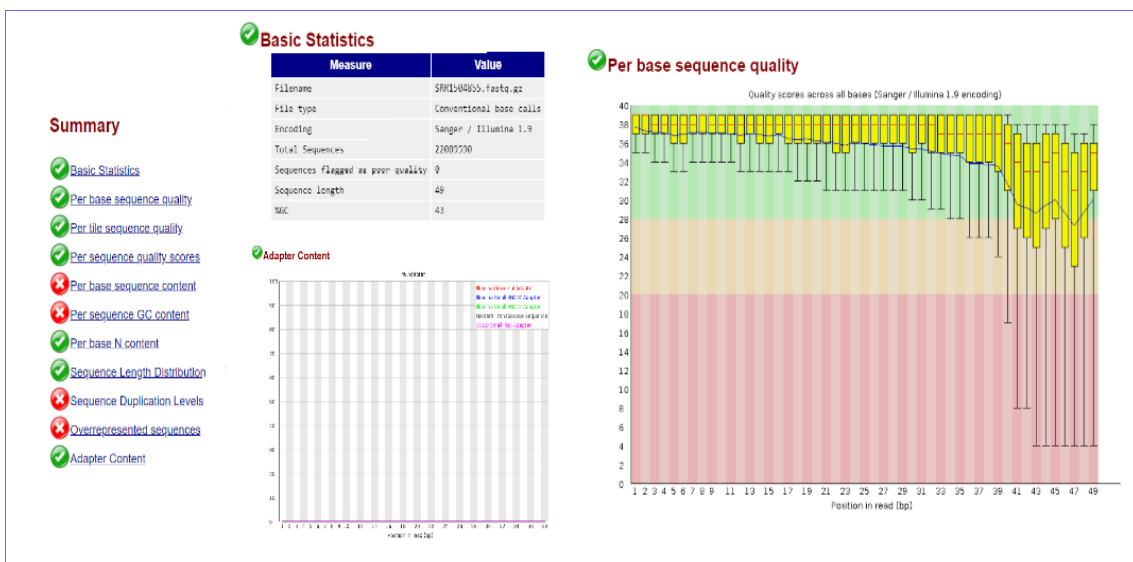


Figura 12. Resumen del informe de calidad SRR1504855\_fastq.html. Con estadísticas básicas, calidad de secuencia por bases y contenido de adapters.

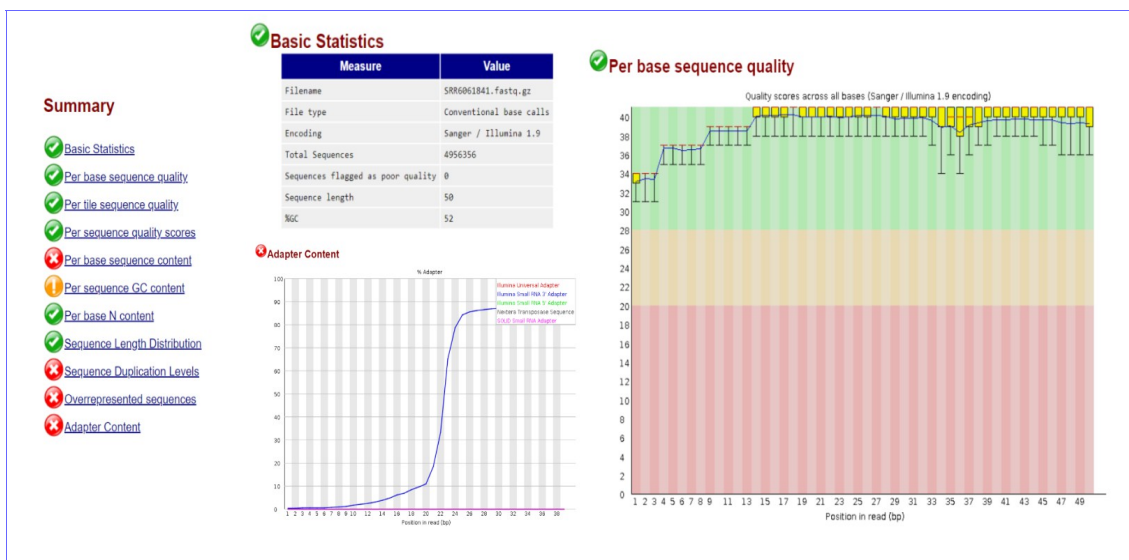


Figura 13. Resumen del informe de calidad SRR6061841\_fastq.html. Con estadísticas básicas, calidad de secuencia por bases y contenido de adapters.

Sin embargo, cuando se realizó el procesamiento de los *reads* de esta forma, eliminando *adapters* 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 los *reads* del experimento 1 no se alinearon, lo que llevó a pensar que en las secuencias del experimento 1 existía algún *adapter* 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

Pero esta secuencia no se encuentra 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 (Fig. 14):

 **Overrepresented sequences**

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)
TCCCTGAGACCCCTAACTTGTGAATCTCGTATGCCGCTTCTGCTTGAAA	69026	0.31248203339219965	TruSeq Adapter, Index 8 (96% over 28bp)
AAGGGATTCTGATGTTGGTATCTCGTATGCCGCTTCTGCTTGAAAAAA	32958	0.1492015017028383	TruSeq Adapter, Index 25 (96% over 28bp)
TGAGGTAGTAGGTTGTGGTTATATCTCGTATGCCGCTTCTGCTTGGA	29199	0.1321844361982273	TruSeq Adapter, Index 25 (96% over 28bp)
AAGGGATTCTGATGTTGGTACATCTCGTATGCCGCTTCTGCTTGAAA	28838	0.13055018223516146	TruSeq Adapter, Index 9 (96% over 28bp)
TCCCTGAGACCCCTTAACTTGCATCTCGTATGCCGCTTCTGCTTGAAA	25240	0.1142619668359621	TruSeq Adapter, Index 2 (96% over 28bp)

Figura 14. 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 los reads del experimento 1, por lo que se decide eliminarla como *adapter* antes de realizar el alineamiento de *reads*.

### 3.4 Anotación de *miRNAs*.

Para la anotación de *miRNA* se ha escogido la herramienta *Miraligner* [12]. Para la definición del proceso se toma como referencia el tutorial *miRNA annotation* [13]. 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*.

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.

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_fich_entrada -o path_fich_salida
```

Los resultados finales se guardan en  
 ~/TFM/datos/collapse/SRRx\_clean\_trimmed.fastq

Una muestra de un registro de este tipo:

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

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*.

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 (Fig. 15).

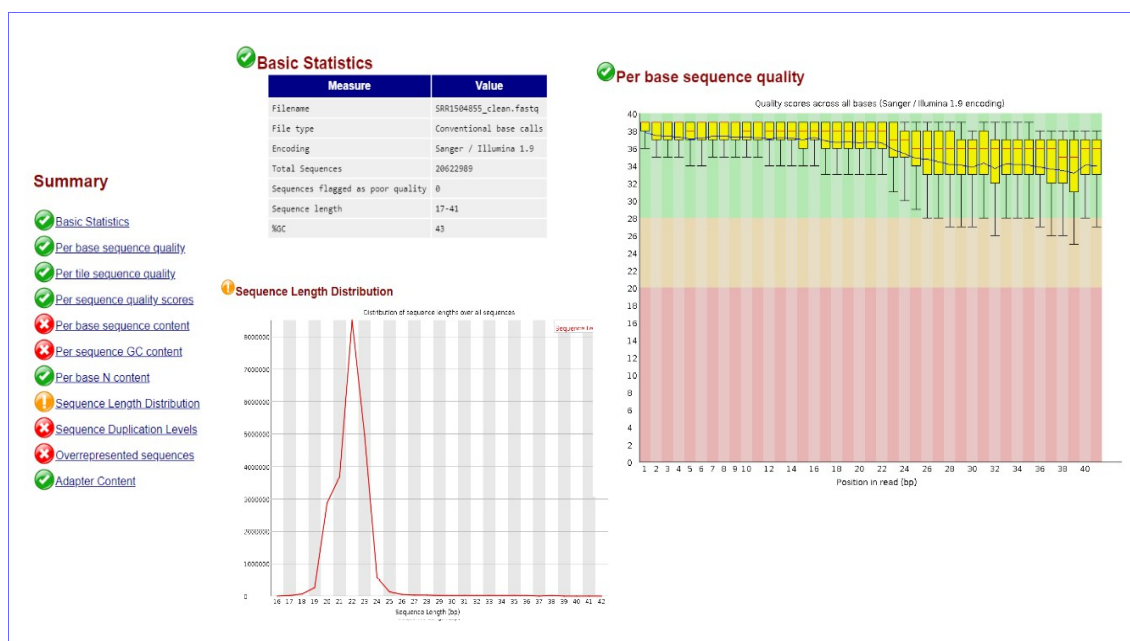


Figura 15. Resumen del informe de calidad *SRR1504855\_clean\_fastqc.html*. Con estadísticas básicas, calidad de secuencia por bases y distribución de secuencias por longitud.

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*.

Posteriormente, sobre este fichero se ejecuta *seqcluster collapse* y se 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*.

Una muestra de un registro de este tipo:

```
@seq_1_x12018
CATTGATCATCGACACTTCGAACGCACTTGCGGCCCGGGTT
+
BBBBEEEEGGGGIIIIIIIIIIIIIIIIIIIIIIHIIGHHHIIGEF
```

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](#). Esta herramienta, 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.

Para preparar 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 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 escrito en 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 5' diferentes de la secuencia anotada en miRBase
t3	Nuc. en 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

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*, 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
```

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;
```

### 3.5 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 RStudio* con paquetes propios del proyecto *Bioconductor* para análisis de experimentos de RNA-seq, como *DESeq2* [14, 15] y otros específicos del análisis de miRNAs/isomiRs, como *isomiRs* [16]. Para mayor detalle sobre el código *R* desarrollado en esta sección, se recomienda consultar el documento técnico *TFM\_FCH\_1\_1\_Documento tecnico*, anexo.

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.

#### 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*.

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* (Fig. 16).

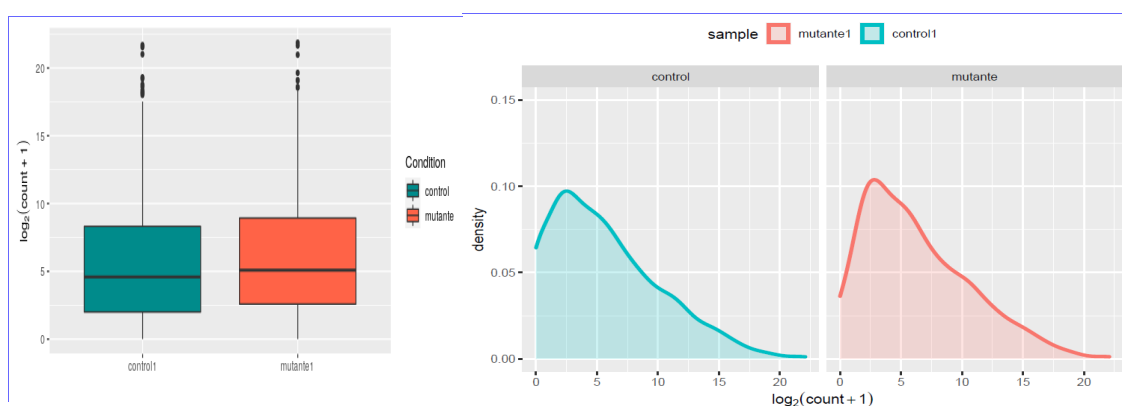


Figura 16. Boxplot y diagrama de densidad de las muestras del experimento 1. Mutante *Mecp2*+/- frente a control (wt).

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). La línea roja representa la desviación de la distribución entre las dos muestras (Fig. 17).

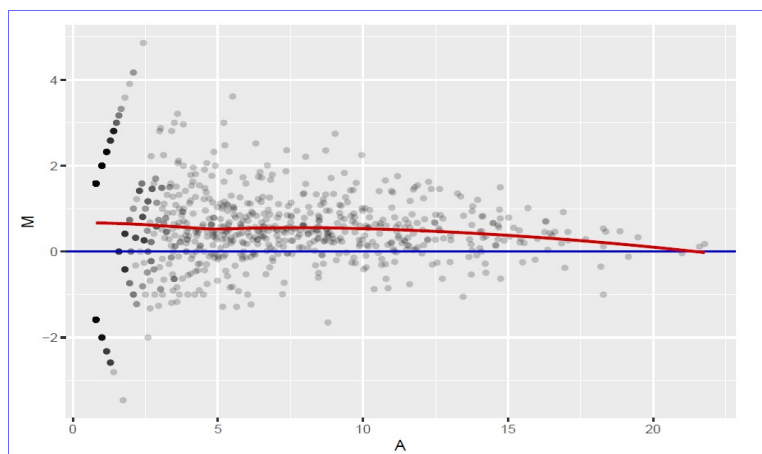


Figura 17. Plot MA de las muestras del experimento 1. Mutante frente a control.

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.

### Experimento 2 (ratones Foxg1+/-)

Se cargan los datos de los ficheros \*.mirna de las muestras y se hace un *boxplot* y un diagrama de densidad de los datos en crudo, utilizando *pseudocounts*. Se hace un diagramas MA entre muestras mutantes, 2 y 3 (se descarta 1 porque muestra un perfil muy diferente) y entre muestras de control, 1 y 2 (Fig. 18).

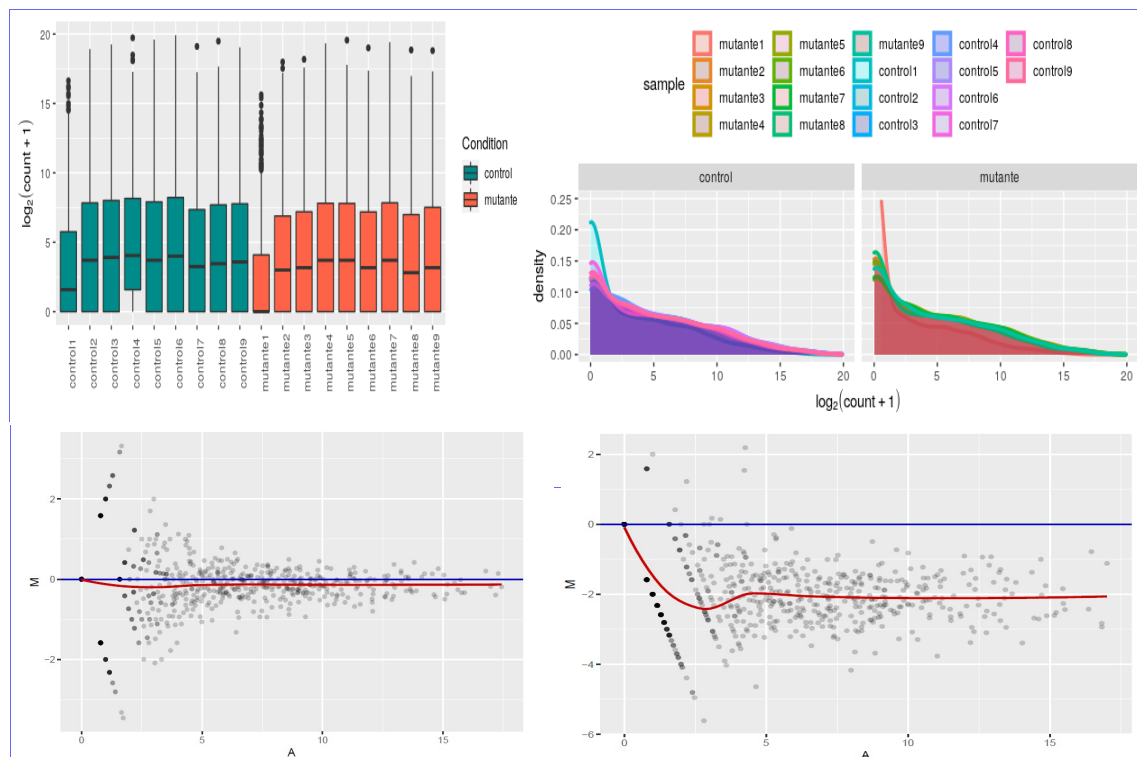


Figura 18. Boxplot, diagrama de densidad y digramas MA (mutante2- variante3 y control1- control2) antes de la normalización de las muestras del experimento 2. Ratones mutantes Foxg1+/- frente a ratones control (wt).

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*. Se observa que las muestras se han normalizado, excepto la mutante1, que muestra un perfil de densidad muy diferente y la cotrol1, 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 (Fig. 19).

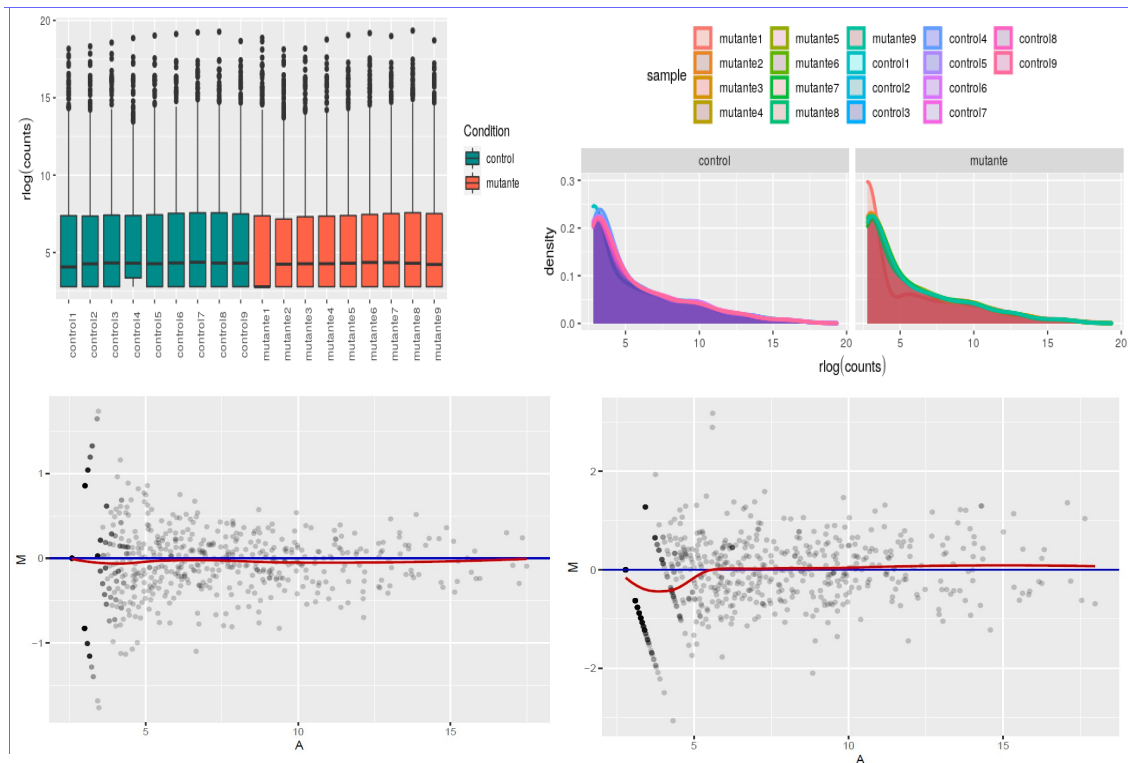


Figura 19. Boxplot, diagrama de densidad y digramas MA (mutante2- mutante3 y control1- control2) después de la normalización de las muestras del experimento 2. Ratones mutantes *Foxg1*+/- frente a ratones control (*wt*).

Se hace una representación MDS de las 2 primeras dimensiones y un dendrograma de las muestras del experimento 2 (Fig. 20).

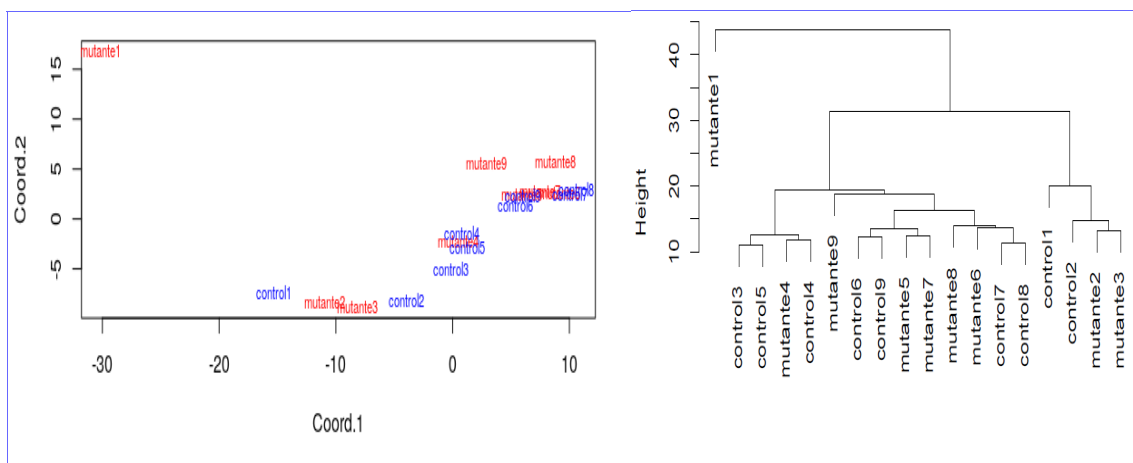


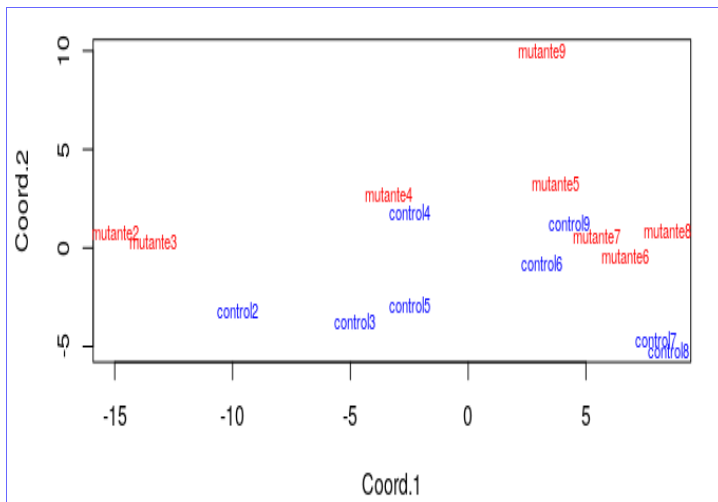
Figura 20. Representación MDS y dendrograma de las muestras del experimento 2 (ratones *Foxg1*+/-).



El propósito del *MDS* (*multidimensional scaling*) es ofrecer una representación visual de los patrones de proximidad entre muestras. El dendrograma es una representación de la distancia entre muestras que permite agruparlas en *clusters*.

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.

Se vuelve a repetir el diagrama MDS después de haber eliminado de la carga las 2 muestras (Fig. 21).



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 21. Diagrama MDS de las muestras del experimento 2 (ratones Foxg1+/-) una vez eliminadas las muestras mutante1 y control1

### 3.6 Análisis de expresión diferencial de *miRNAs/isomiRs*.

#### *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* (Fig. 22).

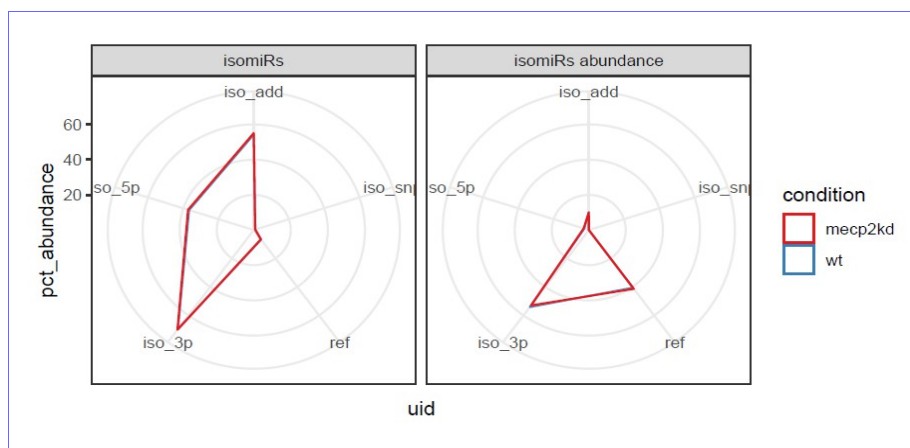


Figura 22. Gráfico isoPlot que muestra la diferenciación de isomiRs entre las 2 condiciones del experimento 1 (ratones Mecp2+/-).

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 comparan 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 [3] el fichero *mecp2.xls*, que contiene las tablas:

Table S1A. RNA-seq Data of MeCP2 Knockout.

Table S1B. RNA-Seq of MeCP2 Overexpression.

Se carga la tabla S1A para comprobar los resultados de *miRNAs* DE que reporta el artículo. Según dice el mismo:

"Normalized with total reads, 314 of all 720 (43.6%) sequenced *miRNAs* from the hippocampus of *Mecp2* 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."

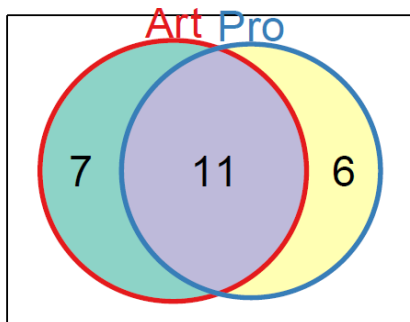
En la tabla aparecen 355 *miRNAs*, todos *up-regulated*; aunque con *logFC* mayor que 1.5 sólo hay 66. Para hacer la comparación con los datos descargados y procesados en este trabajo se tendrán en cuenta los *miRNAs*, sin diferenciar las variantes -3p y -5p; para ello se trunca el nombre del *miRNA*, dejando sólo el nombre principal.

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

```
## [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 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 (Fig. 23).

