

## Documento biológico del TFM

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

**Fernando Hernández Correa**

Máster en bioinformática y bioestadística.  
Bioinformática clínica.

**Guerau Fernández Isern**

**Javier Luis Cánovas Izquierdo**

**10 de Junio de 2020**

# Índice de contenido

1. INTRODUCCIÓN.....	3
2. ANTECEDENTES.....	3
3. PROBLEMÁTICA.....	5
3.1 Síndrome de Rett típico (tRTT) mediado por mutación del gen MECP2.....	7
3.2 Síndrome de Rett atípico (atRTT) mediado por mutación del gen FOXP1.....	10
4. PLANTEAMIENTO.....	12
5. METODOLOGÍA.....	12
5.1 Datos de partida.....	12
5.2 Descarga de datos en crudo y adaptación al formato adecuado.....	13
5.3 Comprobación de la calidad de los reads.....	13
5.4 Anotación de miRNAs.....	14
5.5 Control de calidad de muestras y normalización.....	16
5.6 Análisis de expresión diferencial de miRNAs/isomiRs.....	16
5.7 Proceso de cruce de información de miRNAs/isomiRs DE en ambos experimentos.....	19
5.8 Análisis detallado de isomiRs DE.....	19
5.9 Clasificación supervisada.....	20
5.10 Integración gen-miRNA.....	20
5.11 Enriquecimiento GO.....	23
5.12 Enriquecimiento KEGG.....	35
6. ANÁLISIS DE RESULTADOS.....	39
6.1 Experimento1 (ratones mecp2+/-).....	39
6.2 Experimento2 (ratones foxg1+/-).....	42
6.3 Cruce de los dos experimentos.....	44
7. CONCLUSIONES.....	46
REFERENCIAS.....	62

## 1. INTRODUCCIÓN

El presente documento desarrolla las tareas de análisis, resultados y conclusiones biológicos correspondientes al TFM que lleva por título **Análisis comparativo de smallRNAs en síndrome de Rett mediado por MECP2 y síndrome atípico de Rett mediado por FOXP1**.

En referencia al Plan de Trabajo, este documento recoge las tareas del grupo T2, encaminadas a la consecución del objetivo global 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.

## 2. 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.

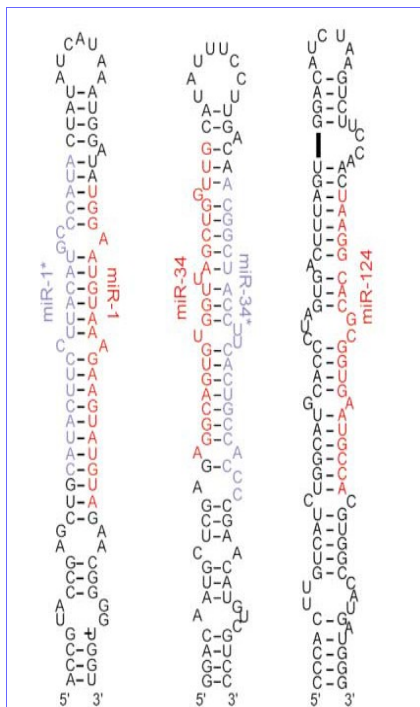


Figura 1. Ejemplos de *miRNAs* en metazoos. Las secuencias corresponden a los precursores (*hairpin*) *mir-1*, *mir-34* y *mir-124* de *C. elegans*, pero se han encontrado secuencias homólogas en mamíferos. (Lagos-Quintana et al., 2001,2002; Lau et al., 2001; Lee and Ambros, 2001) [[1](#)].

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*).

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 [[1](#)].

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

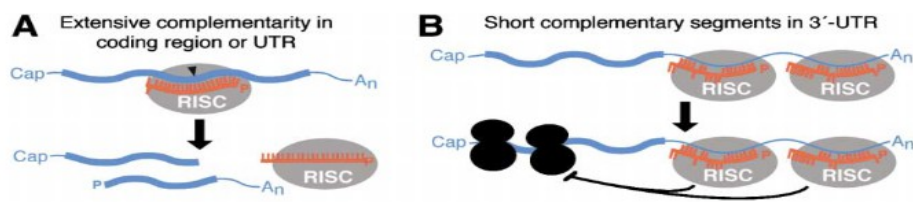
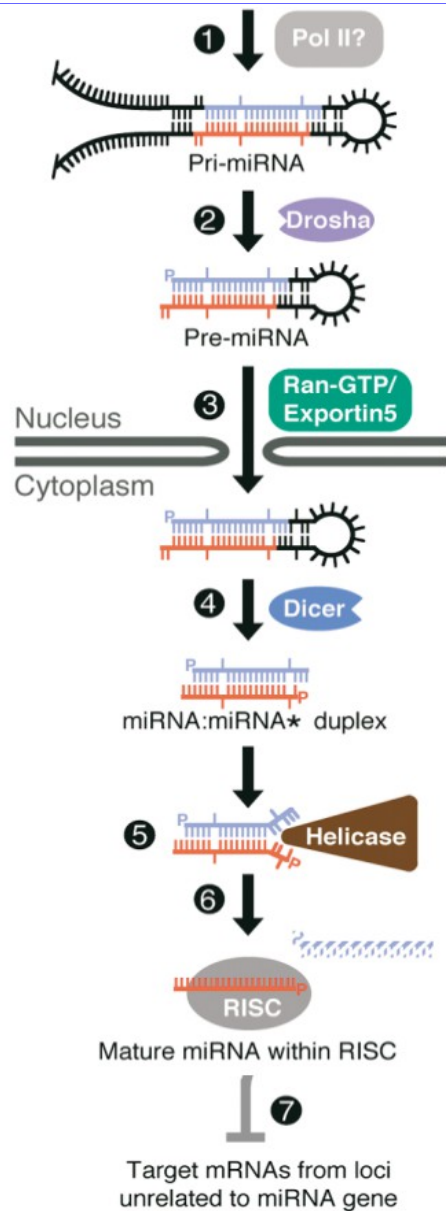


Figura 2. Biogénesis de *miRNA* en mamíferos. 1) División del *pri-miRNA*; se libera un bucle (*stem loop*) de ~60-70 nuc. 2) La endonucleasa *RnaseIII Drosha* corta las hebras del bucle en lugares próximos al *pre-miRNA*. 3) El *pre-miRNA* se transporta al citoplasma. 4) El corte nuclear de *Drosha* 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. [1]

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*.

Se pueden considerar las siguientes variantes de *isomiRs* sobre el *miRNA* maduro:

- *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.

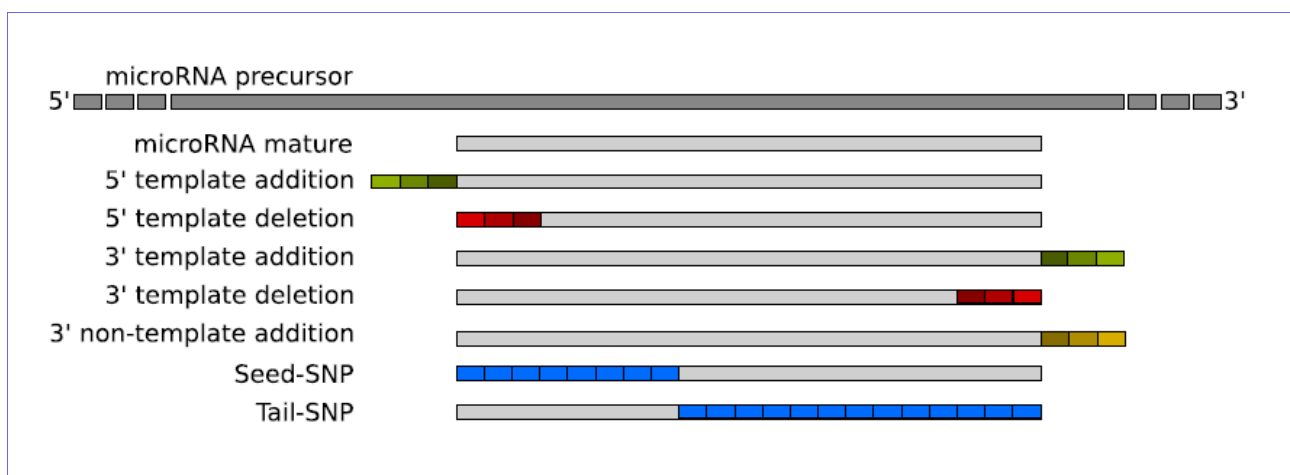


Figura 3. Tipos de *isomiR*. Las cajas verdes representan adición de nucleótidos, las amarillas adiciones *non-template*, las rojas deleciones y las azules *SNPs* (*single nucleotide polymorphism*) [2].

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* [2].

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 [3].

### 3. PROBLEMÁTICA

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 FOXP1, 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 las 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.

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 FOXP1. Las características del síndrome FOXP1 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.

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 FOXP1 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 FOXP1, 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 FOXP1 con un reconocimiento clínico específico.

El fenotipo del síndrome FOXP1, 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 isodisómico. 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 [6]. 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)

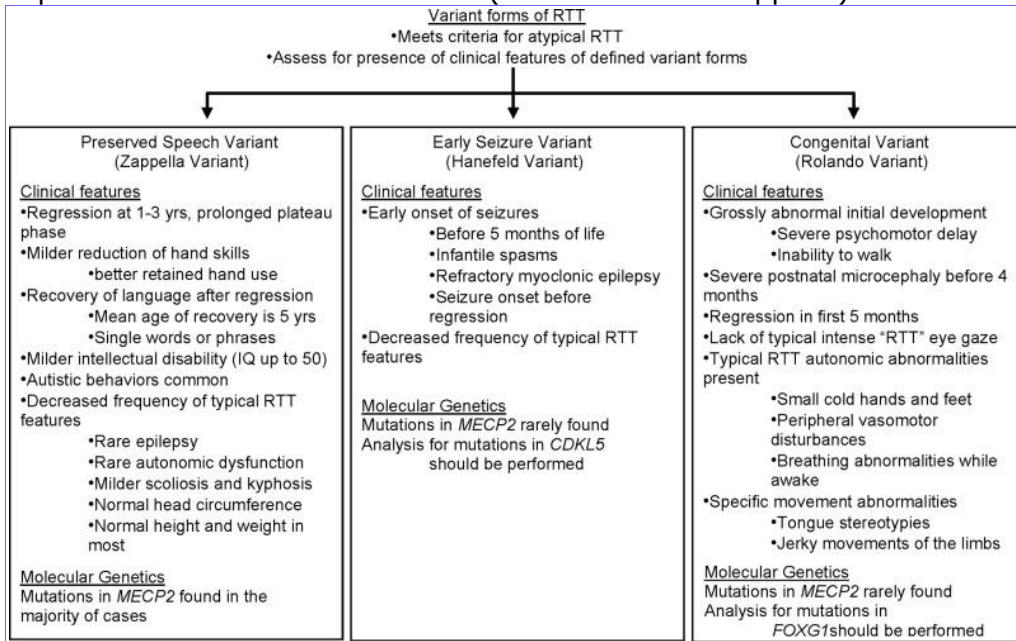


Figura 4. Diagnóstico de variantes de RTT [7].

### 3.1 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 [4].

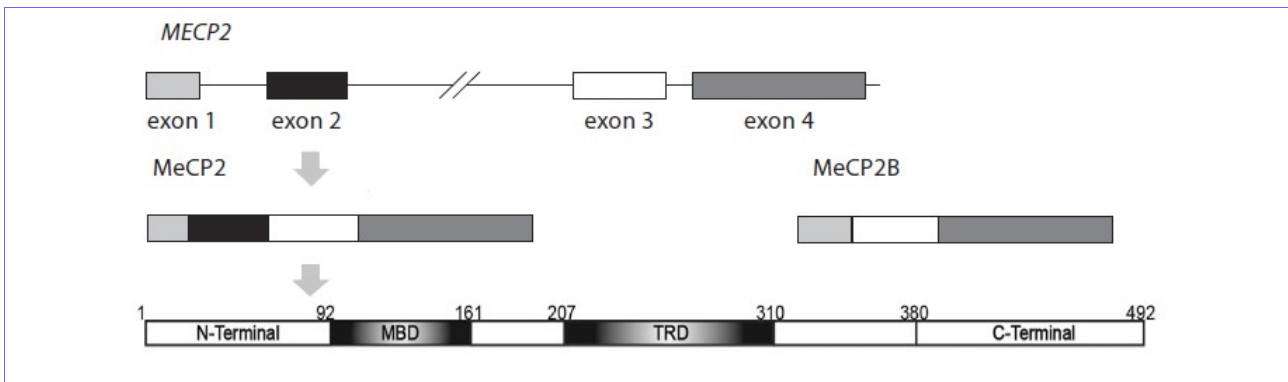


Figura 5. Estructura del gen MECP2 y estructura primaria de la proteína MeCP2 [4 y 5].

La proteína MeCP2 tiene 4 dominios funcionales:

- 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.

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 MDB 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.

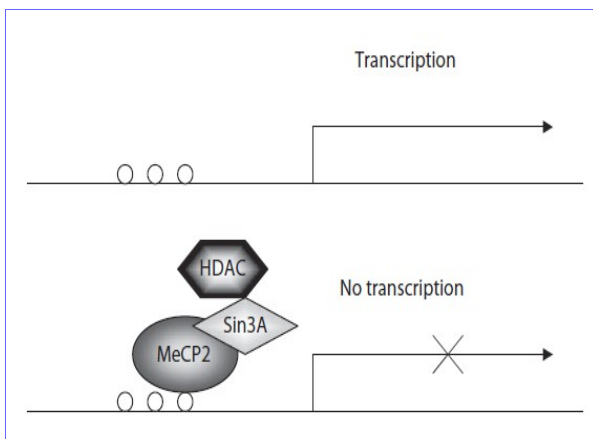


Figura 6. Función de MeCP2. MeCP2 actúa como represor transcripcional. Se une a CpG metilado y a través del co-represor Sin3A recluta a la histona de-acetilasa (HDAC), que provoca la de-acetilación de las histonas H3 y H4. El resultado es una compactación de la cromatina, a la que la maquinaria transcripcional no puede acceder [4].

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 [4]:

- 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.



- *FXYD domain*. Controla la excitabilidad celular por modulación de la actividad  $\text{Na}^+/\text{K}^+/\text{ATPasa}$ .

En el trabajo de referencia [5] 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. 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.

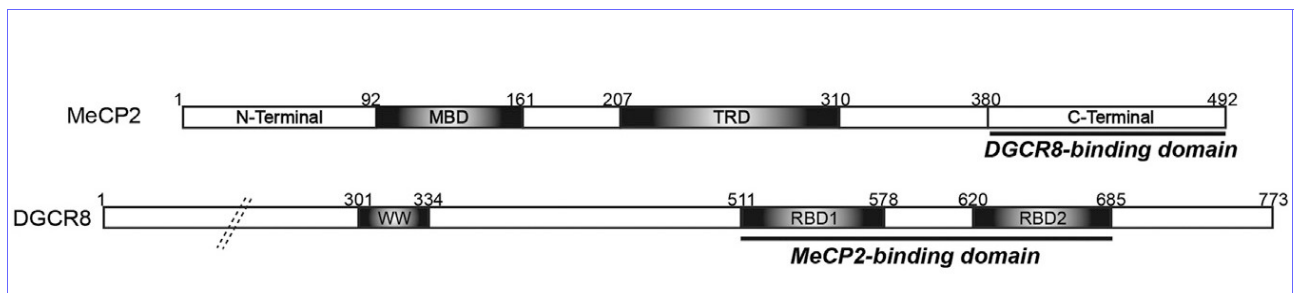
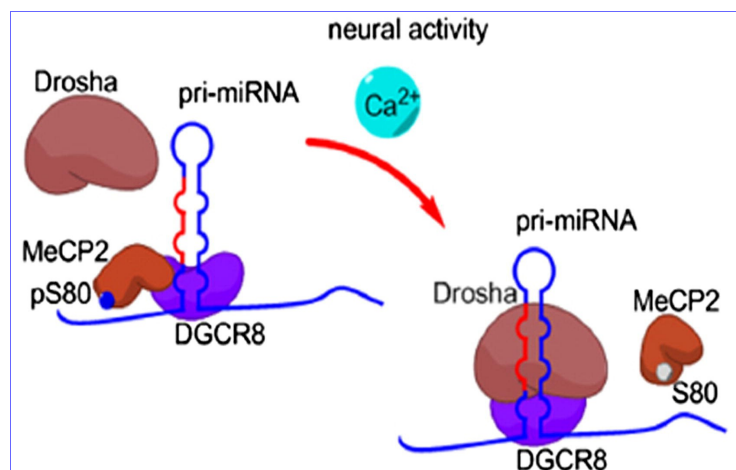


Figura 7. Interacción de MeCP2 con DGCR8. La unión de ambas moléculas depende del extremo C-terminal de MeCP2 y de los dominios de unión a RNA (RBD1 y RBD2) de DGCR8 [5].

