



UNIVERSIDAD
MAYOR

para espíritus emprendedores

Facultad de Ciencias

**ESCUELA DE
BIOTECNOLOGÍA**

**“Identificación de transcritos quiméricos en modelo de Síndrome de Rett:
posibles dianas de intervención terapéutica”**

Guillermo Ignacio Albornoz Guiñez

Proyecto de Tesis para optar al título de Biotecnólogo

Santiago - Chile

2022



UNIVERSIDAD
MAYOR

para espíritus emprendedores

Facultad de Ciencias

**ESCUELA DE
BIOTECNOLOGÍA**

**“Identificación de transcritos quiméricos en modelo de Síndrome de Rett:
posibles dianas de intervención terapéutica”**

Guillermo Ignacio Albornoz Guiñez

Proyecto de Tesis para optar al título de Biotecnólogo

Tutor: Dr. Mauricio Sáez Venegas

Santiago - Chile

2022

*Dedico este trabajo a mi perrita tete,
quien me acompañó durante 17 años,
pero lamentablemente este último año
que paso la tuvimos que dejar ir.*

Siempre te amaré

SOLO USO ACADÉMICO

Agradecimientos

Primero, me gustaría agradecer a todos los docentes que me enseñaron y ayudaron durante todo mi proceso universitario, pero principalmente agradezco al Dr. Mauricio Sáez por ayudarme, enseñarme, guiarme y acompañarme durante este último año.

Agradecimientos a la Dra. Daniela Rivera Rocabado del centro de genómica, ecología y medio ambiente (GEMA) de la universidad Mayor, por su orientación en el análisis estadístico.

Agradezco a mi familia por acompañarme durante todos estos años de estudio, por cada consejo y apoyo emocional que me brindaron y estar ahí cuando más los necesitaba.

No puedo dejar pasar en esta oportunidad, agradecer a mis compañeros de generación, que durante estos años de carrera mutuamente nos estuvimos apoyando, en especial a Sebastián Oyarce, Vicente Colarte y Juan González los cuales además de ser mis amigos, fueron un gran apoyo donde estuvieron presentes para escuchar, entender y guiarme en los momentos más complicados en mi proceso de tesis.

Finalmente agradecer a mi pareja Ailene Contreras por apoyarme y soportarme durante todo este proceso.

SOLO USO ACADÉMICO

Información del laboratorio y centro de investigación

Laboratorio de epigenética, cromatina y neurociencia; Escuela de biotecnología; Centro de Genómica y Bioinformática; Universidad Mayor; Dr. Mauricio Sáez Venegas; bajo financiamiento Fondecyt 11171015

SOLO USO ACADÉMICO

Abreviaturas

- **RTT:** Síndrome de Rett
- **SNC:** Sistema nervioso central
- **DDN:** Discapacidad del neurodesarrollo
- **DEA:** Desorden del espectro autista
- **Mecp2:** Methyl-CpG binding protein 2
- **CDKL5:** Cyclin-dependent kinase-like 5
- **FOXP1:** Forkhead box protein G1
- **MBD:** Methyl-CpG-binding domain
- **HDAC:** Histone deacetylase
- **CREB1:** Cyclic AMP-responsive element-binding protein 1
- **L1:** Long interspersed nuclear elements 1
- **ORF:** Open reading frame
- **KO:** Knockout
- **RDA:** Respuesta a daño del ADN
- **H2AX:** Histona 2AX
- **ENA:** European Nucleotide Archive
- **CTAT:** Trinity Cancer Transcriptome Analysis Toolkit
- **STAR-Fusion:**
- **FFPM:** Lecturas totales por millón de fragmentos de fusión
- **PQ:** Puntos de quiebres
- **TQ:** Transcritos quiméricos
- **CircRNA:** ARN circulares
- **ARNnc:** ARN no codificante
- **SNHG11:** Small nucleolar RNA host gene 11
- **lncRNAs:** Long non-coding RNAs

I. Índice de contenido

• Resumen.....	11
• Summary.....	12
• Capítulo 1: Introducción y Marco teórico	13
1.1 El síndrome de Rett.....	13
1.2 Genes involucrados en RTT.....	16
1.3 Las funciones de <i>MECP2</i>	17
1.4 <i>L1</i> y <i>MECP2</i>	19
1.5 Influencia que tiene <i>MECP2</i> en la respuesta de daño al ADN.....	21
• Capítulo 2: Hipótesis y objetivos.....	24
• Capítulo 3: Materiales y métodos.....	25
3.1 Servidor.....	25
3.2 Descarga de secuenciaciones de RNA-Seq.....	25
3.3 Análisis de calidad y recorte de secuencias.....	26
3.4 Alineamiento de secuencias y análisis utilizando la herramienta computacional ARRIBA.....	26
3.5 Alineamiento de secuencias y análisis utilizando la herramienta computacional STAR-Fusion.....	27
3.6 Filtrado de datos.....	28
3.7 Generación de gráficos y Diagrama de cuerdas.....	29
3.8 Análisis estadísticos.....	31
3.9 Identificación y anotación de dominios proteicos.....	32
• Capítulo 4: Resultados.....	33
4.1 Comparación y cuantificación de transcritos quiméricos reportados en muestras de Síndrome de Rett y muestras silvestre.....	33
4.2 Cuantificación de puntos de quiebres obtenidos de los reportes de los programas utilizados.....	35
4.3 Distribución de los puntos de quiebres extra e intracromosómicos sobre el genoma del ratón.....	39
4.4 <i>Hotspot</i> , genes asociados y dominios proteicos.....	41
4.5 Similitudes entre los resultados obtenidos de los análisis generados por STAR-Fusion y ARRIBA.....	45

4.6	Clasificación de puntos de quiebres de los transcritos quiméricos.....	46
4.7	Cuantificación de puntos de quiebre según tejido.....	52
•	Capítulo 5: Discusión.....	57
5.1	Diferencia entre los programas utilizados.....	57
5.2	ARN circulares.....	59
5.3	Dependencia de <i>Mecp2</i>	60
5.4	<i>SNHG11</i>	61
5.5	Esporádicos, medios y recurrentes.....	62
•	Capítulo 6: Conclusiones.....	64
•	Capítulo 7: Implicancias y proyecciones.....	65
7.1	Análisis en mayor cantidad de tejidos.....	65
7.2	Análisis de linfocitos en enfermedades.....	66
7.3	Detección de patrones en síndrome de Rett.....	66
•	Capítulo 8: Referencias.....	67
•	Capítulo 9: Anexo.....	80

SOLO USO ACADÉMICO

II. Índice de figuras

• Capítulo 1: Introducción y Marco teórico	13
- Figura 1. Funciones Conocidas de <i>MECP2</i>	18
- Figura 2. Expresión de L1 en muestras <i>MECP2-KO</i>	20
- Figura 3. Contenido de LINE 1 en modelos <i>Mus musculus</i>	21
- Figura 4. Sistema de redes que muestra la relación entre proteínas con <i>MECP2</i>	22
• Capítulo 3: Materiales y métodos.....	25
- Figura 5. Diagrama de flujo para el análisis de calidad, “ <i>Trimming</i> ” y alineamiento.....	26
- Figura 6. Alineamiento de secuencias a referencia, obtención y análisis de transcritos quiméricos.	27
• Capítulo 4: Resultados.....	33
- Figura 7. Cuantificación de candidatos identificados por STAR-Fusion y ARRIBA	34
- Figura 8. Cuantificación e identificación de puntos de quiebre existente en los candidatos de transcritos quiméricos.....	37
- Figura 9. Diagrama de cuerdas conectado a los puntos de quiebre de transcritos quiméricos detectados.....	40
- Figura 10. Diagrama de Venn de los puntos de quiebres obtenidos de los resultados de STAR-Fusion y ARRIBA.....	45
- Figura 11. Caracterización de puntos de quiebres en transcritos quiméricos identificados en reportes de STAR-Fusion.....	48
- Figura 12. Caracterización de puntos de quiebres en transcritos quiméricos identificados en reportes de ARRIBA.....	50
- Figura 13. Cuantificación e identificación de puntos de quiebres existentes en los diferentes tejidos analizados.....	54

III. Índice de tablas

• Capítulo 1: Introducción y Marco teórico	13
- Tabla 1. Genes causantes del síndrome de Rett.....	17
• Capítulo 3: Materiales y métodos.....	25
- Tabla 2. Criterios de búsqueda en muestras silvestres.....	25
- Tabla 3. Criterios de búsqueda en muestras mutantes.....	25
• Capítulo 4: Resultados.....	33
- Tabla 4. Cuantificación de los transcritos quiméricos identificados en los programas STAR-Fusion y ARRIBA	41
- Tabla 5. Eventos únicos de transcritos quiméricos encontrados.....	41
- Tabla 6. “Hotspot” encontrados en STAR-Fusion	42
- Tabla 7. “Hotspot” encontrados en ARRIBA.....	43
- Tabla 8. Dominios proteicos identificados de los resultados de STAR-Fusion.....	43
- Tabla 9. Dominios proteicos identificados de los resultados de ARRIBA.....	44

SOLO USO ACADÉMICO

IV. Resumen

El síndrome de Rett (RTT) es un trastorno neuronal, que se caracteriza por la pérdida del habla y de los movimientos intencionados de las manos. La causa principal de este síndrome es la mutación del gen de la proteína de unión a metil-CpG 2 (*MECP2*), cuya función principal está relacionada con la epigenética y la arquitectura de la cromatina, además de tener actividad como represor y activador de genes. Un elemento genético que tiene relación con *MECP2* es Long interspersed nuclear elements-1 (*LINE-1* o *L1*), que es un retrotransposón autónomo, y es conocido por su alta expresión y movilidad en el tejido cerebral y su movimiento de "cortar y pegar", que puede generar translocaciones, deleciones y duplicaciones en el genoma. Esto puede conducir a un aumento de la fusión de genes o de transcripciones quiméricas en los casos existentes de este síndrome.

En este trabajo se utilizaron bases de datos de libre acceso, principalmente el *European Nucleotide Archive*, para obtener un total de 100 muestras de *RNA-seq* de tejido cerebral, de las cuales 50 eran muestras *Mus musculus MECP2-KO* y otras 50 muestras de silvestres. Éstas fueron analizadas mediante 2 herramientas computacionales, las cuales fueron STAR-Fusion y ARRIBA, para la detección de transcritos quiméricos, que utilizaron los informes obtenidos de las herramientas para luego filtrar los datos y ver si hay diferencias en el número de transcritos de las muestras *MECP2-KO* y las muestras silvestres.

Los resultados obtenidos a través del análisis muestran que existe un aumento de transcritos quiméricos en las muestras *MECP2-KO* frente a las muestras silvestres, lo que podría significar que debido al aumento de la expresión de *L1* por la deficiencia de *MECP2*, este retrotransposón provoca un aumento de saltos generando un aumento de transcritos quiméricos en las muestras con síndrome RETT.

V. Summary

Rett syndrome (RTT) is a neuronal disorder, which is characterized by loss of speech and purposeful hand movements, the main cause of this syndrome a mutation in the gene methyl-CpG binding protein 2 (*MECP2*), which main function is related to epigenetics and chromatin architecture, in addition to having activity as both repressor and activator of genes. One of the main genes that have a *MECP2* relationship is Long interspersed nuclear elements-1 (LINE-1 or L1), which is an autonomous retrotransposon, and is known for its high expression and mobility in brain tissue and its "cut and paste" movement, which can generate translocations, deletions, and duplications in the genome. This can lead to an increase in gene fusion or chimeric transcripts within existing cases of this syndrome.

In this work, open access databases, mainly European Nucleotide Archive, were used to obtain a total of 100 RNA-seq samples from brain tissue, of which 50 were *Mus musculus MECP2-KO* models and others 50 wild type samples. These were analyzed using 2 computational tools for the detection of chimeric transcript candidates, which were STAR-Fusion and ARRIBA, which used the reports obtained from the tools to then filter the data to see if there were differences in the number of transcripts from the *MECP2-KO* samples and wild type samples.

The results obtained through the analysis show that there is an increase of chimeric transcripts in *MECP2-KO* samples as opposed to wild type samples, which could mean that due to the increase of L1 expression due to *MECP2* deficiency, this retrotransposon causes an increase of jumps generating an increase of chimeric transcripts in RETT syndrome models.

1. INTRODUCCIÓN

Los trastornos del neurodesarrollo o discapacidad del neurodesarrollo (DDN), son alteraciones graves y crónicas que se producen antes de que el individuo cumpla los 22 años, luego de esta edad lo más probable es que la condición continúe indefinidamente y provoque limitaciones funcionales trascendentales en 3 o más de las siguientes 8 áreas de actividades de la vida: cuidado personal, lenguaje receptivo y expresivo, aprendizaje, movilidad, autodirección, capacidad de vida independiente y autosuficiencia económica (1). De esta manera, las DDNs son desórdenes crónicos que afectan las funciones del sistema nervioso central (SNC) durante el desarrollo de las habilidades motoras. Algunas de estas funciones son el razonamiento, la comunicación y/o el comportamiento (2); estas irregularidades dentro del SNC se pueden generar, por interrupciones o alteraciones de procesos como la regulación de la estructura del ADN y la función neuronal. Los mecanismos epigenéticos son fundamentales en el desarrollo del SNC, por lo que estos procesos pueden causar una serie de trastornos del neurodesarrollo (3).

Dentro de los posibles trastornos que se pueden producir debido a lo mencionado anteriormente, podemos mencionar el desorden del espectro autista (DEA), déficit de atención/desorden de hiperactividad, el síndrome de Down, el síndrome de Rett (RTT), etc (4–6). Todas estas condiciones genéticas son ampliamente estudiadas en el mundo, pero aun así el conocimiento que tenemos de esta enfermedad progresa lentamente. Por este motivo, cada información que se pueda conseguir para poder entenderlas de mejor manera es fundamental para el estudio de cada una de ellas. En el caso de este trabajo de tesis, se hablará de RTT y su implicancia a nivel molecular.

1.1. El síndrome de Rett

El RTT es un desorden neuronal, que se caracteriza por la pérdida del habla y los movimientos propositivos de manos. Este síndrome fue descrito originalmente por Andreas Rett en 1966 (7). Los diagnósticos clínicos que reconocen este síndrome han ido variando ligeramente con el paso del tiempo, lo cual refleja el aumento de la comprensión de este síndrome (8). El RTT es la segunda mayor causa de discapacidad intelectual en las mujeres después del síndrome de Down (9). Este trastorno presenta una incidencia de 1,09/10.000 mujeres (10), aunque los pocos hombres que sufren de este síndrome usualmente presentan encefalopatías congénitas graves y finalmente muerte dentro de los primeros 2 años de vida (11).

Existen 2 clasificaciones de RTT, el RTT clásico y el RTT atípico, el cual puede ser más o menos severo que un caso clásico y no todos los casos atípicos presentan los mismos síntomas (8). Debido a esto se estableció un criterio clínico para describir el diagnóstico de RTT (8). En base a 3 tipos de criterios:

A. Criterios principales

- Pérdida parcial o completa de habilidades propositivas de las manos.
- Pérdida parcial o completa de la habilidad para hablar adquirida.
- Anomalías de la marcha: Deterioro (dispraxia) o ausencia de habilidad.
- Movimientos estereotipados de las manos, como retorcerse o apretarlas, aplaudir o dar golpecitos, hablar en voz alta y automatismos de lavado/limpieza.

B. Criterios de exclusión del RTT clásico

- Lesión cerebral secundaria a un traumatismo (peri o postnatal), a una enfermedad neurometabólica o a una infección grave que cause problemas neurológicos.
- Desarrollo psicomotor groseramente anormal en los primeros 6 meses de vida.

C. Criterios de apoyo para el RTT atípico

- Trastorno de la respiración al estar despierto.
- Bruxismo al estar despierto.
- Deterioro del patrón del sueño.
- Tono muscular anormal.
- Alteraciones vasomotoras periféricas.
- Escoliosis.
- Retraso en el crecimiento.
- Manos y pies pequeños y fríos.
- Risas/gritos inapropiados e inoportunos.
- Disminución de la respuesta al dolor.
- Comunicación ocular intensa – “eye pointing”.

En el RTT existe un periodo de neurodesarrollo normal el cual caracteriza el fenotipo que presentan los pacientes con RTT durante los primeros 6-18 meses de edad (12), durante los 6-18 meses se presenta el periodo de regresión seguido por una recuperación o estabilización, el cual debiese expresar todos

los criterios principales y todos los criterios de exclusión del RTT clásico, aunque a menudo están presentes los criterios de apoyo. En cambio, para diagnosticar un caso de RTT atípico este debe tener un periodo de regresión seguido por una recuperación o estabilización con un mínimo de 2 de los criterios principales y un mínimo de 5 de los criterios de apoyo (8,13). Los criterios recién descritos se van presentando en distintas edades y etapas, de la persona que padece de este síndrome (14).

El síndrome consta de 4 etapas:

- **Etapa 1 o de estancamiento:** Generalmente esta inicia entre los 4 y los 18 meses, este se caracteriza por un estancamiento del desarrollo psicomotor y de la comunicación, además un retraso en el control de la postura y el equilibrio; debido a esto, esta etapa se considera de estancamiento.
- **Etapa 2 o regresión rápida del desarrollo:** esta etapa empieza entre el primer y cuarto año de vida. En esta etapa se pierden habilidades adquiridas como la función de las manos, habilidades motoras finas, balbuceo, algunas niñas presentan síntomas similares a las del autismo. Movimiento estereotipado de las manos, lavarse, aplaudir y llevarse las manos repetidamente a la boca, también se puede producir irregularidades en la respiración, esta etapa se conoce como de regresión.
- **Etapa 3 o periodo pseudoestacionario:** Esta ocurre durante los 2 y 10 años, y se caracteriza por apraxia, problemas motores y convulsiones, además de disfunción variable de la motricidad gruesa, retraso mental y cierta mejoría en la interacción social, en algunos casos las niñas con este síndrome permanecen en esta etapa para el resto de sus vidas, a esta se le considera etapa pseudo estacionaria.
- **Etapa 4 o periodo de deterioro tardío:** finalmente en esta se incluye la reducción de la movilidad y una severa discapacidad física. Algunas niñas que mantuvieron el caminar ahora pueden perder esta capacidad, y los movimientos repetitivos de las manos pueden disminuir y la mirada suele mejorar, a esta etapa se le llama deterioro motor (14–17).

Antes de la publicación de los criterios para diagnosticar este síndrome, los primeros informes sobre RTT se centraron en el rasgo autista que se presentaba en la etapa de regresión (14). Las observaciones que se hicieron posteriormente a estas publicaciones, concluyeron que además del deterioro de la

comunicación y estereotipias severas, los comportamientos de retraimiento social contribuyen a que los individuos cumplieron los criterios del “Manual diagnóstico y estadístico de los trastornos mentales” (DSM-V) para los trastornos generalizados del desarrollo no especificado (PDD-NOS;(18–20)), por lo que durante mucho tiempo este síndrome se consideró como un trastorno del espectro autista (TEA). Aunque desde el 2013, en la quinta edición del DSM-V el RTT ya no se considera como TEA, ya que este se diagnostica cuando se cumplen todos los criterios de diagnóstico para TEA, debido a esto RTT se considera un trastorno genético neurológico y del desarrollo (21).

En 1999 se descubrió que el gen principal que causaba este síndrome era *Methyl-CpG binding protein 2 (MECP2)* (22). Entre el 90-95% de los pacientes que son diagnosticados con RTT clásico, presentan mutaciones en el gen *MECP2*, el cual está ligado al cromosoma X, y provocan la pérdida de la función de este gen (23). Las mutaciones de este gen surgen predominantemente en las células “germinales” del lado paterno (24), y la mayoría de los pacientes de RTT, a los que se les hace un examen genético, es causado por una mutación *de novo* (24). Dentro de los cientos de mutaciones conocidas de *MECP2*, se han identificado 8 mutaciones que constituyen alrededor del 60% de los casos documentados de RTT (25).

1.2. Genes involucrados en RTT

A pesar de las variables fenotípicas, el 95% de los casos de RTT clásicos y aproximadamente el 75% de los casos de RTT atípico tienen una mutación en *MECP2* (26), además de este gen existen otros 2 genes que también pueden causar RTT estos son *Cyclin-dependent kinase-like 5 (CDKL5)* y *Forkhead box protein G1 (FOXP1)* (27,28). Generalmente los pacientes que tienen mutaciones en algunos de estos 2 genes mencionados sufren de un RTT atípico, además de estos genes, se conocen más de 60 genes, que sus mutaciones pueden provocar fenotipos similares al RTT. Estos genes se pueden observar en la **Tabla 1** (29). La función de los genes que fueron identificados en los últimos años, en su mayoría están involucrados en los siguientes procesos biológicos:

- Desarrollo y función neuronal, incluyendo la función de la sinapsis, los canales iónicos y neurotransmisores (30), por ejemplo, *SYNGAP1*, *SCN1A*, *SCN2A*, *SCN8A*, *KCNA2*, *KCNJ10*, *KCNQ2*, *GABBR2*, *GABRD*, *GRIN2A*, *GRIN2B*, *ANKRD31*, *CHRNA5*, *JMJD1C*.
- Modificación de la epigenética, organización de la cromatina (31), por ejemplo, *HDAC1*, *NCOR*, *JMJD1C*.

- Factores de translación/transcripción(29): *ZFX, ZNF238, ZNF620, ZSCAN, TAF1B, EIF1A2, EIF2B2, EIF4G1*.

Tabla 1. Mutación en genes causantes del síndrome de Rett. Mutaciones en genes causantes de síndrome de Rett clásico o atípico, o fenotipos similares descubiertos en los últimos años 4. Extraído desde Ehrhart, Sangani & Curfs, 2018.

Publication	Discovered gene
Gilissen <i>et al.</i> [23]	<i>SMC1A</i>
Saitou <i>et al.</i> [24]	<i>TBL1XR1</i>
Okamoto <i>et al.</i> [25]	<i>GABRD</i>
Baasch <i>et al.</i> [26]	<i>SCN2A</i>
Hara <i>et al.</i> [27***]	<i>SHANK3</i>
Olson <i>et al.</i> [28]	<i>SCN8A, IQSEC2</i>
Hoffjan <i>et al.</i> [29***]	<i>WDR45</i>
Lucariello <i>et al.</i> [30***]	<i>ANKRD31, CHRNA5, HCN1, SCN1A, TCF4, GRIN2B, SLC6A1, MGRN1, BTBD9, SEMA6B, AGAP6, MGRN1, VASH2, ZNF620, GRAMD1A, GABBR2, ATP8B1, HAP1, PDLIM7, SRRM3, CACNA11</i>
Lopes <i>et al.</i> [31**]	<i>TCF4, EIF1A2, STXBP1, ZNF238, SLC35A2, ZFX, SHROOM4, EIF2B2, RHOBTB2, SMARCA1, GABBR2, EIF4G1, HTT</i>
Saez <i>et al.</i> [32]	<i>JMID1C</i>
Lee <i>et al.</i> [33]	<i>SATB2</i>
Yoo <i>et al.</i> [34]	<i>GABBR2</i>
Vidal <i>et al.</i> [35**]	<i>GRIN2B, GABBR2, MEF2C, STXBP1, KCNQ2, SLC2A1, TCF4, SCN2A, SYNGAP1, CACNA11, CHRNA5, HCN1</i>
Sajan <i>et al.</i> [36**]	<i>PWP2, SCG2, IZUMO4, XAB2, ZSCAN12, IQSEC2, FAM151A, SYNE2, SMC1A, ARHGEF10L, HDAC1, TAF1B, KCNJ10, CHD4, LRRC40, LAMB2, GRIN2B, IMPDH2, SAFB2, ACTL6B, STXBP1, TRRAP, WDR45, SLC39A13, FAT3, IQGAP3, NCOR2, GABBR2, TCF4, GRIN2A</i>
Allou <i>et al.</i> [37]	<i>IQSEC2, KCNA2</i>
Huisman <i>et al.</i> [38***]	<i>SMC1A</i>

1.3 Las funciones de *MECP2*

MECP2 se identificó y caracterizó en 1992. Esta posee un *Methyl-CpG-binding domain (MBD)*, debido a este dominio la proteína se puede unir específicamente al ADN metilado (32). La primera función que se describió de *MECP2* fue la represión transcripcional, donde se identificó que esta proteína interactúa con *mSin3A*, la cual media la asociación con *histone deacetylase (HDAC)* (33). Este complejo proteico cumple la función de eliminar los grupos acetilos de los residuos de lisina, para así remodelar la cromatina (34), con una función de represión transcripcional (35,36). Esta proteína además tiene la función de activador transcripcional, debido a que este interactúa con el activador de transcripción *CREB1 (cyclic AMP-responsive element-binding protein 1)* en regiones promotoras que participan en la activación de genes. A este complejo de proteínas también se le une un coactivador y el ARN polimerasa II (37). Luego *MECP2* se describió como un modulador multifuncional en la expresión de los genes y además como un regulador transcripcional para la activación o represión molecular (37). Se ha observado que esta proteína está relacionada tanto en la compactación de la cromatina (38) y participación en la generación de loops de la cromatina (39). En la **Figura 1** se pueden observar las funciones recién descritas.

