

INSTITUTO NACIONAL DE MEDICINA GENÓMICA 17
DE MARZO DEL 2022 – 17 DE SEPTIEMBRE DEL 2022

**ANÁLISIS DE DETERMINANTES TRANSCRIPCIONALES DE
LONGEVIDAD EN DIVERSOS TEJIDOS HUMANOS**

OMAR FRANCO RODRÍGUEZ | 2173064878

UNIVERSIDAD AUTÓNOMA METROPOLITANA UNIDAD
XOCHIMILCO

DIVISIÓN DE CIENCIAS BIOLÓGICAS Y DE LA SALUD
LICENCIATURA EN QUÍMICA FARMACÉUTICA BIOLÓGICA
ASESOR INTERNO: MIRANDA HERNÁNDEZ FRANCISCO
ASESOR EXTERNO: GUTIÉRREZ GONZÁLEZ HUMBERTO



Vo. Bo. Dr. Humberto Gutiérrez G.

Vo. Bo. Dr. Francisco Miranda H.

Resumen

Los genes humanos dirigen el desarrollo, mantenimiento y complejidad funcional del organismo, es decir, todas las funciones de las que es capaz un sistema vivo. Sin embargo, esto no es resultado de genes individuales actuando en aislamiento, sino de la colaboración de cientos o miles de genes actuando en conjunto y en estrecha coordinación funcional. Sin embargo, si una determinada función celular es en realidad el resultado de la colaboración de un conjunto definido de genes, en qué medida esta función se encuentra más relacionada con el nivel colectivo de expresión de estos en contraposición con su nivel de expresión coordinada (co-expresión), es algo que aún se desconoce.

Con objeto de responder esta pregunta, se utilizaron datos de expresión públicamente accesibles para evaluar si un conjunto de genes involucrados en una función bien definida (ciclo celular), muestran un nivel concomitadamente alto, ya sea de expresión colectiva o alternativamente un nivel alto de coexpresión, al comparar entre tejidos humanos con niveles bajos, medios y altos de actividad proliferativa.

Los resultados encontrados en esta investigación muestran que el nivel colectivo de expresión correlacionada de genes asociados al ciclo celular es un indicador más robusto de actividad proliferativa en tejidos humanos que el nivel absoluto de expresión colectiva. Este resultado sugiere que el nivel de actividad coordinada de genes asociados a una determinada función juega un papel más crítico en el reclutamiento funcional de la misma que el nivel absoluto de expresión colectiva del mismo ensamble de genes.

Marco Teórico

Como resultado de una interacción compleja de numerosos genes, se necesitan proteínas para llevar a cabo la mayoría de los procesos biológicos, incluidos formación de complejos moleculares que influyen en las propiedades físicas y químicas de numerosos metabolitos, alterando o regulando las propiedades secundarias y estructuras terciaria de otras proteínas, ensamblando señales complejas y circuitos reguladores o dirigiendo el montaje de grandes estructuras moleculares y anatómicas a nivel del organismo [1]. Estos complejos “cócteles” de proteínas, se requieren para

garantizar que las funciones biológicas se ensamblen cuidadosa, activa y controladamente para lograr funciones celulares y fisiológicas normales. Como resultado, cada proteína y, en consecuencia, cada gen, es tan vital para el éxito de una función como cualquier otro. El cómo interactúan dinámicamente los genes, a nivel de expresión, entre sí para instruir funciones complejas es algo que actualmente se entiende poco.

Un aspecto conocido de los perfiles globales de expresión génica en todos los organismos es que esta se encuentra organizada en subpoblaciones de genes, en donde decenas, cientos o incluso miles de genes muestran un patrón de expresión coordinada. A estas subpoblaciones de genes con expresión altamente correlacionada se les conoce como módulos de coexpresión

La relevancia funcional de estas agrupaciones o módulos de coexpresión se sugiere por el hecho de que grupos de genes conocidos por su involucramiento en alguna función común, tienden a mostrar también patrones de actividad altamente correlacionada.

Si bien se cree que esta estructura de coexpresión refleja, en cierto sentido, la arquitectura reguladora subyacente de la maquinaria genética, la fuente exacta de estos patrones coordinados de expresión sigue sin estar clara. Más específicamente, se supone comúnmente que la coexpresión es el resultado de genes que comparten reguladores comunes (es decir, dirigidos conjuntamente por factores de transcripción compartidos), sin embargo, esta suposición nunca se ha probado formalmente, ya que la expresión correlacionada también podría surgir en principio, incluso en ausencia de proximidad regulatoria [2].

Por otro lado, el nivel absoluto de expresión de genes individuales asociados a una determinada función se ha utilizado tradicionalmente como el indicador maestro de actividad de dicha función.

En este estudio se propone investigar la relación entre la organización funcional de un conjunto de genes que se sabe que están involucrados o relacionados a una función común (modelo) y su patrón colectivo de expresión. Específicamente, examinaremos los cambios en el nivel general de expresión, o el nivel general de actividad coordinada (coexpresión) de los genes asociados a la función de Ciclo Celular contrastando entre tejidos con una baja, media o alta actividad proliferativa.

Si bien el nivel absoluto de expresión génica se ha utilizado tradicionalmente como el indicador maestro del reclutamiento u organización funcionales en los genes individuales, se busca encontrar que métrica refleja mejor el aumento en la organización funcional de un conjunto de genes dados, ya que aún no se ha explorado esta área con profundidad.

Planteamiento del problema

Como ya se ha mencionado, los genes tienen un papel importante en el control de una gama muy amplia de procesos celulares y fisiológicos. Así también sabemos que el nivel de expresión de los genes individuales posee un lugar clave en la regulación de ciertas funciones celulares, sin embargo, somos conscientes de que estos genes no actúan de forma aislada y que cualquier función celular es el resultado de un gran número de estos genes actuando siempre en estrecha coordinación con los demás.

Este trabajo busca investigar cuál es el vínculo que existe entre el nivel de expresión genética, o bien el nivel de coordinación (coexpresión) por un lado, y el nivel de demanda de la función biológica en común por el otro, así como poder responder a la pregunta ¿Cuál de estos dos parámetros (nivel de expresión o nivel de coexpresión) es un mejor indicador de la activación de una función dada?

En cuanto a los genes, la función que se manejará como Modelo, será ciclo celular, (GO:0007049 por su caracterización y anotación en la base de datos Gene Ontology).

Todo esto será posible gracias a las herramientas de la Genómica Computacional con base a la cual se generarán diversos algoritmos de evaluación de expresión y coexpresión, para así obtener los resultados, tanto numéricos como gráficos requeridos para el análisis.

Objetivo General

Determinar si la actividad de una función celular, como es la función proliferativa, se encuentra más asociada a un nivel de expresión concreto; ya sea a la expresión colectiva de genes asociados al ciclo celular, o si en cambio esta se encuentra más fuertemente asociada al nivel colectivo de expresión coordinada (coexpresión).

Objetivos Particulares:

1. Buscar, encontrar y filtrar bases de datos en las que se puedan obtener registros de nivel de expresión de aproximadamente veinte mil genes humanos en una variedad de tejidos con diferentes niveles de actividad proliferativa.
2. Mediante una batería de herramientas computacionales, medir y comparar estadísticamente el nivel de expresión colectiva de un conjunto definido de genes involucrados en una función igualmente definida (ciclo celular), específicamente comparando entre tejidos no neuronales, tejido nervioso (cerebro) y datos de expresión derivados de líneas celulares.
3. Mediante una batería de herramientas computacionales, medir y comparar el nivel de coexpresión (correlación media) colectiva de un conjunto definido de genes involucrados en una función igualmente definida (ciclo celular), específicamente comparando entre tejidos no neuronales, tejido nervioso (cerebro) y datos de expresión derivados de líneas celulares.
4. Llevar a cabo un análisis basado en teoría de redes para determinar cambios en la topología de la red de coexpresión de los genes asociados a ciclo celular al contrastar tejidos con bajo, medio y alto nivel de actividad proliferativa.

