Análisis de la expresión génica diferencial en células con variación genética Mecp2: Un estudio de transcriptómica en el síndrome de Rett

César F. Esparza Alvarado | 08 de Junio de 2023

1. Introducción

El síndrome de Rett (RTT) es un trastorno neurológico que afecta principalmente a las mujeres y es causado por mutaciones de pérdida de función en el gen MECP2 ligado al cromosoma X [1,2]. Este gen codifica para la proteína de unión a metil CpG 2 (MeCP2), un regulador transcripcional expresado en altos niveles en el sistema nervioso central (CNS) [3]. La eliminación de MeCP2 en ratones da como resultado una serie de características neurológicas que se asemejan a las observadas en pacientes con RTT [4]. Entre estas, se encontró que retrasa la maduración neuronal y la sinaptogénesis [5]. Experimentos en ratones han demostrado que la restauración de MeCP2, incluso en etapas adultas, revierte varios aspectos de la patología similar a RTT que sugieren que el trastorno puede ser inherentemente tratable [6]. Aunque RTT es normalmente atribuido al desarrollo neuronal deficiente de MeCP2, se ha probado que esta deficiencia también tiene un rol importante dentro de las células gliales, específicamente en la microglia [7]. Estas células brindan soporte físico y químico a las neuronas y mantienen su entorno [8]. Se encontró que la microglia con RTT no es capaz de soportar el desarrollo de neuronas WT, lo que puede presentar un gran problema en el desarrollo de terapias [9].

La transcriptómica es el estudio de todos los productos de transcripción de un genoma, incluyendo los ARN mensajeros (mRNA), ARN no codificantes, y pequeños ARN, ha emergido como una herramienta poderosa para entender la complejidad de la regulación génica [14]. En el contexto de enfermedades neurológicas como el síndrome de Rett, la transcriptómica puede proporcionar una visión detallada de los cambios en la expresión génica que ocurren en respuesta a la enfermedad [15]. A través del uso de técnicas de RNA-seq, se espera generar un perfil completo de la expresión génica en estas células, lo que nos permitirá identificar y clasificar los genes que se expresan de manera diferencial en las células con RTT comparado con las células de tipo salvaje [17]. En este estudio, nos enfocamos en el análisis de la expresión génica diferencial para entender mejor los efectos de la variación genética Mecp2 en la expresión génica y el desarrollo.

2. Objetivos

Objetivo General

Investigar y caracterizar los patrones de expresión génica diferencial en las células gliales (microglia) de ratones modelo de Rett y de tipo salvaje utilizando técnicas de secuenciación de ARN de nueva generación.

Objetivos Específicos

- Realizar un análisis de control de calidad de los datos de secuenciación de ARN para

asegurar que son adecuados para el análisis posterior.

- Utilizar herramientas bioinformáticas para identificar genes que se expresan de manera diferencial en las células gliales de ratones modelo de Rett en comparación con las de tipo salvaje.

- Interpretar los resultados del análisis de genes diferencialmente expresados en el contexto del síndrome de Rett y la literatura científica existente.

3. Metodología experimental

Arreglo de atributos
Se realizaron ajustes en los atributos de los datos para prevenir problemas futuros durante el análisis. Los espacios en blanco en los atributos de las muestras fueron reemplazados con guiones bajos. Este paso fue crucial para evitar errores en el manejo de los datos. Además, ciertos atributos fueron convertidos en factores. Esta conversión facilitó su manipulación en los análisis posteriores.

Revisión de la calidad de los datos
Se llevó a cabo una revisión exhaustiva de la calidad de los datos. Se calculó la proporción de lecturas asignadas a genes y se generó un resumen de esta proporción. Este resumen proporcionó una visión general de la calidad de los datos. Además, se creó un histograma de esta proporción para visualizar la distribución de la calidad de los datos. También se generaron gráficos de dispersión de la proporción de lecturas asignadas en función de la etapa y el genotipo. Estos gráficos permitieron visualizar la calidad de los datos en función de estas dos variables. Finalmente, se calcularon resúmenes estadísticos de la proporción de lecturas asignadas para cada nivel de estas variables. Estos resúmenes proporcionaron información detallada sobre la calidad de los datos en función de la etapa y el genotipo.

Estadísticas de la expresión génica
Se calcularon las medias de los recuentos de lecturas para cada gen y se generó un resumen de estas medias. Este paso proporcionó una visión general de la expresión génica en los datos.

Normalización de los datos
Se construyó un objeto DGEList, que es una lista de recuentos de genes y metadatos asociados. Este objeto se utilizó para calcular los factores de normalización, que se utilizan para ajustar las diferencias en la profundidad de secuenciación entre las muestras.