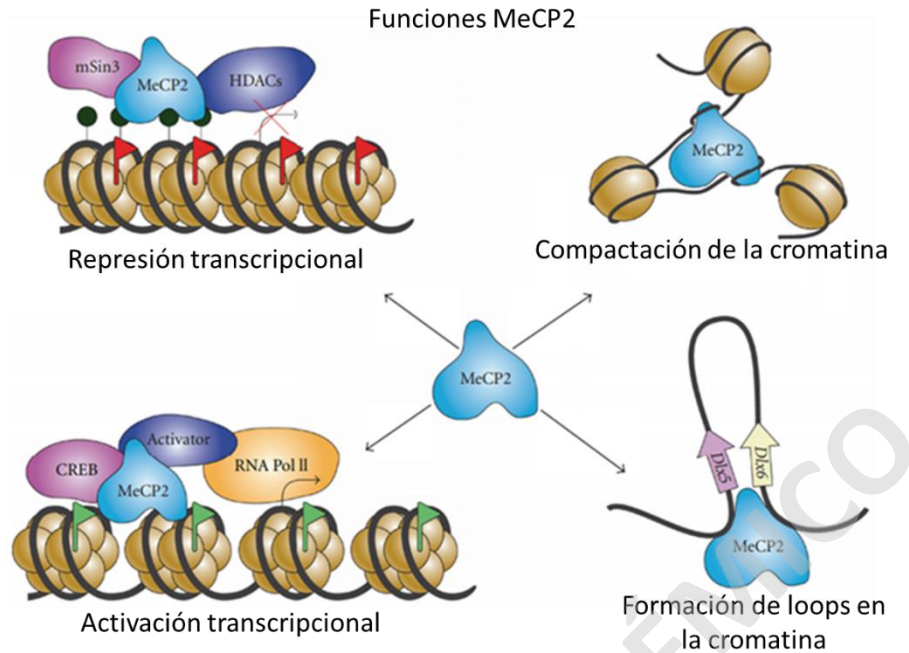


Figura 1. Funciones conocidas de *MeCP2*. Esquema de las diversas funciones de la proteína *MECP2*. Figura adaptada desde Zachariah, R. M., & Rastegar, M; *Neural plasticity*. 2012.

Conociendo las funciones principales de *MECP2*, y además de la relación que tiene con el RTT, es importante señalar que este gen se expresa de forma ubicua en humano, pero los niveles de la proteína de *MECP2* son especialmente altos en el cerebro en donde la expresión comienza en etapas tempranas del desarrollo y los procesos plasticidad sináptica (40).

MECP2 es uno de los moduladores claves de la epigenética en el cerebro (41), el cual es un término que se refiere al control de la expresión génica y los cambios heredables que controla el fenotipo celular y la finalidad de los genes sin alterar la secuencia de ADN (42). En el caso de las neuronas, células postmitóticas no podemos hablar de cambios heredables, pero sí de control epigenético por distintos mecanismos como remodelación en la cromatina, modificaciones post traduccionales de histonas y metilación del ADN, los cuales provocan cambios en la regulación transcripcional y en la expresión de genes de las neuronas (43).

Hasta la fecha se conocen un total de 4.668 variantes de *MECP2*, donde 3.341 mutaciones son patogénicas, y más de 2000 de estas están relacionadas con alguna clase de RTT (44). Las principales complicaciones que hay para estudiar este gen es el modelo de estudio, ya que en modelos de animales

invertebrados o plantas no existe un gen homólogo a *MECP2* (45). Por lo mismo, el modelo que se utiliza es *Mus musculus*, debido a que esta proteína la proteína MECP2 altamente conservada en vertebrados (46).

1.4 L1 y MECP2

Las dianas directas de *MECP2* han sido difíciles de localizar, y los intentos de utilizar Chip de ADN (*microarrays*) para poder identificar estas dianas en el hipocampo y en el córtex cerebral utilizando ratones *MECP2-KO* solo han revelado cambios sutiles en la transcripción de genes (47). Aun así, se han encontrado elementos genéticos que son regulados por *MECP2*, uno de estos es *Long interspersed nuclear elements-1 (LINE-1 o L1)*, donde se demostró que una de las funciones que tiene *MECP2* es regular la actividad de L1 en el sistema nervioso (48).

L1 es un no-LTR (long terminal repeat) retrotransposón autónomo, que contiene un promotor interno de RNA polimerasa II, dos marcos de lectura abiertos (ORF), los cuales son ORF1 y ORF2, y una cola polyA (49). Este retrotransposón es abundante en los mamíferos ya que alcanza casi el 20% de todo el genoma (49,50). L1 es un elemento móvil autónomo, lo que quiere decir que no necesita de otra secuencia o elemento para poder retro transportarse (51). Es importante señalar que uno de los principales tejidos en el que este elemento móvil se retro transpone es el tejido cerebral. En 2009 se hizo un experimento para ver la cantidad de copias de L1 en tejido de cerebro en comparación con tejido de hígado y corazón, en ese experimento se pudo observar que en tejido del hipocampo tenían más de 1.000 copias adicionales de L1, lo que sugiere un aumento teórico de ORF2 de aproximadamente 80 copias por célula (52).

La actividad de L1 durante el desarrollo del cerebro puede tener un impacto en la expresión de genes y la función neuronal, debido a que este aumentaría el mosaicismo genético en el cerebro (53), por el aumento de saltos e inserciones de L1 dentro del genoma, estas pueden tener un gran impacto en la estructura de los genes y en sus funciones a través de mutagénesis insercional en los exones (54,55). Además, *L1* puede provocar problemas en posibles elementos regulatorios, como la interrupción del avance de la RNA polimerasa II, la poliadenilación prematura, acumulación de promotores alternativos, entre otras consecuencias funcionales (56). Lo mencionado anteriormente genera un aumento en los saltos de L1, lo que aumenta el mosaicismo genético, ya que se sabe que al momento de utilizar el

mecanismo de transposición provoca el quiebre de la doble hebra de ADN, lo que puede llegar a generar translocaciones, supresiones y duplicaciones (57).

Un regulador de *L1* es *SOX2*, el cual cumple una función de factor de transcripción, que inhibe la retrotransposición de *L1* durante la diferenciación neuronal, al estar *SOX2* sobre expresado, existen bajos niveles de *L1*. Por el contrario, si disminuye la expresión de *SOX2* aumenta la expresión de *L1* (58). Junto a este gen también actúa *MECP2*, el cual genera una represión en la transcripción de *L1* (59). *MECP2* une a la sección 5'UTR de *L1* lo que provoca la represión dependiente de la metilación (60). En 2010 se descubrió que *MECP2* modula *L1* tanto sobre la actividad del promotor como en la retrotransposición (49). Diversos experimentos han mostrado que *MECP2-KO* en el cerebelo y en el cuerpo estriado de ratones mutantes causa un aumento en la expresión de *L1*, en comparación con ratones silvestres (**Figura 2**).

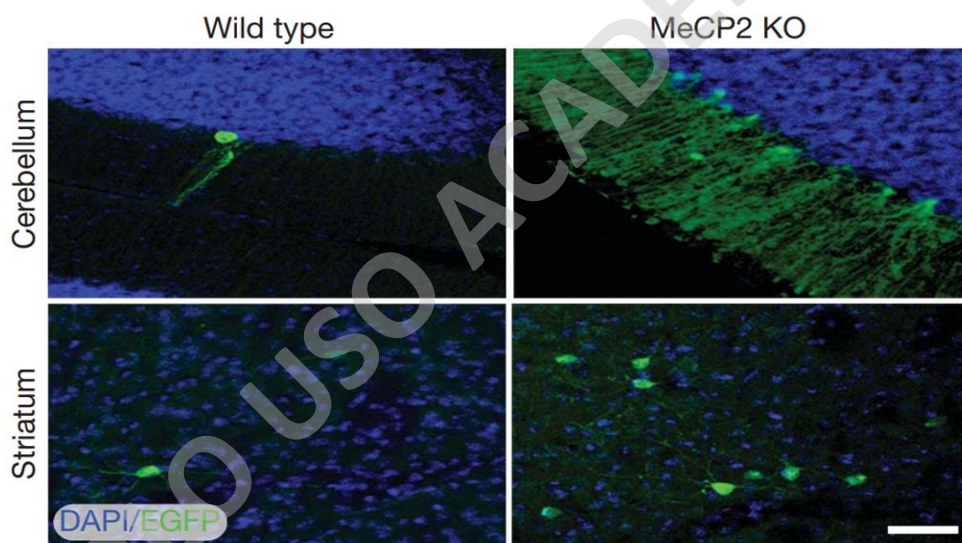


Figura 2. Expresión de L1 en muestras *MeCP2-KO*. EGFP indica la expresión de L1 (verde), y DAPI indica el núcleo celular (azul). En los paneles de la derecha se puede observar un aumento de *EGFP*, dirigido por promotor *L1* en ambos tejidos neuronales, lo que indica un aumento de la expresión de *L1* en muestras *MECP2-KO*. A diferencia con los paneles de la izquierda, donde se observa una menor cantidad de *EGFP* en el núcleo, debido a que estos son tejidos de muestras silvestres. Muotri, Alysson R., et al. Nature Letter, 2010.

Análisis de expresión *post mortem* de pacientes RTT muestran un aumento en la expresión de *L1* en neuronas corticales, (**Figura 3**) (61). Sin embargo, se observó que la retrotransposición de *L1* se genera mayoritariamente en las células progenitoras neuronales (62). Esto indicaría que en la ausencia de *MECP2*, *L1* aumenta su expresión generando un aumento de “saltos”, los que provocan el quiebre de la doble cadena de ADN, que podría contribuir al desarrollo de este síndrome. No obstante, el número

y las ubicaciones de estas recombinaciones que se generan debido a la sobreexpresión de L1, aún permanecen incógnitas, por lo que su estudio podría mejorar el entendimiento en el RTT.

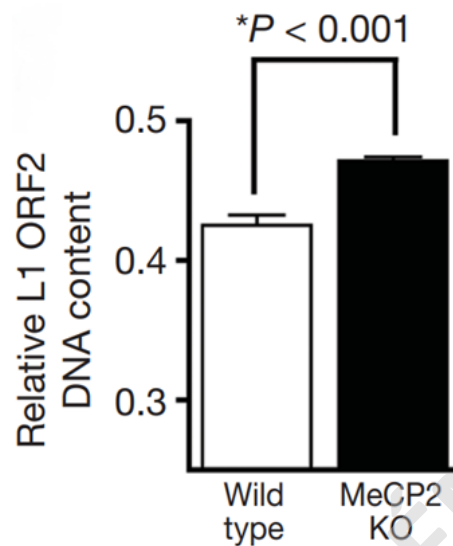


Figura 3. Contenido de LINE 1 en muestras *Mus musculus*. Contenido de L1 en células neuroepiteliales de ratones muestras silvestre y de *MECP2 KO*., imagen obtenida de Muotri, Alysson R., et al. Nature Letter, 2010

1.5 Influencia que tiene *MECP2* en la respuesta de daño al ADN

En los últimos años se ha visto la relación que tiene *MECP2* con la respuesta a daño del ADN (RDA). La histona 2AX (*H2AX*), es un biomarcador natural para medir la respuesta de daño en células eucariotas (63).

Se utilizó *H2AX* para comparar la respuesta a daño en células con una mutación en *MECP2* y células control. En el experimento se reporta que la expresión reducida de *MECP2* está relacionada con un aumento de la senescencia, una reducción de la capacidad de proliferación y una acumulación de ADN sin reparar (64). En la **Figura 4** se muestra que *MECP2* interactúa con múltiples genes que participan en los procesos de reparación del daño al ADN, dentro de estos genes están *Tet1*, *Birc6*, y *Spin1*, mientras que *Rbbp8*, *Rad50*, *Fan1*, *Brca1*, y *Brca2*, que codifican factores que participan en la reparación del quiebre de la doble hebra de ADN (65).

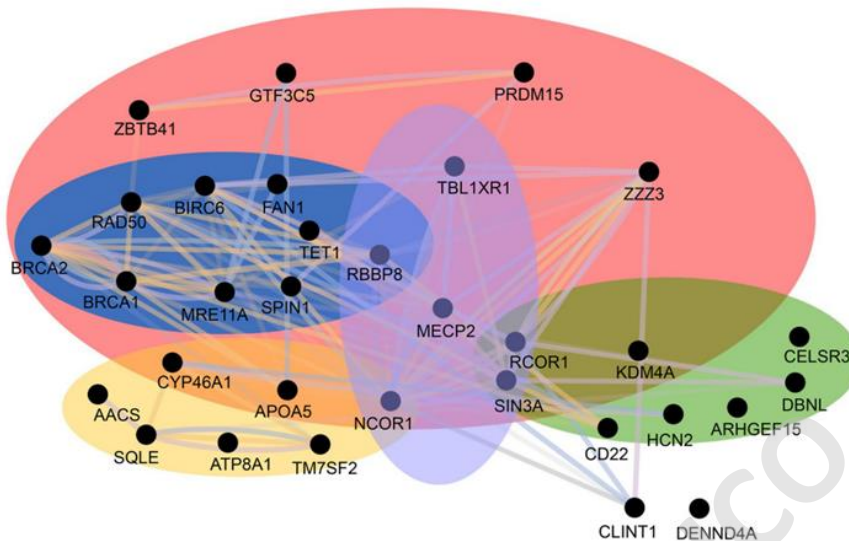


Figura 4. Red de interacciones que muestra la relación entre proteínas con *MECP2*. Las proteínas se representan como puntos negros o nodos, y las relaciones entre ellas se representan como líneas de color o bordes, en los que las líneas púrpuras denotan co-expresión, las líneas rosas denotan interacciones físicas, las líneas naranjas denotan interacciones predichas, las líneas azules denotan co-localización y las líneas verdes denotan dominios proteicos compartidos. La agrupación posterior (indicada con óvalos de colores) reveló que la mayoría de las proteínas se agrupan en las categorías generales de metabolismo lipídico y homeostasis (amarillo), función sináptica (verde) y respuesta al daño del ADN (azul). Varias proteínas también participan en la represión transcripcional (púrpura). El 63% se incluye en la amplia categoría de regulación de la actividad del ADN (rosa). Imagen obtenida de Enikanolaiye, A., Ruston, J., et al. *Genome research*. 2020

Estas evidencias muestran que *MECP2* estaría, al menos, indirectamente relacionado con la RDA, además toda esta información da indicios que mutaciones de *MECP2* en RTT, disminuye la reparación del quiebre de la doble hebra de ADN o incluso no provocaría la RDA.

La expansión de elementos móviles y las alteraciones en la reparación del ADN pueden manifestarse como un aumento en la recombinación del ADN, produciendo genes fusión o genes quiméricos, estos son la yuxtaposición de 2 genes diferentes resultando en una estructura que tiene la “cabeza” de un gen y la “cola” de otro (66). Estos genes pueden ser el resultado de un reordenamiento cromosómico estructural, principalmente, translocación, inversión, amplificación y/o delección, o aberraciones no estructurales causado por un empalme *cis-* y *trans-* (67). El empalme *cis-* es un mecanismo intramolecular que elimina los intrones y une los exones dentro del mismo transcrito de ARN, y el empalme *trans-* es aquel que elimina los intrones u outrones une los exones que no están dentro del mismo transcrito de ARN (68).

Los TQ se han detectado principalmente tanto en tumores mesenquimales, epiteliales y otros tumores sólidos (69), pero la participación de estos en otras enfermedades no se ha estudiado en profundidad. Dependiendo del promotor que dirija la expresión del gene-fusión, se pueden detectar la presencia de TQ con herramientas computacionales específicas durante la etapa de alineamiento a partir de datos crudos de RNA-seq.

Dada la evidencia expuesta este trabajo busca detectar y caracterizar transcritos quiméricos a partir de la vasta base de datos de RNA-seq públicos de RTT, sus modelos y controles, y determinar su presunta relación con el desarrollo de la enfermedad.

SOLO USO ACADÉMICO

2 HIPÓTESIS Y OBJETIVOS

2.1 Hipótesis:

En modelos murinos de Síndrome de Rett existe un aumento en la tasa de recombinaciones genómicas que son detectables por la presencia de transcritos quiméricos.

2.2 Objetivo General:

Identificar y comparar la presencia de transcritos quiméricos en modelos de Síndrome de Rett

Objetivos específicos:

- 2.2.1 Identificar y cuantificar los transcritos quiméricos existentes en los datos de repositorios de libre acceso.
- 2.2.2 Identificar y evaluar los puntos de quiebre recurrentes de los transcritos quiméricos en mutante.

SOLO USO ACADÉMICO

3 MATERIALES Y MÉTODOS

3.1 Servidor

Todas las herramientas informáticas utilizadas, se ejecutaron en los servidores del Centro de Genómica y Bioinformática de la Universidad Mayor. Se empleó el servidor Darwin el cual posee 16 nodos con 28 núcleos de procesamiento cada uno y un total de 62 Gb de memoria RAM en cada nodo.

3.2 Descarga de secuenciaciones de RNA-Seq de muestras utilizadas

Se descargaron 50 secuenciaciones de RNA-seq de modelos murinos de Síndrome de Rett y 50 secuencias de sus respectivos controles, las cuales se seleccionaron según los siguientes criterios: muestra de tejido cerebral, la estrategia de librería fuese RNA-seq, la secuencia fuese “Paired end” y el modelo utilizado fuese *Mus musculus*; Esto se puede ver de mejor manera en las **Tablas 2 y 3**. Todas las secuencias fueron descargadas desde *European Nucleotide Archive* (ENA) (70).

Tabla 2. Criterios de búsqueda en muestras silvestres

Especie	Tejido cerebral	Modelo	Sexo	Edad	Estrategia de librería	Replicas
Mus musculus	microglia	Silvestre	Hembra	5 Semanas	RNA-Seq	3
Mus musculus	Giro dentado	Silvestre	Hembra	8-9 Semanas	RNA-Seq	24
Mus musculus	Corteza cerebral	Silvestre	Hembra	22 Semanas	RNA-Seq	4
Mus musculus	Cerebelo	Silvestre	Hembra	22 Semanas	RNA-Seq	4
Mus musculus	Cultivo cortical	Silvestre	Macho	P0-P1 (10 DIV)	RNA-Seq	9
Mus musculus	Hipocampo	Silvestre	Macho	7-Semanas	RNA-Seq	6

Tabla 3. Criterios de búsqueda en muestras mutantes

Especie	Tejido cerebral	Modelo	Sexo	Edad	Estrategia de librería	Replicas
Mus musculus	microglia	MeCP2-KO	Hembra	5 Semanas	RNA-Seq	3
Mus musculus	Giro dentado	MeCP2-KO	Hembra	8-9 Semanas	RNA-Seq	24
Mus musculus	Corteza cerebral	MECP2 KO	Hembra	22 Semanas	RNA-Seq	4
Mus musculus	Cerebelo	MECP2 KO	Hembra	22 Semanas	RNA-Seq	4
Mus musculus	Cultivo cortical	MECP2 KO	Macho	P0-P1 (10 DIV)	RNA-Seq	9
Mus musculus	Hipocampo	MECP2 KO	Macho	7-Semanas	RNA-Seq	6

Los datos se descargaron en el servidor del Centro de Genómica y Bioinformática de la Universidad Mayor, y se utilizó el programa en Browser Tools (71), bajo los siguientes parámetros:

“enaDataGet -f fastq SRRxxxx”

3.3 Análisis de calidad y recorte de secuencias

Se procedió a evaluar la calidad de las 100 muestras utilizando el programa FASTQC (v.0.11.9)(72), el cual se utiliza para proporcionar de una forma sencilla el control de calidad de los datos de secuencia, utilizando el siguiente comando en la plataforma Unix “Fastqc SRRxxxx –outdir carpeta/de/salida”. Luego de analizar las calidades de las secuencias se continuó con el “trimming” o recorte de las secuencias, esto se hizo para eliminar regiones de la secuencia de baja calidad y eliminar restos de las secuencias adaptadoras de Illumina que pudieron haber quedado durante la secuenciación. Para esto se utilizó la herramienta informática “Trim Galore” (73), utilizamos el siguiente comando “trim_galore –cores 4 –fastqc –paired”, se agregó la opción “–paired”, debido a que todas las secuencias eran “Paired end”. Después de tener las secuencias ya recortadas, se volvió a revisar la calidad de estas, utilizando FASTQC para evaluar el efecto del procesamiento. Las secuencias recortadas se consideraban correctas cuando en el segundo reporte de FASTQC se observa que no tenían contaminación de adaptadores. Finalmente, con las secuencias ya recortadas y de buena calidad, se pasó a alinear las secuencias descargadas. En la **Figura 5** se puede observar un diagrama de los pasos que se siguieron en esta etapa.

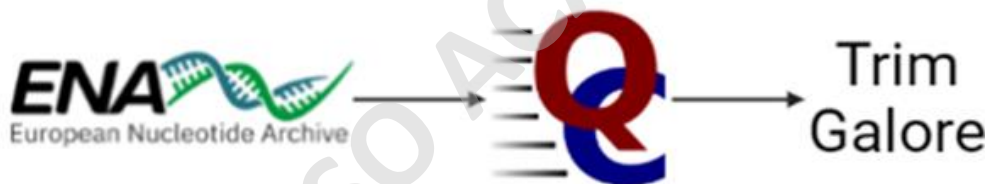


Figura 5. Diagrama de flujo para el análisis de calidad, “Trimming” y alineamiento.

3.4 Alineamiento de secuencias y análisis utilizando la herramienta ARRIBA

En este trabajo de tesis se consideraron 2 herramientas informáticas para procesar las secuencias de RNA-seq, los programas utilizados fueron ARRIBA (74) y STAR-Fusion (75). Ambos programas son herramientas para la detección de TQ en datos de RNA-seq, para la cual alinean las secuencias contra un genoma de referencia para generar los análisis posteriores.