Resultados

Se realizó la obtención de datos públicamente disponibles de expresión de 20 mil genes humanos derivados de la base de datos Brainspan [3], normalizados en lecturas por kilobase por millón (RPKM, por sus siglas en inglés).

Trabajamos únicamente con las muestras disponibles de corteza, cuerpo estriado y cerebelo, en los cuales se practicaron diferentes análisis para poder medir y comparar los niveles de expresión y co-expresión. La elección de estos tejidos corresponde a

tres regiones en que el contenido de células gliales (siendo éstas las células proliferativamente activas en contraste con neuronas diferenciadas), es substancialmente diferente, con la menor proporción presente en cerebelo, seguido de corteza y con la mayor proporción presente en el estriado [4]. En este estudio por lo tanto se asumirá que la actividad proliferativa en cada una de estas regiones se corresponde con la proporción de células gliales respecto a neuronas en cada una de ellas.

Los resultados obtenidos para el cálculo de expresión en los genes del ciclo celular se obtuvieron identificando dichos genes con base en su anotación a la categoría funcional “cell cycle”, del gene ontology project (*GO:0007049*), la lista obtenida fue de 1723 genes involucrados en esta función.

En un análisis preliminar del perfil general de expresión de las muestras de tejidos analizados, llevamos a cabo un examen de componentes principales para determinar si existe una diferencia global entre las muestras. Como se observa en el ejemplo de la figura 1, en que se compara el cerebelo con la corteza la comparación entre estos tejidos muestra un agrupamiento parcialmente separado en sus perfiles globales. Este resultado sugiere que existen diferencias aunque quizás tenues en el nivel global de expresión media entre tejidos con distinto nivel de actividad proliferativa.

Comparación de expresión en Cerebelo y Cuerpo Estriado:

Con objeto de determinar si los tejidos con mayor actividad proliferativa muestran un mayor nivel colectivo de expresión de genes asociados al ciclo celular, se identificaron, en las muestras de cerebelo y cuerpo estriado, los genes relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$), a estos se les determinó la mediana de expresión en cada grupo. Se tomaron 5000 muestras de genes aleatorios del mismo tamaño que el número de genes del ciclo celular presentes en los grupos de muestras ($n=1723$), y se les calculó la mediana, para así obtener la distribución de estas (Figura 2), dado que la distribución esperada se encuentran desplazada en las muestras aleatorias de una estructura con respecto a las del otro grupo de tejidos (Figura 2, paneles A y B), se optó por corregir la expresión total en ambos grupos de tejidos dividiendo por la respectiva mediana de expresión de todos los genes (*Mediana Global de expresión*), resultando así en distribuciones comparables (figura 2, panel C) . En este análisis se encontró que tanto las muestras correspondientes a estriado como a cerebelo, muestran un nivel colectivo de expresión mayor que el esperado al azar (distribución

aleatoria), sin embargo el cerebelo, con un menor contenido relativo de células proliferativas (glía), muestra una mayor actividad media de los genes del ciclo celular que el estriado, con un mayor contenido de células gliales y por lo tanto con una mayor actividad proliferativa.

Comparación de expresión en Cerebelo y Corteza:

De la misma forma se identificaron, en las muestras de cerebelo y corteza los genes relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$) y se les determinó la mediana de expresión en cada grupo. Se tomaron 5000 muestras de genes aleatorios del mismo tamaño que el número de genes del ciclo celular presentes en los grupos de muestras, y se les calculó la mediana, para así obtener la distribución esperada de éstas (*Figura 3*). Dado que las distribuciones esperadas se encuentran desplazadas una respecto a otra en ambos grupos de tejidos se optó por corregir ambos resultados en cada grupo, dividiendo por la respectiva mediana de expresión de todos los genes (*Mediana Global de expresión*), resultando así en distribuciones comparables (*Figura 3 C*). En este análisis se encontró que la expresión colectiva de los genes del ciclo celular en la corteza (con un mayor contenido de células gliales), es mayor que la del cerebelo.

Comparación de expresión en Corteza y Estriado:

Finalmente tomando los genes relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$) ya identificados en las muestras de tejidos corteza y estriado, se les determinó la mediana de expresión en cada grupo. Se tomaron 5000 muestras de genes aleatorios del mismo tamaño que el número de genes del ciclo celular presentes en los grupos de muestras ($n=1723$), y se les calculó la mediana, para así obtener la distribución de estas (*Figura 4; Panel A*). Nuevamente, dado que la distribución esperada se encuentra desplazada una respecto a la otra en ambos grupos de tejidos, se corrigieron los valores de expresión en ambos grupos dividiendo por la mediana de expresión del total de los genes en cada grupo (*Mediana Global de expresión*), resultando así en distribuciones comparables (*Figura 4 A y B*). En esta comparación se encontró un mayor nivel medio de expresión de los genes asociados al ciclo celular en la corteza que en el estriado, siendo que la corteza cuenta con una menor proporción de células gliales que en el estriado.

En estas tres comparaciones se observa que en sólo uno de los casos tiene lugar una mayor expresión colectiva de genes del ciclo celular en un tejido de mayor actividad proliferativa.

Resultados de Co-expresión: Para estos datos se usó el método “*Spearman*”, el cual es independiente de si la relación entre dos variables es o no lineal, procediéndose de la siguiente forma:

Medianas de co-expresión en Cerebelo vs. Cuerpo Estriado:

Se removieron los genes con varianza 0 a lo largo de las muestras de cerebelo y cuerpo estriado, con los genes identificados, los relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$ y 1684 respectivamente), se obtuvo la matriz de correlación entre todos estos genes, y se calculó la mediana de co-expresión (*Mediana del triángulo superior de su matriz de correlación*) para cada grupo. Se tomaron 5000 muestras de genes aleatorios ($n=1723$ y 1684 respectivamente) y de la misma forma se les aplicó correlación para obtener las medianas de co-expresión y su distribución. Dado que la distribución esperada se encuentran desplazada una respecto a la otra entre los grupos de tejidos a comparar, se optó por corregir los valores de co-expresión de todos los pares posibles de genes del ciclo celular, dividiendo por la respectiva mediana de co-expresión global entre todos los genes con varianza > 0 en cada grupo (*Mediana global de co-expresión*) resultando así en distribuciones comparables (Figura 5 A y B)

El resultado muestra una coexpersión colectiva mayor a la esperada al azar en ambos grupos de tejidos, siendo la coexpresión colectiva de las líneas celulares, mayor a la observada en tejidos nerviosos.

Medianas Aleatorias de co-expresión en Cerebelo vs Corteza:

Nuevamente, removiendo los genes con varianza 0 de las muestras de cerebro y corteza, y se identificaron los genes relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$ y 1684 respectivamente), se obtuvo la matriz de correlación entre todos estos genes, y se calculó la mediana de co-expresión (*Mediana del triángulo superior de su matriz de correlación*) para cada grupo. Se tomaron 5000 muestras de genes aleatorios ($n=1723$ y 1684 respectivamente) y de la misma forma se obtuvieron las medianas de co-expresión y su distribución esperada. Dado que la distribución esperada se

encuentran desplazada una respecto a la otra al comparar ambos grupos de tejidos, se optó por corregir ambos resultados dividiendo por la mediana de co-expresión global (todos los genes del transcriptoma) varianza > 0 de cada grupo de tejidos (*Mediana global de co-expresión*), resultando así en distribuciones comparables (Figura 6 A y B).