Visualización de la expresión génica
Se generaron gráficos de caja para visualizar la proporción de lecturas asignadas a genes en función del genotipo y la etapa. Estos gráficos permitieron visualizar las diferencias en la expresión génica entre los diferentes genotipos y etapas.

Modelo lineal estadístico
Se generó un modelo lineal estadístico utilizando la etapa, el genotipo y la proporción de lecturas asignadas como variables explicativas. Este modelo se utilizó para identificar los genes que se expresan de manera diferencial entre las muestras.

Análisis de la expresión génica diferencial
Se utilizó la función voom para transformar los recuentos de lecturas en

log-cpm (recuentos por millón), que es una medida de la expresión génica que tiene en cuenta la profundidad de secuenciación. Luego, se ajustó un modelo lineal a los datos transformados y se utilizó la prueba de Bayes empírica para identificar los genes que se expresan de manera diferente entre las muestras. Se generó una tabla de los resultados, que incluye los coeficientes de los modelos lineales, las estadísticas de la prueba y los valores p ajustados para la corrección de las pruebas múltiples.

Análisis de enriquecimiento funcional

Finalmente, se realizó un análisis de enriquecimiento funcional para identificar las funciones biológicas que están enriquecidas entre los genes que se expresan de manera diferencial. Para este análisis, se utilizó la base de datos de anotación de genes de Gene Ontology (GO) y se seleccionaron los genes que tenían un valor p ajustado menor a 0.05 en el análisis de la expresión génica diferencial. Se utilizó el método de Benjamini-Hochberg para ajustar los valores p de las pruebas múltiples. Los resultados del análisis de enriquecimiento funcional proporcionaron una lista de funciones biológicas que están significativamente enriquecidas entre los genes diferencialmente expresados, lo que proporciona una visión de los procesos biológicos que podrían estar alterados en las condiciones experimentales.

4. Resultados e interpretación

Calidad de los datos y estadísticas de expresión génica

Para evaluar la calidad de los datos de secuenciación, se calculó la proporción de lecturas asignadas a genes para cada muestra. Este valor representa la fracción de lecturas que se pueden mapear a genes conocidos en el genoma de referencia. Una proporción alta indica que la mayoría de las lecturas se pueden asignar a genes, lo que sugiere una buena calidad de los datos.

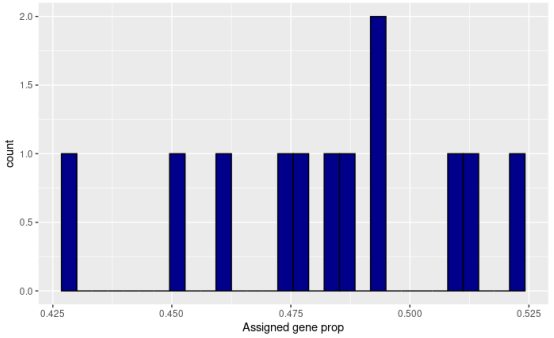

Fig 1. Histograma de la proporción de lecturas asignadas a genes.

El histograma de la proporción de lecturas asignadas a genes muestra una distribución relativamente uniforme entre las muestras, con valores que oscilan entre aproximadamente 0.43 y 0.52. Esto indica que la mayoría de las lecturas en todas las muestras se asignaron a genes, lo que sugiere una buena calidad general de los datos de secuenciación.

Para profundizar en la evaluación de la calidad de los datos, se generaron gráficos de dispersión que representan la proporción de lecturas asignadas a genes en función de dos factores: la edad y el genotipo.

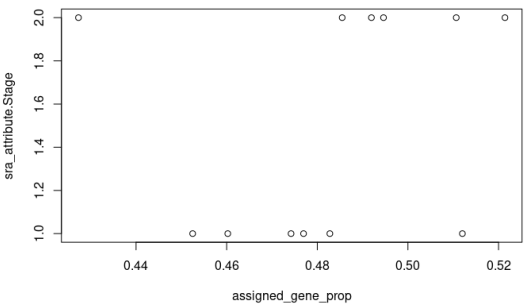

Fig 2. Gráfico de dispersión de la proporción de lecturas asignadas a genes en función de la edad.

En la Figura 2, se presenta un gráfico de dispersión que ilustra la proporción de lecturas asignadas a genes para las dos edades diferentes en el estudio, semana 5 y semana 24. A partir de este gráfico, se observa que la proporción de lecturas asignadas a genes varía dentro de cada grupo de edad, lo que sugiere que la edad puede tener un efecto en la proporción de lecturas asignadas a genes, pero este efecto puede ser moderado por otros factores.