Para los análisis con ARRIBA, se generó un índice, utilizando el genoma de referencia GrCM38 (76), utilizando STAR (77) con el siguiente comando:

```
“STAR --runThreadN 5 --runMode genomeGenerate --genomeDir /ruta/hacia/la/carpeta/star_index --genomeFastaFiles /ruta/hacia/genoma/formato/fastq/de/GrCm38/genome.fa --sjdbGTFfile /ruta/hacia/archivo/fomato/GTF/genes.gtf”.
```

Después se procedió a alinear las secuencias previamente descargadas utilizando el programa STAR (77), y utilizando el siguiente comando, el *output* obtenido de este alineamiento se utilizará solo para analizar las secuencias con ARRIBA.

```
“STAR --runThreadN 5 --runMode alignReads --genomeLoad NoSharedMemory --readFilesCommand zcat --outStd SAM --outSAMtype BAM Unsorted --outSAMunmapped Within --outBAMcompression 0 --outFilterMultimapNmax 50 --peOverlapNbasesMin 10 --alignSplicedMateMapLminOverLmate 0.5 --alignSjstitchMismatchNmax 5 -1 5 5 --chimSegmentMin 10 --chimOutType WithinBAM --chimJunctionOverhangMin 10 --chimScoreDropMax 30 --chimScoreJunctionNonGTAG 0 --chimScoreSeparation 1 --chimSegmentReadGapMax 3 --chimMultimapNmax 50 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0”
```

Luego con las secuencias ya alineadas se prosiguió a utilizar el programa ARRIBA (74). Para el cual se utilizaron el índice y el alineamiento previamente hecho, para ejecutar este programa, utilizando el siguiente comando:

```
“arriba -x Alignes.out.bam -o SRRxxxx.tsv -O fusions.discarded.tsv -a /ruta/hacia/genoma/formato/fasta/de/GrCm38/genome.fa -g /ruta/hacia/archivo/genes/fomato/GTF/genes.gtf”
```

En la **Figura 6** se puede observar un diagrama de los pasos que se siguieron en esta etapa.



Figura 6. Diagrama de flujo muestra etapas de alineamiento de secuencias a referencia, obtención y análisis de transcritos quiméricos.

3.5 Alineamiento de secuencias y análisis utilizando la herramienta STAR-Fusion

Para generar los análisis de los datos de RNA-seq en STAR-Fusion, primero se deben alinear las secuencias previamente descargadas. Para esto STAR-Fusion debe utilizar un programa llamado “*Trinity Cancer Transcriptome Analysis Toolkit*” (CTAT) (78), el cual tiene un kit de herramientas para el análisis transcriptómico para aprovechar el RNA-Seq y poder obtener información sobre la biología de los

transcriptomas del cáncer. Este programa trabaja junto a STAR-Fusion para generar una librería, la cual es requerida por STAR-Fusion, ya que la utiliza para el alineamiento y el análisis de TQ en las secuencias. Se utilizó el genoma de referencia GRCM38 (76) para la generación esta librería Se utilizó el siguiente comando para la generación de esta librería:

```
"prep_genome_lib.pl --genome_fa /ruta/hacia/genoma/formato/fasta/de/GrCm38/genome.fa --gtf ruta/hacia/archivo/GTF/genes.gtf --annot_filter_rule AnnotFilterRule.pm --dfam_db mouse --pfam_db current".
```

La librería generada contiene el índice que se necesita para el alineamiento de las secuencias y además archivos con anotaciones de genes, de TQ, de dominios de proteínas y el genoma de referencia utilizado.

STAR-Fusion utiliza una sola línea de comando para alinear y hacer el análisis de TQ, como se puede observar en el comando más adelante, el nombre de la librería que se generó anteriormente con CTAT es "ctat_genome_lib_build_dir", en este mismo comando de STAR-Fusion, se agregó el parámetro "--no_annotation_filter", para evitar que el resultado del análisis filtre secuencias quiméricas encontradas que no estuvieran anotadas en su base de datos. También se disminuyó FFPM (lecturas totales por millón de fragmentos de fusión), para aumentar la sensibilidad del programa para así ver posibles TQ esporádicos. Se utilizó la siguiente línea de comando para generar el alineamiento y el análisis utilizando STAR-Fusion:

```
"STAR-Fusion --genome_lib_dir ctat_genome_lib_build_dir --left_fq reads_1.fq --right_fq reads_2.fq --no_annotation_filter --min_FFPM 0.01".
```

3.6 Filtrado de datos

Se obtuvieron un total de 200 reportes, 100 por cada programa utilizado, de ellos 50 corresponden a muestras de Síndrome de Rett y sus 50 controles. En todos estos se filtraron los datos obtenidos en función de la distancia de los puntos de empalme obtenidos en los reportes. Se consideraron solamente las secuencias quiméricas donde los sitios dadores y aceptores estuvieran a más de 100.000 pares de base de distancia. También se eliminaron los casos de secuencias intergénicas que se obtenían en los reportes.

Finalmente se tomaron todos los reportes obtenidos, y se fueron analizando los puntos de quiebres (PQ), esto se refiere al punto donde se inicia y termina el empalme entre 2 exones de genes distintos. Se separaron los PQ del lado dador o inicial y los PQ del lado aceptor o final, todos estos datos se filtraron y se cuantifican los PQ en cada uno de los reportes, luego de eso se determinó las ocurrencias de cada PQ en todos los reportes y se compararon entre los reportes de las transcriptómicas de muestras de Síndrome de Rett y sus controles. Estos se filtraron con los siguientes comandos:

```
# Borrar las columnas con información innecesaria para el análisis
cut -f -22 archivo.txt > archivo_2.txt # Un ejemplo del archivo.txt se puede ver en el Anexo 1, al cual se
eliminó desde la columna 22 en adelante, ejemplo de archivo_2.txt en Anexo 2.

# Borrar los transcritos que son intergénicos
awk '!($1==$2&&$2==$1)' archivo_2.txt > archivo_3.txt # Se eliminaron los genes que se repetían en
diferentes columnas Anexo 3.
awk '!($1~$2)' archivo_3.txt > archivo_4.txt # Se consolidó eliminando en cada muestra los transcritos
repetidos, ejemplo de archivo_4.txt en Anexo 4.
awk '!($2~$1)' archivo_4.txt > archivo_5.txt # Se utilizó como reporte final el archivo_5.txt, ejemplo en
Anexo 5

# Cuantificar ocurrencias dentro de solo los PQ de los lados dadores o receptores
awk 'index($5, "puntodequebredador")' archivo_5.txt | wc -l
awk 'index($6, "puntodequebreaceptor")' archivo_5.txt | wc -l
```

3.7 Generación de gráficos y Diagrama de cuerdas

Se graficaron las ocurrencias de los sitios dadores y aceptores para la presentación de los resultados obtenidos, primero se generaron los gráficos con los candidatos identificados por los programas de ARRIBA y STAR-Fusion

A partir de los PQ de los reportes obtenidos de STAR-Fusion y ARRIBA, se determinaron las ocurrencias de los PQ que estaban en las muestras de *MECP2-KO* y las muestras silvestres. Se generaron 4 tablas distintas, las cuales se pueden ver en el Anexo 6 y 7, y se prosiguió a hacer un *scatter plot* de las ocurrencias de todos los PQ utilizando el lenguaje R (79), para la cual se utilizó la interfaz gráfica de *RStudio* (80), y se utilizaron distintas librerías de R como *ggplot2* (81), *ggpointdensity* (82), *viridis* (83), *hrbrthemes* (84).

```

# Se descargaron y se cargaron las librerías
>library(ggpointdensity)
>library(viridis)
>library(ggplot2)
>library(hrbrthemes)

# Leer el archivo Excel y darle una variable
>Variable <- read_excel("Ruta/hacia/el/archivo")

#Generar el Scatter plot y agregar la línea de tendencia
>ggplot(data=Variable, mapping = aes(x = Mecp2ko, y = control)) +
  geom_point(size = 2) + theme_bw(15) +
  geom_smooth(method="loess" , color="red", fill="#69b3a2", level=0.5 ) +4
  xlim(0, 50) +
  ylim(0, 50) +
  geom_pointdensity(size = 4) +
  scale_color_viridis(option = "turbo") +
  theme_ipsum()

```

Los PQ ya graficados se separaron en 3 grupos, PQ esporádicos, medios y recurrentes. Esto se hizo para poder analizar de mejor manera las distintas cantidades de cada punto de quiebre identificado, una vez identificados y contabilizados cada uno de los PQ en las muestras analizadas, se prosiguió a separar los datos en 3 tipos distintos, PQ esporádicos, medios y recurrentes. Los PQ esporádicos son aquellos detectados de 1 y 5 veces en todas las muestras, los PQ medio son aquellos que son detectados entre 5 y 15 veces y finalmente los recurrentes son aquellos que se encuentran más de 15 veces. Cuando un punto de quiebre empalma con más de una región cromosómica dentro de una misma muestra se contabiliza una vez por cada sitio de empalme. Luego se utilizó *GraphPad* (versión 9.3.1) (85) para graficar los datos.

Para visualizar las coordenadas entregadas por los reportes de ambos programas utilizados y apreciar los saltos cromosómicos se generó un gráfico de cuerdas, para esto se utilizó R (79) y el paquete *RCircos* (86), esto se hizo según los siguientes comandos:

```

# Descargar y cargar las librerías
>library(RCircos)

# Cargar los ideogramas cromosómicos
>data ("UCSC.Mouse. GRCm38.CytoBandIdeogram")
>data("UCSC.mm10.chr")
>head (UCSC.Mouse.GRCm38.CytoBandIdeogram)

# Inicializar los componentes esenciales o parámetros de Rcircus
>chr.exclude <- NULL
>cyto.info <- UCSC.Mouse. GRCm38.CytoBandIdeogram
>tracks.inside <- 10
>tracks.outside <- 0

# Representar el ideograma con sus parámetros
>RCircos.Set.Plot.Area()
>RCircos.Set.Core.Components(cyto.info, chr.exclude,
                             tracks.inside, tracks.outside)
RCircos.Set.Plot.Area
RCircos.Chromosome.Ideogram.Plot()

# Cargar la variable de los datos
link.data<- read.table("C:/Users/memo0/Documents/cuerdas/star/WT-STAR-cuerdas-2.txt",
                      sep="\t", quote="", head=T) # anexo de cuerdas

# Seleccionar el carril que se verán los datos
track.num <- 1

# Cargar el idiograma con los datos y parámetros finales
RCircos.Link.Plot(link.data, track.num, by.chromosome = TRUE)

```

3.8 Análisis estadísticos

Se llevó a cabo la prueba estadística de Wilcoxon Mann Whitney para verificar la significancia estadística. Esta prueba realiza una comparación estadística de la media y determina si existe una diferencia en la variable dependiente para dos grupos independientes.

Para determinar si existe una diferencia entre las variables, primero se determina un intervalo de confianza, en este caso usamos 0,05, para la diferencia entre las dos medianas de las poblaciones. Utilizando las medianas de las poblaciones, se obtiene el *P value*, que es indica/representa la probabilidad de que la hipótesis nula sea cierta, el cual luego se compara con el intervalo de confianza, si el *P value* es menor que 0,05, significa que se rechaza la hipótesis nula.

En esta tesis se utilizó este análisis estadístico para comprobar si existen diferencias significativas entre las cantidades y ocurrencias de los TQ obtenidos de los resultados obtenidos por ARRIBA y STAR-Fusion.

3.9 Identificación y anotación de dominios proteicos

Utilizando la secuencia de del PQ con mayor ocurrencia, se averiguo si este pudiera estar generar algún dominio proteico. Por esto se buscó las secuencias de los sitios de empalme iniciales y que su sitio de empalme final, para esto se utilizó *UCSC Genome Browser* (87), la cual es una herramienta en la web que sirve de microscopio multipotente y que permite ver los diferentes cromosomas del genoma, de humano o ratón, a diferentes escalas, esta herramienta se usó para identificar las secuencias de estos PQ para luego traducirlos a sus aminoácidos usando *Expasy Translate* (88), esta es una herramienta en la web que permite traducir la secuencia de nucleótidos (ADN/RNA) en secuencias de proteínas. Finalmente se utilizó SMART (89), el cual es una herramienta que permite identificar y anotar los dominios de proteínas y analizar la arquitectura de los dominios proteicos en base a una secuencia aminoacídica. Estas 3 herramientas se utilizaron para observar los dominios de proteínas que estas secuencias codificaban.

SOLO USO ACADÉMICO

4 Resultados

4.1 Comparación y cuantificación de transcritos quiméricos reportados en muestras de Síndrome de Rett y muestras silvestre

Para poder determinar diferencias de transcritos quiméricos (TQ) entre las muestras de RTT y silvestres, se utilizaron los reportes obtenidos por los programas STAR-Fusion y ARRIBA. Se procedió a identificar los candidatos de TQ o fusión de genes existentes en cada reporte, filtrando aquellos transcritos que fueran intragénicos y que la distancia entre el punto de quiebre dador y receptor fuera menor 100.000 pares de bases. Posteriormente se graficó la cantidad de TQ identificados en cada muestra, por cada programa (**Figura 7**). Ambos gráficos se evaluaron de la misma manera, se puede observar que hay un leve aumento de TQ en las muestras *MECP2-KO* en ambos resultados. Hay que mencionar que en el caso de los datos obtenidos por el programa ARRIBA se obtuvieron una mayor cantidad de TQ por reporte, se puede observar la tabla de mejor manera en el Anexo 8.

Se realizó la prueba de Wilcoxon Mann Whitney para ambos grupos de datos, comparando la cantidad de candidatos de TQ existentes en los resultados de las muestras *MECP2-KO* y las muestras silvestres. Los resultados de STAR-Fusion muestran un resultado de *P value* de 0,0601, y en el caso de ARRIBA el *P value* fue de 0,438, lo que indicaría que en ambos programas no hay una diferencia significativa entre las muestras *MECP2-KO* y las muestras silvestres.

Los resultados entregados por ARRIBA y STAR-Fusion, nos indicarían que no existe una diferencia en los candidatos de TQ entregados por estos programas, por lo que no existiría un aumento de TQ en casos de *MECP2-KO* en comparación con las muestras silvestres.

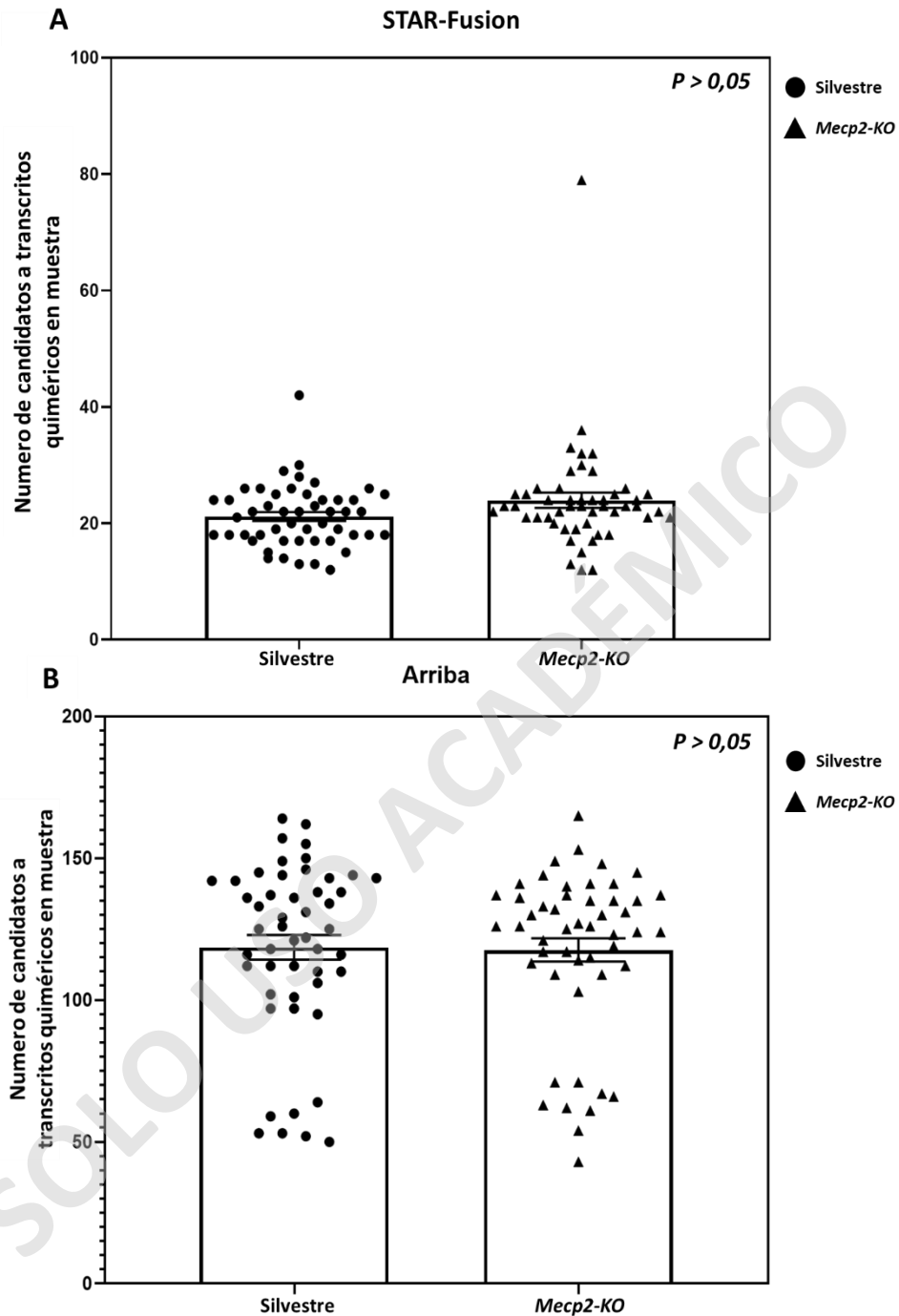


Figura 7. Cuantificación de candidatos TQ identificados por STAR-Fusion y ARRIBA (A y B). Los gráficos representan el número de candidatos de transcritos quiméricos detectados por los reportes de los softwares utilizados. Los puntos circulares representan los TQ de las muestras silvestre y los triángulos de las muestras *MECP2-KO*. Para este análisis se utilizaron secuenciaciones de 6 tejidos de cerebro distintos, los cuales son: Hipocampo, Cultivo cortical, Cerebelo, Giro dentado, Córtex y Microglía (A) gráfico que muestra la cantidad de los candidatos de transcritos quiméricos obtenidos por los reportes generados por STAR-Fusion, estos se separaron en muestras silvestre y los muestras *MECP2-KO*, los análisis estadísticos que se generaron fue una prueba de Wilcoxon Mann Whitney el cual dio un *P value* de 0,0601. (B) El gráfico fue generado con los candidatos de TQ obtenidos con el programa ARRIBA y se separaron entre los datos control y *MECP2-KO*, prueba de Wilcoxon Mann Whitney el cual dio un *P value* de 0,4380.

4.2 Cuantificación de puntos de quiebres obtenidos de los reportes de los programas utilizados

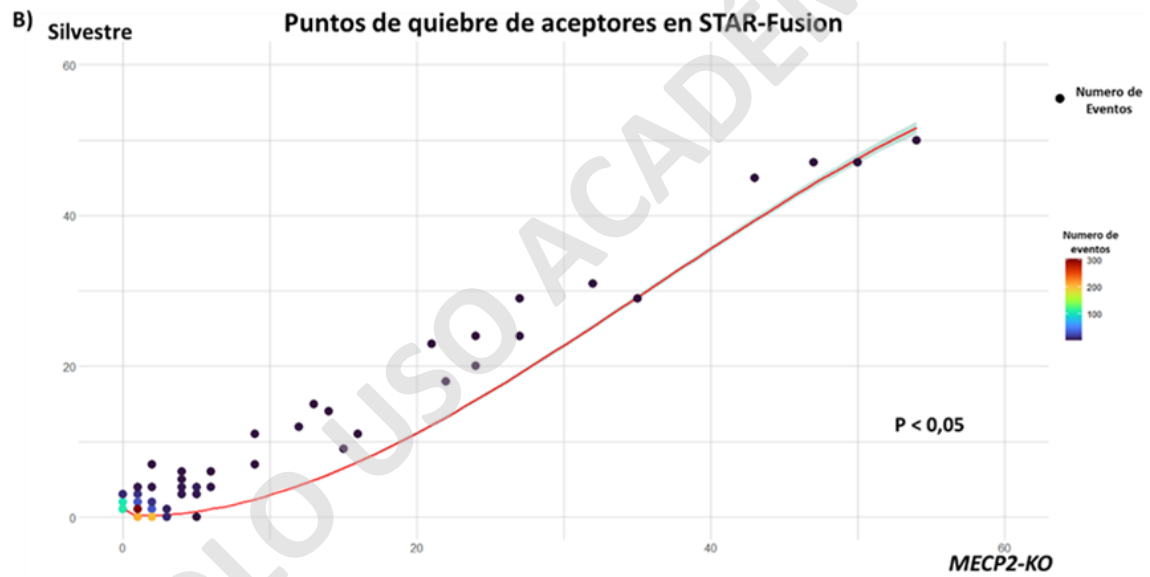
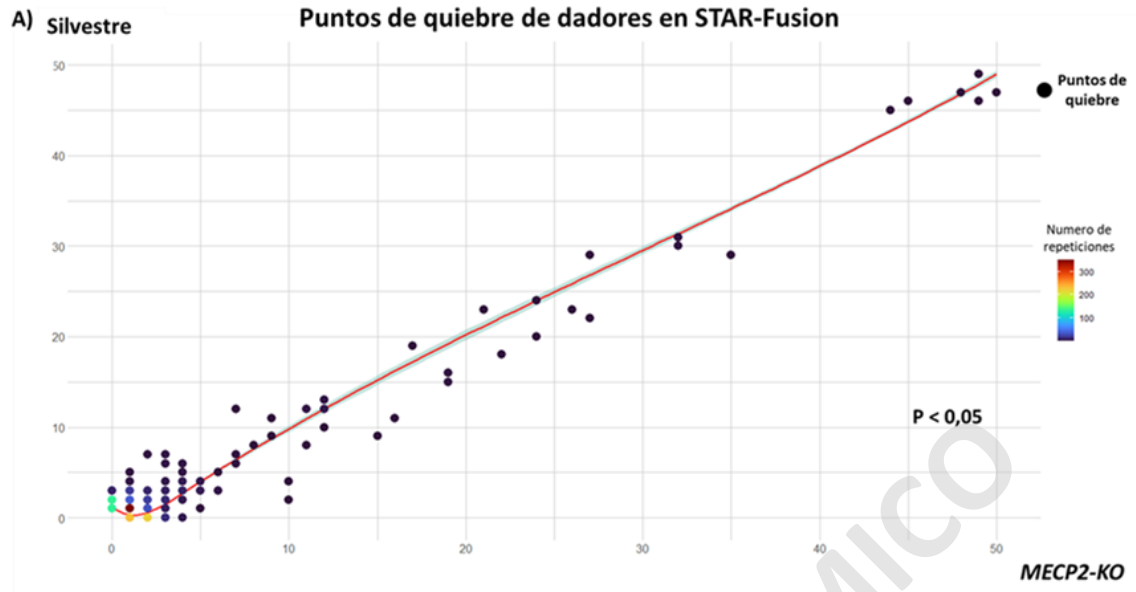
Un sitio dador puede empalmar con múltiples sitios aceptores y generar distintos TQ. Para simplificar el análisis se prosiguió a identificar la cantidad de PQ existentes en cada transcriptómica analizando por separado la ocurrencia del PQ del sitio dador y el PQ aceptor. Con los transcritos identificados, se prosiguió en separar los PQ en los de lado dador y aceptor de las 50 muestras silvestre y las 50 muestras de *MECP2-KO*, luego se determinó la cantidad de muestras que presentan el uso del sitio dador y/o aceptor. Se contaron las veces que cada punto quiebre se repetía dentro de cada reporte de los muestras silvestre y *MECP2-KO*, con estos datos se generaron 4 tablas las cuales se pueden ver en Anexo 6 y 7. En la **Figura 8** se puede ver los 4 gráficos que se generaron, cada uno correspondiente a los datos obtenidos del programa utilizado, se fijaron los límites de los ejes X e Y en 50, debido a que los datos que superan este límite se consideraron outliers, en el caso de los PQ de STAR, en el lado dador no hubieron outliers y en el lado aceptor hubo 3 y en los PQ de ARRIBA el lado dador hubo 4 outliers y en el lado aceptor 5.