Medianas Aleatorias de co-expresión en Corteza vs Cuerpo Estriado;

Para terminar, se removieron los genes con varianza 0 en las muestras, se identificaron los genes relacionados a la función del ciclo celular (*GO:0007049*, $n=1723$ y 1684 respectivamente) y se obtuvo la matriz de correlación entre todos estos genes, y se calculó la mediana de co-expresión (*Mediana del triángulo superior de su matriz de correlación*) para cada grupo. Se tomaron 5000 muestras de genes aleatorios ($n=1723$ y 1684 respectivamente) para obtener las medianas de co-expresión y su distribución. Dado que las distribuciones esperadas se encuentran desplazadas una respecto a la otra en los dos grupos de tejidos, se optó por corregir ambos resultados dividiendo por la respectiva mediana de co-expresión de todos los genes con varianza > 0 en cada grupo (*Mediana global de co-expresión*), resultando así en distribuciones comparables (Figura 7 A y B).

Estos resultados en su conjunto muestran una mayor co-expresión colectiva de los genes involucrados en el ciclo celular en los tejidos de mayor actividad proliferativa (Cuerpo estriado > corteza > cerebelo).

logFC e Intercuartiles: Expresión vs. Co-expresión

logFC e Intercuartiles en análisis de Expresión y Co-expresión de Cerebelo vs

Cuerpo estriado:

Se realizó la prueba de Wilcoxon pareado para comprobar que hay una diferencia significativa, entre los genes del ciclo celular en común ($n=1723$ y $n=1684$ respectivamente) entre las 2 muestras comparadas (*Cerebelo y cuerpo estriado*), en su respectivo análisis (*Datos de expresión o co-expresión*), obteniendo el valor de p . También se calculó el Log Fold Change ($\log FC$), usando como base el grupo cuya mediana (*Expresión o Co-expresión*) sea más pequeña respecto al otro. Previamente se aplicó una corrección consistente en el valor de expresión de todos los genes un valor igual al inverso del FC máximo con objeto de eliminar valores de expresión igual a cero y así evitar la aparición de incrementos (FC) infinitos. . Se graficó en forma de abanico los cambios de expresión de cada uno de los genes del ciclo celular, representando en la punta, o base del abanico, al primer grupo (*El de Mediana menor*), y la dirección de cada línea representa el incremento o decremento (valor del $\log FC$) en la expresión de cada gen. De modo alternativo, se hizo una comparación gráfica entre medianas y rangos intercuartiles de estos genes en los grupos de muestras de cerebelo y cuerpo estriado, (*Expresión y Co-expresión respectivamente*) (Figura 8 A y B). A continuación se hizo una comparación similar utilizando la mediana de co-expresión de cada uno de los genes del ciclo celular con el resto de los mismos y se usaron estas medianas para llevar a cabo una comparación pareada y prueba de wilcoxon entre los dos grupos de tejidos aquí considerados. Los resultados se muestran en la figura 8 C y D.

Como puede observarse, se encontraron incrementos estadísticamente muy significativos tanto en expresión como coexpresión mediana en los tejidos de mayor actividad proliferativa (Líneas celulares > tejidos nerviosos).

logFC e Intercuartiles en análisis de Expresión y Co-expresión de Cerebelo vs Corteza

Usando un procedimiento similar al arriba descrito, se realizó el test de Wilcoxon, para comprobar que hay una diferencia significativa, entre los genes del ciclo celular en común ($n=1723$ y $n=1684$ respectivamente) entre las 2 muestras comparadas (*Cerebro y Corteza*), en su respectivo análisis (*Datos de expresión o co-expresión*), obteniendo el p -value. También se calculó el Log Fold Change ($\log FC$), usando como base el grupo

cuya mediana (*Expresión o Co-expresión*) sea más pequeña respecto al otro. (Figura 9 A y B). A continuación se hizo una comparación correspondiente utilizando la mediana de co-expresión de cada uno de los genes del ciclo celular con el resto de estos y se usaron estas medianas para llevar a cabo una comparación pareada y prueba de wilcoxon entre los dos grupos de tejidos aquí considerados. Los resultados se muestran en la figura 9 C y D.

Como puede observarse, se encontraron incrementos estadísticamente muy significativos tanto en expresión como co-expresión mediana en los tejidos de mayor actividad proliferativa (Cerebelo > Corteza).

logFC e Intercuartiles en análisis de Expresión y Co-expresión de Corteza vs Cuerpo Estriado:

Finalmente se llevó a cabo una comparación idéntica a las dos pruebas anteriores (Figura 10). Tomados en conjunto, estos resultados indican un incremento tanto en el nivel colectivo de expresión como en el de coexpresión de genes funcionalmente involucrados en el ciclo celular, en tejidos de mayor actividad proliferativa (Cuerpo estriado >> Corteza >> Cerevelo).

Generación de Redes Génicas

Con objeto profundizar en la caracterización de los cambios en la dinámica de co-expresión de los genes asociados al ciclo celular, al comparar tejidos de mayor y menor actividad proliferativa, llevamos a cabo un análisis de estos genes utilizando una aproximación de teoría de redes.

Red generada a partir de los Genes del Ciclo Celular:

Se tomó la matriz de co-expresión de las muestras de Cerebelo y se graficaron las interacciones entre pares de genes con correlaciones mayores a 0.8. En la red resultante, los nodos representan los genes y las aristas representan la existencia de una correlación mayor a 0.8. esta se desglosó en una lista de pares, donde se eliminaron los duplicados, con la lista de interacciones se realizó un corte en el umbral establecido (0.8) y se generó la red de Genes de Cerebro con las interacciones resultantes (figura 11).

Por el mismo procedimiento se construyeron las redes de coexpresión correspondientes a corteza y cuerpo estriado (figuras 12 y 13).

Como puede observarse en las figuras 11 a la 13, existe una reorganización en la estructura de la red de coexpresión de los genes del ciclo celular al comparar tejidos de menor y mayor actividad proliferativa, En particular se observa una reducción en el tamaño total de la red en los tejidos de más alta actividad proliferativa , pero esta reducción se encuentra acompañada de un aumento en la densidad (número total de interacciones) de la red resultante, lo cual, como se vio en los análisis precedentes resulta en un aumento neto en la coexpresión total de los genes del ciclo celular en tejidos de mayor actividad proliferativa.

Conclusión:

Los resultados señalan que el nivel colectivo de expresión correlacionada de los genes que identificamos como asociados a la función proliferativa (ciclo celular) resulta ser un indicador más robusto de la actividad proliferativa en los tejidos humanos que el nivel absoluto de la expresión colectiva. Este resultado sugiere que el nivel de actividad coordinada de los genes asociados a una determinada función juega un papel más crítico en el reclutamiento funcional de la misma que el nivel absoluto de expresión colectiva del mismo ensamble de genes.

Gráficas de Resultados

Análisis de Componentes principales:

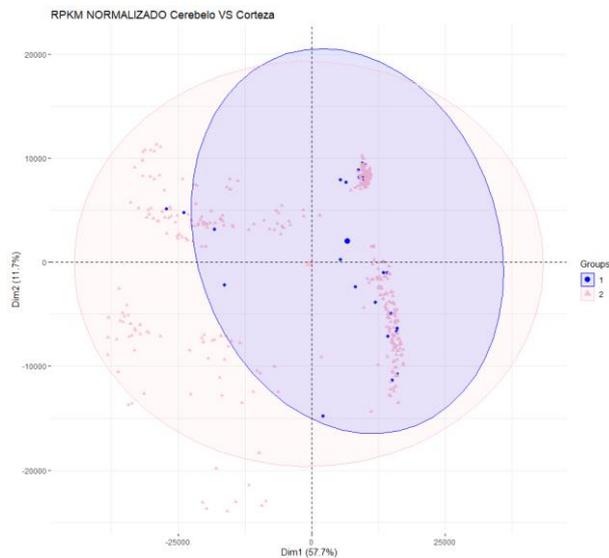


Figura 1. PCA de Cerebelo vs. Corteza; Realizamos el análisis de componentes principales aplicado únicamente a las muestras pertenecientes a cerebelo y corteza, estos normalizados por cuantiles, para así lograr ver una separación.