Se propone también que la interacción entre MeCP2 y DGCR8 está regulada por el estado de fosforilación del residuo Ser80 de MeCP2. 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.

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* [5].



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.

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 [5].

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*.

### 3.2 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.

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  $\beta$* ), 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 [8].

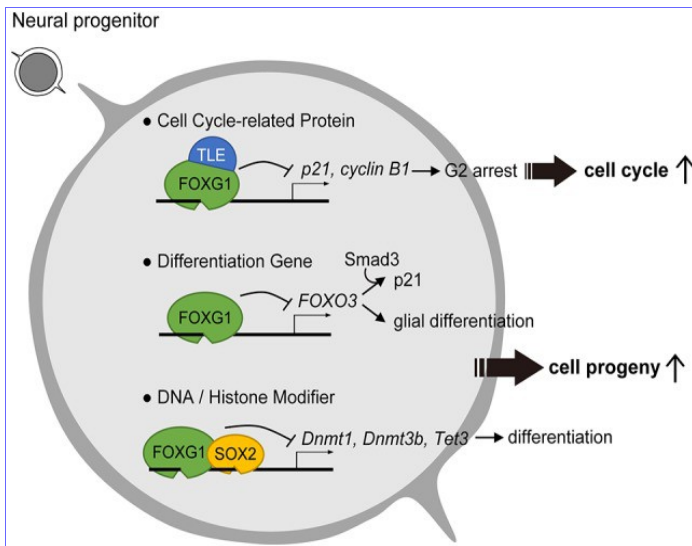


Figura 9. Papel de FOXG1 en el mantenimiento de progenitor celular y en la supresión de diferenciación. Para impedir la salida del ciclo celular, FOXG1 coopera con TLE para inhibir la expresión de proteínas clave en el ciclo celular. FOXG1 también inhibe la diferenciación génica y la modificación DNA/Histona, impidiendo la diferenciación neuronal [8].

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 [9], 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.

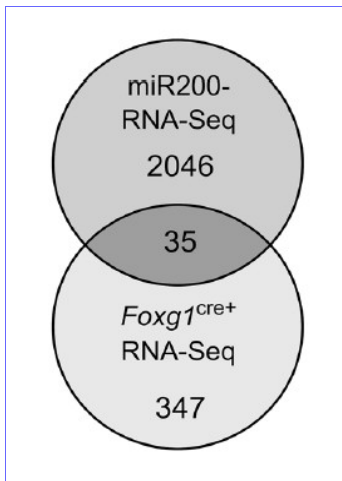


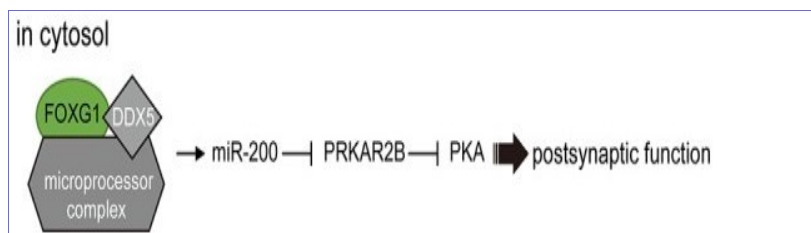
Figura 10. Identificación de *targets* comunes en sobreexpresión de *miR200* y deficiencia en *Foxg1*. Se identificaron 2081 genes diferencialmente expresados en condiciones de sobre-expresión de miembros de ella familia *miR200* y 382 genes en *Foxg1<sup>cre/+</sup>*, de los cuales 35 son comunes [9].

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; 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 *Prkar2r* directamente, además de su regulación post-transcripcional mediada por *miR200*. 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 [8].



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.

#### 4. PLANTEAMIENTO

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-/y 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.

#### 5. METODOLOGÍA

Para ver en más detalle la metodología empleada, es conveniente consultar el documento técnico-bioinformático asociado de este mismo trabajo: TFM\_FHC\_1\_1.

##### 5.1 Datos de partida

El primer estudio cuenta con 2 muestras:

GSM1424567, secuenciación *miRNA* de hipocampo de ratones normales (*wild tipe*)

GSM1424566, secuenciación *miRNA* de hipocampo de ratones *Mecp2-/Y*

Las muestras están disponibles en:

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

El segundo estudio 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/+*

Las muestras están disponibles en:

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

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

## 5.2 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 utiliza la herramienta *fastq-dump*, que transforma los datos *sra* en crudo al formato *fastq* [10].

En los ficheros resultantes de las muestras, cada registro consta de 4 líneas, con el siguiente formato:

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

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

## 5.3 Comprobación de la calidad de los reads.

Para la comprobación de la calidad de los *reads* se utiliza la herramienta *fastqc*. Como resultado de cada fichero *.fastq* se obtienen 2 archivos:

- \*\_fastqc.zip            Resultados del análisis.
- \*\_fastqc.html        Informe en formato html.

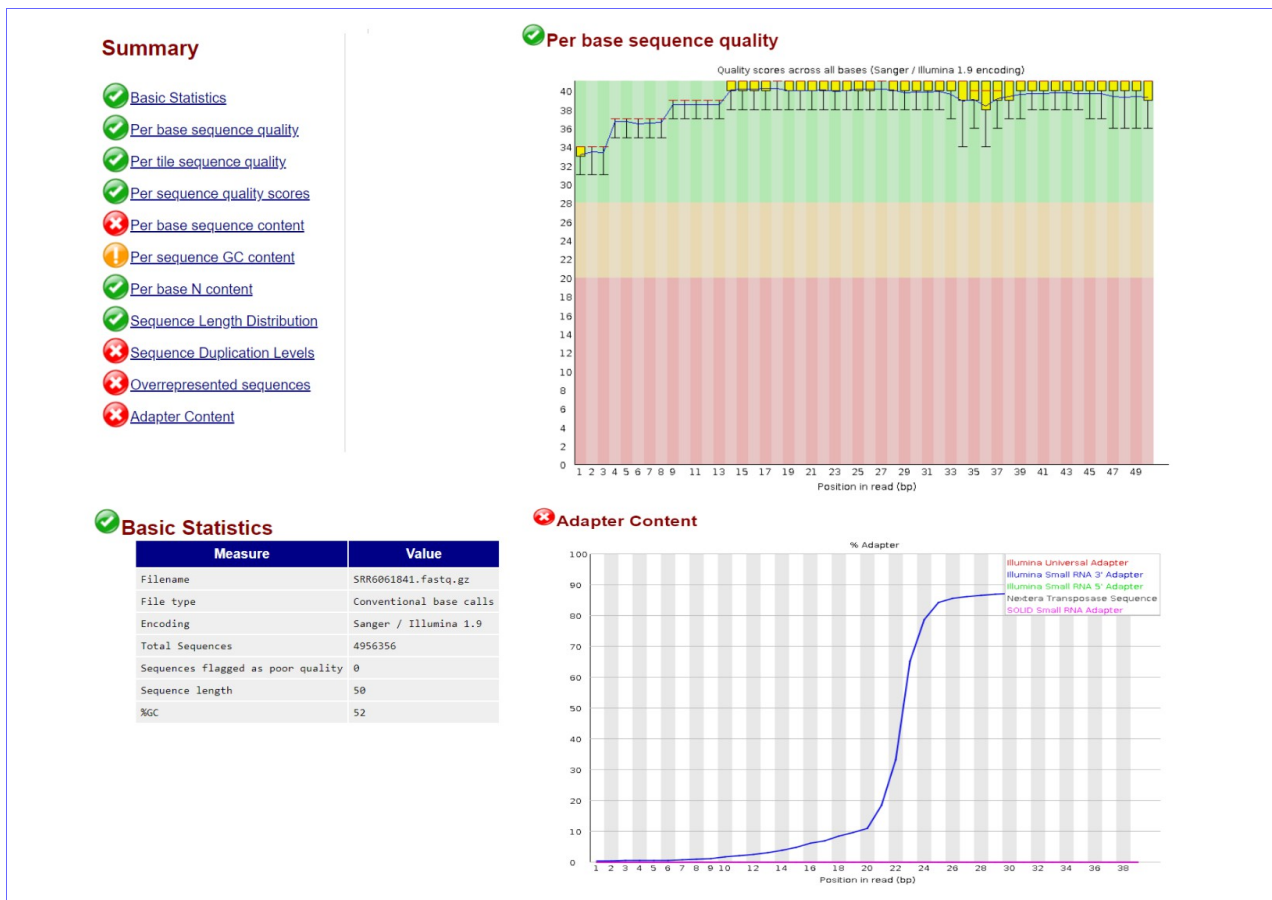


Figura 12. Parte del informe de calidad *fastqc* de una de las muestras del experimento 2.

En los informes de las muestras del primer experimento se aprecia una calidad de las secuencias por bases bastante buena y no se detecta la presencia de adapters. Sin embargo, en la sección *Overrepresented sequences* del informe se detecta una secuencia catalogada como *TruSeq Adapter*:

```
ATCTCGTATGCCGTCTTCTGCTTG
```

En los informes de las muestras del segundo experimento la calidad de las secuencias por bases es muy buena y se detecta la presencia del *Illumina Small RNA 3' Adapter*:

```
5'-TGGAATTCTCGGGTGCCAAGG-3'
```

Estos *adapters*, en el extremo 3' de las secuencias, se eliminarán en el paso siguiente.

## 5.4 Anotación de miRNAs.

Para la anotación de *miRNAs* se ha escogido la herramienta Miraligner [11]. Para la definición del proceso se toma como referencia el tutorial *miR annotation*[12].

### 5.4.1 Eliminación de adapters.

Se eliminan los *adapters* identificados anteriormente con la herramienta *cutadapt*.

### 5.4.2 Unir secuencias iguales en un solo registro.

Se recomienda unir en un solo registro las secuencias iguales, para facilitar su procesamiento. Para ello se utiliza la opción *collapse* de *seqcluster*. Una muestra de un registro de este tipo, una vez realizado *collapse*:

```
@seq_1_x12018
CATTGATCATCGACACTTCGAACGCACTTGCGGCCCGGGTT
+
BBBBEEEEGGGGGIIIIIIIIIIIIIIIIIIIIIIHIIGHHHIIGEF
```

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

### 5.4.3 Preparar la base de datos.

En principio, no es necesario descargar toda la base de datos de miRBase; en general sólo son necesarios los ficheros *hairpin.fa* y *miRNA.str*. Es altamente recomendable filtrar *hairpin.fa* para que contenga sólo la especie recomendada, en este caso ratón (*mmu*).

### 5.4.4 Anotación de miRNA/isomiR con Miraligner.

Para mapear las secuencias de las muestras de los 2 experimentos frente a la base de datos miRBase se utilizará la herramienta Miraligner. A la salida se generan 3 ficheros:

```
*.mirna.opt información del proceso.
*.nomap     secuencias no mapeadas
*.mirna     secuencias mapeadas
```

Ejemplo de un registro de uno de los ficheros \*.mirna:

```

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

```

Los campos del fichero \*.mirna son:

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

La nomenclatura empleada en los *isomiR* es:

- Sustitución de nucleótido:  
*NUMBER/NUCLEOTIDE\_ISOMIR/NUCLEOTIDE\_REFERENCE*
- Adiciones 3':  
*0/NA* indica que no hay modificación.  
*UPPER CASE LETTER* indica adición al final. Son modificaciones post-transcripcionales.
- Cambios 5':  
*0/NA* indica que no hay modificación.  
*UPPER CASE LETTER* indica inserción del nucleótido. La secuencia comienza antes de la posición de *miRNA* maduro.  
*LOWER CASE LETTER* indica delección del nucleótido. La secuencia comienza después de la posición del *miRNA* maduro.
- Cambios 3':  
*0/NA* indica que no hay modificación.  
*UPPER CASE LETTER* indica inserción del nucleótido. La secuencia finaliza después de la posición de *miRNA* maduro.  
*LOWER CASE LETTER* indica delección del nucleótido. La secuencia termina antes de la posición del *miRNA* maduro.

## 5.5 Control de calidad de muestras y normalización.

Los ficheros de salida \*.mirna de la herramienta *Miraligner* se utilizarán para el análisis de miRNAs/isomiR expresados diferencialmente. Para ello se empleará el entorno software R con paquetes propios del proyecto *Bioconductor* para análisis de experimentos de RNA-seq, como *DESeq2* [13, 14] y otros específicos del análisis de miRNAs/isomiRs, como *isomiRs* [15].

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.

### 5.5.1 Experimento 1 (ratones mecp2+/-).

Se cargan los datos de las 2 muestras de este experimento en un objeto *IsomirDataSeq*. Se comprueba mediante la matriz de *pseudocounts* que las muestras son bastante homogéneas en cuanto a distribución. No tiene sentido hacer normalización porque hay una sola muestra por condición experimental.

### 5.5.2 Experimento 2 (ratones foxg1+/-).

En este caso se dispone de 18 muestras, 9 por cada condición: ratones mutantes foxg1+/- y ratones normales de control (wt). En los controles de calidad de muestras se aprecia la necesidad de normalizar las muestras. Tras la normalización, se observa una desviación en la distribución de las muestras mutante1 y control1, por lo que se decide eliminarlas de la carga de datos en el objeto *IsomirDataSeq* que se utilizará para el análisis de expresión diferencial.

## 5.6 Análisis de expresión diferencial de miRNAs/isomiRs.

### 5.6.1 Experimento 1 (ratones mecp2+/-).

En este experimento no hay réplicas biológicas, por lo que no se puede hacer un análisis de expresión diferencial directamente con los paquetes *DESeq2* / *isomiRs*, que necesitan al menos, 3 réplicas por condición.

En el estudio de referencia [5] se menciona:

*"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."*

Sin embargo, en los datos publicados en el artículo aparecen 355 miRNAs, todos *up-regulated*; aunque con *logFC* mayor que 1.5 sólo hay 66. Se hace un análisis previo de los datos descargados y procesados en este trabajo sin diferenciar las variantes -3p y -5p de los miRNAs; para ello se trunca el nombre del miRNA, dejando sólo el nombre principal y se compara el resultado con los datos reportados. El resultado de esta comparación se resume en el siguiente diagrama.

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.



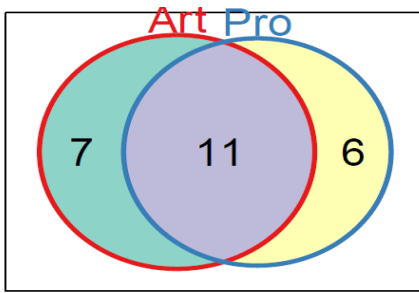


Figura 13. Diagrama de Venn comparando los resultados de miRNAs DE up-reg entre los datos Art) obtenidos del artículo de referencia y Pro) procesados y analizados en este trabajo.

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 utilizaron distintos procedimientos de análisis, todos incluidos en la plataforma Bioconductor.

El paquete *NOISeq* [16] 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*. Con este método, con un umbral de probabilidad de expresión diferencial  $> 0.9$ , se obtiene un conjunto de 43 miRNAs/isomiRs diferencialmente expresados entre ambas condiciones.

Para utilizar el paquete *DESeq2 / isomiRs* se necesitarían, al menos, 3 réplicas por condición. Se generan estas réplicas con el paquete *polyester* [17] y se realiza en análisis de expresión diferencial con *DESeq2 / isomiRs*; el resultado, con un criterio de *p-valor* ajustado  $< 0.1$ , es un conjunto de 14 miRNAs/isomiRs DE entre ambas condiciones.

Se realiza también el análisis con el paquete *RankProd*, que utiliza el test no paramétrico Rank Product (RP). Se seleccionan, con un criterio de *p-valor*  $< 0.05$ , 59 miRNAs/isomiRs down-reg en la clase mutante y 62 up-reg. En este último procedimiento de análisis es donde se obtiene un mayor número de miRNAs/isomiRs DE.

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 [5] por estar sobreexpresado en el síndrome de carencia de proteína MECP2.

### 5.6.2 Experimento 2 (ratones foxg1+/-).

En este experimento, al disponer de suficientes réplicas por condición, se hace el análisis de expresión diferencial directamente con el paquete *isomiRs*. Se seleccionan, con un umbral de *p-valor* ajustado  $< 0.1$ , un conjunto de 19 miRNAs/isomiRs diferencialmente expresados entre ambas condiciones.

Se hace una comparación con los datos publicados en el artículo de referencia [9] y se obtienen los resultados reflejados en el siguiente diagrama.

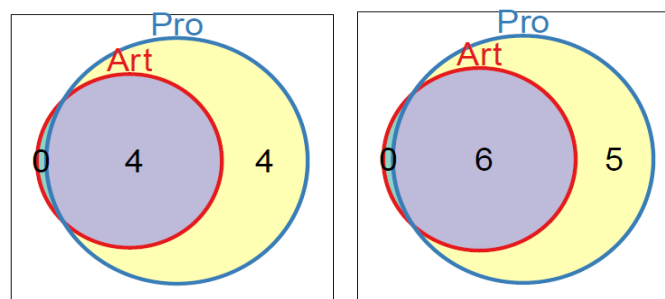


Figura 14. Diagramas de Venn comparando los miRNAs DE seleccionados en Art) Artículo de referencia y Pro) Procesados en este trabajo. A la izquierda los up-reg y a la derecha los down-reg.