En las **Figuras 8 A y B** son los gráficos generados con los datos obtenidos por el programa de STAR-Fusion, en la primera imagen se puede observar que la mayor concentración de datos se encuentra cercano al origen del gráfico (reflejado en el color de la coordenada). Se puede ver que hay un mayor número de PQ que se repiten 1 o 2 veces, esto sucede tanto en los PQ de los sitios dadores y aceptores de los datos recopilados por STAR-Fusion, según el *Heatmap* se puede decir que hay alrededor de 200 eventos que solo se encuentran en las muestras de *MECP2-KO*, también se puede observar que la línea de tendencia de la **Figura 8 A** tiene una pequeña curvatura hacia el eje X, a diferencia de la **Figura 8 B** que muestra una clara curvatura al eje X demostrando que existe una mayor cantidad de PQ en las muestras *MECP2-KO* que tienen una ocurrencia de 1 a 5. Para ambos conjuntos de datos se analizó las diferencias de los datos por una prueba de Wilcoxon Mann Whitney, en el cual ambos obtuvieron un *P value* menor a 0,0001, indicando que existe una diferencia significativa entre muestras silvestre y muestras *MECP2-KO*.

Los resultados de STAR-Fusion nos indicarían que existe una mayor cantidad de ocurrencias de los PQ en las muestras *MECP2-KO* en comparación con las muestras silvestre, lo que nos indicaría que hay una mayor cantidad de PQ que se repiten en los diferentes candidatos de TQ en los resultados obtenidos por STAR-Fusion. Esto nos podría decir que en casos de la existencia de un *MECP2-KO* podría provocar un aumento de TQ, ya existentes en casos silvestres, en comparación con casos de RTT.

En el caso de los PQ obtenidos por el programa ARRIBA (**Figura 8 C y D**), se aprecia la misma tendencia en los eventos esporádicos o presentes en 4 o menos muestras de un tipo, aunque en el caso de este programa hay menos PQ que se repiten 1 o 2 veces en muestras *MECP2-KO* en comparación con el programa STAR-FUSION, aun así, hay una leve mayoría en los PQ del sitio dador y aceptor, en el caso de la línea de tendencia, estas se curvan al inicio y al final del gráfico, a diferencia con los gráficos obtenidos por STAR-FUSION. Los datos obtenidos se evaluaron mediante la prueba de Wilcoxon Mann Whitney correspondiente, el cual dio un *P value* 0,9686 en PQ de los sitios aceptores y un *P value* de 0,3118 En los PQ del sitio dador, se concluye que ninguno de los 2 casos existe una diferencia significativa entre ambas muestras. Los resultados obtenidos por ARRIBA nos indicarían que no existe una diferencia entre la cantidad de ocurrencias de los PQ obtenidos de los reportes entregados por ARRIBA, por lo que no habría ninguna diferencia en la cantidad de TQ en casos de RTT en comparación con casos silvestres.

SOLO USO ACADÉMICO



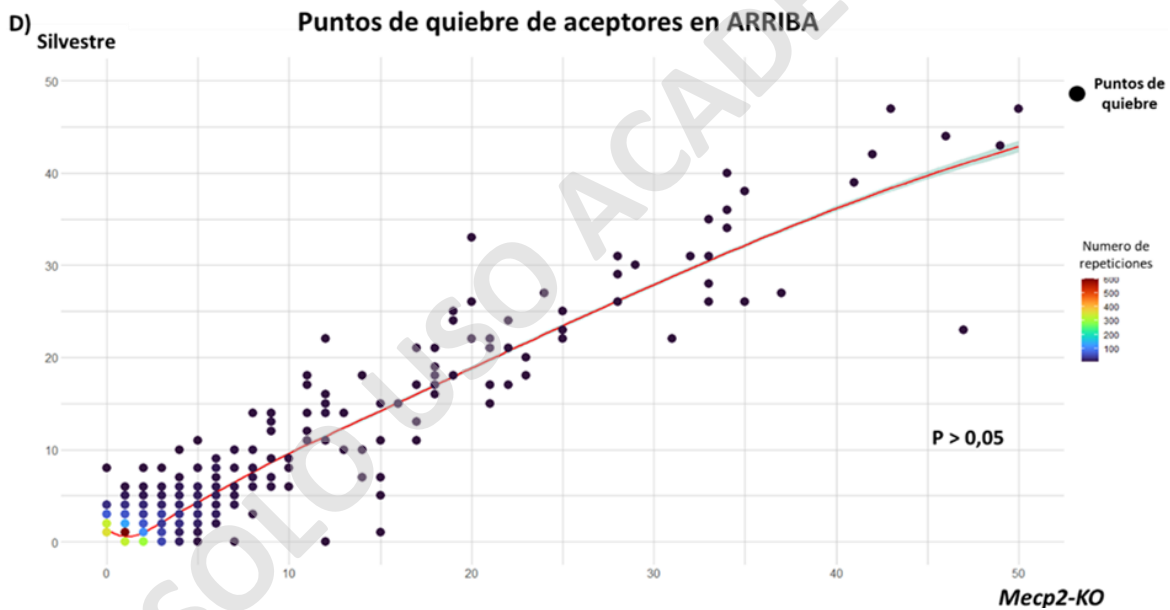
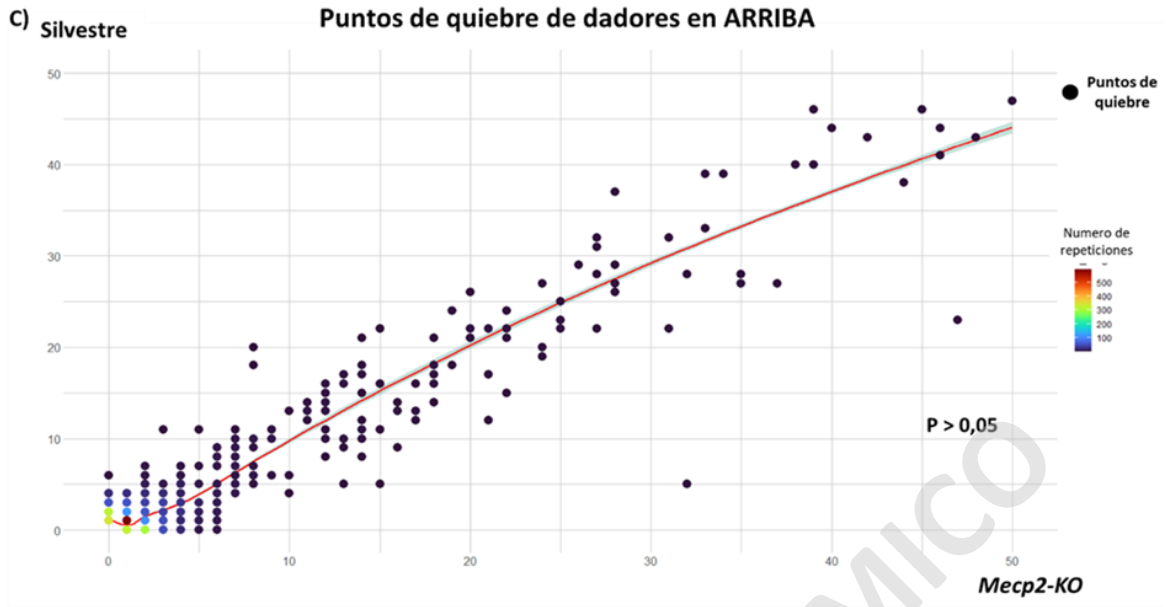


Figura 8. Cuantificación e identificación de puntos de quiebre existentes en los candidatos de transcritos quiméricos (A - D). Se graficaron las ocurrencias de los PQ de los candidatos de transcritos quiméricos obtenidos por los programas utilizados. Se generaron 2 gráficos del tipo *scatter plot* junto a un *heatmap*, se utilizaron solo los PQ del lado dador o aceptor. Para cada punto, el valor en X corresponde a las ocurrencias de determinado punto de quiebre en animales *MECP2-KO* y el valor de Y corresponde a las ocurrencias del punto de quiebre en animales silvestre. Los colores representan la cantidad de PQ que comparten la misma distribución entre mutantes y silvestre, mostrando en negro las coordenadas con un solo punto de quiebre y en rojo las coordenadas con más de 500 PQ. Los gráficos se hicieron con los resultados obtenidos por el programa STAR-Fusion. (A) Se generó con los PQ del lado dador y los cuales se analizaron por la prueba de *Wilcoxon Mann Whitney* el cual dio un *P value* <0,0001. En (B) se generaron con los PQ del lado aceptor y el análisis estadístico fue la prueba de *Wilcoxon Mann Whitney* el cual dio un *P value* <0,0001. A partir de los resultados obtenidos por el programa ARRIBA, se analizaron con los PQ del lado dador y aceptor con el análisis estadística prueba de *Wilcoxon Mann Whitney* el cual no dio diferencias significativas (C y D). La línea de tendencia se generó utilizando el método “*Loess*” el cual utiliza la siguiente fórmula $y \sim \text{poly}(x, 2)$

4.3 Distribución de los puntos de quiebres extra e intracromosómicos sobre el genoma del ratón

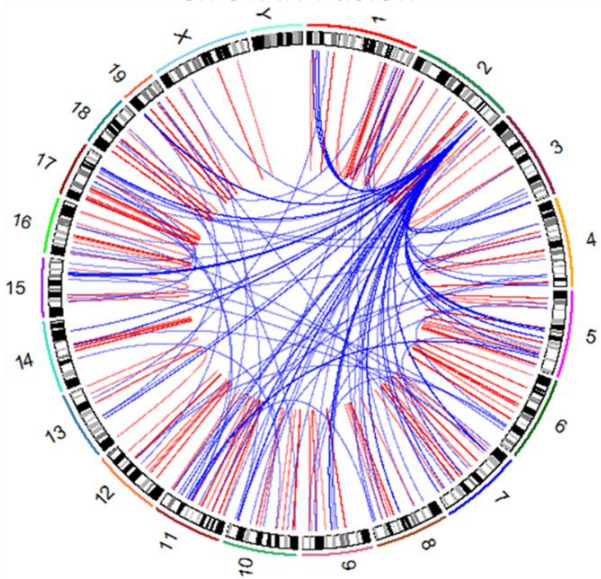
A partir de las coordenadas genómicas de los PQ detectados en los TQ, se realizaron diagramas de cuerdas que conectan los sitios dadores y aceptores. Esto se llevó a cabo a partir de los resultados de STAR-Fusion y ARRIBA. En el diagrama de cuerdas se puede observar que están los 21 cromosomas del ratón (*Mus musculus*). Dentro del círculo se pueden ver 2 tipos de líneas, las de color rojo conectan los PQ candidatos para transcritos quiméricos que son intracromosomal y las líneas de color azul conectan los puntos que son extracromosomales (**Figura 9**).

Los 2 primeros diagramas se generaron con los PQ obtenidos con STAR-Fusion. En la **Figura 9 A** se pueden ver las conexiones de los PQ en las muestras silvestre, las cuales dieron un total de 1058 transcritos quiméricos que se detectaron en las 50 muestras silvestre, de las cuales sólo 247 son de candidatos extracromosomales y los restantes 811 son de intracromosomales. En cambio, en las muestras *MECP2-KO* se detectaron 1197 transcritos quiméricos, de estos 313 son extracromosomales y 884 son intracromosomales, mostrando que existe una mayor cantidad transcritos quiméricos en las muestras de *MECP2-KO* (**Figura 9 A-B**).

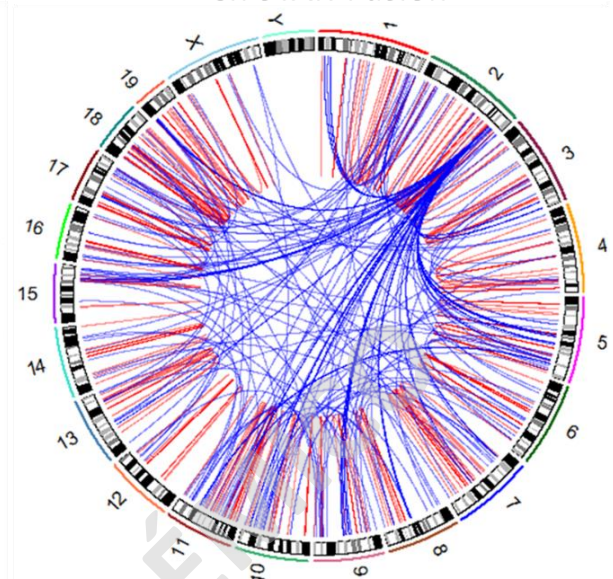
El mismo análisis fue realizado con los resultados de ARRIBA. En el caso de los controles existen un total de 3133 transcritos quiméricos, de los cuales 1518 son casos de intracromosomales y 1615 son extracromosomales, en el diagrama de las muestras *MECP2-KO* dieron un total de 3150 candidatos a transcritos quiméricos de los cuales 1523 son intracromosomales y 1627 son extracromosomales (**Figura 9 C-D**).

Los resultados obtenidos en los diagramas nos demuestran que existen, una mayor cantidad de TQ reportados por ARRIBA en comparación con los resultados de STAR-Fusion, además se pueden ver que existen PQ que tienen una mayor cantidad de conexiones, las cuales pueden representar un “hotspot”, que este alterando un gen en específico, el cual podría alterarse o provocar alguna sintomatología específica en los casos de RTT.

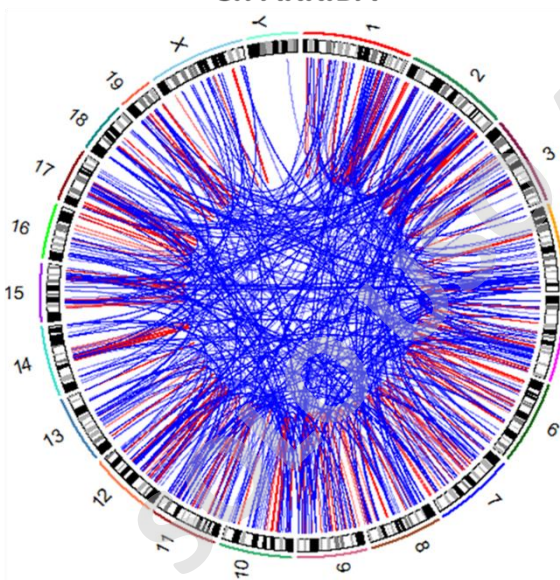
A) Diagrama de cuerdas silvestre en STAR-Fusion



B) Diagrama de cuerdas *Mecp2-KO* en STAR-Fusion



C) Diagrama de cuerdas silvestre en ARRIBA



D) Diagrama de cuerdas control en ARRIBA

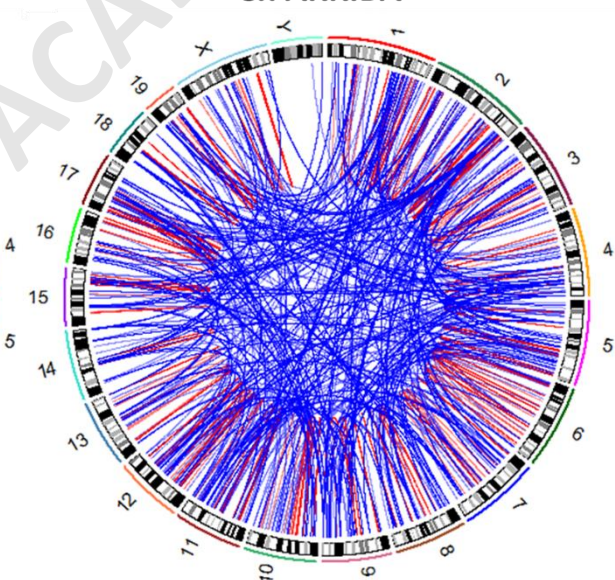


Figura 9. Diagrama de cuerdas conectado a los puntos de quiebre de transcritos quiméricos detectados (A - D). La unión entre los PQ del lado dador y aceptor, obtenidos por los reportes generados por *STAR-Fusion* y *ARRIBA*, las líneas de color azul son las uniones entre PQ de diferentes cromosomas o extracromosomales y las líneas rojas son la unión entre los PQ del mismo cromosoma o intracromosomales. (A - B) Los diagramas se generaron en base a los PQ de los reportes de *STAR-Fusion*. (C - D) Los diagramas se generaron en base a los PQ de los reportes de *ARRIBA*.

4.4 “Hotspot”, genes asociados y dominios proteicos

Se analizó los transcritos intra y extracromosomales de todas las muestras, los cuales se utilizaron para generar el diagrama de cuerdas a partir de los datos obtenidos por los reportes entregados por ambos programas (**Figura 7**). La información obtenida se usó para cuantificar los transcritos quiméricos totales presentes en las muestras y se generaron las siguientes tablas.

Tabla 4. Eventos totales de transcritos quiméricos encontrados. Comparación de transcritos quiméricos totales obtenidos con las 2 herramientas utilizadas en el análisis de las 100 muestras a lo largo del trabajo.

Muestras	STAR-Fusion		Arriba	
	Silvestre	<i>Mecp2-KO</i>	Silvestre	<i>Mecp2-KO</i>
Total	1058	1197	3133	3150
extracromosomales	247	313	1615	1627
intracromosomales	811	884	1518	1523

En línea con lo visto anteriormente, ambos programas dan a conocer que las muestras de *MECP2-KO* tienen una mayor cantidad de TQ en comparación con las muestras silvestres (**Tabla 4**). En el caso de STAR-Fusion hay una diferencia de 11,6% a favor de las muestras mutantes respecto al control; en el caso ARRIBA hay una pequeña diferencia del 0,5% a favor de las muestras mutantes.

En la **Tabla 4** los reportes contabilizan los TQ cada vez que se encuentren en una muestra. Para evaluar la variedad de TQ detectados, se procedió a filtrar los transcritos repetidos encontrados en múltiples muestras. En la **Tabla 5** se puede observar que en el caso de ARRIBA hay una mayor cantidad de transcritos quiméricos únicos en las muestras silvestre en comparación con las muestras *MECP2-KO* teniendo una diferencia del 4,6%, lo que nos dice que este programa detectó una mayor cantidad de TQ únicos en las muestras silvestre.

Tabla 5. Eventos únicos de transcritos quiméricos encontrados. Comparación de transcritos quiméricos únicos obtenidos con las 2 herramientas utilizadas en el análisis de las 100 muestras a lo largo del trabajo.

Muestras	STAR-Fusion		Arriba	
	Silvestre	<i>Mecp2-KO</i>	Silvestre	<i>Mecp2-KO</i>
Total	238	339	739	705
extracromosomales	95	150	290	258
intracromosomales	143	189	449	447

En la **Tabla 5** se puede observar que en el caso de ARRIBA hay una mayor cantidad de transcritos quiméricos únicos en las muestras silvestre en comparación con las muestras MECP2-KO teniendo una diferencia del 4,6%, lo que nos dice que este programa detectó una mayor cantidad de TQ únicos en las muestras silvestre.

En el caso de STAR-Fusion se observa que hay mayor cantidad de TQ en las muestras *MECP2-KO*, tanto como en los resultados extra e intracromosomales. Las muestras *MECP2-KO* tienen un 30% más de TQ quiméricos en comparación con las muestras silvestres.

Los resultados obtenidos de STAR-Fusion, que muestran una mayor cantidad de TQ únicos en las muestras de *MECP2-KO*, estarían sugiriendo que hay una mayor posibilidad de que estos TQ puedan alterar la expresión de diversos genes, y que esto podría provocar las sintomatologías relacionadas con RTT. A diferencia con los resultados obtenidos por ARRIBA, los cuales previamente no han demostrado que existe una diferencia entre ambas muestras.

Luego utilizando los diagramas de cuerdas se observa que en algunos casos existe una acumulación de líneas en un punto específico, los cuales se pueden hallar analizando la ocurrencia de cada uno de los PQ. Para eso se analizó cuáles eran los 3 PQ dadores y aceptores con mayor ocurrencia, esto son considerados “Hotspot”. Estos PQ y sus genes asociados están informados en la **Tabla 6 y Tabla 7**, para STAR-Fusion y ARRIBA respectivamente.

Tabla 6. “Hotspot” encontrados utilizando el programa STAR-Fusion. PQ con mayor ocurrencia del lado dador y aceptor y el gen asociado a este PQ, obtenidos de los resultados entregados por el análisis generado por STAR-Fusion.

Dador STAR-Fusion	Ocurrencias Silvestre	Ocurrencias <i>MECP2-KO</i>	Gen
chr18: 23310271	49	49	Gm19389
chr16: 39984727	47	50	LOC100502943
chr17: 8046616	52	45	Gm1604A
Aceptor STAR-Fusion	Ocurrencias Silvestre	Ocurrencias <i>MECP2-KO</i>	Gen
chr2: 158382504	196	210	Snhg11
chr18: 23475116	95	94	Dtna
chr16: 41533453	77	81	Lsamp

Tabla 7. “Hotspot” encontrados utilizando el programa ARRIBA. Q con mayor ocurrencia del lado dador y aceptor y el gen asociado a este PQ, obtenidos de los resultados entregados por el análisis generado por ARRIBA.

Dador ARRIBA	Ocurrencias Silvestre	Ocurrencias <i>MECP2-KO</i>	Gen
chr4: 42181877	130	79	Fam205a2, Gm2002
chr17: 7079510	47	53	Gm1604b
chr9: 13298561	47	50	Ccdc82(6208), Phxr4(132800)
Aceptor ARRIBA	Ocurrencias Silvestre	Ocurrencias <i>MECP2-KO</i>	Gen
chr2: 158382504	177	220	Snhg11
chr4: 42871557	162	142	Fam205c
chr17: 7227913	55	61	Rps6ka2

En ambas tablas se puede ver que el punto con mayor cantidad de ocurrencias en el aceptor es el Chr2: 158392504, que posee como gen asociado el *Small nuclear RNA host gene 11 (SNHG11)*, los miembros de la familia SNHG han sido reconocidos como *long non-coding RNAs (lncRNAs)* (90).

Utilizando la secuencia de este PQ Chr2:158392504, se averiguo si este TQ podría estar generando proteínas aberrantes. Al ser un *lncRNA*, este no debería codificar a aminoácidos (91), pero al ser un sitio final de empalme rio abajo de un gen codificante este podría estar generando algunas de estas proteínas.