Datos de Expresión:

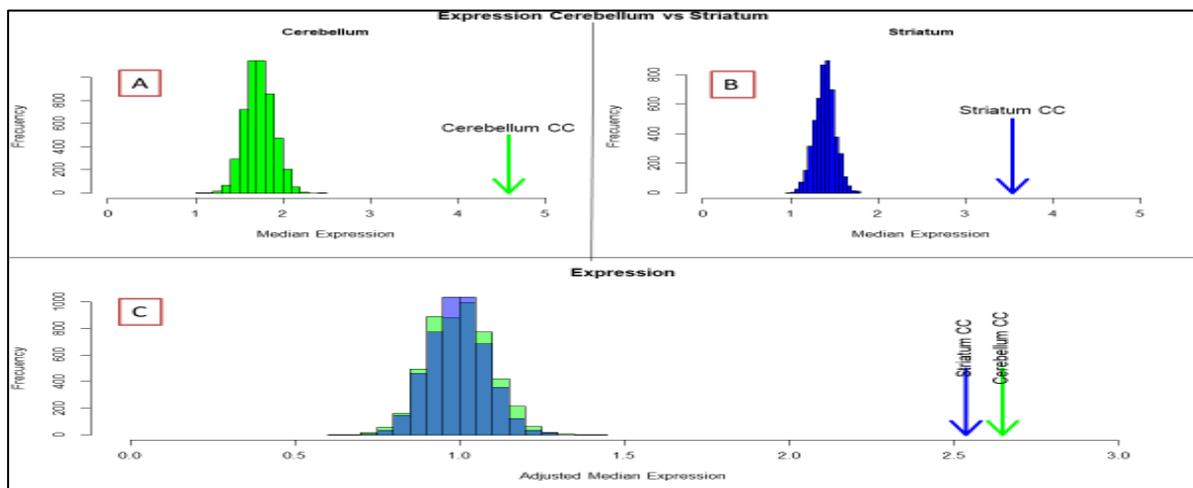


Figura 2. Distribución de Medianas Aleatorias de expresión (Cerebelo vs. Cuerpo Estriado);

A) y B) Se tomaron 5000 muestras aleatorias de genes del fondo (histogramas), para cada grupo de comparación, cerebelo (verde) y cuerpo estriado (azul) y se comparó la mediana de expresión de los genes del ciclo celular (flechas). **C)** Las medianas de expresión de los genes del ciclo celular se ajustaron, dividiéndola entre la mediana de las muestras aleatorias, con el fin de hacer las muestras comparables. Los gráficos ajustados se superpusieron para observar que grupo presentaba mayor expresión en genes del ciclo celular, en este grupo de comparación el cerebelo resulto tener una mediana mayor que la del cuerpo estriado.

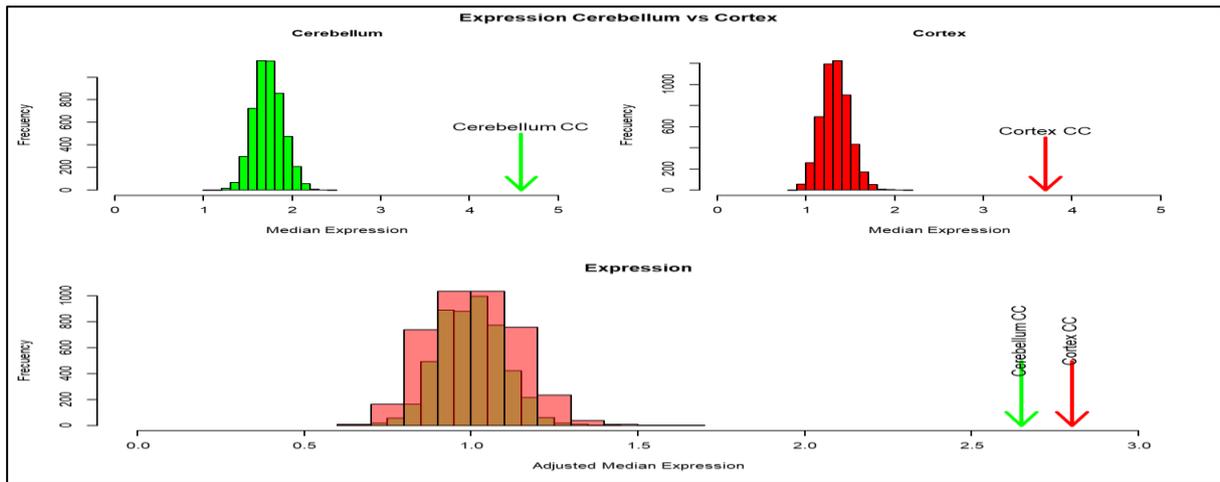


Figura 3. Distribución de Medianas Aleatorias de expresión (Cerebelo vs. Corteza);

A) y B) Se tomaron 5000 muestras aleatorias de genes del fondo (histogramas), para cada grupo de comparación, cerebelo (verde) y corteza (rojo) y se comparó la mediana de expresión de los genes del ciclo celular (flechas). **C)** Las medianas de expresión de los genes del ciclo celular se ajustaron, dividiéndola entre la mediana de las muestras aleatorias, con el fin de hacer las muestras comparables. Los gráficos ajustados se superpusieron para observar que grupo presentaba mayor expresión en genes del ciclo celular, en este grupo de comparación la corteza resulto tener una mediana mayor que la del cerebelo.

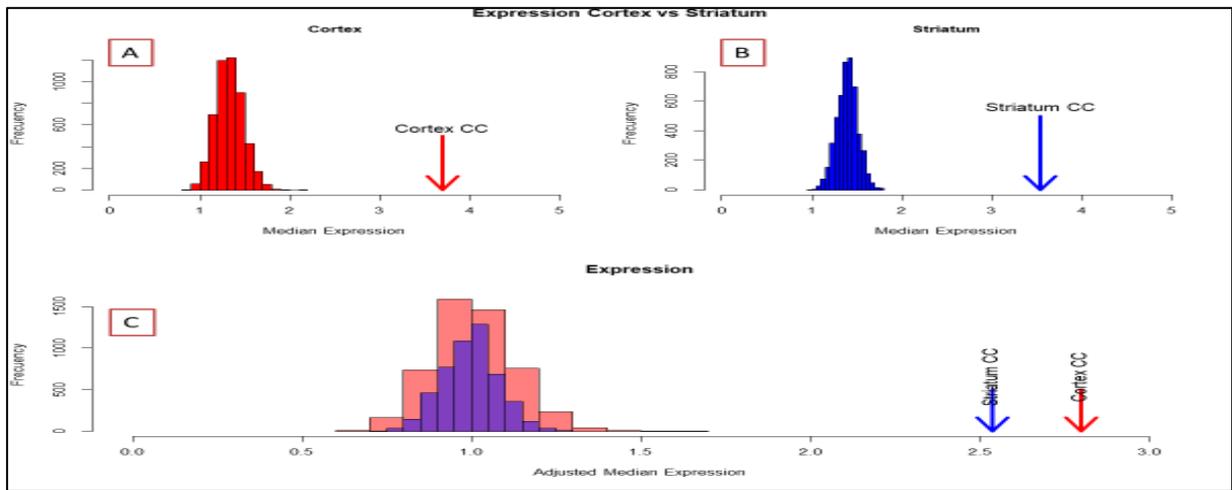


Figura 4. Distribución de Medianas Aleatorias de expresión (Corteza vs. Cuerpo Estriado);

A) y B) Se tomaron 5000 muestras aleatorias de genes del fondo (histogramas), para cada grupo de comparación, corteza (rojo) vs cuerpo estriado (azul) y se comparó la mediana de expresión de los genes del ciclo celular (flechas). **C)** Las medianas de expresión de los genes del ciclo celular se ajustaron, dividiéndola entre la mediana de las 5000 muestras aleatorias, con el fin de hacer las muestras comparables. Los gráficos ajustados se superpusieron para observar que grupo presentaba mayor expresión en genes del ciclo celular, en este grupo de comparación la corteza resulto tener una mediana mayor que la del cerebelo.

