

UACM

Universidad Autónoma
de la Ciudad de México

Nada humano me es ajeno

UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO

COLEGIO DE CIENCIA Y TECNOLOGÍA
LICENCIATURA EN CIENCIAS GENÓMICAS

**BASES GENÓMICAS DE LA EVOLUCIÓN DEL CEREBRO
A PARTIR DEL NÚMERO DE NEURONAS EN
CORTEZA CEREBRAL**

TESIS

QUE PARA OPTAR POR EL TÍTULO DE LICENCIADO EN:

CIENCIAS GENÓMICAS

PRESENTA:

JAIME TENOPALA RODRÍGUEZ

DIRECTORA DE TESIS:

DRA. ARAXI URRUTIA ODABACHIAN

CODIRECTORA:

DRA. PAOLA EVERARDO MARTÍNEZ

**POSGRADO
EN CIENCIAS
GENÓMICAS**

Ciudad de México, marzo de 2026

SISTEMA BIBLIOTECARIO DE INFORMACIÓN Y DOCUMENTACIÓN



UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO COORDINACIÓN ACADÉMICA

RESTRICCIONES DE USO PARA LAS TESIS DIGITALES

DERECHOS RESERVADOS[©]

La presente obra y cada uno de sus elementos está protegido por la Ley Federal del Derecho de Autor; por la Ley de la Universidad Autónoma de la Ciudad de México, así como lo dispuesto por el Estatuto General Orgánico de la Universidad Autónoma de la Ciudad de México; del mismo modo por lo establecido en el Acuerdo por el cual se aprueba la Norma mediante la que se Modifican, Adicionan y Derogan Diversas Disposiciones del Estatuto Orgánico de la Universidad de la Ciudad de México, aprobado por el Consejo de Gobierno el 29 de enero de 2002, con el objeto de definir las atribuciones de las diferentes unidades que forman la estructura de la Universidad Autónoma de la Ciudad de México como organismo público autónomo y lo establecido en el Reglamento de Titulación de la Universidad Autónoma de la Ciudad de México.

Por lo que el uso de su contenido, así como cada una de las partes que lo integran y que están bajo la tutela de la Ley Federal de Derecho de Autor, obliga a quien haga uso de la presente obra a considerar que solo lo realizará si es para fines educativos, académicos, de investigación o informativos y se compromete a citar esta fuente, así como a su autor ó autores. Por lo tanto, queda prohibida su reproducción total o parcial y cualquier uso diferente a los ya mencionados, los cuales serán reclamados por el titular de los derechos y sancionados conforme a la legislación aplicable.

UNIVERSIDAD AUTÓNOMA DE LA CIUDAD DE MÉXICO

PLANTEL DEL VALLE

INTEGRACIÓN DE JURADO

Presidente:	Dra. Ángeles Concepción Tecalco Cruz Posgrado de Ciencias Genómicas, UACM
Secretario:	Dra. Paola Everardo Martínez Posgrado de Ciencias Genómicas, UACM
Vocal:	Dra. Araxi Urrutia Odabachian Instituto de Ecología UNAM

DIRECTORA

DRA. ARAXI URRUTIA ODABACHIAN
INSTITUTO DE ECOLOGÍA, UNAM

CODIRECTORA

DRA. PAOLA EVERARDO MARTÍNEZ
POSGRADO DE CIENCIAS GENÓMICAS, UACM

AGRADECIMIENTOS

El presente trabajo fue elaborado bajo la colaboración de diversas personas e instituciones, a quien agradezco profundamente su ayuda y apoyo incondicional durante mi formación académica.

Primeramente, agradezco a la **Universidad Autónoma de la Ciudad de México**, por permitirme el acceso a la educación superior, así como a la formación académica y profesional adquirida durante estos años.

A todas y todos los profesores del posgrado de ciencias genómicas de la UACM. Por su pasión por la enseñanza, por compartir su conocimiento con paciencia y dedicación, y por inspirarme a cuestionar, aprender y crecer. Cada lección impartida, cada consejo y cada reto académico dejaron una huella invaluable en mi formación. Gracias por fomentar mi curiosidad, y por enseñarme que el verdadero aprendizaje va más allá de los libros.

Al instituto de **Ecología de la UNAM** y en especial al **laboratorio de genómica evolutiva y funcional**, por brindarme y permitirme las facilidades de realizar el presente trabajo en sus instalaciones.

A la **Escuela Nacional de Antropología e Historia** y en especial al **laboratorio de genética molecular**, por permitir el continuo desarrollo académico, por medio del acceso a sus instalaciones, laboratorio, materiales y reactivos, para realizar prácticas de laboratorio de los trabajos de investigación importantes para mi formación académica.

A la **Dra. Araxi Urrutia Odabachian** por permitirme desarrollar la presente investigación en su laboratorio, bajo su dirección y supervisión, además de su paciencia y apoyo a lo largo de este camino en mi formación profesional. Gracias por compartir su conocimiento, por sus valiosos consejos y por motivarme a superar cada desafío con determinación. Le admiro y le estaré muy agradecido.

A la **Dra. Paola Everardo Martínez** por su ayuda, apoyo, e invaluable orientación, no solo durante el desarrollo de este trabajo, sino durante mi formación académica. Su dedicación, paciencia y compromiso han sido fundamentales en mi crecimiento como estudiante y como persona. Gracias por compartir su conocimiento con tanta generosidad, por su apoyo en los momentos de duda y por impulsarme siempre a dar lo mejor de mí. Le tengo una profunda admiración y agradecimiento.

A la **Dra. Alin Patricia Acuña Alonzo** a quien le estoy muy agradecido por la paciencia y apoyo incondicional durante este camino. Por su disponibilidad y ayuda en todo

momento, y aún más la disposición para compartir su conocimiento, resolver dudas y brindar siempre una palabra de aliento. Gracias por su tiempo y su amistad.

A las lectoras de este trabajo **Dra. Ángeles Concepción Tecalco Cruz, Dra. Claudia Selene Zárate Guerra y Dra. Helena Solleiro Villavicencio**, gracias por su valiosa revisión crítica, sus comentarios constructivos y su tiempo dedicado a fortalecer la calidad académica de este trabajo.

A mis amigos de la universidad **Kathia, Berenice, Karina, Yessica, Paulina, Javier, José Ángel, Kevin, Mario, Ricardo y Jorge** compañeros de este viaje de aprendizaje y crecimiento. Gracias por las risas en los momentos de estrés, por las conversaciones que desafiaron nuestras ideas. Cada uno de ustedes ha sido un pilar en este camino, y sin su apoyo, esta etapa no habría sido la misma. Que la ciencia, el conocimiento y la amistad sigan uniendo nuestros caminos.

A mis compañeros del laboratorio **Amairani, Karénina y Benjamín** gracias por su apoyo incondicional, por los consejos en los momentos difíciles y por hacer que cada día fuera una experiencia enriquecedora. Que la curiosidad y el espíritu de colaboración sigan guiando nuestro camino.

A mis amigos de vida **Diego, Ricardo, Ing. Juan, Rubén, Víctor y Neri**, porque más allá de la universidad, el laboratorio y la ciencia, en mi vida personal han sido mi refugio, mi apoyo incondicional y mi mayor fuente de alegría. Gracias por estar en cada momento, por las risas interminables, por los consejos sinceros y por demostrarme que la verdadera amistad trasciende el tiempo y la distancia. Este logro también es suyo, porque sin su compañía y aliento, el camino habría sido mucho más difícil.

DEDICATORIA

"Dedico este trabajo a quienes caminaron a mi lado durante esta travesía. A los viajes que expandieron mis límites y a la incertidumbre que, lejos de detenerme, me enseñó a caminar con paso firme. En este camino, el esfuerzo fue mi motor y la constancia el mapa para no perderme en las dudas. No solo honro aquí mis logros, si no sobre todo mis derrotas, pues en el silencio de las caídas hallé el aprendizaje que me permitió continuar. Hoy celebro con felicidad esta meta alcanzada, pero teniendo claro que el camino no termina aquí: este es solo el preludio de nuevos horizontes, nuevos desafíos y todos los viajes que aún restan por vivir"

A mi mamá **Mary**, dedico este logro, te agradezco por tu apoyo incondicional durante esta travesía, que con el gran amor y devoción que has demostrado a nosotros tus hijos, tienes la fortaleza de salir adelante sin importar los obstáculos que te ha dado la vida, y a pesar de eso habernos formado como personas de bien, por ser mi guía, mi inspiración, por ser la mujer que me dio la vida y que me enseñó a vivirla.

A mi hermana **Natxielly**, este logro también te lo dedico a ti, y aunque muchas veces pareciera que estuviéramos enojados y peleados, siempre había momentos donde la tempestad desaparecía y nos enseñaba el valor de la unión entre hermanos, como sabiamente nos decía mamá: "Apóyense entre ustedes y manténgase unidos para poder alcanzar todos sus objetivos". Gracias no solo por ser parte fundamental de este gran logro, sino que también por todos aquellos momentos bonitos que pasamos a lo largo de todos estos años y por estar cerca incluso en la distancia.

A mis queridas sobrinas **Alison** y **Gisel**. A ti, Alison, que con tu inteligencia y curiosidad me sigues inspirando a seguir aprendiendo y a nunca dejar de cuestionar. Que nunca pierdas la pasión por descubrir el mundo. Y a ti, Gisel, que con tu sonrisa y asombro de todo lo que te rodea, me recuerdas la magia de la vida y el valor de cada pequeño descubrimiento. Que tu camino esté siempre lleno de conocimiento y felicidad.

A mi cuñado **Luis**, valoro todo lo que has hecho por esta familia, tu apoyo ha marcado la diferencia y saber que contamos contigo me da mucha tranquilidad eres una gran persona y un pilar fundamental en nuestra familia.

ÍNDICE DE CONTENIDO

ABREVIATURAS.....	10
RESUMEN	11
I. INTRODUCCIÓN	12
1.1. LA CORTEZA CEREBRAL Y SU PAPEL EN LA EVOLUCIÓN DEL CEREBRO	13
1.2. UNA MIRADA EVOLUTIVA: DEL TELENCEFALO A LA CORTEZA CEREBRAL	13
1.3. LA EXPANSIÓN CORTICAL COMO EJE DE VARIABILIDAD EVOLUTIVA	15
1.4. MÁS ALLÁ DEL TAMAÑO: ¿QUÉ DETERMINA LA EVOLUCIÓN CEREBRAL?	21
1.5. GENES Y EVOLUCIÓN DEL CEREBRO	25
1.6 GLÍA COMO FACTOR EVOLUTIVO DEL CEREBRO	27
II. ANTECEDENTES PARTICULARES.....	28
III. HIPÓTESIS	33
IV. JUSTIFICACIÓN	34
V. OBJETIVOS	35
5.1. OBJETIVO GENERAL.....	35
5.2. OBJETIVOS ESPECÍFICOS	35
VI. ESTRATEGIA EXPERIMENTAL	36
VII. MATERIALES Y MÉTODOS.....	37
7.1. DATOS DE COMPOSICIÓN CELULAR DEL CEREBRO Y DE GENOMAS SECUENCIADOS EN MAMÍFEROS.....	37
7.2. ANOTACIONES DE FAMILIAS DE GENES	38
7.3. REGRESIONES FILOGENÉTICAS ENTRE EL TAMAÑO DE LAS FAMILIAS DE GENES Y LOS PARÁMETROS DE COMPOSICIÓN CELULAR	40
7.3.1 Caracterización funcional de familias génicas mediante dominios proteicos en genes	42
7.4. ANÁLISIS DE ENRIQUECIMIENTO FUNCIONAL EN TÉRMINOS DE ONTOLOGÍA GÉNICA	42
7.5. ANÁLISIS DE EXPRESIÓN GÉNICA EN EL CEREBRO CON RESPECTO A OTROS TEJIDOS CORPORALES A PARTIR DE DATOS DEL PROYECTO <i>FANTOM5</i>	43
7.6. ANÁLISIS DE EXPRESIÓN GÉNICA EN LA CORTEZA CEREBRAL EN COMPARACIÓN CON OTRAS REGIONES EN CEREBRO	44
7.7. ANÁLISIS DE EXPRESIÓN GÉNICA ESPACIO-TEMPORAL	45
7.8. CARACTERIZACIÓN DE GENES PERTENECIENTES A FAMILIAS ASOCIADAS A CORTEZA CEREBRAL	46
7.9. ANÁLISIS DE ENRIQUECIMIENTO FUNCIONAL DE REDES DE INTERACCIÓN PROTEÍNA- PROTEÍNA	48
VIII. RESULTADOS	50
8.1 ANÁLISIS DE REGRESIÓN LINEAL CON CORRECCIÓN FILOGENÉTICA Y CARACTERIZACIÓN FUNCIONAL DE FAMILIAS DE GENES ASOCIADAS CON CAMBIOS EN LA COMPOSICIÓN CELULAR EN CORTEZA CEREBRAL	50
8.1.1. Caracterización funcional de familias génicas mediante dominios proteicos en genes	53
8.1.2. Enriquecimiento funcional en términos de <i>Gene Ontology (GO)</i> entre los GFS de neuronas corticales y de glía cortical: resultados del PGLS doble.....	56
8.1.3. Enriquecimiento funcional en términos de <i>Gene Ontology (GO)</i> entre los GFS del número de neuronas corticales y la masa cortical relativa: resultados del PGLS doble	57
8.1.4. Enriquecimiento funcional en términos de <i>Gene Ontology (GO)</i> entre los GFS, del número de neuronas corticales y el número de neuronas en cerebro completo: resultados del PGLS doble	58
8.2. CARACTERIZACIÓN DE LA EXPRESIÓN GÉNICA ESPACIO-TEMPORAL DE GENES PERTENECIENTES A FAMILIAS DE GENES ASOCIADAS A CORTEZA CEREBRAL CON RESPECTO A OTROS TEJIDOS DEL CUERPO HUMANO	59
8.2.1. Expresión génica en el cerebro, de genes asociados al número de neuronas en corteza cerebral, en comparación con otros tejidos del cuerpo humano, en etapas fetal y adulta	60
8.3. CARACTERIZACIÓN DE LA EXPRESIÓN GÉNICA ESPACIO-TEMPORAL DE GENES PERTENECIENTES A FAMILIAS DE GENES ASOCIADAS A LA CORTEZA CEREBRAL CON RESPECTO A OTRAS REGIONES CEREBRALES.....	61
8.3.1. Expresión génica en la corteza cerebral frente a otras regiones del cerebro, para genes asociados al número de neuronas en corteza cerebral, en etapas fetal, infantil y adulta	61

8.4. ETAPAS CLAVE DEL DESARROLLO CEREBRAL CON MAYOR EXPRESIÓN DE GENES ASOCIADOS A NEURONAS CORTICALES	63
8.4.1. Dinámica espacio-temporal de la expresión génica de genes pertenecientes a familias asociadas significativamente a neuronas en corteza cerebral	63
8.5. CARACTERIZACIÓN FUNCIONAL Y GENÓMICA DE LOS GENES PERTENECIENTES A FAMILIAS ASOCIADAS A CORTEZA CEREBRAL	64
8.5.1. Distribución genómica de los genes pertenecientes a familias de genes asociadas significativamente a neuronas en corteza cerebral	65
8.6. IDENTIFICACIÓN DE ENRIQUECIMIENTO FUNCIONAL DE REDES DE INTERACCIÓN PROTEÍNA- PROTEÍNA	69
8.6.1. Vía de señalización de los receptores de la superficie celular	70
8.6.2. Proceso del sistema inmune	73
8.6.3. Respuesta inmune adaptativa	75
8.6.4. Respuesta inflamatoria	77
8.6.5. Respuesta a estímulos externos	78
8.6.6. Respuesta al estrés	80
IX. DISCUSIÓN	83
9.1 PROPUESTA DE MODELO INTEGRADOR DE LA EXPANSIÓN NEURONAL CORTICAL	89
X. CONCLUSIONES	90
XI. PERSPECTIVAS	92
XII LIMITACIONES	93
XIII. REFERENCIAS	94
XIV. ANEXOS	101
14.1 TABLAS SUPLEMENTARIAS	101
14.2 FIGURAS SUPLEMENTARIAS	104
14.3 CÓDIGO	111
Script 1. PGLS doble de predictor único	111
Script 2. Valores de Significantes	114
Script 3. Expansión y contracción de familia de genes	115
Script 4. Diagramas de Venn	115
Script 5. Enriquecimiento funcional en términos <i>GO</i>	116
Script 6. <i>HeatMap</i>	117
Script 7. Match Familia de genes - Genes	119
Script 8. Expresión de genes (tejidos o región cerebral)	119
Script 9. Expresión de genes espacio-temporal	122
Script 10. Número de copias por familia	124
Script 11. Dominios proteicos por familia de genes	125
Script 12. Comparación de dominios proteicos por familia de genes	125
Script 13. Dominios diferencialmente abundantes	126
14.4 CONSTANCIAS	128
.....	128

ÍNDICE DE TABLAS

TABLA 1. DATOS DE COMPOSICIÓN CEREBRAL EN ESPECIES DISTINTAS DE MAMÍFEROS CON GENOMA DE REFERENCIA DISPONIBLE	37
TABLA 2. DISTRIBUCIÓN DE GENES EN REGIONES ENRIQUECIDAS EN TODO EL GENOMA HUMANO	66
TABLA 3. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0007166 - CELL SURFACE RECEPTOR SIGNALING PATHWAY” IDENTIFICADAS MEDIANTE EL ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA.....	72

TABLA 4. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0002376 - <i>IMMUNE SYSTEM PROCESS</i> ” IDENTIFICADAS MEDIANTE ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA	75
TABLA 5. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0002250 - <i>ADAPTATIVE IMMUNE RESPONSE</i> ” IDENTIFICADAS MEDIANTE ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA	76
TABLA 6. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0006954 – <i>INFLAMMATORY RESPONSE</i> ” IDENTIFICADAS MEDIANTE ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA	78
TABLA 7. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0009605 <i>RESPONSE TO EXTERNAL STIMULUS</i> ” IDENTIFICADAS MEDIANTE ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA	80
TABLA 8. SUBCONJUNTO DE PROTEÍNAS RELEVANTES EN FUNCIONES CEREBRALES, PERTENECIENTES AL PROCESO BIOLÓGICO: “GO:0006950 <i>RESPONSE TO STRESS</i> ” IDENTIFICADAS MEDIANTE ANÁLISIS DE INTERACCIÓN PROTEÍNA-PROTEÍNA	81
TABLA S1. TABLA GENERAL DE FAMILIAS ASOCIADAS SIGNIFICATIVAMENTE CON PARÁMETROS DE COMPOSICIÓN CELULAR DEL CEREBRO, RESULTADO DEL ANÁLISIS DE PGLS SIMPLE DE ÚNICO PREDICTOR	101
TABLA S2. ESTRUCTURAS CEREBRALES CLASIFICADAS PARA EL PRESENTE ESTUDIO	102

ÍNDICE DE FIGURAS

FIGURA 1. VARIABILIDAD EN EL CEREBRO DE LOS VERTEBRADOS.	15
FIGURA 2. ESTRUCTURA DE LA CORTEZA CEREBRAL.	16
FIGURA 3. ESQUEMA MEDIAL EN HEMISFERIO DERECHO CON DELINEACIÓN DE LAS PRINCIPALES REGIONES EN LA CORTEZA CEREBRAL HUMANA.	17
FIGURA 4. ESQUEMA DEL NEOCÓRTEX.	18
FIGURA 5. ESTRUCTURA CEREBRAL.	19
FIGURA 6. VARIABILIDAD DEL TAMAÑO DEL CEREBRO EN MAMÍFEROS.	20
FIGURA 7. EXPANSIÓN DEL CEREBRO DURANTE LA EVOLUCIÓN HUMANA.	22
FIGURA 8. EXPANSIÓN EVOLUTIVA EN CORTEZA CEREBRAL DE MAMÍFEROS.	23
FIGURA 9. MASA ABSOLUTA, NÚMERO DE NEURONAS Y NÚMERO DE CÉLULAS NO NEURONALES EN EL CEREBRO HUMANO ADULTO.	24
FIGURA 10. COMPARATIVA DE TAMAÑO DE CEREBRO Y CONTEO NEURONAL EN CORTEZA CEREBRAL, EN CUATRO ESPECIES DE MAMÍFEROS.	25
FIGURA 11. DUPLICACIÓN DE GENES DURANTE LA EVOLUCIÓN DE LAS FAMILIAS DE GENES DE GLOBINA HUMANA.	27
FIGURA 12. VARIACIONES DEL TAMAÑO DE LAS FAMILIAS DE GENES EN LÍNEA CON EL AUMENTO DE LA ENCEFALIZACIÓN EN MAMÍFEROS.	29
FIGURA 13. RELACIÓN ENTRE VARIACIONES DEL TAMAÑO DE FAMILIA DE GENES ASOCIADAS A ÍNDICE DE ENCEFALIZACIÓN Y EXPRESIÓN GÉNICA EN CEREBRO.	29

FIGURA 14. ANÁLISIS DE ENRIQUECIMIENTO DE ONTOLOGÍA GÉNICA DE FAMILIAS CON VARIACIONES DEL TAMAÑO DE LA FAMILIA GÉNICA (GFS) EN LÍNEA CON EL ÍNDICE DE ENCEFALIZACIÓN (EI) Y LA RELACIÓN DE TAMAÑO DE NEOCORTEZA EN CEREBRO (NR).	30
FIGURA 15. DIFERENCIAS ENTRE ESPECIES E INTRA-CORTEZAS PREDICHAS POR EL MODELO EN EL TIEMPO, LA EXTENSIÓN Y LA ASIGNACIÓN DE CAPAS DE LA PRODUCCIÓN DE NEURONAS CORTICALES.	32
FIGURA 16. EL ESQUEMA MUESTRA LAS MÚLTIPLES CAPAS DE REGULACIÓN QUE SUBYACEN A LA EVOLUCIÓN DEL CEREBRO HUMANO.	32
FIGURA 17. ÁRBOL FILOGENÉTICO (POLITOMÍA NO RESUELTA) DE LAS 18 ESPECIES DE MAMÍFEROS ANALIZADAS, CON GENOMA DE REFERENCIA Y CON DATOS DE COMPOSICIÓN CELULAR CEREBRAL.	39
FIGURA 18. SOBRERREPRESENTACIÓN DE TAMAÑO DE FAMILIAS DE GENES (GFS), EN EXPANSIÓN Y CONTRACCIÓN ASOCIADOS A FENOTIPOS DE COMPOSICIÓN CELULAR, A PARTIR DE RESULTADOS EN ANÁLISIS DE PGLS DOBLE DE PREDICTOR ÚNICO.	51
FIGURA 19. ESPECIFICIDAD DE LAS ASOCIACIONES DEL TAMAÑO DE FAMILIA DE GENES (GFS) CON FENOTIPOS DE COMPOSICIÓN CELULAR, A PARTIR DE RESULTADOS EN ANÁLISIS DE PGLS DOBLE DE PREDICTOR ÚNICO.	52
FIGURA 20. PERFIL RELATIVO DE DOMINIOS PRESENTES POR ESPECIE, EN LAS 385 FAMILIAS DE GENES ASOCIADAS A NEURONAS EN CORTEZA CEREBRAL.	54
FIGURA 21. DISTRIBUCIÓN COMPARATIVA DE DOMINIOS PROTEICOS PRESENTES POR ESPECIE, EN LAS 385 FAMILIAS DE GENES ASOCIADAS A NEURONAS EN CORTEZA CEREBRAL.	55
FIGURA 22. SOBRERREPRESENTACIÓN DE ANOTACIONES FUNCIONALES ENTRE EL GFS ASOCIADAS AL NÚMERO DE NEURONAS EN CORTEZA CEREBRAL Y GLÍA EN CORTEZA CEREBRAL.	56
FIGURA 23. SOBRERREPRESENTACIÓN DE ANOTACIONES FUNCIONALES ENTRE EL GFS ASOCIADAS AL NÚMERO DE NEURONAS EN LA CORTEZA CEREBRAL Y LA MASA CORTICAL RELATIVA.	58
FIGURA 24. SOBRERREPRESENTACIÓN DE ANOTACIONES FUNCIONALES ENTRE EL GFS ASOCIADAS AL NÚMERO DE NEURONAS EN CORTEZA CEREBRAL Y EL NÚMERO DE NEURONAS EN CEREBRO COMPLETO.	59
FIGURA 25. EXPRESIÓN GÉNICA DE GENES ASOCIADOS A NEURONAS EN CORTEZA CEREBRAL, EN COMPARACIÓN CON OTROS TEJIDOS EN EL CUERPO HUMANO EN ETAPAS FETAL Y ADULTA.	61
FIGURA 26. EXPRESIÓN GÉNICA DE GENES ASOCIADOS A NEURONAS EN CORTEZA CEREBRAL, EN COMPARACIÓN CON OTRAS REGIONES DEL CEREBRO HUMANO EN ETAPAS: FETAL, INFANTIL Y ADULTA.	62
FIGURA 27. PATRONES DE TEMPORALIDAD DE EXPRESIÓN GÉNICA DE LAS FAMILIAS DE GENES SIGNIFICATIVOS ASOCIADOS A NEURONAS DE CORTEZA CEREBRAL CON RESPECTO A ETAPAS DEL DESARROLLO CEREBRAL HUMANO (EDAD).	64
FIGURA 28. DISTRIBUCIÓN DE GENES PERTENECIENTES A FAMILIAS DE GENES SIGNIFICATIVOS PARA NEURONAS EN CORTEZA CEREBRAL.	65
FIGURA 29. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0007166 - <i>CELL SURFACE RECEPTOR SIGNALING PATHWAY</i> .	71

FIGURA 30. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0002376 - <i>IMMUNE SYSTEM PROCESS</i> .	74
FIGURA 31. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0002250 - <i>ADAPTATIVE IMMUNE RESPONSE</i> .	76
FIGURA 32. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0006954 <i>INFLAMMATORY RESPONSE</i> .	77
FIGURA 33. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0009605 <i>RESPONSE TO EXTERNAL STIMULUS</i> .	79
FIGURA 34. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA, GO:0006950 <i>RESPONSE TO STRESS</i> .	82
FIGURA 36. MODELO INTEGRADOR DE LA EXPANSIÓN NEURONAL CORTICAL EN MAMÍFEROS.	90
FIGURA 1S. ESPECIFICIDAD DEL TAMAÑO DE FAMILIA DE GENES (GFS), ASOCIADOS CON FENOTIPOS DE COMPOSICIÓN CELULAR, A PARTIR DE RESULTADOS EN ANÁLISIS DE PGLS SIMPLE DE PREDICTOR ÚNICO.	104
FIGURA 2S. DISTRIBUCIÓN DEL NÚMERO DE COPIAS EN FAMILIA DE GENES (N=2777) EN CADA ESPECIE.	105
FIGURA 3S. DISTRIBUCIÓN DEL NÚMERO DE COPIAS EN FAMILIA DE GENES (N= 385) ASOCIADA AL NÚMERO DE NEURONAS EN CORTEZA CEREBRAL, EN CADA ESPECIE.	105
FIGURA 4S. COMPARACIÓN DEL NÚMERO DE COPIAS POR FAMILIA DE GENES ENTRE EL CONJUNTO TOTAL DE FAMILIAS DE GENES (N= 27777) Y LAS FAMILIAS ASOCIADAS AL NÚMERO DE NEURONAS EN CORTEZA CEREBRAL (N= 385).	106
FIGURA 5S. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA SEÑALANDO LAS PROTEÍNAS DE INTERÉS.	107
FIGURA 6S. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA MODIFICADO A ESCALA (1/4) PARA UNA MEJOR OBSERVACIÓN DE LA INTERACCIÓN DE PROTEÍNAS DE INTERÉS Y SELECCIONADA POR REGIÓN.	108
FIGURA 7S. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA MODIFICADO A ESCALA (1/4) PARA UNA MEJOR OBSERVACIÓN DE LA INTERACCIÓN DE PROTEÍNAS DE INTERÉS Y SELECCIONADA POR REGIÓN.	109
FIGURA 8S. REDES DE INTERACCIÓN PROTEÍNA-PROTEÍNA MODIFICADO A ESCALA (1/4) PARA UNA MEJOR OBSERVACIÓN DE LA INTERACCIÓN DE PROTEÍNAS DE INTERÉS Y SELECCIONADA POR REGIÓN.	110

Abreviaturas

ADN:	Ácido Desoxirribonucleico
ADNm:	ADN mitocondrial
ARN:	Ácido ribonucleico
ARNt:	ARN de transferencia
bIP:	Basal Intermediate Progenitor
BP:	Basal Progenitors
CAGE:	Cap Analysis of Gene Expression
CP:	Cortical Plate
EQ:	Encephalization Quotient
FA:	Fibras de Asociación
FANTOM5:	Functional Annotation of the Mammalian Genome 5
FDR:	False Discovery Rate
FP:	Fibras de proyección a tronco y médula
GFS:	Gene Family Size
GO:	Gene Ontology
HLA:	Human Leukocyte Antigen
HSPA:	Heat Shock Protein
IFNR:	Interferon Production Regulator
IZ:	Intermediate Zone
Mpb:	Mega base pairs
MCM:	Minichromosome Maintenance
MHC:	Major Histocompatibility Complex
MTOR:	Mammalian Target of Rapamycin
NK:	Natural Killer
NTPD:	Núcleos Talámicos de Proyección Difusa
pcw:	Post-Conception Weeks
PGLS:	Phylogenetic Generalized Least Squares
PPI:	Protein-Protein Interaction
PIC:	Phylogenetic Independent Contrast
SNC:	Sistema Nervioso Central
STRING:	Search Tool for the Retrieval of Interacting Genes/Proteins
SVZ:	Subventricular Zone

Resumen

El cerebro es el órgano más complejo producido por la evolución biológica y en mamíferos, ha mostrado un notable incremento en la encefalización (tamaño cerebral relativo al cuerpo) en comparación con otros vertebrados. Dentro de este grupo, primates y cetáceos presentan los mayores niveles de encefalización, siendo la corteza cerebral el principal componente de esta expansión, ya que explica cerca del 90% de la variación entre especies. Más allá del tamaño absoluto, la composición celular y el número de conexiones neuronales se consideran indicadores más precisos de complejidad cerebral y la evolución de capacidades cognitivas. El desarrollo de metodologías recientes ha permitido cuantificar de manera confiable el número de neuronas y células gliales, proporcionando estimaciones comparables entre regiones cerebrales y entre distintas especies. El presente trabajo explora las bases genómicas de la evolución de la composición celular en la corteza cerebral de mamíferos mediante un enfoque de genómica comparativa. Se analizó la relación entre el número de neuronas corticales y la variación en el tamaño de familias de genes (GFS, *gene family size*) utilizando datos de 18 especies con genomas de referencia disponibles. Para ello, se implementaron modelos de regresión con corrección filogenética, lo que permitió evaluar asociaciones entre características genómicas y fenotípicas controlando la dependencia evolutiva. En los resultados se identificó 385 familias de genes asociadas significativamente con el número de neuronas en la corteza cerebral. Los estudios funcionales de estas familias revelaron un enriquecimiento en funciones vinculadas a la señalización celular y la respuesta inmune. Adicionalmente, se exploró su expresión espacio-temporal con datos transcriptómicos humanos, localización cromosómica e interacciones proteína-proteína. Los resultados mostraron baja expresión durante la etapa fetal, esto sugiere que estas familias génicas actúan principalmente en etapas posteriores del desarrollo cortical, relacionadas con la maduración y organización neuronal más que con la proliferación celular; una sobreexpresión durante la infancia, en paralelo con procesos de sinaptogénesis y plasticidad neuronal; y niveles altos en la adultez, con un ligero descenso respecto a la infancia, reflejando un equilibrio entre funcionalidad y mantenimiento de redes neuronales. Asimismo, se detectó acumulación de genes en cromosomas específicos, asociados a funciones inmunológicas y de plasticidad sináptica. El análisis de redes de co-expresión y de interacciones proteína-proteína confirmó enriquecimientos funcionales congruentes. Cabe resaltar que la firma genómica identificada no corresponde a un efecto sistémico de todo el cerebro, sino que se restringe específicamente a la corteza cerebral. En conjunto, este

estudio constituye el primer reporte de una firma genómica asociada a la expansión del número de neuronas corticales en mamíferos. Los resultados sugieren una coevolución entre el sistema nervioso y el sistema inmune, ambos desempeñando un papel esencial en la plasticidad, neuroprotección y mantenimiento de la función cerebral. Estos hallazgos ofrecen nuevas perspectivas para comprender los mecanismos genómicos que sustentan la evolución de la corteza cerebral y, en particular, el origen de las capacidades cognitivas avanzadas en el linaje humano.

I. Introducción

El cerebro, es reconocido como una de las estructuras biológicas más fascinantes, sofisticadas y complejas surgidas a lo largo de la evolución del linaje de metazoarios. En mamíferos este órgano se caracteriza por un notable incremento en su tamaño y complejidad en comparación con otras especies, (De Felipe, 2011). Este notable incremento se debe a una región la cual es considerada clave en su evolución, esta estructura es conocida como “corteza cerebral”, la cual facilita funciones cognitivas complejas y contribuye a la diversidad estructural y funcional del cerebro (Rakic, 2000). Las variaciones en la corteza cerebral explican el 90% de las variaciones del tamaño total del cerebro entre mamíferos. Nuestra especie, *Homo sapiens* destaca como la especie de mamífero con el cerebro y corteza cerebral más grande en proporción al tamaño corporal. Sin embargo, diversas especies de primates y cetáceos, también han experimentado un notable incremento del tamaño cerebral, al igual que en humanos, compartiendo características fundamentales en su estructura y función las cuales se encuentran conservadas de igual manera en otras especies, particularmente en mamíferos en general (Marino, 1998). Esto refuerza la hipótesis de un ancestro común.

Más allá del tamaño de la corteza cerebral, algo que distingue al cerebro humano de otros mamíferos es el número excepcionalmente alto de neuronas corticales, que le confiere una mayor capacidad para procesar información, adaptarse a entornos cambiantes y desarrollar habilidades complejas como el lenguaje, la creatividad y la planificación estratégica a largo plazo. Como señala la Dra. Herculano-Houzel (2009). *"Lo que coloca a los humanos en la cima cognitiva no es un cerebro especial, sino simplemente un cerebro de primate ampliado con el mayor número de neuronas en la corteza cerebral, lo que permite una mayor capacidad de procesamiento y flexibilidad"*. Esta afirmación subraya que la ventaja cognitiva humana radica en la escala y la densidad neuronal de la corteza cerebral, más que en diferencias estructurales únicas.

1.1. La corteza cerebral y su papel en la evolución del cerebro

La historia evolutiva del cerebro está estrechamente relacionada con el aumento de la capacidad cognitiva y a la necesidad de adaptarse a ambientes cada vez más complejos durante el paso de los años. Por ejemplo, a medida que los entornos se volvieron más exigentes, esto se tradujo en un aumento progresivo del tamaño cerebral y conectividad sináptica (De Felipe, 2011; Jerison, 1974).

En el linaje de mamíferos se desarrolló y expandió sus estructuras corticales, en particular la neocorteza, que recubre el resto del cerebro. En algunos linajes se observa una expansión más pronunciada de esta estructura resultando en, la formación de pliegues permitiendo una mayor superficie neocortical. Esta expansión y plegamiento ocurre de manera independiente en varios linajes de mamíferos tales como los primates y los cetáceos. La expansión y el plegamiento de la corteza cerebral son características evolutivas clave que han potenciado las capacidades cognitivas superiores en diversas especies, permitiendo una mayor eficiencia en la integración y el procesamiento de información (Rakic, 2000). Estas diferencias morfológicas reflejan diferencias en la composición celular impulsado por la proliferación de células madre y progenitoras neuronales, junto a una compleja regulación genética y ambiental (Namba & Huttner, 2017). Estos procesos, no solo han contribuido significativamente al incremento en el volumen cortical, sino también a una especialización funcional entre los mamíferos.

1.2. Una mirada evolutiva: del telencéfalo a la corteza cerebral

El origen del cerebro o sistema nervioso central (SNC), entendido como una concentración de neuronas y células de apoyo en el organismo anatómicamente localizado cerca de los ojos y la boca, se remonta a la aparición del ancestro común de los bilaterales (incluidos moluscos, artrópodos y vertebrados), hace aproximadamente unos 700 millones de años. Donde un cúmulo de neuronas recibe y procesa estímulos sensoriales, señales internas del organismo, al igual que controla respuestas fisiológicas (Hartenstein & Stollewerk, 2015).

El origen de las neuronas puede trazarse a los primeros metazoarios multicelulares en los que es posible distinguir células especializadas con capacidad de censar el ambiente, procesar señales de potencial eléctrico y emitir señales químicas en el organismo (Watanabe, 2017). En linajes basales como cnidarios ya se encuentran células especializadas con funciones sensoriales conectadas entre sí, pero sin encontrarse concentradas en un órgano especializado (Watanabe, 2017). La organización del sistema nervioso central formado por

el cerebro y la espina dorsal de los vertebrados se puede ya observar en linajes basales de cordados. En la etapa juvenil, por ejemplo, *Ciona intestinalis* presenta una estructura cerebral y una espina dorsal. La actividad de genes cruciales en el desarrollo del cerebro en vertebrados puede ya observarse en este urocordado (Ryan et al., 2016). Es de notar que, en su etapa adulta, *C. intestinalis* pierde las estructuras del cerebro y espina dorsal para vivir como filtrador de partículas en el océano adherido a una roca. Reflejando el gran costo energético que el cerebro representa (Ryan et al., 2016).

Al surgimiento del subfilo *Vertebrata* hace aproximadamente 500 millones de años, es posible reconocer estructuras comunes en todos los vertebrados. Con el arribo de los vertebrados a hábitats terrestres podemos observar un incremento en el tamaño del cerebro, en particular en el “telencéfalo” en reptiles y mamíferos en comparación con anfibios y peces. Particularmente en mamíferos, el telencéfalo desempeña un papel crucial en las funciones cerebrales superiores, como el aprendizaje, la memoria, la percepción sensorial y el control motor. Esta estructura cerebral es altamente compleja y está compuesta por diversas regiones funcionalmente especializadas, entre las cuales destacan:

- La corteza cerebral (palio dorsal), encargada del procesamiento de información sensorial y funciones cognitivas avanzadas.
- El bulbo olfatorio (palio ventral), involucrado en la percepción del olfato.
- El hipocampo (palio medial), esencial para la consolidación de la memoria y la navegación espacial.
- Los ganglios basales y la amígdala (subpalio), participan en la regulación del movimiento y el procesamiento emocional.

De todas estas regiones mencionadas, la corteza cerebral adquiere especial relevancia, ya que constituye más de la mitad del volumen total del cerebro en mamíferos. Su expansión y plegamiento son características clave en la evolución cerebral (Florio & Huttner, 2014). La neocorteza juega un papel clave aquí, ya que está formada por capas de neuronas que recubren al telencéfalo, dicha estructura está presente desde la aparición de los primeros mamíferos y su presencia es evidente a partir de escaneos de fósiles de mamíferos tempranos.

Por ejemplo, la Figura 1 ilustra de manera clara las diferencias notables en la evolución del cerebro en vertebrados. Al observar en peces y anfibios, el telencéfalo es relativamente pequeño, mientras que en reptiles se observa un incremento considerable en

su tamaño. Pero de manera más notable hay un mayor incremento en aves y mamíferos, lo que indica que el telencéfalo adquirió un papel protagónico en la evolución cerebral de los vertebrados (Kverková et al., 2022).

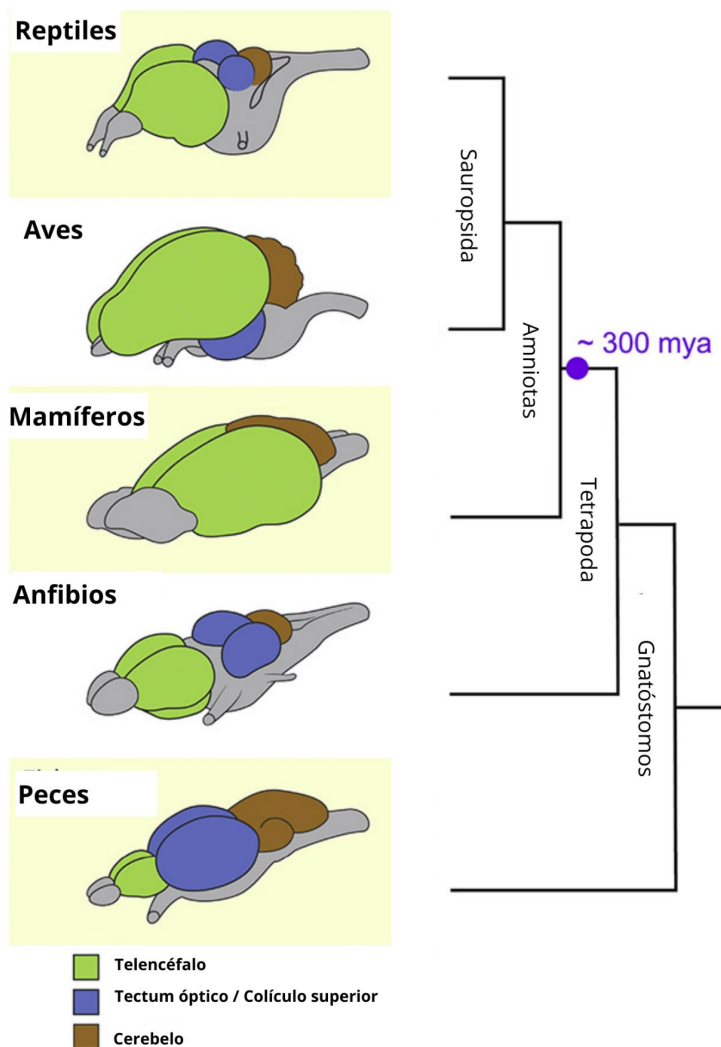


Figura 1. Variabilidad en el cerebro de los vertebrados. Componentes principales del cerebro de los vertebrados. De izquierda a derecha, cerebros esquemáticos y secciones transversales a nivel del telencéfalo. Color verde: telencéfalo; azul: tectum óptico/colículo superior; café: cerebelo. Imagen de: (Pessoa et al., 2019) y modificada para el presente trabajo.

En el caso de las aves, un linaje diverso y paralelo a la evolución cerebral en mamíferos, también experimentó una expansión del telencéfalo, aunque no a través de la evolución de una estructura especializada como la neocorteza sino a través de la expansión del palio.

1.3. La expansión cortical como eje de variabilidad evolutiva

Los mamíferos tienden a presentar los cerebros de mayor tamaño en el reino animal, y como ya se mencionó, la expansión de la corteza cerebral, constituye un factor determinante en la evolución de los mamíferos, en comparación con otros taxones (Lui et al., 2011).

Para entender cómo surge esta expansión cortical debemos observar la anatomía del cerebro seccionado como se muestra en la Figura 2, aquí puede apreciarse una delgada línea de materia color azul que representa la materia gris, correspondiente al manto cortical altamente expandido dentro de la bóveda craneal en mamíferos (Namba & Huttner, 2017).

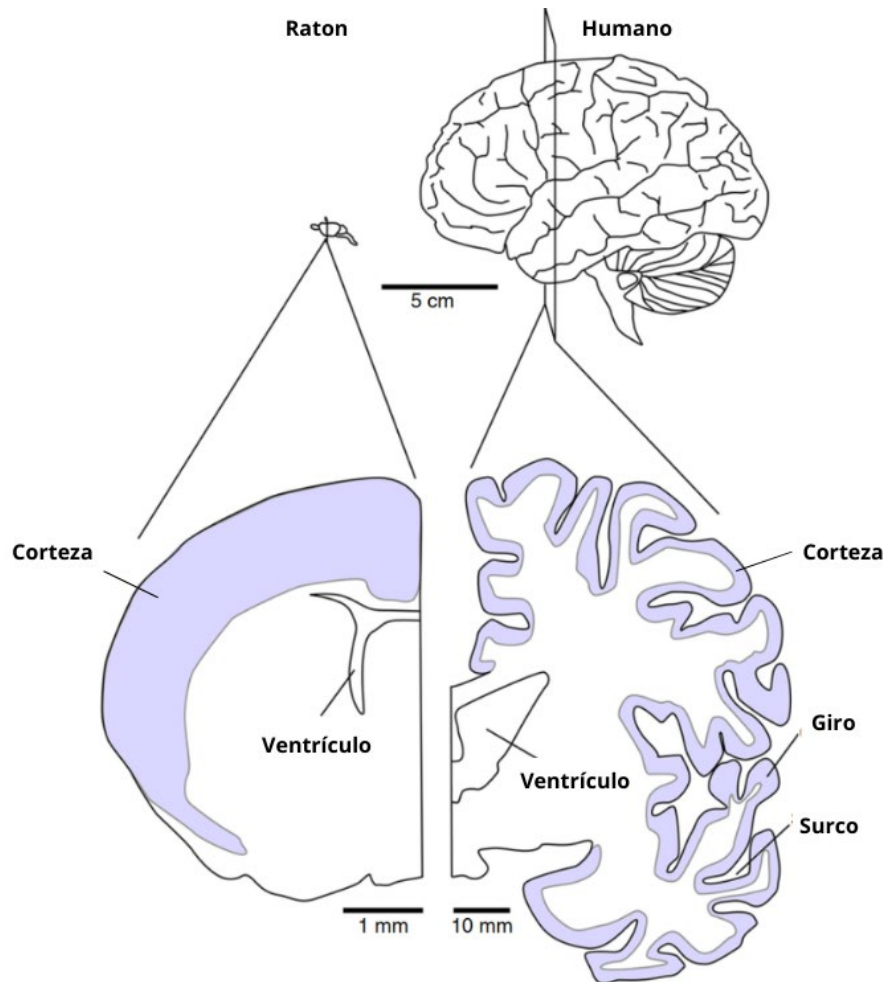


Figura 2. Estructura de la corteza cerebral. Dibujos del cerebro de ratón (izquierda) y humano (derecha) (arriba) y de una sección coronal (abajo). El área azul indica la materia gris. Imagen de (Namba & Huttner, 2017) y modificada para el presente trabajo.

Dentro de la anatomía de la corteza cerebral, es importante señalar que se divide en tres regiones principales (Figura 3), diferenciadas por el número de capas corticales que estas poseen.

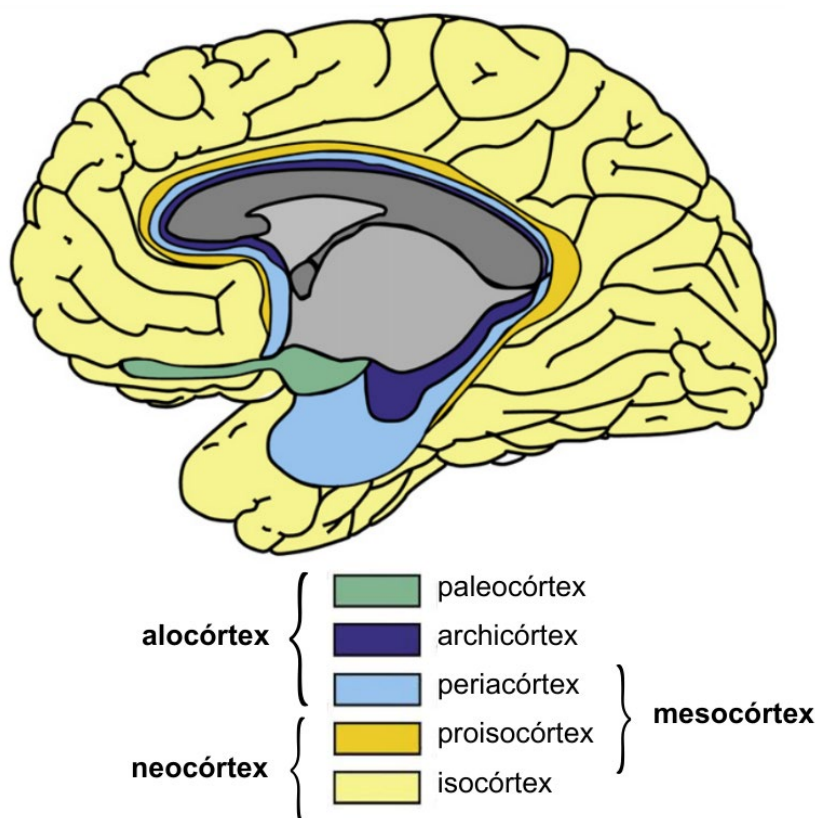


Figura 3. Esquema medial en hemisferio derecho con delineación de las principales regiones en la corteza cerebral humana. Tres regiones corticales principales; alocórtex, mesocórtex y neocórtex, y subregiones. Imagen de: (Zilles & Amunts, 2012) y modificada para el presente trabajo.