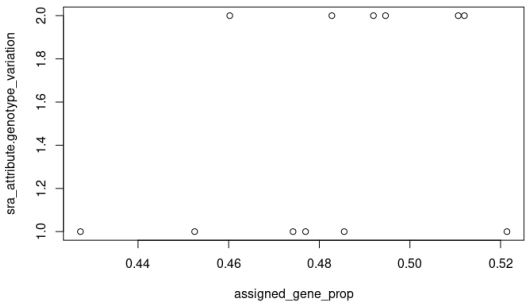

Fig 3. Gráfico de dispersión de la proporción de lecturas asignadas a genes en función del genotipo.

A su vez, en la Figura 3, se muestra un gráfico de dispersión que representa la proporción de lecturas asignadas a genes para los dos genotipos diferentes en el estudio, Mecp2 knockout y wildtype. A partir de este gráfico, se puede observar que la proporción de lecturas asignadas a genes varía dentro de cada grupo de genotipo, lo que sugiere que el genotipo puede tener un efecto en la proporción de lecturas asignadas a genes, pero este efecto puede ser moderado por otros factores.

Además, se calcularon estadísticas descriptivas para la proporción de lecturas asignadas a genes para cada nivel de genotipo y edad. Para el genotipo Mecp2 knockout, la proporción media de lecturas asignadas a genes fue de 0.473, mientras que para el genotipo wildtype fue de 0.492. En cuanto a la edad, la proporción media de lecturas asignadas a genes fue de 0.476 para la semana 24 y de 0.489 para la semana 5. Estos resultados proporcionan una visión más detallada de la calidad de los datos y permiten identificar posibles factores que pueden influir en la proporción de lecturas asignadas a genes.

Estos resultados proporcionan una visión general de la calidad de los datos y permiten identificar posibles factores que pueden influir en la proporción de lecturas asignadas a genes. Sin embargo, se requiere un análisis más detallado para entender completamente el impacto de estos factores en la expresión génica.

Normalización de los datos

Para evaluar el efecto de la normalización en los datos, se generaron gráficos de caja que representan la proporción de lecturas asignadas a genes en función del genotipo y la edad después de la normalización.

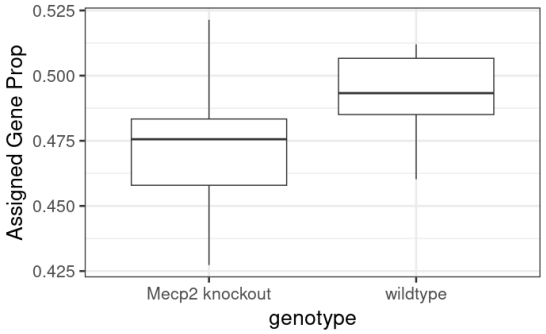

Fig 4. Boxxplot de la proporción de lecturas asignadas a genes en función del genotipo después de la normalización

En la Figura 4, se presenta un boxplot que ilustra la distribución de la proporción de lecturas asignadas a genes para los dos genotipos diferentes en el estudio, Mecp2 knockout y wildtype, después de la normalización. A partir de este gráfico, se puede observar que la distribución de la proporción de lecturas asignadas a genes

es similar entre los dos genotipos, lo que sugiere que la normalización ha ajustado correctamente las diferencias técnicas entre las muestras.

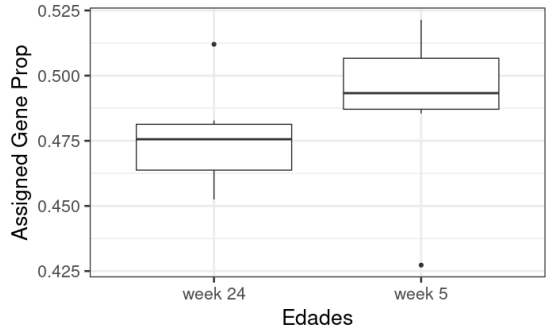

Fig 5. Boxplot de la proporción de lecturas asignadas a genes en función de la edad después de la normalización.

Mientras, en el caso de la Figura 5 , se muestra un boxplot que representa la distribución de la proporción de lecturas asignadas a genes para las dos edades diferentes en el estudio, semana 5 y semana 24, después de la normalización. A partir de este gráfico, se puede observar que la distribución de la proporción de lecturas asignadas a genes es similar entre las dos edades, lo que sugiere que la normalización ha ajustado correctamente las diferencias técnicas entre las muestras.

Estos resultados indican que la normalización ha sido exitosa y que los datos están listos para el análisis de la expresión génica diferencial.