Datos de Co-Expresión:

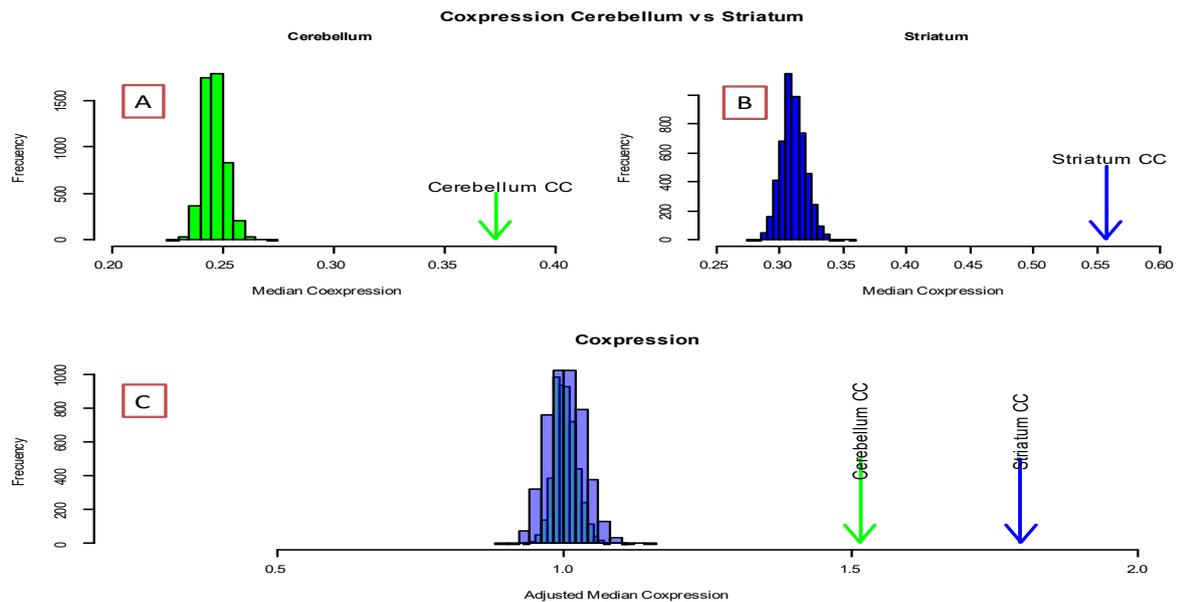


Figura 5. Distribución de Medianas Aleatorias de coexpresión (Cerebelo vs. Cuerpo Estriado); **A) y B)** Antes de realizar la coexpresión para cada grupo, se eliminaron los genes cuya varianza era igual a 0 y se interceptaron los genes del ciclo celular que compartían ambos grupos para obtener una muestra total de 1684 genes. Una vez realizada la correlación para cada grupo, se calculó la mediana de correlación (flechas) y se realizaron 5000 muestras aleatorias, tomando genes del fondo, para después graficarlas en un histograma por cada muestra (Cerebelo = verde y Cuerpo Estriado = azul) **C)** Las muestras aleatorias y medianas de los genes del ciclo celular se ajustaron dividiéndolas entre la mediana de las muestras aleatorias, con el fin de hacer las muestras comparables. Se logró observar que la mediana de los genes del ciclo celular es mayor en el cuerpo estriado comparado con la del cerebelo.

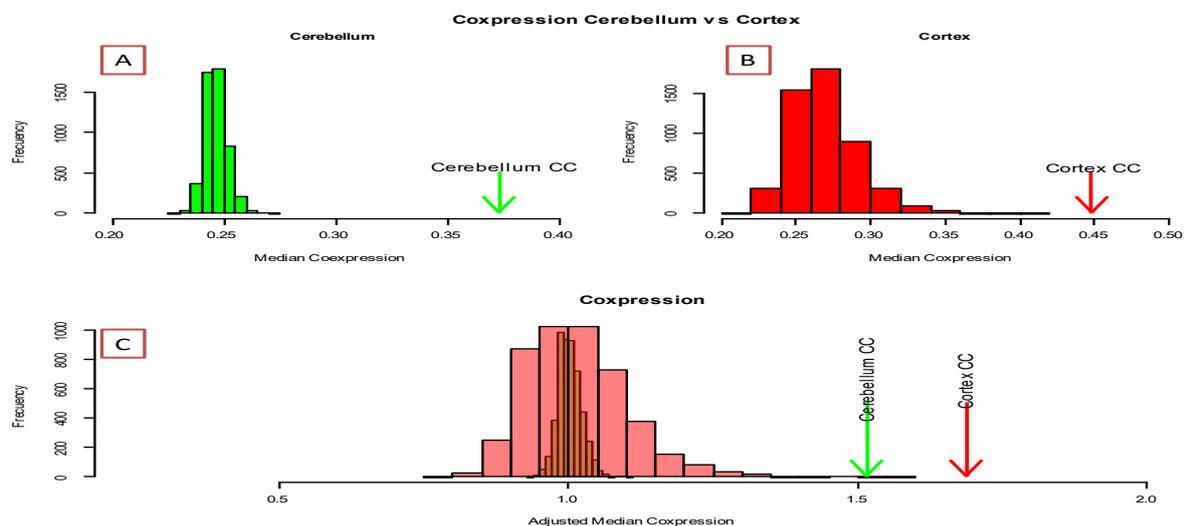


Figura 6. Distribución de Medianas Aleatorias de coexpresión (Cerebelo vs. Corteza); **A) y B)** Antes de realizar la coexpresión para cada grupo, se eliminaron los genes cuya varianza era igual a 0 y se interceptaron los genes del ciclo celular que compartían ambos grupos para obtener una muestra total de 1684 genes. Una vez realizada la correlación para cada grupo, se calculó la mediana de correlación (flechas) y se realizaron 5000 muestras aleatorias, tomando genes del fondo, para después graficarlas en un histograma por cada muestra (Cerebelo = verde y Corteza = rojo) **C)** Las muestras aleatorias y medianas de los genes del ciclo celular se ajustaron dividiéndolas entre la mediana de las muestras aleatorias, con el fin de hacer las muestras comparables. Se logró observar que la mediana de los genes del ciclo celular es mayor en la corteza comparado con la del cerebelo.