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

Para comparar los miRNAs desregulados en ambos experimentos, mientras que en este segundo se comprueba que el procesamiento y análisis de datos en crudo de este trabajo obtiene datos muy similares a los ofrecidos por el artículo de referencia, en el primero se ha evidenciado la dificultad de obtener los mismos resultados y ante la ausencia de réplicas en el estudio se han valorado distintos análisis de expresión diferencial, observando que los resultados obtenidos con *RankProd* ofrecen un amplio espectro de miRNAs sobre el que trabajar. Además, los resultados del análisis con *RankProd* son los que indican una diferencia de expresión más significativa del miR-134 en ratones mecp2kd (aunque por encima del umbral pval establecido); que es uno de los miRNA destacado en el artículo de referencia como más relevante en este síndrome.

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

En el análisis hecho con *RankProd* sobre los datos del segundo experimento, con un umbral de pval < 0.05, se han obtenido 61 miRNAs down-reg en la clase mutante y 63 up-reg.

### 5.6.3 Comparación entre ambos experimentos

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

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

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

## 5.7 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.

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.

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, de manera resumida, son los siguientes:

Cruce de información	MiRNAs seleccionados		
Tabla 1 exp. 1 con tabla 1 exp. 2	miR-12191	miR-200a	miR-871
Tabla 2 exp. 1 con tabla 2 exp. 2	miR-148a	miR-423	miR-92a
Tabla 1 exp. 1 con tabla 2 exp. 2	miR-1298	miR-212	miR-34b(c)
Tabla 2 exp. 1 con tabla 1 exp. 2	-		

(para mostrar la información de manera más compacta, no se han tenido en cuenta las variantes -3p y -5p, que pueden ser diferentes en ambos experimentos).

## 5.8 Análisis detallado de isomiRs DE.

### 5.8.1 Experimento 1 (ratones mutantes *mecp2+/-*)

Para hacer un análisis detallado de isomiRs DE se seleccionan todas las variantes posibles de isomiRs en la función *isoDE*. El resultado es una lista de 7 isomiRs DE. Se consultan estas secuencias en la plataforma *Isomir Bank* [19], 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).

## 5.8.2 Experimento 2 (ratones mutantes *foxg1*+/-)

Se obtienen 28 isomiRs DE. 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).

## 5.9 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 2 p-valor=0,5. En el gráfico se representa la distribución de muestras a lo largo de las 2 primeras componentes (t1 y t2). El color rojo corresponde a las muestras mutantes y el azul a las de control.

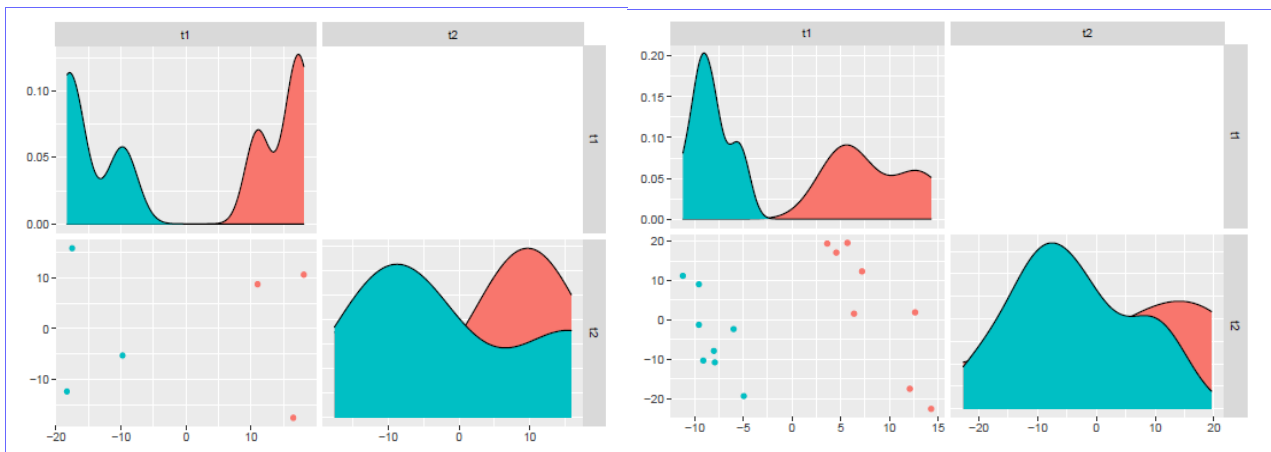


Figura 15. Representación PLS-DA de las muestras del experimento 1 (ratones *mecp2*+/-), a la izquierda y del experimento 2 (ratones *foxg1*+/-), a la derecha.

## 5.10 Integración gen-miRNA.

### 5.10.1 Herramientas de predicción de genes target.

Para la descripción y selección de las herramientas de predicción de *targets* se ha tomado como referencia el artículo [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, empezando desde el extremo 5' y continuando hacia el extremo 3'.

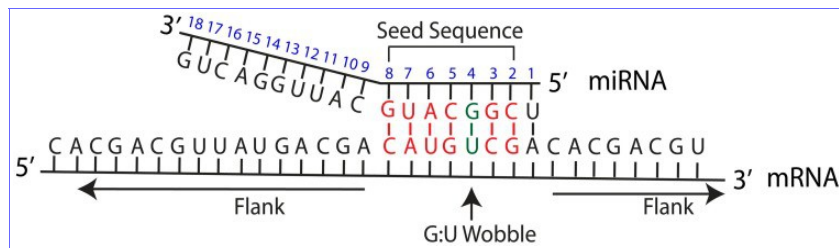


Figura 16. Interacción entre el miRNA y el mRNA target. Se representa la secuencia seed con los emparejamientos de nucleótidos WC (Watson Crick) marcados en rojo y un ejemplo de G-U wobble marcado en verde. Los flancos se refieren a las secuencias del mRNA a cada lado de la región target que hibrida 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).

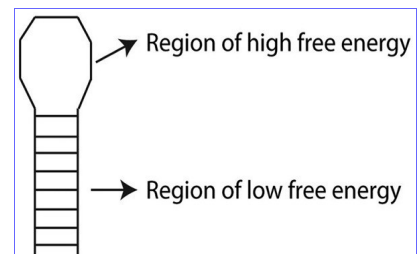


Figura 17. Estructura secundaria de RNA en forma de bucle (hairpin loop) como ejemplo de zonas de alta energía libre (menos estable) y baja energía libre (más estable) [20].

– 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.

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

Se emplearán distintas herramientas para la predicción de genes *target*.

La herramienta **TargetScan** [21] 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.

Para la predicción se utilizó 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. Con estos filtros, se han seleccionado 898 genes *target*.

La herramienta **miRanda** [23] 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 como *data.frame* en R para su tratamiento. 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). Con estos filtros se han seleccionado 589 genes *target*.

**DIANA MicroT-CDS** [24, 25] 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. 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. Con estos filtros han seleccionado 632 genes *target*.

**DIANA es TarBase** [26] a diferencia de las anteriores, basa sus resultados en *targets* miRNA validados experimentalmente en lugar de predicciones. 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*

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

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

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

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*.

## 5.11 Enriquecimiento GO.

Para hacer un enriquecimiento de categorías GO (Gene Ontology) se emplea la lista de genes *target* obtenida por la herramienta TargetScan. Para el proceso se utiliza la librería *clusterProfiler* de Bioconductor [27].

### 5.11.1 Enriquecimiento GO Experimento 1 (ratones mecp2+/-)

En un diagrama de puntos del resultado del enriquecimiento para los genes *target* seleccionados se comprueba que entre las 30 categorías GO más representadas en los se encuentran bastantes referentes a la diferenciación neuronal.

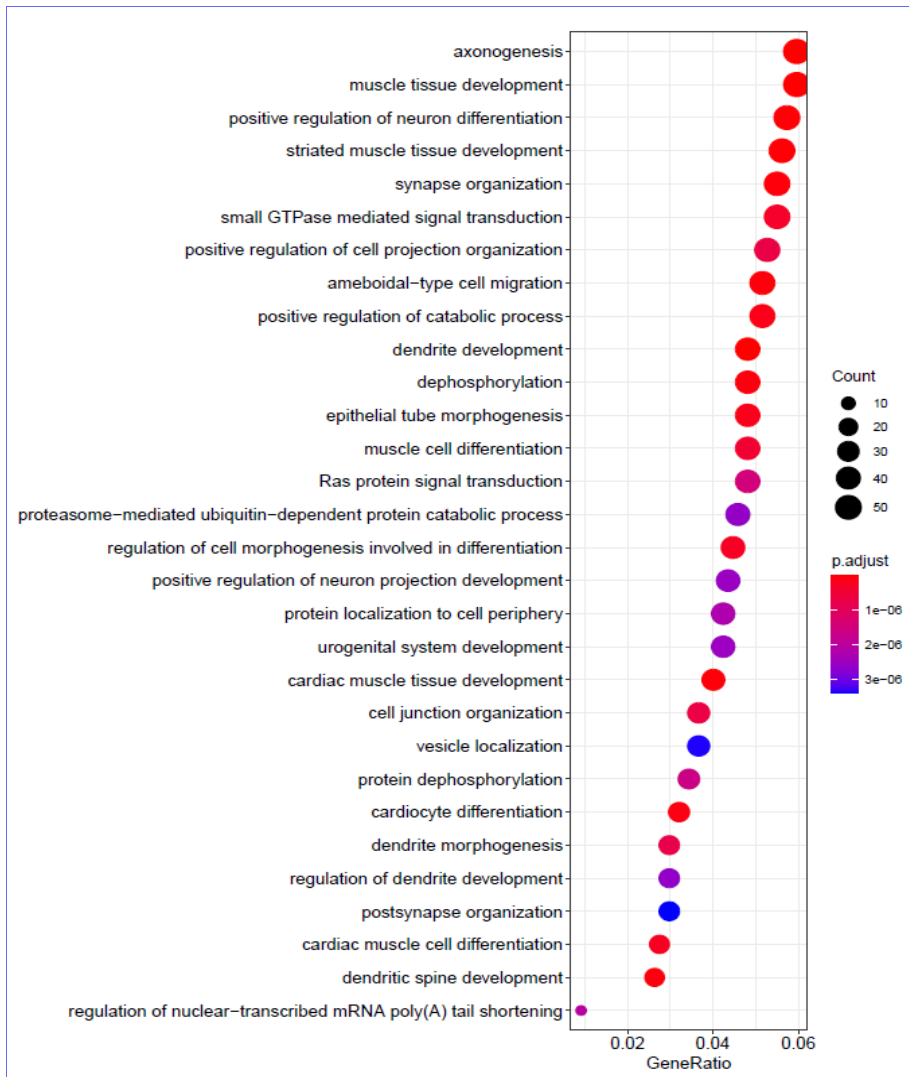


Figura 18. Diagrama de puntos de categorías GO más representativas entre los genes target seleccionados del experimento 1 (ratones mecp2+/-).

En el mapa de enriquecimiento se distinguen 2 redes importantes de categorías relacionadas, una referente a diferenciación neuronal y otra referente al desarrollo muscular.



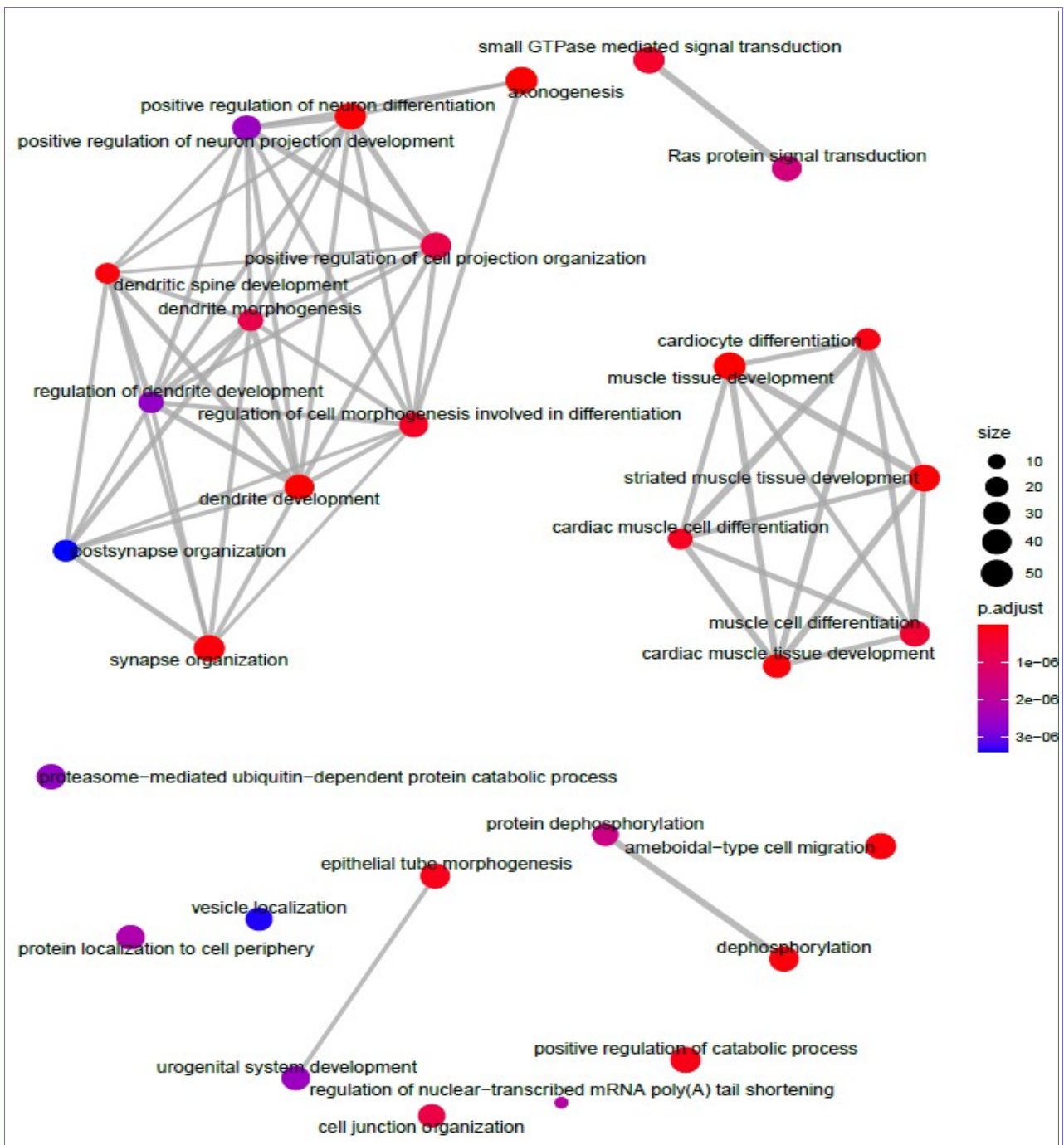


Figura 19. Mapa de enriquecimiento de categorías GO más representativas entre los genes target seleccionados del experimento 1 (ratones mecp2+/-).

En el diagrama de red se ve la relación entre genes target y categorías.