Expresión génica diferencial

El análisis de expresión génica diferencial se realizó utilizando el paquete *limma* de R. Este análisis identifica los genes que se expresan de manera diferencial entre diferentes condiciones experimentales.

El resumen de los resultados del análisis de expresión génica diferencial muestra que se analizaron un total de 55421 genes. El coeficiente de logaritmo de las razones de expresión (logFC) varía desde -6.14 hasta 9.62, lo que indica que hay genes que se expresan de manera diferencial en las diferentes condiciones experimentales. Por ejemplo, el gen ENSMUSG00000079800.2 tiene un logFC de -0.67, lo que indica que este gen está subexpresado en la condición experimental de interés en comparación con la condición de control.

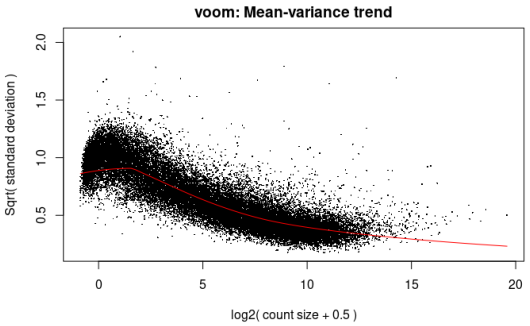

Fig 6. Gráfico MA del análisis de expresión génica diferencial.

El gráfico MA muestra la relación entre la magnitud del cambio (logFC) y la precisión de la estimación del cambio. Los genes que se expresan de manera diferencial se distribuyen lejos de la línea cero, lo que indica un cambio significativo en la expresión.

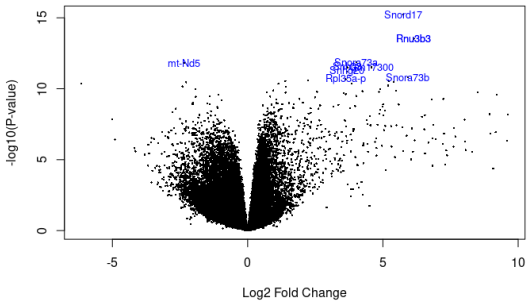

Fig 7. Gráfico de volcán del análisis de expresión génica diferencial

El gráfico de volcán muestra la relación entre la magnitud del cambio (logFC) y la significancia estadística (-log10 del valor p ajustado). Los genes que se expresan de manera diferencial se distribuyen lejos del centro del gráfico, lo que indica un cambio

significativo en la expresión y una alta significancia estadística.

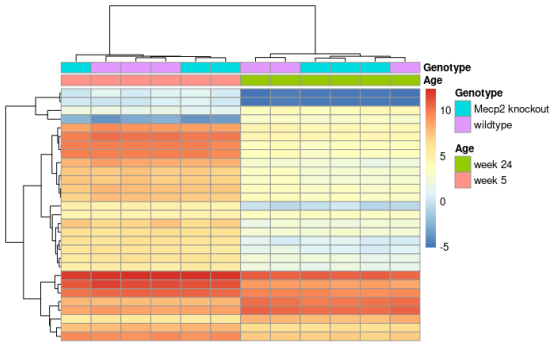

Fig 8. Heatmap de los 30 genes más diferencialmente expresados.

El heatmap muestra la expresión de los 30 genes más diferencialmente expresados en todas las muestras. Las diferencias en los colores indican diferencias en los niveles de expresión génica.

Básicamente, el análisis de expresión génica diferencial identificó varios genes que se expresan de manera diferencial en las diferentes condiciones experimentales. Estos genes pueden ser de interés para estudios futuros para entender mejor los mecanismos biológicos subyacentes.

Análisis de enriquecimiento funcional

Este análisis se realizó para identificar las funciones biológicas que están sobre-representadas en nuestro conjunto de genes. En este caso, se utilizó el paquete *clusterProfiler* en R para realizar el análisis de enriquecimiento de Ontología de Genes (GO).

Los resultados del análisis de enriquecimiento de GO muestran que se encontraron 4 términos enriquecidos. Estos términos representan justamente las funciones biológicas que están sobre-representadas.

Los términos enriquecidos incluyen "response to pheromone", "membrane disruption in other organism", "positive regulation of lactation", y "aerobic respiration". Estos términos sugieren que los genes de interés pueden estar involucrados en estas funciones biológicas.