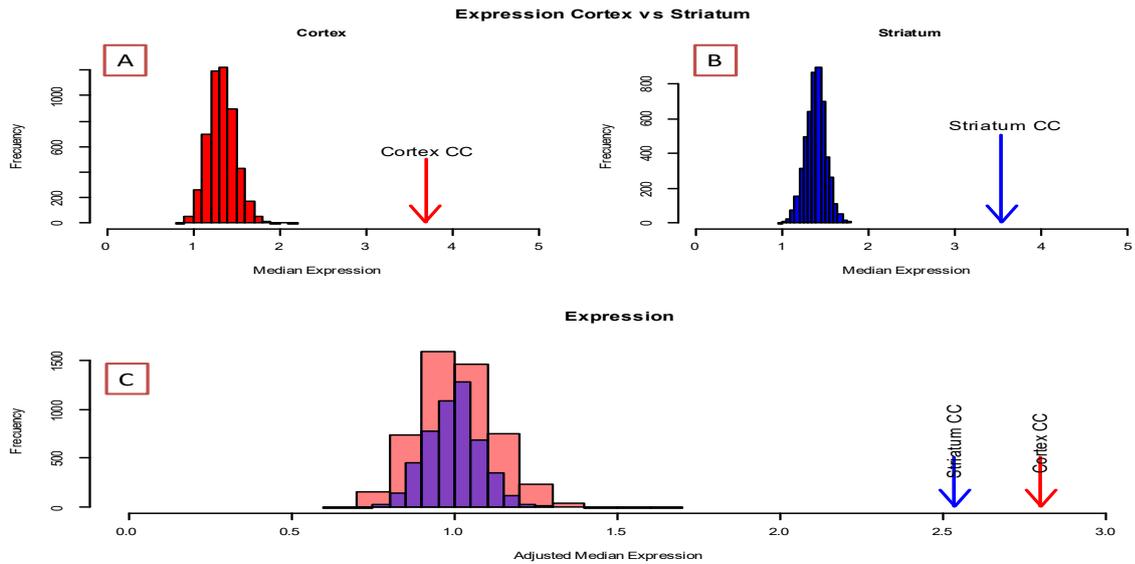


Figura 7. Distribución de Medianas Aleatorias de coexpresión (Cuerpo Estriado vs. Corteza); **A) y B)** Antes de realizar la coexpresión para cada grupo, se eliminaron los genes cuya varianza era igual a 0 y se interceptaron los genes del ciclo celular que compartían ambos grupos para obtener una muestra total de 1684 genes. Una vez realizada la correlación para cada grupo, se calculó la mediana de correlación (flechas) y se realizaron 5000 muestras aleatorias, tomando genes del fondo, para después graficarlas en un histograma por cada muestra (Cuerpo estriado = azul y Corteza = rojo) **C)** Las muestras aleatorias y medianas de los genes del ciclo celular se ajustaron dividiéndolas entre la mediana de las muestras aleatorias, con el fin de hacer las muestras comparables. Se logró observar que la mediana de los genes del ciclo celular es mayor en la corteza comparado con la del cuerpo estriado.

Logaritmo Fold Change (logFC): Expresión vs. Coexpresión

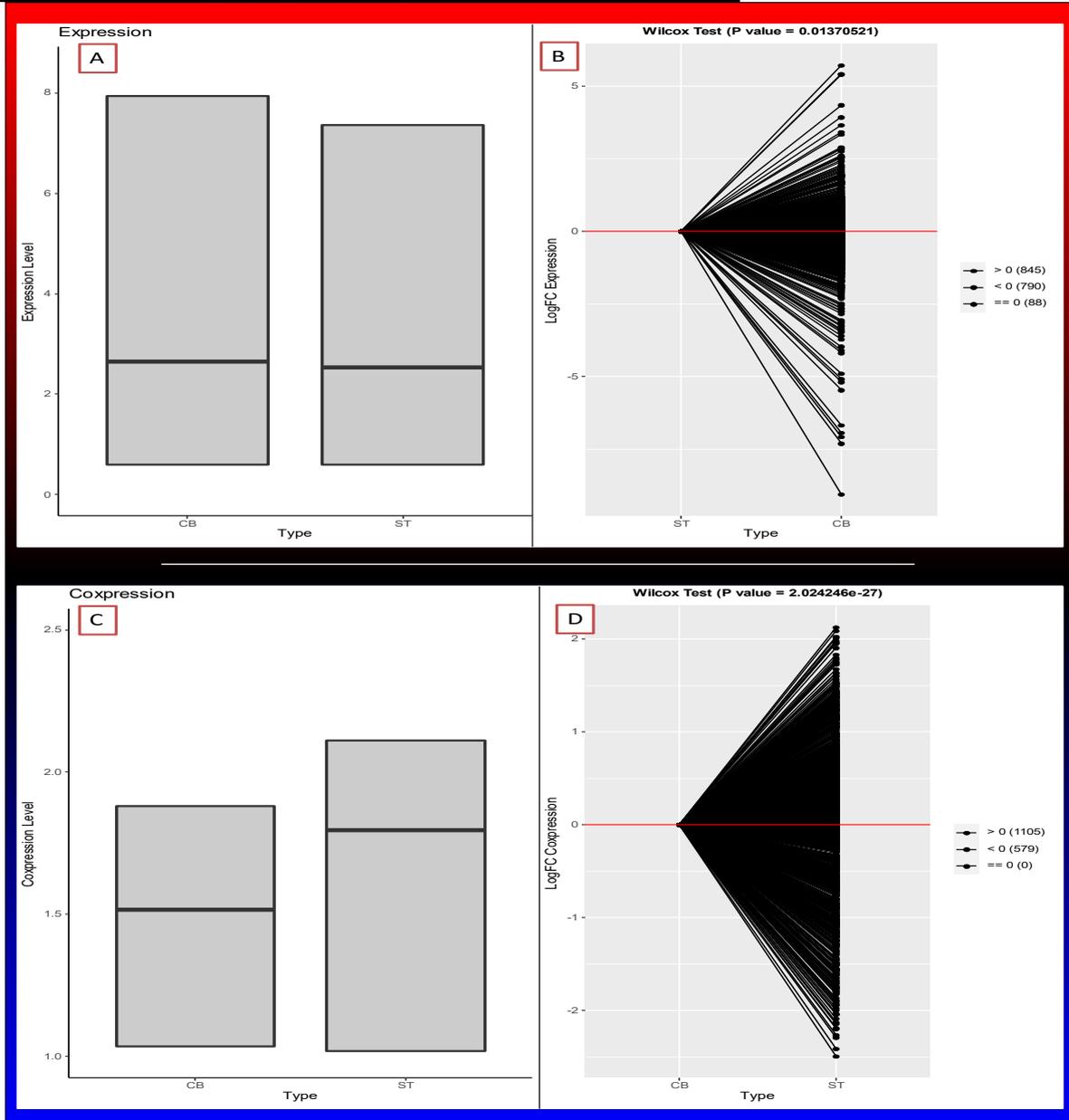


Figura 8. logFC e Intercuartiles en Exp. vs Co-exp. (Cerebelo vs. Cuerpo Estriado);

A) y B) Se hace una comparación entre los intercuartiles de los genes en los grupos de muestras del cerebelo y cuerpo estriado, mostrando también el valor de la mediana (*expresión y coexpresión respectivamente*) **B y D)** Se realizó la prueba de Wilcoxon para comprobar que hay una diferencia significativa entre los genes del ciclo celular en común ($n=1723$ y $n=1684$ respectivamente) entre las 2 muestras comparadas (Cerebelo y Cuerpo estriado), en su respectivo análisis (*Datos de expresión o coexpresión*), el resultado del p-value se muestra en la parte superior. Después se calculó el Log FC con respecto al primer grupo de comparación mostrado en la figura. Previamente aplicando una corrección consistente en: sumar al valor de cada gen el inverso del FC máximo para evitar valores de 0 que resultarían en cambios infinitos. Cada línea en los paneles B y D representa el aumento o decremento en proporción de expresión o coexpresión de cada uno de los genes del CC analizados en las muestras de cerebelo y cuerpo estriado. Los valores de p se encuentran en la parte superior de las figuras B y D