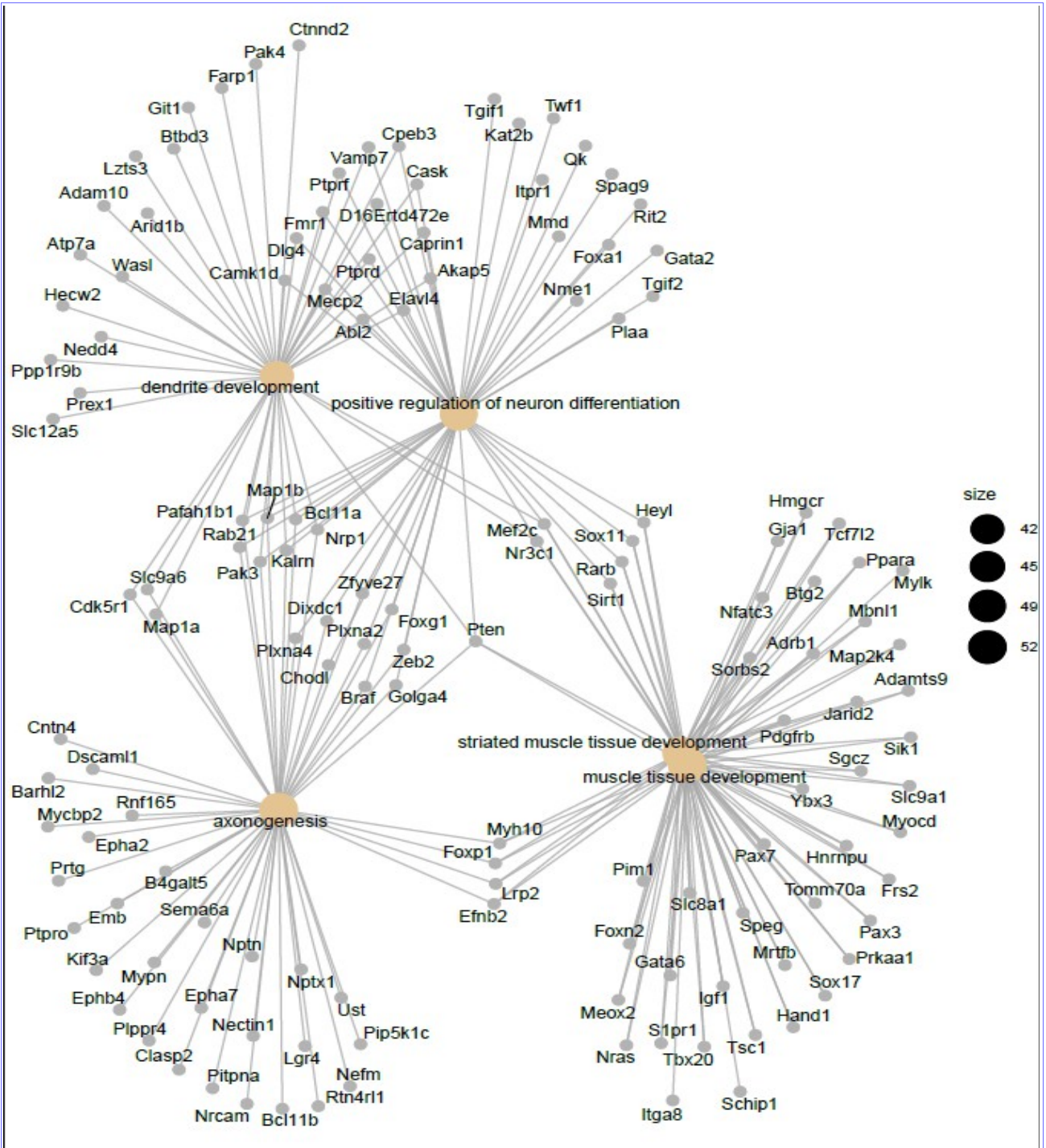


Figura 20. cnetplot de los genes target seleccionados en el experimento 1 (ratones *mecp2*<sup>+/-</sup>), en el que se puede ver la relación entre posibles genes target y procesos biológicos.

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

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

Hay 126 genes *target* diferentes implicados en estas funciones biológicas.

#### 5.11.2 Enriquecimiento GO Experimento 2 (ratones *foxc1+/-*)

Se procede de igual manera que en el experimento 1, se aprecia en el mapa resultante del enriquecimiento GO de los genes *target* seleccionados que aparecen 3 redes diferentes de características relacionadas con el desarrollo del sistema nervioso.

En la primera de ellas se encuentran las categorías:

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

Hay 147 genes *target* diferentes implicados en estas funciones biológicas. Se hace un mapa de red bipartita con estos genes *target* y su relación con miRNAs.

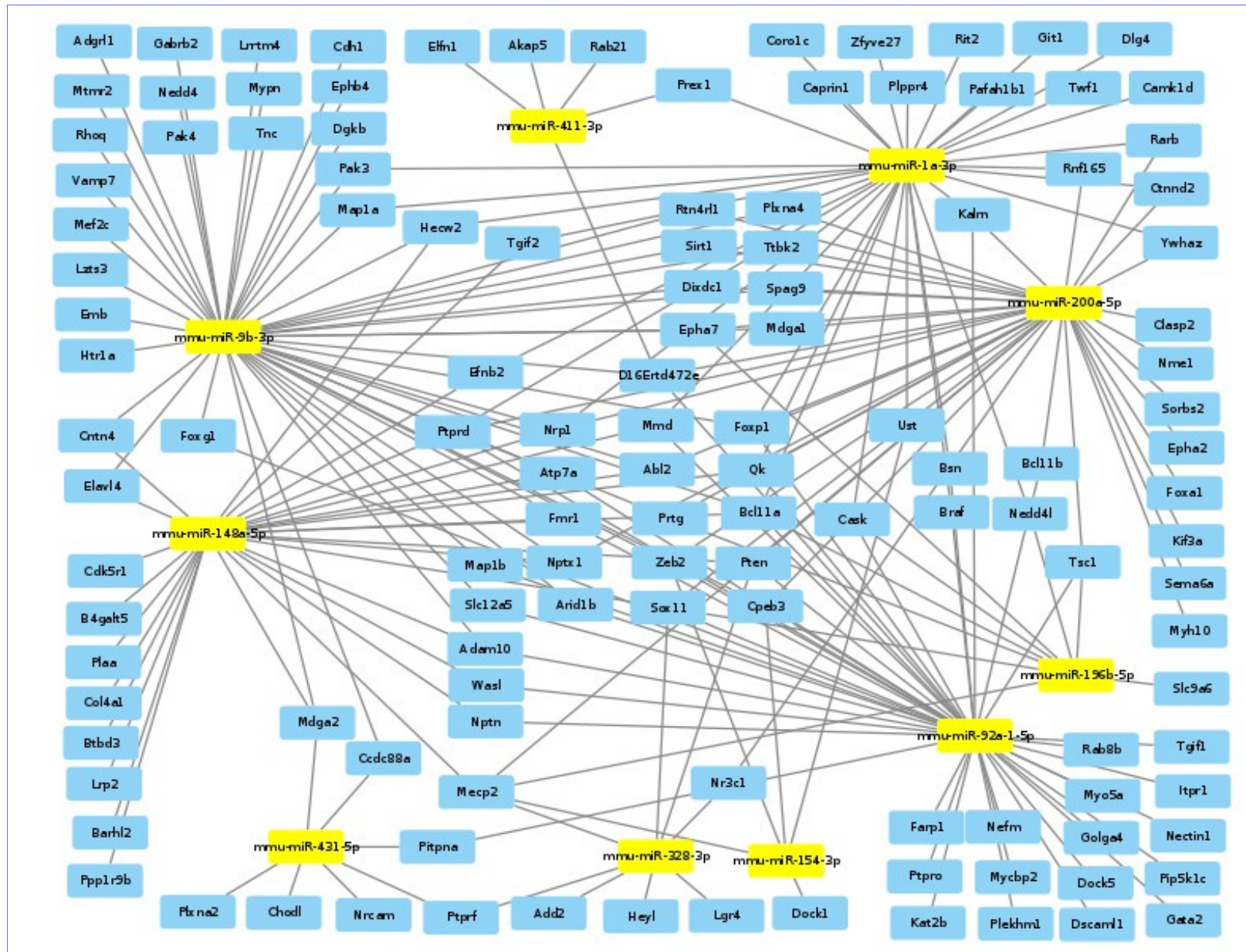


Figura 21. Mapa de red bipartita de miRNAs DE y genes target seleccionados del experimento 1 (ratones mecp2+/-).

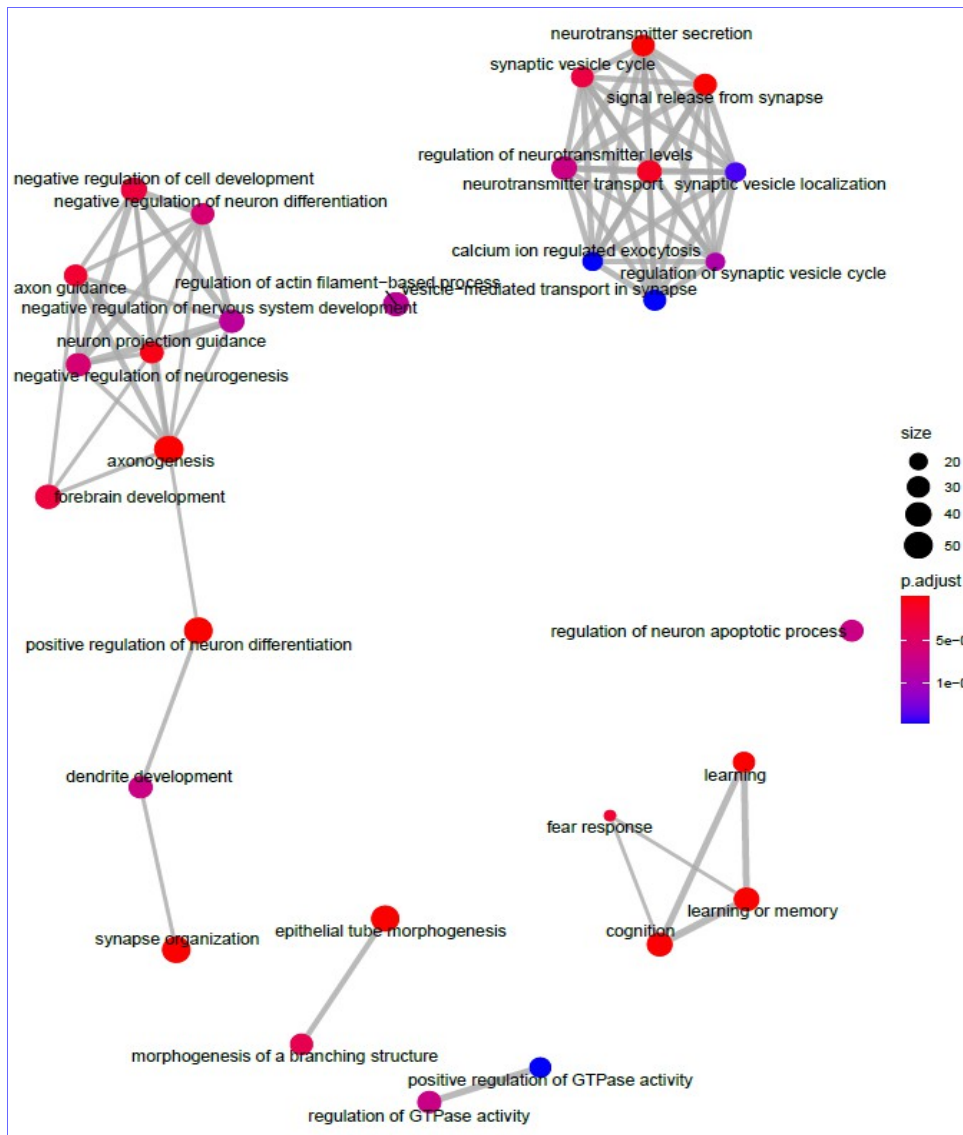


Figura 22. Mapa de enriquecimiento de categorías GO más representativas entre los genes *target* seleccionados del experimento 2 (ratones *foxg1*<sup>+/-</sup>)

En la segunda se encuentran las categorías:

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

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

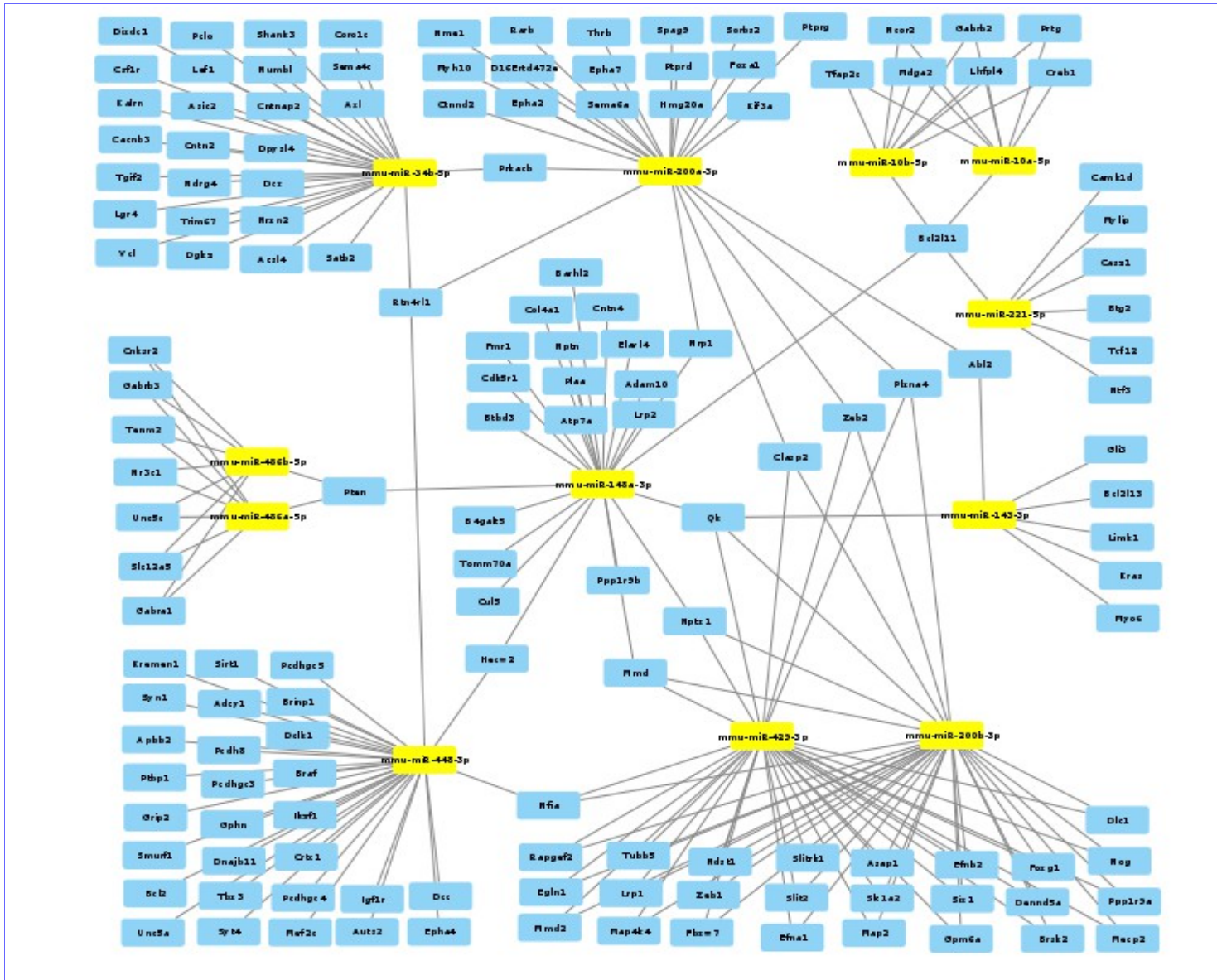


Figura 23. Mapa de red bipartita de miRNAs DE y genes target seleccionados en el grupo 1 del experimento 2 (ratones foxg1+/-).

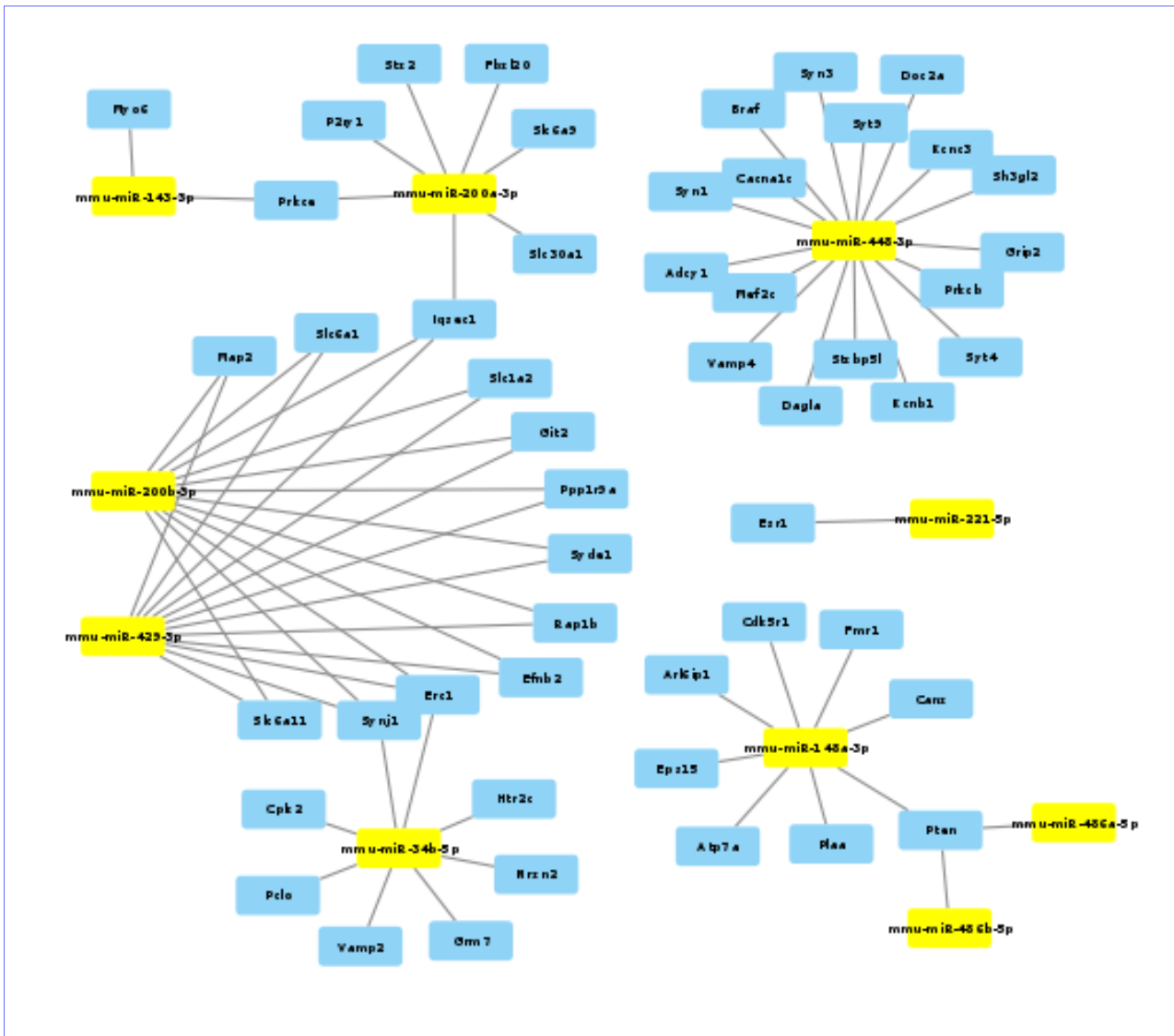


Figura 24. Mapa de red bipartita del grupo 2 del experimento 2 .