1. **Alocórtex.** La primera región y se caracteriza por tener tres capas corticales: molecular, piramidal y polimorfa. A su vez, se subdivide en el archicórtex, que incluye el hipocampo, y el paleocórtex, dónde se encuentra la corteza o bulbo olfatorio.
2. **Mesocórtex.** Representa la segunda región y actúa como una zona de transición entre el alocórtex y el neocórtex. Se divide en dos subregiones:
 - Periarchicórtex, que conecta el alocórtex con el mesocórtex.
 - Proisocórtex, que vincula el mesocórtex con el neocórtex.
 - Dependiendo de la subregión, el mesocórtex puede presentar entre tres y seis capas corticales.
3. **Neocórtex o isocórtex.** La tercer y más extensa región, representando aproximadamente el 90% de la corteza cerebral total. Se distingue por estar compuesta por seis capas (Figura 4) bien definidas (Solé, 2022; Zilles & Amunts, 2012):
 - I. Capa molecular
 - II. Capa granular externa
 - III. Capa piramidal externa: células piramidales pequeñas
 - IV. Capa granular interna: células estrelladas

- V. Capa piramidal interna: células piramidales grandes
- VI. Capa poliforme

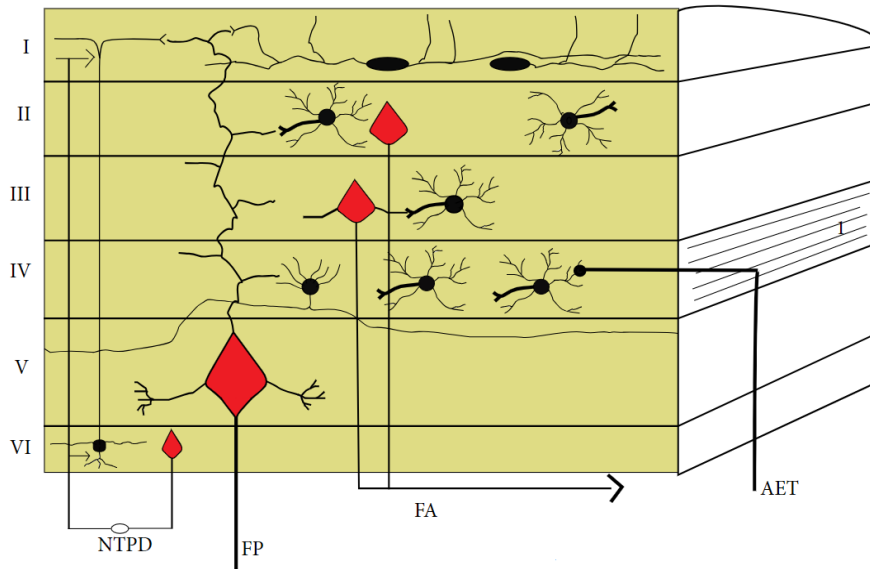


Figura 4. Esquema del neocórtex. Estructura de neocórtex en seis capas que lo componen. AET (aferentes específicas talámicas), FA (fibras de asociación), FP (fibras de proyección a tronco y médula) y NTPD (circuito de ida y vuelta con núcleos talámicos de proyección difusa). Imagen de: (Solé, 2022).

El neocórtex es una estructura de origen evolutivo relativamente reciente en mamíferos (Florio & Huttner, 2014). Su expansión se considera un logro evolutivo, ya que permitió una notable diversificación celular y una mayor sofisticación de los circuitos corticales, favoreciendo el desarrollo de comportamientos complejos (Lodato & Arlotta, 2015). La expansión neocortical es muy evidente en primates antropoides, en especial en humanos, donde el neocórtex representa aproximadamente el 80% de la masa cerebral total (Azevedo et al., 2009). Se estima que está compuesto por alrededor de 100 mil millones de neuronas, acompañadas por una cantidad diez veces mayor de células no neuronales (Herculano-Houzel, 2009).

Cabe recalcar que la variación en corteza cerebral de los mamíferos radica específicamente en la extensión de su superficie, y no en su grosor el cual es casi uniforme entre las diferentes especies (Mota & Herculano-Houzel, 2014). En la Figura 2, se observa claramente, la comparación anatómica de la corteza cerebral entre el ratón y el humano, observándose diferencias notables. Una diferencia muy marcada, es el tamaño de la corteza: ya que, en humano, el volumen es aproximadamente 1,000 veces mayor, que en el ratón (Figura 2). Otra diferencia, es notable en la superficie cortical observada en la Figura 5; aquí el ratón muestra una superficie cortical lisa, es decir, lisencefálica y por el contrario en humano presenta numerosos pliegues en su superficie cortical, ósea girencefálica, formando

un patrón complejo de giros y surcos que permite una mayor expansión de la corteza dentro de la bóveda craneal (Florio & Huttner, 2014).

En estudios anteriores se ha descrito que este plegado cortical, constituye una solución evolutiva al desafío del aumento de la superficie cortical (Kelava et al., 2013). La cual explica que, a medida que las especies de mamíferos aumentaron de tamaño corporal a lo largo de la evolución, sus cerebros también debieron expandirse. Sin embargo, esta expansión no podía ocurrir de manera proporcional, debido a restricciones biomecánicas, por ejemplo, el estrés que sufrirían las vértebras cervicales con el incremento excesivo en el tamaño craneal. Debido a esto, el cráneo no podría escalar isométricamente en relación con el cuerpo, limitando así su crecimiento (Kelava et al., 2013).

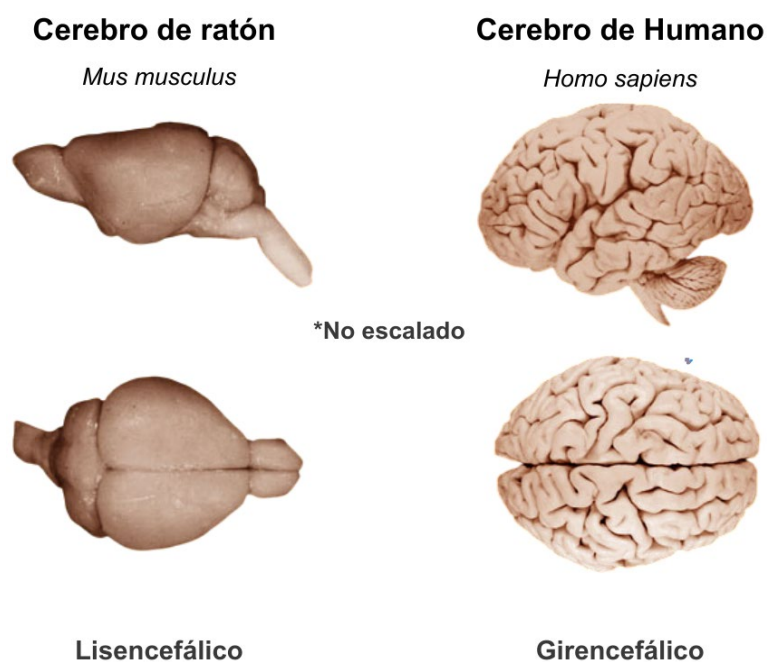


Figura 5. Estructura cerebral. Fotografías de sección lateral y superior del cerebro, en ratón adulto (izquierda) y humano adulto (derecha). Tipos de cerebro: lisencefálico (izquierda) y girencefálico (derecha). Fotografías de cerebros pertenecen a <http://brainmuseum.org> y modificadas para el presente trabajo.

Entonces para que la corteza cerebral pueda plegarse, su grosor no puede aumentar significativamente, esto se debe a restricciones impuestas por el patrón de conectividad, es decir, las interconexiones entre las distintas regiones cerebrales. Lo que explica al mismo tiempo la mínima o inexistente variación en el grosor de la corteza cerebral entre las diferentes especies de mamíferos (De Felipe, 2011). En resumen, el aumento gradual del tamaño corporal requiere un mayor número de neuronas, lo que se logra mediante la expansión de la superficie de la corteza cerebral, que a su vez se pliega para ajustarse a un cráneo relativamente más pequeño.

Actualmente, los mamíferos presentan notables variaciones en su estructura y función cerebral (De Felipe, 2011). Por ejemplo, en la Figura 6 se muestran algunas especies de mamíferos, junto con su respectiva masa cerebral, expresada en gramos.

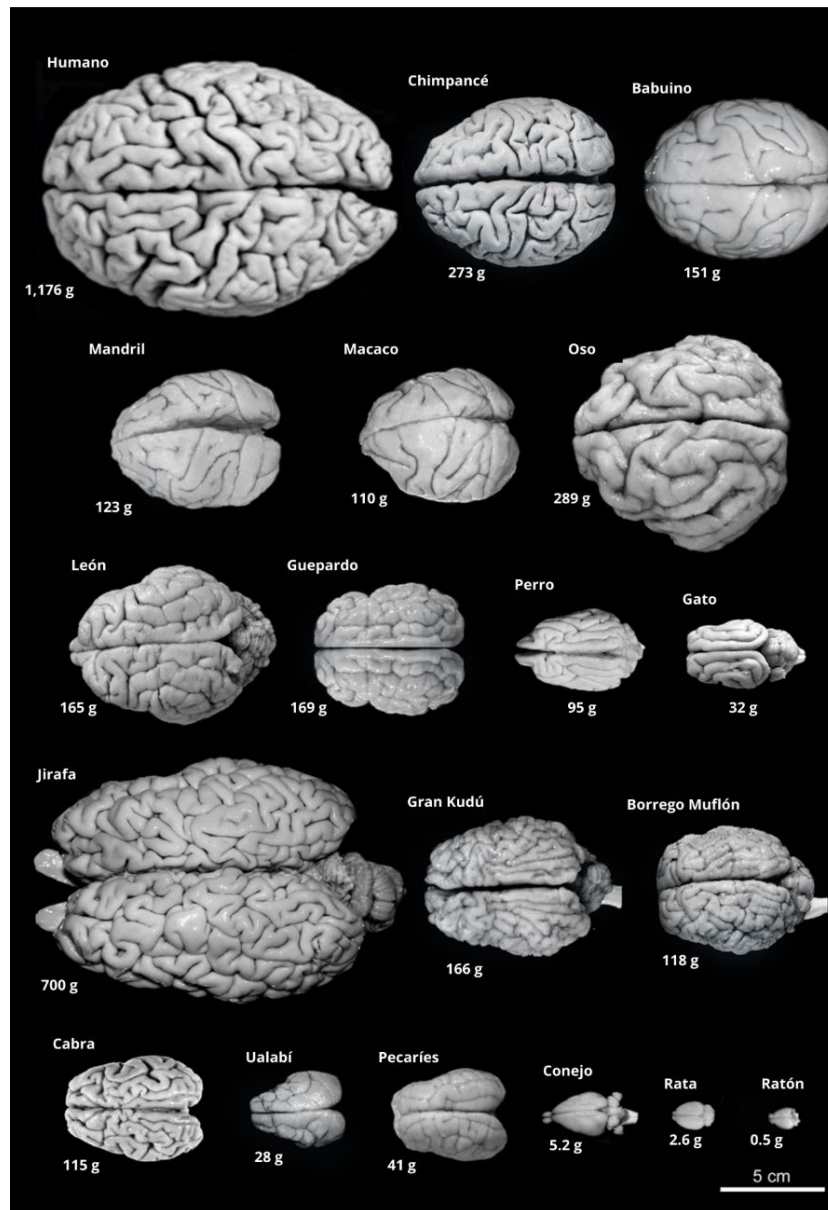


Figura 6. Variabilidad del tamaño del cerebro en mamíferos. Fotografías y pesos de cerebros de diferentes especies. Primates: humano (*Homo sapiens*), chimpancé (*Pan troglodytes*), babuino (*Papio cynocephalus*), mandril (*Mandrillus sphinx*), macaco (*Macaca tonkeana*). Carnívoros: oso (*Ursus arctos*), león (*Panthera leo*), guepardo (*Acynonix jubatus*), perro (*Canis lupus familiaris*), gato (*Felis catus*). Artiodáctilos: jirafa (*Giraffa camelopardalis*), gran kudú (*Tragelaphus strepsiceros*), borrego muflón (*Ovis aries musimon*), Cabra (*Capra aegragus hircus*); pecaríes (*Tayassuidae pecari*). Marsupiales: Ualabí (*Macropus rufogriseus*). Lagomorfos: conejo (*Oryctolagus cuniculus*). Roedores: rata (*Rattus rattus*), ratón (*Mus musculus*). Barra de escala: 5 cm. Imagen de (De Felipe, 2011) y modificada para el presente trabajo.

1.4. Más allá del tamaño: ¿qué determina la evolución cerebral?

Durante décadas, se ha planteado que el aumento del tamaño del cerebro está relacionado con el tamaño corporal, a lo que se le conoce como “encefalización”, un término acuñado por primera vez por Jerison (1974). Dado que las especies con cerebros más grandes suelen tener cuerpos proporcionalmente más grandes, el tamaño cerebral debe ajustarse en función del peso corporal. Para ello, se emplea una fórmula matemática conocida como “cociente o índice de Encefalización” (*EQ*, por sus siglas en inglés: *encephalization quotient*), el cual se define como; el cociente entre la masa del cerebro y el peso corporal, en relación con el peso cerebral esperado (Jerison, 1974).

Si bien esta métrica se ha utilizado ampliamente para cuantificar el tamaño cerebral en diversas especies, e incluso en especímenes fósiles como *Australopithecus afarensis* (3.1–3.6 millones de años) y *Homo neanderthalensis* (30,000 años) por Marino (1998), esta ha sido cuestionada por su incapacidad para explicar variabilidad entre especies, ya que los cambios en tamaño cerebral no siempre siguen relaciones alométricas predecibles con el tamaño corporal. Sugiriendo que EQ no es un indicador fiable para evaluar la evolución de la expansión cerebral (Shultz & Dunbar, 2010).

Sin embargo, estudios anteriores han observado que el número de neuronas corticales muestra una correlación más estrecha con habilidades cognitivas complejas, sugiriendo una posible relación con el aumento del tamaño cerebral (Herculano-Houzel, 2017; Herculano-Houzel et al., 2011). Sin embargo, el estudio de la evolución de la corteza cerebral a nivel celular se ha limitado principalmente a comparaciones entre especies vivas de mamíferos, debido a que el registro fósil de tejidos blandos como el cerebro es muy limitado (Lui et al., 2011).

Por ejemplo, a pesar de que existen análisis de fósiles craneales para estimar el tamaño cerebral en homínidos a lo largo del tiempo, donde incluso se ha observado una expansión en la capacidad volumétrica endocraneal en la evolución humana: desde la aparición del *Homo habilis*, hace aproximadamente 2 millones de años, hasta otros grandes homínidos como *Homo neanderthalensis* y *Homo sapiens* (Florio et al., 2017), como se ilustra en la Figura 7.

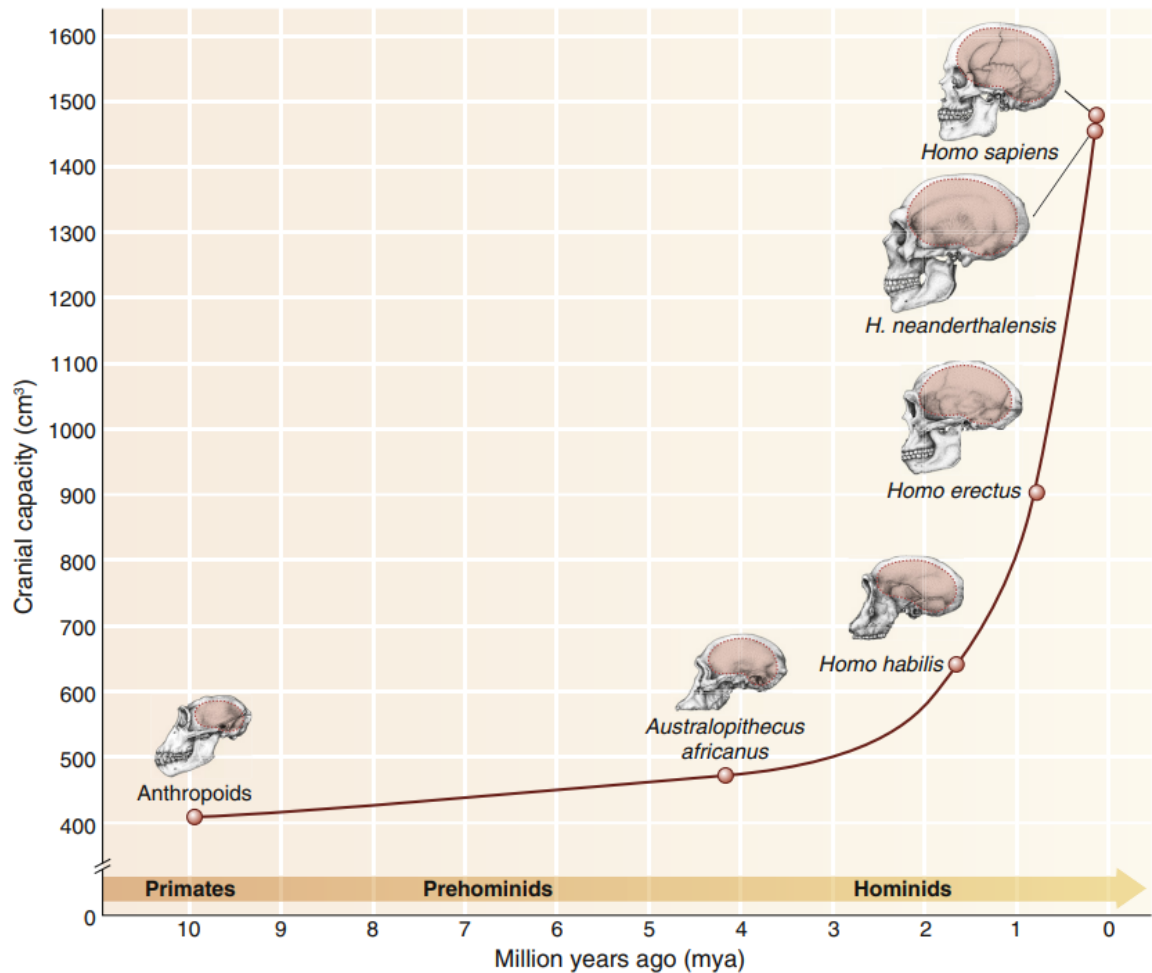


Figura 7. Expansión del cerebro durante la evolución humana. Aumento evolutivo de la capacidad craneal (eje Y, cm^3), a lo largo del tiempo en el linaje de los homínidos (eje X, Millones de años, *mya*). Tendencia hacia un aumento acelerado de la capacidad craneal a partir del *Homo habilis* hasta el *Homo neanderthalensis* y *Homo sapiens*. Imagen de (Florio et al., 2017).

En la última década, se ha propuesto que la expansión cortical está vinculada a la proliferación de progenitores basales (BP), los cuales se localizan en la zona subventricular (SVZ) observado en la figura 8 y presentan una variabilidad significativa en su capacidad proliferativa, entre diferentes especies de mamíferos (Namba & Huttner, 2017).

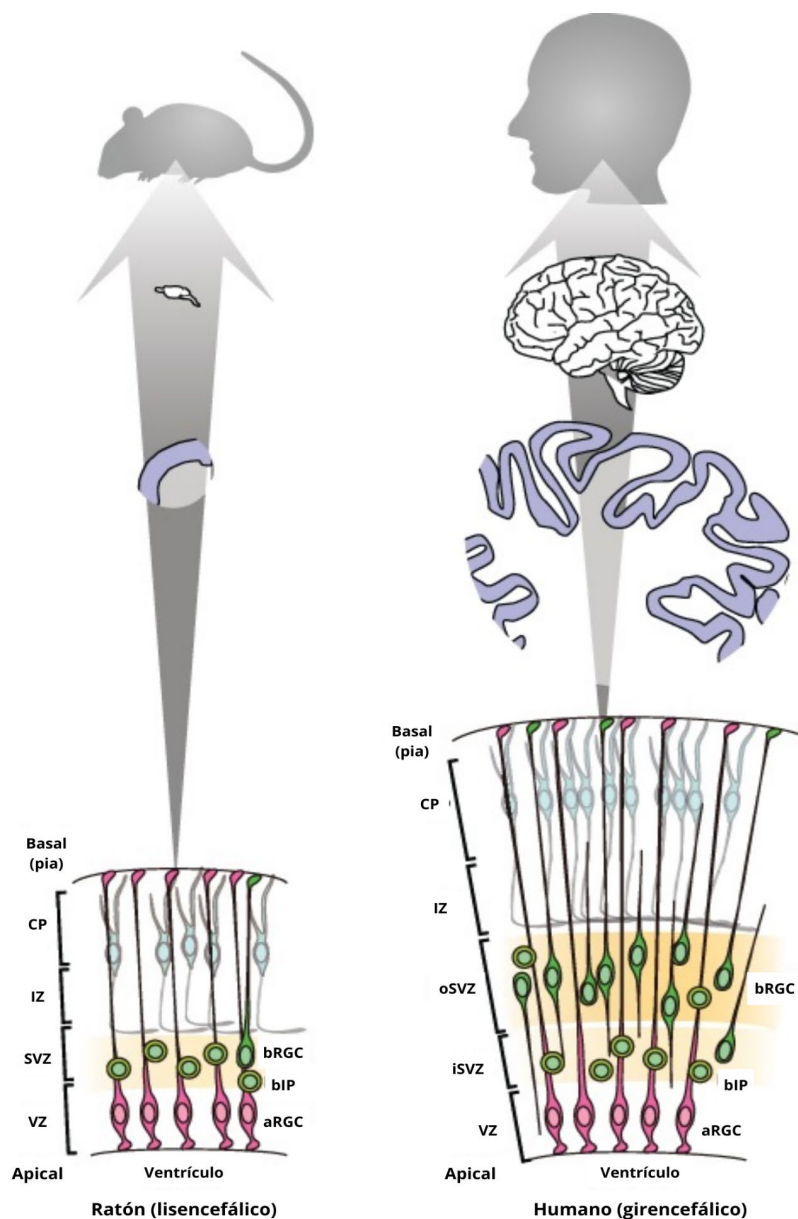


Figura 8. Expansión evolutiva en corteza cerebral de mamíferos. Comparativa entre ratón (izquierda) y humano (derecha). La expansión cortical es por la capacidad proliferativa y neurogénica aumentada y prolongada de células neuronales, acompañada de un mayor plegamiento de la superficie pial, dando lugar a la girusencefalia en humano, en lugar de la lisencefalia en el ratón. Placa cortical (CP), zona intermedia (IZ), zona ventricular (VZ), zona subventricular exterior (oSVZ), zona subventricular interna (iSVZ), células gliales radiales apicales (aRGC), células gliales radiales basales (bRGC) y progenitor intermedio basal (bIP). Imagen de (Namba & Huttner, 2017).

Por ejemplo, en especies con un córtex pequeño y lisencefálico, como el ratón, los BP muestran una baja capacidad de proliferación. Y en contraste, en especies con una corteza cerebral expandida y girusencefálica, como en los humanos, los BP son más prolíficos y experimentan múltiples ciclos de división, lo que genera una mayor cantidad de BP y, en consecuencia, un mayor número de neuronas en corteza cerebral (Florio & Huttner, 2014; Lui et al., 2011).

Así que la capacidad proliferativa diferencial de los progenitores basales en mamíferos, contribuye al aumento del número de neuronas y a la formación de nuevas redes corticales. Sugiriendo que la expansión del cerebro podría estar determinada, en gran medida, por la cuantificación del número de neuronas corticales.

Actualmente, el desarrollo de nuevas metodologías permitió la cuantificación de células en el cerebro de una manera comparable entre especies. En la actualidad se utiliza un método altamente preciso y reproducible para cuantificar neuronas y células no neuronales del cerebro, conocido como “fraccionamiento isotrópico”. Este método altamente estudiado con menos de un 10% de variación entre individuos en sus resultados, lo hace ideal para estudios comparativos ya que proporciona datos fiables y consistentes sobre el número de neuronas y células no neuronales en una amplia variedad de especies (Herculano-Houzel & Lent, 2005).

De hecho, se ha demostrado que los resultados a partir de este método y comparaciones con los valores esperados en humanos, las densidades neuronales y no neuronales, en función de la masa cerebral (Figura 9), se encuentran dentro del rango esperado (Azevedo et al., 2009). Indicando, que utilizar los datos resultantes mediante este método, son fiables y pueden ser utilizados en la investigación.

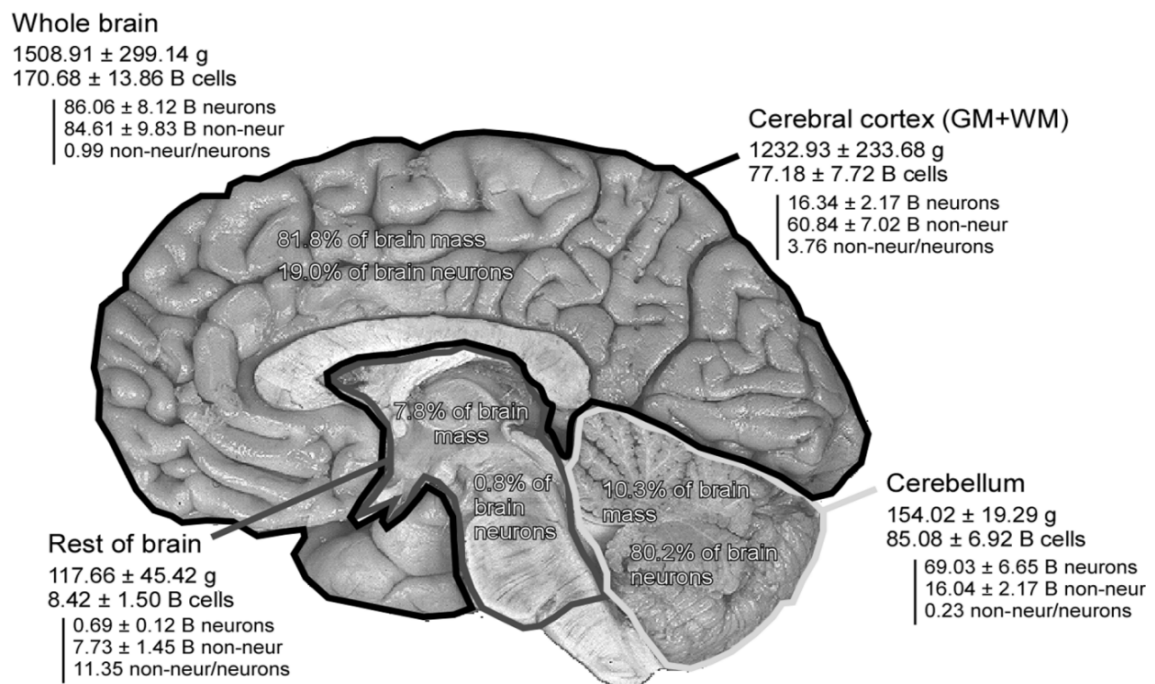


Figura 9. Masa absoluta, número de neuronas y número de células no neuronales en el cerebro humano adulto. Datos de masa absoluta, número de neuronas y número de células no neuronales en todo el cerebro humano. Los valores son la media y se refieren a los dos hemisferios juntos. B = mil millones. Imagen de: (Azevedo et al., 2009).

Investigaciones utilizando los resultados de este método demostraron que, los cerebros más grandes en mamíferos no necesariamente contienen más neuronas y que estas no siempre se distribuyen de manera uniforme. Observando la Figura 10, se comparan cinco cerebros de mamíferos de distintas especies, donde se muestra el peso relativo de la masa cerebral expresada en gramos (g), así como de un conteo aproximado total de neuronas. Por

ejemplo, el cerebro humano es menor en tamaño y peso (~1232 g), con respecto al elefante africano (~2848 g), pero el cerebro humano contiene un mayor número de neuronas (16.3 billones), que en las otras especies de mamíferos de la imagen.

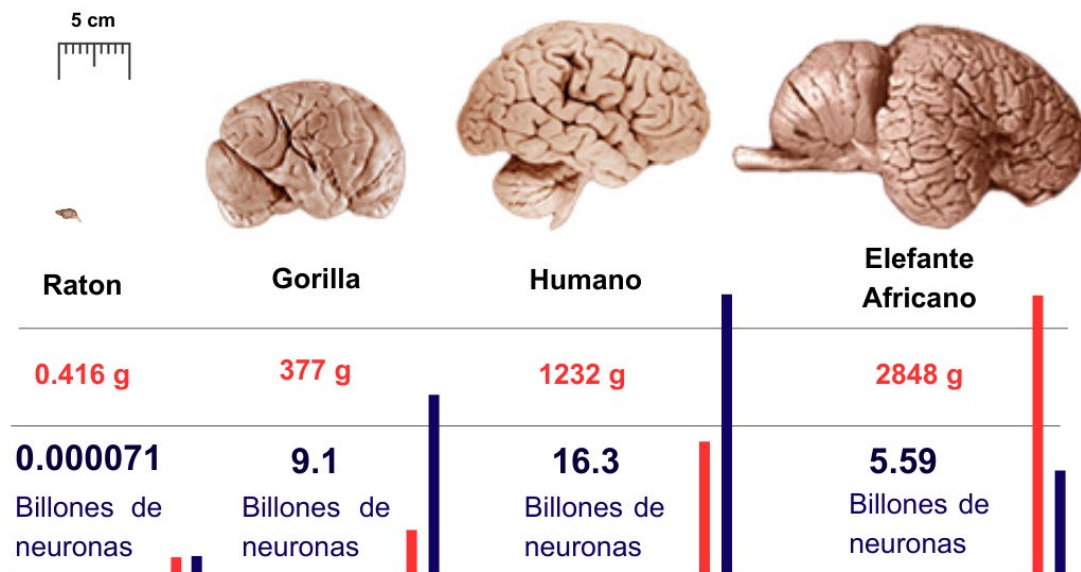


Figura 10. Comparativa de tamaño de cerebro y conteo neuronal en corteza cerebral, en cuatro especies de mamíferos. Se presentan fotografías de cuatro diferentes especies de mamíferos, donde se compara el peso de la masa cerebral total (rojo), expresada en gramos (g), y el conteo celular neuronal (azul marino). Las fotografías se tomaron de <http://brainmuseum.org> y se modificaron para el presente trabajo.

El notable incremento del número de neuronas corticales en humanos se estima que ha estado impulsado por mecanismos evolutivos que aún no se comprenden del todo, pero que probablemente involucraron cambios genómicos clave asociados al desarrollo, la diferenciación neuronal y la regulación de procesos moleculares específicos en la corteza cerebral.

1.5. Genes y evolución del cerebro

Se ha propuesto que la evolución del cerebro ocurrió a través de la expansión de células progenitoras basales, seguida por la formación de circuitos neuronales que resultaron en un mayor número de neuronas. Este proceso sugiere que los genes desempeñan un papel fundamental en la expansión cerebral a lo largo de la evolución, ya que factores génicos influyen directamente en el desarrollo del cerebro (Shaw et al., 2018). Los avances simultáneos en genómica evolutiva, biología del desarrollo y neurociencia han facilitado recientemente la detección de modificaciones genómicas que influyen de manera específica en el desarrollo de la corteza cerebral humana (De Felipe, 2011).

Estudios recientes refuerzan esta idea al demostrar que la regulación génica es crucial en la evolución del cerebro (Florio et al., 2017). Se ha observado que ciertos elementos reguladores específicos presentes en humanos están asociados con genes implicados en la expansión cortical y en la susceptibilidad a enfermedades del neurodesarrollo. Los genes *NOTCH2NL*, que son parálogos específicos del ser humano, promueven el mantenimiento y la expansión de los progenitores corticales, lo que conlleva a un aumento de la producción neuronal y contribuye a la expansión cortical (Suzuki et al., 2018). El gen *ASPM* experimentó una evolución acelerada y está asociado con el tamaño del cerebro, desempeñando un papel importante en la expansión evolutiva del cerebro (Zhang, 2003).

Este conjunto de evidencias refuerza la hipótesis de que una combinación de genes específicos, duplicaciones de genes y elementos reguladores, mejoran la proliferación de progenitores neuronales y el desarrollo cortical. Estos factores génicos no solo contribuyen a la expansión física del cerebro, sino también a la complejidad de las funciones cognitivas, e incluso podrían tener relación con la longevidad.

Estudios comparativos del genoma completo en distintas especies de mamíferos han demostrado asociaciones significativas entre el tamaño de las familias de genes y la encefalización (Castillo-Morales et al., 2014, 2016), respaldando la importancia de la duplicación génica en la evolución del cerebro. Las familias de genes muestran un papel importante, ya que son conjuntos de genes homólogos que comparten un ancestro común y a menudo, desempeñan funciones similares. Las diferencias en el tamaño de estas familias son el resultado de la duplicación de genes específicos y la pérdida de otros, ofreciendo información valiosa sobre las fuerzas evolutivas que han moldeado los genomas de los mamíferos a lo largo del tiempo (Demuth et al., 2006), sugiriendo una vía prometedora para investigar las bases genómicas de la evolución del cerebro. Un ejemplo, de cómo ocurrieron estas duplicaciones génicas, en términos evolutivos de las familias de genes, es el caso de la familia de las globinas humanas (Figura 11):

- La duplicación inicial dio origen a un gen ancestral que, en un linaje, produjo el gen moderno de la mioglobina, la cual es una proteína muscular fijadora de oxígeno, y en otro linaje, los genes que codifican para las globinas. Se estima que esta duplicación ocurrió hace aproximadamente 800 millones de años.
- Posteriormente, los linajes proto- α y proto- β se separaron a través de otra duplicación hace unos 500 millones de años, y las duplicaciones posteriores dentro de las familias α y β ocurrieron durante los últimos 200 millones de años.

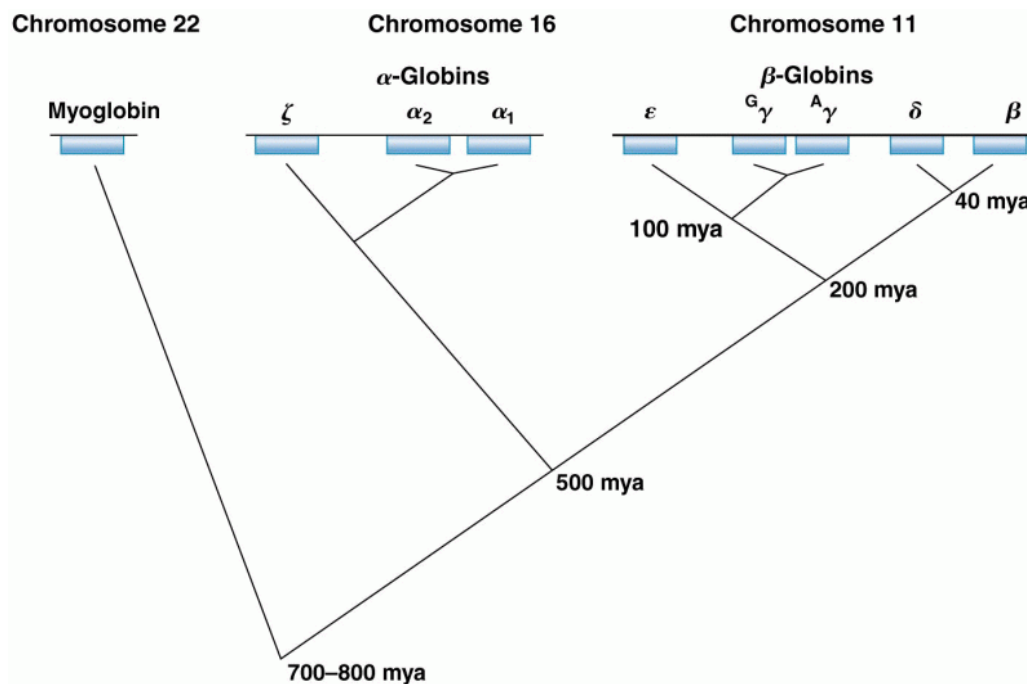


Figura 11. Duplicación de genes durante la evolución de las familias de genes de globina humana. Comparaciones de secuencias de nucleótidos permiten deducir las relaciones evolutivas entre los genes de la globina. Aquí se muestran las fechas de ocurrencia de las duplicaciones clave en millones de años (Myr). Imagen tomada de (Brown, 2002), Chapter 15 “How Genomes Evolve”.

Cabe destacar que estas duplicaciones génicas también estuvieron acompañadas de reubicaciones cromosómicas. Por ejemplo, el gen de la mioglobina se encuentra actualmente en el cromosoma 22, mientras que los genes de la globina α están distribuidos en el cromosoma 16 y los genes de la globina β en el cromosoma 11 (Brown, 2002). Mostrando, cómo la duplicación génica y la reorganización del genoma pueden dar lugar a una mayor diversidad funcional y la especialización de genes. La evolución del cerebro humano, en particular su expansión y complejidad, está impulsada en gran medida por duplicaciones de genes específicos y cambios regulatorios (Suzuki et al., 2018).

1.6 Glía como factor evolutivo del cerebro

Aunque históricamente la evolución del cerebro ha sido interpretada desde una perspectiva centrada en las neuronas, estudios recientes reconoce que las células gliales son arquitectas fundamentales de la complejidad neuronal, lo que reformula fundamentalmente nuestra comprensión de la evolución del cerebro más allá de la visión tradicional centrada en las neuronas (Allen & Lyons, 2018). La microglía participa en procesos de remodelación

sináptica, vigilancia inmunológica y señalización inflamatoria (Sancho et al., 2021), mientras que los oligodendrocitos son esenciales para el metabolismo lipídico y la mielinización axonal, procesos críticos para la eficiencia de la transmisión neuronal (Pękowska et al., 2025). Con estas evidencias la trayectoria evolutiva de la neuroglia muestra una complejidad creciente, con células gliales que emergen temprano en el desarrollo del sistema nervioso y se vuelven progresivamente más especializadas, y propone que la glía constituye un componente clave en la evolución de cerebros complejos en mamíferos (Verkhatsky et al., 2022).

No obstante, dado que el presente trabajo forma parte de un proyecto más amplio en el que la contribución genómica de la glía cortical es analizada de manera independiente, esta tesis se enfoca deliberadamente en las neuronas de la corteza cerebral, con el objetivo de caracterizar los determinantes genómicos asociados a la expansión del número neuronal, sin desconocer la relevancia evolutiva de los linajes gliales.

II. Antecedentes particulares

La evolución del cerebro, y en particular la expansión de la corteza cerebral, continúa siendo un tema complejo e intrigante, influenciado por múltiples factores que determinan su desarrollo y diversidad funcional. Se ha propuesto que el incremento del tamaño cerebral juega un papel clave en la evolución de los mamíferos, diferenciándolos de otros linajes de vertebrados, representando una innovación evolutiva distintiva de este clado (Castillo-Morales et al., 2016).

Estudios previos han asociado dicho incremento del cerebro, con la aparición de comportamientos complejos, y una mayor capacidad adaptativa frente a entornos cambiantes (Lui et al., 2011; Rakic, 2009). En el año 2000, se propuso la hipótesis de la “unidad radial” como un posible mecanismo para explicar este fenómeno, sugiriendo que los cambios en la cinética de proliferación de las células madre en la zona ventricular aumenta el número de unidades columnares radiales. Este modelo anticipa que la identificación de familias de genes y moléculas morforegulatoras serían clave para comprender los procesos que controlan la producción, migración y distribución de neuronas durante el desarrollo cortical (Rakic, 2000).

Más recientemente, estudios de genómica comparativa han mostrado que la encefalización se correlaciona significativamente con variaciones en el tamaño de las familias de genes (GFS, por sus siglas en inglés: *gene family size*), en particular en aquellas

involucradas en funciones biológicas de quimiotaxis, señalización celular y procesos inmunológicos. Estos cambios génicos parecen representar adaptaciones funcionales, ante requerimientos de cerebros más grandes (Figura 12) (Castillo-Morales et al., 2014).

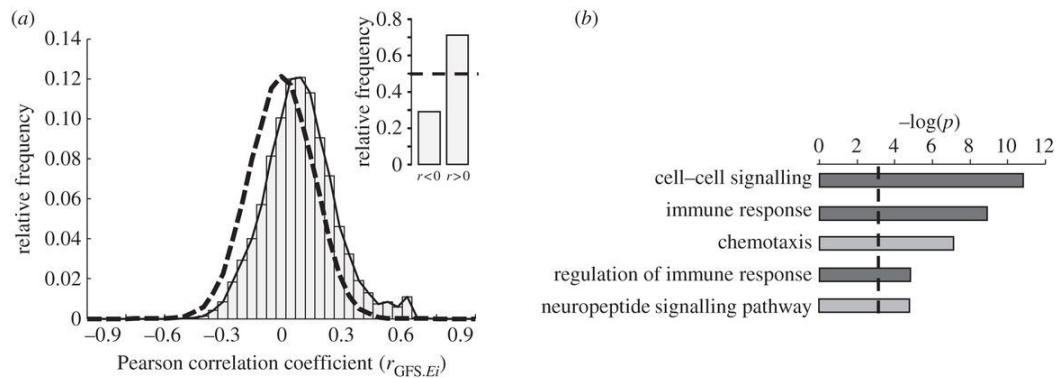


Figura 12. Variaciones del tamaño de las familias de genes en línea con el aumento de la encefalización en mamíferos. (a) Histograma muestra la distribución de coeficientes de correlación para GFS (*gene family size*) y el índice de encefalización (Ei, por sus siglas en inglés: *Encephalization index*) en 12,373 familias de genes que abarcan 39 genomas de mamíferos. La línea discontinua representa una estimación de la distribución esperada basada en la aleatorización. Inserto: distribución de correlaciones positivas y negativas relativas a la distribución esperada (línea discontinua). (b) Desviaciones de las expectativas aleatorias en el coeficiente de correlación media de las familias de genes asociadas con términos GO individuales (expresado como $-\log(p)$ -valor). Sólo se muestran las categorías GO (por sus siglas en inglés: *Gene Ontology*). Imagen de: (Castillo-Morales et al., 2014).

Además, se documentó que los genes pertenecientes a estas familias muestran una expresión particularmente elevada en la corteza cerebral, en comparación con regiones como el cerebelo o áreas subcorticales (Figura 13), reforzando un posible rol en la expansión cortical (Castillo-Morales et al., 2014). En conjunto estos resultados resaltan la posible importancia funcional de las familias génicas asociadas con la encefalización, las cuales representan una respuesta en la evolución del cerebro ante requisitos funcionales que subyacen en el aumento del tamaño cerebral (Castillo-Morales et al., 2014, 2016).

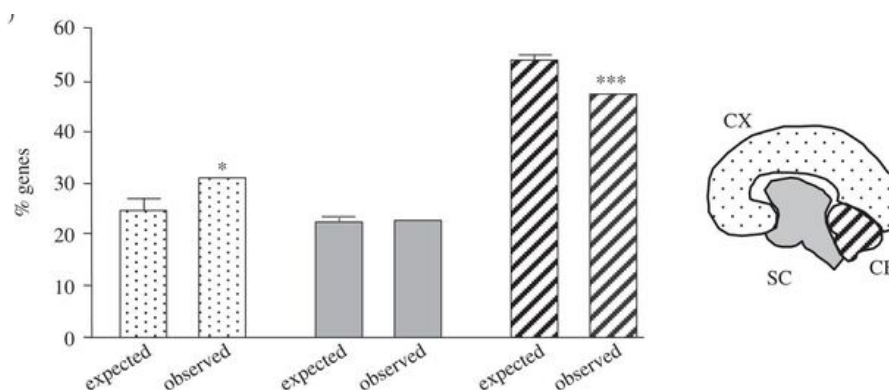


Figura 13. Relación entre variaciones del tamaño de familia de genes asociadas a Índice de encefalización y expresión génica en cerebro. Porcentaje de genes expresados al máximo en las regiones cortical (CX), subcortical (SC) y cerebelosa (CB) respectivamente. Los valores esperados (media $\pm$ s.e.m.) se determinaron numéricamente utilizando muestras aleatorias de tamaño comparable de genes extraídos del grupo más amplio de familias de genes. * $p < 0.01$; ** $p < 0.001$; *** $p < 0.0001$. Imagen tomada de: (Castillo-Morales et al., 2014).

En estudios funcionales posteriores, se demostró que la expansión cortical en relación con el resto del cerebro, está asociada con variaciones en el GFS, en especial, en aquellas familias de genes implicadas en procesos biológicos como: quimiotaxis, señalización célula-célula y funciones del sistema inmune. Los genes pertenecientes a dichas familias, se expresaron de manera relevante en la neocorteza en humanos durante el desarrollo cerebral temprano, en comparación con adultos (Figura 14) (Castillo-Morales et al., 2016).

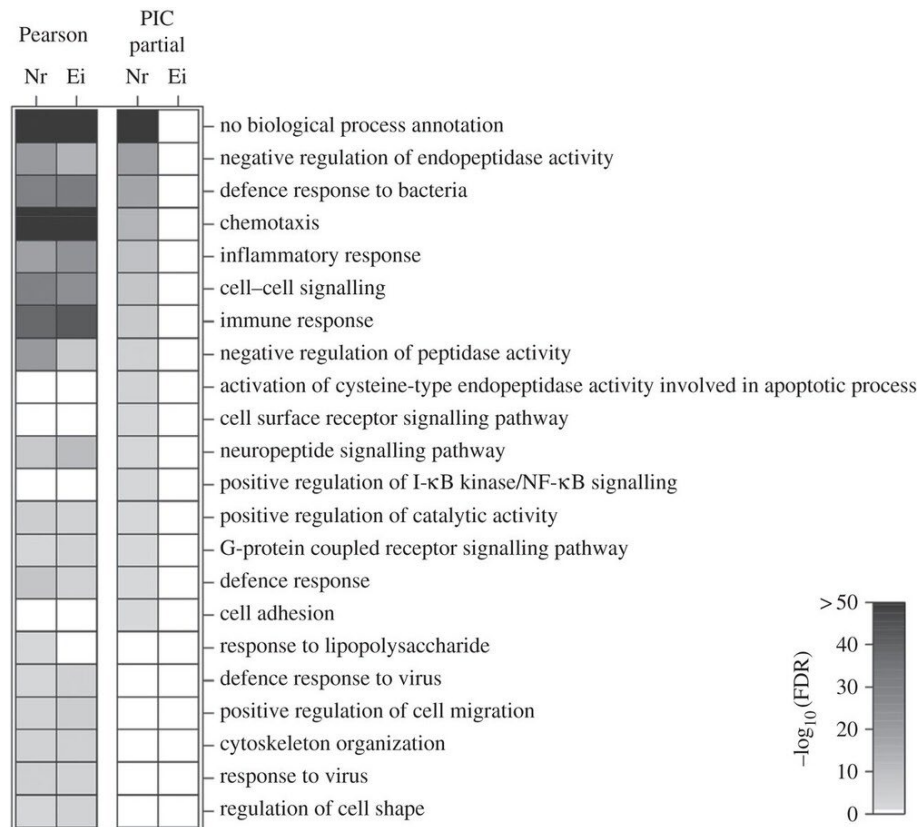


Figura 14. Análisis de enriquecimiento de ontología génica de familias con variaciones del tamaño de la familia génica (GFS) en línea con el índice de encefalización (Ei) y la relación de tamaño de neocorteza en cerebro (Nr). *HeatMap* de la sobrerrepresentación de los términos GO (expresado como p -value corregido por Benjamini-Hochberg (BH)), entre las familias de genes con variaciones de tamaño más significativamente asociadas con Ei y Nr. Las dos primeras columnas corresponden a familias de genes con la asociación más significativa entre GFS y Ei o Nr, respectivamente ($rGFS\ Nr > 0$, $FDR < 0.05$, $n = 440$ y $BenjaminirGFS\ Ei > 0$, $FDR < 0.05$, $n = 321$). Las columnas tercera y cuarta representan términos GO enriquecidos entre familias de genes cuyas variaciones en GFS muestran la asociación más significativa con uno de los fenotipos cerebrales después de tener en cuenta la varianza compartida con el otro fenotipo neural, así como la relación filogenética de las especies analizadas mediante análisis de contraste independiente ($rPIC(Nr \sim Ei + MLSP)$, $PIC(GFS \sim Ei + MLSP) > 0$, $FDR < 0.05$, $n = 272$ y $rPIC(Ei \sim Nr + MLSP)$, $PIC(GFS \sim Nr + MLSP) > 0$, $FDR < 0.05$, $n = 0$ respectivamente). Solo los términos GO enriquecidos significativamente después de la corrección de pruebas múltiples de BH se muestran en las figuras. Imagen tomada de: (Castillo-Morales et al., 2016).