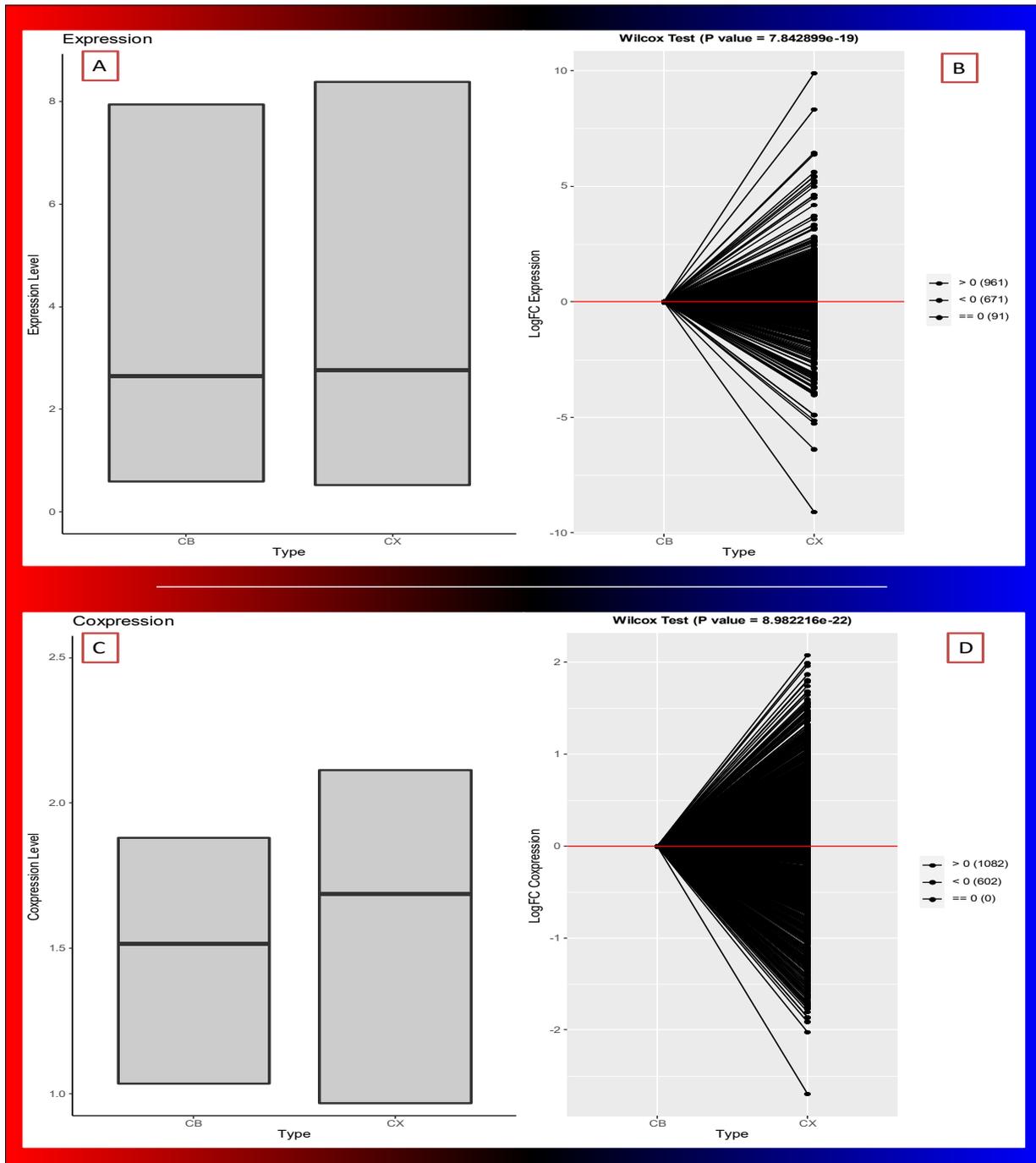


Figura 9. logFC e Intercuartiles en Exp. vs Co-exp. (Cerebelo vs. Corteza);

A) y B) Se hace una comparación entre los intercuartiles de los genes en los grupos de muestras del cerebelo y corteza, mostrando también el valor de la mediana (*expresión y coexpresión respectivamente*)

B y D) Se realizó la prueba de Wilcoxon para comprobar que hay una diferencia significativa entre los genes del ciclo celular en común ($n=1723$ y $n=1684$ respectivamente) entre las 2 muestras comparadas (Cerebelo y Corteza), en su respectivo análisis (*Datos de expresión o coexpresión*), el resultado del p-value se muestra en la parte superior. Después se calculó el Log FC con respecto al primer grupo de comparación mostrado en la figura. Previamente aplicando una corrección consistente en: sumar al valor de cada gen el inverso del FC máximo para evitar valores de 0 que resultarían en cambios infinitos. Cada línea en los paneles B y D representa el aumento o decremento en proporción de expresión o coexpresión de cada uno de los genes del CC analizados en las muestras de cerebelo y corteza. Los valores de p se encuentran en la parte superior de las figuras B y D