En la tercera se encuentran:

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

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

#### 5.11.4 Cruce de ambos experimentos.

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

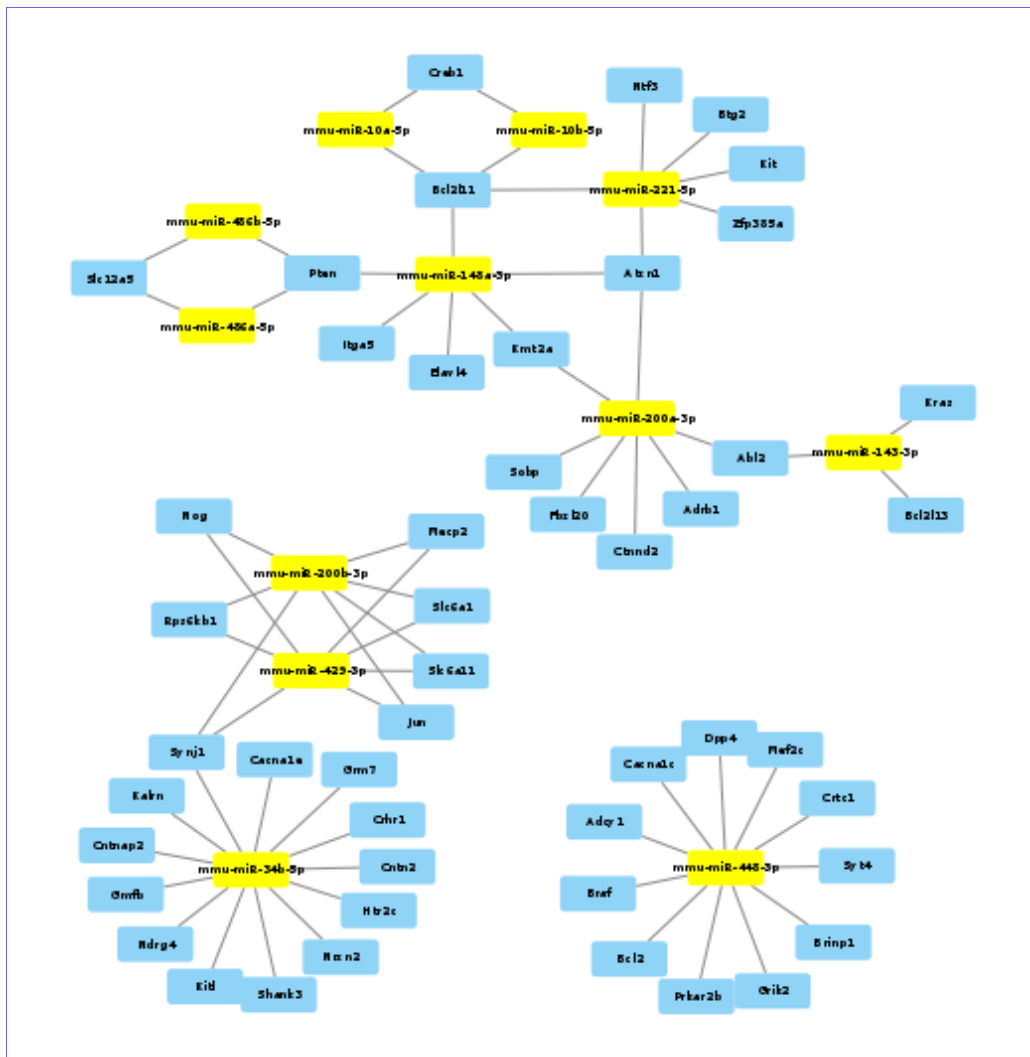


Figura 25. Mapa de red bipartita del grupo 3 del experimento 2 .

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

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

Hay 127 genes target diferentes implicados en estas funciones biológicas.



En el segundo agrupamiento se encuentran las siguientes categorías:

Categoría	Descripción	Nº 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 implicados en estas funciones biológicas.

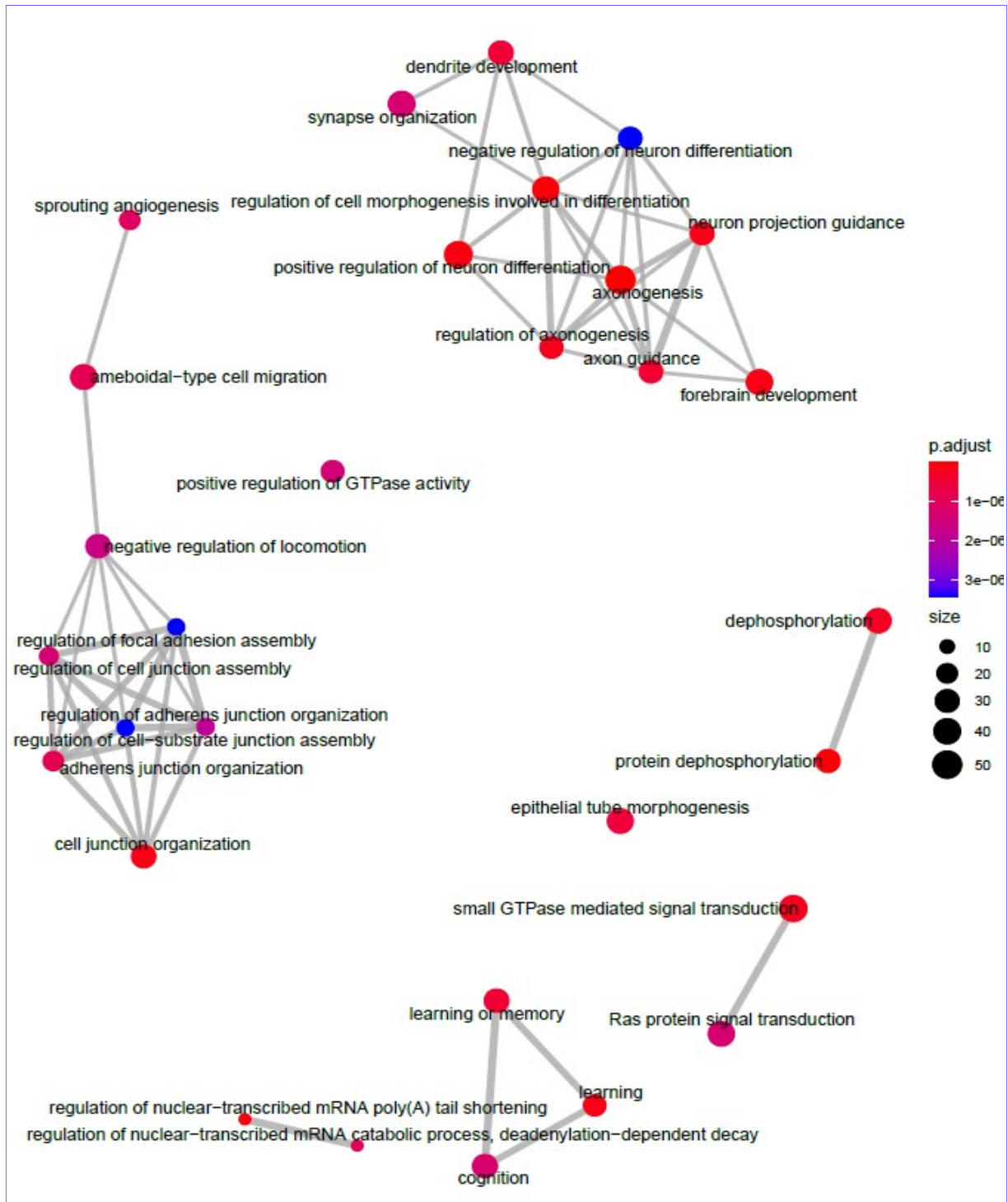


Figura 26. Mapa de enriquecimiento de los genes target seleccionados en el cruce de los 2 experimentos.

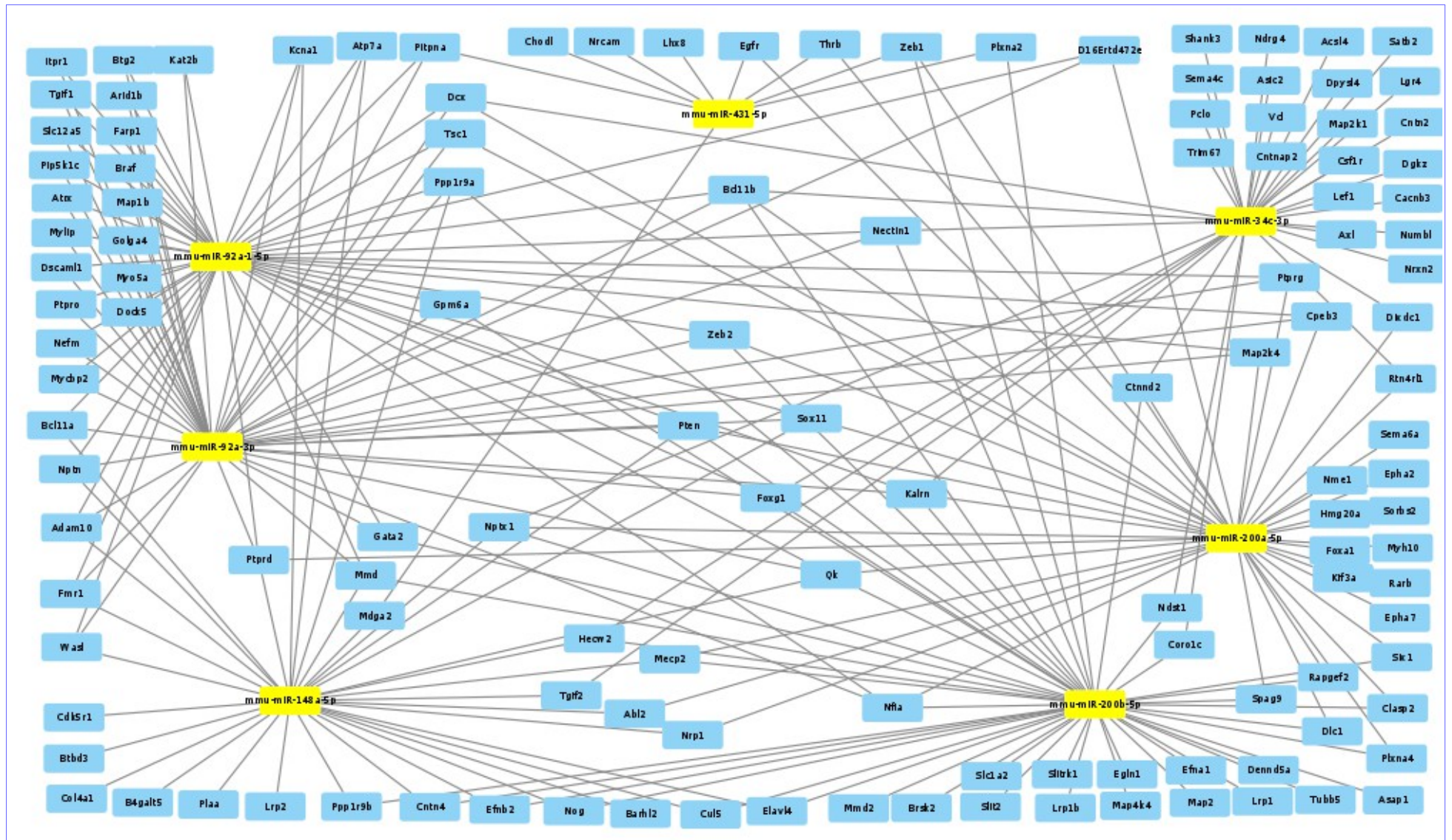


Figura 27. Mapa de red bipartita de los miRNAs y genes target seleccionados en el grupo 1 de enriquecimiento GO del cruce de los 2 experimentos.

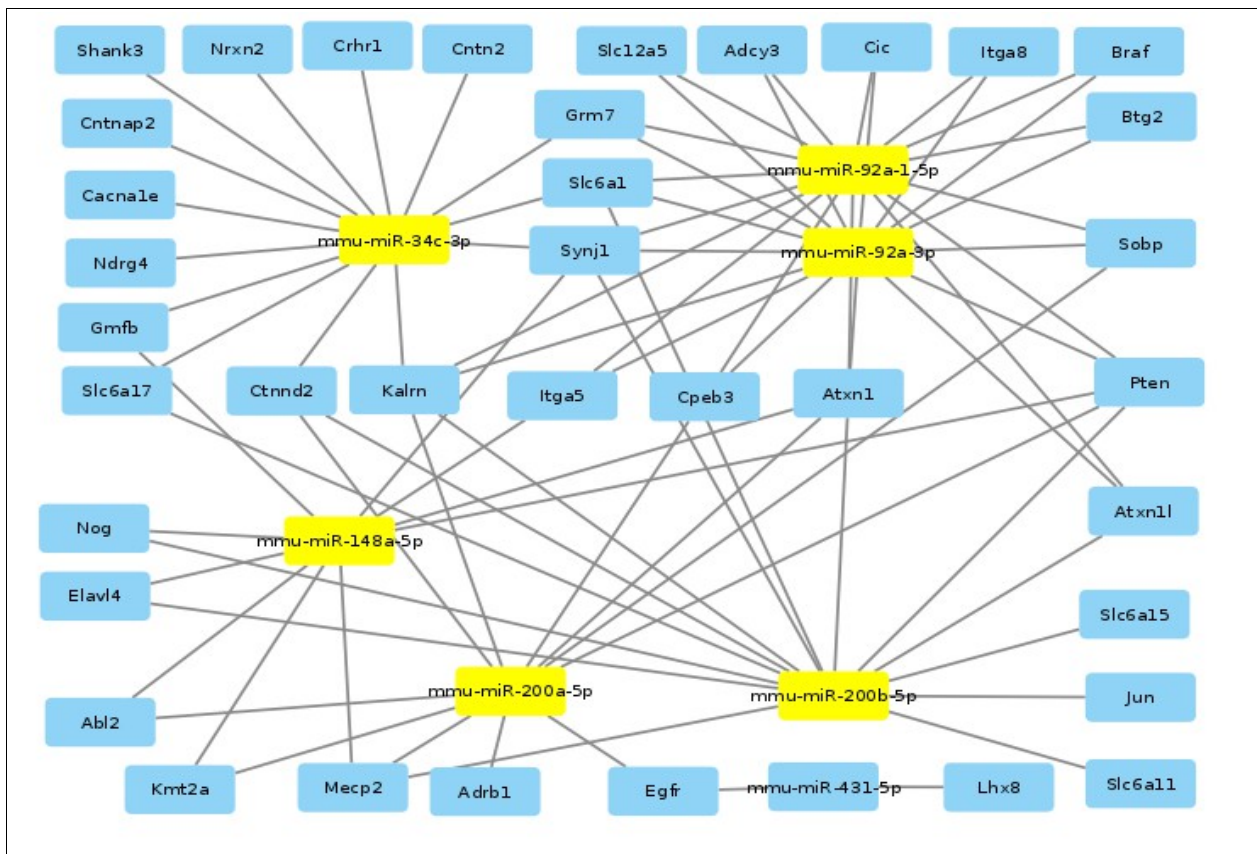


Figura 28. Mapa de red bipartita del grupo 2 de categorías GO en el cruce de los dos experimentos.

## 5.12 Enriquecimiento KEGG.

El enriquecimiento KEGG (<https://www.kegg.jp/>; Kyoto Encyclopedia of Genes and Genomes) [28] también se realiza con el paquete *clusterProfiler* de Bioconductor [27].