Durante el desarrollo postnatal, regiones del cerebro, como la corteza temporal, parietal y frontal, presentan un incremento más marcado en comparación con otras regiones cerebrales, y posiblemente relacionada con diferencias en la maduración sináptica y susceptibilidades a experiencias tempranas aunado a factores ambientales (Lui et al., 2011; Rakic, 2009). Durante la adquisición de habilidades, las regiones cerebrales involucradas en esta tarea inicialmente, aumentan de volumen, seguidas de un proceso de renormalización, en el cual se eliminan los recursos neuronales innecesarios. Este mecanismo dinámico permite que las estructuras cerebrales se adapten eficientemente a nuevas habilidades sin un crecimiento excesivo y perpetuo (Wenger et al., 2017).

A nivel celular, se ha observado que la proporción neurona/glia no aumenta de manera uniforme con el tamaño del cerebro, sino que varía en función de la disminución de la densidad neuronal, es decir que no implica una pérdida de la capacidad cognitiva, sino una reorganización estructural que hace que el cerebro sea más eficiente en el procesamiento, conectividad y plasticidad neuronal (Herculano-Houzel, 2014). El tamaño medio de las células neuronales es altamente variable, y el de las células gliales se mantiene relativamente constante en todas las estructuras cerebrales en diferentes especies (Herculano-Houzel, 2014).

Modelos computacionales también han demostrado que las variaciones en la cinética de la neurogénesis pueden explicar la diversidad del número de neuronas corticales entre especies (Figura 15), particularmente en las capas superiores, que son las más influenciadas durante el desarrollo cortical (Cahalane et al., 2014). La variabilidad en la proporción de células neuronales y gliales, así como en su distribución regional, sugiere que las firmas genómicas que impulsaron la expansión cortical, podrían haber favorecido la generación de redes neuronales más complejas y especializadas (Florio et al., 2018; Florio & Huttner, 2014).

Recientemente se han identificado firmas genómicas responsables de la expansión cortical, destacando genes como *DUF1220*, cuyo número de copias es significativamente mayor en humanos y neandertales en comparación con chimpancés; el gen *FOXP2*, que codifica un factor de transcripción altamente expresado en el neocórtex fetal y adulto; y *SRGAP2C*, una proteína activadora de la GTPasa *Slit-Robo Rho*, que desempeña un papel clave en la duplicación de genes específicos involucrados en el desarrollo y evolución del neocórtex. Sin embargo, estos estudios solo se han centrado en primates, incluidos los humanos, lo cual limita su aplicabilidad a nivel comparativo en otros mamíferos (Florio et al., 2017).

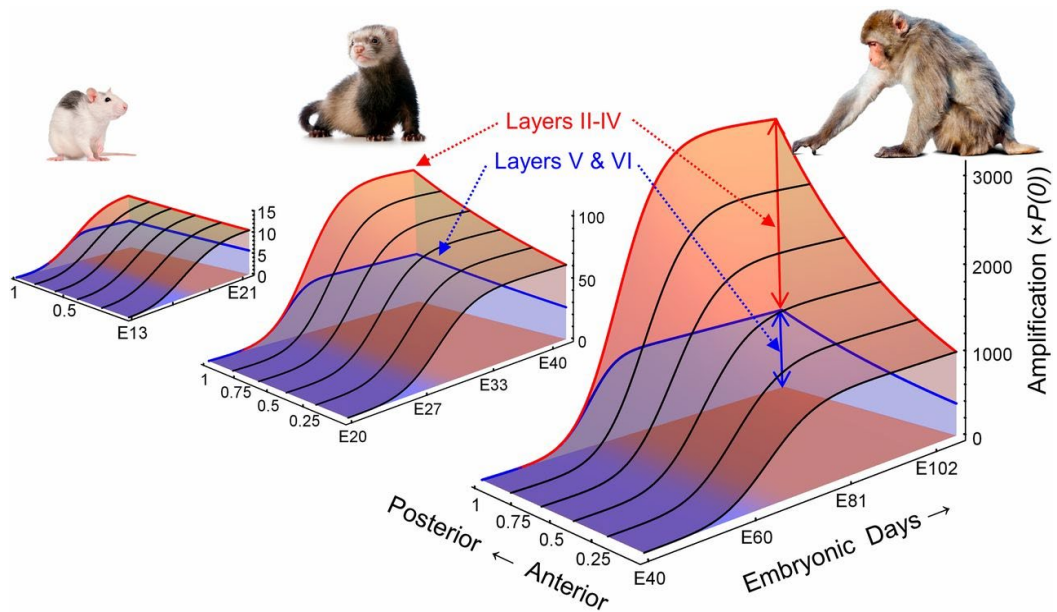


Figura 15. Diferencias entre especies e intra-corteza predichas por el modelo en el tiempo, la extensión y la asignación de capas de la producción de neuronas corticales. La ilustración muestra cantidades predichas de producción neuronal, a través del eje anterior–posterior (espacial) de la corteza en el transcurso de la neurogénesis embrionaria (eje del tiempo), para tres diferentes puntuaciones de corteza (1.0, similar a una rata; 1.75, similar a un hurón; 2.5, similar a un mono macaco). Imágenes de ratas y hurones cortesía de iStockphoto/GlobalP. Imagen de macaco cortesía de iStockphoto/JackF. Imagen tomada de: (Cahalane et al., 2014).

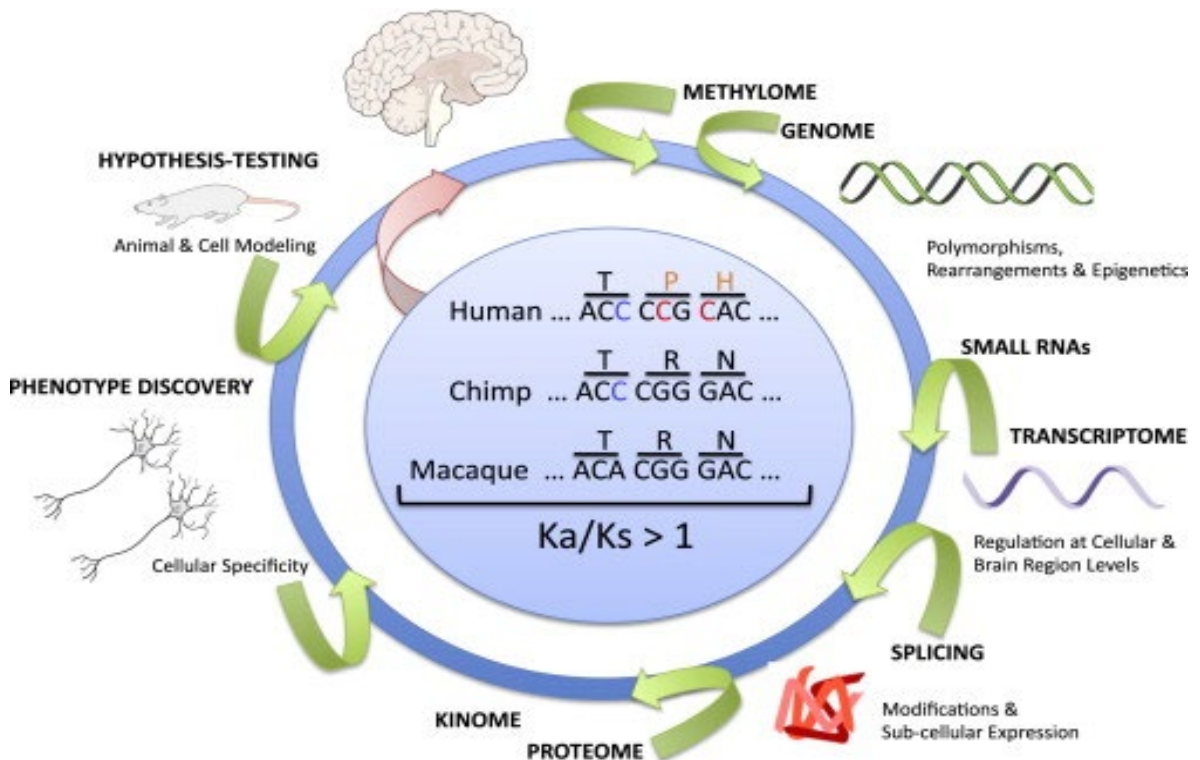


Figura 16. El esquema muestra las múltiples capas de regulación que subyacen a la evolución del cerebro humano. Niveles de regulación genómica que en conjunto podrían develar el funcionamiento de los genes implicados en la evolución del cerebro humano. Imagen tomada de: (Konopka & Geschwind, 2010).

Integrar múltiples niveles de expresión genómica, enfocados en el desarrollo del cerebro con estudios de fenotipos únicos, en una amplia variedad de especies podría revelar en un futuro, un modelo más completo de las bases genómicas que construyeron al cerebro (Konopka & Geschwind, 2010). Es decir, el comprender la evolución del cerebro implica integrar múltiples niveles de regulación genómica, desde la epigenética y el transcriptoma (por ejemplo, el metiloma, así como cambios en las isoformas por *splicing*, modificaciones postraduccionales, etc.), hasta el proteoma y el cinoma, lo que permitirá construir un modelo más completo del desarrollo cerebral, con implicaciones directas en la comprensión de la evolución del cerebro y así mismo, en el estudio de enfermedades del neurodesarrollo (Konopka & Geschwind, 2010).

Este enfoque requiere sumar también, datos de diferentes tipos celulares, tejidos, modelos animales y etapas del desarrollo, así como el uso de análisis funcionales y bioinformáticos que permitan interpretar los datos a múltiples escalas (Figura 16).

A pesar de estos avances, actualmente no existen estudios que aborden de manera integral las bases genómicas subyacentes a la evolución del cerebro a partir del conteo celular en mamíferos. La mayoría de los estudios existentes se han centrado principalmente en comprender los determinantes genómicos en grupos específicos de mamíferos, como los primates, utilizando análisis transcriptómicos específicos para explorar sólo genes específicos implicados en la expansión cortical (Florio et al., 2018), limitando la generalización de los resultados. Por lo tanto, es necesario un enfoque más amplio, que incorpore diversas especies de mamíferos para identificar patrones genómicos robustos y generalizables.

III. Hipótesis

Existe una firma genómica en la forma de expansión y contracción del tamaño de familias de genes asociada a la evolución del número de neuronas en la corteza cerebral en mamíferos la cual no puede ser explicada por la encefalización o tamaño de corteza ni como un proceso sistémico presente al evaluar el número de neuronas en el cerebro completo.

IV. Justificación

Comprender los determinantes genómicos que acompañaron la evolución de la corteza cerebral es fundamental para responder una de las preguntas centrales de la neurociencia evolutiva: ¿qué mecanismos biológicos sustentan el desarrollo de funciones cognitivas superiores en mamíferos, particularmente en humanos? (Florio et al., 2018). Este conocimiento no solo permitirá explicar cómo se configuró la arquitectura cerebral a lo largo de la evolución, sino también establecer comparaciones entre especies que han seguido trayectorias evolutivas distintas.

La genómica comparativa ofrece una herramienta clave para este propósito, ya que permite vincular variaciones genómicas con diferencias fenotípicas, como el tamaño del cerebro y el número de neuronas corticales (Florio et al., 2018; Florio & Huttner, 2014). Los avances recientes en secuenciación de genomas y el uso de organismos modelo no convencionales hacen posible analizar, de manera más amplia, la evolución de familias de genes y sus implicaciones funcionales en múltiples linajes de mamíferos (Konopka & Geschwind, 2010).

El estudio de la corteza cerebral tiene una relevancia doble. En primer lugar, desde la perspectiva evolutiva, porque esta estructura es esencial para el procesamiento de información compleja y la base de funciones cognitivas avanzadas. En segundo lugar, desde la perspectiva biomédica, porque alteraciones en el desarrollo cortical o disfunciones en sus redes neuronales se asocian con diversos trastornos del neurodesarrollo y enfermedades neuropsiquiátricas, cuyas bases moleculares aún se desconocen en gran medida (Lodato & Arlotta, 2015).

Por ello, identificar las firmas genómicas asociadas al incremento en el número de neuronas corticales no solo aporta una comprensión más profunda de la evolución del cerebro, sino que también abre la posibilidad de esclarecer por qué este órgano es particularmente vulnerable a trastornos neurodegenerativos y del desarrollo. Este estudio, por tanto, se justifica tanto por su valor científico en el entendimiento de la evolución cerebral, como por sus implicaciones en la biología de la salud y la enfermedad.

V. Objetivos

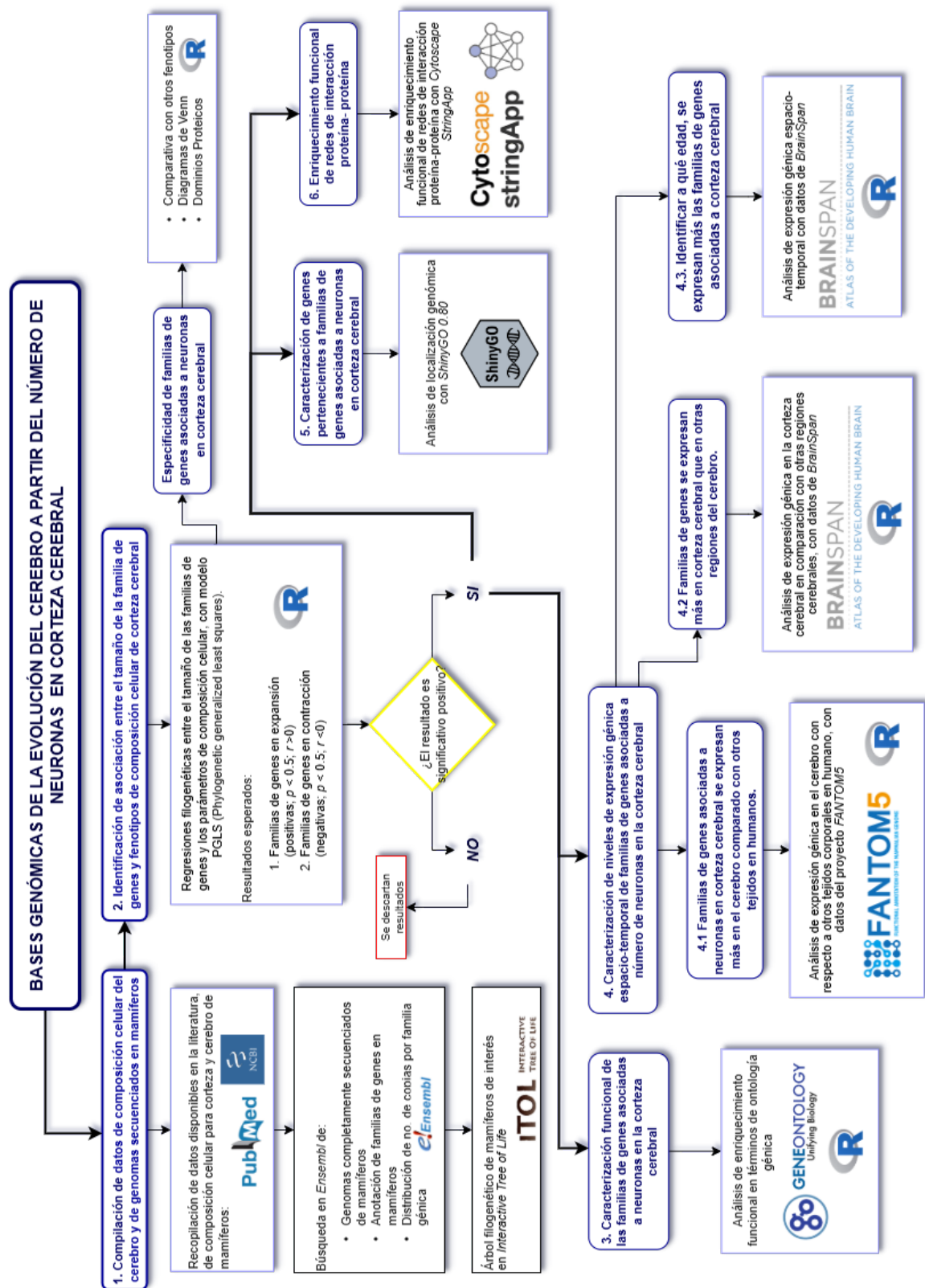
5.1. Objetivo general

Identificar las firmas genómicas que subyacen de la composición celular por conteo de número de neuronas en la evolución de la corteza cerebral.

5.2. Objetivos específicos

1. Compilar datos de composición celular del cerebro en mamíferos y de sus genomas secuenciados.
2. Identificar la asociación entre el tamaño de la familia de genes y la composición celular en la corteza cerebral.
3. Especificidad de familias de genes
4. Caracterizar funcionalmente las familias de genes asociadas a la composición celular en la corteza cerebral. A través de enriquecimiento funcional en términos *GO*.
5. Caracterizar niveles de expresión espacio-temporal de familias de genes asociadas a composición celular en la corteza cerebral.
 - 4.1 Identificar si las familias de genes asociadas a la corteza cerebral se expresan más en el cerebro, comparado con otros tejidos en humano
 - 4.2 Identificar si las familias de genes se expresan más en corteza cerebral que en otras regiones del cerebro.
 - 4.3 Identificar en qué etapa del desarrollo cerebral se expresan más las familias de genes en corteza cerebral.
6. Caracterizar los genes de las familias de genes asociadas a neuronas en corteza cerebral.
 - 5.1 Distribución de las familias de genes de neuronas en corteza cerebral en el genoma.
 - 5.2 Enriquecimiento funcional de redes de interacción proteína- proteína.

VI. Estrategia experimental



VII. Materiales y Métodos

7.1. Datos de composición celular del cerebro y de genomas secuenciados en mamíferos

De la literatura disponible, se recopilaron datos publicados sobre composición celular de áreas específicas del cerebro en mamíferos (ver Tabla 1), considerando únicamente aquellas especies para las cuales se dispone de genoma secuenciado. Los datos incluyeron el número total de células neuronales y no neuronales, así como el peso cerebral (en gramos) por región analizada.

En el caso de las células no neuronales, se empleará el término células gliales o simplemente glía, el cual engloba a diversos tipos celulares: oligodendrocitos, astrocitos, microglías y células endoteliales. Estas células constituyen una proporción significativa del tejido cerebral: se estima que representan entre el 50 % y el 90 % del total de células en el cerebro, dependiendo de la región cerebral y de la especie analizada (Azevedo et al., 2009; Herculano-Houzel, 2014).

Para asegurar una distribución normal de la variable de conteo celular, estos valores fueron transformados logarítmicamente ($\log_{10}$), siguiendo la metodología propuesta por Buzsáki & Mizuseki, 2014. La transformación y procesamiento de los datos se realizó en el entorno de programación R utilizando el paquete “dplyr” (Wickham et al., 2014). Todos los análisis posteriores que involucran esta variable se basan en su versión transformada logarítmicamente.

Tabla 1. Datos de composición cerebral en especies distintas de mamíferos con genoma de referencia disponible

Especie	Neuronas en cerebro ($\times 10^6$)	Masa cortical relativa (gr)	Neuronas en corteza cerebral ($\times 10^6$)	Masa corporal (gr)	Masa cerebral (gr)	Referencias
Tití común (<i>Callithrix jacchus</i>)	635.80	5.561	244.72	361	7.78	(Herculano-Houzel et al., 2007)
Perro doméstico (<i>Canis familiaris</i>)	2252.69	65.491	527.91	19700	86.462	(Jardim-Messeder et al., 2017)
Cuy (<i>Cavia porcellus</i>)	233.56	1.938	43.51	311	3.656	(Herculano-Houzel et al., 2011)
Gato doméstico (<i>Felis catus</i>)	1215.21	24.176	249.83	4500	34.858	(Jardim-Messeder et al., 2017)
Gorila (<i>Gorilla gorilla</i>)	33400	377.2	9100	139842	500	(Herculano-Houzel & Kaas, 2011)
Rata topo (<i>Heterocephalus glaber</i>)	26.88	0.184	6.15	23.3	0.392	(Herculano-Houzel et al., 2011)
Humano (<i>Homo sapiens</i>)	86060	1232.93	16340	65000	1508.91	(Azevedo et al., 2009)

Elefante (<i>Loxodonta africana</i>)	257043.47	2847.594	5593.241033	2750000	4480	(Herculano-Houzel et al., 2014)
Macaco cangrejero (<i>Macaca fascicularis</i>)	3440	36.226	800.96	5700	62.1	(Herculano-Houzel & Kaas, 2011)
Macaco Rhesus (<i>Macaca mulatta</i>)	6376.16	69.832	1710	3900	87.346	(Herculano-Houzel et al., 2007)
Lémur ratón gris (<i>Microcebus murinus</i>)	254.71	0.908	22.31	0.54	1.8	(Gabi et al., 2010)
Ratón (<i>Mus musculus</i>)	67.87	0.173	13.69	40.4	0.402	(Herculano-Houzel et al., 2011)
Hurón (<i>Mustela putorius furo</i>)	404.43	3.123	38.95	809	5.429	(Jardim-Messeder et al., 2017)
Liebre (<i>Oryctolagus cuniculus</i>)	494.20	4.448	71.45	46000	9.1	(Herculano-Houzel et al., 2011)
Galago de Garnet (<i>Otolemur garnettii</i>)	936	6.29	226.09	1314	10.15	(Herculano-Houzel et al., 2007)
Babuino oliva (<i>Papio anubis</i>)	10950	120.214	2880	8000	151.19	(Gabi et al., 2010)
Rata (<i>Rattus norvegicus</i>)	188.87	0.769	31.02	315.1	1.724	(Herculano-Houzel et al., 2011)
Cerdo (<i>Sus scrofa</i>)	2220	35.78	292.96	1000	57.758	(Kazu et al., 2014)

gr = gramos; $lg_{10} = \times 10^6$

7.2. Anotaciones de familias de genes

Se obtuvieron anotaciones de familias de genes de 78 especies de mamíferos a partir de la base de datos especializada *Ensembl* 2020 (Yates et al., 2020). Durante la revisión inicial, se identificó que algunas especies presentaban un número inusualmente bajo de genes codificadores de proteínas, probablemente como consecuencia de una baja cobertura de secuenciación o de anotaciones genómicas incompletas.

Para evaluar si esta observación se debía a una baja calidad del genoma en lugar de a una variación biológica real en el número de genes, se definió un conjunto de genes de núcleo de copia única. Este conjunto consistió en todos aquellos genes presentes en al menos el 90 % de las especies y anotados siempre como una sola copia. Se observó que los genomas con menos de 16,000 genes codificadores de proteínas anotados también carecían de un número desproporcionadamente alto de estos genes núcleo de copia única.

Como medida de calidad, se excluyeron del análisis aquellas especies con menos de 16,000 genes codificadores de proteínas anotados, lo que permitió eliminar gran parte de la asociación entre el número total de genes anotados y el recuento de genes núcleo de copia única. A partir de este filtrado, se seleccionaron 18 especies de mamíferos con genomas de calidad suficiente y datos disponibles de composición celular cerebral. El conjunto final incluyó 3,854 familias de genes, correspondientes a las 18 especies seleccionadas. Para su inclusión en el análisis comparativo, se aplicaron dos criterios adicionales:

1. Cada familia debía estar presente en al menos el 80 % de las especies, con el fin de eliminar familias de genes de evolución reciente, restringidas a linajes específicos.
2. Cada familia debía contener al menos dos genes en una o más especies, para descartar familias sin variación en el número de copias o genes de copia única cuya variación se limitará a eventos aislados de duplicación o delección.
3. Cabe aclarar que las familias de genes no se definieron en función de la presencia de dominios individuales, sino mediante relaciones evolutivas de homología; los dominios proteicos fueron empleados como descriptores funcionales complementarios.

Tras aplicar ambos filtros, el conjunto final quedó compuesto por 2,777 familias de genes. La mayoría de estas familias génicas presentan un bajo número de copias, mientras que un subconjunto reducido muestra expansiones notables, reflejando la distribución esperada de familias génicas en genomas de mamíferos como se muestra en la Figura suplementaria 2S, utilizando el paquete “ggplot2” en R (Wickham, 2016).

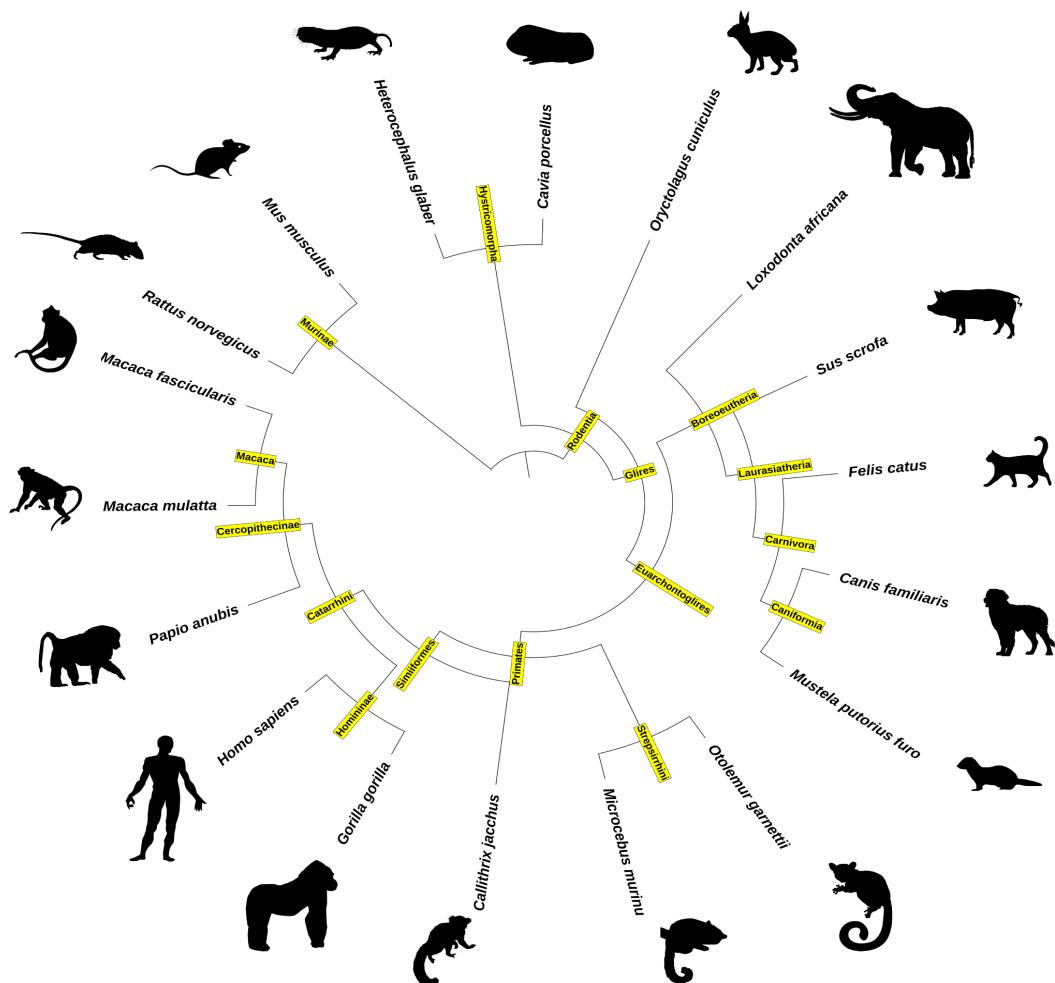


Figura 17. Árbol filogenético (politomía no resuelta) de las 18 especies de mamíferos analizadas, con genoma de referencia y con datos de composición celular cerebral. Las especies son: Tití común (*Callithrix jacchus*), Perro

doméstico (*Canis familiaris*), Cuy (*Cavia porcellus*), Gato doméstico (*Felis catus*), Gorila (*Gorilla gorilla*), Rata topo (*Heterocephalus glaber*), Humano (*Homo sapiens*), Elefante (*Loxodonta africana*), Macaco cangrejero (*Macaca fascicularis*), Macaco Rhesus (*Macaca mulatta*), Lémur ratón gris (*Microcebus murinus*), Ratón (*Mus musculus*), Hurón (*Mustela putorius furo*), Liebre (*Oryctolagus cuniculus*), Galago de Garnet (*Otolemur garnettii*), Babuino oliva (*Papio anubis*), Rata (*Rattus norvegicus*) y Cerdo (*Sus scrofa*). El árbol filogenético fue realizado sin escalas con iTol, las ilustraciones de siluetas son de phylopic.org (Keesey, 2024).

Para representar visualmente las relaciones evolutivas entre las especies incluidas, se generó un árbol filogenético utilizando la herramienta *Interactive Tree of Life (iTOL*; <https://itol.embl.de/>) (Letunic & Bork, 2024). Además, las ilustraciones de las especies fueron obtenidas del repositorio *PhyloPic* (2025) (<https://www.phylopic.org/>) para facilitar la identificación visual de cada especie (Keesey, 2024). Se construyó un árbol filogenético con el fin de representar visualmente las relaciones evolutivas entre las especies seleccionadas, la topología no refleja relaciones filogenéticas, ya que todas las ramas parten de un único nodo común con longitudes iguales, para indicar que no se infirieron relaciones evolutivas en este análisis. El diagrama se incluye únicamente para listar los taxones analizados (Figura 17).

7.3. Regresiones filogenéticas entre el tamaño de las familias de genes y los parámetros de composición celular

Se empleó un modelo de regresión con corrección filogenética para evaluar la asociación entre las variaciones en el tamaño de las familias de genes y la variación fenotípica en composición celular entre especies, descartando el efecto de la ascendencia evolutiva compartida. Específicamente, se implementó un modelo de regresión de mínimos cuadrados generalizados filogenéticamente (PGLS, por sus siglas en inglés: *Phylogenetic Generalized Least Squares*) (Grafen, 1992). Este modelo utiliza una matriz de varianza-covarianza filogenética derivada del árbol filogenético para ajustar la dependencia entre especies y realizar un análisis de regresión lineal más preciso. Las regresiones filogenéticamente corregidas se realizaron con el paquete “nlme” en R (Pinheiro & Bates, 2023), asumiendo un modelo evolutivo de movimiento browniano (modelo evolutivo que describe cómo cambian los rasgos a lo largo del tiempo en un árbol filogenético) y aplicando el método de máxima verosimilitud para estimar los parámetros del modelo.

Tras ejecutar los modelos PGLS, se calcularon los valores de correlación (r) a partir de los valores t , y los valores de significancia (p -value) fueron ajustados por pruebas múltiples mediante el método de Benjamini-Hochberg (Benjamini & Hochberg, 1995), con

el fin de controlar la tasa de descubrimientos falsos (FDR, por sus siglas en inglés: *False Discovery Rate*). Este procedimiento sigue los siguientes pasos:

1. Se ordenan los valores p en orden ascendente
2. Se calcula un valor crítico para cada valor p , utilizando la fórmula:

$$\left(\frac{i}{m}\right) \cdot Q$$

donde i es el rango del p -value, m es el número total de pruebas y Q la FDR deseada.

3. Se identifica el p -value más grande que sea menor al valor crítico correspondiente.
4. Y finalmente se designa a cada p -value que sea menor que el valor crítico como significativo.

Este método es más poderoso que la corrección de Bonferroni, ya que clasifica cada hipótesis de acuerdo a su p -value y posteriormente se asigna un umbral diferente y por consiguiente esto permite controlar la tasa de descubrimientos falsos de una manera más efectiva (Glueck et al., 2008).

Inicialmente, se aplicó un modelo PGLS simple, es decir, con un solo predictor (una variable fenotípica como independiente y el tamaño de familia de genes como dependiente). Sin embargo, al evaluar la especificidad del tamaño de familias de genes (GFS) entre distintos fenotipos, se observó un solapamiento considerable de familias de genes asociadas a más de un fenotipo (ver Anexos, Figura 1S). Esto sugiere una alta correlación entre fenotipos, posiblemente derivada de su evolución filogenéticamente correlacionada.

Para abordar esta covariación, se implementaron regresiones filogenéticas secuenciales (Dormann et al., 2013; Graham, 2003). En este enfoque, se introdujeron simultáneamente dos variables predictoras en el modelo: el fenotipo de interés y los residuos de la regresión entre un segundo fenotipo y el fenotipo principal. Este modelo, que denominamos PGLS doble de predictor único, permite controlar el efecto compartido entre fenotipos y estimar la contribución específica de cada uno a la variación en el tamaño de las familias de genes (Dormann et al., 2013; Graham, 2003).

Las familias de genes que presentaron una correlación positiva y significativa ($p < 0.05$; $r > 0$), fueron clasificadas como familias en expansión. Por el contrario, aquellas con una correlación negativa y significativa ($p < 0.05$; $r < 0$) fueron clasificadas como familias en contracción.

Para evaluar la especificidad de estas asociaciones entre fenotipos, se generaron diagramas de Venn, utilizando el paquete “ggplot2” en R (Wickham, 2016). Así mismo, se construyeron gráficos de barras apiladas o porcentajes, para representar la proporción

relativa de familias de genes en expansión (significativas positivos) y contracción (significativas negativos), por cada fenotipo, utilizando el paquete “ggplot2” en R (Wickham, 2016).

7.3.1 Caracterización funcional de familias génicas mediante dominios proteicos en genes

Para caracterizar la arquitectura proteica de las familias génicas asociadas al número de neuronas en corteza cerebral, se analizaron los dominios *InterPro* presentes en las proteínas codificadas por dichas familias. Se calculó la frecuencia relativa de cada dominio por especie y se realizaron comparaciones entre perfiles de dominios, sin recurrir a una nueva clasificación funcional, de esta manera, la comparación se centró en la arquitectura proteica subyacente y no en la redundancia funcional, utilizando el paquete “ggplot2” en R (Wickham, 2016).

7.4. Análisis de enriquecimiento funcional en términos de ontología génica

Las anotaciones funcionales de términos de ontología génica (*Gene Ontology*, GO) para cada gen de cada especie se obtuvieron de *Gene Ontology Consortium* (www.geneontology.org), un recurso que proporciona información sobre las funciones biológicas, moleculares y celulares de los genes y sus productos en una amplia variedad de especies (Ashburner et al., 2000).

Para este análisis, se utilizaron exclusivamente las familias de genes significativamente asociadas a los fenotipos de interés, derivadas de los resultados del modelo PGLS doble de predictor único. A cada familia de genes se le asignó un término GO si al menos uno de sus genes en cualesquiera de las 18 especies de mamíferos incluidas presentaba dicha anotación (Yates et al., 2020). Con el fin de evitar sesgos debidos a términos GO poco representados, aquellos asociados con menos de 50 familias de genes se agruparon bajo una categoría común denominada "*small GO*". Esta agrupación previene la sobre interpretación de categorías poco informativas y ayuda a conservar el poder estadístico del análisis (Castillo-Morales et al., 2014). Las familias sin ninguna anotación funcional GO fueron clasificadas dentro de una categoría "*unknown GO*".

El análisis de enriquecimiento funcional se realizó comparando la proporción de familias de genes asociadas a cada término GO en el conjunto de interés, frente a su proporción esperada en 1,000 muestras aleatorias de igual tamaño, extraídas del conjunto de fondo (todas las familias de genes disponibles). Para cada término GO, se estimó un *p-value*

a partir de la media y la desviación estándar de su frecuencia en las muestras aleatorias. Estos análisis se realizaron en R utilizando los paquetes “parallel” y “biomaRt” (Bengtsson, 2020; Durinck et al., 2005). Los resultados se visualizaron mediante mapas de calor (*heatmaps*), que representan gráficamente los procesos biológicos enriquecidos en cada uno de los fenotipos.

Adicionalmente, se identificaron los genes humanos pertenecientes a las 385 familias de genes significativamente asociadas a neuronas en corteza cerebral, obtenidas del análisis PGLS doble de único predictor. Para ello, se utilizó la herramienta *BioMart* disponible en Ensembl (<https://www.ensembl.org/biomart>), seleccionando el dataset *Ensembl genes 110*, para *Homo sapiens* (GRCh38.p14), con los atributos *Gene stable ID* y *PANTHER ID*. El filtrado se realizó con los paquetes “dplyr” (Wickham et al., 2014) y “biomaRt” (Durinck et al., 2005) en R, resultando en un total de 3,500 genes humanos pertenecientes a las 385 familias asociadas a neuronas en corteza cerebral.

7.5. Análisis de expresión génica en el cerebro con respecto a otros tejidos corporales a partir de datos del proyecto *FANTOM5*

Con base en los genes identificados dentro de las familias de genes significativamente asociadas de manera positiva con neuronas en la corteza cerebral (derivadas del modelo PGLS doble de único predictor), se llevaron a cabo análisis posteriores de expresión génica utilizando datos del proyecto *FANTOM5* (*Functional Annotation of the Mammalian Genome 5*) (<https://fantom.gsc.riken.jp/5/>).

FANTOM5 proporciona un conjunto integral de datos de expresión génica obtenidos mediante la técnica *CAGE* (*Cap Analysis of Gene Expression*), una metodología de secuenciación de ARN que permite mapear con alta precisión los promotores activos y los sitios de inicio de transcripción (TSS, *Transcription Start Sites*) en todo el genoma de forma no sesgada (Hurst et al., 2015). Esta técnica captura los extremos 5' de los transcritos, proporcionando una visión detallada de la regulación transcripcional. Los datos disponibles incluyen niveles de expresión génica en una amplia gama de tejidos y tipos celulares, tanto en humanos como en ratones. Los perfiles de expresión están organizados en dos categorías principales: etapa fetal y etapa adulta (Lizio et al., 2015).

Para estandarizar los niveles de expresión génica, los datos fueron normalizados mediante *Z-score* utilizando el lenguaje de programación R, siguiendo la metodología

descrita por Hurst et al. (2015), con adaptaciones específicas a los objetivos de este estudio. La fórmula utilizada fue:

$$Z = \frac{(x-\mu)}{\sigma}$$

Donde:

- x es el valor de expresión génica observado en un gen,
- μ es la media de expresión génica esperada para todos los genes
- σ es la desviación estándar de la expresión esperada.

Esta normalización permite comparar directamente la expresión relativa de genes con distintos niveles de expresión basal. Un *Z-score* positivo indica que un gen presenta mayor expresión que la media esperada, mientras que un *Z-score* negativo indica una expresión disminuida. Si se asume que no existen cambios netos en la expresión global entre tejidos, la mediana del *Z-score* tiende a situarse cerca de cero. Para la manipulación, filtrado y organización de los datos, se utilizó el paquete “dplyr” (Wickham et al., 2014) en R, mientras que la visualización de los resultados se realizó mediante gráficos de densidad contruidos con el paquete “ggplot2” (Wickham, 2016) en R.

7.6. Análisis de expresión génica en la corteza cerebral en comparación con otras regiones en cerebro

A partir de los genes identificados en las familias génicas significativamente asociadas con neuronas de la corteza cerebral, se analizaron los patrones de expresión génica espacio-temporal en diferentes regiones del cerebro humano. Para ello, se utilizaron datos transcriptómicos provenientes de la base de datos “*RNA-seq Gencode v10 summarized to genes*” del consorcio *BrainSpan* (<https://www.brainspan.org>) (Miller et al., 2014).

BrainSpan recopila perfiles de expresión génica en 16 regiones del cerebro, que incluyen la corteza cerebelosa, el núcleo mediodorsal del tálamo, el estriado, la amígdala, el hipocampo y once áreas de la corteza cerebral, obtenidos a partir de muestras postmortem de 57 donantes humanos de ambos sexos, de diversas etnias y rangos de edad que abarcan desde el desarrollo embrionario hasta la edad adulta tardía (Kang et al., 2011). Para facilitar la comparación de la dinámica de expresión, los datos fueron organizados en tres grupos etarios, definidos con base en las etapas clave del desarrollo cerebral:

- Fetal: incluye edades expresadas en semanas post-concepción (pcw), correspondientes a etapas prenatales. Las edades reportadas en *BrainSpan* fueron: 8pcw, 9pcw, 12pcw, 13pcw, 16pcw, 17pcw, 19pcw, 21pcw, 24pcw, 25pcw, 26pcw, 35pcw y 37pcw. Esta etapa ha sido asociada con una alta actividad proliferativa y de diferenciación en el sistema nervioso central (SNC) (Dobbing & Sands, 1973).
- Infantil: Comprende las edades de 4 meses a 8 años: 4mos, 10mos, 1yrs, 2yrs, 3yrs, 4yrs y 8yrs. Durante esta etapa, se ha reportado una intensa actividad en la sinaptogénesis y maduración neuronal, especialmente en los primeros cuatro años de vida, con una desaceleración posterior hasta alcanzar el 90% del volumen cerebral final alrededor de los seis años (Huttenlocher et al., 1982).
- Adulto: Agrupa las edades de 18 a 40 años: 18yrs, 19yrs, 21yrs, 23yrs, 30yrs, 36yrs, 37yrs y 40yrs. Esta fase se caracteriza por procesos de mantenimiento cerebral y posterior degeneración progresiva, marcada por un adelgazamiento no lineal de la corteza cerebral (Marsh et al., 2008).

Consideraciones sobre la exclusión de datos. Las edades intermedias 11yrs, 13yrs y 15yrs, también disponibles en *BrainSpan*, fueron excluidas debido a que coinciden con la pubertad, etapa caracterizada por intensos cambios hormonales que pueden inducir variabilidad significativa en los patrones de expresión génica, generando posibles sesgos en el análisis (Cameron, 2004).

Para cada grupo etario se calculó el promedio de los niveles de expresión génica, considerando únicamente estructuras de la corteza cerebral, incluido el hipocampo. La lista completa de las estructuras incluidas se encuentra detallada en los Anexos (Tabla S2). Posteriormente, se realizó un análisis comparativo en R para evaluar los patrones de expresión de las familias de genes significativamente asociadas a neuronas en corteza cerebral, con énfasis en su comportamiento en esta región frente a otras áreas cerebrales. Para la manipulación, clasificación y análisis de los datos se utilizaron los paquetes estadísticos “dplyr” (Wickham et al., 2014) y “ggplot2” (Wickham, 2016). Los resultados se visualizaron mediante gráficos de densidad.

7.7. Análisis de expresión génica espacio-temporal

A partir de los datos transcriptómicos de la base de datos “*RNA-seq Gencode v10 summarized to genes*” del consorcio *BrainSpan* (<https://www.brainspan.org>) (Miller et al.,

2014) y con base en las modificaciones descritas previamente, se diseñó un enfoque adicional para evaluar la dinámica espacio-temporal de la expresión génica asociada a las familias de genes vinculadas con neuronas de la corteza cerebral. Con este propósito, se construyeron dos bases de datos independientes, considerando la distribución espacial (región cerebral) y temporal (edad).

Los datos se clasificaron en grupos etarios específicos: 8pcw, 9pcw, 12pcw, 13pcw, 16pcw, 17pcw, 19pcw, 21pcw, 24pcw, 25pcw, 26pcw, 35pcw, 37pcw, 4mos, 10 mos, 1yrs, 2yrs, 3yrs, 4yrs, 8yrs, 18yrs, 19yrs, 21yrs, 23yrs, 30yrs, 36yrs, 37yrs y 40yrs (*pcw* = *semanas post-concepción*; *mos* = *meses*; *yrs* = *años*; Miller et al., 2014, 2017). Esta segmentación permitió evaluar la evolución de la expresión génica a lo largo del desarrollo cerebral humano, desde etapas embrionarias hasta la adultez, de manera más general.

- Base de datos 1 (normalizada). Los datos de expresión génica fueron normalizados mediante *Z-score*, lo cual permitió una comparación relativa entre genes, independientemente de su nivel absoluto de expresión. Esta transformación estandariza los datos y facilita la identificación de tendencias globales en la actividad génica a lo largo del tiempo.
- Base de datos 2 (sin normalizar). En esta versión, se utilizaron los datos sin normalización, preservando la magnitud original de los niveles de expresión génica. Este enfoque permite observar directamente los valores absolutos de expresión, y puede ser útil para interpretar la relevancia funcional de genes altamente expresados.

Para ambas bases de datos, se emplearon los paquetes estadísticos “dplyr” (Wickham et al., 2014) para el procesamiento y clasificación de los datos, y “ggplot2” (Wickham, 2016) para la generación de gráficos de regresión lineal, con el fin de visualizar los cambios en la expresión génica a lo largo del tiempo.

7.8. Caracterización de genes pertenecientes a familias asociadas a corteza cerebral

Con base en los genes identificados en las familias génicas significativamente asociadas con neuronas en la corteza cerebral, obtenidos a partir del análisis de PGLS doble de predictor único, se realizó una caracterización funcional, utilizando la herramienta bioinformática *ShinyGO 0.80* (<http://bioinformatics.sdstate.edu/go/>). Esta plataforma web permite realizar análisis de enriquecimiento funcional, visualización de datos genómicos y exploración de

redes biológicas y otras funciones relacionadas con genómica y proteómica, en una interfaz web accesible (Ge et al., 2020).

Se realizó un análisis de localización cromosómica para generar un mapa visual de los genes pertenecientes a las familias de genes asociadas a neuronas en corteza cerebral, con el objetivo de identificar regiones específicas del genoma humano enriquecidas con estos genes. Los parámetros establecidos para el análisis fueron los siguientes:

- Especie de referencia: *Homo sapiens*.
- Corrección por pruebas múltiples: Se aplicó un $FDR \leq 0.05$ utilizando el método de Benjamini-Hochberg, para reducir la tasa de falsos positivos (Glueck et al., 2008).
- Número de vías (“*Pathways to show*”): Se configuró para mostrar 20 vías biológicas, permitiendo una visión general y enfocada de los resultados (Ge et al., 2020).
- Tamaño de vía (“*Pathway size Min. / Max.*”). Se estableció un rango de 10 a 500 genes por vía, lo que facilita la inclusión de rutas biológicas con relevancia funcional y evita sesgos hacia vías muy grandes o muy pequeñas dando así una visión más general (Ge et al., 2020).
- Eliminación de Redundancia: Se activó la opción “*Remove Redundancy*” para eliminar términos funcionales redundantes y conservar aquellos más representativos (Ge et al., 2020).
- Visualización de identificadores de vía: Se habilitó la opción “*Show Pathway IDs*” para mostrar los identificadores únicos de cada vía junto a sus nombres.
- Base de datos funcional seleccionada: Se seleccionó “*GO Biological Process*”, lo cual permitió identificar procesos biológicos enriquecidos entre los genes analizados.