Utilizando herramientas computacionales se pudo ver si SNHG11 codificaba para algún dominio proteico, sabiendo que este gen es un *lncRNA*. Utilizando los resultados obtenidos por ARRIBA y STAR-Fusion, se analizó cuál eran los TQ que más se repetían en cada resultado, y que tuvieran a *SNHG11* como PQ aceptor, esto se puede observar en la **Tabla 8 y 9**.

Tabla 8. Dominios proteicos identificados de los resultados de STAR-Fusion. Los primeros 4 TQ con mayor ocurrencia que tengan el PQ Chr2: 158392504 del lado aceptor, con mayor ocurrencia obtenidos de los resultados de STAR-Fusion, para los cuales se tradujo su secuencia nucleotídica a aminoacídica y se obtuvo el dominio proteico al cual esta secuencia traduce.

Genes	Dominio proteico	Repeticiones del dominio
Vps13c--SNHG11	low complexity	2
Srsf11--SNHG11	low complexity	2
Slc23a2--SNHG11	low complexity	2
Spnb3--SNHG11	low complexity	4

Tabla 9. Dominios proteicos identificados de los resultados de ARRIBA. Los primeros 4 TQ con mayor ocurrencia que tengan el PQ Chr2: 158392504 del lado aceptor, con mayor ocurrencia obtenidos de los resultados de ARRIBA, para los cuales se tradujo su secuencia nucleotídica a aminoacídica y se obtuvo el dominio proteico al cual esta secuencia traduce.

Genes	Dominio proteico	Repeticiones del dominio
Vps13c--SNHG11	low complexity	2
Srsf11--SNHG11	low complexity	2
Sptbn2--SNHG11	low complexity	4
Rims1--SNHG11	low complexity	2

En las **Tablas 8 y 9**, se puede observar que utilizando los resultados obtenidos por ambos programas se obtuvieron resultados muy similares, aunque los genes de los TQ que estuviesen más repetidos no fueron los mismos. En todos los TQ que se analizaron, se obtuvieron TQ que codificaban para un dominio proteico llamado “*Low complexity*” (LCRs), en la mayoría de los casos esta secuencia codificada para 2 o 4 de estos dominios proteicos.

Los LCRs también se definen como dominios de composición sesgada, son regiones aminoacídicas que difieren de la composición y complejidad en la mayoría de las proteínas que normalmente se asocia con la estructura globular (92).

Estos resultados nos demuestran que parte del gen *SNHG11*, podría estar codificando para el dominio proteico “*low complexity*”, lo cual estaría traduciendo a proteínas aberrantes y esto puede provocar una acumulación de proteínas intra o extracelular mal plegadas, esto se conoce como agregación de proteínas, el cual es un fenómeno en el que existe un desorden en las proteínas o agregados de proteínas mal plegadas. Se sabe que puede llegar a generar diferentes fenómenos y enfermedades, como el Alzheimer, Parkinson, etc (93), por lo que si en casos de RTT, existe la agregación de proteínas por este TQ, esta agregación de proteínas podría estar provocando alguna posible sintomatología dentro del RTT.

4.5 Similitudes entre los resultados obtenidos de STAR-Fusion y ARRIBA

Luego se Utilizaron los PQ obtenidos de los resultados de STAR-Fusion y ARRIBA para observar cuantos PQ dadores y aceptores se repetían en cada programa.

Se generó un diagrama de Venn (**Figura 10**), para poder observar cuantos PQ se repetían en 1 programa o en ambos.

En el caso de los resultados obtenidos por STAR-Fusion (**Figura 10 A**), los PQ del lado dador fueron un total 441 y del lado aceptor fueron 363. En el caso de ARRIBA (**Figura 10 B**) los PQ obtenidos del lado dador fueron 925 y 937 del lado aceptor.

En el caso de los PQ que se repitieron en ambos programas fueron 141 los PQ del lado dador y 96 PQ del lado aceptor,

Estos resultados nos demuestran la poca similitud que hay entre los resultados que nos dieron ambos programas, ya que entre ambos existen muy pocos PQ del lado dador y aceptor que se repiten.

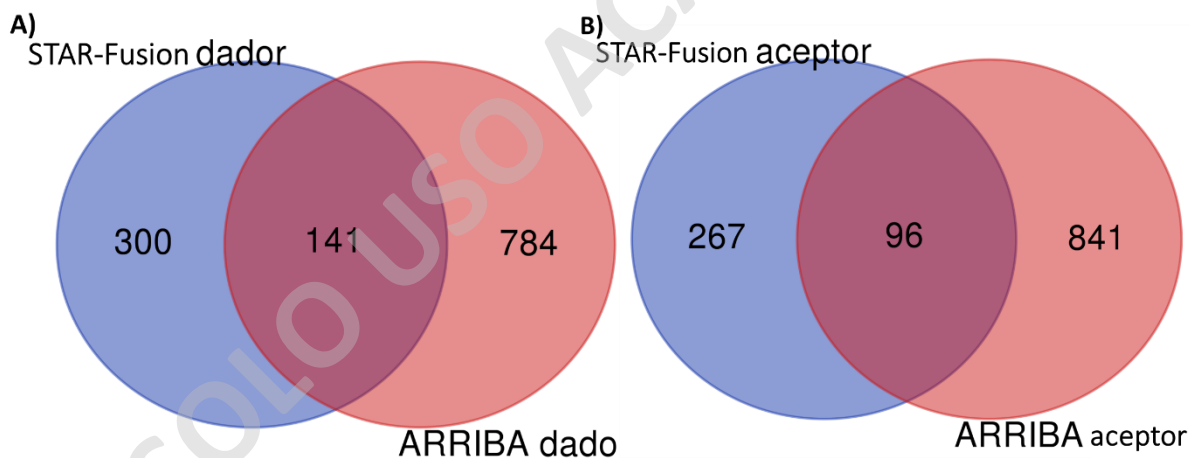


Figura 10. Diagrama de Venn de los puntos de quiebres obtenidos de los resultados de STAR-Fusion y ARRIBA (A y B). En la **Figura 10 A** se observa las cantidades de PQ del lado dador que comparten los resultados obtenidos por STAR-Fusion y ARRIBA, los PQ del lado dador obtenidos de los resultados de STAR-Fusion son 441, mientras que los obtenidos del programa ARRIBA son 895 y comparten 141 PQ. En la **Figura B** se observan las cantidades de PQ del lado aceptor que comparten los resultados obtenidos por STAR-Fusion y ARRIBA, los PQ del lado aceptor obtenidos de los resultados de STAR-Fusion son 363, mientras que los PQ del lado aceptor obtenidos de los resultados de ARRIBA son 937 y ambos comparten 96 PQ del lado aceptor.

4.6 Clasificación de puntos de quiebres de los transcritos quiméricos

Se separaron los PQ en 3 grupos distintos, PQ esporádicos, medios y recurrentes, estos se explicaron anteriormente en metodología. Con estos grupos se generaron 3 gráficos distintos para el lado dador y el lado aceptor, por cada programa (**Figura 11**). En los primeros gráficos de los PQ del lado dador que se generaron con STAR-Fusion, se aprecia que, para los 3 grupos, las muestras *MECP2-KO* tienen mayor ocurrencia de TQ en comparación con las muestras silvestre, siendo significativa la diferencia en esporádicos ($P < 0,0001$) y recurrentes ($P < 0,05$), no se aprecian diferencias significativas para la ocurrencia de sitios de empalme medios (**Figura 11 A, B y C**).

En el caso de los PQ del lado aceptor generados por el mismo programa se puede observar algo similar que con las imágenes anteriores, ya que igualmente los 3 casos se ven una mayor cantidad de repeticiones de los PQ en cada una, el análisis estadístico indicó diferencias significativas en esporádicos ($P < 0,0001$) y en recurrentes ($P < 0,05$), pero no así en los PQ medios (**Figura 11 D, E y F**).

En la **Figura 12** se muestran los resultados de ARRIBA (73), se puede observar que no existe alguna diferencia entre los datos, incluso llega a ser muy similar la cantidad de repeticiones que existen entre las diferentes muestras, por lo que se realizó una prueba de Wilcoxon Mann Whitney dentro de los 3 gráficos. En el caso de los PQ esporádicos el *P value* fue de 0,363, en el caso de los PQ medios el *P value* fue de 0,1347 y finalmente en el caso de los PQ recurrentes el *P value* fue de 0,6419 (**Figura 12 A, B y C**), lo que no hubo una diferencia significativa para ningún PQ del lado dador. En el caso de los datos del lado aceptor (**Figura 12 D, E y F**) se puede observar que solamente los PQ esporádicos tienen una diferencia significativa entre las muestras de *MECP2-KO* y silvestre, pero muestra un resultado opuesto a lo esperado y reportado con STAR-Fusion, ya que en los resultados de STAR-Fusion no demuestran que hay una mayor cantidad de TQ en las muestras *MECP2-KO* y además existe una diferencia significativa entre las muestras silvestres y las muestras *MECP2-KO*. En el caso de los PQ medios y recurrentes no se aprecian diferencias significativas. En los análisis estadísticos los PQ esporádicos el *P value* fue de 0,02424, en el caso de los PQ medios este fue de 0,6523 y los PQ recurrentes el *P value* fue de 0,4926, dando como resultado que solamente los PQ esporádicos del lado aceptor tienen una diferencia significativa.

Los resultados obtenidos hasta el momento nos podrían indicar que un aumento de los PQ esporádicos dentro de las muestras *MECP2-KO*, lo que provocaría desregulación en L1, lo cual aumentaría los saltos de este retrotransposón dentro del genoma provocando un aumento de TQ, lo cual generaría mal funcionamiento de genes y posibles proteínas “aberrantes”, esto provocaría el aumento de la variabilidad fenotípica de RTT.

SOLO USO ACADÉMICO

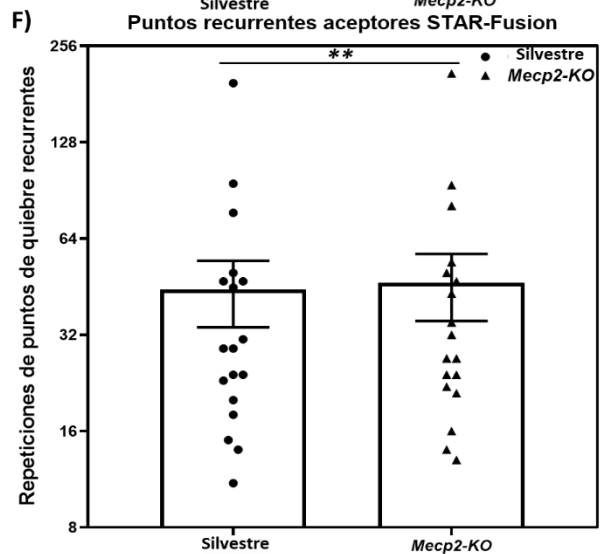
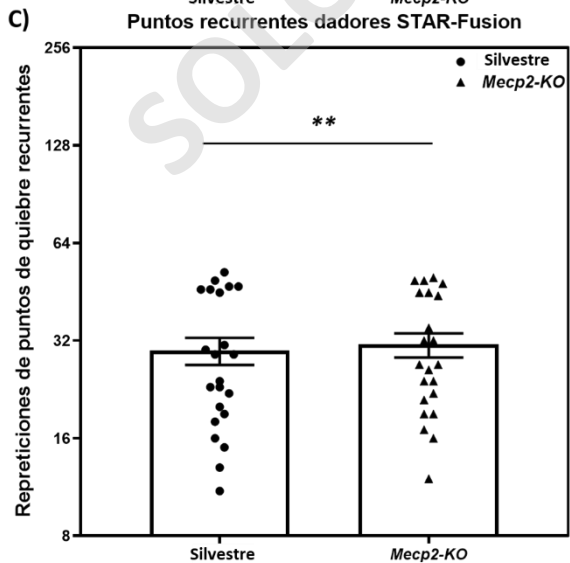
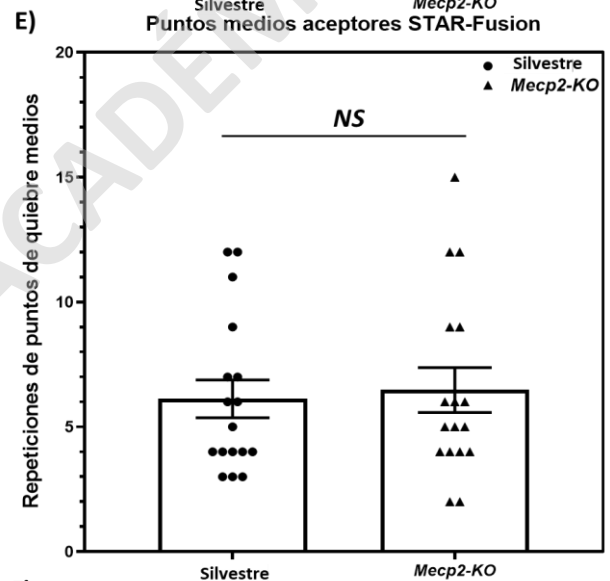
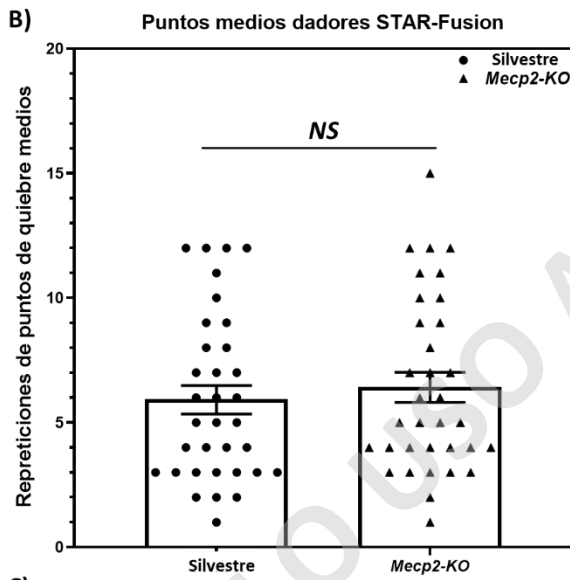
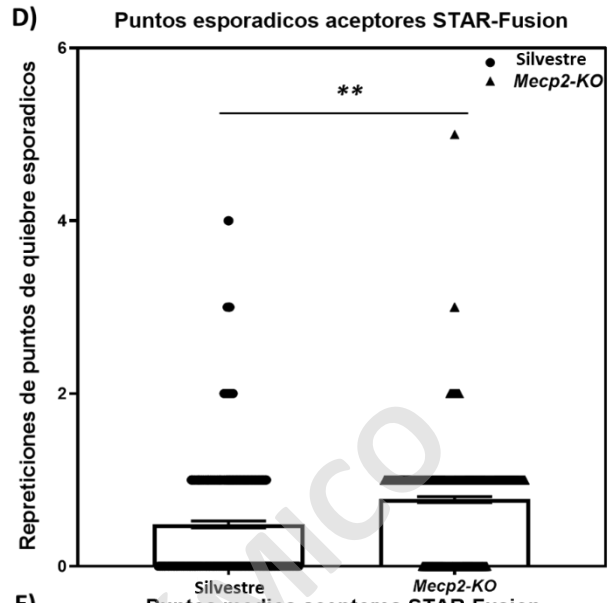
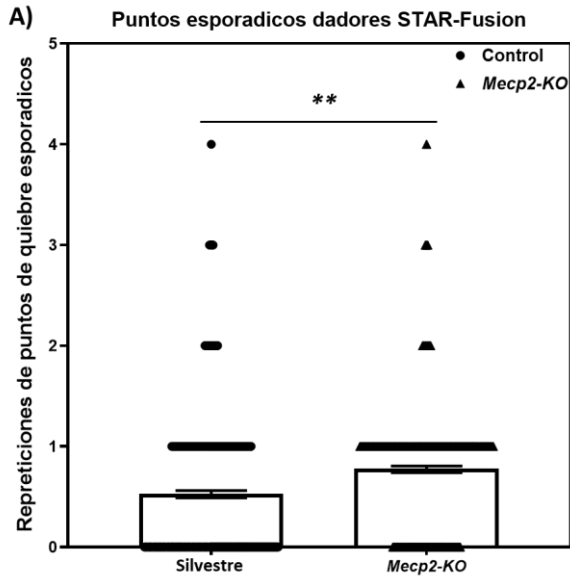


Figura 11. Caracterización de los puntos de quiebres en transcritos quiméricos identificados en reportes de STAR-Fusion (A - F). Se clasificaron los PQ /sitios de empalme en 3 grupos dependiendo de la cantidad de ocurrencias en el total de muestras, estos grupos fueron esporádicos, medios y recurrentes. Los gráficos generados con los grupos esporádicos, medio y recurrentes del lado dador en base de transcritos quiméricos generados por STAR-Fusion, se generaron los análisis estadísticos de los tres grupos mediante una prueba de Wilcoxon Mann Whitney los cuales del grupo esporádico dio un P value <0.0001 , grupo medio fue de no significativo ($P = 0,3669$) y los recurrentes $p < 0,05$ ($P = 0,0452$.) **(A, B y C)**. Los gráficos generados con los grupos esporádicos, medio y recurrentes del lado aceptor en base de transcritos quiméricos generados por STAR-Fusion, se generaron los análisis estadísticos de los tres grupos mediante una prueba de Wilcoxon Mann Whitney los cuales del grupo esporádico dio un P value de $<0,0001$, grupo medio fue de no significativo ($P = 0,5469$) y los recurrentes $p < 0,05$ ($P = 0,038$.) **(D, E y F)**. Los gráficos con NS, es que el P value es no significativo ($P > 0,05$), y los con ** tienen una diferencia significativa ($P < 0,05$).

SOLO USO ACADÉMICO

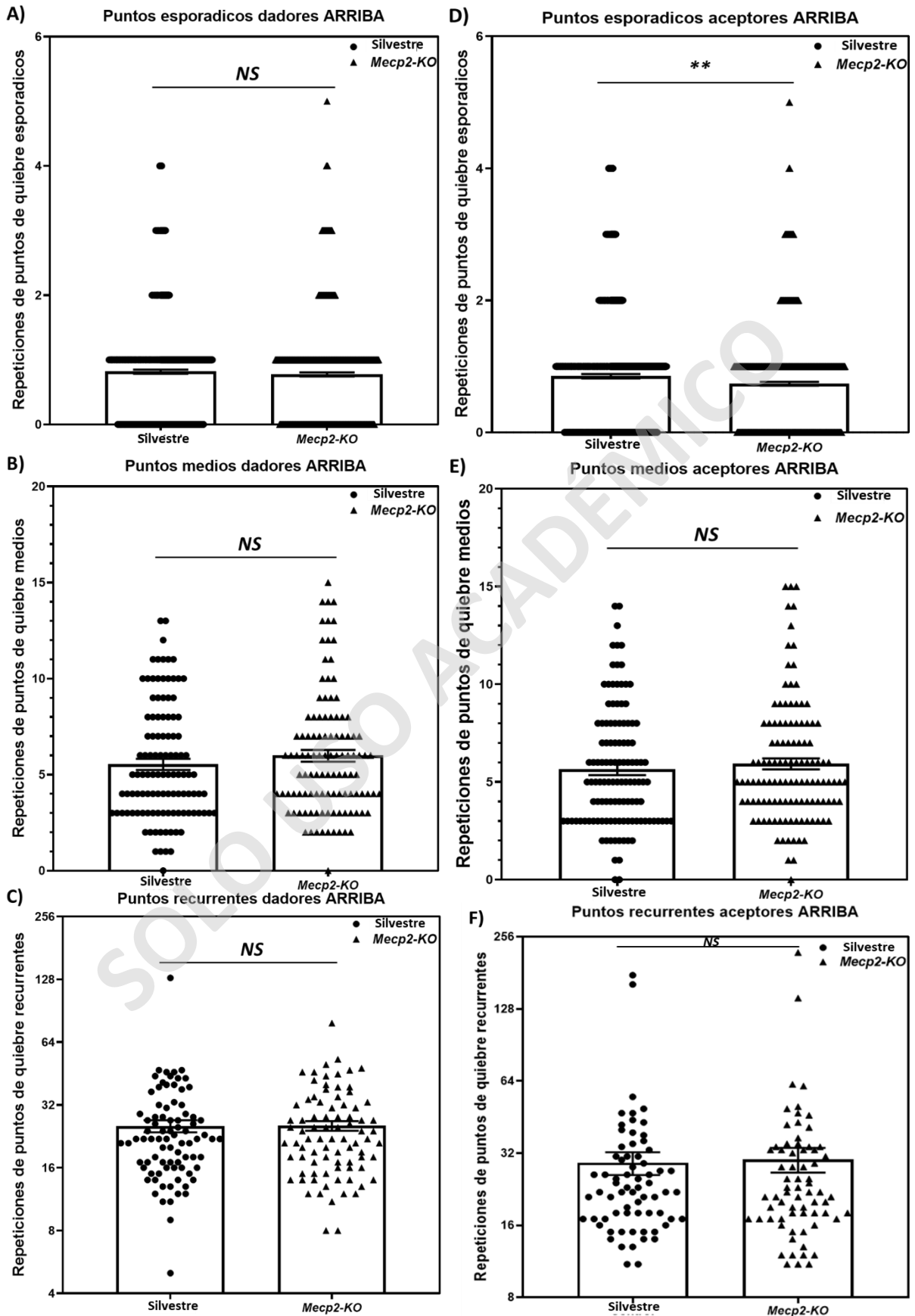


Figura 12. Caracterización de puntos de quiebres en transcritos quiméricos identificados en reportes de ARRIBA (A - F). (A, B y C) Gráficos generados con los grupos esporádicos, medio y recurrentes del lado dador en base a los candidatos de transcritos quiméricos generados por ARRIBA. Se generaron los análisis estadísticos de los tres grupos mediante una prueba de *Wilcoxon Mann Whitney*. El grupo esporádico dio un *P value* de 0,3899, grupo medio fue de 0,1445 y los recurrentes dio 0,5912. (D, E y F). Gráficos generados con los grupos esporádicos, medio y recurrentes del lado aceptor en base a los candidatos de transcritos quiméricos generados por ARRIBA, se generaron los análisis estadísticos de los tres grupos mediante una prueba de *Wilcoxon Mann Whitney* los cuales del grupo esporádico dio un *P value* de 0,0452, grupo medio fue de 0,8531 y los recurrentes dio 0,3565. NS es *P value* no significativo ($P > 0,05$), y los con ** representan una diferencia significativa ($P < 0,05$).

SOLO USO ACADÉMICO

4.7 Cuantificación de puntos de quiebre según tejido

Finalmente, con identificación de los PQ de las muestras, se prosiguió a separarlas por tejido. Se analizaron tejidos de giro dentado, corteza cerebral, cerebelo, cultivo cortical, hipocampo y microglía junto a su control correspondiente. En cada caso se cuantificó la ocurrencia de los PQ identificados tanto en el lado aceptor como en el lado dador. Se utilizaron los datos de STAR-Fusion para generar las siguientes tablas. Los datos utilizados de los PQ del lado dador se pueden observar en el Anexo 9 y los del lado aceptor en el Anexo 10 (**Figura 13**).

Primero se analizaron las 4 muestras de cerebelo, en las cuales se identificaron 71 PQ en el lado dador y 60 en el lado aceptor del total TQ que se identificaron en todas las muestras. Con estos datos se graficaron las **Figuras 13 A y B**, en ambos casos se vio que existe una diferencia significativa entre las muestras *MECP2-KO* y silvestre, obteniendo un *P value* de 0,0118 y 0,01 respectivamente.