### 5.12.1 Enriquecimiento KEGG en el experimento 1 (ratones *mecp2*+/-).

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

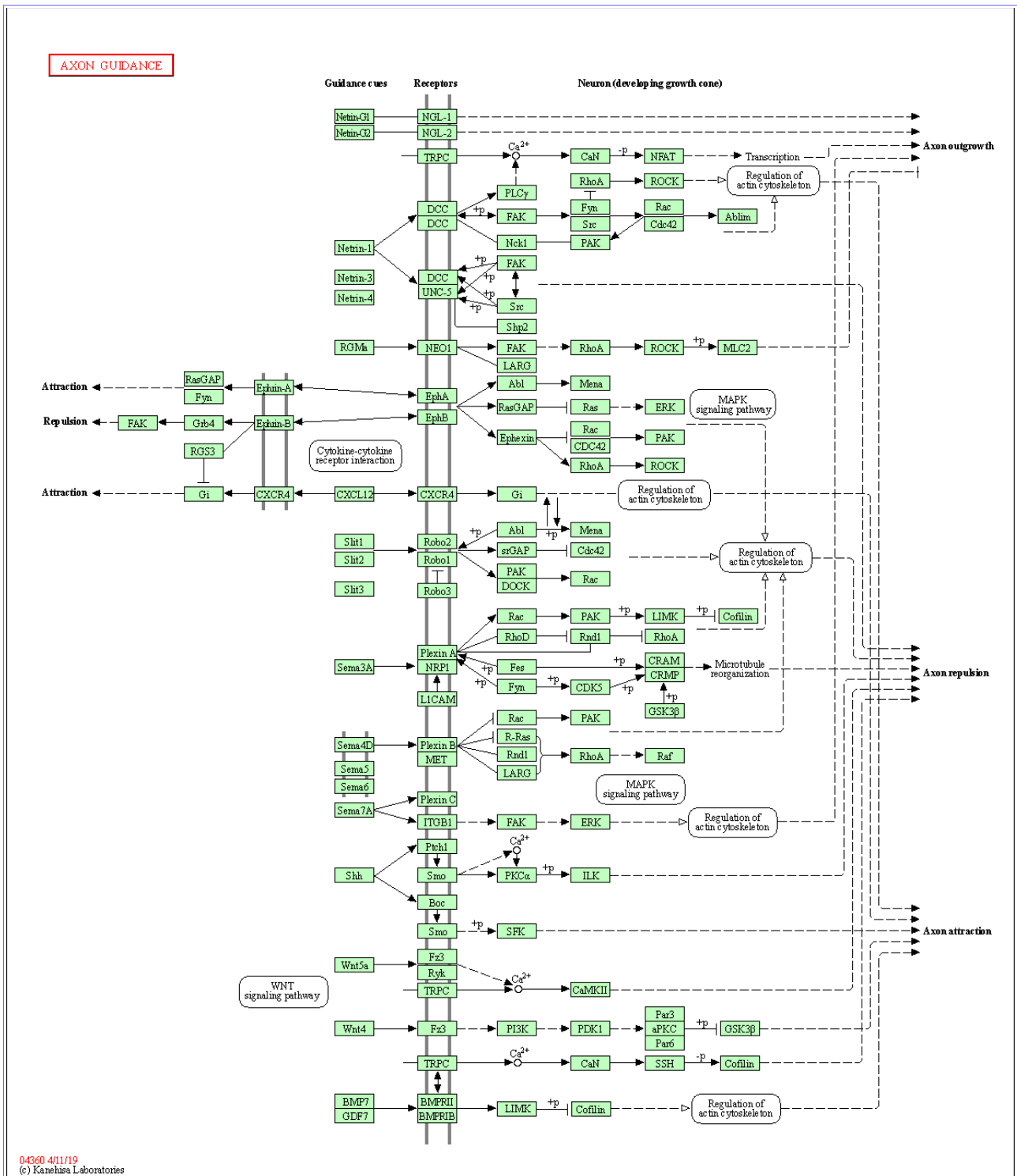


Figura 29. Esquema del pathway KEGG mmu04360, Axon guidance. [28]

En el esquema general *Axon guidance* se puede ver que interviene también el *pathway Regulation of actin cytoskeleton* 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 tirosina 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.

### 5.12.2 Enriquecimiento KEGG en el experimento 2 (ratones *foxg1*+/-).

Se realiza el mismo proceso de enriquecimiento KEGG para los genes *target* seleccionados en el experimento 2, en los 3 grupos de categorías GO señaladas anteriormente. Para el primer grupo 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 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.

En el esquema de este *pathway* se puede ver que 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.

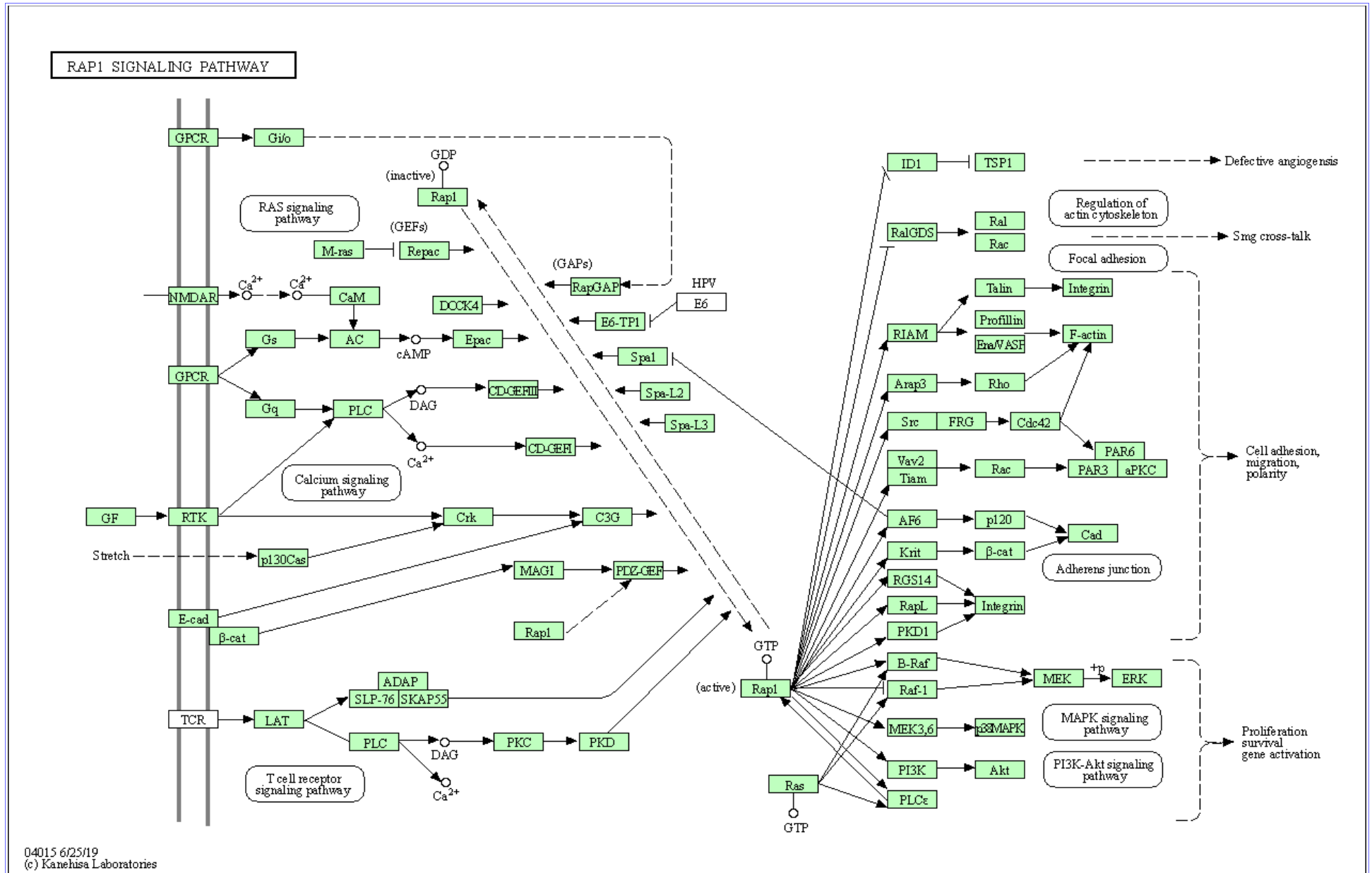


Figura 30. Esquema del pathway KEGG mmu04015, Rap1 signaling pathway [28]

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*.

### 5.12.3 Enriquecimiento KEGG 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>

Estos *pathways* ya se han mencionado anteriormente.

## 6. ANÁLISIS DE RESULTADOS

### 6.1 Experimento1 (ratones *mecp2+/-*)

Para analizar los resultados del experimento 1, lo primero que se hace es acotar los resultados para reducir el número de miRNAs y genes *target* de estos miRNAs, escogiendo los más significativos y de esta manera centrar 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-92a1-5p, 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 *seed match* 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" "Hey1" "Lgr4" "Lrrtm4" "Lzts3" "Map1a" "Mdga1" "Mdga2"
[25] "Mecp2" "Mmd" "Mmd2" "Mtmr2" "Mypn" "Nptx1" "Nr3c1" "Nrpl"
[33] "Pafah1b1" "Pak3" "Pak4" "Pitpna" "Prex1" "Prtg" "Pten" "Rab21"
[41] "Rit2" "Rnf165" "Tgif1" "Tgif2" "Tnc" "Ttbk2" "Twf1" "Ust"
[49] "Ywhaz" "Zeb2" "Zfyve27"
```

El mapa bipartito, reducido a estos genes target es el siguiente:

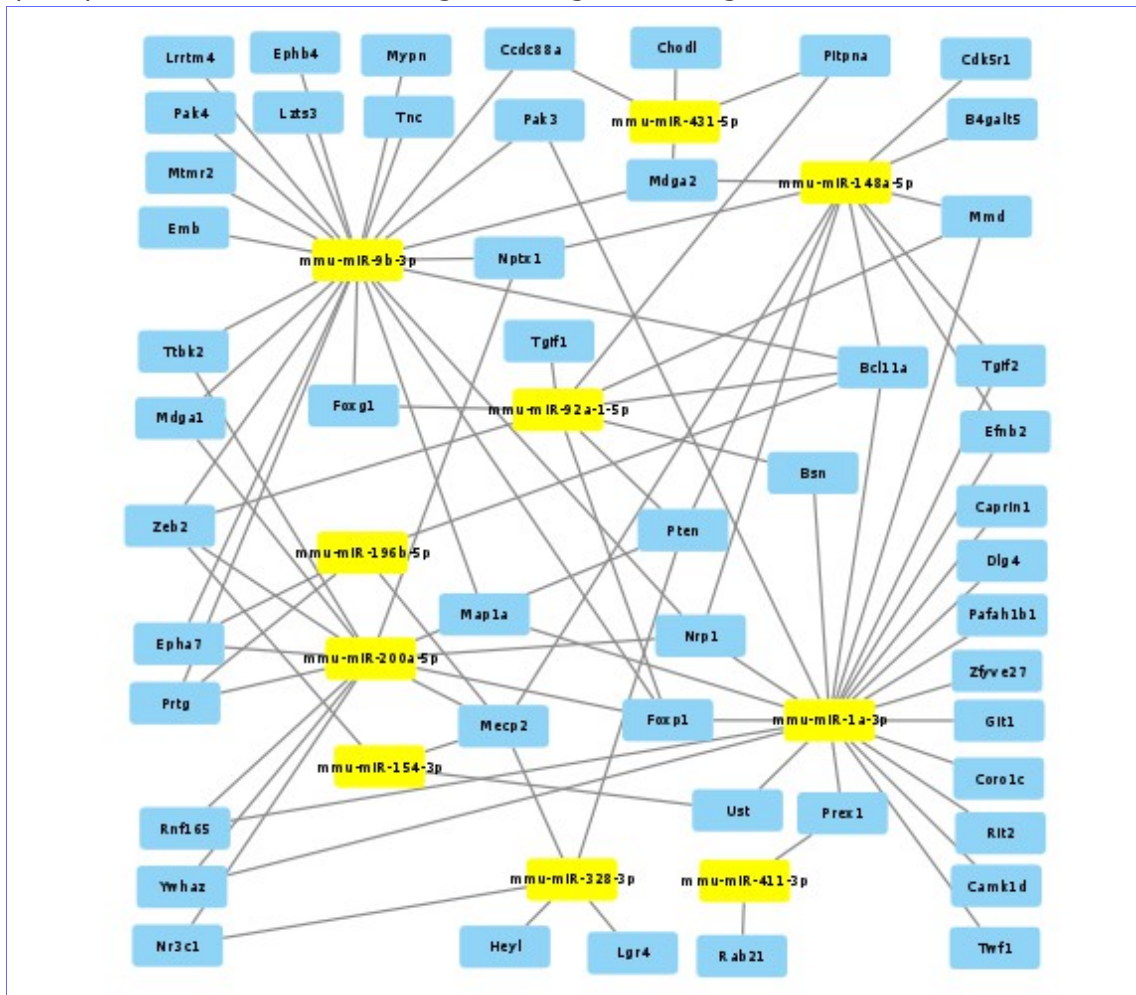


Figura 31. Mapa de red bipartita de miRNAs y genes target seleccionados como más significativos para el experimento 1 (ratones mecp2+/-).

Se hace un análisis de enriquecimiento KEGG con estos 51 genes y se obtiene como significativo los *pathways*:

*Axon guidance*, en el que están implicados 6 de éstos genes: Nrp1, Efnb2, Epha7, Ephb4, Pak3 y Pak4.

*Focal adhesion*, en el que están implicados 4 de éstos genes: Pak3, Pak4, Tnc y Pten.

En el mapa de red bipartita, se puede ver que 6 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* [29]

[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 y Pak4 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.

Además de los resultados de enriquecimiento GO y KEGG obtenidos por clusterProfiler, se ha hecho un análisis de *pathways* con pathfindR [30], 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.

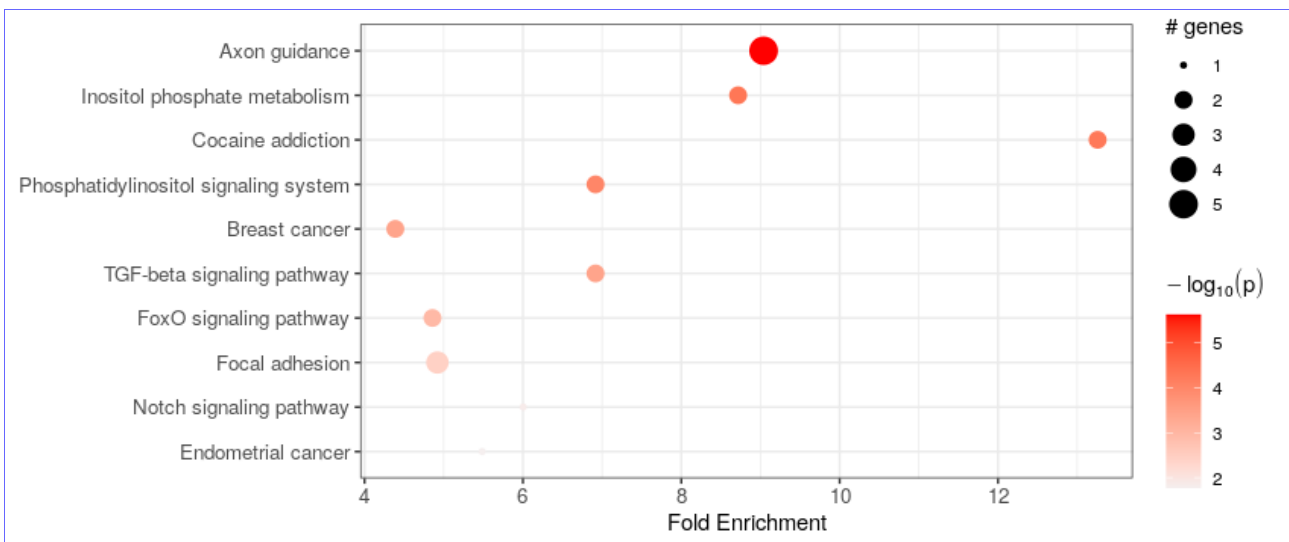


Figura 32. Gráfico de enriquecimiento obtenido por pathfindR para los genes target seleccionados de los miRNAs desregulados en el experimento 1.

A parte del *pathway Axon guidance*, que aparece como el más significativo y que ya se ha detallado anteriormente, aparecen otros de interés.

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.

## 6.2 Experimento2 (ratones foxg1+/-)

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 llega a los siguientes resultados, diferenciados por los grupos establecidos de categorías GO.

Se seleccionan los siguientes 47 genes *target* del grupo 1:

```
[1] "Abl2"    "Apbb2"   "Asap1"   "Auts2"   "Bcl2"    "Brinp1"  "Clasp2"  "Crtc1"   "Dlc1"
[10] "Efnb2"   "Egln1"   "Epha4"   "Fbxw7"   "Foxa1"   "Foxg1"   "Gphn"    "Grip2"   "Hecw2"
[19] "Igflr"   "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"
```

Por un análisis de enriquecimiento KEGG de estos genes se obtienen como más significativo el *pathway Axon guidance*, con los genes implicados Plxna4, Sema6a, Efnb2, Epha4 y Unc5a.

El gen Efnb2 y otros genes codificadores de receptores de efrinas similares a Epha4, se encontraban como genes *target* significativos en el experimento 1.

La proteína codificada por Plxna4 tiene actividad como receptor de semaforinas.

Sema6a codifica una semaforina presente en el tejido neuronal en desarrollo.

Unc5a codifica un receptor que media el efecto quimio-repulsivo de netrin-1, una netrina presente en axones.

Para el grupo 2 se seleccionan los siguientes 20 genes *target*:

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

Que por análisis KEGG, dan como resultado los *pathways*:

*Synaptic vesicle cycle*; genes implicados: Stx2, Slc6a1 y Slc6a11.

*GABAergic synapse*; genes implicados: Slc6a1, Slc6a11 y Cacna1c.

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.

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.

Para el grupo 3 se seleccionan los siguientes 18 genes *target*:

```
[1] "Abl2"    "Adrb1"   "Bcl2"    "Brinp1"  "Cacna1c" "Crtc1"   "Dpp4"    "Grik2"   "Jun"
[10] "Kmt2a"   "Mef2c"   "Nog"     "Prkar2b" "Rps6kb1" "Slc6a1"  "Slc6a11" "Synj1"   "Syt4"
```

Que por análisis KEGG, dan como resultado 7 *pathways*, entre los que el más significativo es *ErbB signaling pathway*, con los genes implicados *Abl2*, *Jun* y *Rps6kb1*.

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

*Jun*. Subunidad de factor de transcripción, con procedencia posiblemente vírica, que interactúa directamente con secuencias DNA target para regular expresión génica.

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

Para un análisis más completo con pathfindR, se prepara el dataset de entrada con los genes agrupados 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.

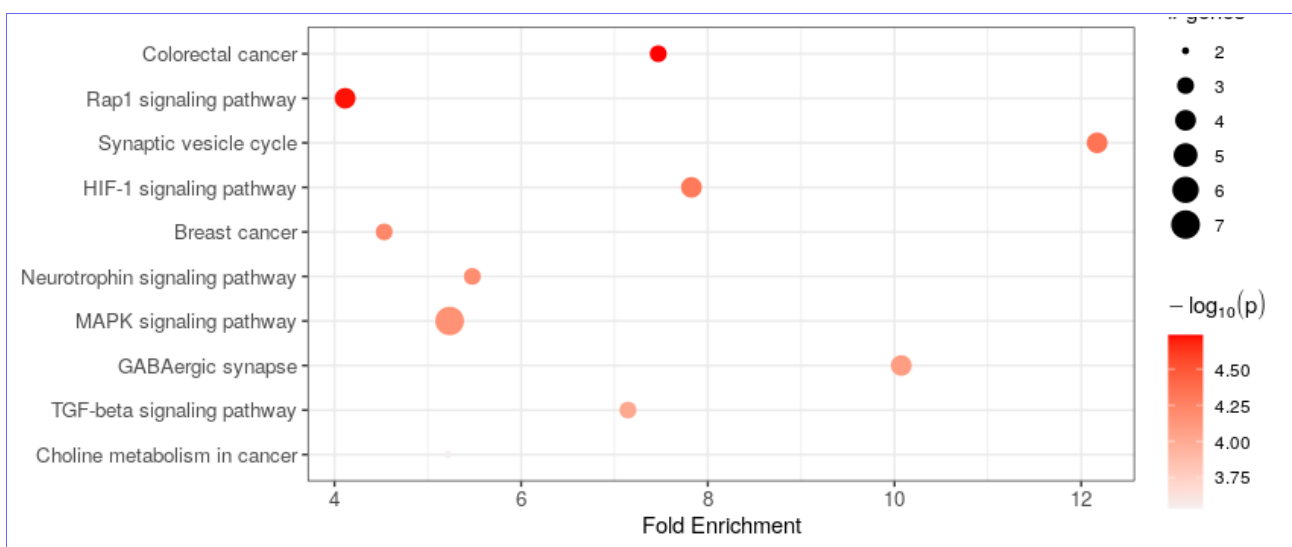


Figura 33. Gráfico de enriquecimiento obtenido por pathfindR para los genes target seleccionados de los miRNAs desregulados en el experimento 2.

A parte de los *pathways Synaptic vesicle cycle* y *GABAergic synapse*, que ya han sido mencionados anteriormente, se destacan los siguientes *pathways*:

*Rap1 signaling pathway*, que ya fue mencionado como *pathway* significativo en los resultados preliminares de este experimento, antes de hacer la selección de miRNAs/genes target más significativos. En él están implicados los genes *Igf1r*, *Rap1b*, *P2ry1* y *Rapgef2*.

*Igf1r* (insulin-like growth factor I). Este receptor tiene actividad tirosina-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.

**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*, *Bcl2*, *Rps6kb1* y *Egln1*.

*Igf1r* ya se mencionó como desregulado en el *Rap1 signaling pathway* y *Rps6kb1* en *ErbB signaling pathway*.

*Bcl2* es un regulador de la apoptosis.

*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 *Bcl2*, *Rap1b* y *Jun*.

**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*, *Mef2c*, *Rapgef2*, *Rap1b*, *Map4k4* y *Jun*.

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

### 6.3 Cruce de los dos experimentos.

Para acotar los resultados de los análisis se procede de forma similar a la descrita en los dos experimentos 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), miR-200b-5p (-3p) y miR-34c-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.

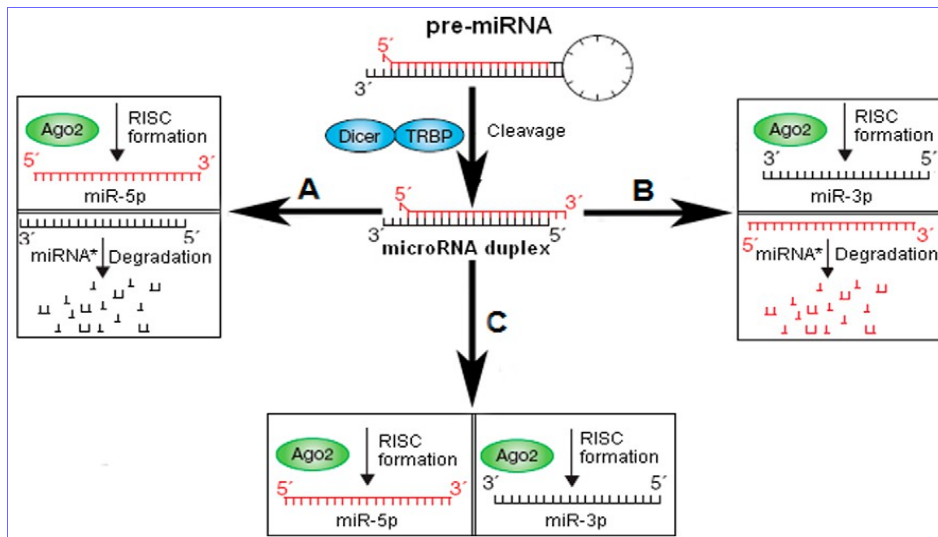


Figura 34. Selección de la cadena de miRNA madura. A) La cadena 5' forma el miRNA-5p maduro y la cadena 3' (miRNA\*) se degrada. B) La cadena 3' forma el miRNA-3p maduro y la cadena 5' (miRNA\*) se degrada. C) Ambas cadenas miRNA-5p y miRNA-3p son maduras y forman complejos RISC de inhibición post-transcripcional de sus genes target [32].

En el proceso de maduración del miRNA, el precursor, una vez procesado por la endonucleasa Drosha 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*).

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 pasajero, sino que se consideran las 2 formas, miRNA-5p y miRNA-3p como activas [32].

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]	"Ab12"	"Acs14"	"Arid1b"	"Asap1"	"Atp7a"	"Ax1"
[7]	"B4gal15"	"Bcl11a"	"Braf"	"Btd3"	"Btg2"	"Cdk5r1"
[13]	"Chod1"	"Clasp2"	"Cntn2"	"Cntnap2"	"Cpeb3"	"Csflr"
[19]	"Cul5"	"Dl6Ert472e"	"Dcx"	"Dixdc1"	"Dlc1"	"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"	"Myliip"
[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.

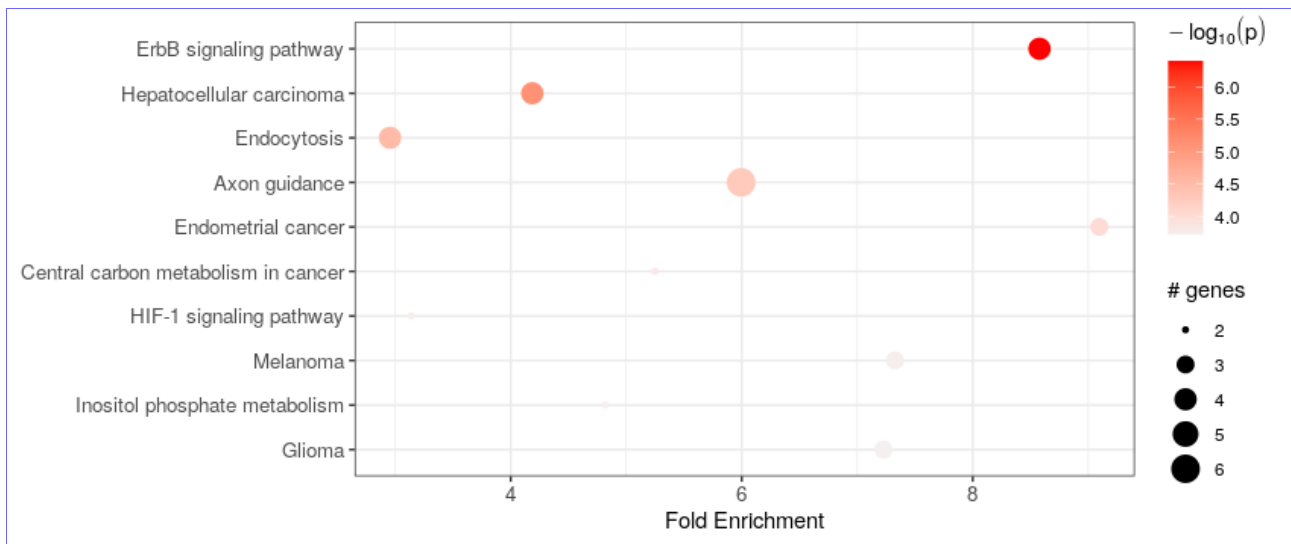


Figura 35. Gráfico de enriquecimiento obtenido por pathfindR para los genes target seleccionados de los miRNAs desregulados en el cruce de ambos experimentos.

Aparece como más significativo el *ErbB signaling pathway*, con los genes implicados Abl2, Egfr, Map2k4 y Braf.

Abl2 ya fue descrito anteriormente dentro de este mismo *pathway*.

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 tirosina 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.

En la figura se puede ver un esquema de este pathway, que 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, Sema6a, Plxna2 y Nrp1.

## 7. CONCLUSIONES.

En el primer experimento se analizaron secuencias de miRNAs en ratones *Mecp2+/-*, comparadas con las de ratones normales. En el estudio de referencia [5] 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 diferencialmente expresados los siguientes:

*Down-reg*: miR-200a

*Up-reg*: miR-328, miR-1a, miR-92a-1, miR411, 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 estudiado con ratones Foxg1+/-.

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.

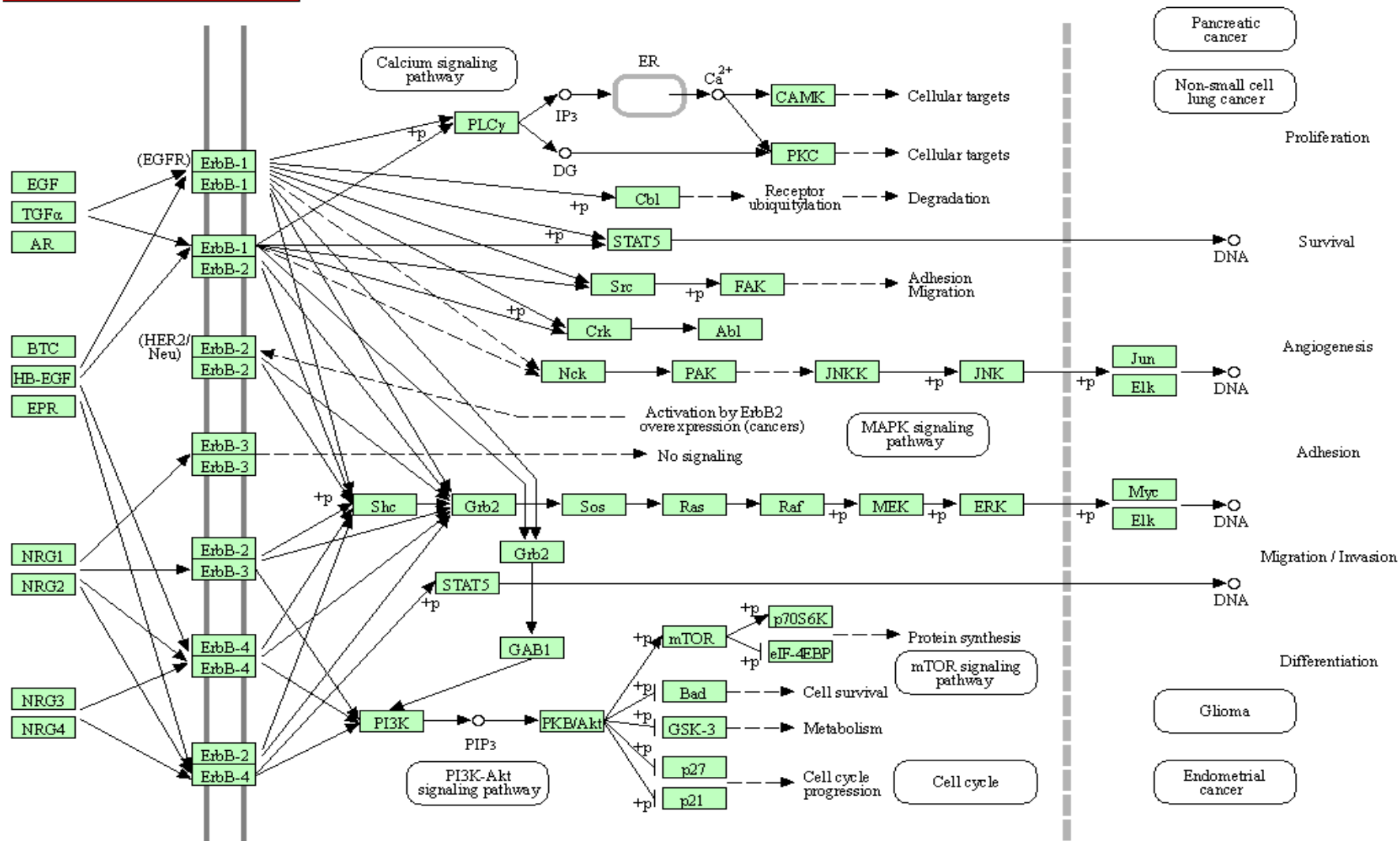
Entre estos *pathways* destaca el guiado del axón (*axon guidance*), con posibles genes *target* implicados como Nrp1, que codifica el neuropilin, un receptor de semaforinas, asociado al neuroma y Efnb2, que codifica un miembro de la familia de las efrinas y aparece también como gen posiblemente desregulado 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 TGF $\beta$  (*Transforming Growth Factor  $\beta$* ), 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 los genes mencionados. 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.

**ERBB SIGNALING PATHWAY**



04012 3/16/18  
(c) Kanehisa Laboratories

Figura 36. Esquema del pathway KEGG mmu04012, ErbB signaling pathway [28]



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. 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	(Nrp1)	Efnb2 Pak3 EphA7 Ephb4 (Nrp1)
PI sig. system		Pten Mtmr2
Focal adhesion		Pten Pak3 Tnc
TGF-beta sig. p.		Tgif1 Tgif2
FoxO sig. path.		Pten Foxg1

En el segundo experimento se analizaron secuencias de miRNAs en ratones *Foxg1+/-*, comparadas con las de ratones normales. En el estudio de referencia [9] 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 (proteín 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. 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	
GABAergic syn.	Slc6a1 Slc6a11	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. path	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.

En tercer lugar, una serie de *pathways* de señalización relacionados.

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* 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 aRTT. 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 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 *target* 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. 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 serina/treonina 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.

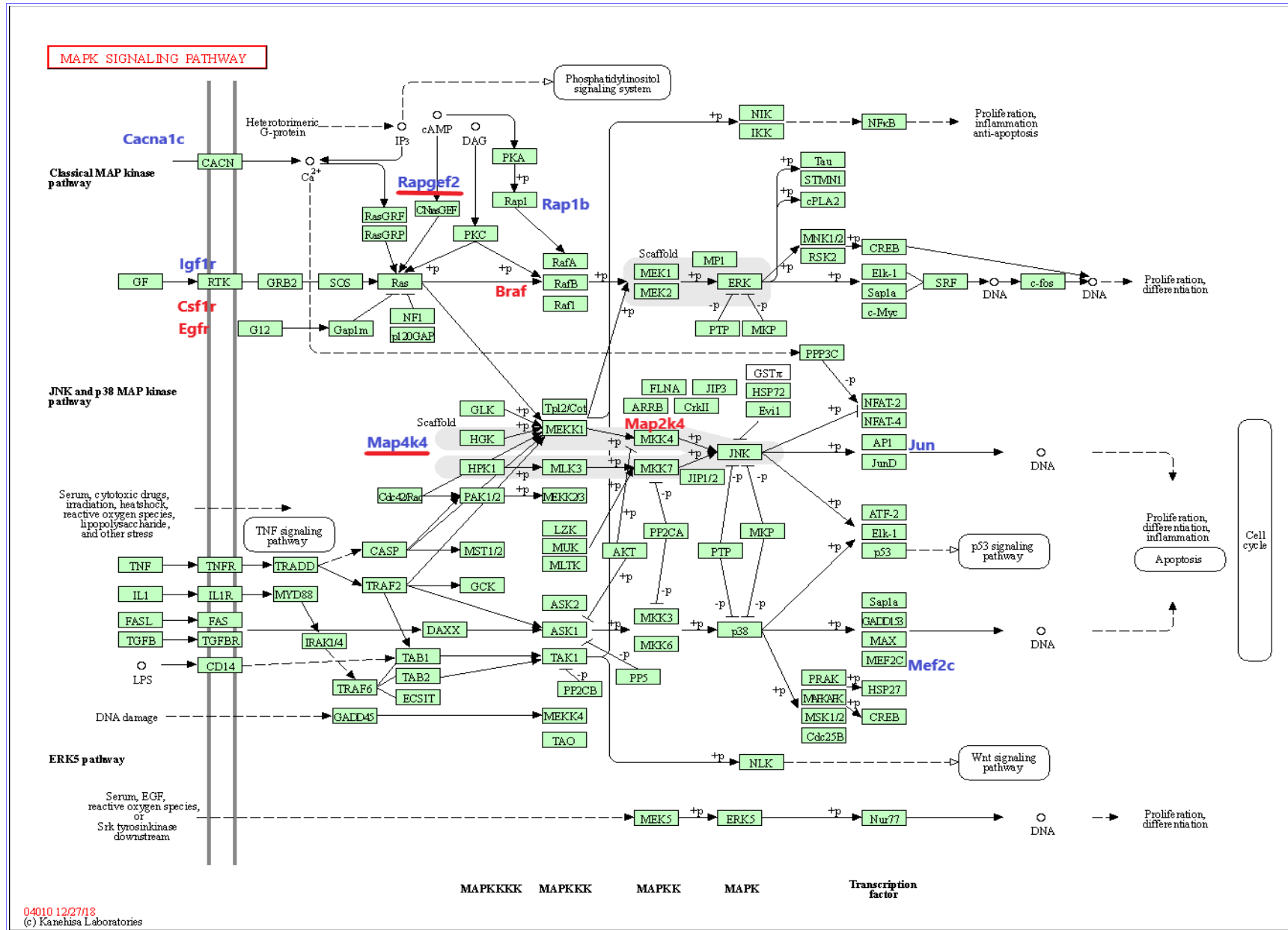


Figura 37. Esquema del pathway KEGG mmu04010, MAPK signaling pathway [28]

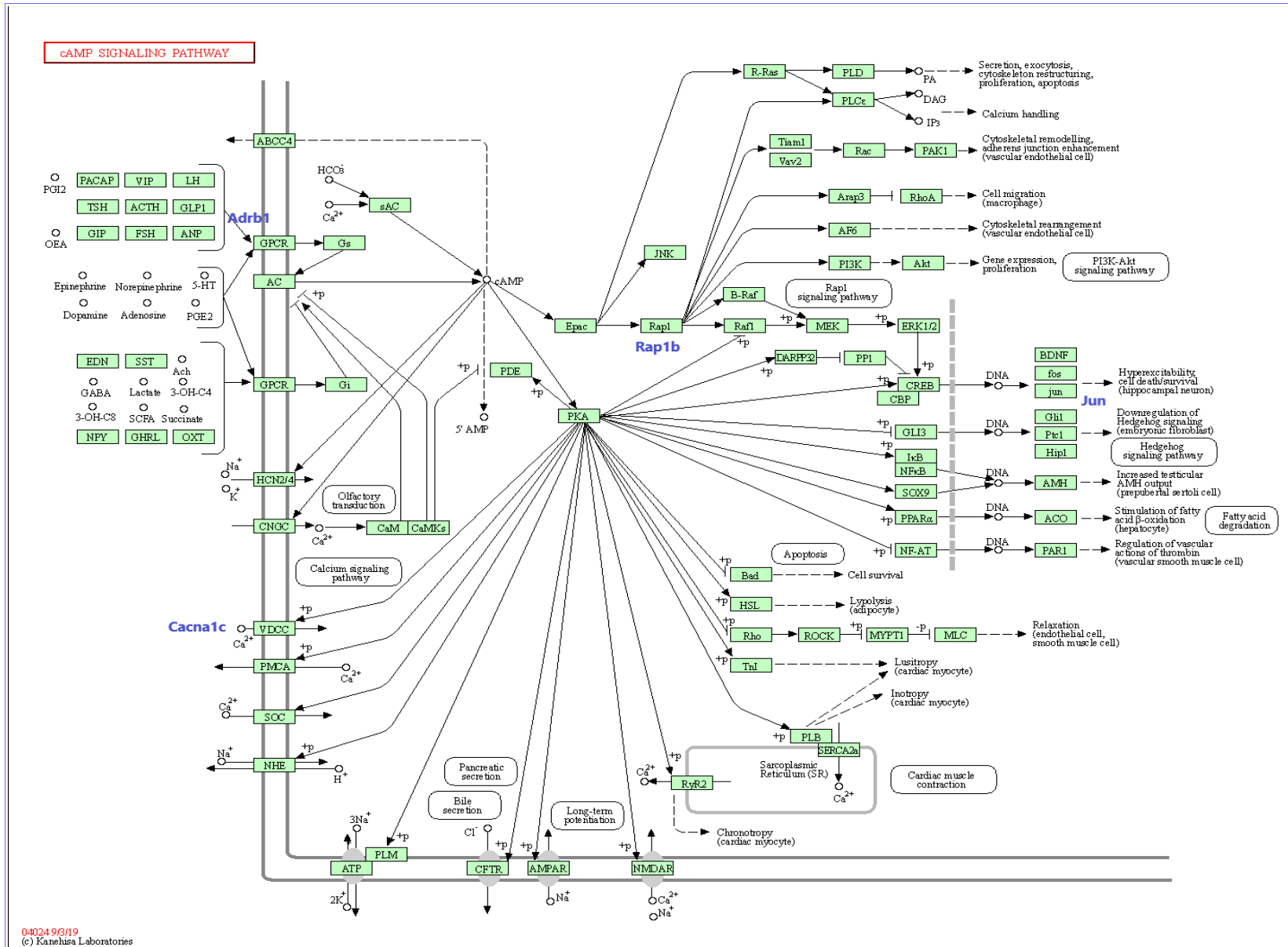


Figura 38. Esquema del pathway KEGG mmu04024, cAMP signaling pathway [28]

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. 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 anormalidad cerebral, neurodegeneración y disosteosclerosis (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 [9], 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 [31]:

- 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. 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 quinasa) 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.

Esta 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.

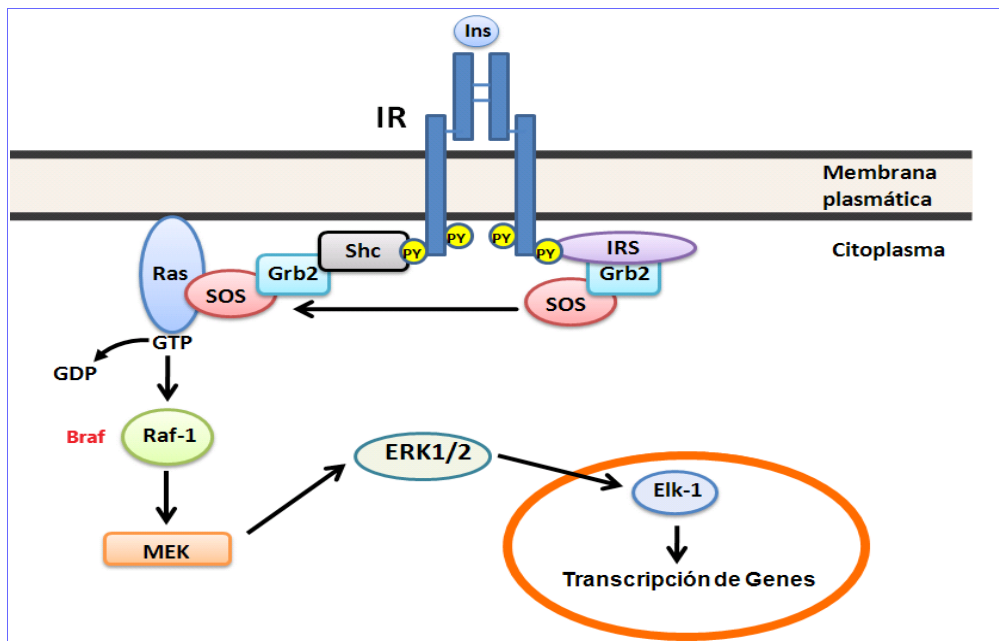


Figura 39. Activación de la vía de las MAPK por la insulina [31]

### 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. 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 alostéricos 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)-P<sub>2</sub>, que son fosforilados generando PIP<sub>2</sub> y PIP<sub>3</sub>. PIP<sub>3</sub> sirve como sitio de unión a serín quinasas 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+/-) [9].



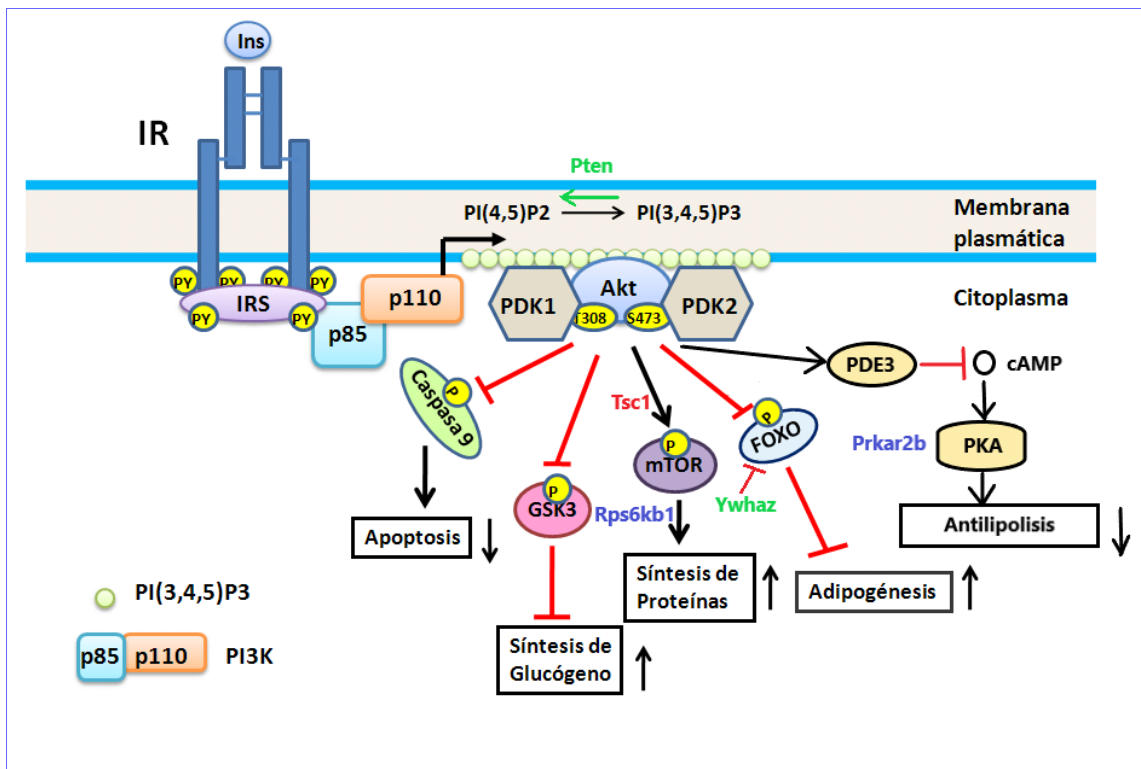


Figura 40. Activación de la vía de las PI3K por la insulina [31]

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 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 targets de miR200b y miR429 (down-reg) y Map2ka4 de miR92a (up-reg). Y por último, el gen Mef2c, 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+/- [9]

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	ACCCGTCCC GTTCGTCCCGG	mmu-miR-1247-5p;iso_3p:a
## 162	ACGCCCTTCCCCCTTCTCA	mmu-miR-1249-3p;ref
## 536	CATCTAACCTGGCCTTTGA	mmu-miR-1981-3p;iso_3p:c
## 792	GCTGCAAACATCCGACTGAAAG	mmu-miR-30f;iso_snp:5CT;iso_5p:GCT;iso_3p:ctc
## 1580	TTTTTCATTATTGCTCCTGACC	mmu-miR-335-3p;ref
## 41	AACATGGAGGAAATTTACGT	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	CAAATCTTATTTGAGCACCTG	mmu-miR-1264-3p;iso_3p:t
## 4341	CAAATCTTATTTGAGCACCTGT	mmu-miR-1264-3p;ref
## 710	AAGCCCTTACCCAAAAAGCAT	mmu-miR-129-2-3p;ref
## 14011	TTCATTGCGCTGTCCAGATGTAA	mmu-miR-1298-5p;iso_add:A
## 14010	TTCATTGCGCTGTCCAGATGTA	mmu-miR-1298-5p;ref
## 1590	ACCATCGACCGTTGAGTGGACA	mmu-miR-181c-3p;iso_add:A;iso_3p:c
## 457	AACATTCAATTGTTGTCGGTGGGA	mmu-miR-181d-5p;iso_add:A;iso_3p:t
## 463	AACATTCAATTGTTGTCGGTGGGT	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	TAACCTTCCCTTGTGCCCTCAGT	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:

```
agUAGUAAUGGUCGCAUAAu
```

El isomiR analizado es una variante -3p a la que se añade la base C en el extremo 3'. En este extremo, las bases AG no se han alineado con la secuencia *target*, por lo que la base añadida puede ser irrelevante. 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	
agUAGUAAUGGUCGCAUAAu		:   :		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 buscan el resto de las secuencias de los isomiRs DE seleccionados en la base de datos miRanda.

Además del miR200b, otro miRNA que se encuentra *down-reg* en los datos del experimento 2 es miR429, perteneciente a la misma familia. La isoforma DE de este miRNA es:

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

Correspondiente a la variante -3p de referencia. Buscando la secuencia complementaria 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 [5], 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 Bdnf 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:

```
AACATGGAGGAAATTTACGT          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.

## REFERENCIAS

- [1] Bartel, David P. "MicroRNAs: genomics, biogenesis, mechanism, and function." *cell* 116.2 (2004): 281-297. <https://www.sciencedirect.com/science/article/pii/S0092867404000455>
- [2] 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/>
- [3] 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/>
- [4] 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>
- [5] 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)
- [6] 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>
- [7] 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/>
- [8] 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/>
- [9] 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>
- [10] SRA handbook. <https://www.ncbi.nlm.nih.gov/books/NBK242621/>
- [11] 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/>
- [12] Pantano, L. miRNA annotation. [https://seqcluster.readthedocs.io/mirna\\_annotation.html](https://seqcluster.readthedocs.io/mirna_annotation.html)
- [13] González, I. "Statistical analysis of RNA-Seq data". Universidad de Toulouse (2014)
- [14] 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>
- [15] Pantano, Lorena. "Characterization of miRNA and isomiR molecules." <https://www.bioconductor.org/packages/devel/bioc/vignettes/isomiRs/inst/doc/isomiRs.html>
- [16] Tarazona, S., Furió-Tari, 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>

- [17] Polyester. Documento de Bioconductor:  
<https://www.bioconductor.org/packages/release/bioc/vignettes/polyester/inst/doc/polyester.html>
- [18] Hong, F.(2010). Bioconductor RankProd Package Vignette.<http://bioconductor.statistik.tu-dortmund.de/packages/3.6/bioc/vignettes/RankProd/inst/doc/RankProd.pdf>
- [19] Zhang, Y., Zang, Q., Xu, B., Zheng, W., Ban, R., Zhang, H., ... & Shi, Q. (2016). IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics*, 32(13), 2069-2071.  
<https://www.ncbi.nlm.nih.gov/pubmed/27153728>
- [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\\_targets\\_in\\_protein\\_coding\\_sequences](https://www.researchgate.net/publication/221785142_Functional_microRNA_targets_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] 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>
- [29] 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).
- [30] 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>

[31] 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>

[32] 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>