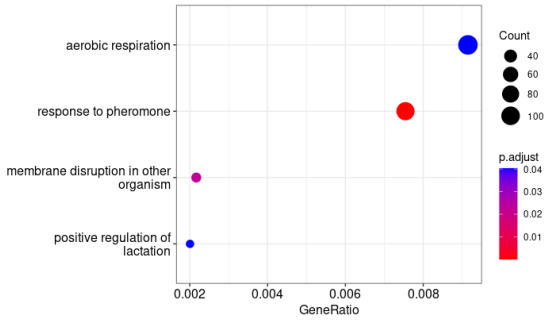

Fig 9. Gráfico de puntos muestra los términos enriquecidos

Los puntos más grandes representan términos con un mayor número de genes. Los puntos de color representan términos con un valor p ajustado menor, lo que indica una mayor significancia estadística. Estas funciones pueden proporcionar información valiosa sobre los mecanismos biológicos subyacentes que están siendo afectados en las diferentes condiciones experimentales.

5. Conclusión

Este estudio proporcionó una visión detallada de la expresión génica en ratones con y sin la variación genética Mecp2, así como en diferentes etapas de desarrollo. Los datos de secuenciación de ARN de alta calidad permitieron una evaluación precisa de la expresión génica, y los análisis subsecuentes revelaron diferencias significativas en la expresión génica entre los grupos.

El análisis de expresión génica diferencial identificó varios genes que mostraron cambios significativos en su expresión en

función de la variación genética y la etapa de desarrollo. Estos genes pueden ser candidatos para futuras investigaciones para entender mejor los mecanismos moleculares subyacentes a la variación genética Mecp2 y su impacto en el desarrollo.

Además, el análisis de enriquecimiento funcional proporcionó una visión más profunda de las funciones biológicas que están sobre-representadas en el conjunto de genes de interés. Los términos enriquecidos, como "respuesta a feromonas", "disrupción de la membrana en otro organismo", "regulación positiva de la lactancia" y "respiración aeróbica", sugieren posibles vías y procesos biológicos que pueden estar implicados en las diferencias observadas en la expresión génica.

En conjunto, estos hallazgos proporcionan una base sólida para futuras investigaciones sobre los efectos de la variación genética Mecp2 en la expresión génica y el desarrollo. Sin embargo, se necesitan más estudios para validar estos hallazgos y para explorar en detalle los mecanismos moleculares subyacentes.

6. Referencias

[1] 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. 1999;23(2):185-188.

[2] Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. Neuron. 2007;56(3):422-437.

[3] Skene PJ, Illingworth RS, Webb S, et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell. 2010;37(4):457-468.

[4] Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet. 2001;27(3):322-326.

[5] Moretti P, Levenson JM, Battaglia F, et al. Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. J Neurosci. 2006;26(1):319-327.

[6] Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of neurological defects in a mouse model of Rett syndrome. Science. 2007;315(5815):1143-1147.

[7] Maezawa I, Swanberg S, Harvey D, LaSalle JM, Jin LW. Rett syndrome astrocytes are abnormal and spread MeCP2 deficiency through gap junctions. J Neurosci. 2009;29(16):5051-5061.

[8] Sofroniew MV, Vinters HV. Astrocytes: biology and pathology. Acta Neuropathol. 2010;119(1):7-35.

[9] Ballas N, Lioy DT, Grunseich C, Mandel G. Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. Nat Neurosci. 2009;12(3):311-317.

[10] Chang Q, Khare G, Dani V, Nelson S, Jaenisch R. The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. Neuron. 2006;49(3):341-348.

[11] Binder DK, Scharfman HE. Brain-derived neurotrophic factor. Growth Factors. 2004;22(3):123-131.

[12] Kohara K, Yasuda H, Huang Y, Adachi N, Sohya K, Tsumoto T. A local reduction in cortical GABAergic synapses after a loss of endogenous brain-derived neurotrophic factor, as revealed by single-cell gene knock-out method. J Neurosci. 2007;27(27):7234-7244.

[13] Kuzumaki N, Ikegami D, Tamura R, et al. Hippocampal epigenetic modification at the brain-derived neurotrophic factor gene induced by an enriched environment. Hippocampus. 2011;21(2):127-132.

[14] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

[15] Johnson R, Zuccato C, Belyaeva NV, Guest DJ, Cattaneo E, Buckley NJ. A microRNA-based gene dysregulation pathway in Huntington's disease. Neurobiol Dis. 2008;29(3):438-445.

[16] De Felice C, Signorini C, Durand T, et al. F2-dihomo-isoprostanes as potential early biomarkers of lipid oxidative damage in Rett syndrome. J Lipid Res. 2011;52(12):2287-2297.

[17] Lioy DT, Garg SK, Monaghan CE, et al. A role for glia in the progression of Rett's syndrome. Nature. 2011;475(7357):497-500.