Figura 23. 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* [17] 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* ( $\log_2$ -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.

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 del documento técnico.

##	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

### 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* [18]. 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*.

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*).

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*.

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

Se cargan los ficheros en un nuevo *IsomirDataSeq*.

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.

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 5 primeros; el listado completo se puede consultar en el ANEXO 1 del documento técnico.

```
## 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
## ...           ...           ...           ...
```

## 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. 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* [19] 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*.

Con la función *plotRP* se ve un gráfico del *pdf* (*percentage of false prediction*) para cada *miRNA*. 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*  $< 0.1$ . 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*  $< 0.05$ .

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 del documento técnico.

```
## 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

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 [3] 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

### 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 (Fig. 24).

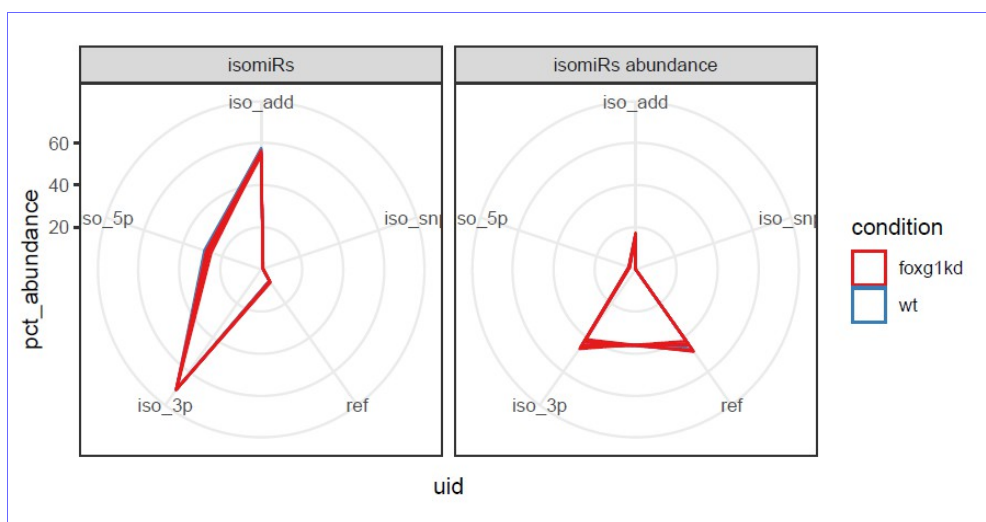


Figura 24. Gráfico isoPlot que muestra la diferenciación de isomiRs entre las 2 condiciones del experimento 2 (ratones *Foxg1*+/-).

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

```
## 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
## ...

```

### Comparación con los resultados del artículo.

Los resultados de *miRNAs* DE relacionados en el artículo de referencia [4] se muestran en el siguiente gráfico recogido del artículo de referencia (Fig. 25).

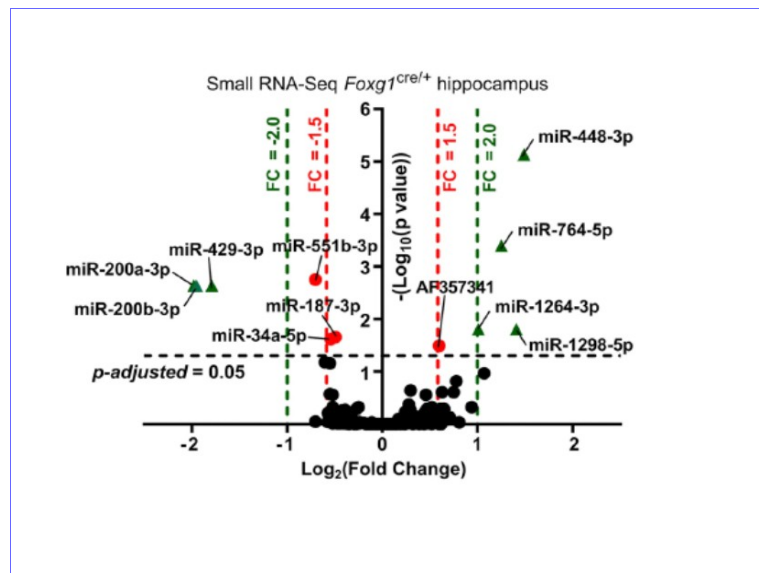


Figura 25. Volcano plot de *miRNAs* DE en el experimento 2 (ratones *Foxg1*+/-) según el artículo de ref. [4].

Comparándolos con los obtenidos por el procesamiento de los datos en crudo, se obtienen los siguientes diagramas de Venn (Fig. 26). Para los isomiRs *up-regulated* y *down-regulated*.

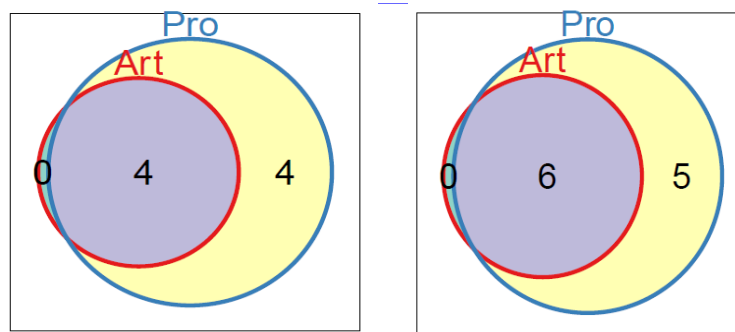


Figura 26. Diagramas de Venn comparando los *miRNAs* DE seleccionados en Art) Artículo de referencia y Pro) Procesados en este trabajo. A la izquierda los *up-reg* y a la derecha los *down-reg*.

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

Al 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 un análisis con *RankProd* de los datos del segundo experimento y hacer una comparación de *miRNAs* desregulados en ambas condiciones 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. 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 del documento técnico.

```
## 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
.....
```

### Comparación entre ambos experimentos.

En ambos experimentos aparecen los siguientes *miRNAs* DE comunes, según los distintos análisis realizados (Tabla 1):

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



Exp1 NOISeq	Exp1 DESeq2/isomiRs	Exp1 RankProduct	Exp2 isomiRs	Exp2 RankProduct
miR-770-5p	miR-200a-5p	miR-181c-5p	miR-200a-3p miR-200b-3p	miR-181c-3p
		miR-1843b-3p		miR-1843b-3p
		miR-1912-3p		miR-1912-3p
		miR-200a-5p		miR-200a-3p
		miR-200b-5p		miR-200b-3p
		miR-212-5p		miR-212-5p
		miR-30(b/e)-5p		miR-30a-3p
		miR-3102-3p		miR-3102-3p
	miR-34b-3p	miR-33-5p	miR-34a-5p miR-448-3p	miR-33-3p
		miR-34b-3p		miR-34b-3p
		miR-423-3p		miR-423-3p
		miR-431-5p		miR-431-3p
		miR-448-3p		miR-448-3p
		miR-455-3p		miR-455-3p
		miR-770-3p		miR-770-5p
		miR-871-3p		miR-871-3p
	miR-92a-3p	miR-92a-3p		

### 3.7 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, desarrollado en *Rstudio*, se detalla en el ANEXO 2 del documento técnico.

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 del documento técnico.

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.

	miR.base	FC.exp1	pfp.exp1	pval.exp1	FC.exp2	pfp.exp2	pval.exp2
mmu-miR-12191-3p	miR-12191	0.3148	0.6412	0.021210	0.7896	7.444e-01	1.747e-02
mmu-miR-200a-5p	miR-200	0.2106	0.4928	0.004208	NA	NA	NA
mmu-miR-200b-5p	miR-200	0.3449	0.8049	0.029210	NA	NA	NA
mmu-miR-200b-3p	miR-200	NA	NA	NA	0.3072	1.284e-07	1.310e-10
mmu-miR-200a-3p	miR-200	NA	NA	NA	0.4042	3.220e-06	6.571e-09
mmu-miR-34c-3p	miR-34	0.2585	0.6310	0.012790	NA	NA	NA
mmu-miR-34b-3p	miR-34	0.4838	0.8359	0.047280	NA	NA	NA
mmu-miR-34a-5p	miR-34	NA	NA	NA	0.7405	7.238e-01	1.846e-02
mmu-miR-871-3p	miR-871	0.2700	0.6374	0.012250	0.7446	5.989e-01	6.723e-03

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

	miR.base	FC.exp1	pfp.exp1	pval.exp1	FC.exp2	pfp.exp2	pval.exp2
mmu-miR-148a-5p	miR-148	5.897	0.3704	0.005929	NA	NA	NA
mmu-miR-148a-3p	miR-148	NA	NA	NA	1.061	0.2018	0.002471
mmu-miR-423-3p	miR-423	4.223	0.5900	0.018890	1.164	0.5793	0.020100
mmu-miR-431-5p	miR-431	5.043	0.4181	0.008032	NA	NA	NA
mmu-miR-431-3p	miR-431	NA	NA	NA	1.094	0.5468	0.017300
mmu-miR-92a-1-5p	miR-92	5.024	0.4517	0.009159	NA	NA	NA
mmu-miR-92a-3p	miR-92	3.426	0.7526	0.040160	0.954	0.5447	0.021120

Tabla 1 del experimento 1 con tabla 2 del experimento 2

	miR.base	FC.exp1	pfp.exp1	pval.exp1	FC.exp2	pfp.exp2	pval.exp2
mmu-miR-1298-3p	miR-1298	0.1608	0.7659	0.001635	1.785	0.0005101	2.082e-06
mmu-miR-1298-5p	miR-1298	0.1829	0.6294	0.002687	2.094	0.0005536	2.824e-06
mmu-miR-212-5p	miR-212	0.2379	0.5850	0.006867	1.337	0.4878000	1.344e-02
mmu-miR-212-3p	miR-212	NA	NA	NA	1.280	0.5617000	2.293e-02
mmu-miR-34b-3p	miR-34	0.4838	0.8359	0.047280	1.363	0.2847000	5.520e-03
mmu-miR-34b-5p	miR-34	NA	NA	NA	1.310	0.3832000	9.385e-03
mmu-miR-34c-5p	miR-34	NA	NA	NA	1.261	0.6870000	3.295e-02
mmu-miR-34c-3p	miR-34	0.2585	0.6310	0.012790	NA	NA	NA

Tabla 2 del experimento 1 con tabla 1 del experimento 2

	miR.base	FC.exp1	pfp.exp1	pval.exp1	FC.exp2	pfp.exp2	pval.exp2
mmu-miR-92a-1-5p	miR-92	5.024	0.4517	0.009159	NA	NA	NA
mmu-miR-92a-3p	miR-92	3.426	0.7526	0.040160	0.954	0.6966	0.02061
mmu-miR-92b-5p	miR-92	NA	NA	NA	0.847	0.8377	0.03761

### 3.8 Análisis detallado de *isomiRs* DE.

#### Experimento 1 (ratones *Mecp2+/-*)

Se hace un análisis de *isomiRs* DE en el experimento 1, con todas las variantes de *isomiRs* indicadas en la función *isoDE*. Se han seleccionado 7 *isomiRs* DE, de los que se muestra el detalle de anotación de la secuencia y el identificador.

##	seq	uid
## 150	ACCCGTCCCGTTCGTCCTCCCGG	mmu-miR-1247-5p;iso_3p:a
## 162	ACGCCCTTCCCCCTTCTTCA	mmu-miR-1249-3p;ref