En las 4 muestras de corteza cerebral se identificaron 77 PQ en el lado dador y 53 en el lado aceptor de las muestras de corteza, las cuales se graficaron y se pueden ver en las **Figuras 13 C y D**, en el análisis estadístico de cada grupo, no existía una diferencia significativa, teniendo un *P value* de 0,4957 y 0,3771.

Por otra parte, las 9 muestras de cultivo cortical se identificaron 117 PQ en el lado dador y 95 del lado aceptor, los datos se graficaron en las **Figuras 13 E y F**, en el análisis estadístico el sitio dador dio un *P value* de 0,0349 teniendo una diferencia significativa y el sitio aceptor un *P value* de 0,1317 siendo su diferencia no significativa.

En el caso de las 24 muestras de la región del giro dentado, se identificaron 117 PQ en el lado dador y 57 en el lado aceptor, se pueden observar en la **Figura 13 G y H**, en este caso el *P value* obtenido fue de 0,3146 en el sitio dador y 0,3377 en el sitio aceptor, por lo que ambos sitios no poseen una diferencia significativa.

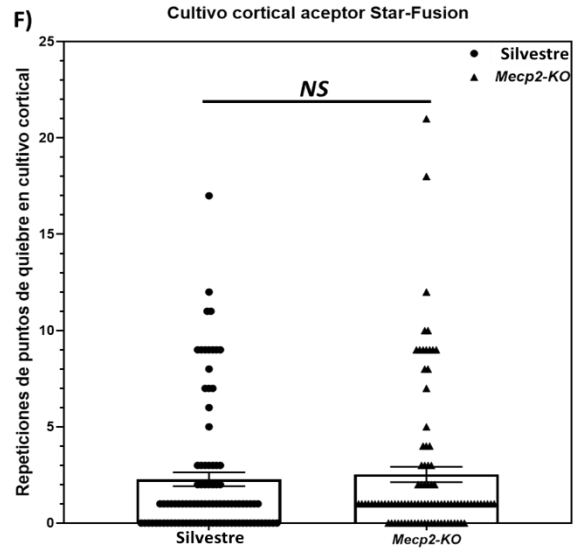
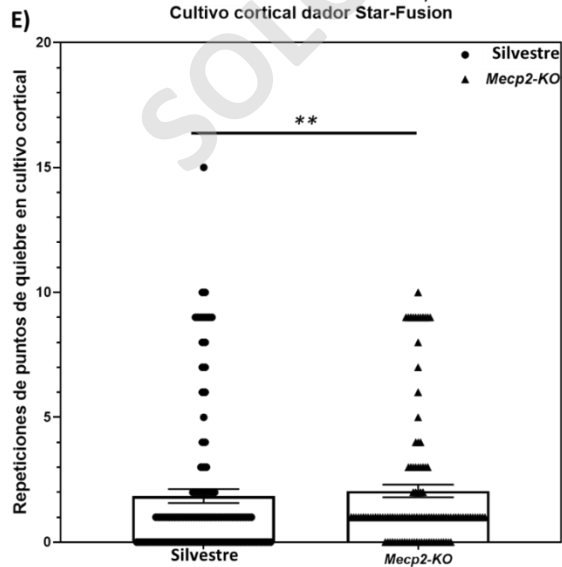
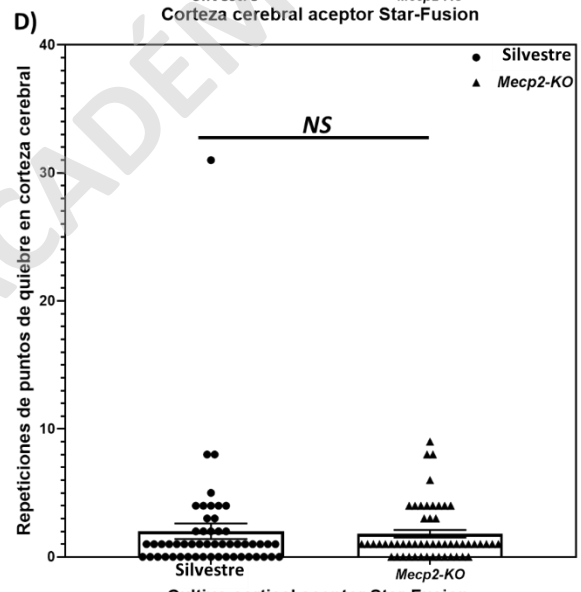
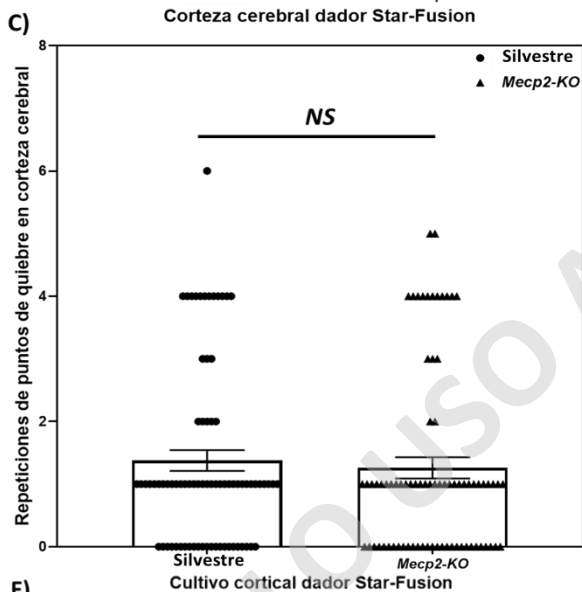
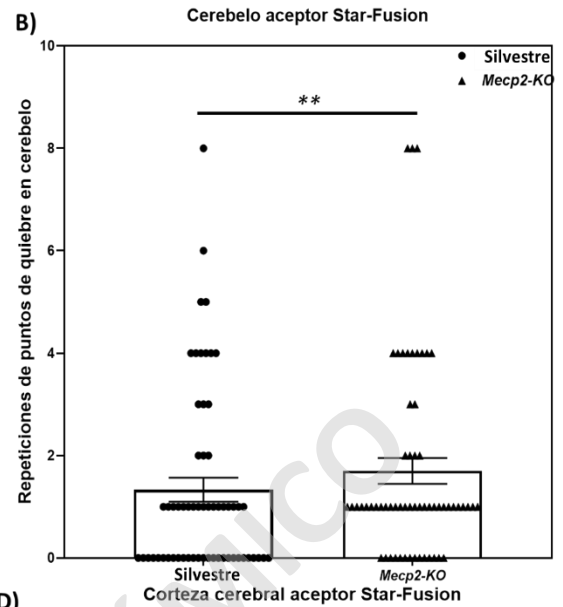
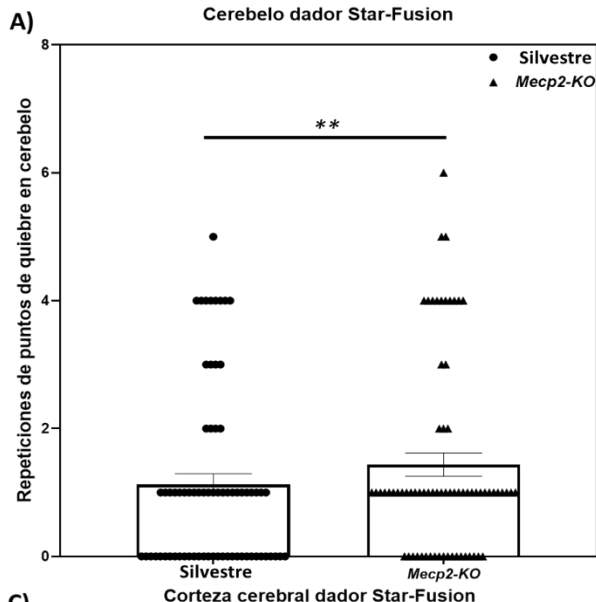
Luego se analizaron las 3 muestras de microglía, las cuales fueron 172 PQ en el sitio dador y 172 en el sitio aceptor, estos gráficos se pueden ver en la **Figura 12 I y J**, donde el *P value* en ambos casos fue de $P < 0,0001$, teniendo ambos una diferencia significativa.

Para concluir se analizaron las 6 muestras de hipocampo, la cual obtuvo 68 PQ en el lado dador y 46 en el lado aceptor, esto se puede ver en las **Figura 13 K y L**, en este caso el *P value* obtenido por el análisis estadístico fue de 0,1566 y 0,2425 respectivamente, lo cual indica que no existe una diferencia significativa.

En las muestras de cerebelo se puede observar que existe una diferencia significativa entre los PQ de obtenidos de las muestras *MECP2-KO* y silvestre, esto podría indicar que en esta zona del cerebro se generan una mayor cantidad de TQ en el RTT lo cual podría contribuir a las alteraciones en los movimientos coordinados y finos del sistema muscular esquelético, lo que estaría relacionado con la pérdida parcial o completa de los movimientos propositivos de las manos, la habilidad para hablar, los movimientos estereotipados de las manos y el desarrollo psicomotor.

En el caso de las muestras de microglía, también existe una diferencia significativa entre los PQ, de ambos lados, entre las muestras *MECP2-KO* y muestras silvestres, aunque en las muestras de microglía eran solamente 3 estos resultados pueden ser poco concluyentes.

SOLO USO ACADÉMICO



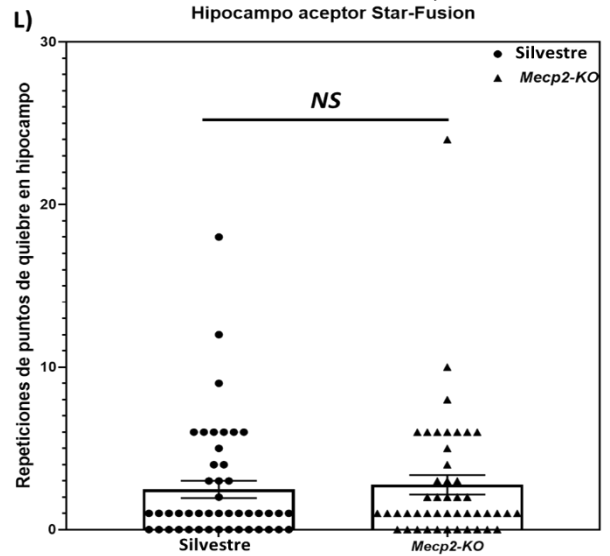
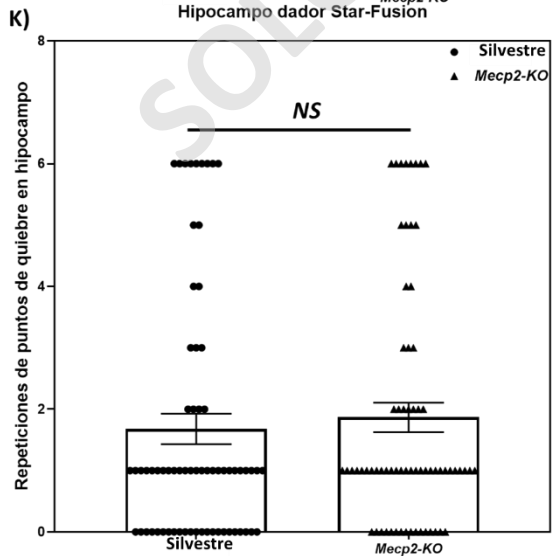
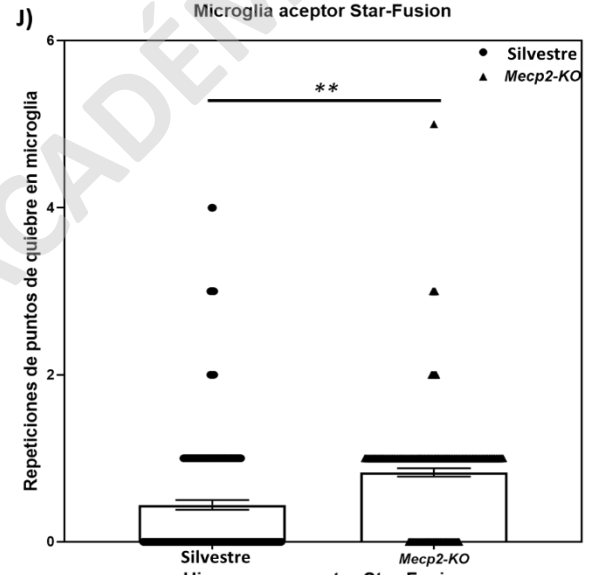
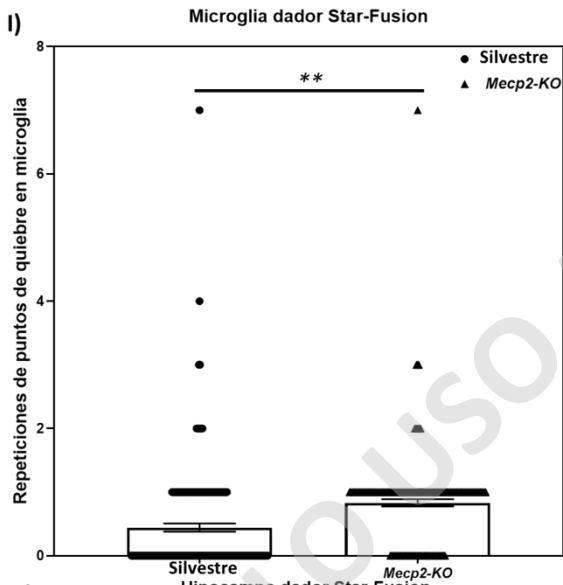
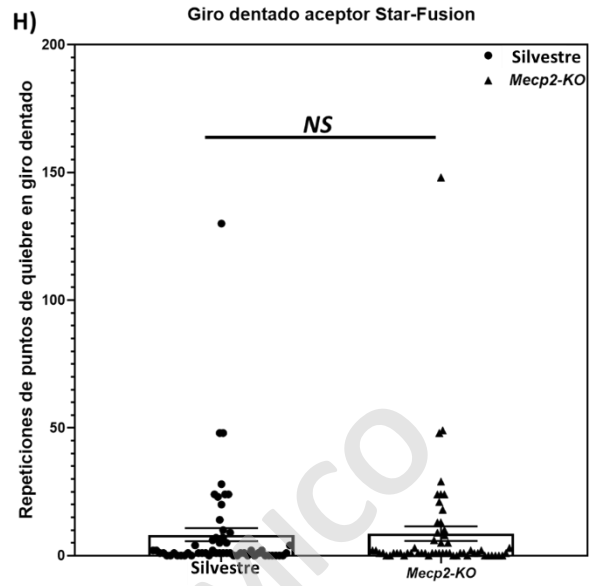
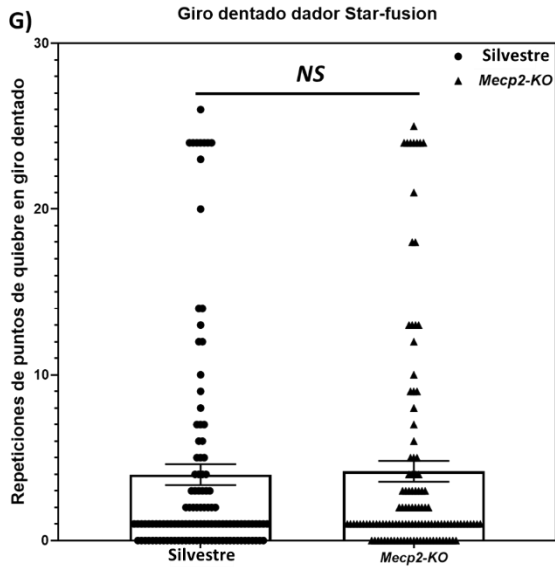


Figura 13. Cuantificación e identificación de los puntos de quiebres existentes en los diferentes tejidos analizados (A - L). Se graficaron las ocurrencias de los PQ de los candidatos de transcritos quiméricos obtenidos por el programa STAR-Fusion. Se generaron 2 gráficos por cada tejido, los cuales son cerebelo, corteza cerebral, cultivo cortical, giro dentado, microglía e hipocampo. Los gráficos se generaron en base a el sitio dador y aceptor de cada uno. Los puntos circulares representan las ocurrencias de muestras silvestre y los triángulos de las muestras *MECP2-KO*. **(A y B)** Representan los gráficos generados en base a los datos recopilados de las muestras de cerebelo, se generaron los análisis estadísticos de ambos gráficos mediante una prueba de *Wilcoxon Mann Whitney* los cuales ambos tenían una diferencia significativa ($P = 0,0118$ y $P = 0,0002$). **(C y D)** Representan los gráficos generados en base a los datos recopilados de las muestras de corteza cerebral, el análisis estadístico se hizo en base a una prueba de *Wilcoxon Mann Whitney* el cual ambos demostraron que no existe una diferencia significativa, obteniendo un *P value* de 0,4957 y 0,3771. **(E y F)** Representan los gráficos generados en base a los datos recopilados de las muestras de cultivo cortical, los análisis estadísticos mediante una prueba de *Wilcoxon Mann Whitney* el cual en el lado aceptor dio un *P value* de 0,0349 y el sitio dador 0,0735. **(G y H)** Estos gráficos se generaron en base a los datos obtenidos de las muestras de giro dentado, el análisis estadístico hecho mediante una prueba de *Wilcoxon Mann Whitney*, el cual demostró que ninguno de los grupos de datos tiene una diferencia significativa dando un *P value* de 0,3146 y 0,3377. **(I y J)** Se generaron los gráficos en base a los datos de las muestras de microglía, mediante una prueba de *Wilcoxon Mann Whitney* se hizo el análisis estadístico en el cual dio un *P value* $P < 0,0001$ en ambos gráficos, lo que demuestra que existe una diferencia significativa. Y en **(K y L)** se generaron los gráficos en base a los datos de hipocampo, mediante una prueba de *Wilcoxon Mann Whitney* se hizo el análisis estadístico en el cual ambos resultados demostró que no existe una diferencia significativa, dando como resultado un *P value* de 0,1566 y 0,2425, respectivamente.

SOLO USO ACADÉMICO

5 DISCUSIÓN

El RTT es un síndrome del neurodesarrollo que afecta severamente habilidades motoras, cognitivas y comunicacionales, las cuales son encabezadas por microcefalia, demora en la generación de nuevas habilidades, ausencia del habla, aparición de características autistas, pérdida de habilidades de manipulación intencionadas que son reemplazadas con movimientos estereotipados de manos y otras anormalidades musculares (94). Como se mencionó anteriormente la principal causa de este síndrome es la deficiencia de *MECP2* (26), para el estudio de RTT el modelo más utilizado es el ratón (*Mus musculus*) debido a que es similar en casos clínicos de humanos (95). Por lo mismo, en esta tesis utilizaron muestras de RTT en *Mus musculus*, debido a que existe una gran cantidad de datos de libre acceso en secuenciaciones de diferentes tejidos cerebrales.

Durante este trabajo de tesis, se analizó la existencia de nuevos y posibles efectos en el genoma en casos de RTT, esto podría ayudarnos a comprender mejor las sintomatologías de este síndrome y ayudarnos a entender de mejor forma la participación del gen *MECP2* en casos de RTT.

5.1 Diferencia entre los programas utilizados

En esta tesis se utilizaron 2 programas para la detección de transcritos de fusión o quiméricos mediante el mapeo de lecturas y en el ensamblaje de transcritos de fusión. Como se mencionó anteriormente los programas que se utilizaron fueron STAR-Fusion y ARRIBA. Se seleccionaron ambos programas debido a un análisis comparativo en el que se evaluaron 23 programas para la detección de transcritos quiméricos (96). Según el estudio que se hizo de los programas utilizados, se seleccionaron los 2 más precisos, eficaces y sensibles, además de tener una combinación entre velocidad y mayor precisión en comparación con los otros 21 programas.

Para poder utilizar STAR-Fusion se requiere usar componentes de CTAT (78), puesto que STAR-Fusion es una herramienta desarrollada como parte de CTAT. STAR-Fusion se incluye en CTAT con un enfoque de mapeo de lectura del genoma de referencia altamente eficiente, y Trinity Fusion para aprovechar el ensamblaje del transcriptoma *de novo* para la detección de fusiones (78). Por esto STAR-Fusion utiliza la base de datos de CTAT para generar la librería del genoma de referencia, en este caso GRCm38 o mm10 y la base de datos de CTAT, por lo que al momento de utilizar este programa es de vital importancia utilizar el comando "--no_annotation_filter", ya que de esa manera se excluye cualquier filtro basado en transcritos quiméricos ya anotados en la base de datos de CTAT.

En el caso de ARRIBA, este no requiere una base de datos, debido a que ARRIBA no es más que una colección de filtros, ya que la generación de candidatos a transcritos quiméricos es manejada casi del todo por STAR (77), debido a que este recolecta toda la evidencia sobre potenciales genes “fusionados”. Esta evidencia se guarda en un archivo de alineamientos quiméricos generado por STAR, y este archivo es la base que utiliza ARRIBA para generar los reportes finales.

Aunque ambos programas presentan cierto nivel de coincidencia, los resultados obtenidos son bastante diferentes, tanto en el número de transcritos quiméricos en muestras *MECP2-KO* y silvestre. En la **Figura 7 A** se ve que los TQ candidatos obtenidos por STAR-Fusion no superan los 40 candidatos por muestra. En cambio, los reportes de ARRIBA, se ve que la mayoría de las muestras superan los 100 candidatos por reporte (**Figura 7 B**), lo que es más del doble que el otro programa. Además, en la **Figura 10** se muestra que hay una gran diferencia entre los PQ que son detectados entre STAR-Fusion y ARRIBA. En el caso de STAR-Fusion solo el 31% de los PQ del lado dador se repetían en ARRIBA, y en ARRIBA apenas un 15% de estos PQ estaban en STAR-Fusion.

En el caso de los PQ del lado aceptor obtenidos de los resultados de STAR-Fusion, solo se compartieron el 26% de los PQ con ARRIBA, y en caso de ARRIBA solo el 10% de los PQ del lado aceptor se repetían en STAR-fusion. Estos resultados muestran una importante diferencia entre los programas utilizados.

Creemos que la diferencia entre las dos herramientas usadas sucede, debido a que STAR-Fusion y ARRIBA utilizan distintas metodologías antes del alineamiento y posterior a este. En el caso de STAR-Fusion, este utiliza el programa CTAT para generar una librería, el cual aprovecha para preparar el genoma objetivo y un conjunto de anotaciones utilizando las herramientas de CTAT. Este programa genera un archivo nombrado “ctat_genome_lib_build_dir”, este archivo contiene una variedad de anotaciones del genoma para la detección de TQ. Gracias a las anotaciones que CTAT utiliza, STAR-Fusion disminuye la imprecisión y utiliza estas anotaciones para poder generar el resultado.

En el caso de ARRIBA, este programa posee tiempos muy cortos y una alta sensibilidad como criterio de análisis (74). La alta sensibilidad a la detección de TQ, provocaría un aumento de candidatos de TQ dentro de los reportes obtenidos. ARRIBA se define como una colección de filtros, ya que la generación de candidatos de TQ está totalmente gestionado por STAR, ya que en el alineamiento recoge todas las pruebas para definir sobre los posibles TQ dentro de las secuencias (74), sugiriendo que ARRIBA tiene

un filtro más permisivo en su algoritmo, lo cual genera una mayor cantidad de candidatos en sus reportes. En cambio, STAR-Fusion filtra de manera más estricta los transcritos intragénicos y la distancia existente entre sus candidatos. Sabiendo esto se considera un tercer programa para zanjar las diferencias encontradas.