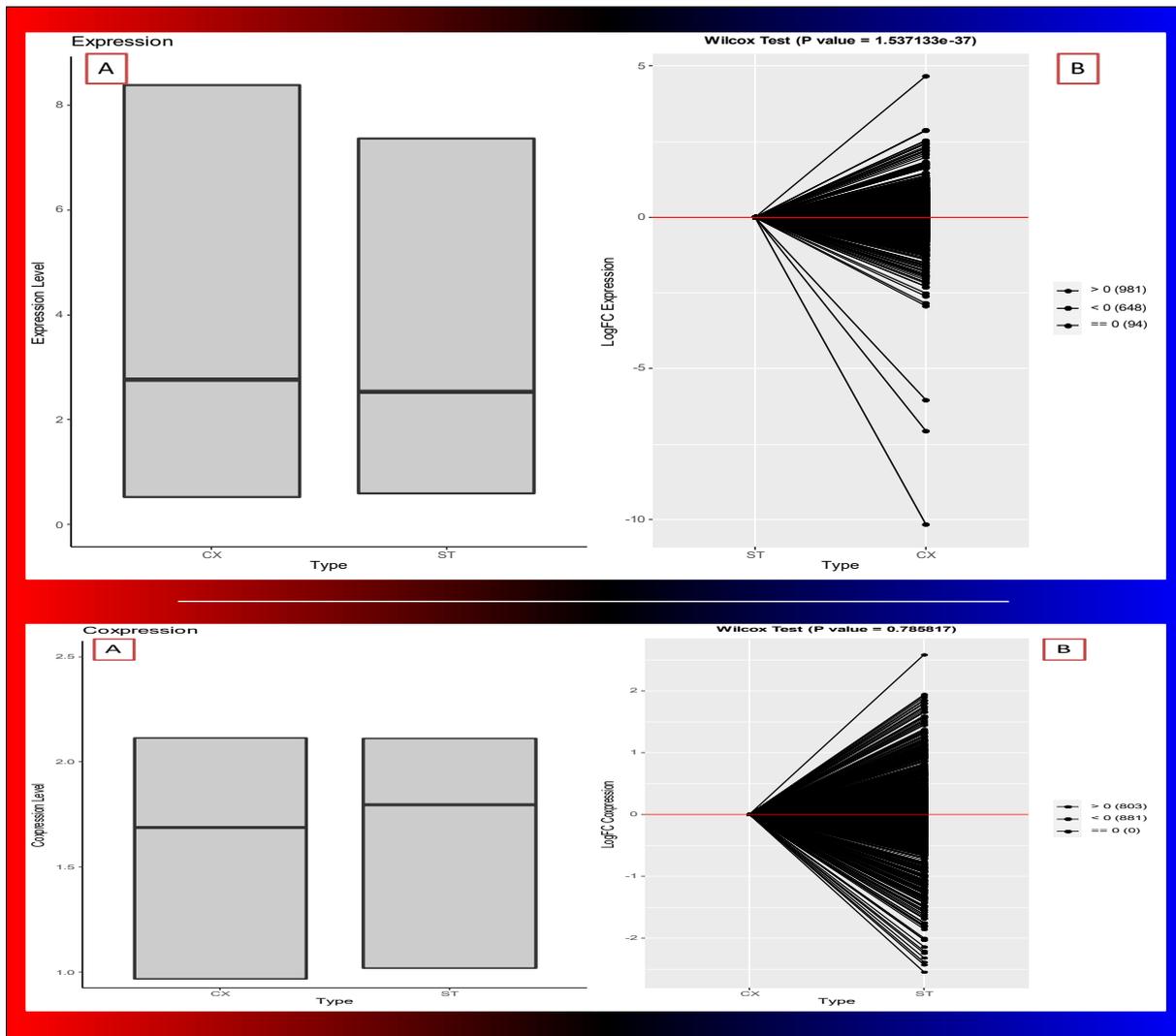


Figura 10. logFC e Intercuartiles en Exp. vs Co-exp. (Corteza vs Cuerpo Estriado);
A) y B) Se hace una comparación entre los intercuartiles de los genes en los grupos de muestras del corteza y cuerpo estriado, mostrando también el valor de la mediana (*expresión y coexpresión respectivamente*)
B y D) Se realizó la prueba de Wilcoxon para comprobar que hay una diferencia significativa entre los genes del ciclo celular en común ($n=1723$ y $n=1684$ respectivamente) entre las 2 muestras comparadas (Cerebelo y Corteza), en su respectivo análisis (*Datos de expresión o coexpresión*), el resultado del p-value se muestra en la parte superior. Después se calculó el Log FC con respecto al primer grupo de comparación mostrado en la figura. Previamente aplicando una corrección consistente en: sumar al valor de cada gen el inverso del FC máximo para evitar valores de 0 que resultarían en cambios infinitos. Cada línea en los paneles B y D representa el aumento o decremento en proporción de expresión o coexpresión de cada uno de los genes del CC analizados en las muestras de corteza y cuerpo estriado. Los valores de p se encuentran en la parte superior de las figuras B y D

Redes: Genes asociados a la función del ciclo celular en diferentes condiciones

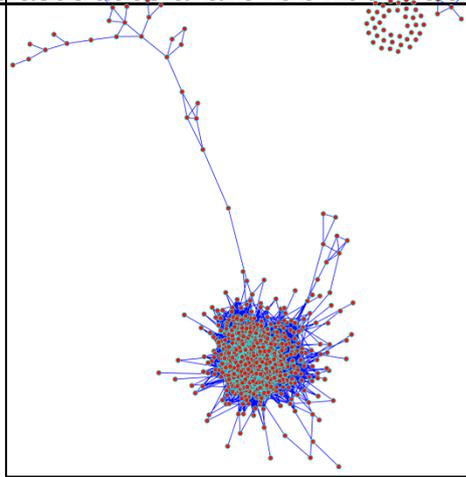


Figura 11. Redes obtenidas a partir de su matriz de coexpresión de los genes del ciclo celular y cuyo valor es mayor a 0.8 en Cerebelo



Figura 12. Red de coexpresión de genes del ciclo celular cuyo valor es mayor a 0.8

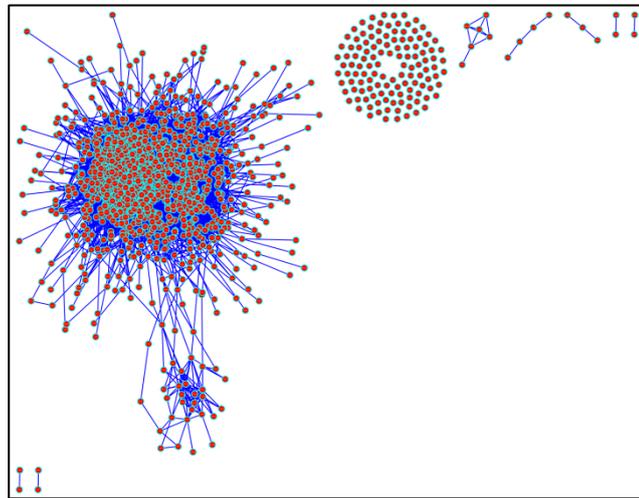


Figura 13. Redes obtenidas a partir de su matriz de coexpresión de los genes del ciclo celular y cuyo valor es mayor a 0.8 en Cuerpo Estriado

Referencias

- [1] Tolia NH, Joshua-Tor L. Strategies for protein coexpression in *Escherichia coli*. *Nat Methods*. 2006 Jan;3(1):55-64. doi: 10.1038/nmeth0106-55. PMID: 16369554.
- [2] Yin W, Mendoza L, Monzon-Sandoval J, Urrutia AO, Gutierrez H. Emergence of co-expression in gene regulatory networks. *PLoS One*. 2021 Apr 1;16(4):e0247671. doi: 10.1371/journal.pone.0247671. PMID: 33793561; PMCID: PMC8016302.
- [3] Brainspan. (s. f.). Atlas of the developing human brain. Recuperado 20 de marzo de 2022, de <https://www.brainspan.org/rnaseq/search/index.html>
- [4] Agustino Martínez, A. (2014). *Revista Hypatia*. Recuperado el 28 de marzo de 2022, de Hypatia: <https://revistahypatia.org/genomica-computacional.html>
- [5] Monroy Jaramillo, N., & Alonso Vilatela, M. E. (2011). La Influencia de los genes en el envejecimiento. *Revista de la Academia Mexicana de Ciencias #62* (enero-marzo), 26-31.
- [6] Castillo-Morales A, Monzón-Sandoval J, Urrutia AO, Gutiérrez H. Postmitotic cell longevity-associated genes: a transcriptional signature of postmitotic maintenance in neural tissues. *Neurobiol Aging*. 2019 Feb;74:147-160. doi: 10.1016/j.neurobiolaging.2018.10.015. Epub 2018 Oct 19. PMID: 30448614.
- [7] Seim I, Ma S, Gladyshev VN. Gene expression signatures of human cell and tissue longevity. *NPJ Aging Mech Dis*. 2016 Jul 7;2:16014. doi: 10.1038/npjamd.2016.14. PMID: 28721269; PMCID: PMC5514998.
- [8] Allocco DJ, Kohane IS, Butte AJ (2004) Quantifying the relationship between co-expression, co regulation and gene function. *BMC bioinformatics* 5: 18. <https://doi.org/10.1186/1471-2105-5-18> PMID: 15053845
- [9] Marco A, Konikoff C, Karr TL, Kumar S (2009) Relationship between gene co-expression and sharing of transcription factor binding sites in *Drosophila melanogaster*. *Bioinformatics* 25: 2473–2477. <https://doi.org/10.1093/bioinformatics/btp462> PMID: 19633094

- [10]Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4. <https://doi.org/10.2202/1544-6115.1128>, PMID: 16646834
- [11]Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, et al. (2011) Spatio-temporal transcriptome of the human brain. *Nature* 478: 483. <https://doi.org/10.1038/nature10523> PMID: 22031440
- [12]Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, et al. (2008) Functional organization of the transcriptome in human brain. *Nature neuroscience* 11: 1271. <https://doi.org/10.1038/nn.2207> ; PMID: 18849986