```
## 536   CATCTAACCTGGCCTTTGA                               mmu-miR-1981-3p;iso_3p:c
## 792   GCTCAAACATCCGACTGAAAG mmu-miR-30f;iso_snp:5CT;iso_5p:GCT;iso_3p:ctc
## 1580  TTTTTCATTATTGCTCCTGACC                               mmu-miR-335-3p;ref
## 41    AACATGGAGGAAATTTACAGT                               mmu-miR-376c-3p;iso_snp:6GA
## 1559  TTTGATGTCTGACTTTGCAGTT                               mmu-miR-7090-3p;iso_add:TT
```

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).

### Experimento 2 (ratones *Foxg1*+/-)

Se hace un análisis de *isomiRs* DE en el experimento 2, con todas las variantes de *isomiRs* indicadas en la función *isoDE*. 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	TCCCTGAGACCCTTTAACCTGTGAA	mmu-miR-125a-5p;iso_add:A
## 4339	CAAATCTTATTTGAGCACCTG	mmu-miR-1264-3p;iso_3p:t
## 4341	CAAATCTTATTTGAGCACCTGT	mmu-miR-1264-3p;ref
## 710	AAGCCCTTACCCAAAAAGCAT	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	AACATTCATTGTTGTCGGTGGGA	mmu-miR-181d-5p;iso_add:A;iso_3p:t
## 463	AACATTCATTGTTGTCGGTGGGT	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	AGGCTGCAGGCCACTTCCCT	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	AATGGCGCCACTAGGGTTGTGT	mmu-miR-652-3p;iso_add:T
## 9121	TAACTCTTCCCTTGTGCCCTCAGT	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).

### 3.9 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).

Se implementa este método usando *Discriminer* dentro de la función *isoPLSDA*. 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*; para el experimento 1 el p-valor=0,6275 y para el experimento 2 el p-valor=0,5. En el gráfico se representa la distribución de muestras a lo largo de las 2 primeras componentes (t1 y t2) (Fig. 27).

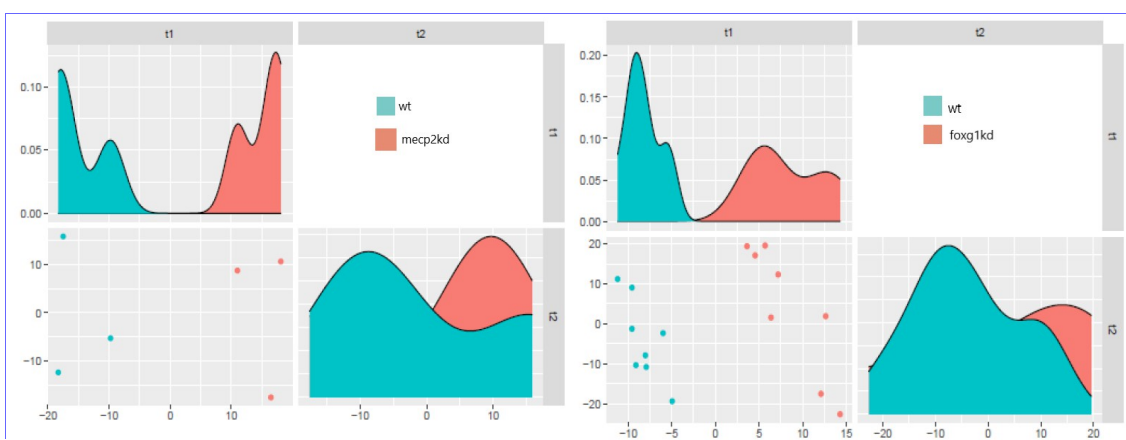


Figura 27. Representación PLS-DA de las muestras del experimento 1 (ratones Mecp2+/-) a la izquierda y del experimento 2 (ratones Foxg1+/-) a la derecha.

### 3.10 Integración gen-*miRNA*.

Para la descripción y selección de las herramientas de predicción de *targets* se ha tomado como referencia el artículo [20]. 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, desde el extremo 5' hacia el extremo 3' (Fig. 28).

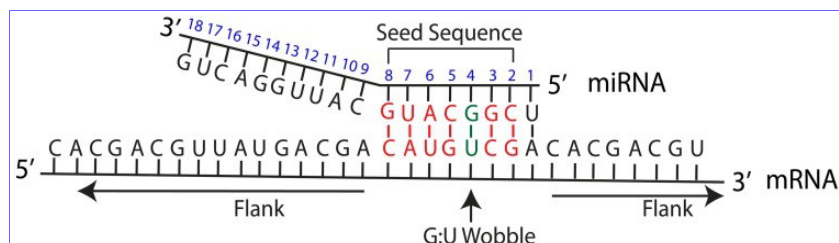


Figura 28. Interacción entre el *miRNA* y el *mRNA* target. Se representa la secuencia *seed* con los emparejamientos de nucleótidos WC (Watson Crick) marcados en rojo y un ejemplo de G-U wobble marcado en verde. Los flancos se refieren a las secuencias del *mRNA* a cada lado de la región target que hibrida con el *miRNA* [20].

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 ( $\Delta G$  positivo) es más inestable que una con baja energía libre ( $\Delta G$  negativo) (Fig. 29).

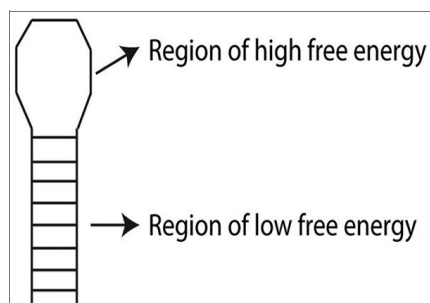


Figura 29. Estructura secundaria de RNA en forma de bucle (hairpin loop) como ejemplo de zonas de alta energía libre (menos estable) y baja energía libre (más estable) [20].

- 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 une 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.

### 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/> [21]. 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* [22] 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  $pfp < 0.5$ . 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/> [23]. *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* y se importa a un *data.frame*.

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. 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> [24, 25]. *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 a un *data-frame*. 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](http://carolina.imis.athena-innovation.gr/diana_tools) [26].

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.

### 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*.

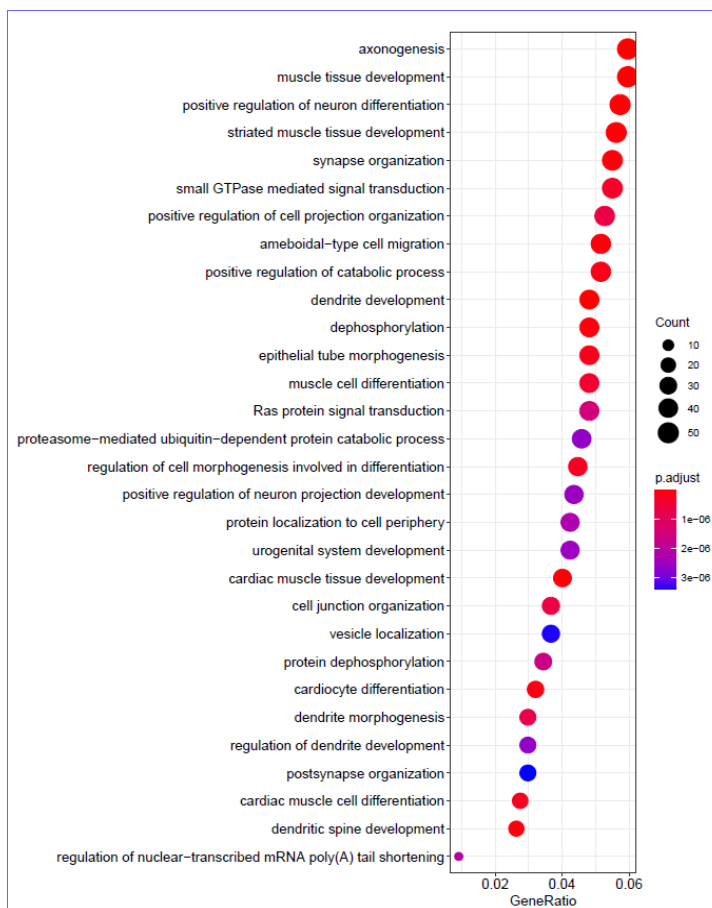
### 3.11 Enriquecimiento GO.

#### Enriquecimiento GO en el 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 el paquete *clusterProfiler* de *Bioconductor* [27].

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

Figura 30. Categorías *GO* más representadas en el experimento 1 (ratones *Mecp2+/-*).





En el mapa de enriquecimiento (Fig. 31) se ve que hay 2 redes importantes de categorías relacionadas, la referente a diferenciación neuronal y otra referente al desarrollo muscular.

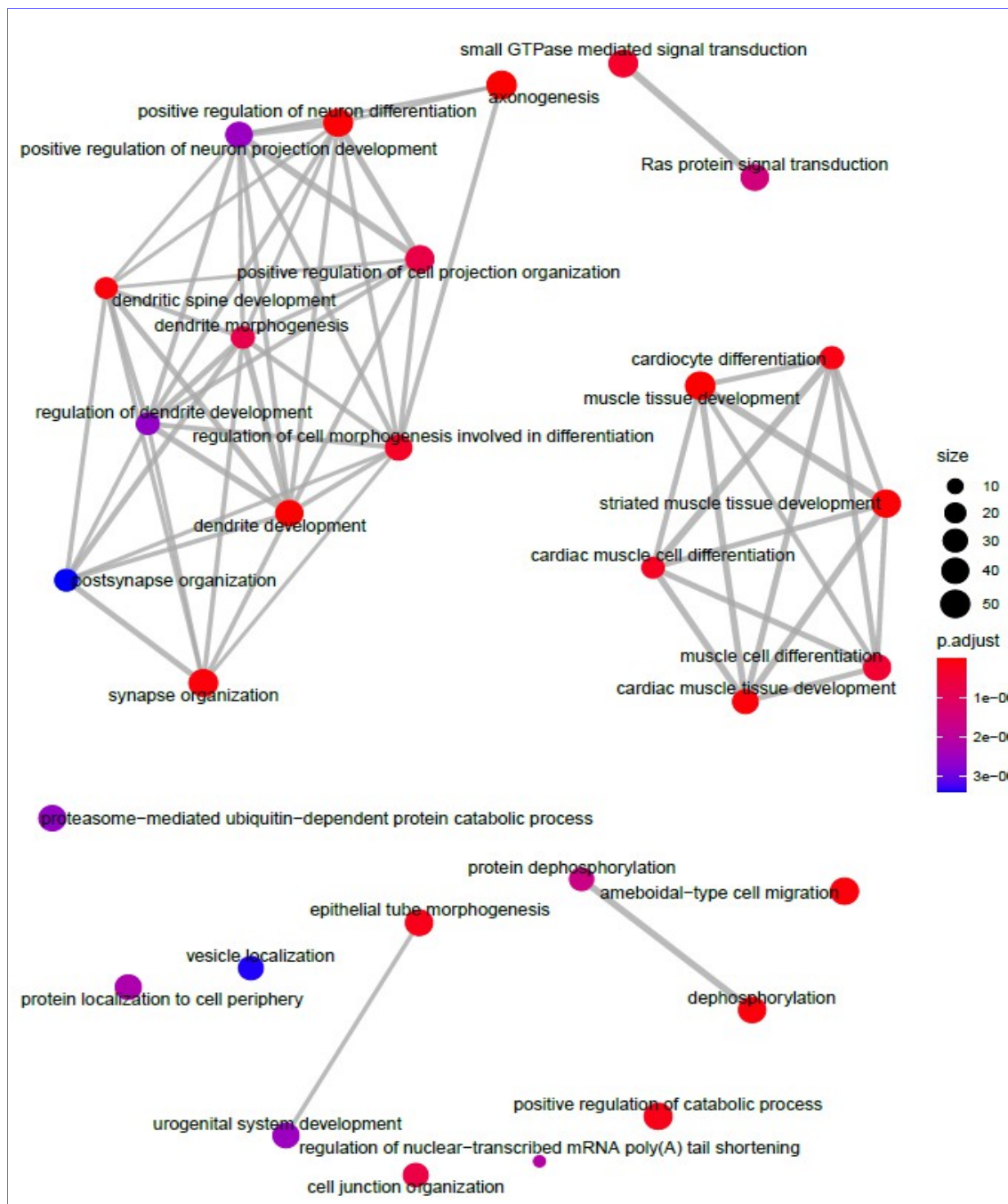


Figura 31. Mapa de enriquecimiento de categorías GO más representativas entre los genes target seleccionados del experimento 1 (ratones Mecp2+/-)

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:

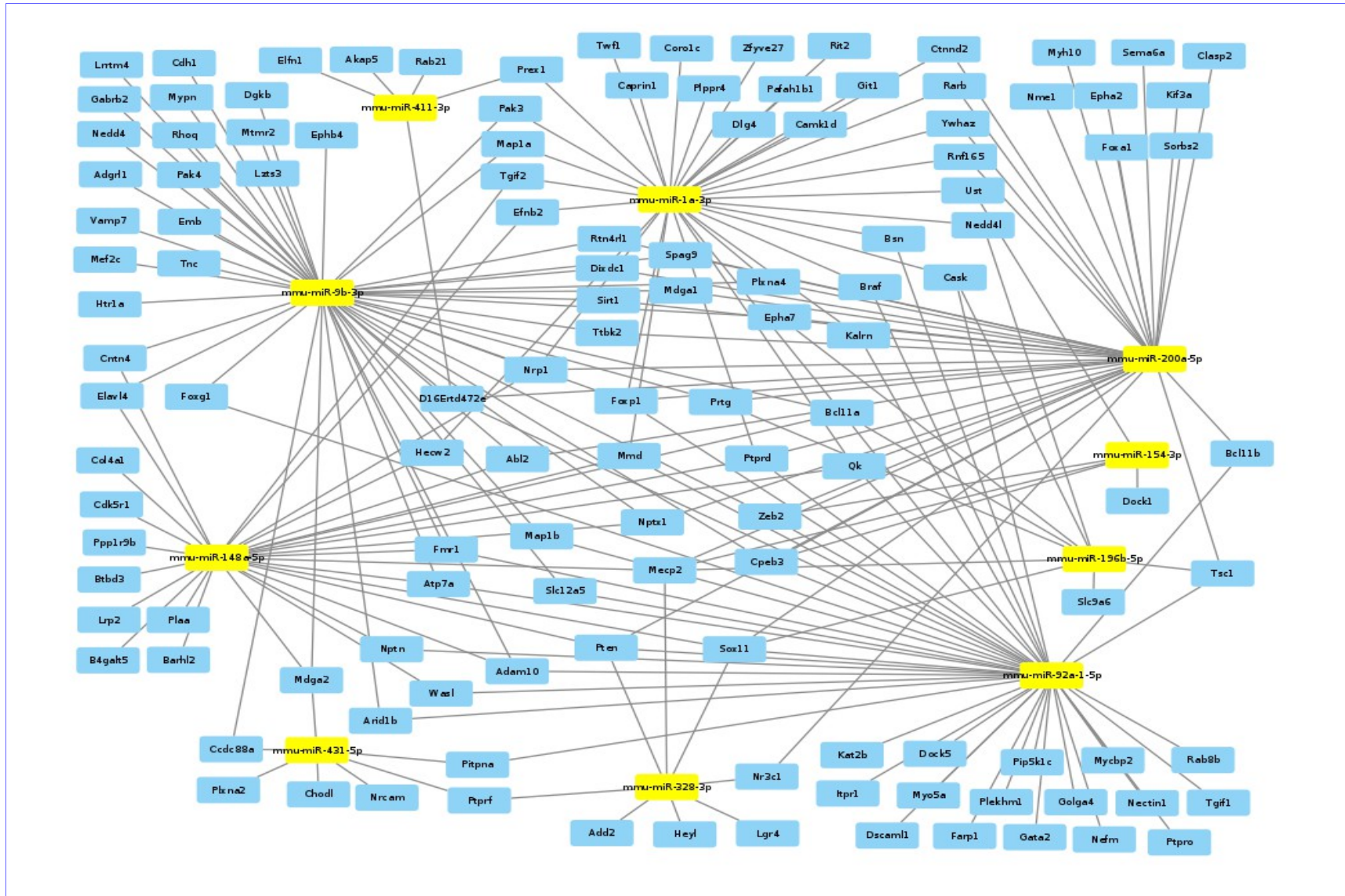


Figura 32. Mapa de red bipartita de miRNAs DE y genes target seleccionados del experimento 1 (ratones Mecp2+/-).

Categoría	Descripción	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* [28] y el modelo se representará con *cytoscape* [29]. El mapa de red bipartita resultante (Fig. 32) presenta 126 genes *target* diferentes implicados en estas funciones biológicas.

#### Enriquecimiento GO en el experimento 2 (ratones *Foxg1*+/-)

Se obtiene el mapa resultante del enriquecimiento GO de los genes *target* seleccionados para este experimento (Fig. 33). Aquí se pueden apreciar 3 redes de características relacionadas con el sistema nervioso. En la primera de ellas se encuentran las categorías:

Categoría	Descripción	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 (Fig. 34) implicados en estas funciones biológicas.

En la segunda se encuentran las categorías:

Categoría	Descripción	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 (Fig. 35) implicados en estas funciones biológicas.

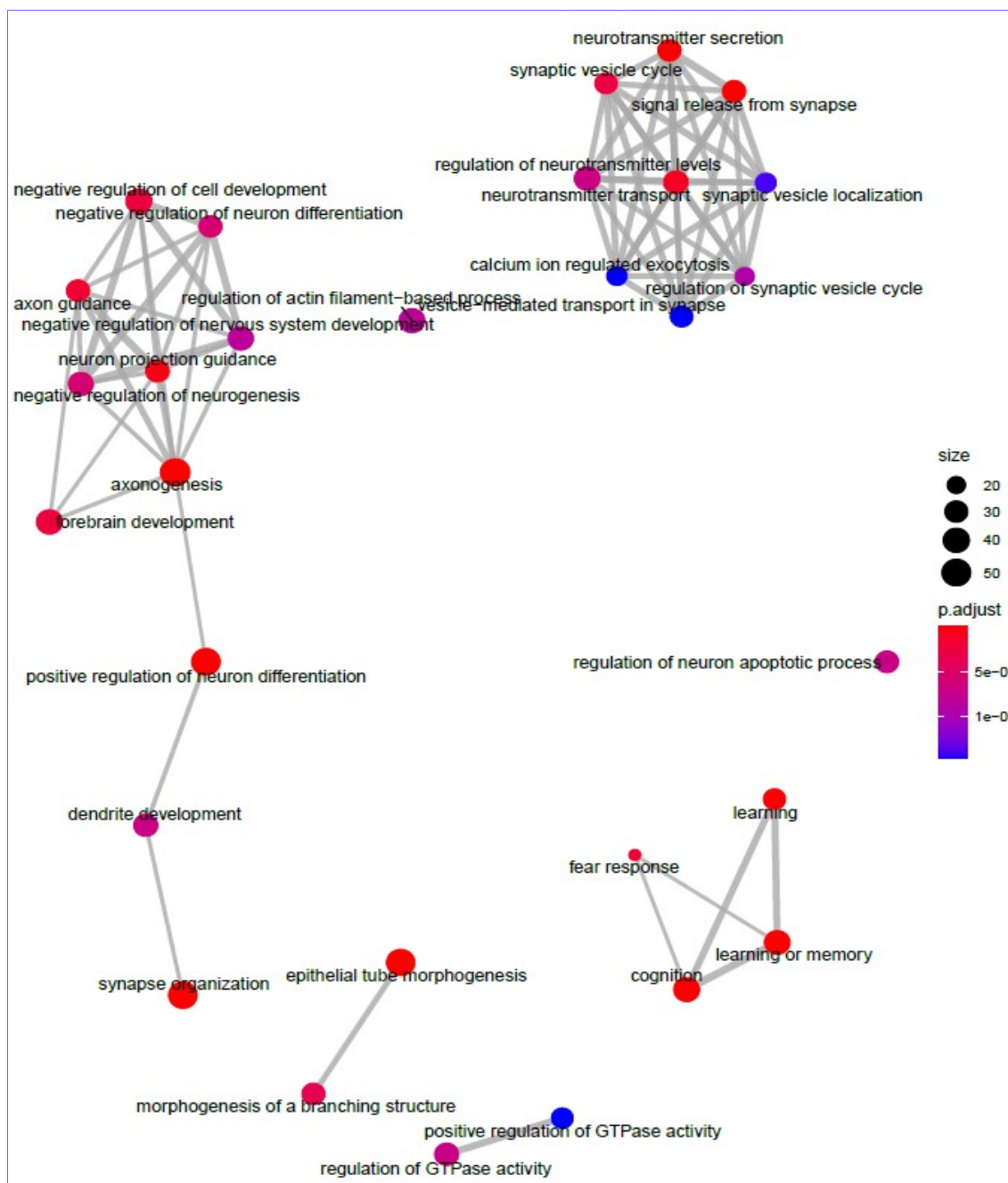


Figura 33. Mapa de enriquecimiento de categorías GO más representativas entre los genes target seleccionados del experimento 2 (ratones Foxg1+/-)

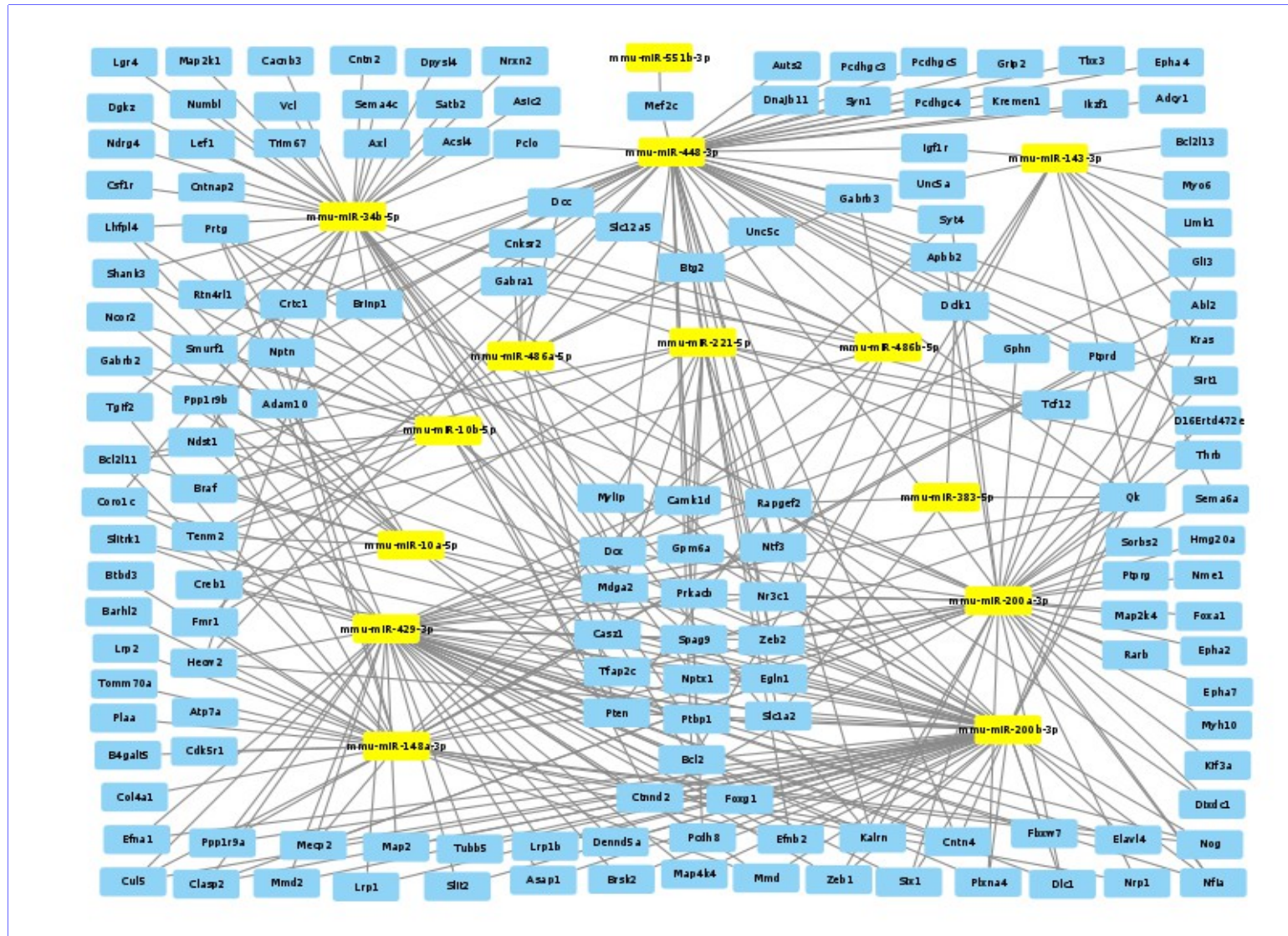


Figura 34. Mapa de red bipartita de miRNAs DE y genes target seleccionados en el grupo 1 del experimento 2 (ratones Foxg1+/-).

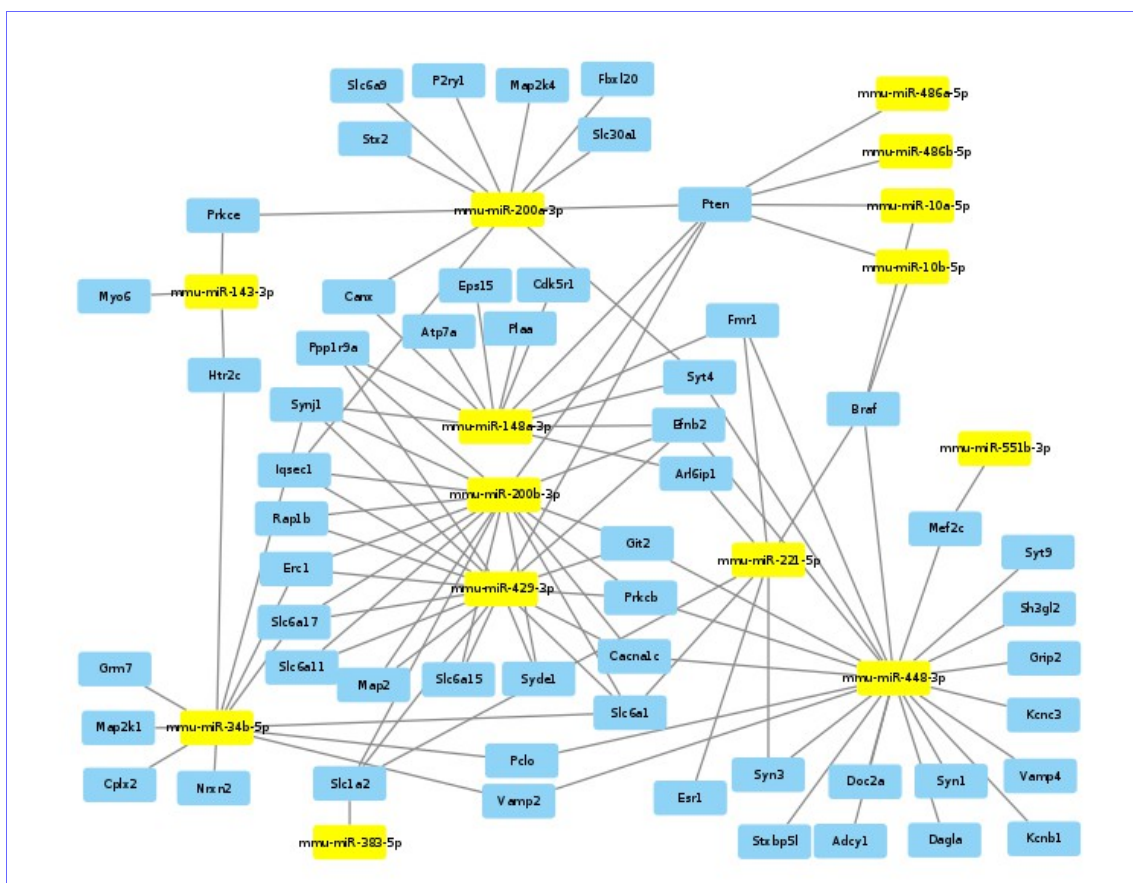


Figura 35. Mapa de red bipartita del grupo 2 del experimento 2 .

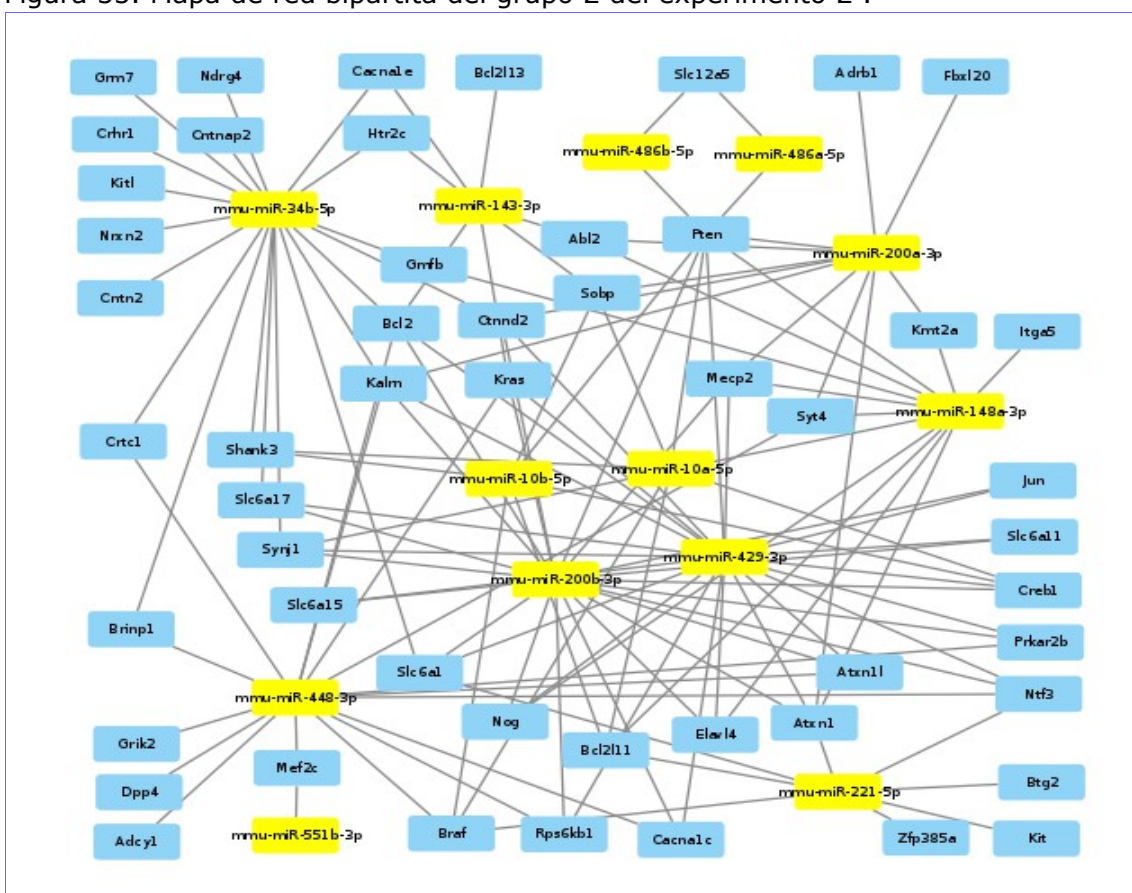


Figura 36. Mapa de red bipartita del grupo 3 del experimento 2 .

En la tercera se encuentran:

Categoría	Descripción	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 (Fig. 36) implicados en estas funciones biológicas.

### Enriquecimiento GO en el cruce de los 2 experimentos

El mapa de enriquecimiento GO de los genes *target* obtenidos en el proceso de cruce de los dos experimentos es el siguiente (Fig. 37).

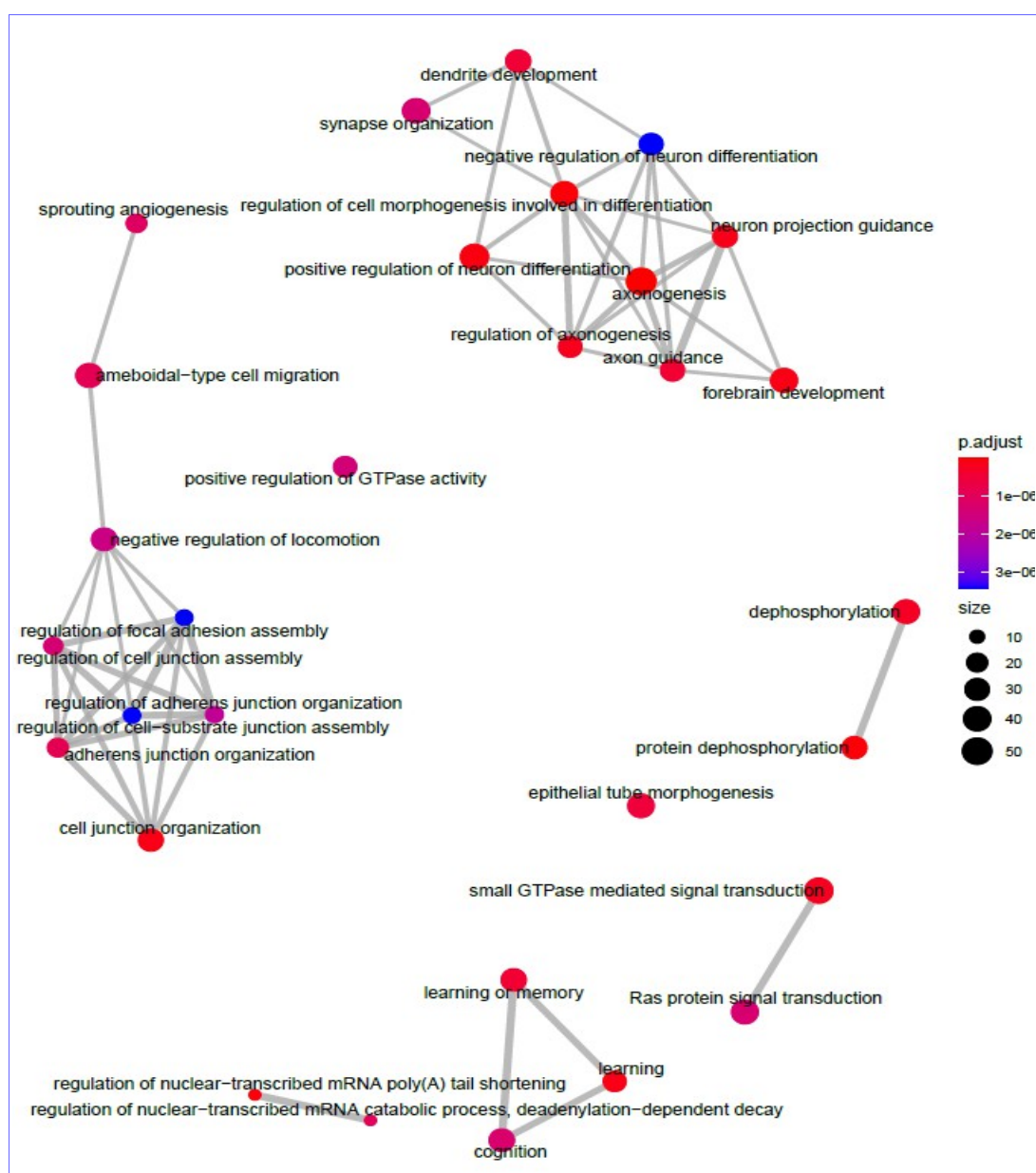


Figura 37. Mapa de enriquecimiento de los genes target seleccionados en el cruce de los 2 experimentos.

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	genes
GO:0016358	<i>Dendrite development</i>	32
GO:0050808	<i>Synapse organization</i>	48
GO:0045665	<i>Negative regulation of neuron differentiation</i>	28
GO:0010769	<i>Regulation of cell morphogenesis involved in differentiation</i>	38
GO:0097485	<i>Neuron projection guidance</i>	32
GO:0045666	<i>Positive regulation of neuron differentiation</i>	49
GO:0007409	<i>Axonogenesis</i>	51
GO:0050770	<i>Regulation of axonogenesis</i>	27
GO:0007411	<i>Axon guidance</i>	31
GO:0030900	<i>Forebrain development</i>	39

Hay 127 genes *target* diferentes (Fig. 38) implicados en estas funciones biológicas.

En el segundo agrupamiento se encuentran las siguientes categorías:

Categoría	Descripción	genes
GO:0007611	<i>Learning or memory</i>	32
GO:0007612	<i>Learning</i>	25
GO:0050890	<i>Cognition</i>	33

Hay 33 genes *target* diferentes (Fig. 39) implicados en estas funciones biológicas.

### 3.12 Enriquecimiento KEGG.

El enriquecimiento KEGG (*Kyoto Encyclopedia of Genes and Genomes*) también se realiza con el paquete *clusterProfile* de Bioconductor [27].

#### 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.

##	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

#### 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 de categorías GO señaladas anteriormente.



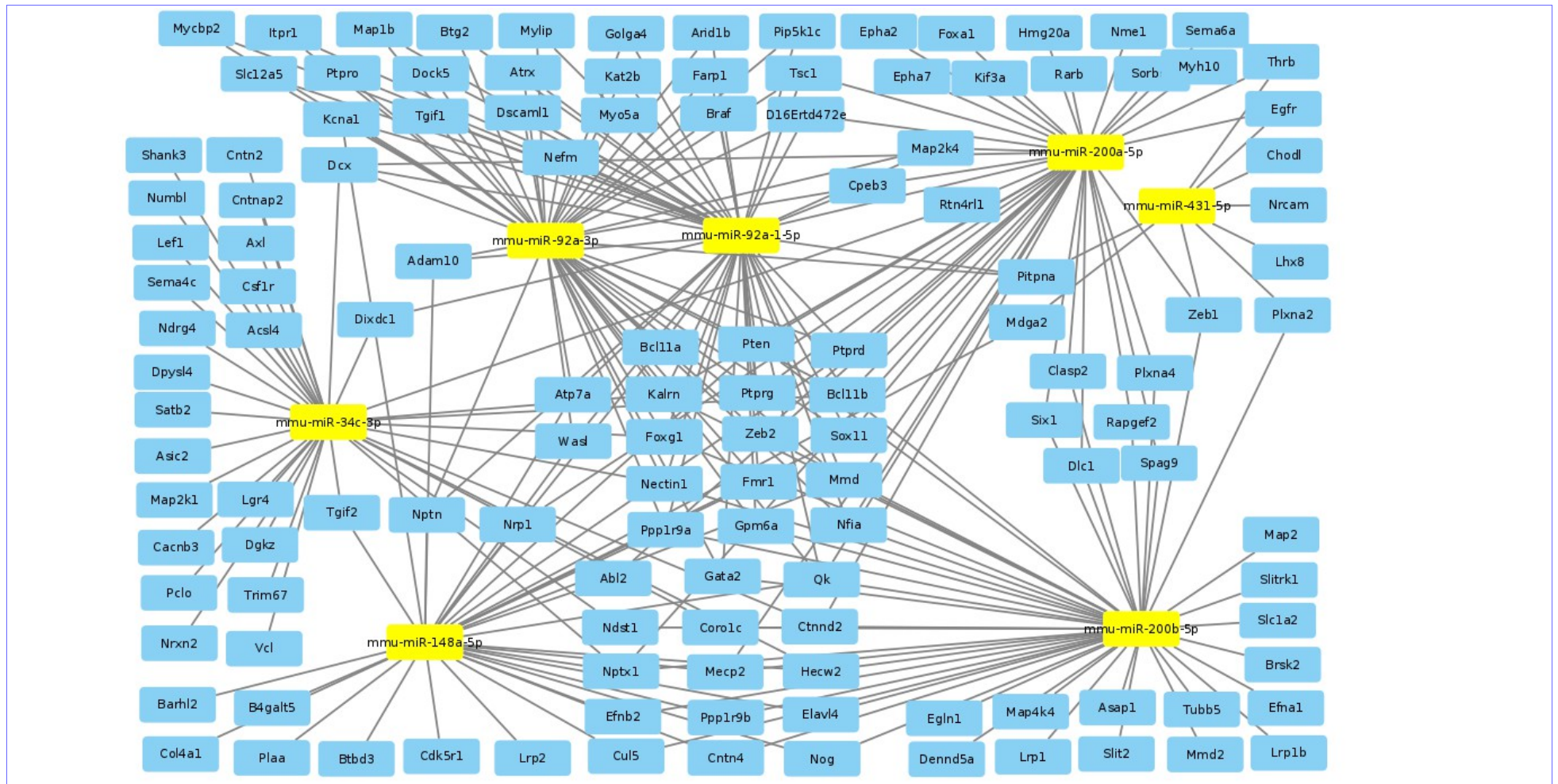


Figura 38. Mapa de red bipartita de los miRNAs y genes target seleccionados en el grupo 1 de enriquecimiento GO del cruce de los 2 experimentos.

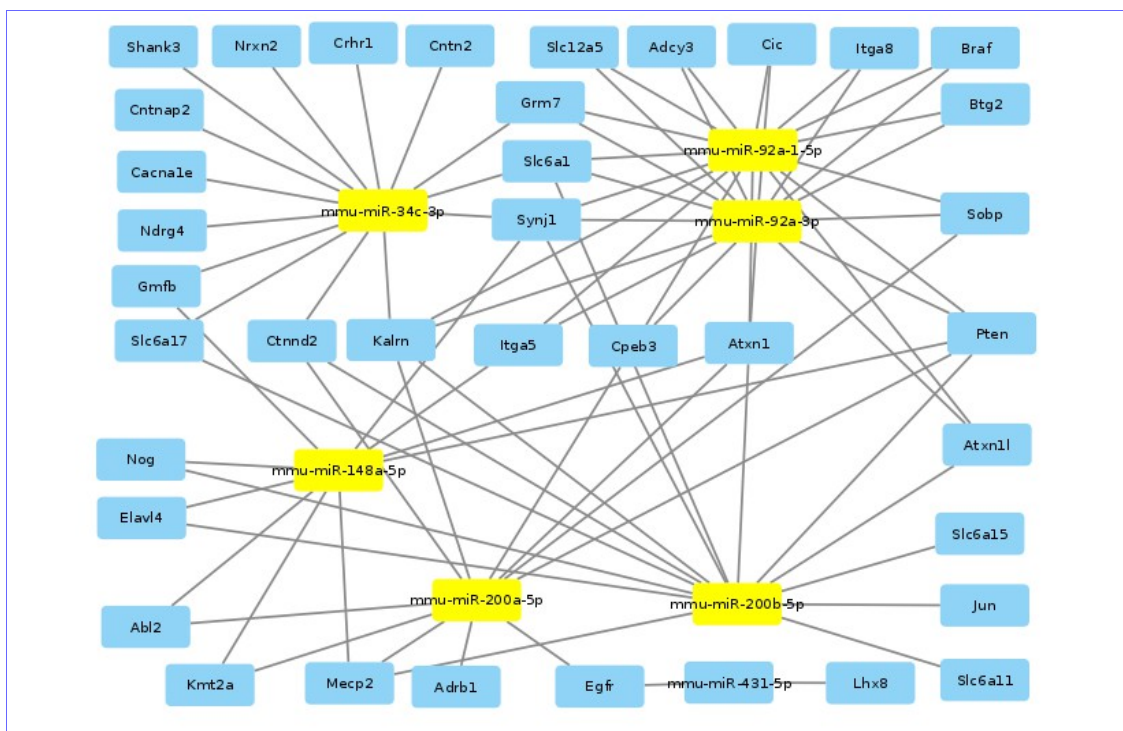


Figura 39. Mapa de red bipartita del grupo 2 de categorías GO en el cruce de los dos experimentos.

Para el primer grupo se seleccionan los *pathways*:

##	ID	Description	GeneRatio	p.adjust
##	mmu04727	GABAergic synapse	16/312	1.411396e-05
##	mmu04015	Rap1 signaling pathway	25/312	1.411396e-05
##	mmu04010	MAPK signaling pathway	30/312	1.411396e-05
##	mmu04211	Longevity regulating pathway	15/312	3.458688e-05
##	mmu01522	Endocrine resistance	15/312	4.285126e-05
##	mmu04151	PI3K-Akt signaling pathway	32/312	4.50833e-05
##	mmu04360	Axon guidance	21/312	4.804338e-05
##	mmu01521	EGFR tyrosine kinase inhibitor resistance	13/312	0.0001029755
##	mmu04014	Ras signaling pathway	22/312	0.0004958467
##	mmu05215	Prostate cancer	13/312	0.0007639823
##	mmu04068	FoxO signaling pathway	15/312	0.0009080047
##	mmu04340	Hedgehog signaling pathway	8/312	0.004438143
##	mmu05206	MicroRNAs in cancer	23/312	0.004586942
##	mmu05166	Human T-cell leukemia virus 1 infection	19/312	0.007043665
##	mmu04510	Focal adhesion	16/312	0.01028232
##	mmu04928	Parathyroid hormone synthesis, sec. and act.	10/312	0.01897064

Para el segundo se seleccionan:

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

Y para el tercero:

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

### Enriquecimiento KEGG en el 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

### 3.13 Relación de herramientas y programas empleados en el *pipeline*.

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 <i>smallRNAs</i> .
miraligner	Herramienta de mapeo y anotación de <i>smallRNAs</i> 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 &amp; 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 R de análisis de enriquecimiento utilizando 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.H s.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.

3.14 Diagrama resumen del *pipeline*.

El *pipeline* desarrollado anteriormente se resume en el siguiente diagrama de flujo (Fig. 40a y 40b).

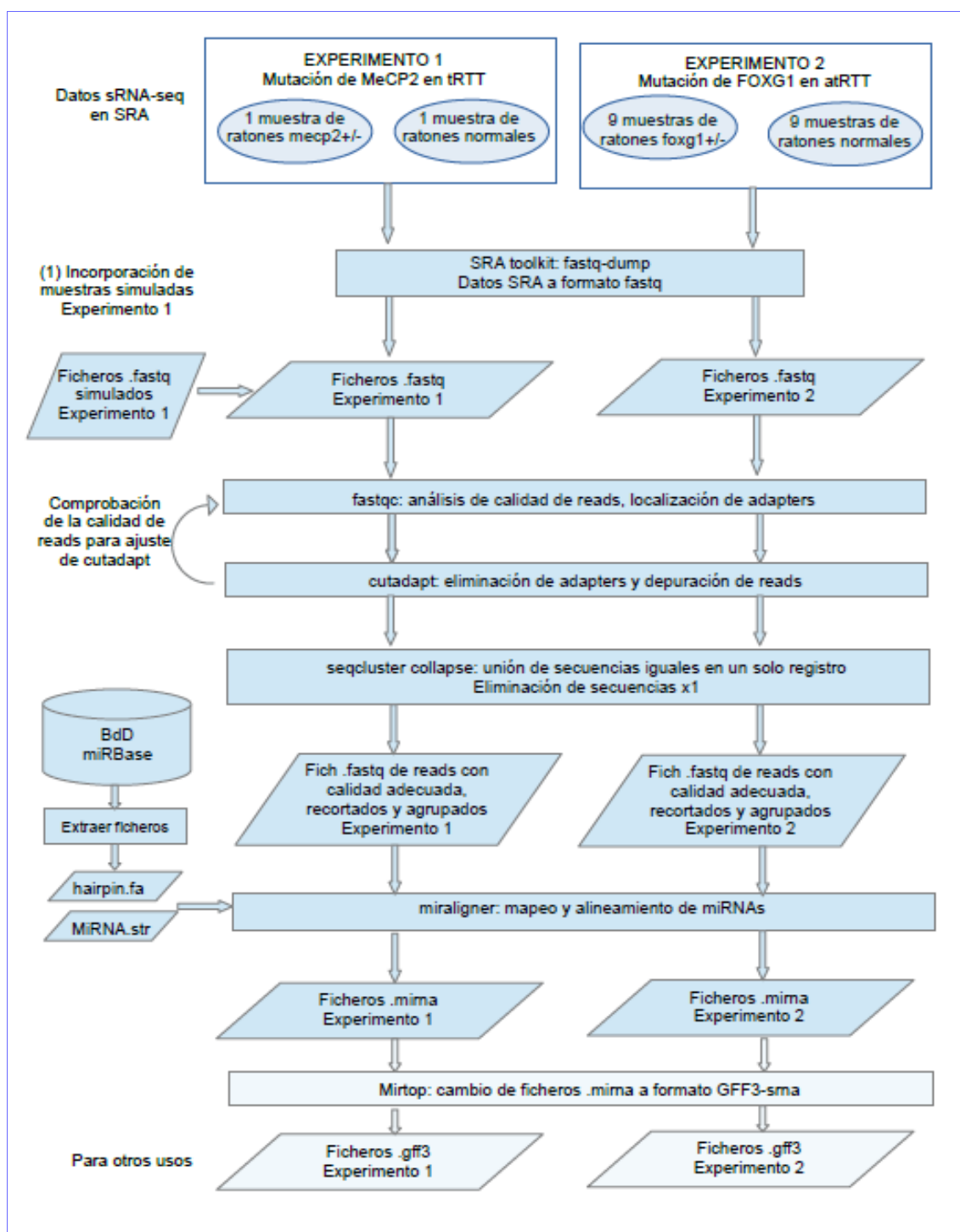


Figura 40a. Primera parte del *pipeline*, que incluye las tareas de descarga de datos en crudo, adaptación de formato, comprobación de la calidad de reads, eliminación de *adapters*, unión de secuencias iguales en un solo registro, preparación de la base de datos de *miRNAs*, alineamiento y anotación de *miRNAs/isomiRs*. Los ficheros \*.mima obtenidos se emplearán como punto de partida en la segunda parte del *pipeline*.

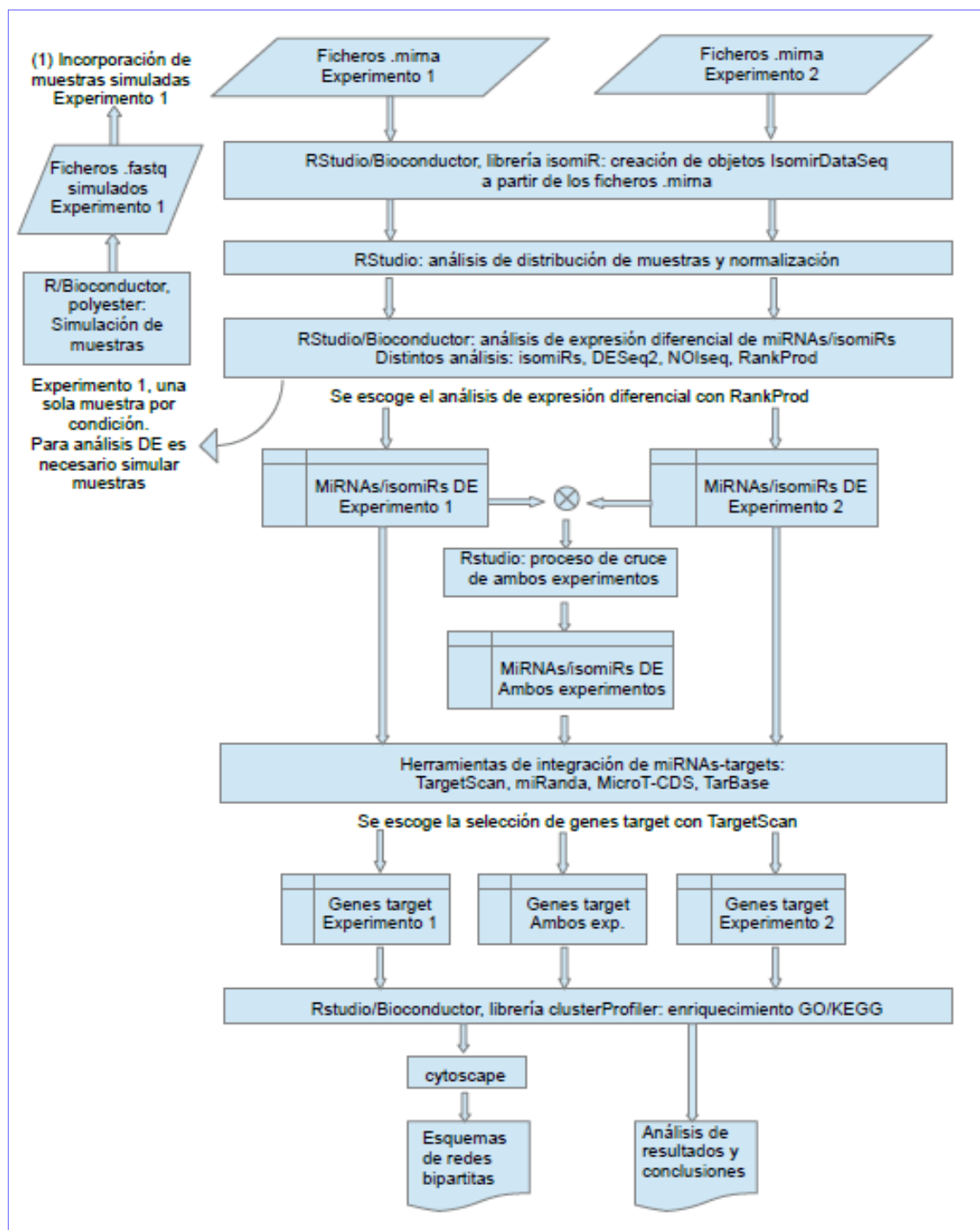


Figura 40b. Segunda parte del pipeline, que incluye las tareas de control de calidad de datos y normalización, análisis de expresión diferencial de *miRNAs/isomiRs* (al hacer este análisis se evidenció el problema de disponer sólo de una muestra por condición en el experimento 1, por lo que fue necesario generar muestras simuladas a partir de las disponibles y tratar estos datos de nuevo desde el principio), proceso de cruce de información de *miRNAs/isomiRs* comunes en ambos experimentos, integración *gen-miRNA*, enriquecimiento GO y enriquecimiento KEGG.

## 4. Análisis de resultados

### 4.1 Análisis de resultados del experimento 1 (ratones *Mecp2+/-*).

En el enriquecimiento *KEGG* [30] del primer experimento, los *pathways* más representativos son los siguientes:

ID	Descripción
mmu04810	<i>Regulation of actin cytoskeleton</i>
mmu04510	<i>Focal adhesion</i>
mmu04151	<i>PI3K-Akt signaling pathway</i>
mmu05206	<i>MicroRNAs in cancer</i>
mmu04360	<i>Axon guidance</i>

El direccionamiento del axón (*axon guidance*) es de gran importancia en la formación de la red neuronal. Los axones se guían por una serie de factores como netrinas, semaforinas, efrinas o *slits* (reguladores negativos en el crecimiento de axones). Estos factores interactúan con receptores de los conos de crecimiento y sus señales de transducción convergen en Rho GTPasas, provocando cambios en la organización del citoesqueleto, que determinan el camino que debe tomar el cono de crecimiento.

El *pathway Axon guidance* tiene relación con el *pathway Regulation of actin cytoskeleton*, que interviene también en el posicionamiento de los axones.

Otro *pathway* importante en los procesos de motilidad celular es *Focal adhesion*, que determina la adhesión molecular en la matriz celular e interviene también en procesos de proliferación y diferenciación celular. Este *pathway* también se relaciona con *Regulation of actin cytoskeleton* y tiene conexiones con *PI3K-Akt signaling pathway* en procesos de supervivencia celular.

La unión de factores de crecimiento a sus receptores tirosín quinasa (*RTK*) o receptores acoplados a proteínas G (*GPCR*) estimulan *PI3K*, que cataliza la producción de *PIP3*, segundo mensajero que colabora en la activación de *Akt* o proteína quinasa B. Esta enzima controla procesos celulares clave como la apoptosis, síntesis de proteínas, metabolismo y ciclo celular, mediante fosforilación.

Para acotar los resultados y poder analizarlos con mayor detalle, se reduce el número de *miRNAs* y genes *target* de estos *miRNAs*, escogiendo los más significativos y de esta manera se centra el análisis en un número de elementos más manejable.

Dentro de la lista de *miRNAs* DE expresados se escogen los que se muestran en el mapa de red bipartita correspondiente, es decir, los que se ha comprobado que hibridan con genes *target* implicados en funciones neurológicas y de este grupo, se seleccionan los más significativos estadísticamente (por valores de FC, pfp y p-value en la tabla de *miRNAs* DE por RankProd). Los *miRNAs* sobre los que se centra el análisis son:

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

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

Para escoger un número reducido de genes *target* que hibridan con estos *miRNAs*, se hace una selección conjunta de los resultados obtenidos por dos de las herramientas de predicción de *targets* empleadas. Se escogen genes *target* de los obtenidos por *TargetScan* que obtengan una puntuación PCT por encima de 0.5 y Seedmatch igual a 8mer y que hayan sido también predichos por *MicroT-CDS*, con valores de miTG.score por encima de 0.7. Excepto para miR-9b-3p, que no encuentra *targets* en *microT-CDS* y se seleccionan sólo por sus resultados en *TargetScan* con PCT por encima de 0,8.

Con estos criterios se llega a una lista de 51 genes *target*:

[1]	"B4galt5"	"Bcl11a"	"Bsn"	"Camk1d"	"Caprin1"	"Ccdc88a"	"Cdk5r1"	"Chod1"
[9]	"Coro1c"	"Dlg4"	"Efnb2"	"Emb"	"Epha7"	"Ephb4"	"Foxg1"	"Foxp1"
[17]	"Git1"	"Heyl"	"Lgr4"	"Lrrtm4"	"Lzts3"	"Map1a"	"Mdga1"	"Mdga2"
[25]	"Mecp2"	"Mmd"	"Mmd2"	"Mtmr2"	"Mypn"	"Nptx1"	"Nr3c1"	"Nrp1"
[33]	"Pafah1b1"	"Pak3"	"Pak4"	"Pitpna"	"Prex1"	"Prtg"	"Pten"	"Rab21"
[41]	"Rit2"	"Rnfl65"	"Tgif1"	"Tgif2"	"Tnc"	"Ttbk2"	"Twf1"	"Ust"
[49]	"Ywhaz"	"Zeb2"	"Zfyve27"					

El mapa bipartito, reducido a estos genes *target* es el siguiente (Fig. 41):

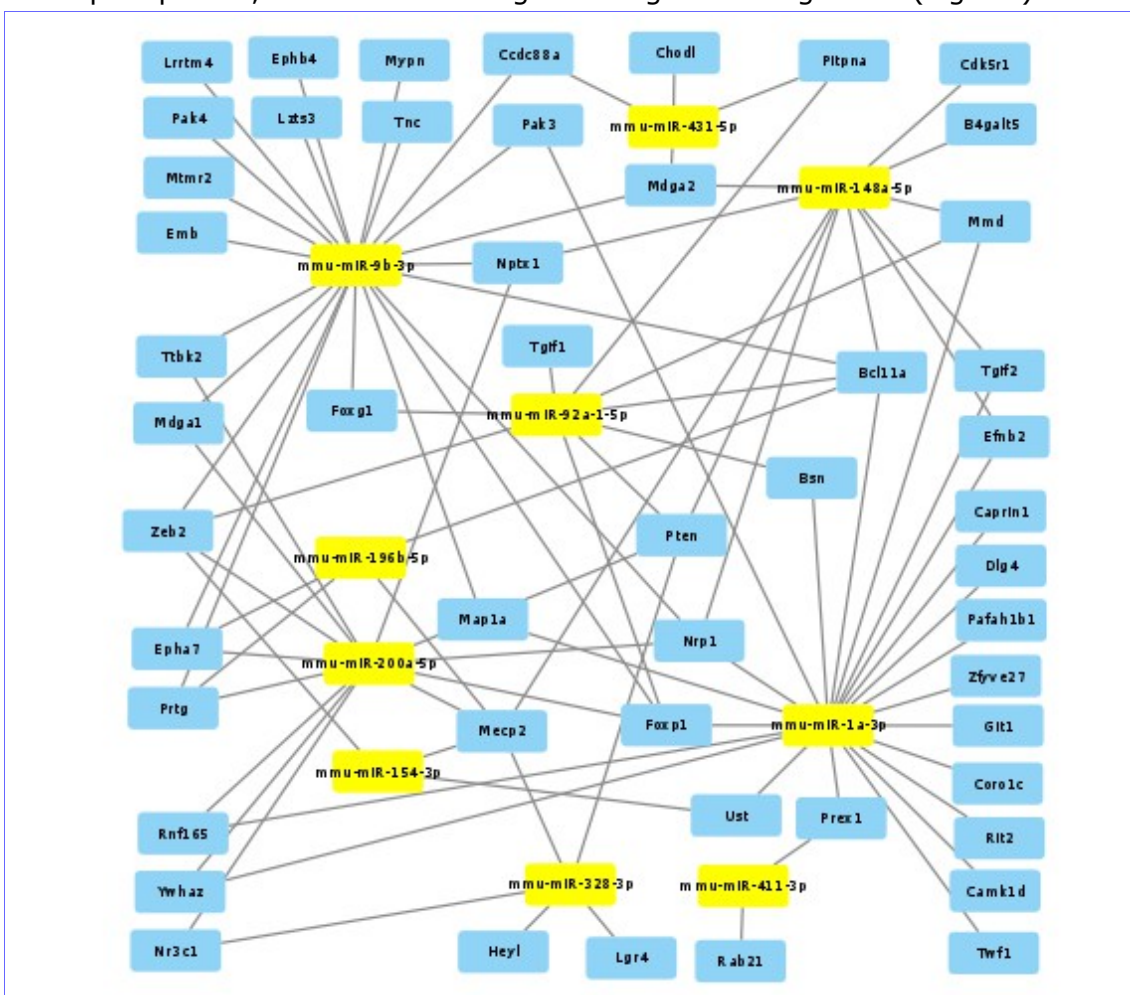


Figura 41. Mapa de red bipartita de miRNAs y genes *target* seleccionados como más significativos para el experimento 1 (ratones *Mecp2*<sup>+/-</sup>).

Con esta selección de 51 genes *target* se hace un análisis de *pathways* con el paquete *pathfindR* [31], 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.

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. El resultado de este análisis se muestra en el siguiente gráfico de enriquecimiento (Fig. 42).

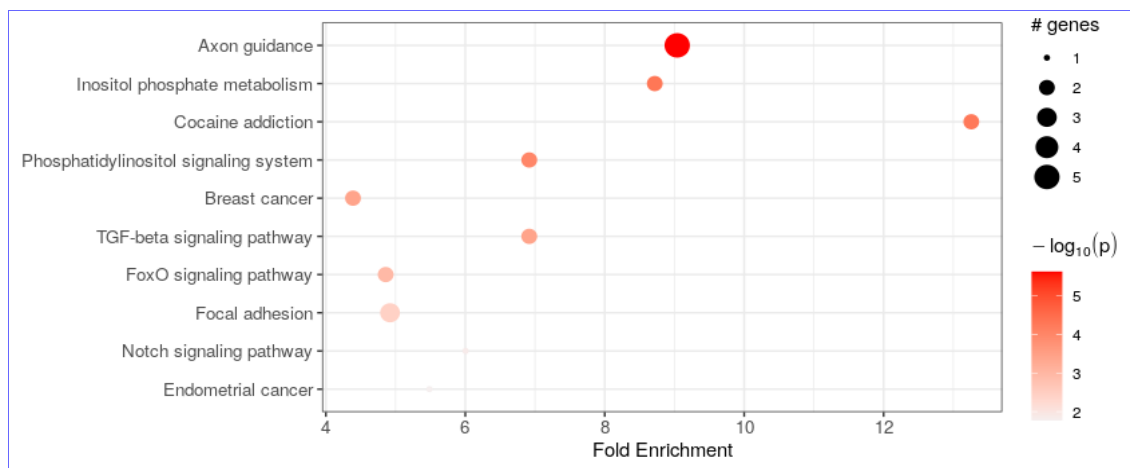


Figura 42. Gráfico de enriquecimiento obtenido por pathfindR para los genes target seleccionados de los miRNAs desregulados en el experimento 1.