El análisis de enriquecimiento funcional se basó en una prueba hipergeométrica, que evalúa si un conjunto de genes está significativamente sobrerrepresentado en una categoría funcional más allá de lo esperado por azar. Para visualizar la distribución de los genes a lo largo del genoma humano, se utilizó la herramienta “*GENOME*” integrada en *ShinyGO 0.80*. Este módulo permite detectar agrupaciones de genes mediante un modelo estadístico de ventana deslizante (“*sliding window*”), que evalúa la distribución de genes a lo largo del genoma, con los siguientes parámetros:

- Tamaño de ventana (“*Window Size*”): Se estableció en 6 Megabases (Mb) para obtener una resolución adecuada que permita identificar agrupaciones de genes a lo largo del genoma (Cirulli et al., 2022).

- Número de pasos por ventana ("*Steps in a window*"): Se configuró en 2, para obtener una visión más general de la distribución genómica (Cirulli et al., 2022).
- Corrección por *FDR*: Se estableció en 0.05, con el método de Benjamini-Hochberg (Glueck et al., 2008).

Este análisis permitió identificar clústeres de genes ubicados en regiones específicas del genoma, lo que sugiere posibles mecanismos de regulación compartida o duplicaciones génicas localizadas.

7.9. Análisis de enriquecimiento funcional de redes de interacción proteína- proteína

Se realizó un análisis de interacción proteína-proteína a partir del conjunto de genes pertenecientes a las familias de genes significativamente asociados a neuronas en corteza cerebral, derivadas del modelo PGLS doble de predictor único. La conversión de identificadores de genes a identificadores de proteínas en formato "*UniProt Swiss-Prot*" se llevó a cabo utilizando el paquete "*Biomart*" (versión 2.50.1) en R (Durinck et al., 2009). Para construir y visualizar las redes de interacción, se empleó la herramienta "*Cytoscape StringApp*" (<https://cytoscape.org/>), que permite la integración de datos muy grandes utilizando el servidor de *STRING* (Doncheva et al., 2019). *STRING* proporciona información sobre interacción proteína-proteína derivada de una amplia variedad de fuentes, incluyendo experimentos de laboratorio, bases de datos de interacciones existentes, publicaciones científicas y métodos computacionales (Szklarczyk et al., 2023). Para la configuración de los parámetros de *Cytoscape StringApp*, se utilizó la siguiente configuración de parámetros:

- Especie de referencia: *Homo sapiens*.
- Tipo de red ("*Network type*"): Se seleccionó la opción "*physical network*", que incluye únicamente interacciones físicas directas entre proteínas, con evidencia experimental, excluyendo asociaciones funcionales directas (Szklarczyk et al., 2023).
- Umbral de confianza ("*Confidence score Cutoff*"): Se estableció en 0.9, para incluir sólo interacciones de alta confianza (Doncheva et al., 2019).
- Número máximo de interactores adicionales ("*Maximum Additional Interactors*"): Configurado en 30, para mantener redes compactas, facilitar la visualización y detectar posibles relaciones no directas con proteínas del conjunto inicial (Szklarczyk et al., 2023)

- Uso de delimitadores inteligentes (“*Use Smart Delimiters*”): Activado para facilitar el procesamiento de múltiples identificadores proteicos (Doncheva et al., 2019).
- Carga de datos de enriquecimiento (“*Load Enrichment Data*”): Activo para incorporar automáticamente datos funcionales relacionados con las proteínas analizadas (Doncheva et al., 2019).

Una vez generada la red de interacción, se aplicó un filtrado funcional utilizando la opción “*Filter STRING Enrichment table*”, configurada con los siguientes parámetros:

- Categoría funcional seleccionada (“*Select category*”): “*GO Biological Process*”, para identificar procesos biológicos enriquecidos, que están relacionados con el conjunto de proteínas de interés (Ashburner et al., 2000).
- Eliminación de términos redundantes (“*Remove redundant terms*”): Activada para mejorar la especificidad del análisis (Supek et al., 2011).
- Corte de redundancia (“*Redundancy Cutoff*”): Se establece en 0.5, eliminando términos que comparten más del 50% de las proteínas asociadas, conservando sólo los más representativos (Supek et al., 2011).

Además, se eliminaron de la red aquellas proteínas que no presentan interacciones, con el objetivo de reducir el ruido de fondo y enfocar el análisis en interacciones biológicamente relevantes (Szklarczyk et al., 2023). A partir de la tabla de enriquecimiento, se seleccionaron los seis procesos biológicos más significativos en el análisis ($p < 0.05$). Para cada proceso, se extrajo el subconjunto de proteínas asociadas y se generaron redes de interacción independientes en *Cytoscape StringApp*. Se aplicó el algoritmo de agrupamiento “*Markov Cluster*” (MCL), incluido en la misma plataforma, que permite identificar clústeres funcionales dentro de la red. Este algoritmo simula el flujo de densidad a través de las interacciones, agrupando nodos densamente conectados, lo que facilita la interpretación de módulos funcionales dentro de cada proceso biológico de interés (Doncheva et al., 2019).

Las redes generadas se visualizan mediante la interfaz gráfica de *Cytoscape StringApp*, lo cual permite una representación clara de las interacciones proteína-proteína y su asociación con funciones biológicas clave.

VIII. Resultados

Para este estudio se seleccionaron 18 especies de mamíferos de las que se dispone tanto de datos publicados sobre la composición celular de la corteza cerebral, así como de un genoma de referencia disponible. Estos datos permitieron evaluar la relación entre los parámetros de composición de células neuronales corticales y el tamaño de 2,777 de las familias de genes. La selección de especies se describe con mayor detalle en la sección de metodología.

8.1 Análisis de regresión lineal con corrección filogenética y caracterización funcional de familias de genes asociadas con cambios en la composición celular en corteza cerebral

Empleando un análisis de regresión de mínimos cuadrados generalizados (PGLS, por sus siglas en inglés: *Phylogenetic Generalised Least Squares*) que aplica análisis de regresión y tomando en cuenta la no independencia de los puntos debido a las relaciones filogenéticas de las especies, identificamos familias de genes, cuyo tamaño (GFS, por las siglas en inglés de *Gene Family Size*) esté asociado al número de neuronas en corteza en distintas especies tras corregir el umbral de significancia por tests múltiples mediante el método de Benjamini-Hochberg. Para descartar que las familias asociadas estén ligadas a la mayor cantidad de células y no a la expansión específica del número de neuronas, se realizó un PGLS de dos predictores incluyendo el número de glía como variable predictora adicional.

Cabe mencionar que el incorporar el número de células gliales como variable independiente tiene el propósito de evaluar la especificidad de la asociación entre el GFS y el número de neuronas en corteza cerebral. Lo que permite descartar que las asociaciones estén influenciadas por un efecto sistémico del total de células corticales y permite comprobar si las asociaciones genómicas son específicas para las neuronas corticales.

Este resultado, identificó 385 familias de genes significativamente asociadas al número de neuronas en la corteza cerebral, así mismo se mostró una expansión en estas familias estadísticamente positiva ($r > 0$; $p < 0.05$; Figura 18-A). En contraste, solo 14 familias de genes mostraron una contracción significativa negativa ($r < 0$; $p < 0.05$; Figura 18-A).

En un estudio anterior, se encontraron asociaciones de tamaño de familias de genes con el tamaño de la corteza relativo al tamaño del resto del cerebro. Para descartar que las asociaciones encontradas con el número de neuronas se explican por el carácter general del tamaño se realizó un PGLS que en esta ocasión incluía tanto número de neuronas en corteza

como el tamaño relativo de la corteza. Se obtuvo un total de 414 familias asociadas al número de neuronas de la corteza y 116 al tamaño relativo de la corteza. En asociaciones negativas se encontró 34 familias asociadas a número de neuronas en corteza y 437 a tamaño de la corteza relativo al tamaño total del cerebro.

Finalmente, para descartar que las asociaciones encontradas con el número de neurona en corteza se explican como un efecto sistémico del aumento en el número de neuronas en todo el cerebro se realizó un tercer PGLS de dos predictores, en esta ocasión incluyendo número de neuronas en corteza y número de neuronas en el cerebro completo. Se encontró un total de 398 familias asociadas con número de neuronas en la corteza y 36 en todo el cerebro. En las asociaciones negativas encontramos 27 familias asociadas al número de neuronas en corteza y 2 con el número de neuronas en cerebro completo.

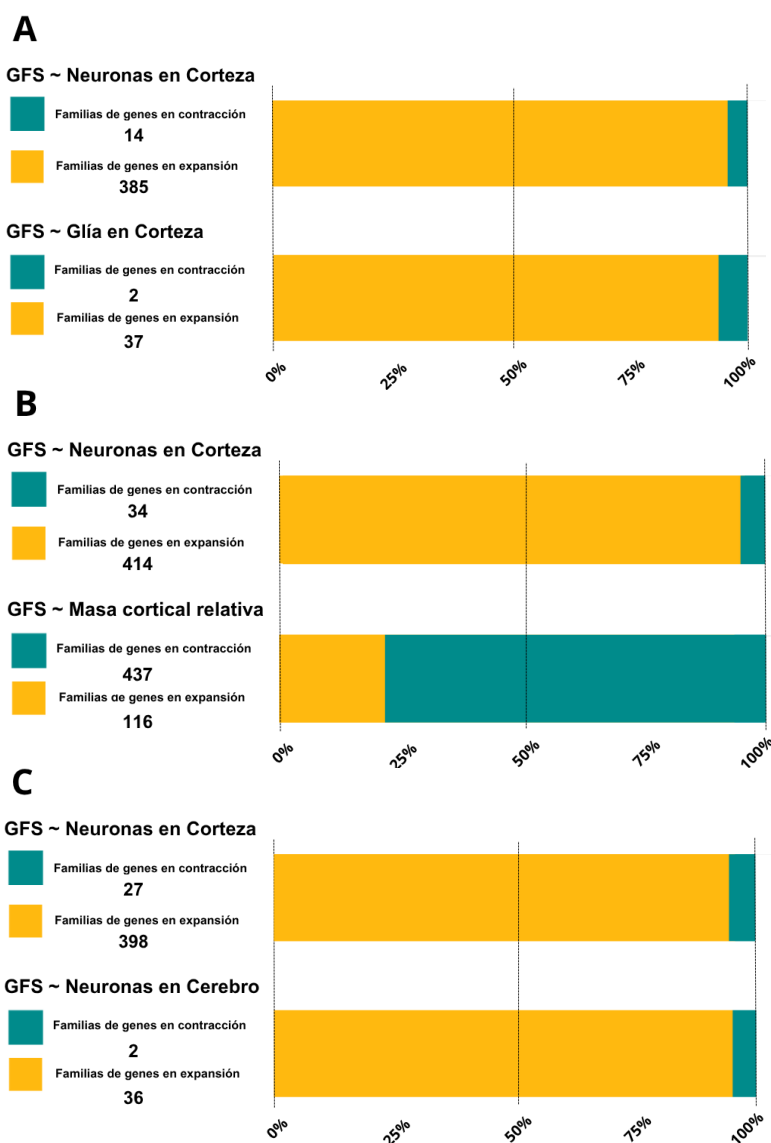


Figura 18. Sobrerrepresentación de tamaño de familias de genes (GFS), en expansión y contracción asociados a fenotipos de composición celular, a partir de resultados en análisis de PGLS doble de predictor único. A) GFS de neuronas y glía en corteza cerebral. **B)** GFS de neuronas en corteza cerebral y masa cortical relativa. **C)** GFS de neuronas en corteza y neuronas en cerebro completo. En color amarillo se muestra el GFS que exhibieron una expansión estadísticamente significativa ($r > 0$, $p < 0.05$). Color verde muestra el GFS que exhibieron una contracción estadísticamente significativa ($r < 0$, $p < 0.05$).

A partir de estos resultados, dado el número reducido de familias asociadas negativamente a los fenotipos evaluados y la dificultad de caracterizar números reducidos, se consideró únicamente las familias de genes con asociaciones significativamente positivas, para los análisis posteriores, ya que estas reflejan una expansión del tamaño de la familia génica.

Evaluando la especificidad de estas asociaciones, se realizaron diagramas de Venn, donde se comparó el GFS asociado a fenotipos de interés. Comparando el GFS asociado al número de neuronas y glía en la corteza cerebral, mostrando un 0% de solapamiento entre ambas (Figura 19-A), indicando que las asociaciones genómicas son específicas para el número de neuronas corticales y no reflejan un efecto general del total de células corticales.

De igual forma, al comparar el GFS de neuronas en corteza cerebral y neuronas en cerebro completo, mostró un 0% de solapamiento (Figura 19-C), descartando también un efecto sistémico generalizado. Y por último comparando el GFS de neuronas en corteza cerebral y la masa cortical relativa, reveló seis familias de genes compartidas entre ambos fenotipos (Figura 19-B).

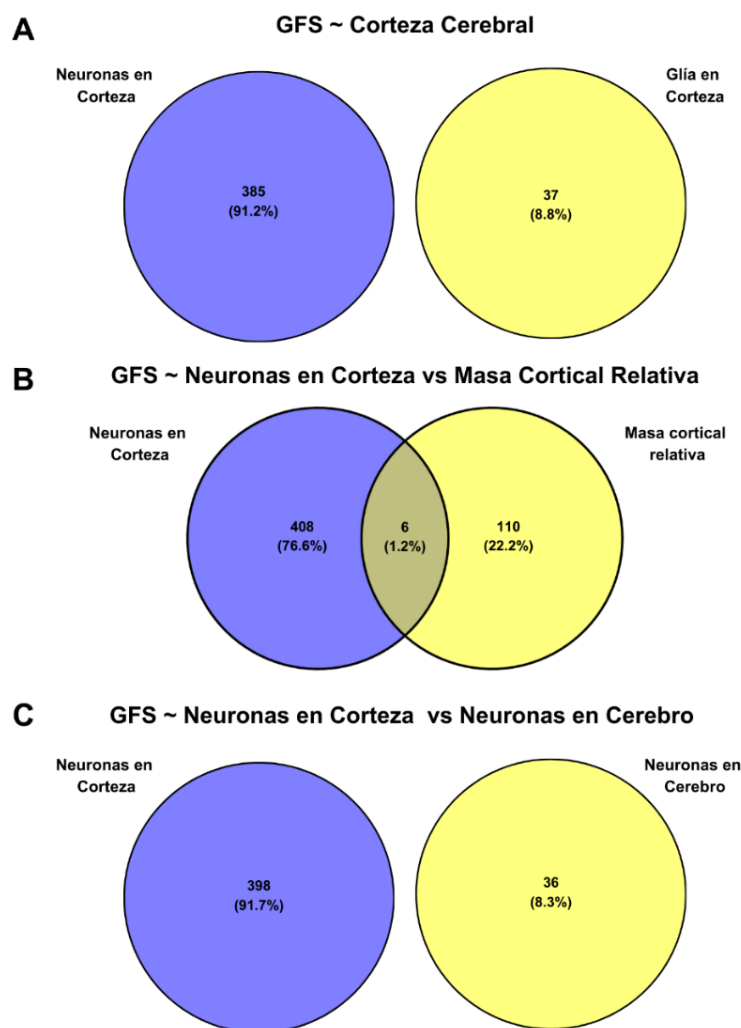


Figura 19. Especificidad de las asociaciones del tamaño de familia de genes (GFS) con fenotipos de composición celular, a partir de resultados en análisis de PGLS doble de predictor único. Se utilizaron los GFS que exhibieron expansión significativa positiva ($p < 0.05$). A) GFS de neuronas y glía en corteza cerebral. B) GFS de número de neuronas en corteza cerebral y masa cortical relativa. C) GFS de neuronas en corteza cerebral y neuronas en cerebro completo.

También se analizó la distribución del número de copias génicas por familia de genes en las especies estudiadas, tanto para el conjunto completo de familias génicas ($n = 2777$; Figura 2S), con un rango que oscila entre 1 y 400 copias. Y para el subconjunto de familias asociadas al número de neuronas en corteza cerebral ($n = 385$; Figura 3S), el rango de copias se redujo significativamente a un intervalo de 5 a 120. Este contraste sugiere que, mientras las expansiones masivas en el genoma general (2777 familias de genes), pueden responder a demandas biológicas diversas y más generales, el set específico (385 familias de genes), opera bajo un régimen de selección más restrictivo y especializado (Figura 4S).

8.1.1. Caracterización funcional de familias génicas mediante dominios proteicos en genes

Con el fin de describir la arquitectura proteica de las familias génicas asociadas al número de neuronas en corteza cerebral, se analizó la distribución de dominios *InterPro* presentes en las proteínas codificadas por dichas familias. Para cada especie se construyó un perfil relativo de dominios, normalizando la frecuencia de aparición de cada dominio por el total de dominios detectados.

El perfil relativo de dominios por especie indica que el aumento de neuronas no se debe a la aparición de nuevos tipos de dominios, sino quizás a cambios en la expresión de genes que ya existían o a mutaciones puntuales (sustituciones de aminoácidos). Dominios como C2, inmunoglobulinas y quinasas, sugieren que el aumento del número de neuronas corticales en mamíferos no es un evento simple, sino que está respaldado por una expansión coordinada de dominios que mejoran la adhesión celular (inmunoglobulinas), la transmisión de señales (quinasas) y el soporte estructural/energético (Figura 20).

El análisis de distribución filogenética mediante el heatmap revela una marcada convergencia evolutiva en la arquitectura de dominios proteicos. Especies con altos conteos neuronales pertenecientes a linajes divergentes muestran un perfil de dominios enriquecido de manera análoga (Figura 21).

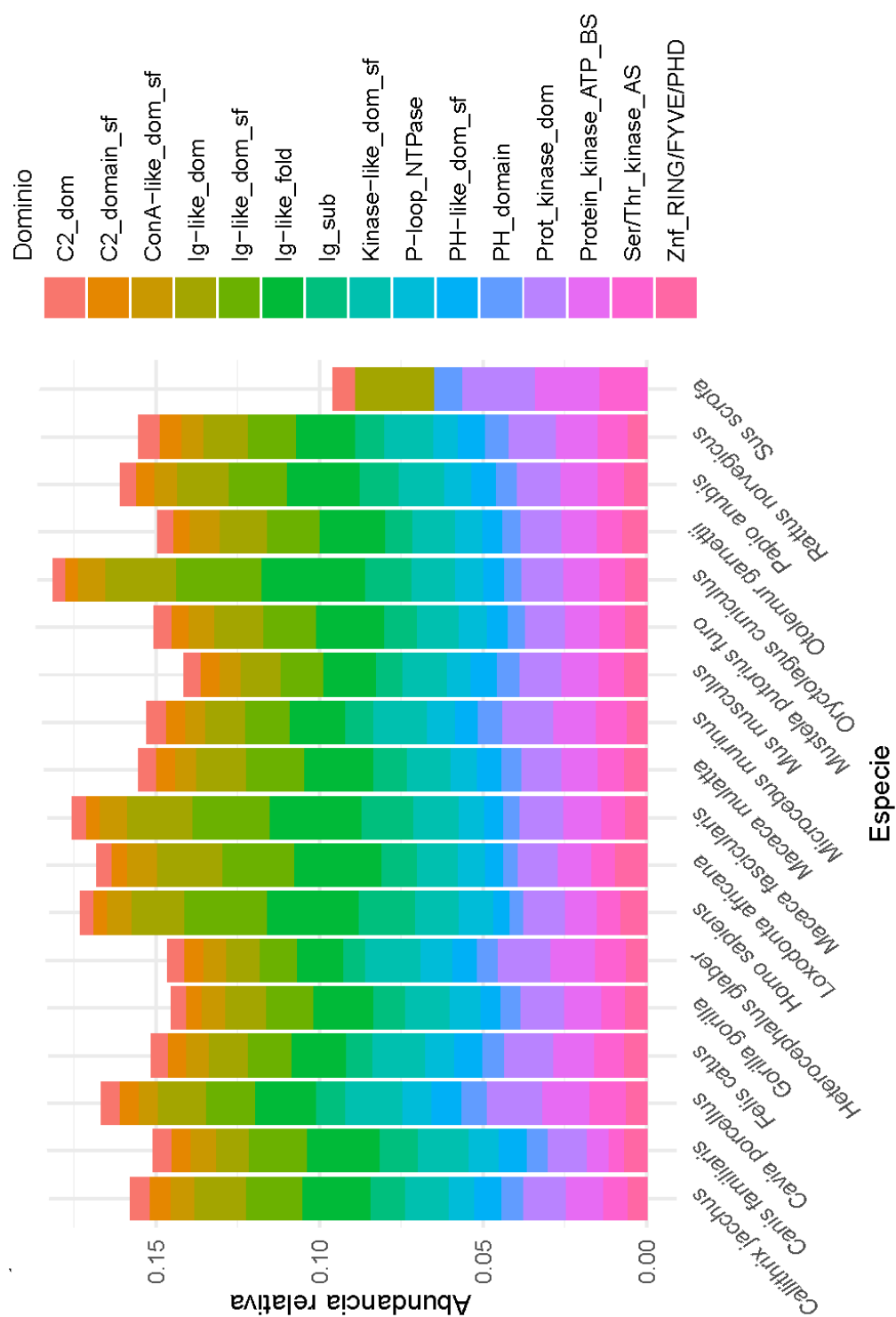


Figura 20. Perfil relativo de dominios presentes por especie, en las 385 familias de genes asociadas a neuronas en corteza cerebral. La grafica de barras muestra la proporción relativa de los dominios más abundantes por especie, normalizada por el total de dominios detectados, comparando la composición estructural del repertorio proteico entre especies independientemente del tamaño absoluto del genoma o del número total de anotaciones.

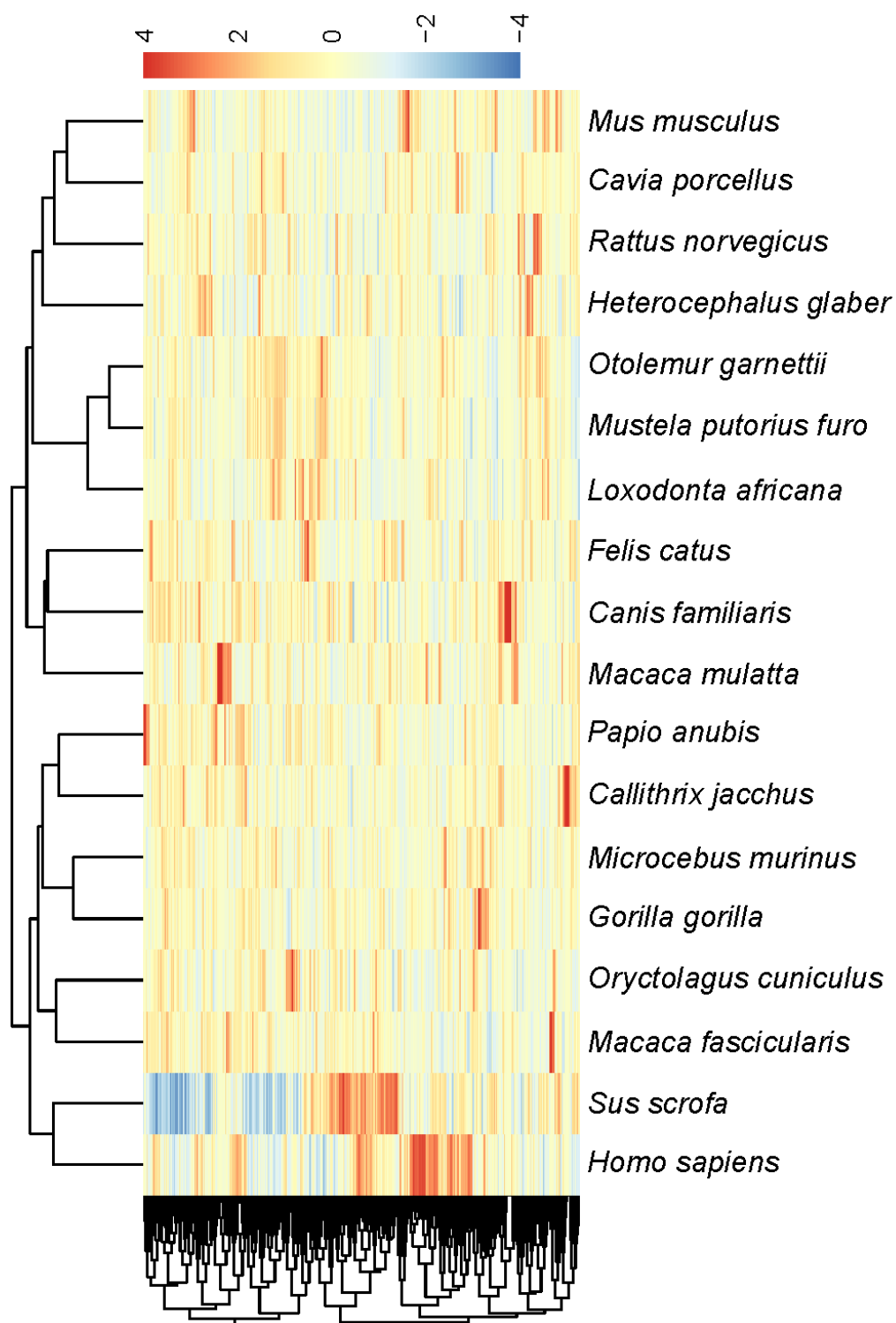


Figura 21. Distribución comparativa de dominios proteicos presentes por especie, en las 385 familias de genes asociadas a neuronas en corteza cerebral. El *heatmap* muestra la abundancia relativa normalizada de dominios *InterPro* por especie. Los colores indican desviaciones relativas respecto al promedio, permitiendo identificar patrones de conservación estructural y variación en la arquitectura proteica entre especies de mamíferos.

8.1.2. Enriquecimiento funcional en términos de *Gene Ontology* (GO) entre los GFS de neuronas corticales y de glía cortical: resultados del PGLS doble

Se realizó un análisis de enriquecimiento funcional utilizando términos de *Gene Ontology* (GO) a partir de las 385 familias de genes significativamente positivas ($p < 0.05$), asociadas con el número de neuronas en la corteza cerebral y las 37 familias de genes significativamente positivas ($p < 0.05$), asociadas a el número de células gliales en corteza cerebral. Mediante un mapa de calor (*HeatMap*), se revelaron procesos biológicos enriquecidos para estas familias de genes (Figura 22).

El GFS de neuronas corticales reveló una sobrerrepresentación de procesos biológicos asociados principalmente a funciones inmunológicas y de comunicación celular, destacando los siguientes términos GO:

- Respuesta de defensa
- Proceso de metabolismo de fosfolípidos
- Sistema inmune adaptativo
- Presentación de antígenos peptídicos exógenos a través del complejo mayor de histocompatibilidad (MHC) de clase II
- Regulación positiva de la respuesta inflamatoria

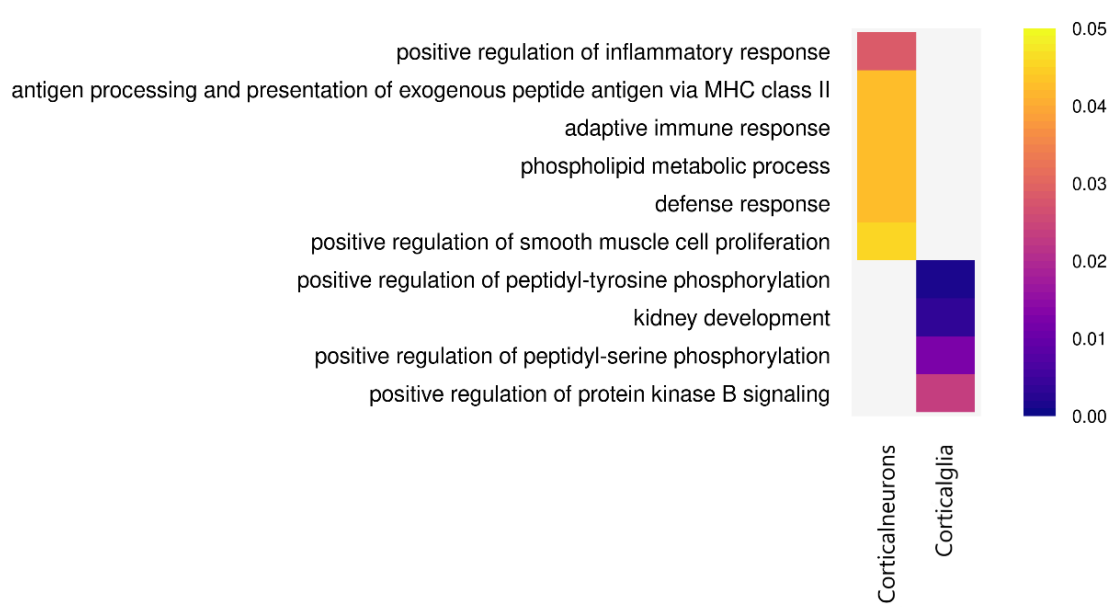


Figura 22. Sobrerrepresentación de anotaciones funcionales entre el GFS asociadas al número de neuronas en corteza cerebral y glía en corteza cerebral. Los términos GO significativamente enriquecidos en el mapa de calor se muestran en colores que van del amarillo ($p = 0.05$) al azul ($p = 0$). Con corrección de FDR de Benjamini-Hochberg.

Y en su contra parte, el GFS de células gliales en corteza cerebral reveló procesos como:

- Regulación positiva de la señalización de la proteína B quinasa
- Regulación positiva de la fosforilación de peptidil-serina
- Desarrollo del riñón

8.1.3. Enriquecimiento funcional en términos de *Gene Ontology* (GO) entre los GFS del número de neuronas corticales y la masa cortical relativa: resultados del PGLS doble

Para determinar si la asociación observada entre GFS y el número de neuronas en la corteza cerebral se podría explicar también por la masa cortical relativa, se realizó un segundo análisis PGLS doble de único predictor, incorporando ambas variables. Los resultados identificaron:

- 414 familias de genes en expansión, significativamente asociadas al número de neuronas en corteza cerebral ($p < 0.05$; Figura 18-B)
- 116 familias de genes en expansión, significativamente asociadas a la masa cortical relativa ($p < 0.05$; Figura 18-B)

El análisis de enriquecimiento funcional en términos *GO* (Figura 23), reveló que el GFS asociado a neuronas corticales están enriquecidas en procesos biológicos relacionados con:

- Regulación positiva de la proliferación celular en músculo liso
- Respuesta de defensa
- Respuesta inmune
- Metabolismo de fosfolípidos
- Respuesta inmune adaptativa.
- Regulación positiva de la respuesta inflamatoria

En contraste, el GFS asociadas a la masa cortical relativa no presentó categorías *GO* significativamente enriquecidas.

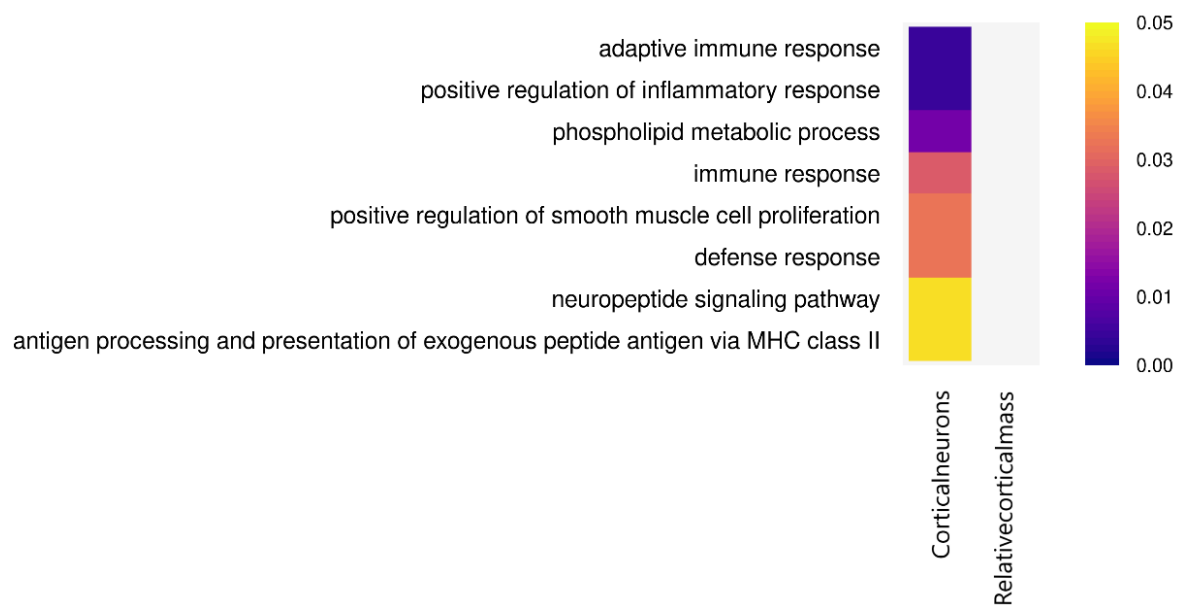


Figura 23. Sobrerrepresentación de anotaciones funcionales entre el GFS asociadas al número de neuronas en la corteza cerebral y la masa cortical relativa. Los términos *GO* significativamente enriquecidos en el mapa de calor se muestran en colores que van del amarillo ($p = 0.05$) al azul ($p = 0$). Con corrección de FDR de Benjamini-Hochberg.

8.1.4. Enriquecimiento funcional en términos de *Gene Ontology* (*GO*) entre los GFS, del número de neuronas corticales y el número de neuronas en cerebro completo: resultados del PGLS doble

Con el objetivo de evaluar si la asociación entre GFS y el número de neuronas en la corteza cerebral representa un patrón específico o un patrón generalizado en el cerebro completo, se implementó un tercer análisis de PGLS doble de predictor único. Utilizando los GFS del número de neuronas en corteza cerebral y el número de neuronas en cerebro completo. Este enfoque permite controlar la colinealidad entre ambos fenotipos y distinguir si las asociaciones encontradas se deben exclusivamente a la corteza o reflejan una tendencia general de expansión neuronal en todo el cerebro. Estos resultados mostraron:

- 398 familias de genes en expansión, significativamente asociadas al número de neuronas en corteza cerebral ($p < 0.05$; Figura 18-C).
- 36 familias de genes en expansión, significativamente asociadas al número de neuronas en cerebro completo ($p < 0.05$; Figura 18-C).

El análisis de enriquecimiento funcional en términos *GO*, reveló para el GFS asociadas a neuronas corticales (Figura 24), procesos biológicos similares al primer análisis en corteza cerebral destacando:

- Presentación de antígenos peptídicos exógenos a través del complejo mayor de histocompatibilidad (MHC) de clase II

- Respuesta de defensa
- Respuesta inmune
- Regulación positiva de la respuesta inflamatoria
- Sistema inmune adaptativo
- Proceso de metabolismo de fosfolípidos

Por otro lado, para el GFS asociados a neuronas en el cerebro completo reveló procesos biológicos como:

- Regulación positiva de la señalización de la proteína B quinasas
- Regulación positiva de la fosforilación de peptidil-tirosina
- Desarrollo del riñón

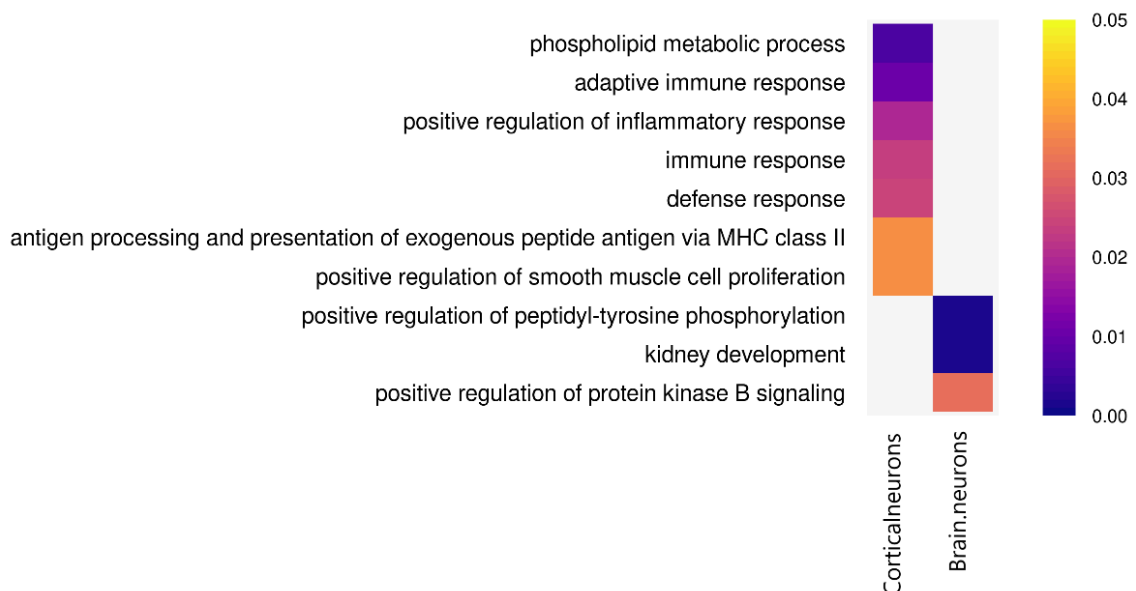


Figura 24. Sobrerepresentación de anotaciones funcionales entre el GFS asociadas al número de neuronas en corteza cerebral y el número de neuronas en cerebro completo. Los términos *GO* significativamente enriquecidos en el mapa de calor se muestran en colores que van del amarillo ($p = 0.05$) al azul ($p = 0$). Con corrección de FDR de Benjamini-Hochberg).

8.2. Caracterización de la expresión génica espacio-temporal de genes pertenecientes a familias de genes asociadas a corteza cerebral con respecto a otros tejidos del cuerpo humano

Una vez observada la especificidad de los resultados, para las 385 familias de genes con asociación al número de neuronas en corteza cerebral identificadas, que constituyen una posible firma genómica, constituyendo la base de los análisis posteriores.

Para lograr una caracterización funcional más precisa, se utilizaron genes individuales (3,500 genes en humano, como se describe en la metodología), pertenecientes a estas familias, en lugar de trabajar con las familias como unidades. Para poder capturar patrones de expresión específicos, los cuales podrían perderse si se analizara únicamente el nivel de familia génica. El objetivo de estos análisis, fue caracterizar detalladamente las 385 familias de genes asociadas a neuronas en corteza cerebral, evaluando si su expresión es específica en el cerebro en comparación con otros tejidos del cuerpo humano. Buscando también determinar si su regulación está restringida a etapas específicas del desarrollo.

El análisis se realizó considerando los dos grupos de donantes reportados en la base de datos *FANTOM5*: en fetos y adultos. Y responder la siguiente pregunta: ¿Los genes pertenecientes a estas familias se expresan preferentemente en el cerebro en comparación con otros tejidos del cuerpo humano?

8.2.1. Expresión génica en el cerebro, de genes asociados al número de neuronas en corteza cerebral, en comparación con otros tejidos del cuerpo humano, en etapas fetal y adulta

Contrario a lo esperado, los resultados mostraron que los genes pertenecientes a las 385 familias significativamente asociadas a neuronas corticales no presentan una expresión diferencial elevada en el cerebro frente a otros tejidos. Tanto en la etapa fetal como en la etapa adulta, la media observada de expresión génica (representada por la línea punteada roja; Figura 25 A y B), fue consistentemente inferior a la media esperada, bajo un escenario de distribución aleatoria (línea continua negra; Figura 25 A y B).

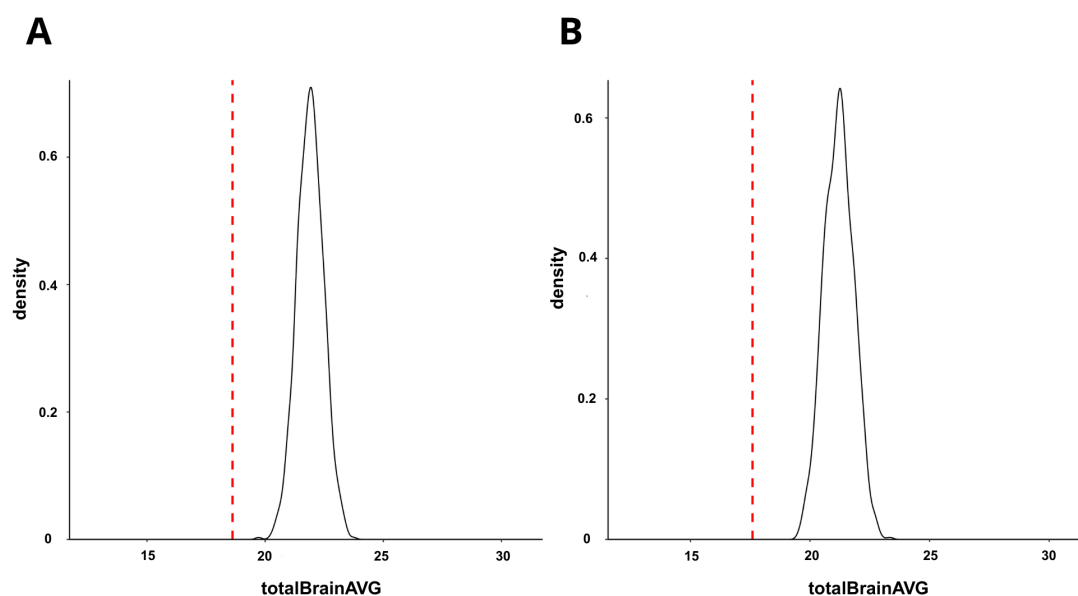


Figura 25. Expresión génica de genes asociados a neuronas en corteza cerebral, en comparación con otros tejidos en el cuerpo humano en etapas fetal y adulta. A) “Neuronas en corteza cerebral en etapa fetal”, B) “Neuronas en corteza cerebral en etapa adulta”. La línea discontinua roja muestra la media real de la expresión génica dada por el promedio del cerebro total, en comparación con otros tejidos del cuerpo humano, de genes asociados a neuronas corticales. La línea negra continua representa la distribución de medias para la clasificación de la expresión génica en el cerebro en comparación con otros tejidos del cuerpo, de muestras de genes seleccionadas aleatoriamente 1000 veces al azar. Eje “y” muestra densidad (*density*), eje “x” muestra expresión génica dada por el promedio del cerebro total (*totalBrainAVG*).

8.3. Caracterización de la expresión génica espacio-temporal de genes pertenecientes a familias de genes asociadas a la corteza cerebral con respecto a otras regiones cerebrales

Aquí se evaluó si los genes de estas 385 familias de genes asociadas a neuronas en la corteza cerebral presentan una expresión diferencial en dicha región con respecto a otras regiones del cerebro humano. Para determinar si la expresión génica está restringida o es predominante en la corteza cerebral, lo que apoyaría su papel específico en el desarrollo y funcionamiento de esta estructura.

Utilizando la base de datos “*RNA-seq Gencode v10 summarized to genes*” de *BrainSpan*, los datos se clasificaron en tres grupos etarios de acuerdo con las etapas reportadas por los donantes; fetal, infantil y adulta como se describe en la metodología. Planteando responder si: ¿Los genes pertenecientes a las familias asociadas al número de neuronas en la corteza cerebral se expresan con mayor intensidad en la corteza cerebral en comparación con otras regiones cerebrales?

8.3.1. Expresión génica en la corteza cerebral frente a otras regiones del cerebro, para genes asociados al número de neuronas en corteza cerebral, en etapas fetal, infantil y adulta

Los resultados revelaron patrones diferenciados según la etapa del desarrollo:

- Etapa fetal: se observó una expresión génica en promedio menor en la corteza cerebral en comparación con lo esperado (Figura 26-A). Este resultado es coherente con el hecho de que, durante esta etapa, el cerebro se encuentra en plena fase de neurogénesis, caracterizada por la proliferación y diferenciación de progenitores neuronales.
- Etapa infantil: se reveló una sobreexpresión significativa de los genes evaluados en la corteza cerebral (Figura 26-B). Esta etapa coincide con un periodo de intensa sinaptogénesis, establecimiento de redes neuronales, y máxima plasticidad cerebral.

- Etapa adulta: aquí la expresión génica se mantuvo elevada, aunque con un ligero descenso con respecto a la etapa infantil (Figura 26-C). Este patrón es consistente con la transición hacia una fase de mantenimiento funcional del cerebro adulto.

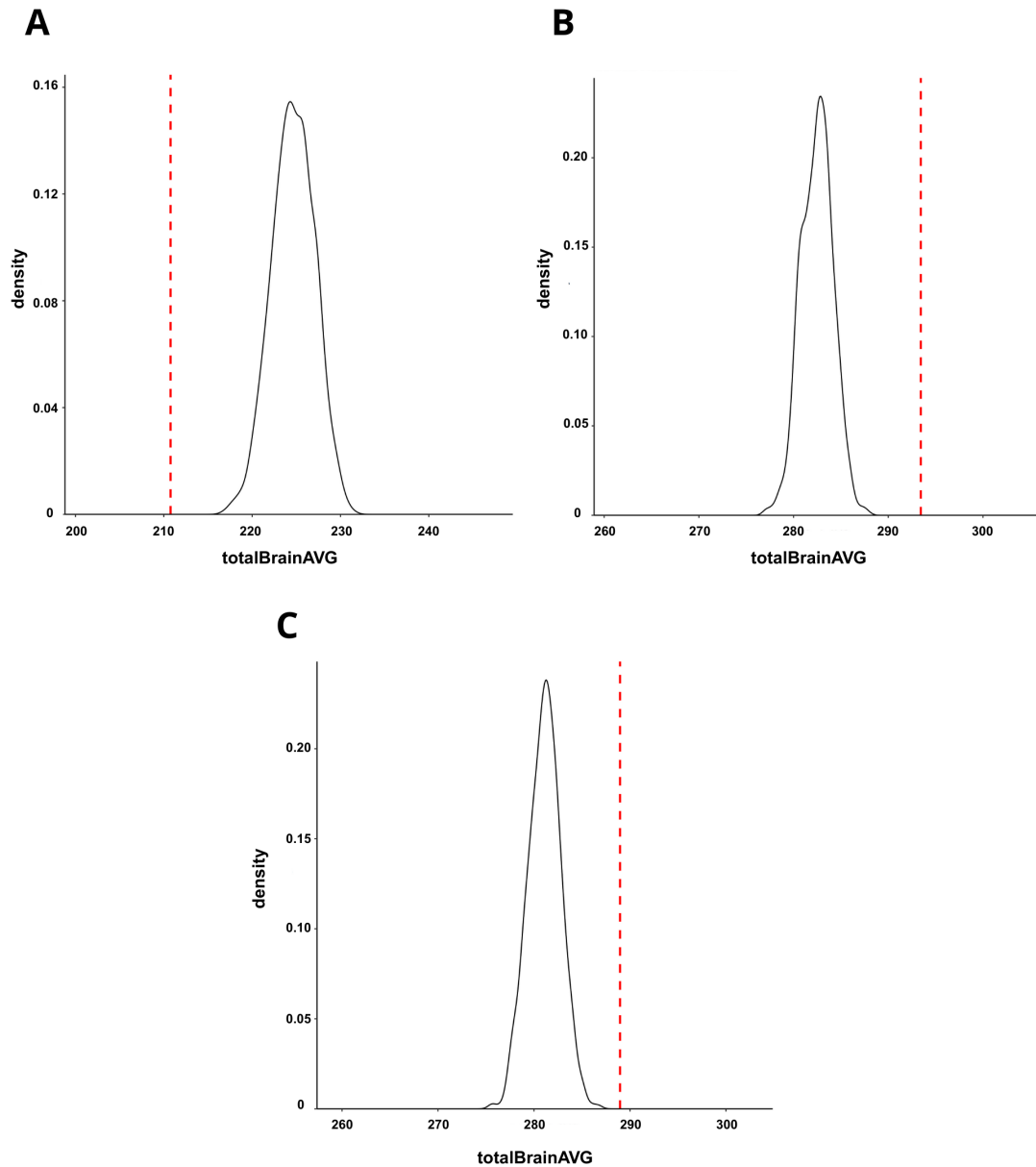


Figura 26. Expresión génica de genes asociados a neuronas en corteza cerebral, en comparación con otras regiones del cerebro humano en etapas: fetal, infantil y adulta. A) Neuronas en corteza cerebral en etapa fetal. B) Neuronas en corteza cerebral en etapa infantil. C) Neuronas en corteza cerebral en etapa adulta. La línea discontinua roja muestra media real de expresión génica dada por el promedio del cerebro total, en comparación con otras regiones del cerebro, de genes asociados a neuronas corticales. La línea negra continua representa la distribución de medias para la clasificación de la expresión génica en el cerebro en comparación con otras regiones del cerebro, de muestras de genes seleccionadas aleatoriamente 1000 veces al azar. Eje “y” muestra densidad (*density*), eje “x” muestra expresión génica dada por el promedio del cerebro total (*totalBrainAVG*).