5.2 ARN circulares

En los reportes de ambos programas no sólo se observa coordenadas del punto de quiebre dador y aceptor, también indica la orientación del mensajero en el cromosoma. Los reportes presentan el sentido de la secuencia en el transcrito, esto lo representaba con un “+” cuando la secuencia del gen iba de 5’ a 3’ y con un “-” cuando la secuencia del gen iba de 3’ a 5’, por lo que en muchos casos se identificaron transcritos quiméricos en que la coordenada o PQ dador era + y el punto de quiebre aceptor era - dentro de las proximidades en un mismo cromosoma. Esto podría indicar la posible aparición de ARN circulares (*circRNAs*), dentro de los reportes de ambos programas, debido a que estos son muy difíciles de distinguir entre duplicaciones genómicas y *circARNs* sin datos de la secuenciación del genoma completo.

Los *circRNAs* son una clase especial de ARN que son producidos por un enlace covalente entre los extremos 5’ y 3’ de una molécula de ARN (97), y estos ARN pertenecen a la categoría de moléculas de ARN no codificante (*ncARN*) (98). Se ha visto que los *circRNAs* se expresan de manera tejido-específica en etapas de desarrollo, tanto en humanos, ratones y moscas (99–101).

En el cerebro humano, se estima que el 20% de los genes producen *circRNAs*, mientras que en el corazón aproximadamente el 9% de los genes expresados producen *circRNAs* (102). Debido a esto, es posible que una parte importante de nuestros resultados correspondan a *circRNAs*.

Recientemente se ha publicado que existe un aumento de la regulación de *circRNAs* durante la maduración y diferenciación neuronal (103), además se ha visto que individuos con RTT han reducido su complejidad dendrítica y densidad en la médula espinal (104). En el experimento de Siqueira et al 2021 (105), se vio que en la existencia de la eliminación de *MECP2*, existe una baja en la expresión de especies de *circRNAs* en el hipocampo de ratones (105). Sabiendo esto, los efectos causados por *MECP2-KO* en los *circRNAs* podría ser crucial para identificar éstos como posibles intermediarios en los defectos del neurodesarrollo en RTT.

5.3 Dependencia de *MECP2*

Teniendo en cuenta los resultados obtenidos, se analizaron los candidatos a transcritos quiméricos que se obtuvieron en los reportes de STAR-Fusion y ARRIBA. Ambos reportes dan los PQ de 3' o sitio dador y el de 5' o sitio aceptor. Como se puede ver en la **Figura 8**, tanto los PQ informados por STAR-Fusion y ARRIBA tienen una mayor ocurrencia en los casos de *MECP2-KO* en comparación con los PQ de las muestras silvestres. Pero como se puede ver en los resultados de los análisis estadísticos se aprecia que los PQ que se obtuvieron con los reportes de STAR-Fusion dieron con una diferencia significativa. Además, se puede observar que en la línea de tendencia esta tiene una curvatura muy clara en la zona inferior izquierda del gráfico de dispersión (**Figura 8 A-B**), indicando que hay una mayor cantidad de repeticiones de PQ en las muestras *MECP2-KO* en comparación con las muestras silvestres.

Como se mencionó anteriormente, una de las funciones que tiene *MECP2* es la represión transcripcional (35, 36) y la activación transcripcional (37), además de la función de regular *L1* (47). Actualmente se sabe que *MECP2* tiene una relación física con *L1*, y que *MECP2* es parte del complejo proteico que provoca la inhibición de *L1*, este es uno de los retrotransposones más abundantes y eficaces dentro del ADN, representa alrededor del 17% del genoma humano (106,107).

Al tener esta información en cuenta, y también viendo los resultados obtenidos por STAR-Fusion, se podría deducir que el aumento de los transcritos quiméricos son debido a la supresión de *MECP2*, esto evitaría que *L1* se inhiba, provocando un aumento en la expresión y abundancia de este retro transposón dentro del ADN, y al mismo tiempo generando un aumento de quiebres de la doble hebra del ADN. Para comprobar esto se podría analizar las anotaciones de las ubicaciones de los elementos *LINES* conocidos y compararlos con los "Hotspot" obtenidos durante esta investigación.

El aumento de la expresión de *L1* provocaría un aumento de eventos esporádicos en las muestras de *MECP2-KO*, que se pudieron observar en la **Figuras 8 A-B**. En ambas imágenes se puede observar que la línea de tendencia tiene una curvatura, mostrando que hay una mayor cantidad de ocurrencia de PQ esporádicos en las muestras *MECP2-KO*. Además, los PQ esporádicos se separaron para poder identificar de mejor manera las coordenadas, como se pueden ver en la **Figura 11 A y D**, en este caso se evidencia una clara discrepancia entre las ocurrencias de las muestras *MECP2-KO* y silvestre, y se determina una diferencia estadística entre ambas muestras y en ambos gráficos.

Otro análisis complementario, que podría apoyar nuestras observaciones sobre TQ, sería su evaluación en modelos del síndrome de duplicación *MECP2*, el cual es un síndrome que es provocado por la duplicación de *MECP2* (108). En estos casos se debería ver un aumento en la expresión de *MECP2* y esto provocaría una disminución en la expresión de *L1*. Se podría utilizar la misma metodología propuesta durante esta tesis y observar si existe una disminución de los TQ en modelos de este síndrome y así compararlos con los resultados que se obtuvieron en este trabajo.

5.4 *SNHG11*

En las **Tablas 6 y 7** se puede ver que el punto con mayores ocurrencias tanto como en STAR-Fusion y ARRIBA fue Chr2: 158392504 *Small nuclear RNA host gene 11 (SNHG11)*, uno de los miembros de la familia *SNHG*, que ha sido reconocida como *long non-coding RNAs (lncRNAs)* (90). Recientemente se ha publicado que *SNHG11* tiene relación con varios tipos de cáncer como el pancreático, gástrico, de próstata, y de mamas (90,109–111), sin embargo, aún no se conoce la función de este gen.

Se ha demostrado que varios *lncRNAs* que participan en la regulación de la expresión de genes pueden jugar un rol crítico en varias funciones biológicas y en enfermedades (112).

Se conoce actualmente que *SNHG11* está relacionado con la vía Wnt/ β -catenina, provocando la ubiquitinación de glucógeno sintetasa quinasa-3 β (*GSK-3 β*) (113), una enzima que interviene en el metabolismo energético el desarrollo de las células neuronales y la formación del patrón corporal (114). Esta vía es activada por la unión de la proteína WNT a el receptor frizzled y al co-receptor LRP, esto genera la supresión de la actividad de *GSK-3 β* , provocando la liberación de β -catenina en el citoplasma, lo que permitiría la asociación de β -catenina con la región genética Tcf/lef en el núcleo celular, lo que gatilla la transcripción de genes WNT provocando el aumento de la proliferación celular (115). En el caso de que WNT no se una al receptor frizzled, se activará el complejo de destrucción de la β -catenina (115), por lo que la proteína WNT y la vía Wnt/ β -catenina es importante para el desarrollo neuronal y el funcionamiento del cerebro(116,117).

Los genes WNT son una gran familia de factores de crecimiento, los cuales, son altamente conservados en la evolución, indicando que son similares en todas las especies (118). Se ha comprobado que en casos de *MECP2-KO* existe una disminución de WNT6, además una disminución de la fosforilación de *GSK-3 β* y un aumento en la fosforilación de β -catenina, pero dentro del mismo estudio se comprobó

que al administrar *WNT6* de manera lentiviral en la amígdala cerebral esta recupera el deterioro locomotor y el déficit del comportamiento generado por la deficiencia de *MECP2* (117).

Los resultados mostrados de la **Tabla 6 y 7**, sugiere que debido al silenciamiento de *MECP2*, durante el RTT, se generaría un aumento de TQ los cuales provocarían el quiebre de la doble hebra de ADN en la posición Chr2: 158392504, afectando directamente al gen *SNHG11*, lo cual provocaría represión en la ubiquitinación de GSK-3 β . Lo que gatillaría una disminución de la vía Wnt/ β -catenina, provocando déficit en el desarrollo neuronal y problemas en el funcionamiento del cerebro adulto, sin embargo, más estudios serían necesarios para comprobar esto.

5.5 Esporádicos, medios y recurrentes

Dentro de los PQ que se mencionaron anteriormente, se consideró la idea de separar estos en 3 grupos distintos, los cuales eran PQ esporádicos, medios y recurrentes. Esto se hizo por las grandes diferencias que había en la cantidad de repeticiones entre todas las coordenadas de los PQ que se obtuvieron, dentro de los 2 programas.

Las ocurrencias encontradas en las muestras silvestre fueron altas, debido a que estas se generan durante el desarrollo, ya que esta es una fuente importante de diversidad genética e innovación reguladora (119). Se conoce que las inserciones *L1* se reproducen a lo largo de la neurogénesis fetal y adulta, así como en las neuronas maduras (120,121). Aun así, se ha visto que no todas las neuronas muestran el mismo nivel de expresión de *L1*, lo que se vería reflejado por subtipos neuronales que aportarían una mayor cantidad de este retrotransposón, según las inserciones que habrán de *L1* existiría una diversidad fenotípica mayor a las que tienen menos inserciones (122).

Los puntos de quiebre esporádicos fueron los que más se mostraron en ambos programas, aunque en el caso de ARRIBA estos no tenían una diferencia significativa. Los antecedentes indicarían que los LTR durante el neurodesarrollo tienen una mayor expresión, lo cual provoca una mayor cantidad de saltos, provocando un aumento en el mosaicismo genético del cerebro. En el caso del RTT esto provocaría un aumento del mosaicismo lo que generaría un aumento de posibles transcritos quiméricos, lo que finalmente provocaría un aumento de TQ únicos.

En ambos programas los puntos medios no tuvieron ninguna diferencia significativa según el análisis estadístico realizado. Aunque la mayor diversidad de PQ se ubican en los puntos esporádicos y los puntos recurrentes.

En los PQ aceptores se encuentran una mayor cantidad de ocurrencia en comparación con los PQ dadores. Esto se puede observar un poco más claro en el diagrama de cuerdas de las muestras *MECP2-KO*, ya que anteriormente se mencionó que el punto de quiebre con mayores ocurrencias es Chr2: 158392504, por lo mismo es uno de los PQ que tiene mayor cantidad de ocurrencias en ambos programas. Esto podría ser un ejemplo claro que existe un aumento de los TQ dentro de las secuencias, y que este síndrome podría tener alguna relación directa con este gen, por lo que su estudio a profundidad en este síndrome presenta una incógnita a responder.

SOLO USO ACADÉMICO

6. Conclusiones

El Síndrome de Rett es uno de los síndromes más comunes dentro de los síndromes poco recurrentes, siendo la segunda mayor causa de déficit mental en niñas después del Síndrome de Down, teniendo 1 caso cada 10.000 mujeres (123). El entendimiento y la detección temprana de este síndrome es muy importante para poder prevenir posibles sintomatologías que puedan poner en riesgo la vida de las personas que sufran de esta condición. Por esta razón es de suma importancia identificar y entender las repercusiones moleculares existentes en estos pacientes o en modelos de la enfermedad.

- La pérdida de función de *MECP2* tienden a un mayor número de transcritos quiméricos por muestra, en comparación con las muestras silvestres, pero no presentan una diferencia significativa
- Hay una diferencia significativa en las ocurrencias de los puntos de quiebre entre las muestras *MECP2-KO* y silvestre, la cual es más evidente en los puntos con menor ocurrencia o esporádicos en los resultados de STAR-Fusion. A diferencia de este, ARRIBA también tiene una acumulación de ocurrencias en los puntos de quiebre dadores y aceptores esporádicos, pero estos no tienen una diferencia significativa en el análisis estadístico, posiblemente por el ruido de su resultado.
- En el Diagrama de cuerdas se aprecia una mayor diversidad de puntos de quiebre en las muestras *MECP2-KO*, lo que deriva en una mayor cantidad de eventos intra y extracromosomales en las muestras *MECP2-KO* en comparación con las muestras silvestre.
- Existen diferencias significativas en la ocurrencia de los sitios dadores y aceptores esporádicos y recurrentes, en los resultados obtenidos por STAR-Fusion, esto nos podría indicar que la existencia de *MECP2-KO*, provocaría desregulación de *L1*, lo que generaría un aumento de TQ dentro del genoma, lo cual sería un aumento de la variabilidad fenotípica de RTT.

7. IMPLICANCIAS, RECOMENDACIONES Y/O PROYECCIONES FUTURAS

En este trabajo se investigó sobre la detección de transcritos quiméricos o genes fusión en modelos de síndrome de Rett. Estos transcritos se han investigado durante mucho tiempo, y gracias a las nuevas tecnologías y el entendimiento de estos transcritos, se ha podido mejorar los programas para la detección de estos. La mayoría de los programas para la detección de TQ son destinados para utilizarlos en modelos de cáncer, debido a que esta enfermedad es una de las más simples para detectar transcritos quiméricos en sus muestras. Pero actualmente se sabe de enfermedades o síndromes que posiblemente pueden ser causadas o donde se acumulen transcritos quiméricos, que podrían ser perjudiciales para la salud. Por lo que sería importante poder mejorar estos programas para que su poder de análisis sea más diverso en la sensibilidad y especificidad de detección, parámetros, uso de índices o librerías y alineadores.

Por lo mencionado anteriormente, en esta investigación se consiguieron algunos resultados concretos, aunque es importante que, para poder tener una conclusión con mayor validez, habría que probar el pipeline utilizado en otros programas para conseguir comparar los resultados obtenidos.

Otra de las proyecciones sería la corrección de expresión que hay en los PQ dadores con los aceptores, ya que en los reportes también se dan los *reads per kilobase of transcript per million reads mapped* (RPKM) o por *transcript per million* (TPM), ambos son unidades de expresión de genes que miden los niveles de expresión de los genes o de los transcritos, de cada transcrito quiméricos identificados y de ambos PQ, por lo que se puede calcular el nivel de expresión que puede tener cada uno de estos y ver qué tan común es la expresión de estos puntos en la muestra.

7.1 Análisis en mayor cantidad de tejidos

Las muestras que se utilizaron solo fueron de *paired end* para facilitar el alineamiento de los *reads* que empalman entre exones, y así mejorar la detección de TQ. Se puede apreciar en la **Figura 13** que no se utilizaron la misma cantidad de muestras por tejidos durante el análisis, esto provoca una mayor dificultad en la comparación entre tejidos. Para análisis futuros se debe aumentar la cantidad de muestras por cada tejido y además de un mayor número de muestras, para así poder obtener resultados más concluyentes en los diferentes tipos de tejidos que existen en el cerebro.

7.2 Análisis de linfocitos en enfermedades

Las herramientas computacionales para la detección de TQ pueden tener un gran impacto en el análisis de diversas enfermedades, tanto para la identificación de posibles transcritos quiméricos en síndromes o también en el análisis de linfocitos para diversas enfermedades. Esto es debido a que los linfocitos o anticuerpos recombinantes son únicos para cada enfermedad y actualmente el método para conocer las recombinaciones que genera el sistema inmune para poder generar estos anticuerpos es costoso y se utilizan modelos de murinos los cuales tardan mucho tiempo en poder ser identificados. Utilizando estas herramientas informáticas se podría identificar la recombinación que se generan en el complejo de histocompatibilidad que produce anticuerpos, de esta manera la transcriptómica de linfocitos en una infección aguda nos permitirá deducir las recombinaciones que producen anticuerpos específicos para una determinada enfermedad. Con todas las implicaciones biotecnológicas que esto conlleva.

7.3 Detección de patrones en síndrome de Rett

En la investigación que se llevó a cabo, se encontraron diferentes transcritos quiméricos que tenían un aumento de repeticiones en las 50 muestras *MECP2-KO* en comparación con las muestras silvestre, esto se hizo en secuenciación masiva de ARN de muestra de cerebro de modelos Rett en *Mus musculus*. El pipeline generado se podría mejorar para poder utilizarlo muestra sanguínea, lo cual podría usarse para observar si existen algunos patrones que se repitan entre cerebro y sangre de ratón *MECP2-KO* y que puedan ser utilizados como biomarcadores de pronóstico temprano. Este puede sentar las bases para la detección temprana de algunas sintomatologías dentro del RTT y utilizar esta técnica y presencia de TQ en evaluación pronóstica del desarrollo de la enfermedad en pacientes reales y tomar las medidas terapéuticas necesarias para evitar o disminuir las sintomatologías que sean perjudiciales de este síndrome.

Además, se podría generar un prueba de detección de severidad en RTT en base a los TQ obtenidos de las diferentes muestras, esto se podría generar en modelos de *Mus musculus* de RTT clásico y atípico, y ver si existen diferencias entre los TQ obtenidos de los análisis generados para estas muestras, esto nos podría ayudar a prevenir sintomatologías que pudieran poner en riesgo la salud de la persona que padece de este síndrome.

8. Referencias

1. Developmental Disabilities Assistance and Bill of Rights Act, Public law 95-602, Stat. 1978 3004-5
2. Mullin AP, Gokhale A, Moreno-De-Luca A, Sanyal S, Waddington JL, Faundez V. Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. *Transl Psychiatry* [Internet]. 2013 Dec 3 [cited 2022 Jan 16];3(12). Available from: <https://pubmed.ncbi.nlm.nih.gov/24301647/>
3. Rangasamy S, D'Mello SR, Narayanan V. Epigenetics, autism spectrum, and neurodevelopmental disorders. *Neurotherapeutics* [Internet]. 2013 Oct [cited 2022 Jan 16];10(4):742–56. Available from: <https://pubmed.ncbi.nlm.nih.gov/24104594/>
4. Doernberg E, Hollander E. Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. *CNS Spectr* [Internet]. 2016 Aug 1 [cited 2022 Jan 16];21(4):295–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/27364515/>
5. Bull MJ. Down Syndrome. Ropper AH, editor. <https://doi.org/10.1056/NEJMra1706537> [Internet]. 2020 Jun 10 [cited 2022 Jan 16];382(24):2344–52. Available from: <https://www.nejm.org/doi/full/10.1056/NEJMra1706537>
6. Gold WA, Krishnaraj R, Ellaway C, Christodoulou J. Rett Syndrome: A Genetic Update and Clinical Review Focusing on Comorbidities. *ACS Chem Neurosci* [Internet]. 2018 Feb 21 [cited 2022 Jan 16];9(2):167–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/29185709/>
7. Rett A. [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wiener Medizinische Wochenschrift* (1946) [Internet]. 1966 Sep 1 [cited 2022 Jan 16];116(37):723–6. Available from: <https://europepmc.org/article/med/5300597>
8. Hagberg B, Hanefeld F, Percy A, Skjeldal O. An update on clinically applicable diagnostic criteria in Rett syndrome. Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany, 11 September 2001. *Eur J Paediatr Neurol* [Internet]. 2002 [cited 2022 Jan 16];6(5):293–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/12378695/>
9. Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. *Neuron* [Internet]. 2007 Nov 8 [cited 2022 Jan 16];56(3):422–37. Available from: <https://pubmed.ncbi.nlm.nih.gov/17988628/>

10. Laurvick CL, de Klerk N, Bower C, Christodoulou J, Ravine D, Ellaway C, et al. Rett syndrome in Australia: a review of the epidemiology. *J Pediatr* [Internet]. 2006 Mar [cited 2022 Jan 16];148(3):347–52. Available from: <https://pubmed.ncbi.nlm.nih.gov/16615965/>
11. Schüle B, Armstrong DD, Vogel H, Oviedo A, Francke U. Severe congenital encephalopathy caused by MECP2 null mutations in males: central hypoxia and reduced neuronal dendritic structure. *Clin Genet* [Internet]. 2008 Aug [cited 2022 Jan 16];74(2):116–26. Available from: <https://pubmed.ncbi.nlm.nih.gov/18477000/>
12. Tapia1+ P, Kouro2+ R, Pérez1-3 M, Torres4 R, Espinoza1 S, Kerr1 B, et al. RETT SYNDROME AND ENVIRONMENTAL ENRICHMENT AS A POTENTIAL THERAPY FOR ATTENUATING THE PATHOLOGY. *Physiological Mini Reviews*. 2019;12.
13. Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, et al. Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. *Ann Neurol* [Internet]. 2010 Dec [cited 2022 Jan 16];68(6):944. Available from: </pmc/articles/PMC3058521/>
14. Hagberg B. Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* [Internet]. 2002 [cited 2022 Jan 16];8(2):61–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/12112728/>
15. Posadas Sosa R, Ríos Medina, E. Verganza, Brenner S. C. SINDROME DE RETT REVISION BIBLIOGRAFICA Y PRESENTACION DE CASOS. en línea. valencia; 2014 Jun.
16. Lavås J, Slotte A, Jochym-Nygren M, van Doorn J, Engerström IW. Communication and eating proficiency in 125 females with Rett syndrome: The Swedish Rett center survey. <https://doi.org/101080/09638280600554868> [Internet]. 2009 Oct [cited 2022 Jan 16];28(20):1267–79. Available from: <https://www.tandfonline.com/doi/abs/10.1080/09638280600554868>
17. Didden R, Korzilius H, Smeets E, Green VA, Lang R, Lancioni GE, et al. Communication in Individuals with Rett Syndrome: an Assessment of Forms and Functions. *J Dev Phys Disabil* [Internet]. 2010 [cited 2022 Jan 16];22(2):105–18. Available from: <https://pubmed.ncbi.nlm.nih.gov/20339577/>
18. Witt-Engerström I, Gillberg C. Rett syndrome in Sweden. *J Autism Dev Disord* [Internet]. 1987 Mar [cited 2022 Jan 16];17(1):149–50. Available from: <https://pubmed.ncbi.nlm.nih.gov/3571139/>

19. Percy AK, Zoghbi HY, Lewis KR, Jankovic J. Rett syndrome: qualitative and quantitative differentiation from autism. *J Child Neurol* [Internet]. 1988 [cited 2022 Jan 16];3 Suppl(SUPPL.). Available from: <https://pubmed.ncbi.nlm.nih.gov/3198904/>
20. Tsai LY. Is Rett syndrome a subtype of pervasive developmental disorders? *J Autism Dev Disord* [Internet]. 1992 Dec [cited 2022 Jan 16];22(4):551–61. Available from: <https://pubmed.ncbi.nlm.nih.gov/1483976/>
21. Vahia VN. Diagnostic and statistical manual of mental disorders 5: A quick glance. *Indian Journal of Psychiatry* [Internet]. 2013 Jul [cited 2022 Jan 16];55(3):220. Available from: </pmc/articles/PMC3777342/>
22. Amir RE, van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* [Internet]. 1999 Oct [cited 2022 Jan 16];23(2):185–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/10508514/>
23. Ip JPK, Mellios N, Sur M. Rett syndrome: insights into genetic, molecular and circuit mechanisms. *Nat Rev Neurosci* [Internet]. 2018 Jun 1 [cited 2022 Jan 16];19(6):368–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/29740174/>
24. Trappe R, Laccione F, Cobilanschi J, Meins M, Huppke P, Hanefeld F, et al. MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin. *Am J Hum Genet* [Internet]. 2001 [cited 2022 Jan 16];68(5):1093–101. Available from: <https://pubmed.ncbi.nlm.nih.gov/11309679/>
25. Neul JL, Fang P, Barrish J, Lane J, Caeg EB, Smith EO, et al. Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* [Internet]. 2008 [cited 2022 Jan 16];70(16):1313–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/18337588/>
26. Neul JL, Fang P, Barrish J, Lane J, Caeg EB, Smith EO, et al. Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* [Internet]. 2008 [cited 2022 Jan 16];70(16):1313–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/18337588/>
27. Kalscheuer VM, Freude K, Musante L, Jensen LR, Yntema HG, Gécz J, et al. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet* [Internet]. 2003 Dec [cited 2022 Jan 16];35(4):313–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/14634649/>

28. Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, et al. FOXP1 Is Responsible for the Congenital Variant of Rett Syndrome. *American Journal of Human Genetics* [Internet]. 2008 Jul 11 [cited 2022 Jan 16];83(1):89. Available from: [/pmc/articles/PMC2443837/](#)
29. Ehrhart F, Sangani NB, Curfs LMG. Current developments in the genetics of Rett and Rett-like syndrome. *Curr Opin Psychiatry* [Internet]. 2018 Mar 1 [cited 2022 Jan 16];31(2):103–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/29206688/>
30. Iourov IY, Vorsanova SG, Yurov YB, Bertrand T. VIII World Rett Syndrome Congress & Symposium of rare diseases, Kazan, Russia. *Mol Cytogenet* [Internet]. 2018 Dec 24 [cited 2022 Jan 16];11(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/30603047/>
31. Sajan SA, Jhangiani SN, Muzny DM, Gibbs RA, Lupski JR, Glaze DG, et al. Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. *Genet Med* [Internet]. 2017 Jan 1 [cited 2022 Jan 16];19(1):13–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/27171548/>
32. Meehan R, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research* [Internet]. 1992 Oct 11 [cited 2022 Jan 16];20(19):5085. Available from: [/pmc/articles/PMC334288/?report=abstract](#)
33. Cowley SM, Iritani BM, Mendrysa SM, Xu T, Cheng PF, Yada J, et al. The mSin3A Chromatin-Modifying Complex Is Essential for Embryogenesis and T-Cell Development. *Molecular and Cellular Biology* [Internet]. 2005 Aug 15 [cited 2022 Jan 16];25(16):6990. Available from: [/pmc/articles/PMC1190252/](#)
34. Rao R, Fiskus W, Ganguly S, Kambhampati S, Bhalla KN. HDAC inhibitors and chaperone function. *Adv Cancer Res* [Internet]. 2012 [cited 2022 Jan 16];116:239–62. Available from: <https://pubmed.ncbi.nlm.nih.gov/23088873/>
35. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* [Internet]. 1998 May 28 [cited 2022 Jan 16];393(6683):386–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/9620804/>
36. Jones PL, Veenstra GJC, Wade PA, Vermaak D, Kass SU, Landsberger N, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* [Internet]. 1998 [cited 2022 Jan 16];19(2):187–91. Available from: <https://pubmed.ncbi.nlm.nih.gov/9620779/>
37. Chahrour M, Sung YJ, Shaw C, Zhou X, Wong STC, Qin J, et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* [Internet]. 2008 May 30

- [cited 2022 Jan 16];320(5880):1224–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/18511691/>
38. Georgel PT, Horowitz-Scherer RA, Adkins N, Woodcock CL, Wade PA, Hansen JC. Chromatin Compaction by Human MeCP2: ASSEMBLY OF NOVEL SECONDARY CHROMATIN STRUCTURES IN THE ABSENCE OF DNA METHYLATION. *Journal of Biological Chemistry*. 2003 Aug 22;278(34):32181–8.
 39. Horike SI, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* [Internet]. 2005 Jan [cited 2022 Jan 16];37(1):31–40. Available from: <https://pubmed.ncbi.nlm.nih.gov/15608638/>
 40. Olson CO, Zachariah RM, Ezeonwuka CD, Liyanage VRB, Rastegar M. Brain Region-Specific Expression of MeCP2 Isoforms Correlates with DNA Methylation within Mecp2 Regulatory Elements. *PLOS ONE* [Internet]. 2014 Mar 3 [cited 2022 Jan 16];9(3):e90645. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090645>
 41. Zachariah RM, Rastegar M. Linking epigenetics to human disease and Rett syndrome: the emerging novel and challenging concepts in MeCP2 research. *Neural Plast* [Internet]. 2012 [cited 2022 Jan 16];2012. Available from: <https://pubmed.ncbi.nlm.nih.gov/22474603/>
 42. Delcuve GP, Rastegar M, Davie JR. Epigenetic control. *Journal of Cellular Physiology* [Internet]. 2009 May 1 [cited 2022 Jan 16];219(2):243–50. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.21678>
 43. Borrelli E, Nestler EJ, David Allis C, Sassone-Corsi P. Decoding the Epigenetic Language of Neuronal Plasticity. *Neuron*. 2008;60(6):961–74.
 44. Krishnaraj R, Ho G, Christodoulou J. RettBASE: Rett syndrome database update. *Hum Mutat* [Internet]. 2017 Aug 1 [cited 2022 Jan 16];38(8):922–31. Available from: <https://pubmed.ncbi.nlm.nih.gov/28544139/>
 45. Guy J, Cheval H, Selfridge J, Bird A. The role of MeCP2 in the brain. *Annu Rev Cell Dev Biol* [Internet]. 2011 [cited 2022 Jan 16];27:631–52. Available from: <https://pubmed.ncbi.nlm.nih.gov/21721946/>
 46. Ehrhart F, Coort SLM, Cirillo E, Smeets E, Evelo CT, Curfs LMG. Rett syndrome – biological pathways leading from MECP2 to disorder phenotypes. *Orphanet Journal of Rare Diseases* 2016 11:1 [Internet]. 2016 Nov 25 [cited 2022 Jan 16];11(1):1–13. Available from: <https://ojrd.biomedcentral.com/articles/10.1186/s13023-016-0545-5>

47. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* [Internet]. 1998 May 28 [cited 2022 Jan 16];393(6683):386–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/9620804/>
48. Tudor M, Akbarian S, Chen RZ, Jaenisch R. Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci U S A* [Internet]. 2002 Nov 26 [cited 2022 Jan 16];99(24):15536–41. Available from: <https://pubmed.ncbi.nlm.nih.gov/12432090/>
49. Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, et al. L1 retrotransposition in neurons is modulated by MeCP2. *Nature* [Internet]. 2010 Nov 18 [cited 2022 Jan 16];468(7322):443–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/21085180/>
50. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 2004 428:6982 [Internet]. 2004 Apr 1 [cited 2022 Jan 16];428(6982):493–521. Available from: <https://www.nature.com/articles/nature02426>
51. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2003 420:6915 [Internet]. 2002 Dec 5 [cited 2022 Jan 16];420(6915):520–62. Available from: <https://www.nature.com/articles/nature01262>
52. Moran J v., Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH. High frequency retrotransposition in cultured mammalian cells. *Cell* [Internet]. 1996 Nov 29 [cited 2022 Jan 16];87(5):917–27. Available from: <https://pubmed.ncbi.nlm.nih.gov/8945518/>
53. Muotri AR, Zhao C, Marchetto MCN, Gage FH. Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* [Internet]. 2009 Oct [cited 2022 Jan 16];19(10):1002–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/19771587/>
54. Singer T, McConnell MJ, Marchetto MCN, Coufal NG, Gage FH. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences*. 2010 Aug 1;33(8):345–54.
55. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* [Internet]. 1988 [cited 2022 Jan 17];332(6160):164–6. Available from: <https://pubmed.ncbi.nlm.nih.gov/2831458/>

56. Miki Y, Miki Y, Nishisho I, Horii A, Ai E, Res C, et al. Disruption of the APC gene by retrotransposal insertion of L1 sequence in a colon cancer. *CANC RES* [Internet]. 1992 [cited 2022 Jan 17];643--645. Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.327.4703>
57. Faulkner GJ, Garcia-Perez JL. L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends Genet* [Internet]. 2017 Nov 1 [cited 2022 Jan 17];33(11):802–16. Available from: <https://pubmed.ncbi.nlm.nih.gov/28797643/>
58. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* [Internet]. 2006 Apr 14 [cited 2022 Jan 17];357(5):1383–93. Available from: <https://pubmed.ncbi.nlm.nih.gov/16490214/>
59. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran J v., Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* [Internet]. 2005 Jun 16 [cited 2022 Jan 17];435(7044):903–10. Available from: <https://pubmed.ncbi.nlm.nih.gov/15959507/>
60. Tchénio T, Casella JF, Heidmann T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Research* [Internet]. 2000 Jan 15 [cited 2022 Jan 17];28(2):411. Available from: </pmc/articles/PMC102531/>
61. Yu F, Zingler N, Schumann G, Strätling WH. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Research* [Internet]. 2001 Nov 1 [cited 2022 Jan 17];29(21):4493. Available from: </pmc/articles/PMC60185/>
62. Shpyleva S, Melnyk S, Pavliv O, Pogribny I, Jill James S. Overexpression of LINE-1 Retrotransposons in Autism Brain. *Mol Neurobiol* [Internet]. 2018 Feb 1 [cited 2022 Jan 17];55(2):1740–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/28220356/>
63. Kuo LJ, Yang LX. γ -H2AX - A Novel Biomarker for DNA Double-strand Breaks. *In Vivo* [Internet]. 2008 May 1 [cited 2022 Jan 17];22(3):305–9. Available from: <https://iv.iiarjournals.org/content/22/3/305>
64. Alessio N, Riccitiello F, Squillaro T, Capasso S, del Gaudio S, di Bernardo G, et al. Neural stem cells from a mouse model of Rett syndrome are prone to senescence, show reduced capacity to cope with genotoxic stress, and are impaired in the differentiation process. *Exp Mol Med* [Internet]. 2018 [cited 2022 Jan 17];50(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/29563495/>
65. Enikanolaiye A, Ruston J, Zeng R, Taylor C, Schrock M, Buchovecky CM, et al. Suppressor mutations in Mecp2-null mice implicate the DNA damage response in Rett syndrome pathology.

- Genome Research [Internet]. 2020 Apr 1 [cited 2022 Jan 17];30(4):540–52. Available from: [/pmc/articles/PMC7197480/](#)
66. Panagopoulos I, Heim S. Interstitial Deletions Generating Fusion Genes. *Cancer Genomics Proteomics* [Internet]. 2021 May 1 [cited 2022 Mar 8];18(3):167–96. Available from: <https://pubmed.ncbi.nlm.nih.gov/33893073/>
 67. Dai X, Theobard R, Cheng H, Xing M, Zhang J. Fusion genes: A promising tool combating against cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2018 Apr 1;1869(2):149–60.
 68. Segni G di, Gastaldi S, Tocchini-Valentini GP. Cis-and trans-splicing of mRNAs mediated by tRNA sequences in eukaryotic cells. 2008 [cited 2022 Apr 28]; Available from: www.pnas.org/cgi/content/full/
 69. Heim Sverre, Mitelman Felix. *Cancer Cytogenetics : Chromosomal and Molecular Genetic Aberrations of Tumor Cells*. 2015;645.
 70. ENA Browser [Internet]. [cited 2022 Jan 17]. Available from: <https://www.ebi.ac.uk/ena/browser/home>
 71. Harrison PW, Ahamed A, Aslam R, Alako BTF, Burgin J, Buso N, et al. The European Nucleotide Archive in 2020. *Nucleic Acids Res* [Internet]. 2021 Jan 8 [cited 2022 Jan 17];49(D1):D82–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/33175160/>
 72. Andrews Simon, Krueger Felix, Segonds-Pichon Anne, Biggins Laura, Krueger Christel, Montgomery Jo. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. [cited 2022 Jan 17]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 73. GitHub - FelixKrueger/TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data [Internet]. [cited 2022 Jan 17]. Available from: <https://github.com/FelixKrueger/TrimGalore>
 74. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Research* [Internet]. 2021 Jan 1 [cited 2022 Jan 17];31(3):448–60. Available from: [/pmc/articles/PMC7919457/](#)
 75. Haas BJ, Dobin A, Stransky N, Li B, Yang X, Tickle T, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv* [Internet]. 2017 Mar 24 [cited 2022 Jan 17];120295. Available from: <https://www.biorxiv.org/content/10.1101/120295v1>
 76. GRCm38 - mm10 - Genome - Assembly - NCBI [Internet]. [cited 2022 May 5]. Available from: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/

77. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2013 Jan [cited 2022 Jan 17];29(1):15–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/23104886/>
78. Haas B, Tickle T, Pochet N, Sun J, Hsu P, Stewart C, et al. Fast and accurate fusion transcript detection using the Trinity Cancer Transcriptome Analysis Toolkit. *Cancer Research* [Internet]. 2016 Jul 15 [cited 2022 Jan 17];76(14 Supplement):5266–5266. Available from: https://cancerres.aacrjournals.org/content/76/14_Supplement/5266
79. R Core Development Team. R: a language and environment for statistical computing, 3.2.1. Document freely available on the internet at: <http://www.r-project.org> [Internet]. 2015 [cited 2022 Jan 17]; Available from: <https://www.yumpu.com/en/document/view/6853895/r-a-language-and-environment-for-statistical-computing>
80. RStudio - RStudio [Internet]. [cited 2022 Jan 17]. Available from: <https://www.rstudio.com/products/rstudio/>
81. Villanueva RAM, Chen ZJ. *ggplot2: Elegant Graphics for Data Analysis* (2nd ed.). <https://doi.org/10.1080/1536636720191565254> [Internet]. 2019 Jul 3 [cited 2022 Jan 17];17(3):160–7. Available from: <https://www.tandfonline.com/doi/abs/10.1080/15366367.2019.1565254>
82. Package “ggpointdensity” Type Package Title A Cross Between a 2D Density Plot and a Scatter Plot. 2019 [cited 2022 Jan 17]; Available from: <https://github.com/LKremer/ggpointdensity/issues>
83. viridis: Colorblind-Friendly Color Maps for R version 0.6.2 from CRAN [Internet]. [cited 2022 Jan 17]. Available from: <https://rdrr.io/cran/viridis/>
84. Package “hrbrthemes” Type Package Title Additional Themes, Theme Components and Utilities for “ggplot2.” 2020 [cited 2022 Jan 17]; Available from: <https://orcid.org/0000-0001-5670-2640>
85. Home - GraphPad [Internet]. [cited 2022 Jan 17]. Available from: <https://www.graphpad.com/>
86. Zhang H, Meltzer P, Davis S. RCircos: An R package for Circos 2D track plots. *BMC Bioinformatics* [Internet]. 2013 Aug 10 [cited 2022 Jan 17];14(1):1–5. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-244>
87. UCSC Genome Browser Home [Internet]. [cited 2022 Jan 19]. Available from: <https://genome.ucsc.edu/index.html>
88. SIB Swiss Institute of Bioinformatics | Expasy [Internet]. [cited 2022 Jan 19]. Available from: <https://www.expasy.org/>

89. Letunic I, Khedkar S, Bork P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*. 2021 Jan 8;49(D1):D458–60.
90. Yu L, Zhang W, Wang P, Zhang Q, Cong A, Yang X, et al. LncRNA SNHG11 aggravates cell proliferation and migration in triple-negative breast cancer via sponging miR-2355-5p and targeting CBX5. *Exp Ther Med* [Internet]. 2021 Jun 18 [cited 2022 Jan 18];22(2). Available from: <https://pubmed.ncbi.nlm.nih.gov/34257707/>
91. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*. 2009 Mar;10(3):155–9.
92. Mier P, Paladin L, Tamana S, Petrosian S, Hajdu-Soltész B, Urbanek A, et al. Disentangling the complexity of low complexity proteins. *Briefings in Bioinformatics*. 21(2):458–72.
93. de Felice FG, Vieira MNN, Nazareth M, Meirelles L, Morozova-Roche LA, Dobson CM, et al. Formation of amyloid aggregates from human lysozyme and its disease-associated variants using hydrostatic pressure. *The FASEB Journal* [Internet]. 2004 Jul 1 [cited 2022 May 21];18(10):1099–101. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1096/fj.03-1072fje>
94. Percy AK. Rett syndrome: exploring the autism link. *Arch Neurol* [Internet]. 2011 Aug [cited 2022 Jan 17];68(8):985–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/21825235/>
95. Shah RR, Bird AP. MeCP2 mutations: progress towards understanding and treating Rett syndrome. *Genome Med* [Internet]. 2017 Feb 17 [cited 2022 Jan 17];9(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/28212680/>
96. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology* [Internet]. 2019 Oct 21 [cited 2022 Jan 17];20(1):1–16. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1842-9>
97. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* [Internet]. 2013 Feb [cited 2022 Jan 17];19(2):141–57. Available from: <https://pubmed.ncbi.nlm.nih.gov/23249747/>
98. Adelman K, Egan E. Non-coding RNA: More uses for genomic junk. *Nature* [Internet]. 2017 Mar 8 [cited 2022 Jan 17];543(7644):183–5. Available from: <https://pubmed.ncbi.nlm.nih.gov/28277509/>
99. Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, et al. Genomewide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent

- neural accumulation. *Cell Rep* [Internet]. 2014 Dec 11 [cited 2022 Jan 17];9(5):1966. Available from: [/pmc/articles/PMC4279448/](https://pubmed.ncbi.nlm.nih.gov/27958329/)
100. Rybak-Wolf A, Stottmeister C, Glažar P, Jens M, Pino N, Giusti S, et al. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Molecular Cell*. 2015 Jun 4;58(5):870–85.
 101. Gruner H, Cortés-López M, Cooper DA, Bauer M, Miura P. CircRNA accumulation in the aging mouse brain. *Sci Rep* [Internet]. 2016 Dec 13 [cited 2022 Jan 17];6. Available from: <https://pubmed.ncbi.nlm.nih.gov/27958329/>
 102. Aufiero S, Reckman YJ, Pinto YM, Creemers EE. Circular RNAs open a new chapter in cardiovascular biology. *Nat Rev Cardiol* [Internet]. 2019 Aug 1 [cited 2022 Jan 17];16(8):503–14. Available from: <https://pubmed.ncbi.nlm.nih.gov/30952956/>
 103. Rybak-Wolf A, Stottmeister C, Glažar P, Jens M, Pino N, Giusti S, et al. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol Cell* [Internet]. 2015 Dec 22 [cited 2022 Jan 17];58(5):870–85. Available from: <https://pubmed.ncbi.nlm.nih.gov/25921068/>
 104. Armstrong D, Dunn JK, Antalffy B, Trivedi R. Selective dendritic alterations in the cortex of Rett syndrome. *J Neuropathol Exp Neurol* [Internet]. 1995 [cited 2022 Jan 17];54(2):195–201. Available from: <https://pubmed.ncbi.nlm.nih.gov/7876888/>
 105. Siqueira E, Obiols-Guardia A, Jorge-Torres OC, Oliveira-Mateos C, Soler M, Ramesh-Kumar D, et al. Analysis of the circRNA and T-UCR populations identifies convergent pathways in mouse and human models of Rett syndrome. *Molecular Therapy - Nucleic Acids* [Internet]. 2022 Mar 8 [cited 2022 Jan 17];27:621–44. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2162253121003334>
 106. Hancks DC, Kazazian HH. Roles for retrotransposon insertions in human disease. *Mob DNA* [Internet]. 2016 May 6 [cited 2022 Jan 17];7(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/27158268/>
 107. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001 409:6822 [Internet]. 2001 Feb 15 [cited 2022 Jan 17];409(6822):860–921. Available from: <https://www.nature.com/articles/35057062>
 108. Ramocki MB, Tavyev YJ, Peters SU. The MECP2 duplication syndrome. *American Journal of Medical Genetics, Part A*. 2010 May;152(5):1079–88.

109. Wu Q, Ma J, Wei J, Meng W, Wang Y, Shi M. lncRNA SNHG11 Promotes Gastric Cancer Progression by Activating the Wnt/ β -Catenin Pathway and Oncogenic Autophagy. *Mol Ther* [Internet]. 2021 Mar 3 [cited 2022 Jan 18];29(3):1258–78. Available from: <https://pubmed.ncbi.nlm.nih.gov/33068778/>
110. XIE Q, ZHAO S, KANG R, WANG X. lncRNA SNHG11 facilitates prostate cancer progression through the upregulation of IGF-1R expression and by sponging miR-184. *Int J Mol Med* [Internet]. 2021 Sep 1 [cited 2022 Jan 18];48(3). Available from: <https://pubmed.ncbi.nlm.nih.gov/34328198/>
111. Fang X, Cai Y, Xu Y, Zhang H. Exosome-mediated lncRNA SNHG11 regulates angiogenesis in pancreatic carcinoma through miR-324-3p/VEGFA axis. 2021;
112. Peng WX, Koirala P, Mo YY. lncRNA-mediated regulation of cell signaling in cancer. [cited 2022 Jan 18]; Available from: <http://www.noncode.org>
113. Wu Q, Ma J, Wei J, Meng W, Wang Y, Shi M. lncRNA SNHG11 Promotes Gastric Cancer Progression by Activating the Wnt/ β -Catenin Pathway and Oncogenic Autophagy. 2021 [cited 2022 May 14]; Available from: <https://doi.org/10.1016/j.ymthe.2020.10.011>.
114. Hülken J, Birchmeier W, Behrens J. E-cadherin and APC compete for the interaction with β -catenin and the cytoskeleton. *J Cell Biol* [Internet]. 1994 Dec [cited 2022 May 14];127(6 Pt 2):2061–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/7806582/>
115. Pai SG, Carneiro BA, Mota JM, Costa R, Leite CA, Barroso-Sousa R, et al. Wnt/ β -catenin pathway: modulating anticancer immune response.
116. van Amerongen R, Nusse R. Towards an integrated view of Wnt signaling in development. *Development* [Internet]. 2009 Oct 1 [cited 2022 May 15];136(19):3205–14. Available from: <https://journals.biologists.com/dev/article/136/19/3205/65324/Towards-an-integrated-view-of-Wnt-signaling-in>
117. Kwan V, Unda BK, Singh KK. Wnt signaling networks in autism spectrum disorder and intellectual disability. *Journal of Neurodevelopmental Disorders* 2016 8:1 [Internet]. 2016 Dec 5 [cited 2022 May 15];8(1):1–10. Available from: <https://jneurodevdisorders.biomedcentral.com/articles/10.1186/s11689-016-9176-3>
118. Nusse R. Wnt signaling in disease and in development. *Cell Research* 2005 15:1 [Internet]. 2005 [cited 2022 May 15];15(1):28–32. Available from: <https://www.nature.com/articles/7290260>

119. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* 2016 18:2 [Internet]. 2016 Nov 21 [cited 2022 Jan 23];18(2):71–86. Available from: <https://www.nature.com/articles/nrg.2016.139>
120. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran J v., Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 2005 435:7044 [Internet]. 2005 Jun 16 [cited 2022 Jan 23];435(7044):903–10. Available from: <https://www.nature.com/articles/nature03663>
121. Macla A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzaouad M, et al. Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Research* [Internet]. 2017 Mar 1 [cited 2022 Jan 23];27(3):335–48. Available from: <https://genome.cshlp.org/content/27/3/335.full>
122. L1 Mosaicism in Mammals: Extent, Effects, and Evolution | Elsevier Enhanced Reader [Internet]. [cited 2022 Jan 23]. Available from: <https://reader.elsevier.com/reader/sd/pii/S0168952517301130?token=0AF622A74B426200E9EC43B62EE9817F0EA061BBAE9F4D89B0B0E089C6257538DA92BE37BA1C95E1E9FA8C7DCE7F387&originRegion=us-east-1&originCreation=20220124155119>
123. Padrón González J, Pérez Mejías R, Padrón Fernández L, Padrón Fernández L. Síndrome de Rett. Presentación de un caso. *MediSur* [Internet]. 2012 [cited 2022 Jan 23];10(1):69–71. Available from: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1727-897X2012000100013&lng=es&nrm=iso&tlng=es

9. Anexo

1. [Anexo Tesis Guillermo Albornoz](#)

SOLO USO ACADÉMICO