La información completa de los *pathways* seleccionados por *pathfindR*, con sus genes relacionados, se encuentra en el punto 2.12.4 *Enriquecimiento KEGG por subredes activas*, del documento técnico.

Los *pathways* *Axon guidance*, que aparece como el más significativo y *Focal adhesion*, ya fueron comentados anteriormente. Con *Axon guidance* están relacionados los genes *Nrp1* (*up-reg*) y *Pak3*, *Efnb2*, *Epha7* y *Ephb4* (*dow-reg*). Y con *Focal adhesion*, *Tnc*, *Pten* y *Pak3* (*down-reg*). En el mapa de red bipartita, se puede ver que 5 de éstos genes hibridan con miR-9b-3p; por lo que este *miRNA* parece ser de especial relevancia en los resultados.

Para ver la función de cada gen se utiliza *GeneCards – the human gene database* [32] [www.genecards.org](http://www.genecards.org)

*Nrp1* codifica una neuropilina, que atúa como co-receptor del receptor tirosín quinasa para miembros de la familia de las semaforinas.

*Efnb2* codifica un miembro de las efrinas, mientras que *Epha7* y *Ephb4* codifican receptores de las efrinas. Ambas (*Eph*/efrinas) son proteínas de membrana que regulan varios procesos biológicos durante el desarrollo embrionario, incluyendo el guiado de los conos de crecimiento del axón.

*Pak3* pertenecen a la familia de proteínas PAK, que son efectores críticos que se unen a Rho GTPasas en la reorganización del citoesqueleto y la señalización nuclear.

*Tnc* codifica una proteína de la matriz extracelular implicada en la guía y migración de neuronas, así como del axón durante el desarrollo de la plasticidad sináptica y la regeneración neuronal.

*Pten* codifica una proteína con un dominio catalítico similar al de la tirosina fosfatasa, que participa en la regulación del *pathway* de señalización AKT/PKB. Aparecen como significativos otros *pathways*, como *TGF-beta signaling pathway*, que regula un amplio espectro de funciones celulares como proliferación, apoptosis, diferenciación y migración. Se encuentran implicados los genes *Tgif1* y *Tgif2*.

Y *FoxO signaling pathway*, relacionado con el anterior. La familia de factores de transcripción FOXO (*forkhead box O*) regula la expresión de genes en eventos fisiológicos que incluyen apoptosis, control del ciclo celular, metabolismo de la glucosa, resistencia al estrés oxidativo y longevidad. Se encuentran implicados los genes *Pten* y *Foxg1*.

#### 4.2 Análisis de resultados del experimento 2 (ratones Foxg1+/-).

En el enriquecimiento *KEGG* que se hizo para los datos del experimento 2, en el primer grupo de genes seleccionados se destacan los *pathways*:

ID	Descripción
mmu04727	<i>GABAergic synapse</i>
mmu04015	<i>Rap1 signaling pathway</i>
mmu04010	<i>MAPK signaling pathway</i>
mmu04151	<i>PI3K-Akt signaling pathway</i>
mmu04360	<i>Axon guidance</i>
mmu04014	<i>Ras signaling pathway</i>
mmu04068	<i>FoxO signaling pathway</i>
mmu04510	<i>Focal adhesion</i>

Para el segundo grupo se selecciona:

ID	Descripción
mmu04721	<i>Synaptic vesicle cycle</i>

Y para el tercero:

ID	Descripción
mmu04010	<i>MAPK signaling pathway</i>
mmu04151	<i>PI3K-Akt signaling pathway</i>

En *Rap1 signaling pathway*, *Rap1* es una pequeña GTPasa que controla distintos procesos, tales como adhesión celular, formación de uniones entre células y polaridad celular. Como todas las proteínas G, *Rap1* oscila entre dos configuraciones, una inactiva, unida a GDP y otra activa, unida a GTP. Ciertas señales extracelulares, como *GEFs* (*guanine exchange factors*) o *GAPs* (*GTPase activating proteins*), controlan este ciclo. *Rap1* juega un papel importante en interacciones célula a célula y de la matriz celular por regulación de la función de integrinas y otras moléculas de adhesión; también regula la actividad de la quinasa *MAP* (*MAPK*) de una manera altamente dependiente del tipo de célula.

Este *pathway* tiene relación con otros procesos seleccionados, como *MAPK signaling pathway*, *PI3K-Akt signaling pathway*, *Ras signaling pathway* y *Focal adhesion*.

Se observan también en estos grupos de genes *target*, *pathways* relacionados con procesos sinápticos, como *GABAergic synapse* y *Synaptic vesicle cycle*.

El *FoxO signaling pathway* se refiere a la familia de factores de transcripción *FOXO (forkhead box O)*, entre los que se encuentra *Foxg1*, elemento principal de este experimento. Hay otros *pathways* comunes al experimento 1 que ya se han mencionado, como *MAPK signaling pathway* y *PI3K-Akt signaling pathway*, que también están relacionados con *FoxO signaling pathway*.

Para acotar los resultados de los análisis se procede de forma similar a la descrita en el experimento 1 y se consideran los siguientes *miRNAs*:

*Up-reg*: miR-448-3p y miR-10a-5p.

*Down-reg*: 10b-5p, miR-429-3p, miR-200a-3p, miR-200b-3p.

Seleccionado los posibles genes *target* de estos *miRNAs* con unos parámetros de predicción más ajustados (valores en *TargetScan* de *SeedMatch* = 8mer y *PCT* > 0.8 y valor en *MicroT-CDS* de *miTG.score* > 0.9), se seleccionan los siguientes genes *target*, diferenciados por los grupos derivados del análisis de enriquecimiento GO.

#### 47 genes *target* del grupo 1:

[1]	"Ab12"	"Apbb2"	"Asap1"	"Auts2"	"Bcl2"	"Brinp1"	"Clasp2"	"Crtc1"	"Dlc1"
[10]	"Efnb2"	"Egln1"	"Epha4"	"Fbxw7"	"Foxa1"	"Foxg1"	"Gphn"	"Grip2"	"Hecw2"
[19]	"Igf1r"	"Ikzf1"	"Lrp1b"	"Map2"	"Map4k4"	"Mef2c"	"Mmd"	"Nfia"	"Nog"
[28]	"Pcdh8"	"Plxna4"	"Ptbp1"	"Ptprg"	"Qk"	"Rapgef2"	"Rtn4rl1"	"Sema6a"	"Six1"
[37]	"Slc1a2"	"Slitrk1"	"Smurf1"	"Syn1"	"Syt4"	"Tbx3"	"Thrb"	"Tubb5"	"Unc5a"
[46]	"Zeb1"	"Zeb2"							

#### 20 genes *target* del grupo 2:

[1]	"Cacna1c"	"Dagla"	"Doc2a"	"Efnb2"	"Git2"	"Grip2"	"Kcnb1"	"Kcnc3"	"Map2"
[10]	"Mef2c"	"P2ry1"	"Rap1b"	"Slc30a1"	"Slc6a1"	"Slc6a11"	"Stx2"	"Syde1"	"Syn1"
[19]	"Synj1"	"Syt4"							

#### 18 genes *target* del grupo 3:

[1]	"Ab12"	"Adrb1"	"Bcl2"	"Brinp1"	"Cacna1c"	"Crtc1"	"Dpp4"	"Grik2"	"Jun"
[10]	"Kmt2a"	"Mef2c"	"Nog"	"Prkar2b"	"Rps6kb1"	"Slc6a1"	"Slc6a11"	"Synj1"	"Syt4"

Para un análisis completo con *pathfindR*, se prepara el *dataset* de entrada con los genes conjuntos de los tres grupos, considerando que los *miRNAs down-regulated* provocarán que sus genes *target* estén sobre-expresados y que los genes *target* de los *miRNAs up-reg* estarán infra-expresados.

El resultado de este análisis se muestra en el siguiente gráfico de enriquecimiento (Fig. 43). La información completa de los *pathways* seleccionados con sus genes relacionados, se encuentra en el punto 2.12.4 *Enriquecimiento KEGG por subredes activas*, del documento técnico.

*Rap1 signaling pathway* aparece como *pathway* más significativo y ya fue mencionado en los resultados preliminares de este experimento. En él están implicados los genes *Igf1r (up-reg)* y *Rap1b*, *P2ry1* y *Rapgef2 (down-reg)*.

*Igf1r (insulin-like growth factor I)*. Este receptor tiene actividad tirosín-quinasa y juega un papel importante en eventos de transformación.

*Rap1b* pertenece a una familia semejante a *RAS*, que regulan múltiples procesos celulares, incluyendo adhesión celular, crecimiento y diferenciación. *P2ry1* pertenece a la familia de receptores acoplados a proteínas G. *Rapgef2* es un miembro de la subfamilia *RAS* con función GTPasa en señales transductoras. Actúa como factor de intercambio del nucleótido guanina (*GEFs*, *Guanine nucleotide exchange factors*), sirviendo de enlace entre los receptores de superficie celular y la activación de *RAS*.

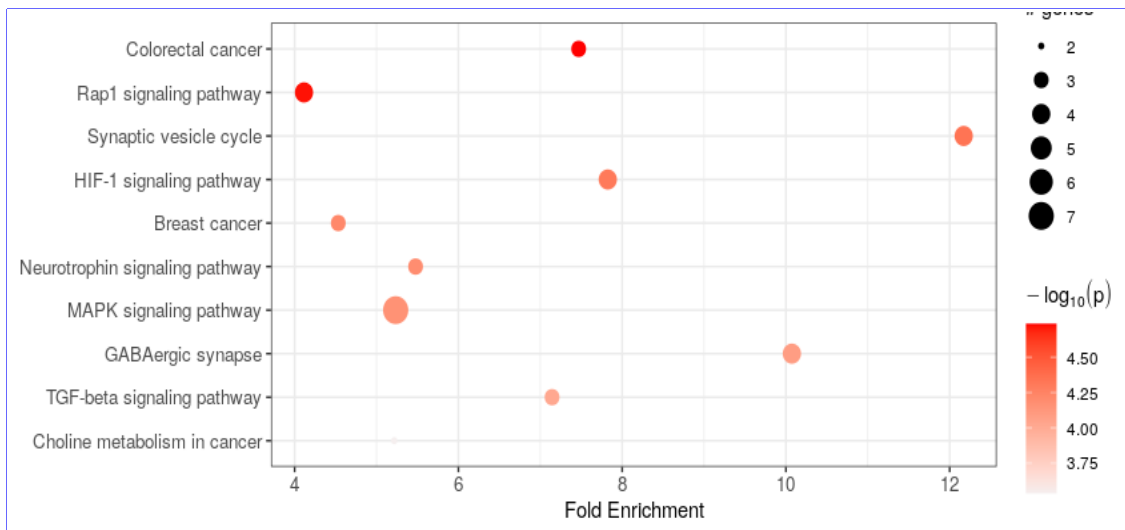


Figura 43. Gráfico de enriquecimiento obtenido por pathfindR para los genes target seleccionados de los miRNAs desregulados en el experimento 2.

También son significativos los *pathways* *Synaptic vesicle cycle*, con los genes implicados *Stx2*, *Slc6a1*, *Slc6a11* y *Slc1a2* (*up-reg*). Y *GABAergic synapse*, con los genes implicados *Slc6a1* y *Slc6a11* (*up-reg*) y *Gphn* y *Cacna1c* (*down-reg*). *Stx2* codifica una sintaxina, proteína implicada en la atracción y fusión de vesículas de transporte en células epiteliales.

*Slc6a1* y *Slc6a11* codifican transportadores de GABA (ácido gamma-aminobutírico; un neurotransmisor) en la sinapsis. Y *Slc1a2* codifica un transportador de glutamato, otro neurotransmisor.

*Cacna1c* codifica un canal de calcio, que media el flujo de iones calcio hacia el interior de la célula a través de la polarización de membrana.

*Gphn* codifica una proteína que ancla receptores de neurotransmisores inhibidores en el citoesqueleto post-sináptico.

*HIF-1 signaling pathway*. El factor 1 inducido por hipoxia (*HIF-1*) es un factor de transcripción que funciona como un regulador maestro de la homeostasis de oxígeno. En este *pathway* están implicados los genes *Igf1r* y *Bcl2* (*down-reg*) y *Rps6kb1* y *Egln1* (*up-reg*). *Igf1r* ya se mencionó como desregulado en el *Rap1 signaling pathway*.

*Bcl2* es un regulador de la apoptosis.

*Rps6kb1* codifica un miembro de la familia de serín/treonín quinasas S6, que responde a la señalización *mTOR* (*mammalian target of rapamycin*) promoviendo síntesis de proteína, crecimiento y proliferación celular.

*Egln1* induce cambios en las proteínas *HIF*.

*Neurotrophin signaling pathway*. Las neurotrofinas son una familia de factores tróficos implicados en la diferenciación y supervivencia de células neuronales.

Genes implicados en este *pathway* son *Bcl2n(down-reg)* y *Rap1b* y *Jun (up-reg)*.

*MAPK signaling pathway*. La cascada *MAPK (mitogen-activated protein kinase)* es un módulo altamente conservado implicado en varias funciones celulares como proliferación, diferenciación y migración. Genes implicados en este *pathway* son *Cacna1c*, *Igf1r* y *Mef2c (down-reg)* y *Rapgef2*, *Rap1b*, *Map4k4* y *Jun (up-reg)*.

*TGF-beta signaling pathway* regula un amplio espectro de funciones celulares como proliferación, apoptosis, diferenciación y migración. Se encuentran implicados los genes *Smurf1 (down-reg)* y *Nog* y *Rps6kb1 (up-reg)*.

Aunque con menos significación, también aparecen como posiblemente desregulados los *pathways Focal adhesion*, con los genes implicados *Igf1r* y *Bcl2 (down-reg)* y *Rap1b* y *Jun (up-reg)*. Y *Axon guidance*, con los genes *Unc5a* y *Epha4 (down-reg)* y *Efnb2*, *Plxna4* y *Sema6a (up-reg)*. Que fueron los *pathways* más significativos en el análisis de datos del experimento 1.

#### 4.3 Análisis de resultados del cruce de los 2 experimentos.

En los genes *target* seleccionados en el cruce de los 2 experimentos, se observan los siguientes *pathways*:

ID	Descripción
mmu04151	<i>PI3K-Akt signaling pathway</i>
mmu04010	<i>MAPK signaling pathway</i>
mmu04015	<i>Rap1 signaling pathway</i>
mmu04810	<i>Regulation of actin cytoskeleton</i>
mmu04510	<i>Focal adhesion</i>
mmu04360	<i>Axon guidance</i>

Para acotar los resultados de los análisis se procede de forma similar a la descrita anteriormente y se consideran los siguientes *miRNAs* desregulados en ambos experimentos:

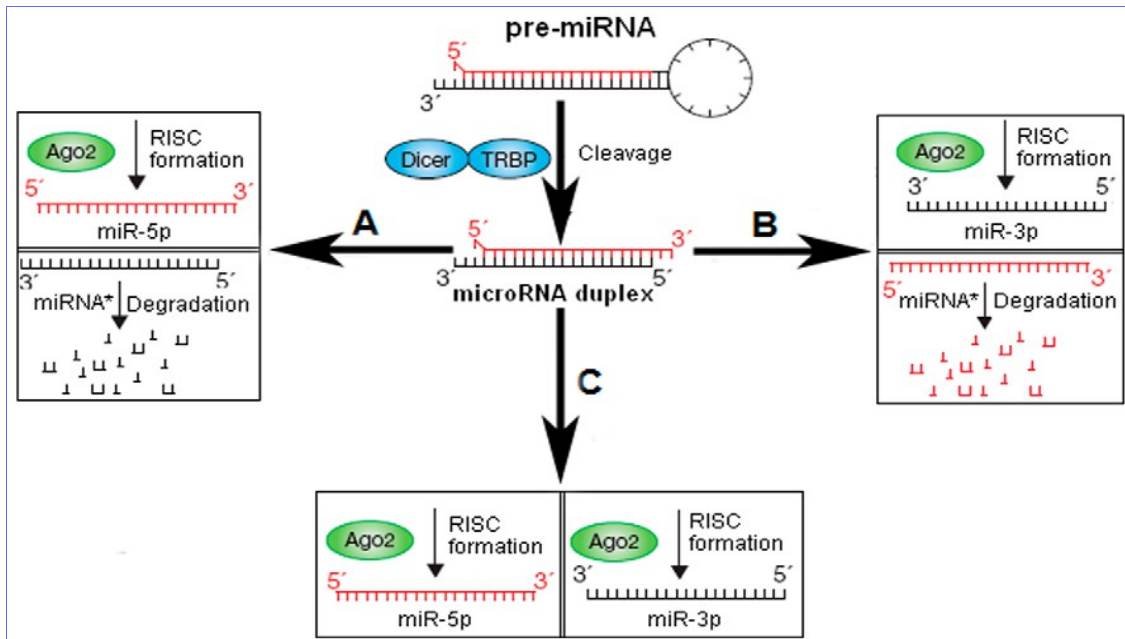
*Up-reg*: miR-148a-5p (-3p), miR-431-5p (-3p), miR-92a-3p y miR-92a-1-5p.

*Down-reg*: miR-200a-5p (-3p) y miR-200b-5p (-3p).

En el experimento 1 se presentan normalmente las variedades -5p de estos *miRNAs* como DE y en el experimento 2 las variedades -3p.

En el proceso de maduración del *miRNA*, el precursor, una vez procesado por la endonucleasa Droscha en el núcleo, migra al citoplasma donde su estructura en horquilla (*hairpin*) se corta en posiciones relativamente fijas por la enzima Dicer; este corte produce un fragmento corto de *RNA* de doble cadena. Aunque las dos cadenas complementarias se producen en la misma cantidad, su presencia no es la misma en el estado estable. Debido a la estabilidad termodinámica de cada extremo del dúplex, una de las hebras se degrada y la otra pasa a ser la cadena activa que forma el *miRNA* maduro que se une al complejo *RISC* por mediación de proteínas Ago, que estabilizan su estructura.

La cadena activa puede ser la del extremo 5' o 3', denominándose *miRNA-5p* o *miRNA-3p* en cada caso; mientras que la cadena degradada se denomina *miRNA\** (*star*) o pasajera (Fig. 44).



(Figura 44. Selección de la cadena de miRNA madura. A) La cadena 5' forma el miRNA-5p maduro y la cadena 3' (*miRNA\**) se degrada. B) La cadena 3' forma el miRNA-3p maduro y la cadena 5' (*miRNA\**) se degrada. C) Ambas cadenas miRNA-5p y miRNA-3p son maduras y forman complejos RISC de inhibición post-transcripcional de sus genes target [33].

Sin embargo, recientemente, se ha observado la presencia de *miRNA\** como forma de *miRNA* maduro activa; con unas expresiones abundantes y con una proporción de *miRNA/miRNA\** que puede variar según el estado de desarrollo. En estos casos no se habla de *miRNA\** o cadena pasajera, sino que se consideran las 2 formas, *miRNA-5p* y *miRNA-3p* como activas [33].

En el presente trabajo, aparecen las 2 formas de *miRNA* maduro como funcionales en la mayoría de los *miRNAs* DE seleccionados. Esta diferencia no parece ser debida al estado de desarrollo, las muestras proceden de ratones de 30 días en el experimento 1 y de 6 semanas en el 2; ni tampoco por el tipo de tejido, hipocampo en ambos experimentos. Se comprueba en los resultados de *TargetScan* que ambas formas hibridan en los mismos genes *target* con valores de parámetros de ajuste (SeedMatch y PCT) muy similares.

Otro punto importante a tener en cuenta es que los *miRNAs* seleccionados en el cruce de los dos experimentos tienen el mismo sentido (*down-reg* o *up-reg*) en ambos. Inicialmente se identificaron *miRNAs* comunes que se expresaban como *down-reg* en el primer experimento y como *up-reg* en el segundo (miR1298, miR212, miR34b/c), pero que no son demasiado significativos por los valores de pfp y p.value resultantes en el análisis por *RankProd* y no han sido tenidos en cuenta en esta comparación.

Seleccionando los posibles genes *target* de estos *miRNAs* con unos parámetros de predicción más ajustados (valores en *TargetScan* de SeedMatch = 8mer y PCT > 0.8 y no se consideran predicciones en *MicroT-CDS*), se llega al siguiente listado de 85 genes, sin diferenciar en grupos por categorías GO:

[1]	"Abl2"	"Acs14"	"Arid1b"	"Asap1"	"Atp7a"	"Axl"
[7]	"B4gal5"	"Bcl11a"	"Braf"	"Btd3"	"Btg2"	"Cdk5r1"
[13]	"Chod1"	"Clasp2"	"Cntn2"	"Cntnap2"	"Cpeb3"	"Csflr"
[19]	"Cul5"	"D16Ert472e"	"Dcx"	"Dixdc1"	"Dlcl"	"Dock5"
[25]	"Dpysl4"	"Efnb2"	"Egfr"	"Egln1"	"Elavl4"	"Farp1"
[31]	"Fmr1"	"Foxa1"	"Foxg1"	"Gata2"	"Hecw2"	"Kalrn"
[37]	"Kif3a"	"Lhx8"	"Lrp1"	"Map1b"	"Map2"	"Map2k4"
[43]	"Map4k4"	"Mdga2"	"Mecp2"	"Mmd"	"Mycbp2"	"Mylip"
[49]	"Ndst1"	"Nfia"	"Nog"	"Nptx1"	"Nrcam"	"Nrpl"
[55]	"Numb1"	"Pclo"	"Pip5k1c"	"Pitpna"	"Plaa"	"Plxna2"
[61]	"Plxna4"	"Pten"	"Ptprd"	"Ptprg"	"Ptpro"	"Qk"
[67]	"Rapgef2"	"Rtn4r11"	"Sema6a"	"Shank3"	"Six1"	"Slc12a5"
[73]	"Slc1a2"	"Slit2"	"Slitrk1"	"Sox11"	"Spag9"	"Tgif2"
[79]	"Thrb"	"Trim67"	"Tsc1"	"Tubb5"	"Was1"	"Zeb1"
[85]	"Zeb2"					

Se hace el análisis con *pathfindR*. El resultado se resume en el diagrama de enriquecimiento (Fig. 45). Los resultados completos están en el punto 2.12.4 *Enriquecimiento KEGG por subredes activas*, del documento técnico.

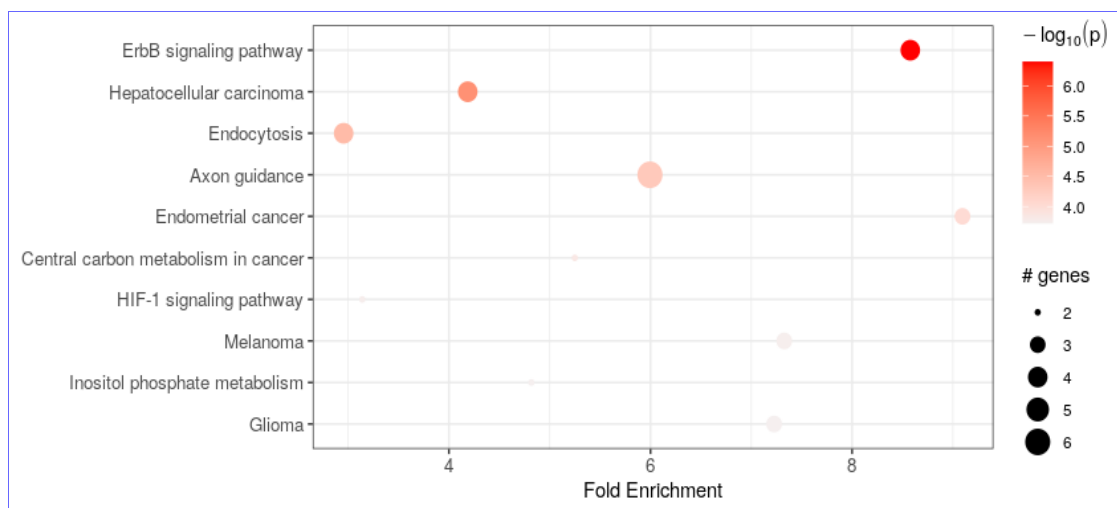


Figura 45. Gráfico de enriquecimiento obtenido por *pathfindR* para los genes target seleccionados de los miRNAs desregulados en el cruce de ambos experimentos.

Aparece como más significativo el *ErbB signaling pathway*, con los genes implicados *Abl2* (*up-reg*) y *Egfr*, *Map2k4* y *Braf* (*down-reg*).

*Abl2* codifica un miembro de la familia Abelson, de no-receptores de tirosín quinasa. Juega un papel importante en la reorganización del citoesqueleto.

La proteína codificada por *Egfr* es una glicoproteína transmembrana, miembro de la superfamilia de proteínas quinasa, que actúa como receptor de factor epidérmico de crecimiento (*EGFR*, *epidermal growth factor*).