8.4. Etapas clave del desarrollo cerebral con mayor expresión de genes asociados a neuronas corticales

Para comprender con mayor precisión los patrones temporales de expresión génica de los genes pertenecientes a las 385 familias de genes asociadas a neuronas en la corteza cerebral, se planteó la siguiente interrogante: ¿A qué edad estos genes pertenecientes a estas familias muestran su mayor nivel de expresión en la corteza cerebral?

Empleando nuevamente los datos de la base de “RNA-seq Gencode v10 summarized to genes” de *BrainSpan*, con una clasificación espacio-temporal específica descrita ya en la metodología, se desarrollaron dos enfoques analíticos complementarios:

1. Base de datos 1: Los valores de expresión génica fueron normalizados mediante *Z-score*, con el objetivo de facilitar la comparación entre genes con diferentes niveles de expresión relativa a lo largo del tiempo.
2. Base de datos 2: Se conservaron los valores de expresión sin normalizar, permitiendo evaluar la magnitud real de la expresión génica en cada etapa del desarrollo.

Esta estrategia dual permite analizar la expresión desde un enfoque comparativo (variabilidad relativa entre edades), también desde una perspectiva absoluta, revelando tendencias generales y picos de actividad transcripcional clave en el desarrollo de la corteza cerebral.

8.4.1. Dinámica espacio-temporal de la expresión génica de genes pertenecientes a familias asociadas significativamente a neuronas en corteza cerebral

Los resultados del análisis espacio-temporal de la expresión génica, revelaron lo siguiente:

- Análisis con base de datos 1 (normalizada con *Z-score*): Se observó una pendiente negativa con un coeficiente de determinación de $R^2 = 0.149$ (14.9%; Figura 27-A). Indicando una ligera disminución en la expresión génica a lo largo del tiempo. Al estar normalizados, los valores se centran alrededor de la media, permitiendo una mejor comparación relativa entre genes, pero sin conservar su magnitud real de expresión.
- Análisis con base de datos 2 (sin normalizar): También mostró una pendiente negativa, pero con un coeficiente de determinación más alto, de $R^2 = 0.271$ (27.1%) (Figura 27-B). Reflejando la magnitud real de la expresión génica, mostrando una

relación más robusta entre la edad y la disminución de la expresión génica en la corteza cerebral.

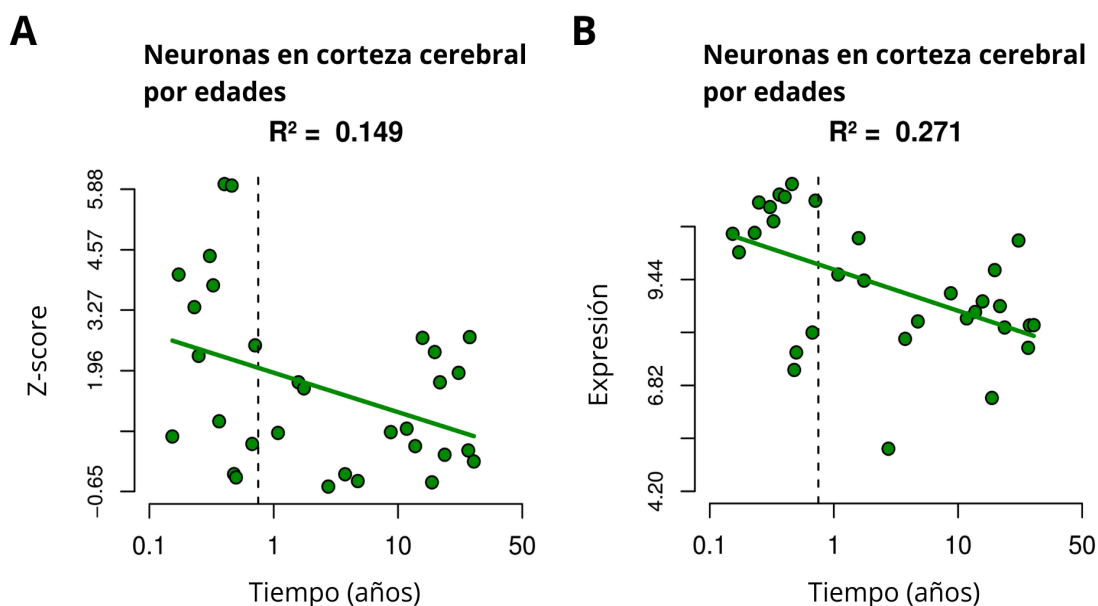


Figura 27. Patrones de temporalidad de expresión génica de las familias de genes significativos asociados a neuronas de corteza cerebral con respecto a etapas del desarrollo cerebral humano (edad). A) Niveles de expresión de genes asociados a neuronas en corteza cerebral (datos normalizados por *Z-score*). B) Niveles de expresión de genes asociados a neuronas en corteza cerebral (datos sin normalizar). Línea discontinua indica el punto de nacimiento. Línea continua verde indica la línea de tendencia. Puntos indican cada uno de los grupos por edad que van desde las 8 pcw hasta los 40 años.

8.5. Caracterización funcional y genómica de los genes pertenecientes a familias asociadas a corteza cerebral

Para caracterizar de manera más precisa las propiedades biológicas y genómicas de los genes pertenecientes a estas 385 familias de genes asociadas a neuronas en corteza. Se realizó un análisis funcional utilizando la herramienta bioinformática *ShinyGO 0.80*. Con la finalidad de explorar su distribución en el genoma humano, y detectar regiones con una concentración de enriquecimiento significativa de estos genes o su dispersión a lo largo del genoma. Para inferir posibles relaciones evolutivas o funcionales a partir de la proximidad cromosómica o la organización en clústeres.

Dado que la formación de familias de genes es el resultado de eventos de duplicación génica a partir de genes ancestrales, se espera que aquellas familias con funciones similares puedan encontrarse agrupadas físicamente en regiones específicas del genoma, lo cual podría reflejar procesos de co-expresión o regulación coordinada. Por otro lado, una distribución más dispersa sugeriría una especialización funcional o una pérdida de vínculos genómicos a

lo largo del tiempo evolutivo. Los análisis realizados permitieron integrar esta información en tres ejes: ubicación genómica, agrupamiento funcional y asociación con procesos biológicos.

8.5.1. Distribución genómica de los genes pertenecientes a familias de genes asociadas significativamente a neuronas en corteza cerebral

Los resultados del análisis identificaron 17 regiones genómicas enriquecidas que mostraron una distribución no aleatoria de genes a lo largo del genoma humano. Estas regiones se localizan en varios cromosomas, destacando los siguientes cromosomas; 5, 6, 7, 15, 17, 19, 21, 22 y X (Figura 28).

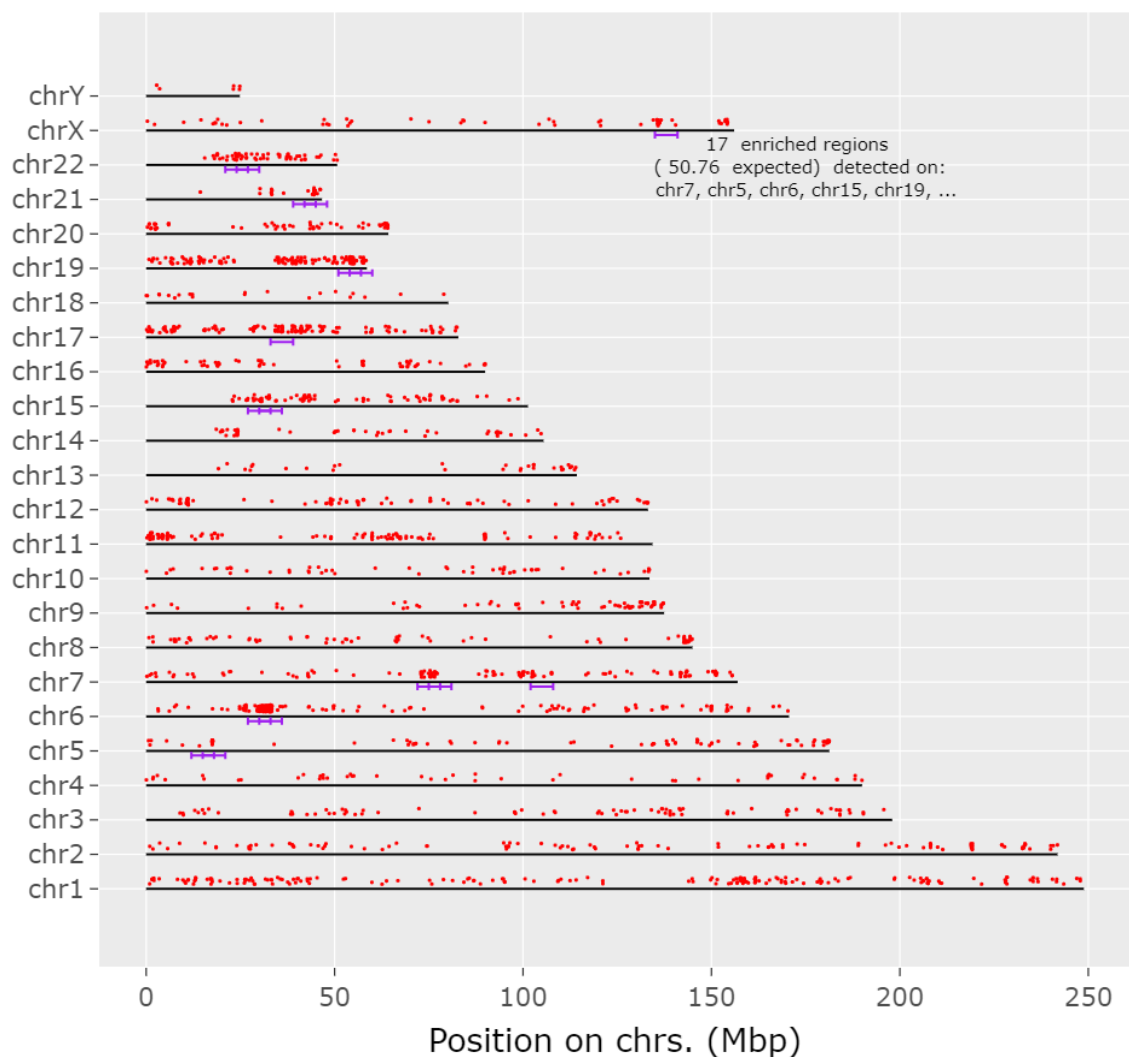


Figura 28. Distribución de genes pertenecientes a familias de genes significativos para neuronas en corteza cerebral. Las líneas violetas indican regiones donde los genes están más enriquecidos. Los puntos rojos representan los genes de interés distribuidos a lo largo del genoma. El eje “X” muestra cromosomas individuales, mientras que el eje “Y” indica la posición de los genes en mega pares de bases. Con corrección FDR de Benjamini-Hochberg $p \leq 0.05$. chr= Cromosoma, Mbp= Mega pares de bases.

La Tabla 2, describe de manera general, el contenido genético de estas 17 regiones específicas, destacando los genes más representativos que podrían estar implicados en procesos relevantes en el cerebro.

Tabla 2. Distribución de Genes en regiones enriquecidas en todo el genoma Humano

Chr	Región (Mbp)	Familia	Gen	Función principal
5	14.14 – 17.63	TRIPLE FUNCTIONAL DOMAIN PROTEIN	<i>TRIO</i>	Regulación del citoesqueleto, migración y morfogénesis neural
		TRANSCRIPTION INITIATION FACTOR TFIIID SUBUNIT 11	<i>TAF11L2, TAF11L3, TAF11L4, TAF11L5, TAF11L6, TAF11L7, TAF11L8, TAF11L9, TAF11L10, TAF11L11, TAF11L12,</i>	Transcripción basal, coactivador de ARN polimerasa II
6	25.96 – 36.94	E3 UBIQUITIN-PROTEIN LIGASE TRIM	<i>TRIM10, TRIM15, TRIM26, TRIM27, TRIM31, TRIM38, TRIM39, TRIM39-RPP21, TRIM40</i>	Ubiquitinación, respuesta inmune
		BUTYROPHILIN SUBFAMILY 3 MEMBER A1	<i>BTN3A1, BTN3A2, BTN3A3, BTN1A1, BTN2A1, BTN2A2</i>	Activación de células T $\gamma\delta$
		TUMOR NECROSIS FACTOR	<i>TNF</i>	Citocina proinflamatoria multifuncional.
		DEATH DOMAIN-ASSOCIATED PROTEIN 6	<i>DAXX</i>	Apoptosis y control del ciclo celular
		E3 UBIQUITIN-PROTEIN LIGASE PPP1R11	<i>PPP1R11, PPP1R18</i>	Inhibidor específico de la proteína fosfatasa-1 (PP1)
		RAS_RAP GTPASE-ACTIVATING PROTEIN SYNGAP	<i>SYNGAP1</i>	Plasticidad sináptica
		HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, A ALPHA CHAIN	<i>HLA-A, HLA-B, HLA-C, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5, HLA-E,</i>	Presentación antigénica
		MHC CLASS I POLYPEPTIDE-RELATED SEQUENCE A-RELATED	<i>MICA, MICB</i>	Inmunidad epitelial
		MEDIATOR OF DNA DAMAGE CHECKPOINT PROTEIN 1	<i>MDC1</i>	Citocina producida por los linfocitos.
		NF-KAPPA-B INHIBITOR-LIKE PROTEIN 1	<i>NFKBIL1</i>	Inhibidor de NFkB, su función aún no se ha determinado aun
		HSPA1	<i>HSPA1A, HSPA1B, HSPA1L</i>	Proteína de choque térmico
		COMPLEMENT C4-A-RELATED	<i>C4A, C4B</i>	Codifica la forma ácida del factor 4 del complemento, parte de la vía de activación clásica.

7	(Región 1) 73.01 – 77.19	LYMPHOCYTE ANTIGEN 6 COMPLEX LOCUS PROTEIN	<i>LY6G5B, LY6G5C, LY6G6C, LY6G6D, LY6G6F, LY6G6F-LY6G6D</i>	Activación inmune
		SUPPRESSOR APC DOMAIN- CONTAINING PROTEIN 1	<i>SAPCD1</i>	Proliferación celular
		ADVANCED GLYCOSYLATION END PRODUCT-SPECIFIC RECEPTOR	<i>AGER</i>	Inflamación crónica
		LYMPHOTOXIN- ALPHA/BETA	<i>LTA, LTB</i>	Citocina producida por los linfocitos.
		SPEEDY PROTEIN E10- RELATED	<i>SPDYE10, SPDYE11, SPDYE12, SPDYE13, SPDYE14, SPDYE15, SPDYE16, SPDYE17, SPDYE18, SPDYE5.</i>	Ciclo celular (fase G1/S)
	(Región 2) 102.66 – 107.58	E3 UBIQUITIN-PROTEIN LIGASE TRIM	<i>TRIM50, TRIM73</i>	Con actividad idéntica de unión a proteínas, participa en la respuesta inmunitaria innata.
		NUCLEAR ENVELOPE PORE MEMBRANE PROTEIN POM	<i>POM121C, POMZP3</i>	Participa en la exportación de ARN desde el núcleo y la importación de proteínas al núcleo
		C-C MOTIF CHEMOKINE 3- LIKE 1	<i>CCL24, CCL26</i>	Citocinas, proteínas secretadas que participan en procesos inmunorreguladores e inflamatorios.
		CLAUDIN-3/4	<i>CLDN3, CLDN4</i>	Uniones celulares, desarrollo prenatal, síndrome de Williams- Beuren
		DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB11-A-RELATED	<i>POLR2J2</i>	Transcripción génica
15	26.54 – 36.88	TRNA-DIHYDROURIDINE (20A_20B) SYNTHASE [NAD(P)+]-LIKE	<i>DUS4L, DUS4L-BCAP29</i>	Modificación de ARNt
		GAMMA-AMINOBUTYRIC ACID RECEPTOR SUBUNIT GAMMA	<i>GABRA5, GABRB3, GABRG3</i>	Neurotransmisión inhibitoria (receptor GABA A)
		CHRNA7-FAM7A FUSION PROTEIN	<i>CHRFAM7A, CHRNA7</i>	Receptores nicotínicos de acetilcolina (nAChR), canales iónicos regulados por ligando que median la transmisión rápida de señales en las sinapsis.
		RHO GTPASE-ACTIVATING PROTEIN 11A	<i>ARHGAP11A, ARHGAP11A-SCG5, ARHGAP11B</i>	Migración celular, ciclo celular
		HOMEBOX PROTEIN MEIS2	<i>MEIS2</i>	Morfogénesis y diferenciación

17	34.25 – 40.17	C-C MOTIF CHEMOKINE 3-LIKE 1	<i>CCL1, CCL11, CCL13, CCL14, CCL15, CCL16, CCL18, CCL2, CCL23, CCL3, CCL3L1, CCL4, CCL4L2, CCL5, CCL7,</i>	Quimiocinas, respuesta inflamatoria
		PROTEIN AF-17	<i>MLLT6</i>	Regulación epigenética de transcripción
		PHOSPHATIDYLINOSITOL 5-PHOSPHATE 4-KINASE TYPE-2 BETA	<i>PIP4K2B</i>	Transducción de señales
		ADP-RIBOSYLATION FACTOR-LIKE PROTEIN 5C-RELATED	<i>ARL5C</i>	Tráfico vesicular
		E3 UBIQUITIN-PROTEIN LIGASE RIFIFYLIN	<i>RFFL</i>	Permite la actividad de unión a enzimas; la actividad de unión a p53; y la actividad de la ubiquitina proteína ligasa.
		VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL SUBUNIT BETA-1	<i>CACNB1</i>	Canal de calcio tipo L, excitabilidad neuronal
		CYCLIN-DEPENDENT KINASE 12	<i>CDK12</i>	Ciclo celular, transcripción génica
		RECEPTOR TYROSINE-PROTEIN KINASE ERBB-2	<i>ERBB2</i>	Receptor tirosina quinasa, desarrollo de tejidos
		RAP GUANINE NUCLEOTIDE EXCHANGE FACTOR-LIKE 1	<i>RAPGEFL1</i>	Señalización acoplada a proteína G, desarrollo neural
19	50.02 – 58.34	SYNAPTOTAGMIN-3	<i>SYT3, SYT5</i>	Plasticidad neuronal
		SH3 AND MULTIPLE ANKYRIN REPEAT DOMAINS PROTEIN 1	<i>SHANK1</i>	Formación de sinapsis
		LEUKOCYTE IMMUNOGLOBULIN-LIKE RECEPTOR SUBFAMILY A	<i>LILRA1, LILRA2, LILRA4, LILRA5, LILRA6, LILRB1, LILRB2, LILRB3, LILRB4, LILRB5</i>	Activación inmune innata
		RET FINGER PROTEIN-LIKE 4A-RELATED	<i>RFPL4A, RFPL4AL1</i>	Participa en la respuesta inmunitaria innata y la regulación de la expresión génica.
		NATURAL CYTOTOXICITY TRIGGERING RECEPTOR 1	<i>NCRI</i>	Inmunidad antiviral
		KILLER CELL IMMUNOGLOBULIN-LIKE RECEPTOR	<i>KIR2DL1, KIR2DL3, KIR2DL4, KIR3DL1, KIR3DL2, KIR3DL3</i>	Tolerancia inmunológica
		Protein kinase C gamma type	<i>PRKCG</i>	Plasticidad sináptica
		PATERNALLY-EXPRESSED GENE 3 PROTEIN	<i>PEG3</i>	Apoptosis y desarrollo neuronal

21	36.46 – 46.23	KERATIN-ASSOCIATED PROTEIN	<i>KRTAP10-1, KRTAP10-10, KRTAP10-11, KRTAP10-12, KRTAP10-2, KRTAP10-3, KRTAP10-4, KRTAP10-5, KRTAP10-6, KRTAP10-7, KRTAP10-8, KRTAP10-9.</i>	Integridad epitelial
		HOMEBOX PROTEIN PKNOX1	<i>PKNOX1</i>	Angiogénesis y desarrollo
		GERMINAL-CENTER ASSOCIATED NUCLEAR PROTEIN	<i>MCM3AP</i>	Estabilidad genómica en SNC
		INTEGRIN BETA-2	<i>ITGB2</i>	Adhesión celular, inmunidad innata
22	21.54 – 31.16	IMMUNOGLOBULIN LAMBDA-LIKE POLYPEPTIDE	<i>IGLL1, IGLL5</i>	Producción de anticuerpos
		MACROPHAGE MIGRATION INHIBITORY FACTOR	<i>MIF</i>	Inmunorregulación e inflamación
		BETA-ADRENERGIC RECEPTOR KINASE 2	<i>GRK3</i>	Señalización adrenérgica
		Ret finger protein-like 1 RFPL1	<i>RFPL1</i>	Control del ciclo celular
		E3 UBIQUITIN-PROTEIN LIGASE ZNRF3	<i>ZNRF3</i>	Regulación de la vía Wnt
X	137.56 – 140.50	SAGE1-LIKE PROTEIN-RELATED	<i>SAGE1</i>	Antígeno tumoral
		CD40 LIGAND	<i>CD40LG</i>	Regula la función de los linfocitos B mediante la interacción con el CD40 en su superficie
		TRANSCRIPTION FACTOR SOX-3	<i>SOX3</i>	Desarrollo neuronal
		ADHESION G-PROTEIN COUPLED RECEPTOR G4	<i>ADGRG4</i>	GPCR en tejidos neuronales

Elaborada a partir de datos en (*Gene - NCBI*, 2025). Chr= Cromosoma; Mbp= Mega pares de bases.

8.6. Identificación de enriquecimiento funcional de redes de interacción proteína- proteína

Para profundizar en la caracterización funcional de estos genes pertenecientes a las 385 familias de genes asociadas a neuronas en la corteza cerebral, se llevó a cabo un análisis de interacción proteína-proteína mediante la herramienta bioinformática *Cytoscape StringApp*. Para visualizar interacciones moleculares entre las proteínas codificadas por estos genes, proporcionando un panorama general y funcional más amplio que complementa los análisis previos de ontología génica. Para examinar la red de interacciones físicas entre proteínas, e

identificar módulos funcionales, proteínas clave, y estructuras organizativas dentro del sistema biológico estudiado.

Los resultados se reflejaron primeramente en una red global de interacciones. Observándose una red con 2,955 proteínas, de las cuales 1,878 fueron nodos activos (esferas) y 1,130 interacciones representadas como aristas (líneas conectantes) la cual puede observarse en la figura suplementaria 5S, para una mejor observación de la red general se modificó a escala (1/4) y se clasificó en tres regiones en las figuras suplementarias S6; S7 y S8.

Aquí se observaron zonas densamente conectadas (módulos), que podrían estar relacionadas con funciones específicas del cerebro. También se detectaron áreas dispersas, con proteínas de baja conectividad, posiblemente implicadas en funciones especializadas dentro del contexto neuronal. Se identificaron proteínas *hub*, es decir, proteínas con un alto número de interacciones, lo cual sugiere un rol central en la integración y coordinación de funciones moleculares dentro del sistema neuronal. Estas proteínas son candidatas clave para futuros estudios funcionales, ya que su alteración podría tener efectos sistémicos sobre las redes de señalización y mantenimiento neuronal.

Los procesos biológicos más significativos y revelados por esta red de interacción proteína-proteína, así como los *hubs* proteicos y proteínas clave asociadas a cada uno de ellos se describen con mayor detalle a continuación.

8.6.1. Vía de señalización de los receptores de la superficie celular

Uno de los procesos biológicos más relevantes identificados en el análisis de interacción proteína-proteína fue la vía de señalización de receptores de superficie celular (GO:0007166 - *Cell surface receptor signaling pathway*). En este proceso participaron 379 genes pertenecientes a las familias génicas asociadas significativamente a neuronas en corteza cerebral (Figura 29). Esta vía abarca una secuencia de eventos moleculares que se inicia cuando un ligando se une a un receptor en la membrana celular, lo que activa una cascada de señalización intracelular culminando en una respuesta específica por parte de la célula (*Gene Ontology Consortium*, 2025b).

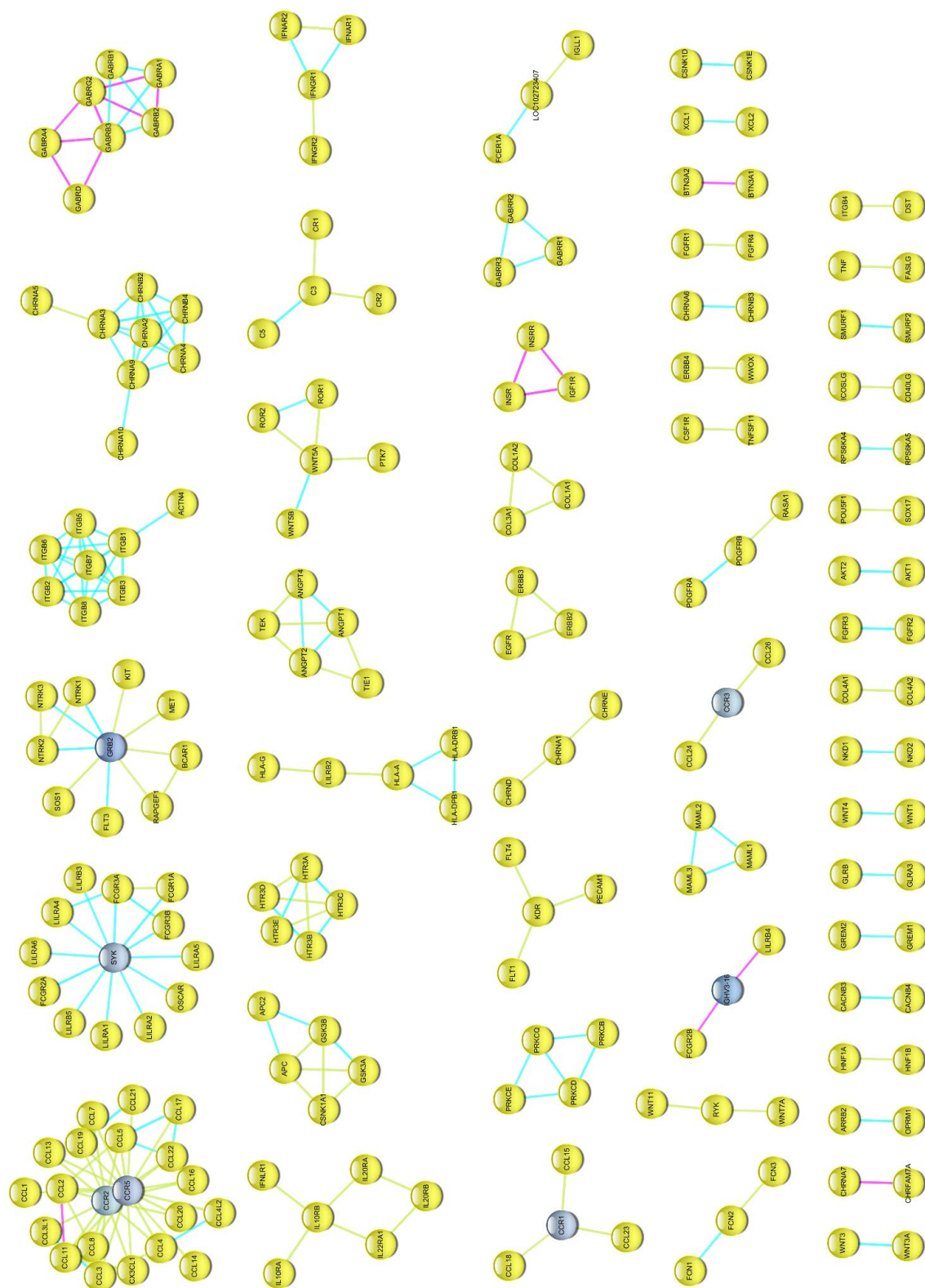


Figura 29. Redes de interacción proteína-proteína, GO:0007166 - *Cell surface receptor signaling pathway*. Relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

Dentro del conjunto total de proteínas funcionales identificadas en este análisis de interacción proteína-proteína, se identificó un subconjunto de proteínas, las cuales tienen

mención en la literatura con asociación a funciones cerebrales, dicho subconjunto se describe en la Tabla 3.

Tabla 3. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0007166 - *Cell surface receptor signaling pathway*” identificadas mediante el análisis de interacción proteína-proteína

Familia de genes	Gen	Proteína	Descripción	Referencia
Quimiocinas tipo CC	<i>CCL2, CCL3, CCL5</i> <i>CCL11, CCL24, CCL26</i>	C-C motif Chemokine ligand	Quimiocinas clave en migración, comunicación intercelular y neurotransmisión.	(Huber et al., 2018)
Receptores TRK (neurotrofinas)	<i>NTRK1, NTRK2, NTRK3</i>	High affinity nerve growth factor receptor	Participan en proliferación, diferenciación y supervivencia neuronal, y tienen un origen evolutivo anterior a los vertebrados.	(Hallböök et al., 2006)
Integrinas beta	<i>ITGB1, ITGB2</i>	Integrin beta-2	Clave en la adhesión celular, sinaptogénesis y organización del neuroepitelio.	(Rakic, 2009)
Receptores nicotínicos ACh (α)	<i>CHRNA7, CHRNBI, CHRNB2, CHRNB4, CHRNB5</i>	Neuronal acetylcholine receptor subunit alpha/beta	Codifican receptores nicotínicos involucrados en neurotransmisión, memoria y sistemas de recompensa.	(Pedersen et al., 2019)
Receptores GABA	<i>ÁlvarGABRA1, GABRA2, GABRA3, GABRA4, GABRA6</i>	Gamma-aminobutyric acid type alpha receptor	Receptores esenciales para la neurotransmisión inhibitoria.	(Shaw et al., 2018)
Receptores de interleucina 10	<i>IL10RA, IL10RB</i>	Interleukin-10 receptor subunit alpha/beta	Asociado a la formación sináptica y supervivencia neuronal en contextos inflamatorios, como en la enfermedad de Alzheimer	(Luo et al., 2022)
Supresor tumoral / señalización WNT	<i>APC</i>	Adenomatous polyposis coli protein	Más allá de su rol como supresor tumoral, regula el crecimiento axonal y la función sináptica.	(Stegmüller & Bonni, 2005)
Quinasas glicógeno sintasa	<i>GSK3A, GSK3B</i>	Glycogen synthase kinase-3 alpha/beta	En especial la isoforma GSK-3 β , es un regulador central en plasticidad sináptica, proliferación neuronal y neuroinflamación	(Lai et al., 2023)
Receptores de serotonina tipo 3	<i>HTR3A</i>	5-hydroxytryptamine receptor 3A	Receptores serotoninérgicos involucrados en procesamiento emocional y respuesta al estrés.	(Gatt et al., 2010)

Wnt ligandos	<i>WNT2, WNT3A, WNT5A, WNT7A, WNT7B</i>	Protein Wnt	Participan en diferenciación neuronal, plasticidad sináptica y morfogénesis cortical.	(Oliva et al., 2013)
Receptores de interferón	<i>IFNAR1, IFNAR2</i>	Interferon alpha/beta receptora	Regulan la respuesta inmune y la eliminación de neuronas innecesarias durante el desarrollo.	(Sarmiento et al., 2019)
Receptor tirosina quinasa relacionada a WNT	<i>RYK</i>	Tyrosine-protein kinase RYK	Receptor implicado en la organización de circuitos neuronales, al modular la morfogénesis dendrítica mediante la vía WNT	(Oliva et al., 2013)

8.6.2. Proceso del sistema inmune

En el proceso biológico GO:0002376 – *Immune System Process*, se identificaron 344 genes pertenecientes a las familias génicas significativamente asociadas a neuronas en corteza cerebral (Figura 30). Este término engloba cualquier mecanismo implicado en la función del sistema inmunológico, incluyendo tanto la respuesta inmune innata como la adaptativa (Gene Ontology Consortium, 2025a). Además de su papel en la defensa contra patógenos y la eliminación de células anómalas, hoy se reconoce que los procesos inmunológicos también influyen activamente en el desarrollo cerebral, la respuesta al estrés y diversas patologías neurológicas.

Varias de las proteínas asociadas a genes identificados también aparecieron previamente en la categoría de señalización por receptores de superficie celular (GO:0007166), como CCL, ITGB, IFNR y FCGR. Sin embargo, en este proceso emergen también familias con una función inmunitaria más directa, cuya relevancia se detalla a continuación. Se identificó un subconjunto de proteínas, con mención en la literatura con asociación a funciones cerebrales, dicho subconjunto se resume en la Tabla 4.

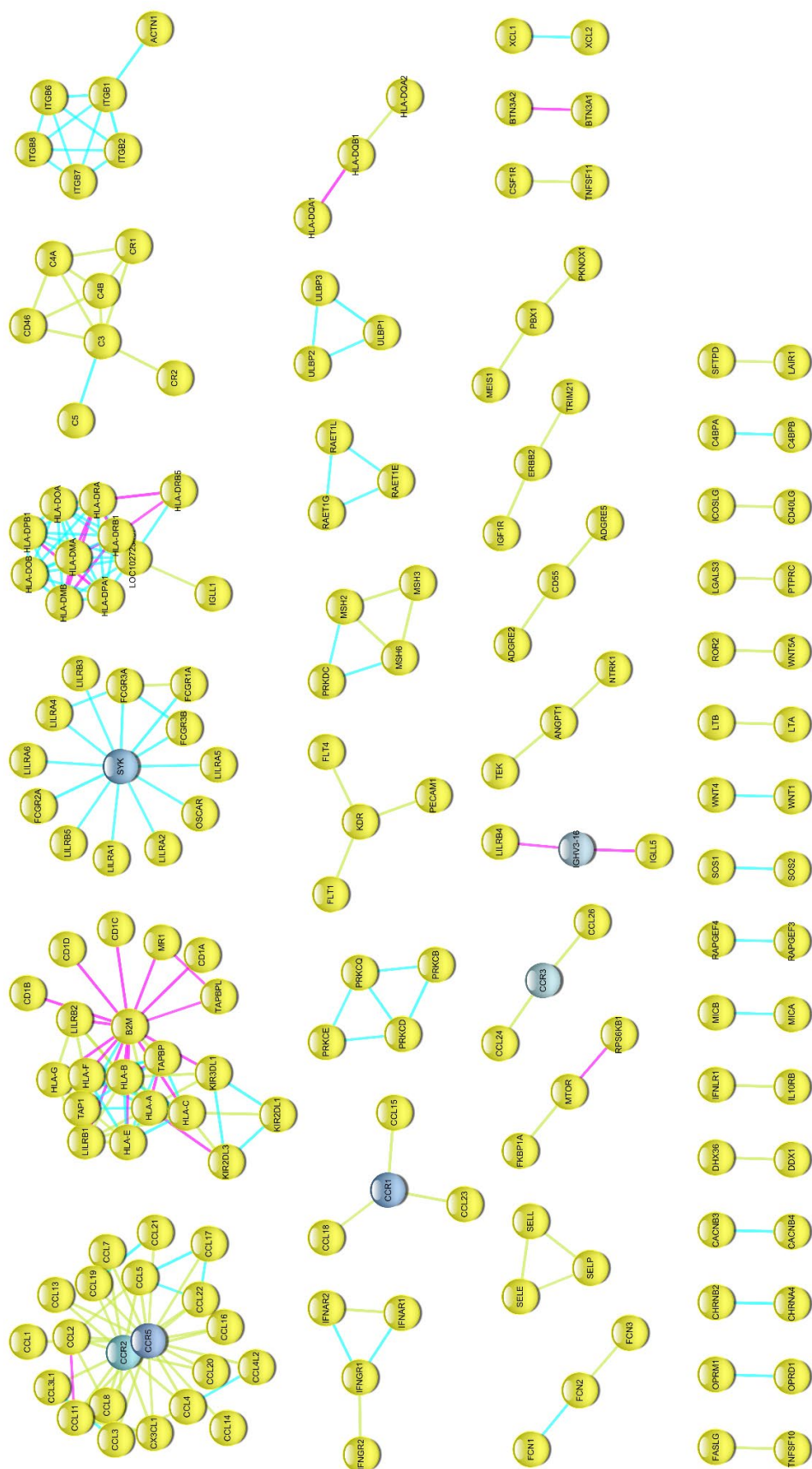


Figura 30. Redes de interacción proteína-proteína, GO:0002376 - *Immune system process*. Se observa la relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

Tabla 4. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0002376 - *Immune system process*” identificadas mediante análisis de interacción proteína-proteína

Familia de genes	Gen	Proteína	Descripción	Referencia
Complejo mayor de histocompatibilidad (MHC) clase I/II	<i>HLA A, HLA B, HLA C, HLA E, HLA F HLA G / HLA DRB1, HLA DQA1, HLA DQB1, HLA DRA</i>	HLA class I/II histocompatibility antigen	Proteínas fundamentales en la presentación antigénica y la plasticidad sináptica.	(Nataf, 2017)
Beta-2 microglobulina	<i>B2M</i>	Beta-2-microglobulin	Componente esencial del MHC clase I. Su acumulación se asocia con deterioro cognitivo y reducción de la neurogénesis en el hipocampo.	(Li et al., 2016)
Genes del sistema del complemento	<i>C3, C4A, C4B</i>	Complement proteins	Participan en la opsonización y activación de la respuesta inflamatoria.	(Janssen et al., 2005)
Receptores tipo inmunoglobulina leucocitarios (LILR)	<i>LILRA1, LILRA2, LILRA4, LILRA5, LILRA6 / LILRB1, LILRB2, LILRB3, LILRB4</i>	Leukocyte immunoglobulin-like receptor subfamily A/B	Regulan tanto la inmunidad como funciones neuronales.	(Brown, 2002)
Familia de proteínas quinasas C (PKC)	<i>PRKCB, PRKCD, PRKCQ, PRKCE / PKN1</i>	Protein kinase C / Serine/threonine-protein kinase	Cinasa dependiente de calcio/fosfolípidos, con funciones en proliferación celular, señalización y memoria. Su evolución implicó duplicaciones génicas, generando doce genes en mamíferos distribuidos en cinco familias con expresión específica por región cerebral.	(García Concejo & Larhammar, 2021)
Selectinas	<i>SELE, SELP, SELL</i>	Selectins	Moléculas de adhesión implicadas en el reclutamiento leucocitario.	(Angiari & Constantin, 2013)

8.6.3. Respuesta inmune adaptativa

Dentro del proceso biológico GO:0002250 – *Adaptive Immune Response*, se identificaron 76 genes pertenecientes a las familias génicas asociadas a neuronas en la corteza cerebral (Figura 31). Este proceso comprende cualquier mecanismo inmunológico que involucre el reconocimiento específico de antígenos por parte de linfocitos T y B, permitiendo una respuesta más precisa frente a patógenos y la generación de memoria inmunológica, que facilita respuestas más rápidas ante exposiciones subsecuentes (European Bioinformatics Institute, 2024a).

Al restringir el análisis a genes directamente implicados en la respuesta inmune adaptativa, se identificaron nuevamente proteínas previamente asociadas a procesos inmunitarios generales, tales como HLA, LILR, B2M, PRKC, FCGR y componentes del complemento. Adicionalmente, se destacaron otras familias de proteínas con funciones específicas en este tipo de respuesta. También se muestra un subconjunto de proteínas, las cuales cuentan con mención en la literatura asociadas a funciones cerebrales, dicho subconjunto se describe en la Tabla 5.

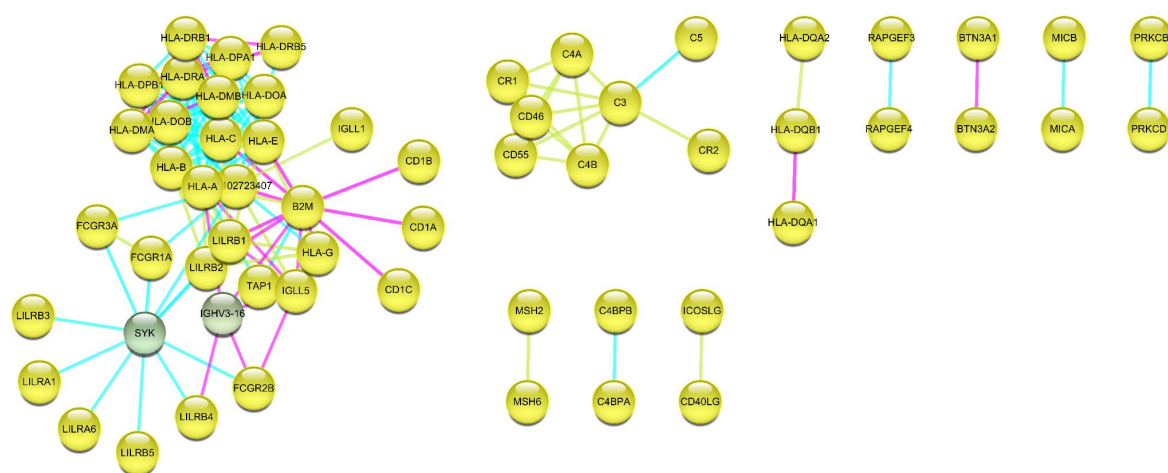


Figura 31. Redes de interacción proteína-proteína, GO:0002250 - *Adaptive immune response*. Se observa la relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

Tabla 5. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0002250 - *Adaptive immune response*” identificadas mediante análisis de interacción proteína-proteína

Familia de genes	Gen	Proteína	Descripción	Referencia
CD1 (cluster of differentiation 1)	<i>CD1A, CD1B, CD1C, CD1E</i>	T-cell surface glycoprotein	(CD1 protein): Familia de moléculas presentadoras de antígenos no clásicos, encargadas de presentar lípidos y glicolípidos a linfocitos T y células NK.	(Hughes, 1991)
MHC clase I relacionado (MHC-Ib)	<i>MICA, MICB</i>	MHC class I polypeptide-related sequence A/B	Antígenos inducidos por estrés, reconocidos por células T $\gamma\delta$ a través del receptor V δ 1.	(Elsner et al., 2001)
Complejo mayor de histocompatibilidad (MHC)	<i>MSH2, MSH6</i>	DNA mismatch repair protein	Aunque principalmente reconocidas por su rol en la reparación de errores de apareamiento de ADN (MMR), estas proteínas también participan en la regulación del sistema inmune.	(Edelbrock et al., 2013)

8.6.4. Respuesta inflamatoria

En el proceso biológico GO:0006954 – *Inflammatory Response*, se identificaron 82 genes pertenecientes a diversas familias asociadas a neuronas en la corteza cerebral (Figura 32). Esta categoría funcional describe la respuesta del sistema inmunológico frente a estímulos nocivos como infecciones, lesiones tisulares o irritantes, y se caracteriza por eventos moleculares como vasodilatación, aumento de la permeabilidad vascular, migración de leucocitos y liberación de citoquinas y mediadores proinflamatorios (European Bioinformatics Institute, 2024c).

Entre las familias de proteínas identificadas en este proceso se encuentran varias ya descritas previamente como CCL, proteínas del complemento, SEL, ITGB, IFNGR y FCGR, así como otras relevantes que se detallan a continuación. Posteriormente se muestra un subconjunto de proteínas, las cuales cuentan con mención en la literatura con asociación a funciones cerebrales, el subconjunto se describe en la Tabla 6.

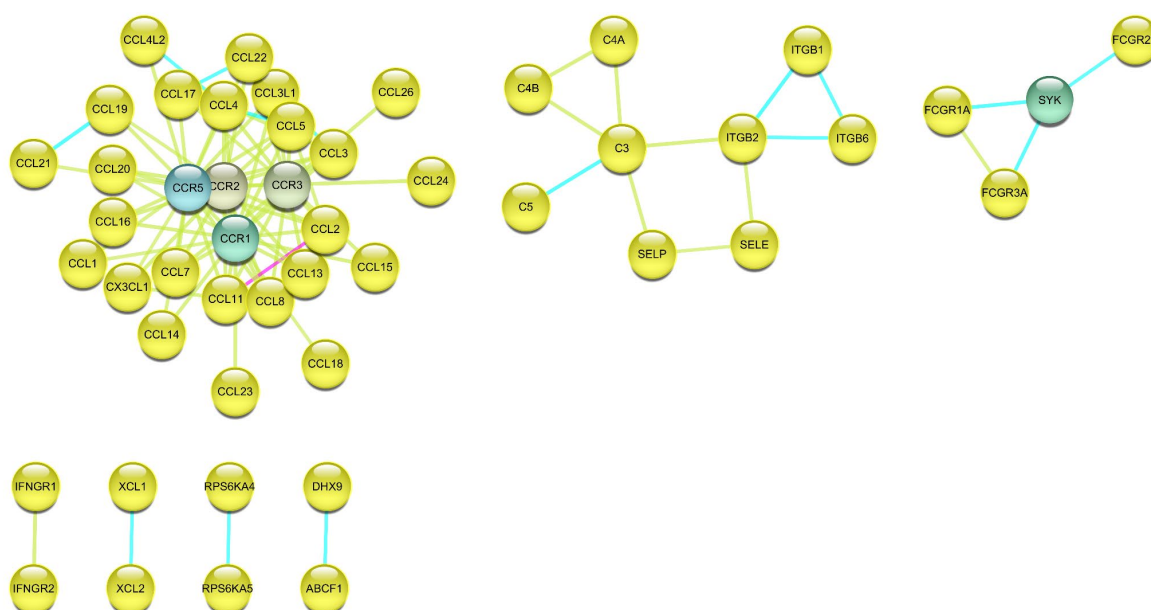


Figura 32. Redes de interacción proteína-proteína, GO:0006954 *Inflammatory response*. Se observa la relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

Tabla 6. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0006954 – *Inflammatory Response*” identificadas mediante análisis de interacción proteína-proteína

Familia de genes	Gen	Proteína	Descripción	Referencia
Ribosomal protein S6 kinase A (RSK)	<i>RPS6KA4, RPS6KA5</i>	Ribosomal protein S6 kinase alpha-1	Familia de cinasas implicadas en la señalización celular y en la regulación de la plasticidad neuronal.	(Decourt et al., 2023)
Helicasas tipo DExH-box (DEAH-box helicasas)	<i>DHX9</i>	ATP-dependent RNA helicase A	Helicasa de ARN dependiente de ATP, esencial en el metabolismo del ARN y la homeostasis neuronal.	(Chakraborty et al., 2018)

8.6.5. Respuesta a estímulos externos

En el proceso biológico GO:0009605 – *Response to external stimulus*, se identificaron 319 genes pertenecientes a las familias asociadas a neuronas en corteza cerebral (Figura 33). Este término hace referencia a cualquier cambio en el estado o actividad de una célula u organismo ante señales del entorno, como variaciones en el movimiento, secreción, producción de enzimas, expresión génica u otras funciones celulares (*European Bioinformatics Institute*, 2024d).

Se identificaron diversas familias de proteínas previamente discutidas, como CCL, sistema del complemento, HLA, LILRA, IFNGR, ITGB, B2M, IL10R, ULBP, PRKC, FCN, RAET, WNT y SEL, además de otras relevantes que se describen a continuación por su papel en la respuesta a estímulos externos.