*Map2k4* codifica un miembro de la familia *MAPK* (*mitogen-activated protein kinase*), que actúa como punto de integración para múltiples señales bioquímicas implicadas en una amplia variedad de procesos celulares, tales como proliferación, diferenciación, regulación de la transcripción y desarrollo.

*Braf* codifica una proteína de la familia *RAF* (*serine/threonine protein kinases*), que juega un papel importante en la regulación del *MAP kinase/ERK signaling pathway*, que afecta a la división celular, diferenciación y secreción.

*ErbB signaling pathway* ya apareció como significativo en el análisis de resultados del experimento 2. Miembros de la familia *ErbB* de receptores tirosín quinasa (*RTKs*) se acoplan con factores de crecimiento extracelulares ligados a

rutas de señalización intracelular que regulan diversos procesos biológicos que incluyen proliferación, diferenciación, motilidad celular y supervivencia. Este *pathway* tiene conexiones con *PI3K-Akt signaling pathway*, señalado anteriormente como significativo en ambos experimentos.

En los resultados de *pathways* significativos en el cruce de ambos experimentos también aparece *Axon guidance*, con los genes implicados *Efnb2*, *Slit2*, *Plxna4* y *Sema6a* (*up-reg*) y *Plxna2* y *Nrp1* (*down-reg*).

*HIF-1 signaling pathway* también aparece como significativo en el cruce de los dos experimentos, con los genes implicados *Egln1* (*up-reg*) y *Egfr* (*down-reg*).



## 5. Conclusiones

En el primer experimento se analizaron secuencias de *miRNAs* en ratones *Mecp2*<sup>+/-</sup>, comparadas con las de ratones normales. En el estudio de referencia [3] se establece que *MeCP2* ejerce una acción supresora en el procesamiento de *miRNA*. En los ratones *Mecp2* mutantes, la baja presencia de esta proteína hace que aumenten los niveles de ciertos *miRNAs* (*up-reg*); entre los que se destacan miR-134, miR-383, miR-382 y miR-182; haciendo especial referencia al miR-134 por su actividad reguladora de ciertos genes *target* críticos en el desarrollo neuronal y la plasticidad, como *CREB*, *LIMK1* y *Pumilio2*.

Sin embargo, en este trabajo no se ha constatado la desregulación de estos *miRNAs* con los datos de secuenciación de *miRNAs* disponibles y la metodología empleada. En su lugar, se han identificado como *miRNAs* DE destacados los siguientes:

*Down-reg*: miR-200a

*Up-reg*: miR-328, miR-1a, miR-92a-1, miR-411, miR-9b y miR-148a

Se evidencia una mayor presencia de *miRNAs up-reg* por la disminución de actividad supresora en el procesamiento de *miRNAs* que ejerce *MeCP2*.

Otro dato importante es que se manifiesta la presencia de miR-200a como *down-reg*, que es uno de los *miRNAs* más significativos señalados en el segundo experimento analizado con ratones *Foxg1*<sup>+/-</sup>.

El análisis de posibles genes *target* de estos *miRNAs* DE, ha llevado a determinar posibles rutas metabólicas (*pathways*) alteradas por la acción reguladora de estos *miRNAs* cuya maduración ha podido ser desestabilizada por una baja expresión de la proteína *MeCP2*.

Entre estos *pathways* destaca el guiado del axón (*axon guidance*), con posibles genes *target* implicados como *Nrp1*, que codifica el neuropilin, un receptor de semaforinas, asociado al neuroma y *Efnb2*, que codifica un miembro de la familia de las efrinas y aparece también como gen posiblemente desregulado en los datos del segundo experimento. Los datos de predicción de *targets* apuntan a que estos 2 genes se puedan hallar regulados principalmente por el miR-1a, aunque *Nrp1* aparece también como regulado por miR-200a.

Otros posibles genes *target* del miR-1a son *Tgif1* y *Tgif2*; implicados en el *pathway* de señalización de *TGFβ* (*Transforming Growth Factor β*), que aparece también como posiblemente desregulado en el análisis de datos del segundo experimento. Estos genes codifican miembros de la familia *homeobox TALE* (*three-amino acid loop extension*), que son reguladores de la transcripción altamente conservados y actúan como co-represores activos de *SMAD2*, participando en la transmisión de señales nucleares durante el desarrollo y en el adulto. Mutaciones en estos genes están asociadas a la holoprosencefalia tipo 4, una anomalía estructural del cerebro.

Otro *pathway* posiblemente desregulado es el de señalización *FoxO*, en el que están implicados los genes *Pten*, que ejerce una regulación negativa en el *pathway* de señalización *AKT/PKB* y está asociado con el autismo y *Foxg1*, que es el principal gen implicado en la variante congénita del síndrome de Rett,

estudiada en el experimento 2. El gen *Pten* puede verse regulado principalmente por miR-92a-1 o por mir-148a; mientras que las predicciones muestran que *Foxg1* puede verse regulado por miR-9b.

Los *miRNAs* mencionados anteriormente, miR-1a, miR-92a-1, miR-148a y miR-9b se han manifestado como *up-reg* en el experimento 1, por lo que se entiende que habrá una represión en la expresión de sus genes *target*. Si bien *Nrp1* y *Pten* (con una predicción menos ajustada) también aparecen como *targets* del miR-200a, que se encuentra *down-reg*; por lo que se supone una gran complejidad en estas redes reguladoras de expresión genética por parte de *miRNAs*.

Otro elemento que llama la atención en los resultados del análisis de este primer experimento es que el propio gen *Mecp2*, que se encuentra con bajos niveles en ratones *Mecp2+/-*, se predice como gen *target* de ciertos *miRNAs* desregulados en estos ratones, concretamente las predicciones más ajustadas apuntan a miR-328 (*up-reg*) y con menor evidencia a miR-3084 (no predicho por *TargetScan*, pero sí con un ajuste alto por *microT-CDS*) y miR-148 (ambos *up-reg*) y con mucha menos evidencia a miR-200a (*down-reg*). Por lo que se puede pensar que los bajos niveles de proteína *MeCP2* en estos ratones, se verán deprimidos aún más por efecto de los *miRNAs* desregulados a causa de la mutación.

Los posibles *pathways* desregulados con sus genes *target* más significativos se resumen en la siguiente tabla (Tabla 2).

Tabla 2. Genes target con predicciones más ajustadas de los *miRNAs* desregulados más significativos en el experimento 1 (ratones *Mecp2+/-*).

pathways	MiRNAs down-reg: miR200a	MiRNAs up-reg: miR-328, miR-1a, miR-92a, miR411, miR-9b y miR-148a
Axon guidance	( <i>Nrp1</i> )	<i>Efnb2</i> <i>Pak3</i> <i>Epha7</i> <i>Ephb4</i> ( <i>Nrp1</i> )
PI sig. system		<i>Pten</i> <i>Mtmr2</i>
Focal adhesion		<i>Pten</i> <i>Pak3</i> <i>Tnc</i>
TGF-beta sig. p.		<i>Tgif1</i> <i>Tgif2</i>
FoxO sig. path.		<i>Pten</i> <i>Foxg1</i>

En el segundo experimento se analizaron secuencias de *miRNAs* en ratones *Foxg1+/-*, comparadas con las de ratones normales. En el estudio de referencia [4] se establece que *FOXG1* afecta a la expresión de *PRKAR2B* a múltiples niveles, por un lado por represión transcripcional directa y por otro lado a través de la familia miR200 que tiene por objetivo (*target*) directo los transcritos de *PRKAR2B*. Sugiriendo que ambos, *FOXG1* y la familia miR200, forman parte de una red multinivel que equilibra la expresión de la subunidad reguladora de *PKA* (*protein kinase A*), *PRKAR2B*. De este modo, *FOXG1* y *MECP2* pueden afectar un *pathway* común dependiente de *PKA* para ajustar la función neuronal en el hipocampo.

Los análisis realizados en el presente trabajo muestran un grupo de *miRNAs* diferencialmente expresados en ratones *Foxg1+/-* que coinciden en gran

medida con los presentados en el estudio de referencia, si bien el gen *Prkar2b* no se ha detectado como target de la familia miR200 (*down-reg*), sino del miR-448 (*up-reg*), que muestra unos valores muy ajustados de predicción con las herramientas *TargetScan* y *microT-CDS* y valores mucho más débiles para la hibridación de *Prkar2b* con miR-200b y miR-429 (de la misma familia).

En el presente trabajo 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 (Tabla 3):

Tabla 3. Genes target con predicciones más ajustadas de los miRNAs desregulados más significativos en el experimento 2 (ratones Foxg1+/-).

pathways	MiRNAs down-reg: miR429,miR200b, miR200a	MiRNAs up-reg: miR448
Axon guidance	Efnb2 Plxna4 Sema6a1	Epha4 Unc5a
Synaptic vesicle	Stx2 Slc6a1 Slc6a11 Slc1a2	
GABAergic syn.	Slc6a1 Slc6a11	Gphn Cacna1c
MAPK sig. path.	Jun Rap1b Rapgef2 Map4k4	Igf1r Cacna1c Mef2c
ErbB sig. path.	Jun Rps6kb1 Abl2	
Rap1 sig. path.	Rap1b Rapgef2 P2ry1	Igf1r
Neurotrophin s.p.	Jun Rap1b	Bcl2
HIF-1 sig. path.	Rps6kb1 Egln1	Igf1r Bcl2
TGF-beta sig. p.	Rps6kb1 Nog	Smurf1
Insulin sig. p.	Rps6kb1	Prkar2b

En primer lugar el *pathway Axon guidance*, con el gen *Efnb2* como posible *target*, común al experimento 1 y el gen *Epha4* como codificador de efrina, semejante a los predichos para el experimento 1.

En segundo lugar los *pathways Synaptic vesicle cycle* y *GABAergic synapse*, con genes target comunes *Slc6a1* y *Slc6a11* y el gen *Cacna1c* común con el siguiente *pathway*.

En el tercer bloque, destaca el *MAPK signaling pathway*, con el que están relacionados el resto de *pathways* y que presenta varios genes target comunes con ellos: *Jun*, *Rap1b*, *Rapgef2* e *Igf1r*. Los siguientes *pathways* también presentan algunos genes *target* comunes como *Rps6kb1* y *Bcl2*.

La cascada *MAPK (mitogen-activated protein kinase)* es un módulo altamente conservado implicado en varias funciones celulares como proliferación, diferenciación y migración. En mamíferos, se expresan al menos cuatro grupos regulados distintos de *MAPKs*: quinasas relacionadas con señales extracelulares (*ERK, extracellular signal-related kinases*)-1/2, quinasas amino-terminales *Jun* (*JNK, Jun amino-terminal kinases*), proteínas *p-38* y *ERK5*.

En la figura correspondiente al *MAPK signaling pathway* (Fig. 46) se han marcado en azul los 7 genes target más significativos detectados. *Cacna1c* codifica el canal de calcio, *Igf1r* el receptor de membrana de factor de crecimiento (*GF*), *Rapgef2* y *Rap1b* actúan como receptores del segundo mensajero *cAMP* para activar la cadena de señalización de proteínas *Ras/Raf*, todos ellos en la ruta clásica de la *MAPK (ERK-1/2)*. Los genes *Map4k4*, que codifica una *MAPKKKK* y *Jun* el factor de transcripción *AP1*; ambos en la cadena de señalización *JNK*. Y el gen *Mef2c*, que codifica un factor potenciador de la transcripción (*enhancer*) en la vía de las proteínas *p-38*.

Este *pathway* se encuentra relacionado con *cAMP signaling pathway*, que aparece también como posiblemente desregulado en el análisis hecho con *pathfindR* de los datos del experimento 2, con 4 de sus genes target relacionados: *Rap1b*, *Jun* y *Cacna1c*, más el gen *Adrb1*, que codifica un receptor adrenérgico que media los efectos fisiológicos de la hormona epinefrina y del neurotransmisor norepinefrina y está relacionado con síndrome de sueño corto, una de las características de *tRTT* y *atRTT*. En la figura que resume este *pathway* se pueden ver los lugares de actuación de estos 4 genes. El *cAMP* actúa como segundo mensajero y su formación está promovida por activación de la adenilato ciclasa (*AC*) después de la unión de recetores acoplados a proteína G (*GPCRs*) con ligandos que incluyen hormonas, neurotransmisores y otras moléculas de señalización. *cAMP* regula procesos fisiológicos cruciales (Fig. 47) y actúa sobre tres objetivos (*targets*) principales: proteína quinasa a (*PKA*), proteína de intercambio activada por *cAMP (Epac)* y canales cíclicos de iones cerrados por nucleótidos (*CNGCs*). *PKA* modula, vía fosforilación, diversos sustratos celulares, entre los que se encuentran factores de transcripción, como *CREB*, uno de los genes señalados en el estudio del experimento 1 como *target* del miR-134, que a su vez modula la transcripción del gen *Jun*. Finalmente, el producto del gen *Jun* es uno de los que intervienen en el proceso de hiperexcitabilidad de muerte o supervivencia en células del hipocampo.

Como ya se ha comentado en un análisis inicial, no se ha detectado el gen *Prkar2b* como target de miR-200 en algún proceso de ajuste de la función neuronal mediado por *PKA*. Sin embargo, se acusa su presencia, junto con *Rps6kb1*, en el *Insulin signaling pathway* probablemente desregulado en el análisis de los datos del experimento 2.

Uno de los efectos de *cAMP* es la activación de la proteína quinasa dependiente de *cAMP*. La quinasa inactiva es un tetrámero compuesto de dos subunidades reguladoras unidas a dos subunidades catalíticas. *cAMP* provoca la disociación de la holoenzima inactiva en un dímero de subunidades reguladoras y dos unidades monoméricas catalíticas libres. La proteína codificada por *Prkar2b* es una de las subunidades reguladoras de la proteína quinasa dependiente de *cAMP*. Se ha comprobado que esta subunidad puede ser fosforilada por la actividad de la subunidad catalítica e interactúa con el factor de transcripción *CREB (cAMP responsive element binding protein)*, suprimiendo su actividad.

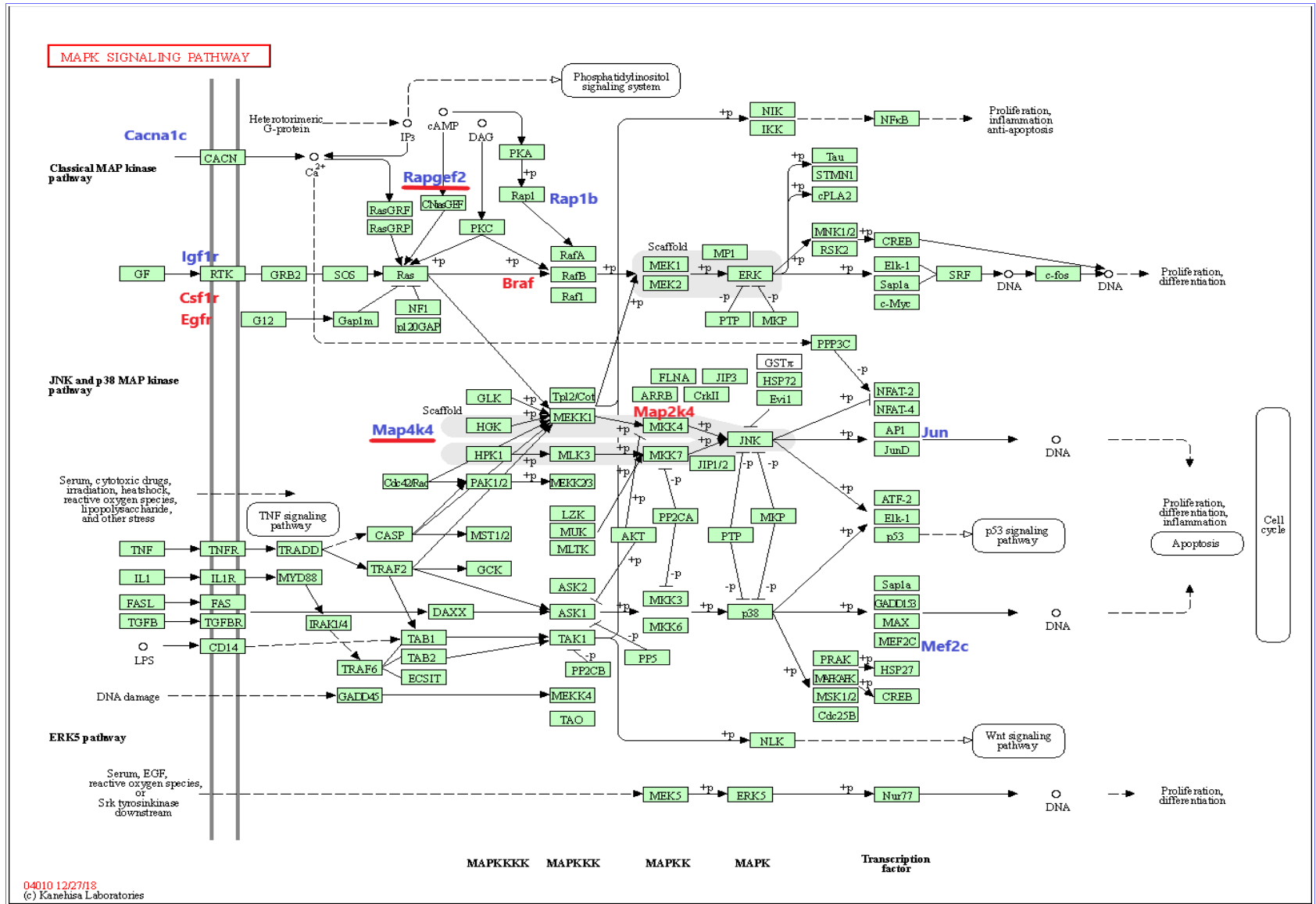


Figura 46. Esquema del pathway KEGG mmu04010, MAPK signaling pathway [30]

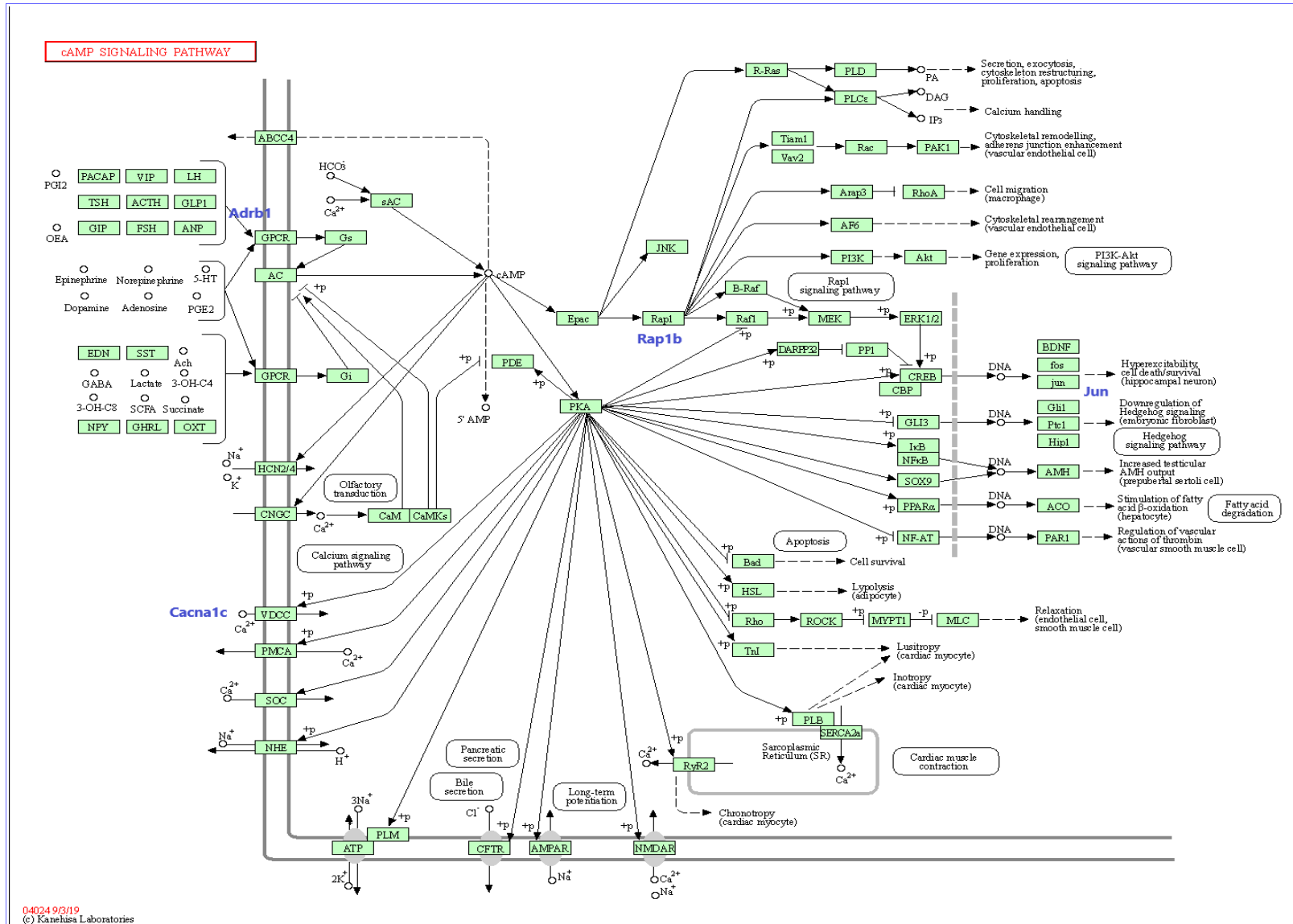


Figura 47. Esquema del pathway KEGG mmu04024, cAMP signaling pathway [30]

En la predicción de targets realizada en este trabajo, *Prkar2b* aparece relacionado con miR-448 (*up-reg*) con un grado de predicción elevado y con menos evidencia con los miR-200b y miR-429 (*down-reg*).

*Rps6kb1* ya se mencionó como miembro de la familia de serín/treonín quinasas S6 , presente también en *ErbB signaling pathway*, *HIF-1 signaling pathway* y *TGF-beta signaling pathway*, como *pathways* probablemente desregulados.

El receptor *IGF1R* (*insulin-like growth factor 1 receptor*) tiene una afinidad alta por *IGF1* y débil por *IGF2* e insulina. Su activación está implicada en el crecimiento celular y en el control de la supervivencia. La unión al ligando activa la receptora quinasa, conduciendo a la autofosforilación y la tirosina fosforilación de múltiples sustratos que funcionan como adaptadores de la señalización, incluyendo los sustratos receptores de insulina (*IRS1/2 insulin-receptor substrates*). La fosforilación de las proteínas *IRSs* conducen a la activación de dos rutas de señalización importantes: *PI3K-AKT/PKB pathway* y *Ras-MAPK pathway*. El resultado de la activación de *MAPK pathway* es el incremento de la proliferación celular, mientras que la activación del *PI3K pathway* inhibe la apoptosis y estimula la síntesis de proteínas. *PI3K-Akt signaling pathway* ha aparecido en el análisis como posiblemente desregulado en ambos 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 (Tabla 4):

Tabla 4. Genes target con predicciones más ajustadas de los *miRNAs* desregulados más significativos en el cruce de los 2 experimentos.

pathways	MiRNAs down-reg: miR200b, miR200a	MiRNAs up-reg: miR92a, miR148a, miR431
Axon guidance	Efnb2 Slit2 Plxna4 Sema6a	Plxna2 Nrpl
Focal adhesion		Egfr Braf Pten Pip5k1c
PI sig. system		Pten Pip5k1c
MAPK sig. path.	Rapgef2 Csf1r Map4k4	Egfr Braf Map2k4
ErbB sig. path.	Abl2	Egfr Braf Map2k4
HIF-1 sig. path.	Egln1	Egfr
FoxO sig. path.	Foxg1	Egfr Braf Pten
Insulin sig. p.		Braf Tsc1
Rap1 sig. path.	Rapgef2 Csf1r	Egfr Braf
TGF-beta sig. p.	Nog Tgif2	

El *pathway Axon guidance* parece desregulado en ambos experimentos con la presencia del gen target *Efnb2* en los dos, aunque con un signo de regulación

probablemente distinto. En el primer experimento estaría regulado principalmente por miR-1a (*up-reg*) y posiblemente infra-expresado y en el segundo experimento por miR-200b y miR-429 (*down-reg*) y por tanto sobre-expresado.

También aparece el *pathway Phosphatidylinositol signaling system* como posiblemente desregulado. El fosfatidilinositol es un componente fosfolipídico de la membrana celular que actúa como sustrato de múltiples proteínas de señalización. Los genes target *Egfr*, *Braf* y *Pten* aparecen también en otros *pathways* de señalización celular.