Se muestra un subconjunto de proteínas, las cuales tienen mención en la literatura con asociación a funciones cerebrales, este subconjunto se describe en la Tabla 7.

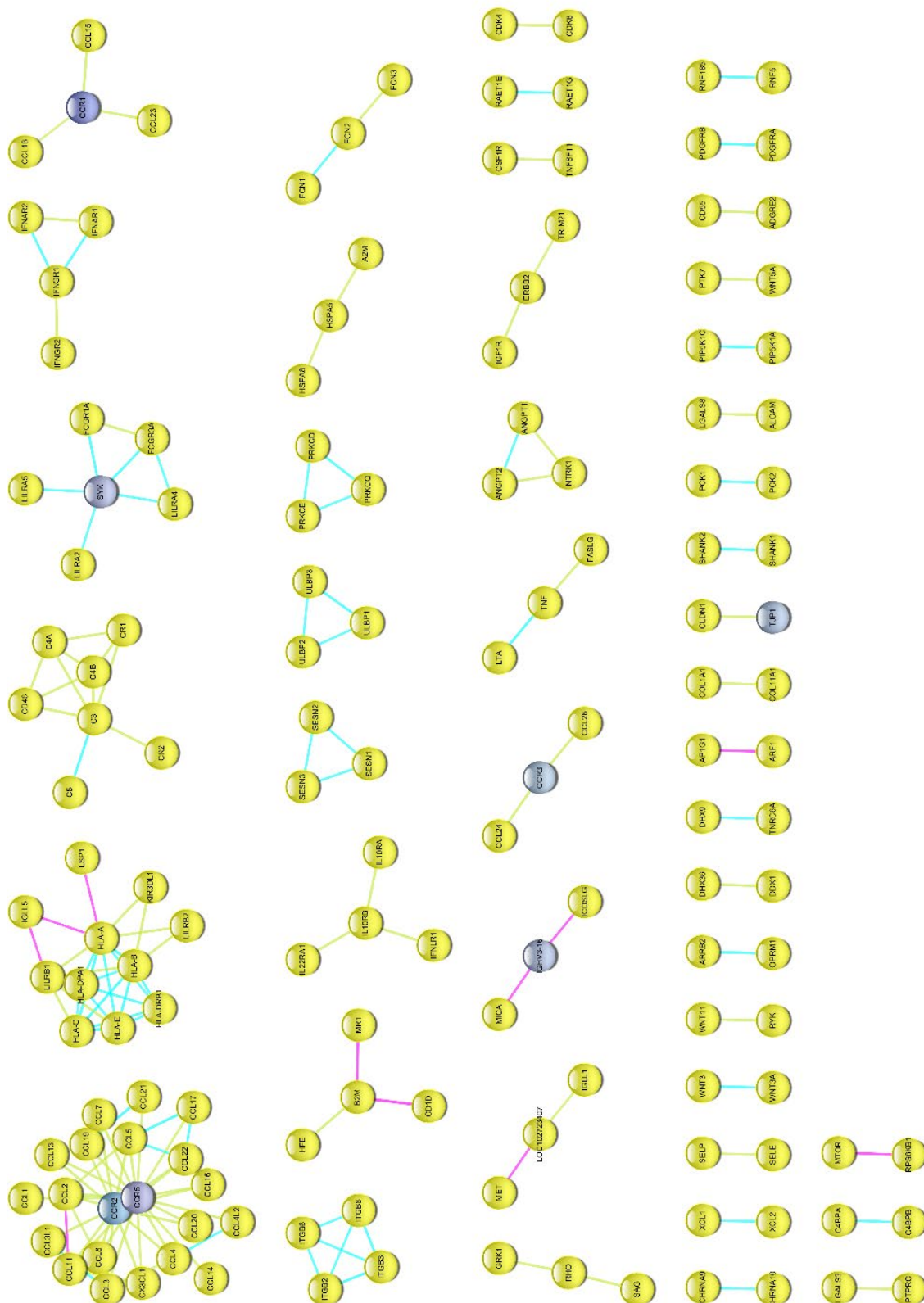


Figura 33. Redes de interacción proteína-proteína, GO:0009605 *Response to external stimulus*. Se observa la relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

Tabla 7. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0009605 *Response to external stimulus*” identificadas mediante análisis de interacción proteína-proteína

Familia de genes	Gen	Proteína	Descripción	Referencia
Sestrinas	<i>SESN1, SESN2, SESN3</i>	Sestrin	Grupo de proteínas que participan en la respuesta al estrés oxidativo y la hipoxia, regulando el metabolismo mitocondrial y promoviendo la homeostasis celular	(Kovaleva et al., 2020)
SH3 and multiple ankyrin repeat domains	<i>SHANK1, SHANK2</i>	SH3 and multiple ankyrin repeat domains protein	Proteínas de andamiaje postsináptico esenciales para la formación y plasticidad sináptica.	(Monteiro & Feng, 2017)
Factor de necrosis tumoral (TNF)	<i>TNF</i>	Tumor necrosis factor	Citocina clave en procesos inflamatorios y en el desarrollo del SNC, incluyendo expansión neuronal y diferenciación regional	(Twohig et al., 2011)
Receptores tirosina quinasa tipo ERBB	<i>ERBB2</i>	Receptor tyrosine-protein kinase erbB-2	Receptor tirosina-quinasa con expresión cerebral abundante, que participa en la diferenciación glial, la proliferación neuronal y la plasticidad sináptica.	(Ghashghaei et al., 2007)
Quinasas dependientes de ciclina (CDK)	<i>CDK4, CDK6</i>	Cyclin-dependent kinase	Enzimas reguladoras del ciclo celular en células progenitoras neurales.	(Grison et al., 2018)
Quinasas relacionadas a PI3K / mTOR	<i>MTOR</i>	Serine/threonine-protein kinase (Mammalian Target of Rapamycin)	Quinasa reguladora del metabolismo, neurogénesis y plasticidad sináptica.	(Kan et al., 2018)

8.6.6. Respuesta al estrés

El proceso biológico GO:0006950 – *Response to stress*, incluye a 401 genes pertenecientes a las familias de genes asociadas a neuronas en la corteza cerebral (Figura 34). Este término engloba cualquier cambio en el estado o actividad de una célula u organismo en respuesta a condiciones adversas o amenazantes, como factores ambientales, físicos, químicos o biológicos (*European Bioinformatics Institute*, 2024b). La función principal de este proceso es permitir la adaptación, protección y reparación del organismo frente al estrés. Las respuestas al estrés están altamente conservadas evolutivamente y abarcan complejas

interacciones entre el cerebro y otros sistemas fisiológicos, como el sistema inmune y cardiovascular.

Entre las proteínas identificadas, destacan numerosas familias previamente descritas, tales como CCL, HLA, KIR3DL, LILR, sistema del complemento, B2M, MSH, PRKD, NTRK, FGFR, ITG, ACTN, ERBB, IFNGR, PRKC, HSPA, CHRNA, FCN, TNF, SESN, ANGPT, ULBP, FCGR, DHX, APC, SEL, RAET, IRGB, RNF y MTOR, además de otras relevantes descritas a continuación por su papel en la respuesta al estrés.

Y por último se describe, un subconjunto de proteínas, que cuentan con mención en la literatura con asociación a funciones cerebrales, el subconjunto se describe en la Tabla 8.

Tabla 8. Subconjunto de proteínas relevantes en funciones cerebrales, pertenecientes al proceso biológico: “GO:0006950 *Response to stress*” identificadas mediante análisis de interacción proteína-proteína

Familia de genes	Gen	Nombre de Proteína	Descripción	Referencia
HSP70 (Heat Shock Protein 70)	<i>HSPA6, HSPA1A, HSPA1B, HSPA1L</i>	Heat shock 70 kDa proteins	Son chaperonas moleculares inducidas por estrés térmico u otros factores.	(Nguyen et al., 2017)
Receptores tirosina quinasa tipo III	<i>KIT</i>	Mast/stem cell growth factor receptor Kit	Crítico para el desarrollo neural. Su ausencia causa hipoplasia cerebral y aumento en la apoptosis durante el desarrollo embrionario	(Widłak, 2013)
Adhesion G-protein coupled receptors (ADGR)	<i>ADGRE2</i>	Adhesion G protein-coupled receptor E2	Receptor acoplado a proteína G que participa en la adhesión celular, desgranulación y quimiotaxis, especialmente en macrófagos, promoviendo la liberación de citocinas inflamatorias como TNF e IL-8	(Boyden et al., 2016)
Factores de transcripción pluripotentes	<i>SOX2</i>	Transcription factor SOX-2	Factor de transcripción fundamental para la pluripotencia y el desarrollo neuronal.	(Stevanovic et al., 2021)

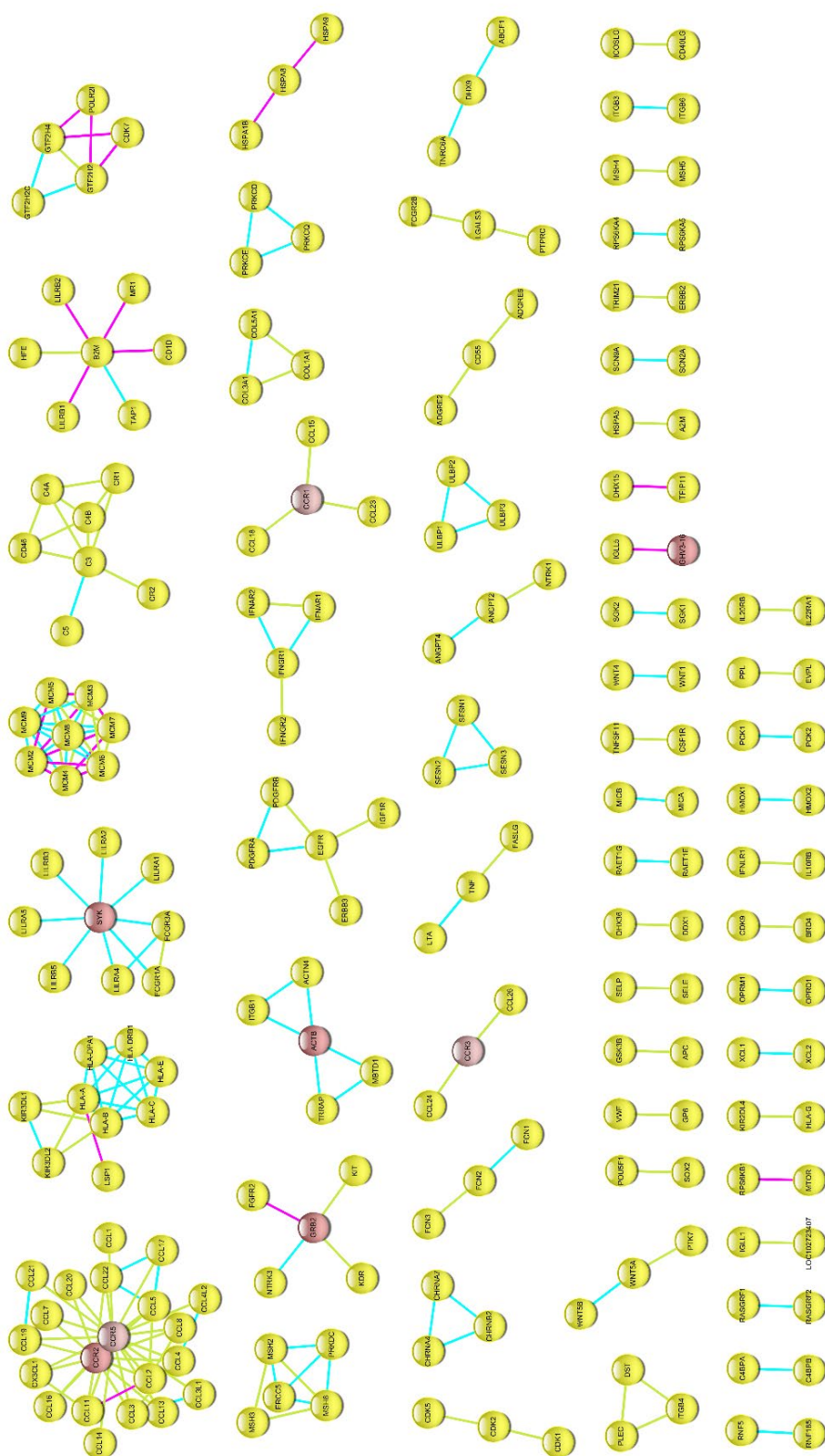


Figura 34. Redes de interacción proteína-proteína, GO:0006950 *Response to stress*. Se observa la relación entre proteínas y vías enriquecidas en términos GO. Nodo amarillo: proteínas de listado de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

IX. Discusión

En mamíferos, el cerebro presenta una mayor plasticidad neuronal, rasgo que se ha asociado con la capacidad de adaptarse a entornos variables. Sin embargo, esta plasticidad implica un alto costo metabólico y una mayor inversión parental durante el desarrollo, debido a la demanda energética asociada con el mantenimiento y funcionamiento del tejido neuronal. Por esta razón, las presiones evolutivas que impulsan el crecimiento cerebral y la expansión cortical han sido objeto de gran interés en estudios evolutivos. La evolución del tamaño y la composición celular de la corteza cerebral en mamíferos está estrechamente relacionada con cambios en la organización celular y, de manera clave, con alteraciones de la expansión y contracción en el tamaño de las familias de genes asociadas a la corteza cerebral. Este estudio refuerza la importancia de la genómica comparativa como herramienta para identificar bases moleculares que subyacen en la especialización neuronal y la plasticidad cortical, proporcionando una visión más profunda de los mecanismos que han moldeado la evolución del cerebro en mamíferos.

El presente trabajo identificó 385 familias de genes que representan una posible firma genómica clave en la evolución de la corteza cerebral de los mamíferos, ya que su expansión se relaciona con el aumento en el número de neuronas corticales, un rasgo distintivo en la evolución del cerebro de este clado. Esta asociación se mantuvo significativa incluso tras controlar el efecto del parentesco filogenético mediante un análisis PGLS, indicando que la relación entre estas familias génicas y el conteo neuronal no es atribuible únicamente a la historia evolutiva compartida entre especies.

El análisis del número de copias génicas por familia indica que la mayoría de las familias, incluidas aquellas asociadas al número de neuronas en corteza cerebral, sugiere que el incremento en la complejidad celular de la corteza no solo depende de cambios en la regulación genética, sino también de una expansión estructural del genoma (CNV). El rango observado, que oscila entre 5 y 120 copias, refleja diferentes estrategias de adaptación genómica: mientras que las familias con un bajo número de copias (5-20) podrían representar componentes estructurales altamente conservados, aquellas que presentan una expansión moderada (40-60) en la mayoría de los mamíferos sugieren un umbral funcional necesario para el mantenimiento de redes neuronales básicas. Sin embargo, la trayectoria evolutiva hacia un mayor número de neuronas parece demandar una 'escalabilidad genómica' no lineal. Las expansiones extremas observadas en linajes con alta densidad neuronal alcanzando hasta

120 copias, podrían interpretarse como una respuesta a la demanda metabólica y de conectividad que imponen los cerebros de gran tamaño.

El análisis de enriquecimiento funcional en términos de *GO*, reveló una sobrerrepresentación de procesos biológicos relacionados con la respuesta inmune, inflamatoria y el metabolismo celular. La presencia destacada de procesos inmunológicos es consistente con estudios previos que han propuesto un posible vínculo entre la expansión cortical y la modulación del sistema inmune, lo que sugiere una relación co-evolutiva entre estos dos sistemas (Filiano et al., 2015). Se ha sugerido que genes involucrados en procesos inmunológicos también pueden desempeñar funciones cruciales en la regulación de la neurogénesis y la plasticidad sináptica, actuando en la protección y mantenimiento neuronal, aspectos esenciales para la plasticidad sináptica, respuesta al estrés y la neuroprotección (Bilbo & Schwarz, 2012).

Además, se observó que estas familias no se solapan con aquellas asociadas a otros fenotipos cerebrales (como el número de neuronas totales en el cerebro o células gliales en corteza cerebral), reforzando su especificidad funcional para el desarrollo cortical. Cabe resaltar que la asociación de estas familias génicas con el número de neuronas en la corteza cerebral no podría atribuirse al tamaño relativo de la masa cortical. Ya que los resultados muestran una asociación más fuerte con el conteo neuronal en la corteza cerebral que con el volumen cortical relativo. Y, el análisis de enriquecimiento funcional reveló que las familias de genes asociadas al conteo neuronal en corteza cerebral vuelven a integrar procesos biológicos funcionales en el metabolismo celular y sistema inmune, mientras que en la masa cortical relativa no se identificaron procesos biológicos funcionales claros.

El análisis de dominios proteicos complementó el estudio funcional a nivel génico al aportar una visión estructural de las familias asociadas a el número de neuronas en corteza cerebral. En conjunto, la distribución comparativa y los perfiles relativos de dominios indican que las familias génicas asociadas al número de neuronas en corteza cerebral comparten una arquitectura de dominios 'proteicos conservada y convergente. Esta convergencia molecular indica que, independientemente del ancestro común, perfiles proteicos similares convergen para resolver las demandas de adhesión , señalización y metabolismo energético que exige una corteza densamente poblada.

El análisis de expresión en el cerebro con respecto a otros tejidos humanos, indicó una baja expresión de estas familias en cerebro, tanto en etapa fetal como en adulta, sugiriendo la existencia de mecanismos reguladores que restringen su activación a contextos

espacio-temporales específicos, o bien que su función está condicionada a etapas puntuales del desarrollo o a procesos altamente especializados (Florio et al., 2018; Florio & Huttner, 2014).

En el análisis de expresión en la corteza cerebral con respecto a otras regiones del cerebro, se observó un patrón de expresión diferencial en distintas etapas del desarrollo cerebral. Destacando una sobreexpresión de estos genes durante la etapa infantil, coincidiendo con períodos críticos de establecimiento y refinamiento de los circuitos neuronales (Stern et al., 2012).

La baja expresión de estos genes en la etapa fetal, podría deberse a que, en esta etapa del desarrollo, el cerebro está predominantemente orientado hacia la proliferación celular y la migración neuronal (Florio et al., 2018). Y dado que estos genes se asocian en mayor parte a funciones biológicas relacionadas con el sistema inmune, es posible que su papel se vuelva más relevante en etapas posteriores, cuando el cerebro madura y comienza a interactuar con estímulos externos. Por ejemplo, se ha sugerido que, durante la gestación el sistema inmune fetal se encuentra metabólicamente restringido para evitar respuestas inmunes energéticamente costosas y potencialmente perjudiciales en esta etapa temprana (Kan et al., 2018). O bien, estos genes podrían desempeñar un papel diferente a los procesos biológicos que ocurren durante el desarrollo fetal.

En la etapa adulta, se observó una expresión más cercana a la media esperada. Esto sugiere que los procesos biológicos regulados por estos genes siguen siendo esenciales para el mantenimiento neuronal y la respuesta a estímulos en el cerebro maduro. Sin embargo, su regulación parece ser más equilibrada en comparación con la infancia, sugiriendo un ajuste en la plasticidad sináptica y la estabilidad neuronal. Además, es importante considerar que en esta etapa comienza un equilibrio entre plasticidad y estabilidad neuronal permitiendo el mantenimiento funcional del cerebro, favoreciendo la longevidad y preservando las funciones cognitivas (Oliva et al., 2013).

El análisis de expresión génica espacio-temporal (región cerebral - grupo etario), reveló resultados que indican que la expresión de estas familias de genes disminuye progresivamente con la edad en la corteza cerebral. El declive gradual en la expresión de estas familias de genes con la edad podría reflejar la transición de un cerebro altamente plástico (capacidad de cambiar y adaptarse en respuesta a su aprendizaje) y adaptable en las primeras etapas de la vida, a uno más especializado y estable en la adultez, con una menor necesidad de reorganización sináptica (Stern et al., 2012).

El análisis de la distribución cromosómica reveló un patrón no aleatorio, con una acumulación significativa en los cromosomas 5, 6, 7, 15, 17, 19, 21, 22 y X. Sugiriendo que estos cromosomas y ciertos *loci* podrían desempeñar un papel central en los procesos biológicos relacionados con función cortical. La concentración de genes en ciertas regiones podría reflejar eventos evolutivos que favorecieron la co-localización de familias génicas con funciones complementarias, promoviendo así su regulación coordinada y contribuyendo a la expansión cortical. Entre los genes identificados, destacan aquellos vinculados tanto a la plasticidad neuronal como a la respuesta inmunológica. Por ejemplo, en el cromosoma 6 se localizan *HLA*, *BTN3* y *TRIM*, genes clave en la regulación inmune y la presentación antigénica, refuerza la idea de que el sistema inmunológico ha influido directamente en el desarrollo cortical y en funciones cognitivas (Bilbo & Schwarz, 2012). Asimismo, genes como *TRIO* (cromosoma 5) y *SYNGAP* (cromosoma 6) están estrechamente relacionados con la organización sináptica y la regulación de la plasticidad neuronal (Ba et al., 2016; Gamache et al., 2020). En el cromosoma X, genes como *SOX3* y *CD40LG* ilustran nuevamente esta convergencia entre procesos neuronales e inmunológicos, al estar implicados en la diferenciación neuronal y en la regulación de la respuesta inmune (Sarmiento et al., 2019; Stevanovic et al., 2021).

Finalmente, el análisis de redes de interacción proteína-proteína reveló módulos funcionales y *hubs* proteicos relevantes para este estudio. Este mapeo de interacciones ofrece una visión global de las proteínas involucradas y permite identificar módulos funcionales que sugieren la participación coordinada de grupos de proteínas en funciones biológicas clave para la organización y funcionamiento cortical.

- GO:0007166 - Vía de señalización de los receptores de superficie celular. Se ha documentado que esta vía desempeña un papel crucial en la capacidad del cerebro para recibir, procesar y responder a estímulos ambientales (Widłak, 2013). La familia de proteínas WNT se encuentra profundamente implicada en procesos como la neurogénesis, la diferenciación celular y la formación de redes neuronales (Oliva et al., 2013).
- GO:0002376 - Proceso del sistema inmune. Indicó una convergencia funcional entre inmunidad y neurobiología resaltando una posible co-evolución de ambos sistemas. Los procesos inmunológicos están estrechamente vinculados con el cerebro, ya que el sistema inmune no solo cumple funciones de defensa frente a patógenos, sino que también influye en el desarrollo neuronal, la plasticidad sináptica y la regulación de

la actividad neuronal (Bilbo & Schwarz, 2012). Este vínculo indica que el sistema inmune y el comportamiento también han co-evolucionado para optimizar la supervivencia frente a amenazas ambientales (Nataf, 2017).

- GO:0002250 - Respuesta inmune adaptativa. La respuesta inmune adaptativa, tradicionalmente reconocida por su papel en la protección frente a patógenos específicos, también ha sido implicada en la regulación de la inflamación y en la promoción de la homeostasis neuronal (Nataf, 2017).
- GO:0006954 - Respuesta inflamatoria. Más allá de su función clásica en la defensa frente a agresiones externas, la inflamación actúa como un mecanismo central en el desarrollo cerebral y la plasticidad neuronal. Su efecto puede ser beneficioso o perjudicial dependiendo del entorno fisiológico o patológico en el que se expresen, regulando así procesos clave para el desarrollo cerebral y el comportamiento (Borsini et al., 2015).
- GO:0009605 - Respuesta a estímulos externos. Destaca la importancia de la plasticidad cerebral y la capacidad adaptativa del sistema nervioso para responder a señales del entorno. Los genes implicados están estrechamente relacionados con la modulación de la actividad sináptica, subrayando su papel en la adaptación neuronal frente a estímulos ambientales cambiantes. La evolución del cerebro en mamíferos ha estado profundamente influenciada por la necesidad de percibir, procesar e integrar estímulos externos de manera eficiente (Kaas, 1989).
- GO:0006950 - Respuesta al estrés. Muestra una amplia gama de mecanismos celulares que permiten a las neuronas corticales detectar, responder y adaptarse a condiciones adversas, incluyendo variaciones ambientales, estrés oxidativo y señales proinflamatorias. Diversas investigaciones han documentado que estructuras cerebrales clave como el hipocampo y la corteza prefrontal sufren remodelaciones estructurales bajo condiciones de estrés, lo que repercute en la fisiología cerebral y en el comportamiento (McEwen, 2013). Las proteínas de choque térmico (HSPA), que actúan como chaperonas moleculares protegiendo a las células del daño inducido por estrés térmico y oxidativo, son esenciales para la integridad neuronal y su actividad resulta crítica en la prevención de la agregación proteica en enfermedades neurodegenerativas como el Alzheimer y el Parkinson (Nataf, 2017).

Entre las principales implicaciones de estos hallazgos destacan:

- Integración del sistema inmunológico y la función cerebral. La presencia de proteínas inmunitarias refuerza la hipótesis de una co-evolución entre el sistema inmune y la plasticidad neuronal, contribuyendo tanto al desarrollo como al mantenimiento funcional de la corteza cerebral.
- Modulación de la comunicación célula-célula. Los resultados evidencian la participación de proteínas implicadas en la señalización sináptica y en la plasticidad neuronal, procesos fundamentales para la consolidación de circuitos neuronales complejos.
- Regulación de la plasticidad sináptica. Los procesos vinculados a la respuesta al estrés permiten la reorganización de las conexiones sinápticas, facilitando la adaptación cerebral a entornos cambiantes.
- Protección celular y neurogénesis. Las vías asociadas a la respuesta al estrés también actúan como mecanismos de defensa, protegiendo las células corticales durante periodos críticos de desarrollo y reorganización sináptica.

Es importante subrayar que estos últimos resultados, no solo corroboran los procesos biológicos identificados previamente mediante el análisis de enriquecimiento funcional de términos *GO* mostrados en los *HeatMaps*, sino que también revelan nuevas funciones biológicas que no habían emergido en análisis anteriores.

En resumen, este estudio proporciona evidencia sólida de que la firma genómica asociada al número de neuronas en la corteza cerebral está compuesta por familias de genes que codifican proteínas funcionalmente especializadas, involucradas en procesos clave como la inmunidad y la señalización intercelular. Esta interacción entre mecanismos inmunológicos y neuronales habría facilitado el desarrollo de funciones cognitivas avanzadas, contribuyendo significativamente a la especialización funcional y a la expansión de la corteza cerebral en mamíferos. Así mismo, esta coincidencia funcional también podría sugerir un posible fenómeno de reutilización de genes, en el cual genes originalmente asociados con la defensa inmunológica habrían sido cooptados durante la evolución para cumplir funciones clave en la organización y mantenimiento de la corteza cerebral.

9.1 Propuesta de Modelo integrador de la expansión neuronal cortical

Con base en los análisis realizados, se propone un modelo integrador que vincula la expansión neuronal cortical con reorganizaciones genómicas y estructurales en familias de genes conservadas en mamíferos. Bajo este marco, la duplicación y expansión de estas familias incrementan el repertorio molecular disponible, mientras que la adopción de una arquitectura de dominios proteicos convergente emerge como el mecanismo clave para la innovación funcional. Esta estrategia permite diversificar las capacidades celulares sin requerir la aparición de dominios de novo, basándose en la amplificación estratégica de unidades funcionales específicas en linajes no relacionados. Dicha modularidad potencia la capacidad regulatoria y de interacción molecular, factores críticos para la especialización de la corteza cerebral (Figura 36).

En este contexto, la asociación de dominios vinculados a procesos metabólicos y de comunicación inmunológica con el número de neuronas no sugiere una sustitución de las funciones gliales. Por el contrario, refleja una asimilación neuronal de capacidades metabólicas e inmunorregulatorias, probablemente como una adaptación evolutiva a las demandas energéticas y de señalización extremas que impone la expansión de la red neuronal. En conjunto, este modelo postula que la evolución cerebral en mamíferos es el resultado de una sinergia entre la expansión génica, la convergencia modular proteica y la sofisticación regulatoria; procesos coordinados que permitieron alcanzar los umbrales de complejidad y densidad neuronal característicos de los cerebros de mayor tamaño."

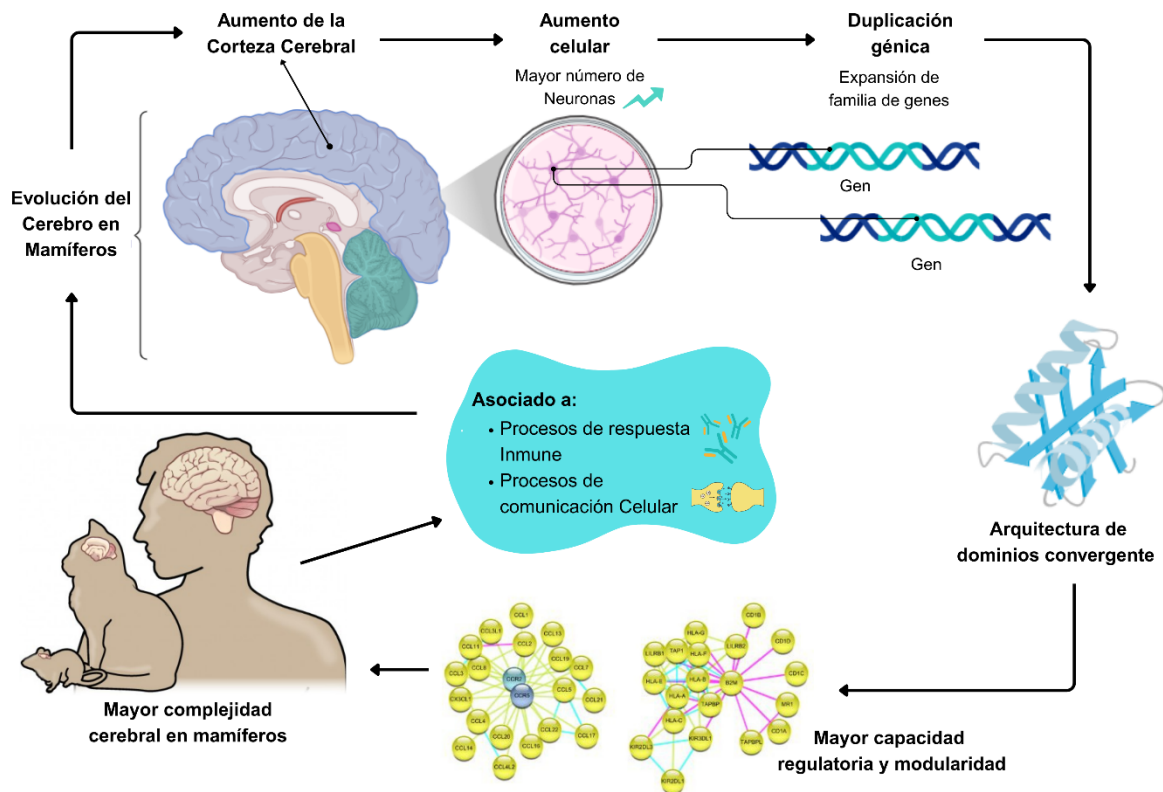


Figura 36. Modelo integrador de la expansión neuronal cortical en mamíferos. El esquema resume la sinergia entre duplicación génica, expansión de familias de genes y una arquitectura de dominios proteicos convergente, como mecanismos asociados al aumento del número de neuronas en la corteza cerebral y la complejidad estructural de la corteza cerebral en mamíferos.

X. Conclusiones

- Los análisis realizados en este estudio revelan la existencia de una posible firma genómica asociada a la expansión del número de neuronas en la corteza cerebral de mamíferos. Se identificaron 385 familias de genes cuya expansión se relaciona principalmente con procesos inmunológicos y de comunicación célula-célula, lo que sugiere que dichos mecanismos han desempeñado un papel clave en la especialización funcional de la corteza cerebral.
- La variabilidad en el número de copias, desde un nivel basal de 5 copias hasta una expansión de 120 copias, revela que la expansión cortical del cerebro ha sido impulsada por una expansión selectiva de familias génicas específicas.
- La caracterización de las familias génicas asociadas al número de neuronas en corteza cerebral mediante dominios *InterPro* permitió describir su organización estructural desde una perspectiva complementaria al análisis funcional. Los resultados indican que estas familias comparten una arquitectura proteica convergente. La abundancia de dominios clave (Ig-like, Kinasas, P-loop NTPase) en

especies no relacionadas revela un determinismo molecular, sugiriendo que la expansión del cerebro requiere un set proteico universal para gestionar la conectividad y el metabolismo neuronal."

- El análisis de expresión génica espacio-temporal mostró que estas familias de genes presentan patrones diferenciales de actividad a lo largo del desarrollo cerebral: baja expresión en etapa fetal, sobreexpresión durante la infancia (período crítico para la formación de circuitos neuronales) y una estabilización en la adultez. Estos resultados refuerzan la hipótesis de que la expansión cortical ha estado favorecida por una regulación espacio-temporal precisa de la expresión génica.
- En cuanto a su distribución cromosómica, se observó que estas familias se agrupan en regiones específicas del genoma, reflejando una regulación coordinada entre genes relacionados o de la misma familia. La presencia de zonas con alta actividad de duplicación sugiere regiones sometidas a presiones evolutivas que habrían favorecido la diversificación funcional y la optimización de procesos neuronales complejos. Sin embargo, la dispersión de algunas familias en diferentes regiones del genoma indica también la existencia de regulaciones independientes o la pérdida de vínculos funcionales ancestrales, evidenciando procesos de especialización génica.
- El análisis de redes de interacción proteína-proteína (PPI) demostró que las proteínas codificadas por estas familias participan en procesos biológicos esenciales para la expansión y funcionalidad de la corteza cerebral. Entre ellos destacan la señalización neuronal, la plasticidad sináptica, la neuroprotección, la remodelación de circuitos neuronales, la respuesta inmune y la adaptación a estímulos externos, todos fundamentales para la consolidación de funciones cognitivas complejas.
- Uno de los hallazgos más relevantes de este estudio es la asociación entre procesos inmunológicos y funciones neuronales, lo que podría representar un caso de cooptación de genes. En este escenario, genes originalmente implicados en la defensa inmunológica habrían sido reutilizados durante la evolución para modular funciones como la plasticidad sináptica, la homeostasis neuronal y la poda sináptica. Esta convergencia funcional entre los sistemas inmune y nervioso habría optimizado la conectividad cerebral y la eficiencia de las redes corticales, favoreciendo la emergencia de funciones cognitivas superiores y una mayor capacidad de adaptación a entornos cambiantes.

- Cabe mencionar que los resultados indican que la expansión del número de neuronas en la corteza cerebral estuvo acompañada no solo por cambios en funciones neuronales clásicas, sino también por la incorporación de capacidades metabólicas e inmunorregulatorias. Implicando una reorganización funcional profunda de las neuronas, en interacción estrecha con otros tipos celulares, particularmente la glía, cuya contribución específica sería abordada en estudios independientes.

En conjunto, los resultados sugieren que la expansión cortical en mamíferos no estuvo impulsada únicamente por la duplicación y expansión génica, sino también por la modulación funcional de proteínas con roles tanto inmunológicos como neuronales. Esto habría potenciado la capacidad del cerebro para procesar información compleja, sostener funciones cognitivas avanzadas y adaptarse a un ambiente dinámico, constituyendo un paso clave en la evolución del cerebro de los mamíferos.

XI. Perspectivas

Este estudio constituye el primer esfuerzo en identificar una firma genómica asociada con el número de neuronas en la corteza cerebral, conformada por 385 familias de genes. Aunque los resultados ofrecen una base sólida, aún es necesario profundizar en estudios funcionales que permitan esclarecer el impacto específico de estas familias en el desarrollo y la especialización cortical.

La expresión diferencial de estas familias a lo largo del desarrollo sugiere un papel regulado espacio-temporal, con sobreexpresión en la infancia y estabilización en la adultez. No obstante, estos patrones podrían estar influidos por factores externos como el ambiente, el estado fisiológico o el contexto patológico, aspectos difíciles de controlar en conjuntos de datos públicos. Esto resalta la necesidad de validaciones experimentales en condiciones controladas y de análisis comparativos entre especies para distinguir señales evolutivamente conservadas de respuestas contextuales.

La identificación de regiones genómicas sujetas a duplicaciones y expansiones sugiere la existencia de zonas evolutivamente dinámicas clave para la especialización cortical. Sin embargo, aún se desconoce cómo las variaciones estructurales y regulatorias han modelado estos procesos a lo largo de la evolución.

El análisis de redes de interacción proteína-proteína evidenció la participación de módulos funcionales relacionados con la plasticidad sináptica, la señalización neuronal y la respuesta inmune. Un reto pendiente es comprender cómo estas redes se han diversificado

en distintos linajes de mamíferos y cómo contribuyen a la variabilidad de capacidades cognitivas entre especies.

Asimismo, la posible co-evolución entre el sistema inmune y el cerebro plantea que genes originalmente vinculados con defensa inmunológica pudieron haber sido reutilizados para modular procesos neuronales clave, como la poda sináptica, la homeostasis y la plasticidad sináptica. Dilucidar cómo ocurrió esta convergencia funcional es esencial para entender la emergencia de funciones cognitivas avanzadas.

Finalmente, para lograr una visión integral será necesario implementar enfoques multi-ómicos que integren datos genómicos, transcriptómicos, epigenómicos y proteómicos. Este tipo de aproximaciones permitirá caracterizar con mayor precisión los mecanismos moleculares y evolutivos que dieron forma a la corteza cerebral.

En conjunto, los resultados de este trabajo abren nuevas vías de investigación en neurociencia evolutiva, neurobiología del desarrollo y biomedicina, ofreciendo un marco para explorar los vínculos entre genómica, plasticidad neuronal y evolución cortical en mamíferos.

XII Limitaciones

Este estudio se basa en un enfoque de genómica comparativa, por lo que las asociaciones detectadas entre familias génicas, dominios proteicos y el número de neuronas corticales representan correlaciones evolutivas a gran escala y no implican causalidad directa. Asimismo, los análisis dependen de anotaciones funcionales y estructurales disponibles en bases de datos públicas, cuya cobertura puede variar entre especies.

El trabajo se centra de manera específica en neuronas de la corteza cerebral, excluyendo deliberadamente el análisis directo de otros tipos celulares, como la glía, esta delimitación responde a una decisión metodológica explícita y al objetivo central del trabajo, y no pretende minimizar la relevancia funcional o evolutiva de otros componentes celulares. Finalmente, los análisis de dominios y de interacción proteína–proteína reflejan patrones generales de organización funcional, pero no capturan dinámicas contextuales específicas.

XIII. Referencias

1. Allen, N. J., & Lyons, D. A. (2018). Glia as architects of central nervous system formation and function. *Science*, 362(6411), 181–185. <https://doi.org/10.1126/science.aat0473>
2. Angiari, S., & Constantin, G. (2013). Selectins and Their Ligands as Potential Immunotherapeutic Targets in Neurological Diseases. *Immunotherapy*, 5(11), 1207–1220. <https://doi.org/10.2217/imt.13.122>
3. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
4. Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Filho, W. J., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532–541. <https://doi.org/10.1002/cne.21974>
5. Ba, W., Yan, Y., Reijnders, M. R. F., Schuurs-Hoeijmakers, J. H. M., Feenstra, I., Bongers, E. M. H. F., Bosch, D. G. M., De Leeuw, N., Pfundt, R., Gilissen, C., De Vries, P. F., Veltman, J. A., Hoischen, A., Mefford, H. C., Eichler, E. E., Vissers, L. E. L. M., Nadif Kasri, N., & De Vries, B. B. A. (2016). *TRIO* loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Human Molecular Genetics*, 25(5), 892–902. <https://doi.org/10.1093/hmg/ddv618>
6. Bengtsson, H. (2020). *parallelly: Enhancing the 'parallel' Package* (p. 1.38.0) [Dataset]. <https://doi.org/10.32614/CRAN.package.parallelly>
7. Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
8. Bilbo, S. D., & Schwarz, J. M. (2012). The immune system and developmental programming of brain and behavior. *Frontiers in Neuroendocrinology*, 33(3), 267–286. <https://doi.org/10.1016/j.yfrne.2012.08.006>
9. Borsini, A., Zunszain, P. A., Thuret, S., & Pariante, C. M. (2015). The role of inflammatory cytokines as key modulators of neurogenesis. *Trends in Neurosciences*, 38(3), 145–157. <https://doi.org/10.1016/j.tins.2014.12.006>
10. Boyden, S. E., Desai, A., Cruse, G., Young, M. L., Bolan, H. C., Scott, L. M., Eisch, A. R., Long, R. D., Lee, C.-C. R., Satorius, C. L., Pakstis, A. J., Olivera, A., Mullikin, J. C., Chouery, E., Mégarbané, A., Medlej-Hashim, M., Kidd, K. K., Kastner, D. L., Metcalfe, D. D., & Komarow, H. D. (2016). Vibratory Urticaria Associated with a Missense Variant in *ADGRE2*. *New England Journal of Medicine*, 374(7), 656–663. <https://doi.org/10.1056/NEJMoa1500611>
11. Brown, T. A. (Ed.). (2002). *Genomes* (2nd ed). Wiley-Liss. <https://www.ncbi.nlm.nih.gov/books/NBK21128/>
12. Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: How skewed distributions affect network operations. *Nature Reviews Neuroscience*, 15(4), 264–278. <https://doi.org/10.1038/nrn3687>
13. Cahalane, D. J., Charvet, C. J., & Finlay, B. L. (2014). Modeling local and cross-species neuron number variations in the cerebral cortex as arising from a common mechanism. *Proceedings of the National Academy of Sciences*, 111(49), 17642–17647. <https://doi.org/10.1073/pnas.1409271111>
14. Cameron, J. L. (2004). Interrelationships between Hormones, Behavior, and Affect during Adolescence: Understanding Hormonal, Physical, and Brain Changes Occurring in Association with Pubertal Activation of the Reproductive Axis. Introduction to Part III. *Annals of the New York Academy of Sciences*, 1021(1), 110–123. <https://doi.org/10.1196/annals.1308.012>
15. Castillo-Morales, A., Monzón-Sandoval, J., De Sousa, A. A., Urrutia, A. O., & Gutierrez, H. (2016). Neocortex expansion is linked to size variations in gene families with chemotaxis, cell–cell signalling and immune response

- functions in mammals. *Open Biology*, 6(10), 160132. <https://doi.org/10.1098/rsob.160132>
16. Castillo-Morales, A., Monzón-Sandoval, J., Urrutia, A. O., & Gutiérrez, H. (2014). Increased brain size in mammals is associated with size variations in gene families with cell signalling, chemotaxis and immune-related functions. *Proceedings of the Royal Society B: Biological Sciences*, 281(1775), 20132428. <https://doi.org/10.1098/rspb.2013.2428>
 17. Chakraborty, P., Huang, J. T. J., & Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nature Communications*, 9(1), 4346. <https://doi.org/10.1038/s41467-018-06677-1>
 18. Cirulli, E. T., Schiabor Barrett, K. M., Bolze, A., Grzymalski, J. J., Lee, W., & Washington, N. L. (2022). *A power-based sliding window approach to evaluate the clinical impact of rare genetic variants*. Genetic and Genomic Medicine. <https://doi.org/10.1101/2022.07.29.22278171>
 19. De Felipe, J. (2011). The Evolution of the Brain, the Human Nature of Cortical Circuits, and Intellectual Creativity. *Frontiers in Neuroanatomy*, 5. <https://doi.org/10.3389/fnana.2011.00029>
 20. Decourt, C., Schaeffer, J., Blot, B., Paccard, A., Excoffier, B., Pende, M., Nawabi, H., & Belin, S. (2023). The RSK2-RPS6 axis promotes axonal regeneration in the peripheral and central nervous systems. *PLOS Biology*, 21(4), e3002044. <https://doi.org/10.1371/journal.pbio.3002044>
 21. Demuth, J. P., Bie, T. D., Stajich, J. E., Cristianini, N., & Hahn, M. W. (2006). The Evolution of Mammalian Gene Families. *PLoS ONE*, 1(1), e85. <https://doi.org/10.1371/journal.pone.0000085>
 22. Dobbing, J., & Sands, J. (1973). Quantitative growth and development of human brain. *Archives of Disease in Childhood*, 48(10), 757–767. <https://doi.org/10.1136/adc.48.10.757>
 23. Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, 18(2), 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>
 24. Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
 25. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16), 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>
 26. Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8), 1184–1191. <https://doi.org/10.1038/nprot.2009.97>
 27. Edelbrock, M. A., Kaliyaperumal, S., & Williams, K. J. (2013). Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 743–744, 53–66. <https://doi.org/10.1016/j.mrfmmm.2012.12.008>
 28. Elsner, H. -A., Schroeder, M., & Blasczyk, R. (2001). The nucleotide diversity of MICA and MICB suggests the effect of overdominant selection. *Tissue Antigens*, 58(6), 419–421. <https://doi.org/10.1034/j.1399-0039.2001.580612.x>
 29. European Bioinformatics Institute. (2024a). *GO:0002250 adaptive immune response*. <https://www.ebi.ac.uk/QuickGO/term/GO:0002250>
 30. European Bioinformatics Institute. (2024b). *GO:0006950 response to stress*. <https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0006950#term=history>
 31. European Bioinformatics Institute. (2024c). *GO:0006954 inflammatory response*.

- <https://www.ebi.ac.uk/QuickGO/term/GO:0006954>
32. European Bioinformatics Institute. (2024d). *GO:0009605 response to external stimulus*. <https://www.ebi.ac.uk/QuickGO/term/GO:0009605>
 33. Filiano, A. J., Gadani, S. P., & Kipnis, J. (2015). Interactions of innate and adaptive immunity in brain development and function. *Brain Research*, 1617, 18–27. <https://doi.org/10.1016/j.brainres.2014.07.050>
 34. Florio, M., Borrell, V., & Huttner, W. B. (2017). Human-specific genomic signatures of neocortical expansion. *Current Opinion in Neurobiology*, 42, 33–44. <https://doi.org/10.1016/j.conb.2016.11.004>
 35. Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., Wimberger, P., Huttner, W. B., & Hiller, M. (2018). Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife*, 7, e32332. <https://doi.org/10.7554/eLife.32332>
 36. Florio, M., & Huttner, W. B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. *Development*, 141(11), 2182–2194. <https://doi.org/10.1242/dev.090571>
 37. Gabi, M., Collins, C. E., Wong, P., Torres, L. B., Kaas, J. H., & Herculano-Houzel, S. (2010). Cellular Scaling Rules for the Brains of an Extended Number of Primate Species. *Brain, Behavior and Evolution*, 76(1), 32–44. <https://doi.org/10.1159/000319872>
 38. Gamache, T. R., Araki, Y., & Huganir, R. L. (2020). Twenty Years of SynGAP Research: From Synapses to Cognition. *The Journal of Neuroscience*, 40(8), 1596–1605. <https://doi.org/10.1523/JNEUROSCI.0420-19.2020>
 39. Garcia Concejo, A., & Larhammar, D. (2021). Protein kinase C family evolution in jawed vertebrates. *Developmental Biology*, 479, 77–90. <https://doi.org/10.1016/j.ydbio.2021.07.013>
 40. Gatt, J. M., Williams, L. M., Schofield, P. R., Dobson-Stone, C., Paul, R. H., Grieve, S. M., Clark, C. R., Gordon, E., & Nemeroff, C. B. (2010). Impact of the *HTR3A* gene with early life trauma on emotional brain networks and depressed mood: Research Article: *HTR3A*, Trauma, Emotional Brain, and Mood. *Depression and Anxiety*, 27(8), 752–759. <https://doi.org/10.1002/da.20726>
 41. Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
 42. Gene Ontology Consortium. (2025a). *GO:0002376—Immune system process*. <https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0002376#term=history>
 43. Gene Ontology Consortium. (2025b). *GO:0007166—Cell surface receptor signaling pathway*. <https://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0007166#term=history>
 44. *Gene—NCBI*. (2025, January 10). <https://www.ncbi.nlm.nih.gov/gene/>
 45. Ghashghaei, H. T., Weimer, J. M., Schmid, R. S., Yokota, Y., McCarthy, K. D., Popko, B., & Anton, E. S. (2007). Reinduction of ErbB2 in astrocytes promotes radial glial progenitor identity in adult cerebral cortex. *Genes & Development*, 21(24), 3258–3271. <https://doi.org/10.1101/gad.1580407>
 46. Glueck, D. H., Mandel, J., Karimpour-Fard, A., Hunter, L., & Muller, K. E. (2008). Exact Calculations of Average Power for the Benjamini-Hochberg Procedure. *The International Journal of Biostatistics*, 4(1). <https://doi.org/10.2202/1557-4679.1103>
 47. Grafen, A. (1992). The uniqueness of the phylogenetic regression. *Journal of Theoretical Biology*, 156(4), 405–423. [https://doi.org/10.1016/S0022-5193\(05\)80635-6](https://doi.org/10.1016/S0022-5193(05)80635-6)
 48. Graham, M. H. (2003). CONFRONTING MULTICOLLINEARITY IN ECOLOGICAL MULTIPLE REGRESSION. *Ecology*, 84(11), 2809–2815. <https://doi.org/10.1890/02-3114>
 49. Grison, A., Gaiser, C., Bieder, A., Baranek, C., & Atanasoski, S. (2018). Ablation of cdk4 and cdk6 affects proliferation of basal progenitor cells in the developing dorsal and ventral forebrain. *Developmental Neurobiology*, 78(7), 660–670. <https://doi.org/10.1002/dneu.22588>
 50. Hallböök, F., Wilson, K., Thorndyke, M., & Olinski, R. P. (2006). Formation and Evolution of the Chordate Neurotrophin and Trk Receptor Genes. *Brain, Behavior and Evolution*, 68(3), 133–144.

<https://doi.org/10.1159/000094083>

51. Hartenstein, V., & Stollewerk, A. (2015). The Evolution of Early Neurogenesis. *Developmental Cell*, 32(4), 390–407. <https://doi.org/10.1016/j.devcel.2015.02.004>
52. Herculano-Houzel, S. (2009). The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3. <https://doi.org/10.3389/neuro.09.031.2009>
53. Herculano-Houzel, S. (2014). The glia/neuron ratio: How it varies uniformly across brain structures and species and what that means for brain physiology and evolution. *Glia*, 62(9), 1377–1391. <https://doi.org/10.1002/glia.22683>
54. Herculano-Houzel, S. (2017). Numbers of neurons as biological correlates of cognitive capability. *Current Opinion in Behavioral Sciences*, 16, 1–7. <https://doi.org/10.1016/j.cobeha.2017.02.004>
55. Herculano-Houzel, S., Avelino de Souza, K., Neves, K., Porfirio, J., Messeder, D., Mattos Feij³, L., Maldonado, J., & Manger, P. R. (2014). The elephant brain in numbers. *Frontiers in Neuroanatomy*, 8. <https://doi.org/10.3389/fnana.2014.00046>
56. Herculano-Houzel, S., Collins, C. E., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences*, 104(9), 3562–3567. <https://doi.org/10.1073/pnas.0611396104>
57. Herculano-Houzel, S., & Kaas, J. H. (2011). Gorilla and Orangutan Brains Conform to the Primate Cellular Scaling Rules: Implications for Human Evolution. *Brain, Behavior and Evolution*, 77(1), 33–44. <https://doi.org/10.1159/000322729>
58. Herculano-Houzel, S., & Lent, R. (2005). Isotropic Fractionator: A Simple, Rapid Method for the Quantification of Total Cell and Neuron Numbers in the Brain. *The Journal of Neuroscience*, 25(10), 2518–2521. <https://doi.org/10.1523/JNEUROSCI.4526-04.2005>
59. Herculano-Houzel, S., Ribeiro, P., Campos, L., Valotta Da Silva, A., Torres, L. B., Catania, K. C., & Kaas, J. H. (2011). Updated Neuronal Scaling Rules for the Brains of Glires (Rodents/Lagomorphs). *Brain, Behavior and Evolution*, 78(4), 302–314. <https://doi.org/10.1159/000330825>
60. Huber, A. K., Giles, D. A., Segal, B. M., & Irani, D. N. (2018). An emerging role for eotaxins in neurodegenerative disease. *Clinical Immunology*, 189, 29–33. <https://doi.org/10.1016/j.clim.2016.09.010>
61. Hughes, A. (1991). Evolutionary origin and diversification of the mammalian CD1 antigen genes. *Molecular Biology and Evolution*, 8(2), 185–201. <https://doi.org/10.1093/oxfordjournals.molbev.a040640>
62. Hurst, L. D., Ghanbarian, A. T., Forrest, A. R. R., FANTOM consortium, & Huminiecki, L. (2015). The Constrained Maximal Expression Level Owing to Haploidy Shapes Gene Content on the Mammalian X Chromosome. *PLOS Biology*, 13(12), e1002315. <https://doi.org/10.1371/journal.pbio.1002315>
63. Huttenlocher, P. R., De Courten, C., Garey, L. J., & Van Der Loos, H. (1982). Synaptogenesis in human visual cortex—Evidence for synapse elimination during normal development. *Neuroscience Letters*, 33(3), 247–252. [https://doi.org/10.1016/0304-3940\(82\)90379-2](https://doi.org/10.1016/0304-3940(82)90379-2)
64. Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B., & Gros, P. (2005). Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature*, 437(7058), 505–511. <https://doi.org/10.1038/nature04005>
65. Jardim-Messeder, D., Lambert, K., Noctor, S., Pestana, F. M., De Castro Leal, M. E., Bertelsen, M. F., Alagaili, A. N., Mohammad, O. B., Manger, P. R., & Herculano-Houzel, S. (2017). Dogs Have the Most Neurons, Though Not the Largest Brain: Trade-Off between Body Mass and Number of Neurons in the Cerebral Cortex of Large Carnivorous Species. *Frontiers in Neuroanatomy*, 11, 118. <https://doi.org/10.3389/fnana.2017.00118>
66. Jerison, H. (1974). Evolution of the Brain and Intelligence. *Science*, 184(4137), 677–679.
67. Kaas, J. H. (1989). The evolution of complex sensory systems in mammals. *Journal of Experimental Biology*, 146(1), 165–176. <https://doi.org/10.1242/jeb.146.1.165>
68. Kan, B., Michalski, C., Fu, H., Au, H. H. T., Lee, K., Marchant, E. A., Cheng, M. F., Anderson-Baucum, E.,

- Aharoni-Simon, M., Tilley, P., Mirmira, R. G., Ross, C. J., Luciani, D. S., Jan, E., & Lavoie, P. M. (2018). Cellular metabolism constrains innate immune responses in early human ontogeny. *Nature Communications*, 9(1), 4822. <https://doi.org/10.1038/s41467-018-07215-9>
69. Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A. M. M., Pletikos, M., Meyer, K. A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M. B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S. N., Vortmeyer, A., ... Sestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, 478(7370), 483–489. <https://doi.org/10.1038/nature10523>
 70. Kazu, R. S., Maldonado, J., Mota, B., Manger, P. R., & Herculano-Houzel, S. (2014). Cellular scaling rules for the brain of Artiodactyla include a highly folded cortex with few neurons. *Frontiers in Neuroanatomy*, 8. <https://doi.org/10.3389/fnana.2014.00128>
 71. Keesey, T. M. (2024, September 30). *PhyloPic*. <https://www.phylopic.org>
 72. Kelava, I., Lewitus, E., & Huttner, W. B. (2013). The secondary loss of gyrencephaly as an example of evolutionary phenotypical reversal. *Frontiers in Neuroanatomy*, 7. <https://doi.org/10.3389/fnana.2013.00016>
 73. Konopka, G., & Geschwind, D. H. (2010). Human Brain Evolution: Harnessing the Genomics (R)evolution to Link Genes, Cognition, and Behavior. *Neuron*, 68(2), 231–244. <https://doi.org/10.1016/j.neuron.2010.10.012>
 74. Kovaleva, I. E., Tokarchuk, A. V., Zheltukhin, A. O., Dalina, A. A., Safronov, G. G., Evstafieva, A. G., Lyamzaev, K. G., Chumakov, P. M., & Budanov, A. V. (2020). Mitochondrial localization of SESN2. *PLOS ONE*, 15(4), e0226862. <https://doi.org/10.1371/journal.pone.0226862>
 75. Kverková, K., Marhounová, L., Polonyiová, A., Kocourek, M., Zhang, Y., Olkowicz, S., Straková, B., Pavelková, Z., Vodička, R., Frynta, D., & Němec, P. (2022). The evolution of brain neuron numbers in amniotes. *Proceedings of the National Academy of Sciences*, 119(11), e2121624119. <https://doi.org/10.1073/pnas.2121624119>
 76. Lai, S., Wang, P., Gong, J., & Zhang, S. (2023). New insights into the role of GSK-3 β in the brain: From neurodegenerative disease to tumorigenesis. *PeerJ*, 11, e16635. <https://doi.org/10.7717/peerj.16635>
 77. Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, 52(W1), W78–W82. <https://doi.org/10.1093/nar/gkac268>
 78. Li, L., Dong, M., & Wang, X.-G. (2016). The Implication and Significance of Beta 2 Microglobulin: A Conservative Multifunctional Regulator. *Chinese Medical Journal*, 129(4), 448–455. <https://doi.org/10.4103/0366-6999.176084>
 79. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S., Mungall, C. J., Arner, E., Baillie, J. K., Bertin, N., Bono, H., De Hoon, M., Diehl, A. D., Dimont, E., Freeman, T. C., ... Kawaji, H. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biology*, 16(1), 22. <https://doi.org/10.1186/s13059-014-0560-6>
 80. Lodato, S., & Arlotta, P. (2015). Generating Neuronal Diversity in the Mammalian Cerebral Cortex. *Annual Review of Cell and Developmental Biology*, 31(1), 699–720. <https://doi.org/10.1146/annurev-cellbio-100814-125353>
 81. Lui, J. H., Hansen, D. V., & Kriegstein, A. R. (2011). Development and Evolution of the Human Neocortex. *Cell*, 146(1), 18–36. <https://doi.org/10.1016/j.cell.2011.06.030>
 82. Luo, S., Dong, X., Guo, S., Wang, Q., Dai, X., Jiang, Y., Zhu, W., Zhou, W., & Song, W. (2022). Regulation of the Human IL-10RB Gene Expression by Sp8 and Sp9. *Journal of Alzheimer's Disease*, 88(4), 1469–1485. <https://doi.org/10.3233/JAD-220321>
 83. Marino, L. (1998). A Comparison of Encephalization between Odontocete Cetaceans and Anthropoid Primates. *Brain, Behavior and Evolution*, 51(4), 230–238. <https://doi.org/10.1159/000006540>
 84. Marsh, R., Gerber, A. J., & Peterson, B. S. (2008). Neuroimaging Studies of Normal Brain Development and Their Relevance for Understanding Childhood Neuropsychiatric Disorders. *Journal of the American Academy of*

- Child & Adolescent Psychiatry*, 47(11), 1233–1251. <https://doi.org/10.1097/CHI.0b013e318185e703>
85. McEwen, B. S. (2013). The Brain on Stress: Toward an Integrative Approach to Brain, Body, and Behavior. *Perspectives on Psychological Science*, 8(6), 673–675. <https://doi.org/10.1177/1745691613506907>
 86. Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., Ebbert, A., Riley, Z. L., Royall, J. J., Aiona, K., Arnold, J. M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R. A., ... Lein, E. S. (2014). Transcriptional landscape of the prenatal human brain. *Nature*, 508(7495), 199–206. <https://doi.org/10.1038/nature13185>
 87. Monteiro, P., & Feng, G. (2017). SHANK proteins: Roles at the synapse and in autism spectrum disorder. *Nature Reviews Neuroscience*, 18(3), 147–157. <https://doi.org/10.1038/nrn.2016.183>
 88. Mota, B., & Herculano-Houzel, S. (2014). All brains are made of this: A fundamental building block of brain matter with matching neuronal and glial masses. *Frontiers in Neuroanatomy*, 8. <https://doi.org/10.3389/fnana.2014.00127>
 89. Namba, T., & Huttner, W. B. (2017). Neural progenitor cells and their role in the development and evolutionary expansion of the neocortex. *WIREs Developmental Biology*, 6(1), e256. <https://doi.org/10.1002/wdev.256>
 90. Nataf, S. (2017). Autoimmunity as a Driving Force of Cognitive Evolution. *Frontiers in Neuroscience*, 11, 582. <https://doi.org/10.3389/fnins.2017.00582>
 91. Nguyen, P., Kdeiss, B., Ord, S., Hess, K., Oliverio, R., & Nikolaidis, N. (2017). Evolution and natural variation of *HSPA1A*, the major stress inducible gene, in humans. *The FASEB Journal*, 31(S1). https://doi.org/10.1096/fasebj.31.1_supplement.604.18
 92. Oliva, C. A., Vargas, J. Y., & Inestrosa, N. C. (2013). Wnts in adult brain: From synaptic plasticity to cognitive deficiencies. *Frontiers in Cellular Neuroscience*, 7. <https://doi.org/10.3389/fncel.2013.00224>
 93. Pedersen, J. E., Bergqvist, C. A., & Larhammar, D. (2019). Evolution of vertebrate nicotinic acetylcholine receptors. *BMC Evolutionary Biology*, 19(1), 38. <https://doi.org/10.1186/s12862-018-1341-8>
 94. PEkowska, A., Verkhatsky, A., & Falcone, C. (2025). Evolution of neuroglia: From worm to man. In *Handbook of Clinical Neurology* (Vol. 209, pp. 7–26). Elsevier. <https://doi.org/10.1016/B978-0-443-19104-6.00004-8>
 95. Pessoa, L., Medina, L., Hof, P. R., & Desfilis, E. (2019). Neural architecture of the vertebrate brain: Implications for the interaction between emotion and cognition. *Neuroscience & Biobehavioral Reviews*, 107, 296–312. <https://doi.org/10.1016/j.neubiorev.2019.09.021>
 96. Pinheiro, J., & Bates, D. (2023, November 27). *nlme: Linear and Nonlinear Mixed Effects Models, R package version 3.1-164*. <https://cran.r-project.org/web/packages/nlme/>
 97. Rakic, P. (2000). Radial Unit Hypothesis of Neocortical Expansion. In G. R. Bock & G. Cardew (Eds.), *Novartis Foundation Symposia* (1st ed., Vol. 228, pp. 30–45). Wiley. <https://doi.org/10.1002/0470846631.ch3>
 98. Rakic, P. (2009). Evolution of the neocortex: A perspective from developmental biology. *Nature Reviews Neuroscience*, 10(10), 724–735. <https://doi.org/10.1038/nrn2719>
 99. Ryan, K., Lu, Z., & Meinertzhagen, I. A. (2016). The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *eLife*, 5, e16962. <https://doi.org/10.7554/eLife.16962>
 100. Sancho, L., Contreras, M., & Allen, N. J. (2021). Glia as sculptors of synaptic plasticity. *Neuroscience Research*, 167, 17–29. <https://doi.org/10.1016/j.neures.2020.11.005>
 101. Sarmiento, L., Svensson, J., Barchetta, I., Giwerzman, A., & Cilio, C. M. (2019). Copy number of the X-linked genes TLR7 and CD40L influences innate and adaptive immune responses. *Scandinavian Journal of Immunology*, 90(2), e12776. <https://doi.org/10.1111/sji.12776>
 102. Shaw, R. E., Kottler, B., Ludlow, Z. N., Buhl, E., Kim, D., Morais Da Silva, S., Miedzik, A., Coum, A., Hodge, J. J., Hirth, F., & Sousa-Nunes, R. (2018). In vivo expansion of functionally integrated GABAergic interneurons by targeted increase in neural progenitors. *The EMBO Journal*, 37(13), e98163. <https://doi.org/10.15252/embj.201798163>

103. Shultz, S., & Dunbar, R. (2010). Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences*, 107(50), 21582–21586. <https://doi.org/10.1073/pnas.1005246107>
104. Solé, H. (2022). Capítulo 9: Corteza Cerebral. *Revista Argentina de Neurocirugía*, 36(04). <https://doi.org/10.59156/revista.v36i04.481>
105. Stegmüller, J., & Bonni, A. (2005). Moving past proliferation: New roles for Cdh1–APC in postmitotic neurons. *Trends in Neurosciences*, 28(11), 596–601. <https://doi.org/10.1016/j.tins.2005.09.003>
106. Sterner, K. N., Weckle, A., Chugani, H. T., Tarca, A. L., Sherwood, C. C., Hof, P. R., Kuzawa, C. W., Boddy, A. M., Abbas, A., Raaum, R. L., Grégoire, L., Lipovich, L., Grossman, L. I., Uddin, M., Goodman, M., & Wildman, D. E. (2012). Dynamic Gene Expression in the Human Cerebral Cortex Distinguishes Children from Adults. *PLoS ONE*, 7(5), e37714. <https://doi.org/10.1371/journal.pone.0037714>
107. Stevanovic, M., Drakulic, D., Lazic, A., Ninkovic, D. S., Schwirtlich, M., & Mojsin, M. (2021). SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Frontiers in Molecular Neuroscience*, 14, 654031. <https://doi.org/10.3389/fnmol.2021.654031>
108. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, 6(7), e21800. <https://doi.org/10.1371/journal.pone.0021800>
109. Suzuki, I. K., Gacquer, D., Van Heurck, R., Kumar, D., Wojno, M., Bilheu, A., Herpoel, A., Lambert, N., Cheron, J., Polleux, F., Detours, V., & Vanderhaeghen, P. (2018). Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell*, 173(6), 1370–1384.e16. <https://doi.org/10.1016/j.cell.2018.03.067>
110. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646. <https://doi.org/10.1093/nar/gkac1000>
111. Twohig, J. P., Cuff, S. M., Yong, A. A., & Wang, E. C. Y. (2011). The role of tumor necrosis factor receptor superfamily members in mammalian brain development, function and homeostasis. *Reviews in the Neurosciences*, 22(5). <https://doi.org/10.1515/RNS.2011.041>
112. Verkhatsky, A., Arranz, A. M., Ciuba, K., & Pękowska, A. (2022). Evolution of neuroglia. *Annals of the New York Academy of Sciences*, 1518(1), 120–130. <https://doi.org/10.1111/nyas.14917>
113. Watanabe, H. (2017). Back Through Time: How Cnidarians and Basal Metazoans Shed Light on Ancient Nervous Systems. In S. Shigeno, Y. Murakami, & T. Nomura (Eds.), *Brain Evolution by Design* (pp. 45–75). Springer Japan. https://doi.org/10.1007/978-4-431-56469-0_3
114. Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and Renormalization of Human Brain Structure During Skill Acquisition. *Trends in Cognitive Sciences*, 21(12), 930–939. <https://doi.org/10.1016/j.tics.2017.09.008>
115. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
116. Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2014). *dplyr: A Grammar of Data Manipulation* (p. 1.1.4) [Dataset]. <https://doi.org/10.32614/CRAN.package.dplyr>
117. Wiślak, W. (2013). Cell Signaling Pathways. In W. Wiślak (Ed.), *Molecular Biology* (Vol. 8248, pp. 121–138). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-45361-8_8
118. Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., ... Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, gkz966.

<https://doi.org/10.1093/nar/gkz966>

119. Zhang, J. (2003). Evolution of the Human *ASPM* Gene, a Major Determinant of Brain Size. *Genetics*, 165(4), 2063–2070. <https://doi.org/10.1093/genetics/165.4.2063>
120. Zilles, K., & Amunts, K. (2012). Architecture of the Cerebral Cortex. In *The Human Nervous System* (pp. 836–895). Elsevier. <https://doi.org/10.1016/B978-0-12-374236-0.10023-9>

XIV. ANEXOS

14.1 Tablas suplementarias

Tabla S1. Tabla general de familias asociadas significativamente con parámetros de composición celular del cerebro, resultado del análisis de PGLS simple de único predictor

Fenotipo	Familia de genes positivo	Familia de genes negativo
Neuronas en cerebro	37	2
Glía en cerebro	35	2
Neuronas en corteza	403	11
Glía en corteza	348	5
Masa corporal	39	2
Masa cerebral	38	3
Masa cortical	34	3
Densidad neuronal	0	0
Densidad neuronal en corteza	0	0
Masa cortical relativa	123	478
Encefalización	0	0

Nota: Tamaño de familias de genes con asociación significativa positiva ($p < 0.05$; $r > 0$), y con asociación significativa negativa ($p < 0.05$; $r < 0$), para cada fenotipo de composición celular en distintas regiones del cerebro, resultado de análisis de PGLS simple de predictor único con resultados independientes.

Tabla S2. Estructuras cerebrales clasificadas para el presente estudio

Grupo	Estructuras de corteza cerebral e hipocampo (utilizadas)	Estructuras de otras áreas del cerebro (descartadas)
<i>Fetus</i>	Dorsolateral prefrontal cortex Ventrolateral prefrontal cortex Anterior (rostral) cingulate (medial prefrontal) cortex Hippocampus (hippocampal formation) Primary visual cortex (striate cortex, area v1/17) Orbital frontal cortex Posterior (caudal) superior temporal cortex (area 22c) Posteroventral (inferior) parietal cortex Primary auditory cortex (core) Inferolateral temporal cortex (area tev, area 20) Primary motor cortex (area m1, area 4) Primary somatosensory cortex (area s1, areas 3,1,2) Occipital neocortex Parietal neocortex Temporal neocortex	Amygdaloid complex Striatum Cerebellar cortex Mediodorsal nucleus of thalamus Dorsal thalamus Primary motor-sensory cortex (samples) Cerebellum Caudal ganglionic eminence Lateral ganglionic eminence Medial ganglionic eminence Upper (rostral) rhombic lip
<i>Infant</i>	Posterior (caudal) superior temporal cortex (area 22c) Inferolateral temporal cortex (area tev, area 20) Dorsolateral prefrontal cortex Primary visual cortex (striate cortex, area v1/17) Ventrolateral prefrontal cortex Anterior (rostral) cingulate (medial prefrontal) cortex Posteroventral (inferior) parietal cortex Hippocampus (hippocampal formation) Orbital frontal cortex Primary auditory cortex (core) Primary motor cortex (area m1, area 4) Primary somatosensory cortex (area s1, areas 3,1,2)	Cerebellar cortex Amygdaloid complex Mediodorsal nucleus of thalamus Striatum

<i>Adult</i>	Hippocampus (hippocampal formation)	Cerebellar cortex
	Inferolateral temporal cortex (area tev, area 20)	Amygdaloid complex
	Orbital frontal cortex	Mediodorsal nucleus of thalamus
	Posterior (caudal) superior temporal cortex (area 22c)	Striatum
	Posteroventral (inferior) parietal cortex	
	Primary auditory cortex (core)	
	Primary motor cortex (area m1, area 4)	
	Primary somatosensory cortex (area s1, areas 3,1,2)	
	Ventrolateral prefrontal cortex	
	Anterior (rostral) cingulate (medial prefrontal) cortex	
	Dorsolateral prefrontal cortex	
	Primary visual cortex (striate cortex, area v1/17)	

4.2 Figuras suplementarias

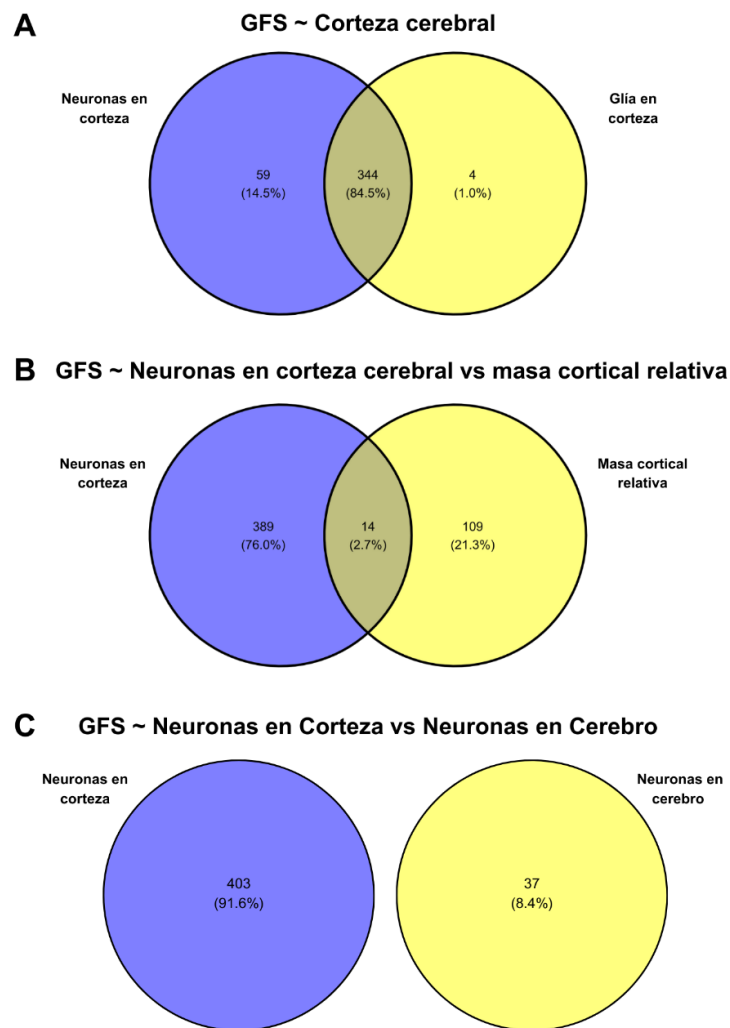


Figura 1S. Especificidad del tamaño de familia de genes (GFS), asociados con fenotipos de composición celular, a partir de resultados en análisis de PGLS simple de predictor único. Diagramas de Venn mostrando el porcentaje de especificidad del GFS, con asociación a fenotipos de composición celular en distintas regiones del cerebro. A) Corteza cerebral. B) Número de neuronas en corteza cerebral y masa cortical relativa. C) Número de neuronas en corteza cerebral y neuronas en cerebro completo.

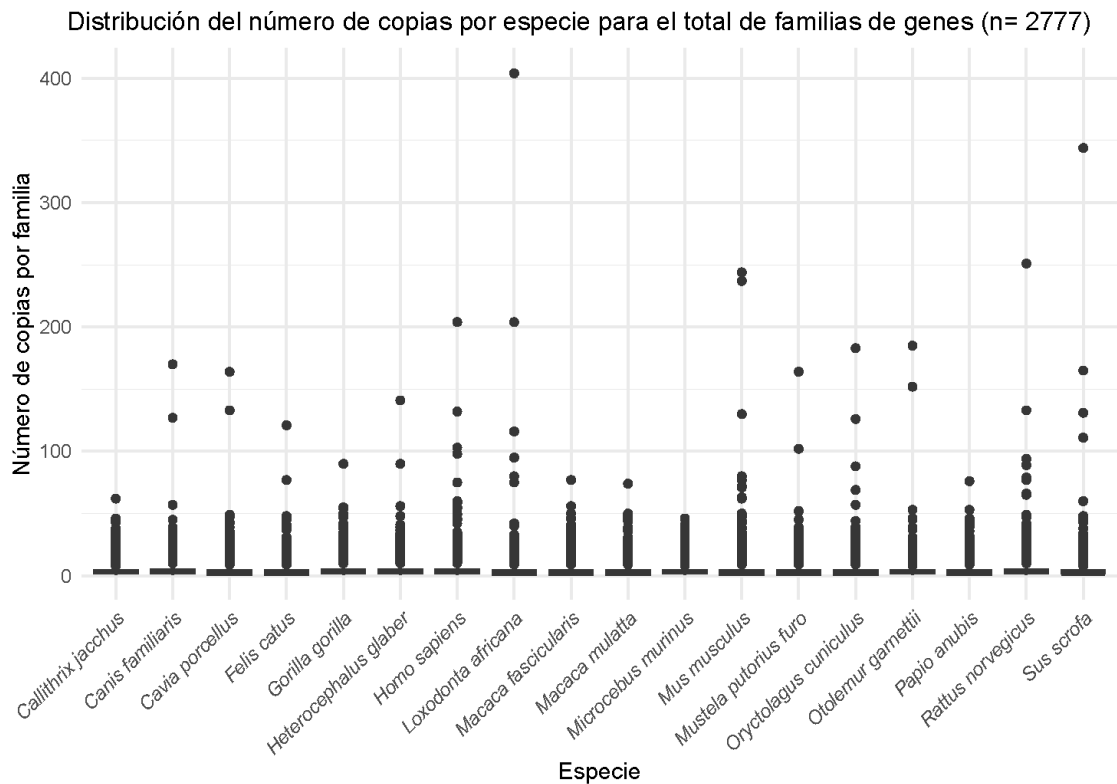


Figura 2S. Distribución del número de copias en familia de genes (n=2777) en cada especie. Boxplots que muestran la variación en el número de copias génicas por familia de genes en las especies analizadas.

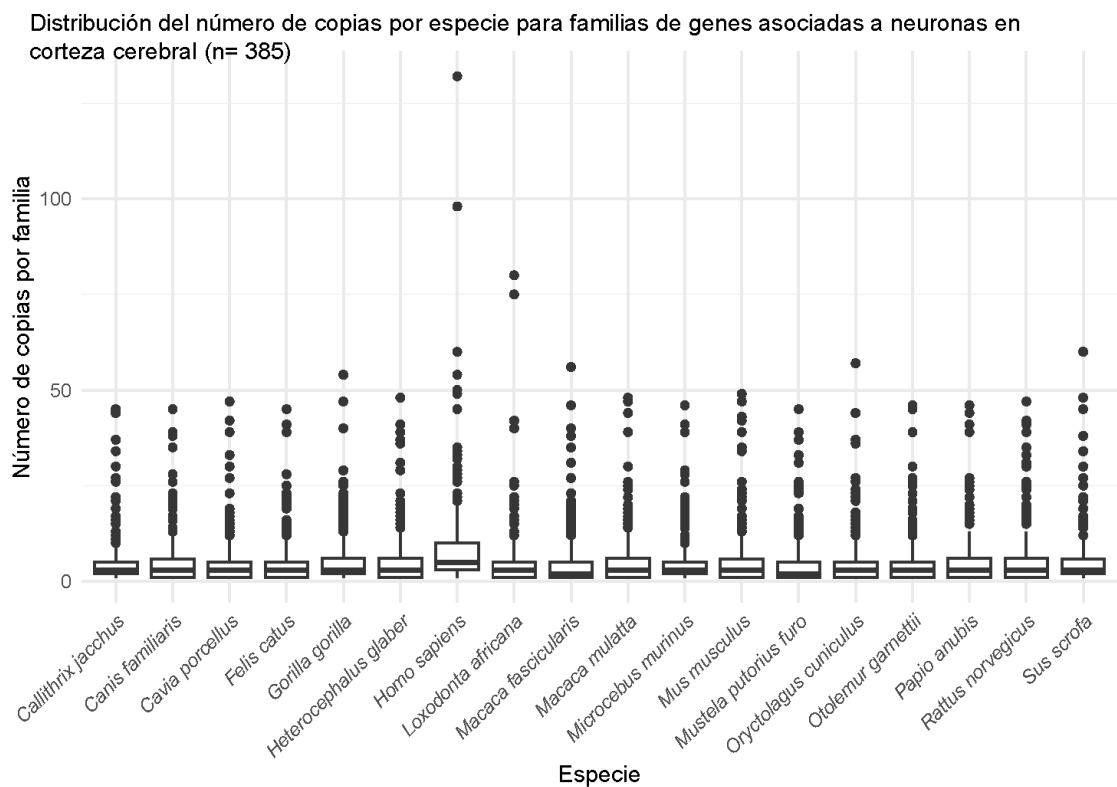


Figura 3S. Distribución del número de copias en familia de genes (n= 385) asociada al número de neuronas en corteza cerebral, en cada especie. Distribución del número de copias génicas por familia en las especies estudiadas para el subconjunto de familias asociadas a el número de neuronas en corteza cerebral.

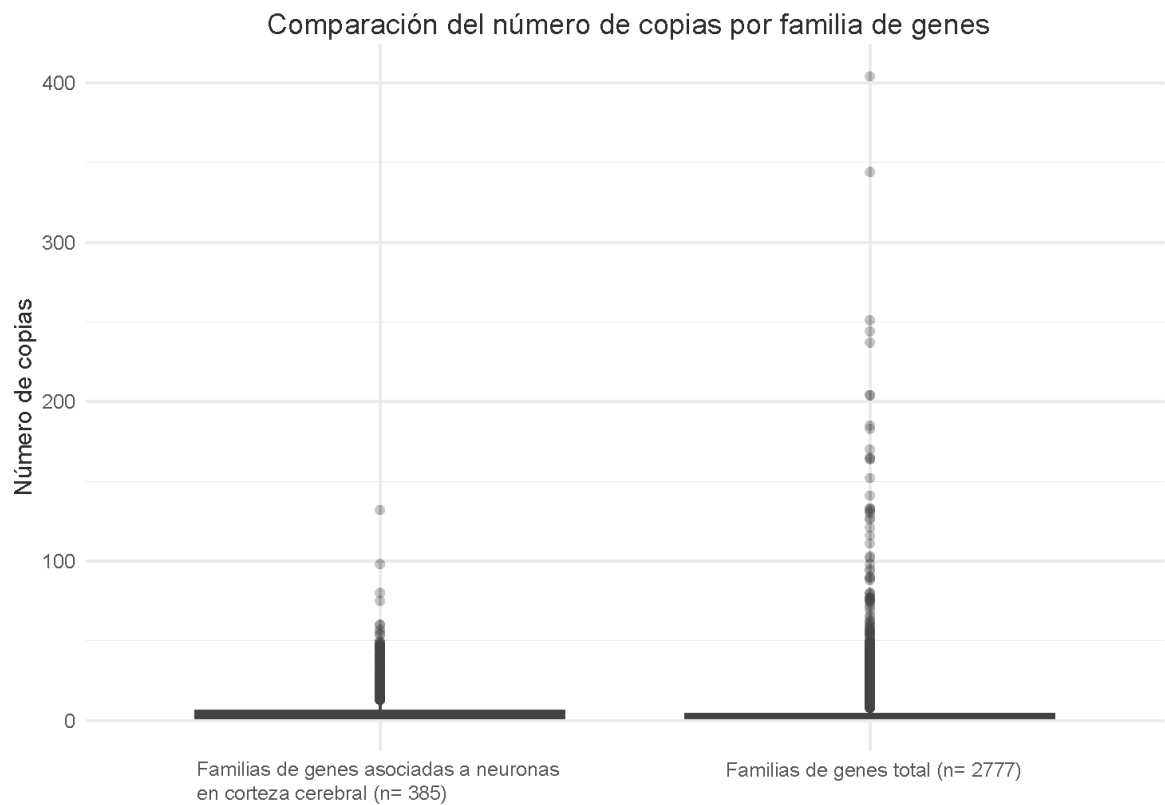


Figura 4S. Comparación del número de copias por familia de genes entre el conjunto total de familias de genes (n= 2777) y las familias asociadas al número de neuronas en corteza cerebral (n= 385). Comparación de las distribuciones del número de copias génicas por familia de genes. Las familias de genes asociadas al número de neuronas en corteza cerebral muestran una distribución comparable al fondo genómico, con ligeras diferencias que sugieren un posible papel de la duplicación génica en funciones de la expansión cortical.

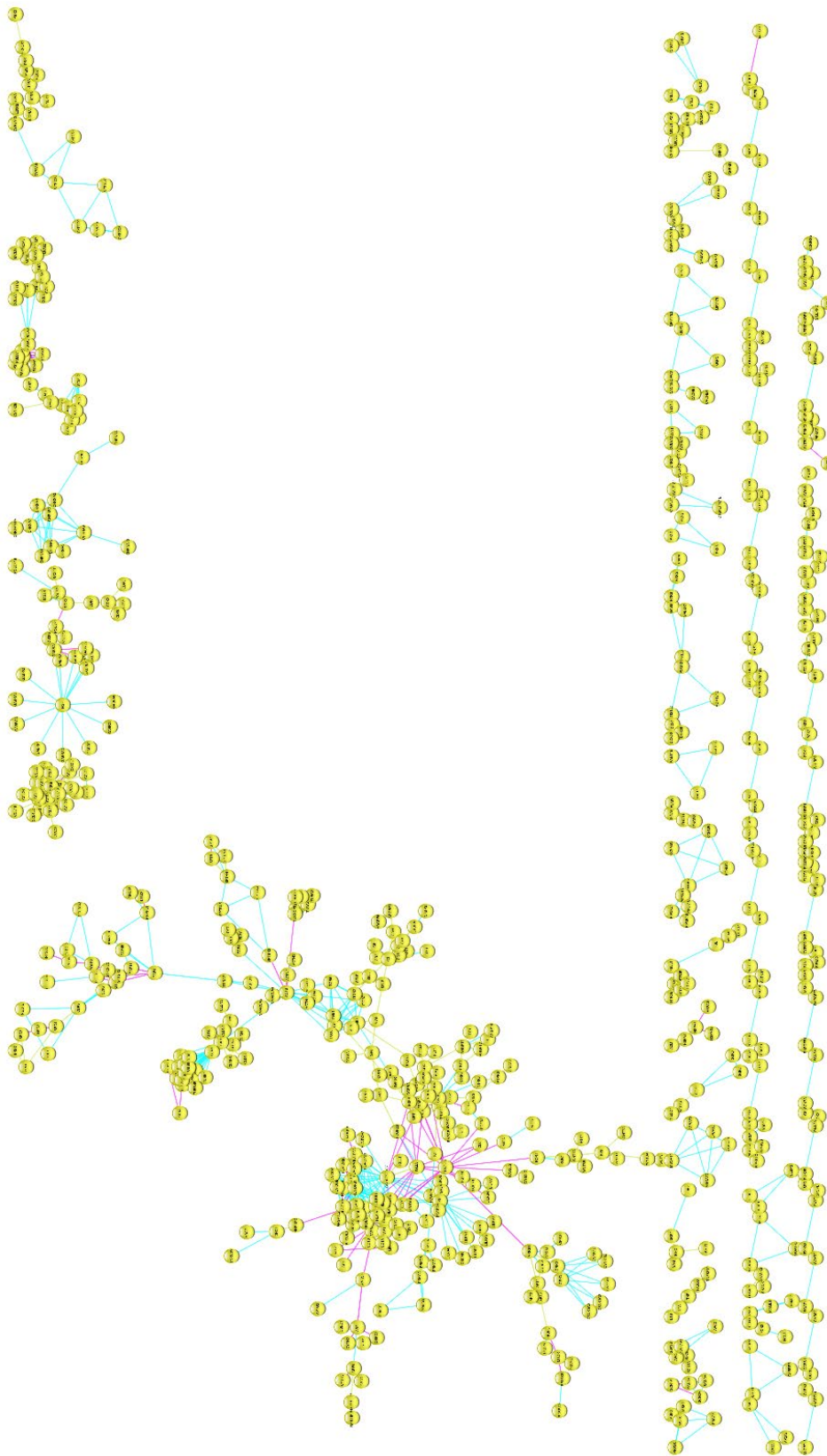


Figura 5S. Redes de interacción proteína-proteína señalando las proteínas de interés. Relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de consulta. Nodo rojo: proteínas predichas añadidas por *STRING*. Nodo gris: proteínas de segundo orden o de interacción indirecta.

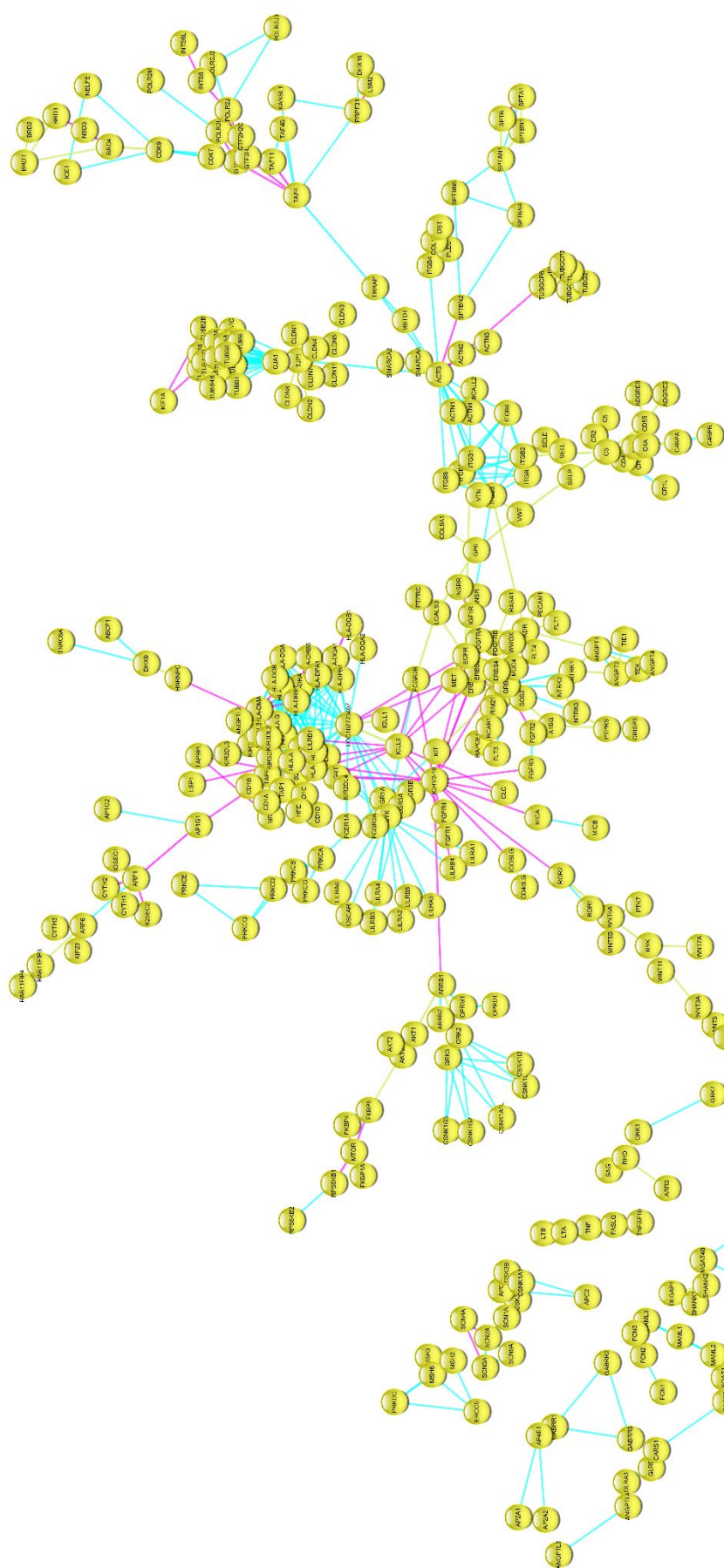


Figura 6S. Redes de interacción proteína-proteína modificado a escala (1/4) para una mejor observación de la interacción de proteínas de interés y seleccionada por región. Relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de consulta en región 1.

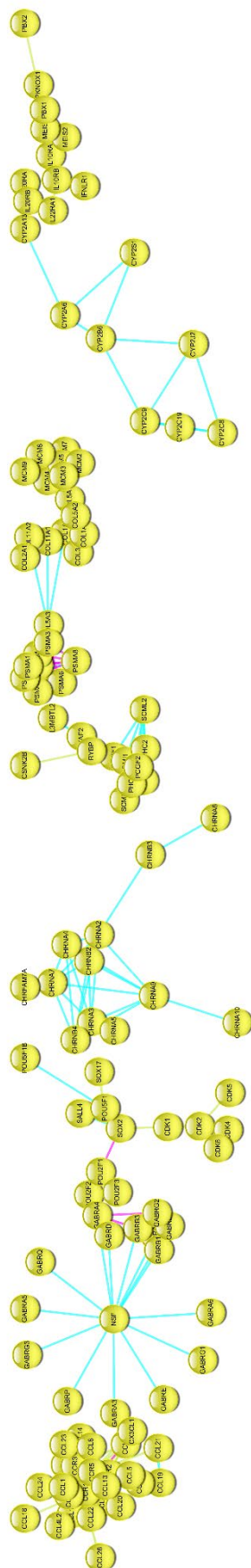


Figura 7S. Redes de interacción proteína-proteína modificado a escala (1/4) para una mejor observación de la interacción de proteínas de interés y seleccionada por región. Relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de consulta en región 2.