Entre ellos destaca *MAPK signaling pathway*, ya mencionado en los dos experimentos como *pathway* presente en muchas de las rutas metabólicas probablemente desreguladas por ambas condiciones experimentales. En este *pathway*, además de *Egfr* y *Braf*, aparecen implicados los genes target *Rapgef2* y *Csf1r*, también presentes en *Rap1 signaling pathway*. En la figura del *MAPK signaling pathway* se han marcado en rojo los 6 genes target que aparecen como posiblemente desregulados en el cruce de datos de los 2 experimentos.

El gen *Csf1r* no ha sido mencionado anteriormente. Codifica un receptor transmembrana tirosín quinasa. Se han asociado mutaciones de este gen con una predisposición a enfermedades mieloides. Está implicado en anomalía cerebral, neurodegeneración y disostesclerosis (BANDDOS).

La desregulación de *FoxO signaling pathway* también aparece en ambos experimentos, con la presencia del gen *Foxg1*, en el que se localiza la mutación de estudio del experimento 2. Como ya se comentó, según el estudio de referencia del segundo experimento [4], la regulación del ciclo celular por *FOXG1* está mediada por su unión al complejo *FOXO/SMAD*, que antagoniza los efectos de la señalización *TGF $\beta$*  (*Transforming Growth Factor  $\beta$* ), otro de los *pathways* posiblemente desregulado en ambos experimentos, con la presencia de los genes target *Nog* y *Tgif2*.

La actividad del gen *Tgif2* ya fue mencionada en el análisis de los datos del experimento 1. El polipéptido codificado por *Nog* se une a miembros de la superfamilia de proteínas señalizadoras *TGF $\beta$* , inactivándolas. Su principal función es crear gradientes morfogénicos durante las fases de desarrollo, interviniendo en numerosos procesos, como la fusión del tubo neural y la formación de articulaciones.

La desregulación de estos 2 *pathways* de señalización (*FoxO* y *TGF-beta*) tiene implicaciones en la alteración del ciclo celular de las células madre neuronales.

Otro *pathway* interesante, posiblemente desregulado en ambas condiciones experimentales es *Insulin signaling pathway*, con la presencia del gen *Braf* y de *Tsc1*.

*Tsc1* codifica la proteína hamartina inhibidora de crecimiento, que interactúa con la proteína activadora GTPasa tuberina, estabilizándola. El complejo hamartina-tuberina regula negativamente el complejo de señalización que tiene por objetivo la rapamicina en mamíferos (*mTORC1*; *mammalian target of rapamycin complex 1*), que es un regulador importante en el crecimiento anabólico de la célula.



La vía de la señalización de la insulina se puede entender como desdoblada en dos vías principales de transducción de señales moleculares [34]:

- La vía de las quinasa activadas por mitógenos (*MAPK signaling pathway*)
- La vía del fosfatidilinositol 3 quinasa (*PI3K-Akt signaling pathway*)

Estas dos rutas ya se han descrito anteriormente, ahora se verán en más detalle bajo el punto de vista de su activación por la insulina.

#### Vía de señalización de las MAPK.

Los efectos de la insulina en la regulación de la síntesis de proteínas son mediados principalmente por esta vía (Fig. 48). El receptor de la insulina (*IR*) se activa por fosforilación al unirse a la insulina y promueve la asociación de la proteína *Shc*, que se une al complejo *Grb2/SOS*. *SOS* es un factor que intercambia nucleótidos de guanina (*GEF*), que activa a *Ras*, pasando a la forma activada *GTP-Ras* e iniciando la cascada de señalización *MAPK*. *GTP-Ras* activa a *Raf-1*, que recluta y activa a *MEK* (quinasa de la *MAP* quinasa; *MAPKK*) y ésta, a su vez, a *ERK1* y *ERK2* (quinasas reguladas extracelularmente).

Alternativamente, se pueden activar *ERK1* y *ERK2* (*MAP* quinasas) por la insulina de forma independiente a *Shc*, por medio de la activación de *IRS* (sustrato del receptor de la insulina). *IRS* activado se une al complejo *Grb2/SOS* y a partir de aquí se entra en la cadena de señalización *MAPK*.

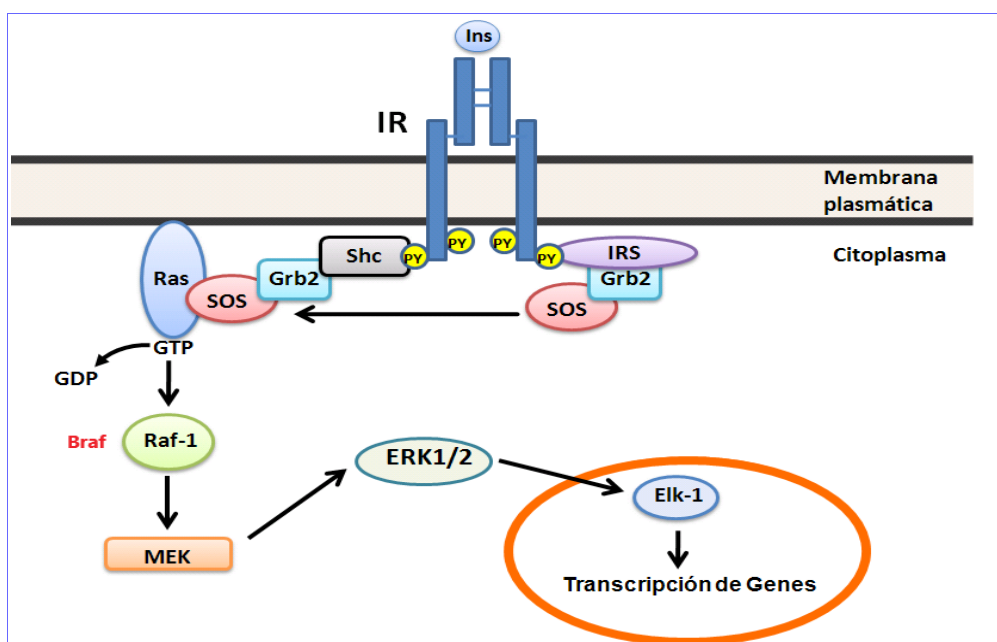


Figura 48. Activación de la vía de las MAPK por la insulina [34]

#### Vía de señalización de la PI3K.

La vía *PI3K* es el principal mecanismo por el que la insulina ejerce sus funciones en el metabolismo de la glucosa y los lípidos (Fig. 49). El receptor activado fosforila *IRS*, provocando la activación de sitios de unión de *IRS* a ciertas proteínas con dominios *SH2* (de homología al dominio 2 de la proteína *Src*), entre las que se encuentra *PI3K*, que actúa como proteína adaptadora. Las

*PI3Ks* tienen una subunidad reguladora, que contiene el dominio *SH2* de unión a *IRS-1* y una subunidad catalítica. La unión provoca cambios halostéricos en la subunidad reguladora que activan la subunidad catalítica de *PI3K*. Esto provoca la activación de la proteína *p110* cerca de la membrana plasmática, donde tiene acceso al sustrato *PI(4)-P* y *PI(4,5)-P2*, que son fosforilados generando *PIP2* y *PIP3*. *PIP3* sirve como sitio de unión a serín quinazas como *PDK1* y *Akt*, también llamada proteína quinasa B (*PKB*). La quinasa *Akt* se une a la membrana y se fosforila en 2 residuos. La fosforilación de Ser473 ocurre primero por acción del complejo *mTOR/Rictor*, también conocido como *PDK2*. Esta fosforilación parece promover la interacción entre el motivo hidrofóbico del carboxilo terminal de *Akt* y la quinasa *PDK1*, que fosforila a Thr308. Estas 2 fosforilaciones son importantes para que *Akt* se active completamente. La enzima *Akt* regula varios efectos metabólicos de la insulina a través de la fosforilación de una lista creciente de sustratos que propagan la respuesta a la insulina, como *mTor*, *FOXO*, *GSK3* y *caspasa 9*.

La vía se resume en la siguiente figura, donde se señalan, en azul, los genes *target* identificados como desregulados en el análisis de datos del experimento 2 y en rojo, los que se obtienen del cruce de datos de los 2 experimentos. Se ha añadido la activación de *PDE3* (Fosfodiesterasa 3), que hidroliza el *cAMP*, inhibiendo la activación de *PKA* (proteína quinasa A), una de cuyas funciones es regular el metabolismo de lípidos; en este último proceso está implicado el gen *Prkar2b*, que codifica una de las subunidades reguladoras de *PKA* y es uno de los genes señalado como clave en los resultados publicados del experimento 2 (ratones *Foxg1+/-*) [4].

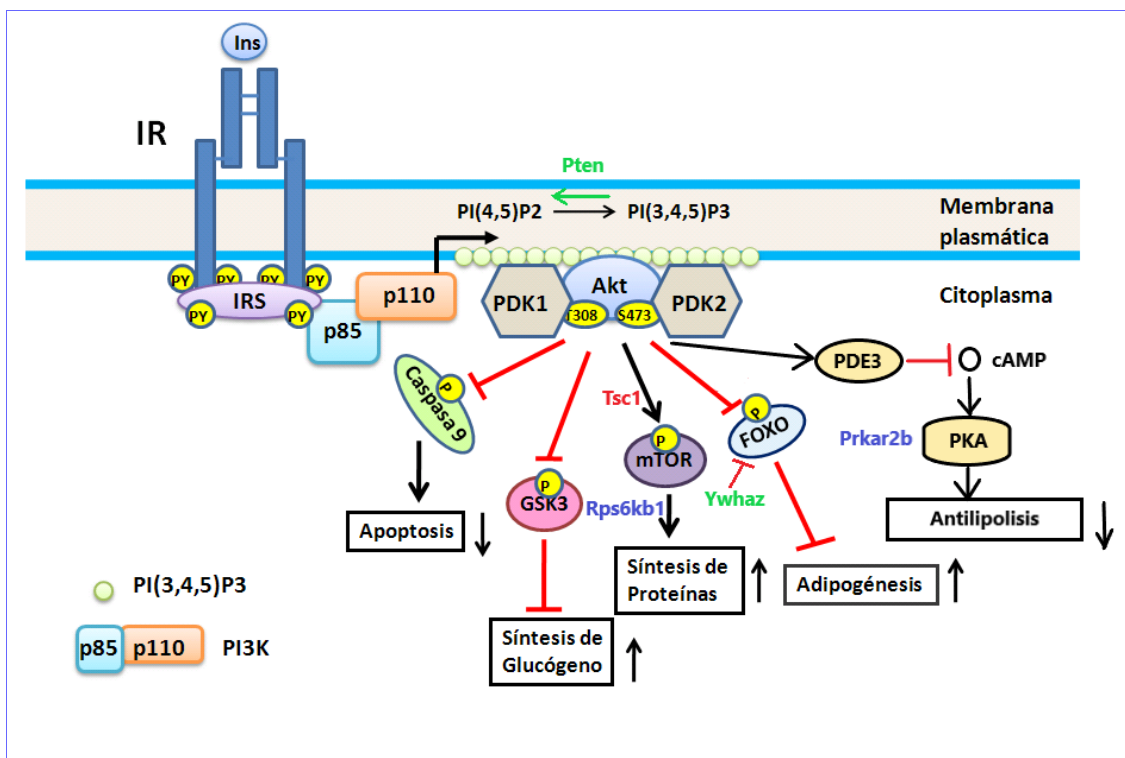


Figura 49. Activación de la vía de las PI3K por la insulina [34]

En los primeros resultados de enriquecimiento *KEGG* de los datos del experimento 1 (ratones *Mecp2+/-*), aparecía el *PI3K-Akt signaling pathway* como posiblemente desregulado, con la presencia de 32 genes *target* implicados. Sin embargo, al hacer una selección más ajustada de *miRNAs* DE y sus genes *target* para un análisis de resultados más acotado, no se encuentra esta vía de señalización como posiblemente desregulada para el experimento 1. Sí aparece como vía común en los datos de los 2 experimentos la de la insulina que, como se ha visto, participa de la vía *PI3K*. Retomando los datos iniciales se comprueba que entre los genes *target* implicados en esta vía hay 2 que aparecen entre los seleccionados posteriormente como más ajustados, son *Pten* e *Ywhaz*, que se han marcado en verde en la figura de la vía *PI3K* activada por insulina.

El gen *Pten* ya se ha mencionado anteriormente como presente, de manera común en los 2 experimentos, en la vía de adhesión focal, la de señalización del fosfatidilinositol y la de señalización *FOXO*. Concretamente, su producto actúa invirtiendo el proceso de *PIP(4,5)2* a *PIP(3,4,5)3*, por lo que regula negativamente la formación de *PI(3,4,5)P3* inhibiendo la activación de *Akt*.

El producto de *Ywhaz* actúa regulando la sensibilidad a la insulina por su interacción con *IRS1*. Aquí se muestra inhibiendo la señalización *FOXO*.

En vista de los datos anteriores, se puede concluir que el *pathway Axon guidance* está posiblemente desregulado en ambos experimentos, con los genes *target* comunes implicados *Nrp1* y *Efnb2*, entre otros. Mientras que *Efnb2* se identifica como *target* de *miR-200b*, *down-reg* en los 2 experimentos; la interacción con *Nrp1* puede ser más compleja. En el experimento 1 con ratones *Mecp2+/-*, *Nrp1* es *target*, además de *miR-200a* (*down-reg* en ambos), de *miR-148a* (*up-reg* en ambos) y de *miR-1a* (*up-reg* en el experimento 1); por lo que la expresión de este gen puede variar significativamente en las 2 condiciones experimentales comparadas, la de ratones *Mecp2+/-*, donde son más significativos los *miRNAs up-reg*, en general y la de ratones *Foxg1+/-*, donde son más significativos los *miRNAs down-reg*.

En el segundo experimento, con ratones *Foxg1+/-*, aparecen como posiblemente desregulados los *pathways Synaptic vesicle cycle* y *GABAergic synapse*, donde aparecen como *targets* los genes comunes a ambos *pathways* *Slc6a1* y *Slc6a11*, hibridados por *miR200b* y *miR429* (*down-reg*) y el gen *Cacna1c*, implicado en *GABAergic synapse*, como *target* de *miR448* (*up-reg*). Estos *pathways* no aparecen como significativos en el análisis de los datos del experimento 1, por lo que el estudio de la posible desregulación de estos genes y su influencia en el correcto funcionamiento de las sinapsis puede revelar algún elemento diferencial en los dos síndromes.

También aparecen una serie de vías de señalización celular posiblemente alteradas en ambos experimentos; entre las que cabe destacar *MAPK signaling pathway* y *PI3K-Akt signaling pathway*, que participan en la mayoría de las vías destacadas y que en un principio no aparecieron como desregulados en el análisis de datos, pero que están relacionadas con el *Insulin signaling pathway*, que aparece en el análisis de los *miRNAs* DE comunes en los dos experimentos. Dentro de los genes implicados en *MAPK signaling pathway* se pueden destacar *Rapgef2*, *Egfr* y *Braf* como posiblemente desregulados en ambos experimentos

y participantes en la vía *MAPK* clásica; *Rapgef2* es posible *target* de miR200b (*down-reg*), *Egfr* de miR200a (*down-reg*) y de miR148a (*up-reg*) y *Braf* de miR92a (*up-reg*). Los genes *Jun*, *Map4k4* y *Map2k4* como posiblemente desregulados en ambos experimentos y participantes de la vía *JNK*; *Jun* y *Map4k4* son posibles *targets* de miR200b y miR429 (*down-reg*) y *Map2ka4* de miR92a (*up-reg*). Y por último, el gen *Mef2c*, posible *target* de miR448 (*up-reg*), que forma parte de la vía de proteínas *p-38* y que aparece sólo en el análisis de los datos del segundo experimento, por lo que puede ofrecer un elemento de diferenciación del síndrome *FOXG1* con respecto al *MECP2*.

*PI3K-Akt signaling pathway* apareció como vía de señalización posiblemente desregulada en ambos experimentos en el análisis de datos previo, antes de realizar la selección de genes *target* y el enriquecimiento *KEGG* con la herramienta *pathfindR* y se revela como vía de señalización clave en *Insulin signaling pathway*, donde cabe destacar la presencia de los genes *Pten* y *Tsc1*, probablemente desregulados en ambos experimentos; *Pten* es posible *target* de miR92a y miR148a (*up-reg*) y *Tsc1* es posible *target* de miR92a (*up-reg*). En esta vía de señalización destaca la presencia del gen *Prkar2b*, uno de los principales motivos de estudio del trabajo de referencia en ratones *Foxg1+/-* [4].

Hasta ahora, se ha realizado un análisis de resultados teniendo en cuenta *miRNAs* DE. Sin embargo, en el punto 5.8 Análisis de *isomiRs* DE, se hizo un análisis detallado por *isomiRs*, que puede ofrecer un mayor detalle de las secuencias más significativas en una búsqueda más precisa de posibles genes *target*.

La lista detallada de los 7 *isomiRs* DE en los datos del experimento 1 es la siguiente.

##	seq	uid
## 150	ACCCGTCCCGTTCGTCCTCCCGG	mmu-miR-1247-5p;iso_3p:a
## 162	ACGCCCTTCCCCCTTCTTCA	mmu-miR-1249-3p;ref
## 536	CATCTAACCTGGCCTTTGA	mmu-miR-1981-3p;iso_3p:c
## 792	GCTGCAACATCCGACTGAAAG	mmu-miR-30f;iso_snp:5CT;iso_5p:GCT;iso_3p:ctc
## 1580	TTTTTCATTATTGCTCCTGACC	mmu-miR-335-3p;ref
## 41	AACATGGAGGAAATTTACAGT	mmu-miR-376c-3p;iso_snp:6GA
## 1559	TTTGATGTCTGACTTTGCAGTT	mmu-miR-7090-3p;iso_add:TT

Ninguno de ellos se corresponde con los *miRNAs* DE con los que se ha trabajado hasta ahora.

La lista detallada de los 28 *isomiRs* DE en los datos del experimento 2 es la siguiente.

##	seq	uid
## 1680	ACCCTGTAGATCCGAATTTGTG	mmu-miR-10a-5p;iso_5p:t
## 1681	ACCCTGTAGATCCGAATTTGTGA	mmu-miR-10a-5p;iso_add:A;iso_5p:t
## 10930	TCCCTGAGACCCTTTAACCTGTGAA	mmu-miR-125a-5p;iso_add:A
## 4339	CAAATCTTATTGAGCACCTG	mmu-miR-1264-3p;iso_3p:t
## 4341	CAAATCTTATTGAGCACCTGT	mmu-miR-1264-3p;ref
## 710	AAGCCCTTACCCAAAAAGCAT	mmu-miR-129-2-3p;ref
## 14011	TTCATTCGGCTGTCCAGATGTAA	mmu-miR-1298-5p;iso_add:A
## 14010	TTCATTCGGCTGTCCAGATGTA	mmu-miR-1298-5p;ref
## 1590	ACCATCGACCGTTGAGTGACA	mmu-miR-181c-3p;iso_add:A;iso_3p:c
## 457	AACATTCATTGTTGTCGGTGGA	mmu-miR-181d-5p;iso_add:A;iso_3p:t

## 463	AACATTCATTGTTGTCGGTGGGT	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	AGGCTGCAGGCCCACTTCCCT	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	GCGACCCATACTGGTTTCAGT	mmu-miR-551b-3p;iso_3p:T
## 1189	AATGGCGCCACTAGGGTTGTGT	mmu-miR-652-3p;iso_add:T
## 9121	TAACTCTTCCCTTGTGCCCTCAGT	mmu-miR-6944-3p;iso_add:T

Aquí sí que se observan isoformas DE de algunos de los *miRNAs* que se han considerado como relevantes en los resultados del experimento 2.

De las herramientas que se han usado para la localización de genes target de los diferentes *miRNAs*, la única que da detalle de la hibridación de la secuencia del *miRNA* con su correspondiente gen *target* es *miRanda*, por lo que se comparará la secuencia del *isomiR* DE con esta base de datos para tratar de localizar las predicciones de genes target más ajustadas.

Se puede comenzar con el miR200b, que es uno de los más relevantes en este experimento con ratones *Foxg1*+/- . El *isomiR* DE correspondiente es:

TAATACTGCCTGGTAATGATGAC	mmu-miR-200b-3p;iso_3p:C
-------------------------	--------------------------

Para localizar la secuencia en la base de datos miRanda, es necesario cambiar el sentido a 3'-5' y cambiar T por U, por lo que la secuencia a buscar sería:

CAGUAGUAAUGGUCCGUCAUAAU

Dentro de las secuencias analizadas por miRanda para el miR200b, la que más se ajusta es:

agUAGUAAUGGUCCGUCAUAAu

El *isomiR* analizado es una variante -3p a la que se añade la base C en el extremo 3'. Se seleccionan los target de la base de datos *miRanda* correspondientes a esta secuencia. Como ejemplo, se muestra el que obtiene una menor puntuación mirsvr (a menor puntuación, mayor estabilidad del híbrido *miRNA*:gen *target*).

mirbase_acc	mirna_name	gene_id	gene_symbol	transcript_id	ext_transcript_id
MIMAT0000233	mmu-miR-200b	216197	Ckap4	uc007gkq.1	NM_175451
mirna_alignment		alignment		gene_alignment	
AgUAGUAAUGGUCCGUCAUAAu		:    :		guAUGGCUUCUUUACAGUAAUu	
mirna_start	mirna_end	gene_start	gene_end	genome_coordinates	
2	21	1104	1125	[mm9:10:83989090-83989111:-]	
conservation	align_score	seed_cat	energy	mirsvr_score	
0.8551	144	7	-8.34	-1.3245	

Se seleccionan los genes con mirsvr < -1.2 (un valor por debajo de -1.2 indica una buena predicción). De los 34 genes *target* seleccionados, hay 2 de los identificados anteriormente en términos *GO* y *pathways* KEGG relevantes: *Slit2* y *Zeb1*.

*Slit2* codifica un miembro de la familia de proteínas *slit*, que juegan un papel importante en el guiado del axón y en la migración neuronal.

*Zeb1* codifica un factor de transcripción del tipo dedos de zinc (*zinc finger*), que juega un papel importante en la regulación de la interleucina 2, una citoquina con diferentes funciones en el sistema inmune.

Del mismo modo se busca el resto de las secuencias de los *isomiRs* DE seleccionados en la base de datos *miRanda*. Para el miR-429:

```
TAATACTGTCTGGTAATGCCGT          mmu-miR-429-3p;ref
```

Corresponde a la variante -3p de referencia. Buscando la secuencia equivalente en *miRanda* se seleccionan 11 posibles genes *target*, de los que *Slitrk1* coincide con los seleccionados anteriormente como relevantes. Este gen codifica un miembro de la familia *slitrk* que puede estar implicado en la formación de yemas de crecimiento en las neuronas.

Dentro de los miRNAs que se encuentran *up-reg* en los datos del experimento 2 se encuentra miR10a, con 2 *isomiRs* DE:

```
ACCCTGTAGATCCGAATTTGTG          mmu-miR-10a-5p;iso_5p:t
ACCCTGTAGATCCGAATTTGTGA mmu-miR-10a-5p;iso_add:A;iso_5p:t
```

Son variantes -5p prácticamente iguales, en las que se ha eliminado una T en el extremo 5' y en la segunda se ha añadido una A en 3'. Seleccionando la secuencia en *miRanda*, se obtiene un solo gen con *mirsvr* < 1.2, *Bdnf*, que no se encuentra entre los genes seleccionados anteriormente. Sin embargo, en el artículo de referencia del experimento 1 [3], se menciona que *MeCP2* silencia transcripcionalmente la expresión de este gen, observándose que *Bdnf* se halla *up-reg* en condiciones *Mecp2 kd*. Por lo que el hecho de que *Bdnf* pueda ser *target* del miR10a y éste se encuentre *up-reg* en los datos del experimento 2, puede suponer que esté regulado en sentidos inversos en ambos síndromes. *Bdnf* (*Brain derived neurotrophic factor*) es un factor de crecimiento del nervio implicado en la plasticidad neuronal, aprendizaje y memoria.

El miR448, *up-reg*, también presenta 2 *isomiRs* DE:

```
TTGCATATGTAGGATGTCCCA          mmu-miR-448-3p;iso_3p:t
TTGCATATGTAGGATGTCCCT mmu-miR-448-3p;iso_add:T;iso_3p:at
```

Son variantes -3p prácticamente iguales. En la primera se ha eliminado una T del extremo 3'; en la segunda se ha eliminado AT de 3' y se añade T. La búsqueda en *miRanda* ofrece una lista de 5 posibles *targets*; ninguno relacionado con los procesos analizados.

El análisis de *isomiRs* en los datos del experimento 1 no parece tan evidente. De los 7 *isomiRs* seleccionados, ninguno corresponde a los *miRNAs* DE analizados anteriormente. Haciendo una búsqueda de estas secuencias en la base de datos *miRanda*, únicamente se encuentran secuencias del miR376c. Al que corresponde el *isomiR* DE:

```
AACATGGAGGAAATTTACGTC          mmu-miR-376c-3p;iso_snp:6GA
```

Es una variante SNP de -3p en la que hay un cambio de A por G en el nucleótido 6. No se encuentra una secuencia equivalente en *miRanda*. Se

obtienen los *targets* de las secuencias más similares y entre los 15 genes *target* resultantes no se encuentra ninguno de los analizados anteriormente.

Al no encontrar referencias válidas, se hace un análisis por separado de estos 7 *miRNAs* de manera similar a como se ha hecho con los *miRNAs* DE analizados en cada experimento. Se obtiene una predicción de genes *target* con la herramienta *TargetScan* (utilizando la función *mirna2targetscan* del paquete *isomiRs* de Bioconductor) y se hace un análisis de enriquecimiento *GO* y *KEGG* de los genes seleccionados como posibles *target*. No se obtiene ningún término *GO* ni ninguna ruta *KEGG* enriquecidos para este listado de genes.

## 6. Discusión

El objetivo principal del trabajo consistía en hacer un análisis detallado, sobre datos disponibles de experimentos de *smallRNA-seq*, de la caracterización de *isomiRs* en dos anomalías genéticas, el síndrome típico de Rett provocado por mutación del gen *MECP2* y una variante atípica del síndrome provocada por mutación del gen *FOXG1*, para encontrar elementos diferenciales y puntos en común entre ambos síndromes.

Para ello, se ha detallado un *pipeline*, utilizando herramientas de análisis de *miRNA/isomiR*, que ha permitido cruzar datos de ambos orígenes de una manera sistematizada, para comparar las dos situaciones y hallar vías genéticas y rutas metabólicas comunes.

En cuanto al análisis de *isomiRs*, no se ha realizado con demasiada profundidad, por el hecho de que el análisis de *miRNAs* DE ha ofrecido datos muy prometedores en la comparación de ambos síndromes. El planteamiento inicial, por los datos disponibles en los estudios publicados, era que sería difícil encontrar *miRNAs* DE comunes a los 2 experimentos, por lo que un análisis más detallado de *isomiRs* podría desvelar ciertas relaciones que, de otro modo, podrían pasar desapercibidas. Sin embargo, al haberse encontrado *miRNAs* que aparecen con cierta significación como DE en ambos experimentos, el análisis detallado de estas vías genéticas y metabólicas comunes ha centrado buena parte del trabajo.

En el diseño y desarrollo del *pipeline* se han empleado herramientas y programas informáticos de uso libre, sin necesidad de adquisición de licencias. En algunas fases del *pipeline* se ha hecho una labor de comparación de distintas herramientas disponibles, como en el caso de análisis de *miRNAs/isomiRs* DE o la integración gen-*miRNA*; sin embargo, en otras secciones, bien por falta de tiempo o bien por encontrar en la literatura información relevante sobre el rendimiento de ciertas herramientas, se ha decidido emplear directamente una en concreto, como ha sido el caso del alineamiento de *miRNAs*. Como conclusión, se ha creado un *pipeline* bien fundamentado y sistematizado, al que se han añadido pequeños *scripts* y subprogramas, como el de cruce de información de *miRNAs* DE entre ambos experimentos, diseñados por el estudiante.

Dentro de los dos experimentos seleccionados para hacer el análisis comparativo, el primero de ellos, estudio de ratones mutantes *Mecp2+/-*, consta sólo de una muestra por condición, por lo que hacer un análisis de expresión diferencial con las herramientas disponibles habría sido un poco limitado. Para solventar esta limitación se crearon réplicas técnicas a partir de las muestras biológicas disponibles. Esta tarea no estaba prevista inicialmente y ha añadido complejidad al proyecto. Además, el hecho de trabajar con muestras simuladas, limita el alcance de los resultados obtenidos.

De hecho, los resultados de *miRNAs* DE obtenidos en el presente trabajo para este primer experimento no coinciden exactamente con los referidos en el estudio [3]. Esto puede ser debido al problema mencionado de falta de réplicas, sumado al distinto método empleado para la obtención de los *miRNAs* DE, que



en el estudio de referencia ha consistido en seleccionarlos solamente por su valor de *fold-change* y en el presente trabajo ha consistido en la comparación de distintos métodos paramétricos y no paramétricos teniendo en cuenta también la dispersión de los resultados y ofreciendo parámetros de contraste más completos, para escoger finalmente uno de ellos.

La metodología empleada ha sido puramente bioinformática. Se ha partido de datos en crudo de experimentos de *smallRNA-seq* de libre acceso y se ha diseñado un *pipeline* de análisis adaptado a las circunstancias concretas de los dos experimentos, pero aplicable a la metodología general de análisis de datos de otros experimentos similares. Las conclusiones obtenidas en dicho análisis se han contrastado por métodos bioestadísticos, pero no han sido sometidas a verificación experimental, por lo que su alcance es limitado.

Para identificar experimentalmente las uniones *miRNA* con genes target se puede utilizar el método *PAR-CLIP*, mencionado anteriormente como base de cálculo de parámetros de predicción de *targets* en *MicroT-CDS*. *PAR-CLIP* (*photoactivable-ribonucleoside-enhanced crosslinking and immunoprecipitation*) es un método bioquímico, similar a *CLIP-seq*, utilizado para identificar la unión de los sitios *RNA* de unión a proteínas (*RPBs*; *RNA-binding proteins*) con complejos ribonucleoproteicos que contienen *miRNA* (*miRNPs*; *microRNA-containing ribonucleoprotein complexes*) (Fig. 50). En primer lugar, la célula se trata con 4-SU (*4-thiouridine*), un ribonucleósido fotoactivable que se incorpora al *RNA* durante la transcripción. La radiación con luz UV cataliza la formación de enlaces covalentes entre los sitios de unión del *RNA* y los complejos ribonucleoproteicos. La célula se rompe y se aísla la proteína de interés por inmunoprecipitación. Se retira el *RNA* no específico y se ligan adaptadores en el extremo 3' de los fragmentos de *RNA* de interés. Se digiere la proteína con proteasa K para dejar sólo un péptido que identifica el nucleótido de unión. Después de unir *linkers* al extremo 5' del *RNA* de interés, se sintetiza *cDNA* por transcripción inversa. Esto permite la secuenciación del *read* por PCR y su mapeo en el genoma, identificando el target del complejo ribonucleoproteico [35].

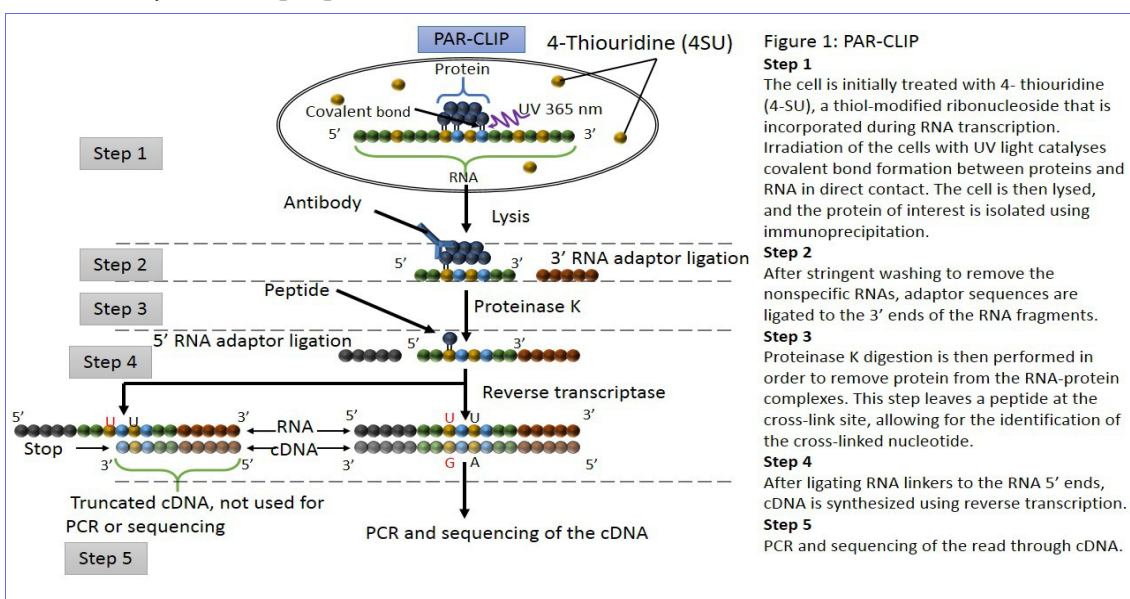


Figura 50. Verificación experimental de targets con PAR-CLIP [35].

Se puede tomar como ejemplo el gen *Pten*, que se ha predicho como gen *target* de distintos *miRNAs* en ambos experimentos y que interviene en la vía de señalización del *PI3K*, implicada en la señalización de la insulina y otros procesos que se han evidenciado como posiblemente desregulados en ambos síndromes. En el análisis se ha tomado como base la predicción hecha por *TargetScan* y se han observado posibles regulaciones de este gen por los *miRNAs* mir-92a y miR-148a (*up-reg* en ambos experimentos) y en menor medida por miR-200b y miR-200a (*down-reg* en ambos experimentos). Predicciones por *TargetScan*:

	Seedmatch	PCT	mir	gene.name
674	8mer	0.97	mmu-miR-92a-1-5p	Pten
675	8mer	0.97	mmu-miR-92a-3p	Pten
676	7mer-1a	0.36	mmu-miR-200a-5p	Pten
677	7mer-m8	0.44	mmu-miR-200a-5p	Pten
678	7mer-m8	0.73	mmu-miR-148a-5p	Pten
679	8mer	0.85	mmu-miR-148a-5p	Pten
680	7mer-m8	0.76	mmu-miR-200b-5p	Pten

Sin embargo, las predicciones hechas con *MicroT-CDS* arrojan resultados diferentes. Para los *miRNAs* DE del primer experimento:

	Ensembl	gene.symbol	miRNA	miTG.score
2405	ENSMUSG00000013663	Pten	mmu-miR-335-3p	0.8945879
3864	ENSMUSG00000013663	Pten	mmu-miR-129b-5p	0.8457793
5389	ENSMUSG00000013663	Pten	mmu-miR-3071-3p	0.8092509
5756	ENSMUSG00000013663	Pten	mmu-miR-411-3p	0.8009941

Y para el segundo experimento:

	Ensembl	gene.symbol	miRNA	miTG.score
1543	ENSMUSG00000013663	Pten	mmu-miR-200a-3p	0.8719079
4142	ENSMUSG00000013663	Pten	mmu-miR-429-3p	0.7118903

La puntuación miTG más alta (0.895) corresponde al miR-335 que se ha señalado con alta significación como *up-reg* en los datos del experimento 1 y no se ofrecen predicciones de que *Pten* sea *target* de mir-92a o miR-148a y sí de miR-200a con una puntuación relativamente alta (miTG.score=0.872).

Buscando las predicciones *target* para el gen *Pten* de los *miRNAs* desregulados en *miRanda* se obtiene una puntuación mirsvr=-0.75 para el miR-200a y el resto muestran puntuaciones poco significativas (por encima de -0.35) o no aparecen en la base de datos.

Una comprobación experimental de la interacción de estos *miRNAs* con el gen *target Pten* podría arrojar luz sobre las complejas redes de regulación que se evidencian en el análisis de resultados de ambos experimentos.

En este estudio se han tenido en cuenta solamente las posibles hibridaciones de los *miRNAs* desregulados con sus genes *target*, provocando una regulación post-transcripcional de los mismos. Algunos de los genes *target* identificados son a su vez inhibidores o potenciadores de la transcripción, que ejerceran control sobre otros productos genéticos.

Un ejemplo de ello es el gen *Mef2c*, que codifica un potenciador de la transcripción (enhancer) y aparece como posible gen *target* de los *miRNAs* DE

miR-448 y miR-551b en el experimento 2 y puede suponer una vía de diferenciación en ambos síndromes. Además de comprobar experimentalmente la hibridación de estos miRNAs en el target (Mef2c), habría que comprobar su diferencia de expresión en ambas situaciones y cómo puede afectar a posteriores vías de regulación.

## 7. Glosario

Adapter	Pequeño fragmento de DNA que se une a las secuencias fraccionadas en un experimento de secuenciación para ligarlas a la superficie sobre la que se va a realizar la PCR (reacción en cadena de la polimerasa).
Ago	Las proteínas Ago o Argonautas se unen a pequeños fragmentos de RNA para formar un complejo ribonucleoproteico de silenciamiento.
Akt	Proteína quinasa B. Juega un papel importante en la señalización celular en mamíferos.
CDS	Región codificadora de un gen.
DESeq2	Método paramétrico para análisis de expresión diferencial que utiliza el ajuste a la distribución binomial negativa.
DGCR8	(DiGeorge syndrome chromosomal or critical region 8). Esta proteína forma parte del complejo microprocesador de micro RNA.
Dicer	Ribonucleasa que interviene en el proceso de corte de los precursores de pequeñas moléculas de RNA. Actúa principalmente en el citoplasma.
Drosha	Ribonucleasa que interviene en el proceso de corte de los fragmentos primarios de RNA que darán lugar a precursores de pequeñas moléculas de RNA. Actúa principalmente en el núcleo celular.
ErbB	Familia de proteínas que comprende cuatro receptores tirosín quinasa, relacionadas estructuralmente con EGFR (epidermal growth factor receptor) y que tienen gran importancia en procesos de desarrollo.
FOXG1	(Forkhead box protein G1). Factor inhibidor de la transcripción que juega un importante papel en la subdivisión regional del desarrollo del cerebro y en el desarrollo del telencéfalo.
Frameshift mutation.	Mutación por desplazamiento o desfase. Puede estar causada por una delección o inserción de un número de nucleótidos no múltiplo de 3 que provoca un desplazamiento del marco de lectura.
GABA	Ácido $\gamma$ -aminobutírico. Aminoácido que actúa como principal neurotransmisor inhibitor en el sistema nervioso central (SNC) de mamíferos.
GTPasa	Guanosina trifosfatasa. Superfamilia de enzimas que regula diversas funciones biológicas. La GTPasa está activa cuando está unida al GTP e inactiva cuando está unida al GDP.

Hairpin	Estructura secundaria del precursor de micro RNA en forma de horquilla (hairpin).
isomiR	Isoforma del micro RNA que presenta variaciones con respecto a la secuencia de referencia.
MAPK	Mitogen-Activated Protein Kinases. Proteínas quinasas activadas por mitógenos, es una ruta de transducción de señales moleculares que modifica la actividad de diferentes proteínas, incluyendo factores de transcripción que modulan la expresión de distintos genes.
MDS	Multidimensional scaling. El escalado multidimensional es un conjunto de técnicas estadísticas que se usan para reducir la dimensionalidad de un conjunto de datos, permitiendo una apreciación visual de las relaciones estructurales subyacentes contenidas en ellos.
MECP2	Methyl CpG binding protein 2. Proteína que puede actuar como activador o represor transcripcional, dependiendo del contexto y que es esencial en el normal funcionamiento de las células nerviosas.
Missense mutation.	Mutación con cambio de sentido, de un solo nucleótido que provoca la aparición de un codón que codifica un aminoácido diferente.
miRNA	micro ARN. ARN monocatenario de pequeña longitud, entre 21 a 25 nucleótidos, que tiene la capacidad de regular la expresión de otros genes.
mRNA	ARN mensajero. Fragmento de ARN procedente de la transcripción del ADN cromosómico que se traduce en la secuencia de la proteína codificada por el gen.
mTOR	mammalian Target of Rapamycin. La diana de rapamicina en células de mamíferos es una proteína con importantes funciones como el control del inicio de la transcripción del mRNA, la organización del citoesqueleto y regulación del crecimiento, proliferación y muerte celular.
NGS	Next-Generation Sequencing. Técnica de secuenciación de ADN de alto rendimiento.
Nonsense mutation.	Mutación sin sentido. Provoca la aparición de un codón de terminación donde antes había uno de especificación de aminoácido por sustitución de un nucleótido.
PIP3	Fosfatidilinositol trifosfato. Es un fosfolípido que reside en la membrana plasmática y que activa componentes de señalización molecular como la proteína quinasa Akt.
PLS-DA	Partial least squares-discriminant analysis. Es un método estadístico de regresión lineal por mínimos cuadrados parciales que permite reducir la dimensionalidad en un conjunto de datos de alta complejidad.

pre-miRNA	Precursor del micro RNA. Procede del procesamiento realizado por el complejo microprocesador, en el que se incluye Drosha, sobre el micro RNA primario. Este precursor tiene forma de horquilla (hairpin) y será procesado por Dicer para generar el micro RNA maduro.
pri-miRNA	micro RNA primitivo. Es el transcrito primario procedente del gen codificante de miRNA y que será procesado por el complejo microprocesador para generar el precursor de miRNA.
PRKAR2B	cAMP-dependent protein kinase type II-beta regulatory subunit. Componente de la proteína quinasa A (PKA) que transduce señales moleculares a múltiples objetivos (targets).
Proteínas G	Proteínas fijadoras de nucleótidos de Guanina. Familia de proteínas transductoras de señales desde el receptor al que están acopladas hasta una o más proteínas receptoras que dependen del GTP para su activación.
Proteínas quinasa	Tipo de enzimas que modifican otras moléculas (sustratos) por fosforilación, normalmente activándolos.
Rank Product	Contraste no paramétrico para la detección de genes (u otros fragmentos de RNA) diferencialmente expresados en distintas condiciones experimentales.
RAS	Conjunto de interruptores-reguladores moleculares muy importantes en una gran variedad de rutas de transmisión de señales celulares que controlan diferentes eventos celulares, como integridad del citoesqueleto, proliferación, diferenciación, adhesión y migración celular y apoptosis.
Reads	Lecturas de los fragmentos de RNA que resultan de un proceso de secuenciación. Estos reads se deben alinear contra secuencias de referencia para encontrar su posición en el genoma.
RISC	RNA-induced silencing complex. Complejo ribonucleoproteico de silenciamiento genético en cuya formación interviene un micro RNA específico.
RTT	Síndrome de Rett. Transtorno del sistema nervioso que se caracteriza por microcefalia y pérdida de habilidades después de un periodo de normalidad inicial.
tRTT	Síndrome de Rett típico. Causado normalmente por mutaciones de MECP2.
atRTT	Síndrome de Rett atípico. Causado por mutaciones de otros genes, como FOXP1.
Seed	Región germen o semilla del micro RNA. La constituyen los 8 primeros nucleótidos de la secuencia madura del miRNA y es la más relevante en la unión al gen target.
siRNA	small interfering RNA. RNA pequeño de interferencia, es una clase de RNA bicatenario, similar al micro RNA, que provoca el silenciamiento de mRNA diana (target) específico.

smallRNA	RNA de pequeño tamaño de menos de 200 nucleótido de longitud, normalmente no codificantes, entre los que se incluyen los micro RNA (miRNA) y los pequeños RNA de interferencia (siRNA).
Splicing mutation.	Mutaciones de inserción, delección o cambio en sitios específicos de corte (splicing) en el proceso de maduración de RNA para dar lugar al RNA mensajero (mRNA).
Target	mRNA de genes objetivo que hibridan con el miRNA, provocando normalmente su inactivación.
TGF $\beta$	Transforming growth factor beta. Factor de crecimiento transformante beta, familia de factores de crecimiento, denominados citoquinas, involucradas en procesos celulares como hematopoyesis, proliferación celular, angiogénesis, diferenciación, migración y apoptosis celular.
XCI	X chromosome inactivation. Es la inactivación preferencial de uno de los cromosomas X en mujeres.

## 8. Bibliografía

- [1] Matijevic, T., Knezevic, J., Slavica, M., & Pavelic, J. (2009). Rett syndrome: from the gene to the disease. *European neurology*, 61(1), 3-10.  
<https://www.karger.com/Article/FullText/165342>
- [2] Kortüm, Fanny, et al. "The core FOXP1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis." *Journal of medical genetics* 48.6 (2011): 396-406.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5522617>
- [3] 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)
- [4] Weise, Stefan C., et al. "FOXP1 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>
- [5] Bartel, David P. "MicroRNAs: genomics, biogenesis, mechanism, and function." *cell* 116.2 (2004): 281-297.  
<https://www.sciencedirect.com/science/article/pii/S0092867404000455>
- [6] 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/>
- [7] Guo, Li, et al. "A comprehensive analysis of miRNA/isomiR expression with gender difference." *PloS one* 11.5 (2016).  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4864079/>
- [8] Neul, J. L., Kaufmann, W. E., Glaze, D. G., Christodoulou, J., Clarke, A. J., Bahi-Buisson, N., ... & Renieri, A. (2010). Rett syndrome: revised diagnostic criteria and nomenclature. *Annals of neurology*, 68(6), 944-950. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3058521/>
- [9] Hou, P. S., hAilín, D. Ó., Vogel, T., & Hanashima, C. (2020). Transcription and Beyond: Delineating FOXP1 Function in Cortical Development and Disorders. *Frontiers in Cellular Neuroscience*, 14, 35.  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7052011/>
- [10] Lai, Mong-Hsun Tsai. "Common applications of next-generation sequencing technologies in genomic research." (2013).  
<http://tcr.amegroups.com/article/view/962/html>
- [11] SRA handbook. <https://www.ncbi.nlm.nih.gov/books/NBK242621/>
- [12] 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/>
- [13] Pantano, L. "miRNA annotation".  
[https://seqcluster.readthedocs.io/mirna\\_annotation.html](https://seqcluster.readthedocs.io/mirna_annotation.html)



- [14] González, I. "Statistical analysis of RNA-Seq data". Universidad de Toulouse (2014)
- [15] 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>
- [16] Pantano, Lorena. "Characterization of miRNA and isomiR molecules." <https://www.bioconductor.org/packages/devel/bioc/vignettes/isomiRs/inst/doc/isomiRs.html>
- [17] 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>
- [18] Polyester. Documento de Bioconductor: <https://www.bioconductor.org/packages/release/bioc/vignettes/polyester/inst/doc/polyester.html>
- [19] Hong, F.(2010). Bioconductor RankProd Package Vignette. <http://bioconductor.statistik.tu-dortmund.de/packages/3.6/bioc/vignettes/RankProd/inst/doc/RankProd.pdf>
- [20] 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>
- [21] 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>
- [22] Csardi G (2013). *targetscan.Hs.eg.db: TargetScan miRNA target predictions for human*. R package version 0.6.1.
- [23] 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>
- [24] 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/>
- [25] 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\\_target\\_s\\_in\\_protein\\_coding\\_sequences](https://www.researchgate.net/publication/221785142_Functional_microRNA_target_s_in_protein_coding_sequences)
- [26] 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>

[27] 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)

[28] Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *InterJournal*, *Complex Systems*, 1695. <http://igraph.org>.

[29] 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>

[30] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27, 29-34 (1999).

<https://www.ncbi.nlm.nih.gov/pubmed/9847135>

[31] 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>

[32] Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan - Golan Y, Kohn A, Rappaport N, Safran M, and Lancet D. *The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis*, *Current Protocols in Bioinformatics*(2016), 54:1.30.1 - 1.30.33.doi: [10.1002 / cpbi.5](https://doi.org/10.1002/cpbi.5).

[33] Guo L, Lu Z (2010) The Fate of miRNA\* Strand through Evolutionary Analysis: Implication for Degradation As Merely Carrier Strand or Potential Regulatory Molecule? *PLoS ONE* 5(6):

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011387>

[34] Reyes, J. A. O., & Plancarte, A. A. (2008). Bases moleculares de las acciones de la insulina. *Revista de educación bioquímica*, 27(1), 9-18.

<https://www.medigraphic.com/pdfs/revedubio/reb-2008/reb081c.pdf>

[35] Cross-linking immunoprecipitation. [https://www.wikiwand.com/en/Cross-linking\\_immunoprecipitation#/miRNA\\_target\\_detection](https://www.wikiwand.com/en/Cross-linking_immunoprecipitation#/miRNA_target_detection)