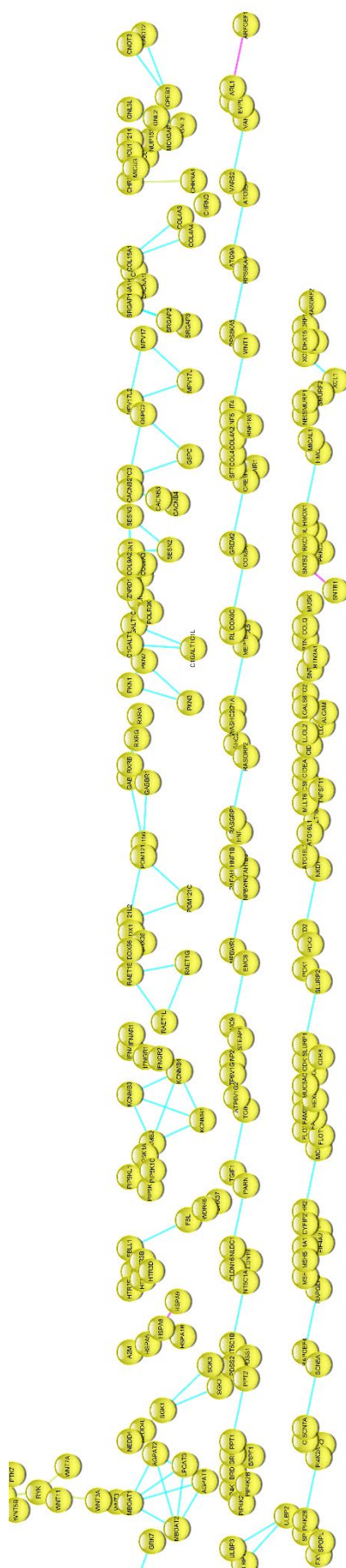


Figura 8S. Redes de interacción proteína-proteína modificado a escala (1/4) para una mejor observación de la interacción de proteínas de interés y seleccionada por región. Relación entre proteínas y vías enriquecidas en términos *GO*. Nodo amarillo: proteínas de consulta en región 3

14.3 Código

Script 1. PGLS doble de predictor único

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
# Load the necessary libraries
library(parallel)
library(nlme)
library(ape)
library(phytools)
library(caper)
library(dplyr)
library(logr)

cl <- makeCluster(detectCores()-1, type = "FORK")
#Read the input file (flat file with values separated by commas) with the phenotype data
traits<-read.csv("Mammalian_traits_Completo_Ctx_022023test.csv", stringsAsFactors = F)
# FILTERING
cat(green("Starting to get the LOG values and Filtering values: Longevity, Body Mass, Brain Mass and cEi..... \n"))
traits$log.longevity<-log10(traits$Maximum.longevity..yrs.)
### Change dir tree.mwk
tree <-read.tree("all_mammals_species_tree.nwk")
#Remove Genus.Species without phenotypic or sequencing data
traits <- traits[!is.na(traits$Cortical_neurons_x_10.6),]
traits <- traits[!is.na(traits$Neurons_brain_x_10.6),]
traits <- traits[!is.na(traits$Genus_species),]
row.names(traits)<-traits$Genus_species
#Read input file with the GFS per family per specie
list_orthogroups_counts<-read.csv("PGLS_46spp_4121GFS_MLSP_Ei_BUSCO_and_GeneFamily_Filtering.csv",row.names = 1)
list_orthogroups_counts<-list_orthogroups_counts[, colnames(list_orthogroups_counts)]
##### IN THE CASE OF NORMALIZED DATA USE THE gene_numbers...
traits<-subset(traits, traits$Genus_species %in% colnames(list_orthogroups_counts))
#Remove families with a zero variance in gene family size in the analysed Genus.Species
tree2<-drop.tip(tree,setdiff(tree$tip.label,row.names(traits)))
# All values should be TRUE
row.names(traits) %in% tree2$tip.label
cat(green("Filtering is DONE! \n"))
name.check(tree2, traits)
cat(red("Getting ready for PGLS \n"))
#####PGLS#####
Genome.completion.filtering<-function(FamData,Filt.Near.Uni, ssp.treshhold){
  want.percent<- Filt.Near.Uni/100
  NumCols1<-length(list_orthogroups_counts)
  # Filter the ortho counts by the single-copy orthologs. We are going to remove the orthogroups that have 0 or 1 number of genes per family
  list1<- list_orthogroups_counts[1:NumCols1] %>% filter_all(all_vars(<.2)) # all orthogroups have 0 or 1 gene
  list1$Total<- rowSums(list1) # summatory of all the sigle-copy genes in new col.
  ## Nearly universal genes calculation (filtering to get our core sigle-copy gene set)
  # NumCols1+1 (is the column with the total gene counts per orthogroup)
  # Why do I need to filter the orthogroups to keep the ones that have as many genes as the 90% or 95% of the number of species?
  filtering.90.1<-> filter(list1, Total >= round(NumCols1*want.percent))
  filtering.90.1<-> filter(list1, Total >= round(NumCols1*want.percent))
  #filtering.90.1$Total<- NULL
  # keep the spp that have 90% of these single-copy genes orthogroups.
  # Find the treshhold for the spp
  # All the spp most have at least the number of genes designated by the treshhold.
  treshhold <- round(nrow(filtering.90.1)*(ssp.treshhold/100))
  log_print(paste(round(nrow(filtering.90.1))))
  log_print(paste("All the spp most have at least the number of genes designated by the treshhold:", treshhold, "genes", sep = " ", hide_notes = T))
  # Count the number of single-copy genes per spp and remove those that does not have at least the number of genes dictated by the treshhold.
  filtering.90.1<-> filtering.90.1[,colSums(filtering.90.1) >= treshhold]
  filtering.90.1<-> filtering.90.1[,colSums(filtering.90.1) >= treshhold]
  log_print(paste("Which spp or columns were removed? ", setdiff(colnames(list_orthogroups_counts),colnames(filtering.90.1))), hide_notes = T)
  ### Jackpot! the final table that will be used to run the PGLS with all the spp that we need.
  gene_numbers<-list_orthogroups_counts[colnames(list_orthogroups_counts) %in% colnames(filtering.90.1)]
```

```

log_print(paste("You have", NumCols1, "spp that passed the genome completeness filter", sep = " "),hide_notes = T )
log_print("#### Jackpot! Now we have finished the filtering to know if the genomes used are complete ####", hide_notes = T)
log_print("The dataframe gene_numbers was created", hide_notes = T)
log_print("The filtering.90.1 object was created", hide_notes = T)
write.csv(filtering.90.1, paste("nearly.universal.",Filt.Near.Uni,".percent.",ssp.threshold,"ssp.threshold",".csv", sep = ""))
log_print(paste("File", "nearly.universal.",Filt.Near.Uni,".percent.",ssp.threshold,"ssp.threshold",".csv", " was created", sep = ""))
list.40_orthogroups_counts<-list_orthogroups_counts[colnames(list_orthogroups_counts) %in% colnames(filtering.90.1)]
sppnumero<-dim(list.40_orthogroups_counts)[2]
gene_numbers <- list.40_orthogroups_counts
write.csv(list.40_orthogroups_counts, paste("Total.counts.",sppnumero,"spp.csv", sep = ""))
log_print(paste("The file Total.counts.",sppnumero,"spp.csv was created", sep = ""))
}

### function to filter orthogroups with gene counts
Genome.completion.filtering(FamData = list_orthogroups_counts, Filt.Near.Uni = 90, ssp.threshold = 90)
colnames(x = traits)[1:8]
Trait.orthoGr.filtering<-function(traits, pheno.col.nums, spp.col.num, gene.numbers){
  phenotypes<- colnames(x = traits)[pheno.col.nums]
  for (i in 1:length(phenotypes)) {
    traits.filtered<-traits[!is.na(traits[,phenotypes[i]]),]
  }
  spp.name<-colnames(x = traits)[spp.col.num]
  traits.filtered<-traits.filtered[traits.filtered[,spp.name] %in% colnames(gene.numbers),]
  gene.numbers.filtered<-gene_numbers[,colnames(gene_numbers) %in% traits.filtered[,spp.name],]
  log_print("*****", hide_notes = T)
  log_print("****Spp missing in traits but present in gene.numbers:****", hide_notes = T)
  log_print("*****", hide_notes = T)
  log_print(setdiff(colnames(gene.numbers),traits.filtered[,spp.name]))
  log_print("*****", hide_notes = T)
  log_print("****Spp missing in gene.numbers but present in traits:****", hide_notes = T)
  log_print("*****", hide_notes = T)
  log_print(setdiff(traits$Spp.Name, traits.filtered$Spp.Name))
  if (dim(traits.filtered)[1] == 0) {
    log_print("The col. names from gene.number might be different to traits spp names", hide_notes = T)
  }
  log_print("*****", hide_notes = T)
  log_print("Removing the orthogroups (gene families) that have 0 genes in 20% of the spp", hide_notes = T)
  log_print("*****", hide_notes = T)
  zeros20perc <- ncol(gene.numbers.filtered)*0.2
  gene.numbers.filtered<-gene.numbers.filtered[rowSums(gene.numbers.filtered == 0) <= zeros20perc, ]
  log_print("*****", hide_notes = T)
  log_print("Removing any orthogroups (gene families) with variance of 0", hide_notes = T)
  log_print("*****", hide_notes = T)
  #[huk]Remove any fams with variance of zeros.
  gene.numbers.filtered<-gene.numbers.filtered[apply(gene.numbers.filtered, 1, var) > 0,]
  # to avoid gene families present in only one sp.
  log_print("*****", hide_notes = T)
  log_print("Removing orthogroups (gene families) of 1 gene or less present in only 1 spp", hide_notes = T)
  log_print("*****", hide_notes = T)
  gene.numbers.filtered<-gene.numbers.filtered[apply(gene.numbers.filtered, 1, max) > 2,]
  log_print(paste("The file ", paste("Filtered.GeneNum.",phenotypes,".csv", " was created", sep = "")))
  write.csv(gene.numbers.filtered, paste("Filtered.GeneNum.",phenotypes,".csv", sep = ""))
}

##### Change parameters!!!!!! #####
Trait.orthoGr.filtering(traits = traits, pheno.col.nums = c(20, 13), spp.col.num = 1, gene.numbers = gene_numbers)
gene_numbers <- gene.numbers.filtered
##### PGLS
out<-t(apply(gene_numbers,1,function(x){
  pheno1<<- "Cortical_neurons_x_10.6" #add phenotypes as needed
  pheno2<<- "Neurons_brain_x_10.6"
  traits2<-cbind(traits[,c("Genus_species", "Cortical_neurons_x_10.6", "Neurons_brain_x_10.6")],as.numeric(x)) #add phenotypes as needed
  names(traits2)[dim(traits2)[2]]<-"GFS"
  #next bit removes Genus.Species with NAs in events from tree, only necessary for losses, deletion, not for GFS
  traits2$Species <- row.names(traits2)
  tree2<- drop.tip(tree2,names(x)[is.na(x)])
  pglModel<-try(gls(GFS ~ Cortical_neurons_x_10.6 + Neurons_brain_x_10.6, correlation = corBrownian(1, form = ~Species, phy = tree2), method = "ML", data = traits2))
  #add phenotypes as needed

```

```

if (inherits(pglsModel, "try-error"))
  return(c(NA,NA))
else
  return(c(anova(pglsModel)$`F-value`[1], anova(pglsModel)$`p-value`[1], anova(pglsModel)$`F-value`[2], anova(pglsModel)$`p-value`[2],
    anova(pglsModel)$`F-value`[3], anova(pglsModel)$`p-value`[3], summary(pglsModel)$tTable[1,1],
    summary(pglsModel)$tTable[2,1], summary(pglsModel)$tTable[3,1], summary(pglsModel)$tTable[1,2], summary(pglsModel)$tTable[2,2],
    summary(pglsModel)$tTable[3,2], summary(pglsModel)$tTable[1,3], summary(pglsModel)$tTable[2,3], summary(pglsModel)$tTable[3,3],
    summary(pglsModel)$tTable[1,4], summary(pglsModel)$tTable[2,4], summary(pglsModel)$tTable[3,4]))
  )))
colnames(out)<-c("F_value Intercept", "p_value Intercept", paste("F_value", pheno1, sep=" "), paste("p_value", pheno1, sep=" "),
  paste("F_value", pheno2, sep=" "), paste("p_value", pheno2, sep=" "), "Coefficient Intercept", paste("Coefficient", pheno1, sep=" "),
  paste("Coefficient", pheno2, sep=" "), "Coef SE Intercept", paste("Coef SE", pheno1, sep=" "), paste("Coef SE", pheno2, sep=" "),
  "Coef tval Intercept", paste("Coef tval", pheno1, sep=" "), paste("Coef tval", pheno2, sep=" "), "Coef pval Intercept",
  paste("Coef pval", pheno1, sep=" "), paste("Coef pval", pheno2, sep=" "))
nameout <- paste("PGLS_", ncol(gene_numbers), "spp_", nrow(gene_numbers), "GFS_Longevity_New_filtering_", pheno1, pheno2, sep="")
cat(red("PGLS is DONE! \n PGLS folder has been created in the directory \n"))
##### Rs calculation from ts
sps<-nrow(traits)
DeFr <- sps - (2+1)
out<-as.data.frame(out)
Coef.tval1<-toString(paste("Coef tval", pheno1, sep=" "))
R1 <- toString(paste("R_t_value", pheno1, sep=" "))
R_t_value1 <- out[[Coef.tval1]] / (sqrt(((out[[Coef.tval1]]^2) + DeFr)) # tvalue / square root of(tvalue^2 + DF))
out[[R1]] <- R_t_value1
Coef.tval2<-toString(paste("Coef tval", pheno2, sep=" "))
R2 <- toString(paste("R_t_value", pheno2, sep=" "))
R_t_value2 <- out[[Coef.tval2]] / (sqrt(((out[[Coef.tval2]]^2) + DeFr)) # tvalue / square root of(tvalue^2 + DF))
out[[R2]] <- R_t_value2
write.csv(out, paste(nameout, "with_Rs", ".csv", sep=""))
cat(yellow("Rs Calculation is DONE! \n File created in the directory \n"))
##### BENJAMINI CALCULATION #####
bh <-read.csv(paste(nameout, "with_Rs", ".csv", sep=""))
bh$P.adjusted.GFS.vs.Cortical_neurons_x_10.6 <- p.adjust(bh$P_value.Cortical_neurons_x_10.6, method = "fdr")
bh$P.adjusted.GFS.vs.Neurons_brain_x_10.6 <- p.adjust(bh$P_value.Neurons_brain_x_10.6, method = "fdr")
write.csv(bh, paste(nameout, "with_Rs_Benja", ".csv", sep=""), row.names=FALSE)
cat(magenta("Benjamini Calculation is DONE! \n File created in the directory \n"))
#####PERMS#####
cat(red("Processing Permutation..... \n"))
clusterExport(cl,c("gene_numbers", "traits", "gls", "tree2", "corBrownian", "sps"))
perms<- parLapply(cl, 1, function(i,...){
  pheno1<<- "Cortical_neurons_x_10.6" #add phenotypes as needed
  pheno2<<- "Neurons_brain_x_10.6"
  y<-sample(c(1:sps))
  gene_numbers2 <- gene_numbers[y]
  outperm1<-apply(gene_numbers2,1,function(x){
    traits2<-cbind(traits[,c("Genus_species", "Cortical_neurons_x_10.6", "Neurons_brain_x_10.6")],x) #add phenotypes as needed
    names(traits2)[dim(traits2)[2]]<-"GFS"
    pglsModel<-gls(GFS ~ Cortical_neurons_x_10.6 + Cortical_glia_x_10.6, correlation = corBrownian(phy = tree2), method = "ML", data = traits2) #add phenotypes as needed
    return(c(anova(pglsModel)$`F-value`[1], anova(pglsModel)$`p-value`[1], anova(pglsModel)$`F-value`[2], anova(pglsModel)$`p-value`[2], anova(pglsModel)$`F-value`[3],
    anova(pglsModel)$`p-value`[3], summary(pglsModel)$tTable[1,1], summary(pglsModel)$tTable[2,1], summary(pglsModel)$tTable[3,1], summary(pglsModel)$tTable[1,2],
    summary(pglsModel)$tTable[2,2], summary(pglsModel)$tTable[3,2], summary(pglsModel)$tTable[1,3], summary(pglsModel)$tTable[2,3], summary(pglsModel)$tTable[3,3],
    summary(pglsModel)$tTable[1,4], summary(pglsModel)$tTable[2,4], summary(pglsModel)$tTable[3,4]))
  })
  outperm1<-t(outperm1)
  colnames(outperm1)<-c("F_value Intercept", "p_value Intercept", paste("F_value", pheno1, sep=" "), paste("p_value", pheno1, sep=" "), paste("F_value", pheno2, sep=" "),
  paste("p_value", pheno2, sep=" "), "Coefficient Intercept", paste("Coefficient", pheno1, sep=" "), paste("Coefficient", pheno2, sep=" "), "Coef SE Intercept", paste("Coef SE",
  pheno1, sep=" "), paste("Coef SE", pheno2, sep=" "), "Coef tval Intercept", paste("Coef tval", pheno1, sep=" "), paste("Coef tval", pheno2, sep=" "), "Coef pval Intercept",
  paste("Coef pval", pheno1, sep=" "), paste("Coef pval", pheno2, sep=" "))
  return(outperm1)
})
perms <- as.data.frame(do.call(rbind, perms), rownames=TRUE)
dim(perms)
pheno1<<- "Cortical_neurons_x_10.6"
nameout2 <- paste("OnePerm_New_Filtering", ncol(gene_numbers), "spp", nrow(gene_numbers), pheno1, sep=" ")
write.csv(perms, paste(nameout2, ".csv", sep=""))
##### Rs calculation from ts

```

```

infile<-read.csv(paste(nameout2,".csv", sep=""))
outfile<-paste(nameout2,"_con_Rs", ".csv", sep="")
R_t_value <- (((infile$Coef.tval.Cortical_neurons_x_10.6)/sqrt((((infile$Coef.tval.Cortical_neurons_x_10.6)^2) + DeFr )) # square root of (tvalue^2 / (tvalue^2 + DF))
infile$R_Coef.tval.Cortical_neurons_x_10.6 <- R_t_value
R_t_value <- (((infile$Coef.tval.Neurons_brain_x_10.6)/sqrt((((infile$Coef.tval.Cortical_glia_x_10.6)^2) + DeFr )) # square root of (tvalue^2 / (tvalue^2 + DF))
infile$R_Coef.tval.Cortical_glia_x_10.6 <- R_t_value
write.csv(infile, outfile, row.names=F)
#####THOUSAND PERMS#####
cat(red("Thousand permutation is processing ..... \n"))
clusterExport(cl,c("gene_numbers", "traits", "gls", "tree2", "corBrownian", "sps"))
perms<- parLapply(cl, 1:1000, function(i,...){
  pheno1<<- "Cortical_neurons_x_10.6" #add phenotypes as needed
  pheno2<<- "Neurons_brain_x_10.6"
  y<-sample(c(1:sps))
  gene_numbers2 <- gene_numbers[,y]
  outperm1<-apply(gene_numbers2,1,function(x){
    traits2<-cbind(traits[,c("Genus_species", "Cortical_neurons_x_10.6", "Neurons_brain_x_10.6")],x) #add phenotypes as needed
    names(traits2)[dim(traits2)[2]]<-"GFS"
    pglsModel<-gls(GFS ~ Cortical_neurons_x_10.6 + Neurons_brain_x_10.6, correlation = corBrownian(phy = tree2), method = "ML", data = traits2) #add phenotypes as needed
    return(c(anova(pglsModel)$'F-value'[1], anova(pglsModel)$'p-value'[1], anova(pglsModel)$'F-value'[2], anova(pglsModel)$'p-value'[2], anova(pglsModel)$'F-value'[3],
    anova(pglsModel)$'p-value'[3], summary(pglsModel)$tTable[1,1], summary(pglsModel)$tTable[2,1], summary(pglsModel)$tTable[3,1], summary(pglsModel)$tTable[1,2],
    summary(pglsModel)$tTable[2,2], summary(pglsModel)$tTable[3,2], summary(pglsModel)$tTable[1,3], summary(pglsModel)$tTable[2,3], summary(pglsModel)$tTable[3,3],
    summary(pglsModel)$tTable[1,4], summary(pglsModel)$tTable[2,4], summary(pglsModel)$tTable[3,4]))
  })
  outperm1<-t(outperm1)
  colnames(outperm1)<-c("F_value Intercept", "p_value Intercept", paste("F_value", pheno1, sep=" "), paste("p_value", pheno1, sep=" "), paste("F_value", pheno2, sep=" "),
  paste("p_value", pheno2, sep=" "), "Coefficient Intercept", paste("Coefficient", pheno1, sep=" "), paste("Coefficient", pheno2, sep=" "), "Coef SE Intercept", paste("Coef SE",
  pheno1, sep=" "), paste("Coef SE", pheno2, sep=" "), "Coef tval Intercept", paste("Coef tval", pheno1, sep=" "), paste("Coef tval", pheno2, sep=" "), "Coef pval Intercept",
  paste("Coef pval", pheno1, sep=" "), paste("Coef pval", pheno2, sep=" "))
  return(outperm1)
})
system("say Just finished!")
perms <- as.data.frame(do.call(rbind, perms), rownames=TRUE)
dim(perms)
pheno1<<- "Cortical_neurons_x_10.6"
nameout2 <- paste("OnePerm_New_Filtering", ncol(gene_numbers), "spp", nrow(gene_numbers), pheno1, sep="_")
write.csv(perms, paste(nameout2, ".csv", sep=""))
#####Rs calculation from ts#####
infile<-read.csv(paste(nameout2,".csv", sep=""))
outfile<-paste(nameout2,"_con_Rs", ".csv", sep="")
R_t_value <- (((infile$Coef.tval.Cortical_neurons_x_10.6)/sqrt((((infile$Coef.tval.Cortical_neurons_x_10.6)^2) + 56 )) # square root of (tvalue^2 / (tvalue^2 + DF))
infile$R_Coef.tval.Cortical_neurons_x_10.6 <- R_t_value
R_t_value <- (((infile$Coef.tval.Neurons_brain_x_10.6)/sqrt((((infile$Coef.tval.Neurons_brain_x_10.6)^2) + 56 )) # square root of (tvalue^2 / (tvalue^2 + DF))
infile$R_Coef.tval.Neurons_brain_x_10.6 <- R_t_value
write.csv(infile, outfile, row.names=F)
cat(red("Permutation is DONE! \n New file created in the directory"), yellow("\n ***** SCRIPT EXECUTED SUCCESFULLY !!! *****"))

```

Script 2. Valores de Significantes

```

#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
##### Select significant gene families #####
infile <- read.csv("PGLS.Cortical_neurons_x_10.6.18.Spp.2777.GeneFams.Rs_benjamini.csv", stringsAsFactors = F)
## Select significant p values with R Positive or R Negative##
pos <- infile[infile$adjusted.GFS.vs.Cortical_neurons_x_10.6 < 0.05 & infile$R.t.value.Cortical_neurons_x_10.6 > 0,]
neg <- infile[infile$adjusted.GFS.vs.Cortical_neurons_x_10.6 < 0.05 & infile$R.t.value.Cortical_neurons_x_10.6 < 0,]
# CHANGE phenotype on file name (example: "Cortical_neurons_x_10.6")##
write.csv(pos[,1], paste("Cortical_neurons_x_10.6.PGLS.18.Spp.", dim(pos)[1], ".POS.", "GeneFams.Rs_benjamini.csv", sep = ""), row.names=F)
write.csv(neg[,1], paste("Cortical_neurons_x_10.6.PGLS.18.Spp.", dim(neg)[1], ".NEG.", "GeneFams.Rs_benjamini.csv", sep = ""), row.names=F)

```

Script 3. Expansión y contracción de familia de genes

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")

# Load libraries
library(ggplot2)
library(dplyr)
library(tidyr)

# Read the data
expansion <- read.csv("expansion.csv")
contraccion <- read.csv("contraccion.csv")

# Merge the data into a single dataframe
datos <- full_join(expansion, contraccion, by = "Fenotipo")

# Replace NA with 0 if there are missing values
datos[is.na(datos)] <- 0

# Transform data into long format for ggplot2
datos_largo <- datos %>%
  pivot_longer(cols = c(Familias_Exp, Familias_Con),
    names_to = "Categoria", values_to = "Cantidad")

# Calculate percentage
datos_largo <- datos_largo %>%
  group_by(Fenotipo) %>%
  mutate(Percent = Cantidad / sum(Cantidad) * 100)

# Reorder the phenotype levels in the desired order
orden_fenotipos <- c("Neuronas_Corteza", "Glia en Corteza")

# Convert Phenotype into a factor with defined order
datos_largo$Fenotipo <- factor(datos_largo$Fenotipo, levels = orden_fenotipos)

# Rename categories for the legend
datos_largo$Categoria <- recode(datos_largo$Categoria,
  "Familias_Exp" = "Expansión",
  "Familias_Con" = "Contracción")

# Create the graph with horizontal bars
p <- ggplot(datos_largo, aes(x = Fenotipo, y = Percent, fill = Categoria)) +
  geom_bar(stat = "identity", position = "fill") + # Black borders for better separation
  scale_fill_manual(values = c("darkcyan", "darkgoldenrod1")) +
  #geom_hline(yintercept = c(0, 0.5, 1), linetype = "solid", color = "black", linewidth = 0.8) + # Line in 0%, 50% y 100%
  scale_y_continuous(labels = scales::percent_format()) +
  scale_x_discrete(limits = rev(orden_fenotipos)) +
  labs(title = "Número de Familias de Genes en Expansión y Contracción",
    x = "Fenotipo",
    y = "Porcentaje",
    fill = "Categoria") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip() + # ♦ Hace las barras horizontales
  annotate("text", x = Inf, y = 0.25, label = "Contracción de familia de genes", vjust = -1, size = 5, fontface = "bold") +
  annotate("text", x = Inf, y = 0.75, label = "Expansión de familias de genes", vjust = -1, size = 5, fontface = "bold")

# Save the chart as a PNG
ggsave("grafico_familias_genes2.png", plot = p, width = 10, height = 6, dpi = 300)
```

Script 4. Diagramas de Venn

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")

# Load the necessary libraries
library(readr)
library(ggplot2)
library(ggvenn)

# Upload CSV files
file1 <- "Cortical_neurons_x_10.6.PGLS.18.Spp.POS.GeneFams.Rs_benjamini.csv"
file2 <- "Cortical_glia_x_10.6.PGLS.18.Spp.POS.GeneFams.Rs_benjamini.csv"

# Reading CSV files
df1 <- read_csv(file1)
```

```
df2 <- read_csv(file2)
# Obtain the lists of IDs
ids1 <- df1$X
ids2 <- df2$X
# Create a list with the IDs
id_list <- list(Cortical_Neurons = ids1, Cortical_glia = ids2)
# Create the Venn diagram using ggvenn
gg <- ggvenn(id_list, columns = c("Cortical_Neurons", "Cortical_Glia"))
# Display the graph
print(gg)
# Save the chart in a PNG file
ggsave("Diagrama_venn_doble_cortical_neurons_glia_RCM.png", plot = gg, width = 6, height = 6, dpi = 1200)
```

Script 5. Enriquecimiento funcional en términos *GO*

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")

##### POSITIVES

library(parallel)

cl <- makeCluster(detectCores()-1)

filename1<-toString(paste("../02-PGLS/MLSP+Ei/PGLS_46spp_4121GFS_Longevity_New_filtering_clog.longevityEiwith_Rs_Benja")) #Change PGLS results *****
filenameout <- gsub("(?:[^\s]*){1}",'',filename1)#to exclude everything before the 1rd space and keep the name's part of interest to name the output files
filename2<-toString(paste(filename1, "POSITIVES.csv", sep=""))
infile<-read.csv(filename2,stringsAsFactors=FALSE)

head(infile)

phenotype<-toString(paste("MLSP+Ei"))#Put Phenotype of interest#####
rm(list=setdiff(ls(), c("infile", "GOanott", "phenotype", "filenameout"))))

#setwd("/Users/Administrator/Documents/PhD/Projecto/GliaN_Paper_Obs/Pearson/PIC")

library(parallel)

cl <- makeCluster(detectCores()-1)

filterbp<-toString(paste("benjamini BP50"))#Put filter and BP characteristics for name#####
outfile <- toString(paste("GO_", phenotype, "_", filterbp, sep=""))
outfile1 <- toString(paste(outfile, ".csv", sep=""))

numreps<-1000

mlsp_assoc <- infile[infile$P.adjusted.GFS.vs.clog.longevity < 0.05 & infile$R_t_value.clog.longevity > 0,]

head(mlsp_assoc)

dim(mlsp_assoc)

Familyset <- as.data.frame(infile[infile$X %in% mlsp_assoc$X, 1])

head(Familyset)

dim(Familyset)

names(Familyset)<- c("FamilyID")

dim(Familyset)

#Population of all gene families studied (Background)

dat<-as.data.frame(infile[,1])

names(dat)<- c("FamilyID")

GOanott<-read.csv("GO_Terms_per_family_Ens99BP_50.csv",header=TRUE,stringsAsFactors=FALSE) #Gene ontology annotations per family#####
GOsBPn<-unique(GOanott[,-1])

GOsBP<-GOsBPn[,1]

GOanott2<-matrix(NA,dim(GOanott)[1],2,dimnames = list(c(1:dim(GOanott)[1]),c("GO.ID", "FamilyID")))

GOanott2[,1]<-GOanott[,2]

GOanott2[,2]<-GOanott[,1]

counts<-vector("numeric")

annotation<-vector("numeric")

join<-merge(Familyset,GOanott2,by="FamilyID")

w <- dim(Familyset)[1]

zscore<-matrix(NA,length(GOsBP),1,dimnames=list(GOsBP,"Z-score"))

countssmpl2<-matrix(NA,length(GOsBP),numreps)

zscore2<-matrix(NA,length(GOsBP),1,dimnames=list(GOsBP,"Z-score"))

counts<-sapply(GOsBP, function(x) #Counts in Familyset per GO

  sum(join[,2]==x))

annotation<-apply(Familyset,1, function(x) #Annotation density of each family in Familyset

  sum(join[,1]==x))

counts2<-counts/mean(annotation) #Counts in Familyset per GO corrected against annotation density

clusterExport(cl, c("dat", "GOanott2", "GOsBP", "w"))
```

```

run3 <- function(c) {
  smpl <- as.data.frame(sample(dat[,1],w),stringsAsFactors=FALSE)
  names(smpl)<- c("FamilyID")
  joinsmpl<-merge(smpl,GOanott2,by="FamilyID")
  countssmpl<-sapply(GOsBP, function(x) #Cuentas por GO Term de genes en la smpl
    sum(joinsmpl[,2]==x))
  annotationsmpl<-apply(smpl,1, function(x) #Densidad de anotacion en reales
    sum(joinsmpl[,1]==x))
  return(list(countssmpl=countssmpl, annotationsmpl=annotationsmpl))
}
MCsmpl<-parSapply(cl,rep(1,numreps),run3)
stopCluster(cl)
countssmpl<-do.call(cbind,MCsmpl[1,])
annotationsmpl<-do.call(cbind,MCsmpl[2,])
means<-rowMeans(countssmpl)
meansannotation<-colMeans(annotationsmpl)
for(p in c(1:numreps))
  countssmpl2[,p]<-countssmpl[,p]/meansannotation[p]
means2<-rowMeans(countssmpl2)
SEMs<-apply(countssmpl,1,sd)
SEMs2<-apply(countssmpl2,1,sd)
zscore[,1]<-(counts-means)/SEMs
zscore2[,1]<-(counts2-means2)/SEMs2
pval<-pnorm(-abs(zscore))
adjpval<-p.adjust(pval,"fdr")
pval2<-pnorm(-abs(zscore2))
adjpval2<-p.adjust(pval2,"fdr")
numpval<-rowSums(countssmpl>=counts)/numreps
numpval2<-rowSums(countssmpl2>=counts2)/numreps
adjnumpval<-p.adjust(numpval,"fdr")
adjnumpval2<-p.adjust(numpval2,"fdr")
out1<-as.data.frame(cbind(GOsBPn[,1],GOsBPn[,2],counts,means,SEMs,zscore,pval, adjpval, numpval, adjnumpval),stringsAsFactors=FALSE)
names(out1)<-c("GOID","GO Term","Gene Families in the GO","Expected families per GO","Standard Deviation of the samples","Z-score","One tail p-value","FDR", "Numeric
p-value", "Numeric FDR")
#write.csv(out1, outfile, row.names=F)
write.csv(out1, outfile1, row.names=F)

```

Script 6. *HeatMap*

```

#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
rm(list=ls())
library(colorout)
nameout <- paste("All_Groups")####put name of file TodosM4top20
filterchar<-toString(paste("BP50_benja"))#Put filter and BP characteristics for name#####*****
outfilename <- toString(paste("GO-HM_", nameout, "_", filterchar, sep=""))
#####If one of the "archives" is modified before, then quit manually single or double quotes ( ' or " ) from the ID names so it doesn't complain when reading the
table.#####
##fill the ones to omit with NAs###
archivos <- list.files(pattern = "GO_")
noms<-sapply(strsplit(archivos,"_"),function(x) length(x)) # BP 4, # MF and CC 5. La parte del nombre cambiante/significativa
tb1<-read.table(archivos[1],sep=" ",header=TRUE)
ps<-matrix(NA,dim(tb1)[1],length(archivos),dimnames=list(tb1$GO.Term,noms))
zs<-matrix(NA,dim(tb1)[1],length(archivos),dimnames=list(tb1$GO.Term,noms))
obs<-matrix(NA,dim(tb1)[1],length(archivos),dimnames=list(tb1$GO.Term,noms))
for (i in c(1:length(archivos))) {
  tb1<-read.table(archivos[i],sep=" ",header=TRUE)
  ps[,i]<-tb1$FDR
  zs[,i]<-tb1$Z.score
  obs[,i]<-tb1$Gene.Families.in.the.GO
}
GOn<-apply(zs,2,function(x) length(which(x>0))) #valores de zscore positivos (por lo tanto solo categorias overrepresented), tanto significativos como no
GOn2<-apply(ps,1,function(x) length(which(x<0.05)))
ind <- which(GOn2==0)

```

```

newnames<-paste(colnames(ps),GOn)
colnames(ps)<-newnames
ps[which(ps>0.05)]<-NA
ps[which(zs<0)]<-NA
ps[which(obs<3)]<-NA #filter that more than 1 Gene Family is observed
ps <- ps[-ind, ]
ps <- ps[order(-ps[,2], -ps[,1]), ] #para ordenar (ascending) primero con el p value de la 1a columna, luego por el de la 2a y finalmente el de la 3a
ps <- ps[apply(ps,1,function(x)any(!is.na(x))),] #TRUE para las filas que no son todas NA
##Save "ps" and manually change the order as desired for the graphic
## First invert the order to see it as the final output graphic
ps<-ps[seq(dim(ps)[1],1),]
tablename<-toString(paste(outfilename, "_Table", sep=""))
write.csv(ps, paste(tablename, ".csv", sep=""))
###CHANGE the order as desired for the graphic
ps<-read.csv(toString(paste(tablename, ".csv", sep="")))
ps<-read.csv("GO-HM_All_Groups_BP50_benja_Table.csv")
ps<-ps[seq(dim(ps)[1],1),] #re-invert to the original
ps2<-ps[,-1]
rownames(ps2) <- ps[,1]
ps<-ps2
# Actual Heatmap plot #
library(reshape)
library(lattice)
library(RColorBrewer)
colores<-c(rev(brewer.pal(9,"YlOrRd")[2:9]))
paleta<-colorRampPalette(colores, space = "Lab")
colores<-paleta(50)
interv<-seq(0,0.05,by=.001)
pdffname <- toString(paste("heatmap_paper.pdf", sep=""))
pdf(pdffname, paper="a4r",width=8, height=8) # BP 11,6 CC 8,8
levelplot(t(ps), main="", xlab="", ylab="", col.regions=colores, cuts=100, at=interv, scales=list(x=list(rot=90,cex=.7), y=list(rot=0,cex=0.7), draw=TRUE) ,
colorkey=list(TRUE,space="right",contour=FALSE,pretty=TRUE), par.settings = list(axis.line = list(col = 0)),
  panel = function(...) {
    panel.fill(col = "gray96")
    panel.levelplot(...)
  })
dev.off()
png("heatmap_for_paper.png",units="px", width=5760, height=5760, res=600)
levelplot(t(ps), main="", xlab="", ylab="", col.regions=colores, cuts=100, at=interv, scales=list(x=list(rot=90,cex=1.5), y=list(rot=0,cex=1.5), draw=TRUE) ,
colorkey=list(labels=list(cex=1.25, TRUE,space="right",contour=FALSE,pretty=TRUE)), par.settings = list(axis.line = list(col = 0)),
  panel = function(...) {
    panel.fill(col = "gray96")
    panel.levelplot(...)
  })
dev.off()
library(viridis)
colvir <- viridis(100)
pdffname <- toString(paste("heatmap_paper_choosed_groups_LAST.pdf", sep=""))
pdf(pdffname, paper="a4r",width=8, height=8) # BP 11,6 CC 8,8
levelplot(t(ps), main="", xlab="", ylab="", col.regions=plasma(150), cuts=100, at=interv, scales=list(x=list(rot=90,cex=.7), y=list(rot=0,cex=0.7), draw=TRUE) ,
colorkey=list(TRUE,space="right",contour=FALSE,pretty=TRUE), par.settings = list(axis.line = list(col = 0)),
  panel = function(...) {
    panel.fill(col = "gray96")
    panel.levelplot(...)
    #panel.text(x,y, a, cex=.6)
  })
dev.off()
png("heatmap_for_paper_choosed_groups_LAST.png",units="px", width=3000, height=3000, res=350)
levelplot(t(ps), main="", xlab="", ylab="", col.regions=plasma(150), cuts=100, at=interv, scales=list(x=list(rot=90,cex=1), y=list(rot=0,cex=1), draw=TRUE) ,
colorkey=list(TRUE,space="right",contour=FALSE,pretty=TRUE), par.settings = list(axis.line = list(col = 0)),
  panel = function(...) {
    panel.fill(col = "gray96")
    panel.levelplot(...)
    #panel.text(x,y, a, cex=.6)
  })
dev.off()

```

Script 7. Match Familia de genes - Genes

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
# Match ID's Panther-Ensembl
####Database filtering of Panther_ID family IDs###
# Remove suffixes from Panther ID
setwd("/home/james/Escritorio/match") #Direccionar carpeta de trabajo #Address work folder
# Read CSV file
datos <- read.csv("PGLS_18spp_2777GFS_Longevity_New_filtering_Cortical_neurons_x_10.6Cortical_glia_x_10.6with_Rs_Benja.csv") #Cambio de
archivo el cual se filtrara
# Check the structure of the data
str(datos)
# Remove '_SF' suffixes from ID_Panther column
datos$PANTHER_ID <- gsub("_SF.*", "", datos$PANTHER_ID)
# Check for updated data
head(datos)
#### Match for IDs ###
# Loading the library
library(dplyr)
# Remove duplicate rows
datosfilter <- datos[!duplicated(datos), ]
datosfilter <- as.data.frame(datosfilter)
colnames(datosfilter) <- "PANTHER_ID"
# Read the databases
base1 <- read.csv("Biomart290224.csv")
base2 <- datosfilter
# Perform the join between the databases using column 'A'.
base_final2 <- merge(base1, base2, by = "PANTHER_ID")
# Delete of duplicate rows and generated column
datosfilter2 <- base_final2[!duplicated(base_final2), ]
# Save the result
write.csv(datosfilter2, "/home/james/Escritorio/match/PGLS_complete_Neuronscortical_and_Gliacortical_ENS.csv", row.names = FALSE)
```

Script 8. Expresión de genes (tejidos o región cerebral)

```
#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
rm(list=ls())
library(logr)
library(ggplot2)
library(dplyr)
## positive or negative
testtype<- "positive"
## brain, f_brain or many others
tissueOFinterest <- "Cortex"
## adult, fetus, infant
mono<- "adult"
### 1 up to how much time you want to lose. e.g.1000
perms= 1000
tmp <- file.path(getwd(), paste("RankingAnalysis.", mono, ".", testtype, ".associated.genes.", tissueOFinterest, ".tissue.log", sep = ""))
lf <- log_open(tmp)
if (testtype == "negative") {
  log_print("Captain's log starts...")
  log_print("Using negatively selected genes", hide_notes = T)
  fantomData <- read.csv("z_score_brainspan_data1.csv") #Database phantom or brainspan
  if (mono == "adult") {
    fantom <- fantomData
```

```

log_print("Using Adult selected genes", hide_notes = T)
} else if (mono == "fetus") {
  feto <- grep(x = colnames(fantomData), pattern = "f_")
  fantom <- subset(fantomData, select = c(1, feto))
  log_print("Using fetal selected genes", hide_notes = T)
} else if (mono == "infant") {
  infant <- grep(x = colnames(fantomData), pattern = "i_")
  fantom <- subset(fantomData, select = c(1, infant))
  log_print("Using infant selected genes", hide_notes = T)
} else {
  # Manejar otros valores de `mono`
  stop("Invalid value for mono")
}

pgls_background <- read.csv("PGLS_complete_Neuronscerebellar_and_Gliacerebellar_ENS.csv") #PGLS complete
pgls_assoc <- read.csv("PGLS_significant_NEG_Neuronscerebellar.csv") #PGLS significant negative
pgls_notAssoc <- read.csv("PGLS_nonsignificant_NEG_Neuronscerebellar.csv") #PGLS nonsignificant negative
} else if (testtype == "positive") {
  log_print("Captain's log starts...")
  log_print("Using positively selected genes", hide_notes = T)
  fantomData <- read.csv("z_score_brainspan_data1.csv") #Database phantom or brainspan
  if (mono == "adult") {
    fantom <- fantomData
    log_print("Using Adult selected genes", hide_notes = T)
  } else if (mono == "fetus") {
    feto <- grep(x = colnames(fantomData), pattern = "f_")
    fantom <- subset(fantomData, select = c(1, feto))
    log_print("Using fetal selected genes", hide_notes = T)
  } else if (mono == "infant") {
    infant <- grep(x = colnames(fantomData), pattern = "i_")
    fantom <- subset(fantomData, select = c(1, infant))
    log_print("Using infant selected genes", hide_notes = T)
  } else {
    # Manejar otros valores de `mono`
    stop("Invalid value for mono")
  }

  pgls_background <- read.csv("PGLS_complete_Neuronscerebellar_and_Gliacerebellar_ENS.csv") #PGLS complete
  pgls_assoc <- read.csv("PGLS_significant_POS_Neuronscerebellar.csv") #PGLS significant negative
  pgls_notAssoc <- read.csv("PGLS_nonsignificant_POS_Neuronscerebellar.csv") #PGLS nonsignificant negative
}

fantom <- read.csv("z_score_brainspan_data1.csv") #Database phantom or brainspan
number_of_genes <- dim(pgls_assoc)[1] ##Change this
#Filter FANTOM data to PGLS background genes
fantom_background <- fantom[fantom$Gene_stable_ID %in% pgls_background$Gene_stable_ID,] #Gene.stable.ID or Gene_stable_ID
fantom_background <- fantom_background[!duplicated(fantom_background$Gene_stable_ID),]
colnames(fantom_background)
row.names(fantom_background) <- fantom_background$Gene_stable_ID
fantom_background$Gene_stable_ID <- NULL
names <- row.names(fantom_background)
fantom_background$total <- (rowSums(fantom_background, na.rm=FALSE))
fantom_background$total <- rowSums(select_if(fantom_background, is.numeric), na.rm = TRUE)
fantom_background <- fantom_background[!fantom_background$total == 0,]
fantom_background$total <- NULL

### creating the index column tio extract the genes from the back ground and use them in the analysis
pgls_assoc$Gene_stable_ID <- make.unique(pgls_assoc$Gene_stable_ID)
rownames(pgls_assoc) <- pgls_assoc$Gene_stable_ID
NEGgenes <- fantom_background[(rownames(fantom_background)) %in% rownames(pgls_assoc),]
number_of_genes <- dim(NEGgenes)[1]

#### permutations
fa1R <- data.frame(rownames(NEGgenes))
fa1RAvg <- data.frame()
BGlistR <- list()
for (i in 1:dim(fantom_background)[1]) {
  genesR <- t(fantom_background[i,])
  GenenameR <- colnames(genesR)
  genesR[tissueOfInterest,]
  brainrank <- data.frame(sum(genesR[,1] < genesR[tissueOfInterest,]))

```

```

rownames(brainrank)<- tissueOFinterest
colnames(brainrank)<- GenenameR
BGlistR[[i]]<- brainrank
if (i == dim(fantom_background)[1]) {
  genesR<- do.call("cbind", BGlistR)
  genesRALLgenesvsBrain<- t(genesR)
  genesR<-rowMeans(genesR)
  # log_print(cat("#####\n#### average mean perms ###\n#####"))
  log_print("#####", hide_notes = T)
  log_print("##### average mean perms ###", hide_notes = T)
  log_print("#####", hide_notes = T)
  log_print("Done creating the Background means .___.", hide_notes = T)
  log_print("Object genesRALLgenesvsBrain and genesR were created", hide_notes = T)
  log_print(paste("The average rank for ", tissueOFinterest," is: ", genesR, sep = ""), hide_notes = T)
  for (j in 1:perms) {
    Cerebro<-paste(tissueOFinterest,j, sep = "")
    die<-genesRALLgenesvsBrain[,tissueOFinterest]
    fa1R[,Cerebro]<- sample(die, size = number_of_genes, replace = TRUE)
    if (j == perms) {
      rownames(fa1R)<- fa1R[,1]
      fa1R[,1]<- NULL
      fa1RAvg<-data.frame(t(fa1R))
      fa1RAvg$totalBrainAVG<-rowMeans(fa1RAvg)
      fa1RAvg<-select(fa1RAvg, contains("totalBrainAVG"))
      log_print("Object fa1Ang and fa1R were created", hide_notes = F)
      # print("Object fa1Ang and fa1R were created", quote=F)
    }
  }
}
}
}
}
#### real means
BGlist<- list()
for (i in 1:dim(NEGgenes)[1]) {
  NEGgenesR<- t(NEGgenes[i,])
  Genename<- colnames(NEGgenesR)
  NEGgenesR[tissueOFinterest,]
  brainrank <- data.frame(sum(NEGgenesR[,1] < NEGgenesR[tissueOFinterest,]))
  rownames(brainrank)<- tissueOFinterest
  colnames(brainrank)<- Genename
  BGlist[[i]]<- brainrank
  if (i == dim(NEGgenes)[1]) {
    NEGgenesR<- do.call("cbind", BGlist)
    NEGgenesALLgenesvsBrain<- t(NEGgenesR)
    NEGgenesR<-rowMeans(NEGgenesR)
    log_print("#####", hide_notes = T)
    log_print("##### real means for focal tissue ####", hide_notes = T)
    log_print("#####", hide_notes = T)
    log_print("Done creating the Real means .___.", hide_notes = T)
    log_print("Object NEGgenesALLgenesvsBrain and NEGgenesR were created", hide_notes = T)
    log_print(paste("The average rank for ", tissueOFinterest," is: ", NEGgenesR, sep = ""), hide_notes = T)
  }
}
}
min(fa1RAvg)
max(fa1RAvg)
NEGgenesR
log_print(paste(tissueOFinterest," tissue has an average rank bigger than random rank in :",
  data.frame(sum(NEGgenesR > fa1RAvg)), "/", dim(fa1RAvg)[1], sep = ""), hide_notes = T)
log_print(paste("Plot name: RankingAnalysis.", mono, ".", testtype, ".associated.genesfamily.", tissueOFinterest, ".tissue.pdf", sep = ""))
log_close()
writeLines(readLines(lf))
#### creating flot
densidad<-density(fa1RAvg$totalBrainAVG)
pdf(paste("RankingAnalysis.", mono, ".", testtype, ".associated.genesfamily.", tissueOFinterest, ".tissue.pdf", sep = ""))
# Filled Density Plot
p <- ggplot(fa1RAvg, aes(x=totalBrainAVG)) +

```

```

geom_density() +
labs(title = "Cerebellar Neurons in adult")+ # Agregar el título con fenotipo y con grupo
theme(plot.title = element_text(hjust = 0.5)) #posicion del titulo
# Add mean line
p + scale_color_grey() + theme_classic() +
  geom_vline(aes(xintercept=NEGgenesR), color="red", linetype="dashed", linewidth=1) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, max(densidad$y) + 0.01)) +
  scale_x_continuous(expand = c(0, 0), limits = c(min(fa1RAvg$totalBrainAVG) - 18.0, max(fa1RAvg$totalBrainAVG) + 18.0 ))
dev.off()

```

Script 9. Expresión de genes espacio-temporal

```

#Set working directory to the one where we have the input files
setwd("/datoslab/usuarios/studContainer/jtr23/PGLS2/")
library(dplyr)
whereToSave<- getwd()
dir.create(file.path(whereToSave,'cellulargliacortical'))
setwd(file.path(whereToSave,'cellulargliacortical'))
rm(list=ls())
##Scatterplot for BrainSpan data. You can find the code to plot the mean expression or the Z-scores.
#Df.edades
load("/home/james/Documentos/Carpetasintitulo/temporal/New/Relacion_Id_Edades(years).RData")
##### Zscore por etapa en las 4 listas distintas #####
print("#####")
print("##### Data in #####")
print("#####")
###Change the phenotype to be studied and add whether "positive" or "negative" data are used###
plotname<- "PGLS_nonsignificant_POS_Cortical_gliila"
print("Creating table with data of the genes associated to cell count in cortex", quote =F) #para nosotros son los genes asociados a componente celular
#datos con base de datos modificados con Z-score
DataBS<- read.csv("/home/james/Documentos/Carpetasintitulo/temporal/New/z_score_brainspan_data.csv") # base2
print(paste("you have ", dim(DataBS)[1], " genes with brain span data z-score"), quote = F)
#datos con base de datos real
DataBS_r<- read.csv("/home/james/Documentos/Carpetasintitulo/temporal/New/Brainspan_RNA-Seq_genes_Developmental_Transcriptome_22042024_ALL.csv") # base2
print(paste("you have ", dim(DataBS_r)[1], " genes with brain span data real"), quote = F)
AGenes<- read.csv("/home/james/Documentos/Carpetasintitulo/temporal/New/PGLS_nonsignificant_POS_Cortical_gliila.csv")
print(paste("Your AGenes object looks like this:"), quote = F)
print( head(AGenes, 3))
##Function to obtain the Z-scores
FunZscore<-function(GENES, Rvalscol, PERMSnum){
  ##erase the id from the genes
  print("Reading variables...")
  GENES<- GENES
  Rvalscol <- Rvalscol
  TablaLong<-DataBS[DataBS[,1] %in% GENES[,Rvalscol],-(1:2)]
  dim<-dim(TablaLong)[1]
  print("Filtering brain life span data vs your PGLS or analysis selected genes with r values")
  print(dim)
  print("calculating mean of the data selected")
  meanLong<-apply(TablaLong, 2, mean)
  print(head(meanLong))
  meanRandom<-rep()
  print("Starting randomization", quote = F)
  print("Sampling the genes from DataBS by the number of your focal genes", quote =F)
  for (i in 1:PERMSnum){ ##change number
    #sample the genes and keep only the numeric info
    x<-DataBS[DataBS[,1] %in% (sample(DataBS[,1],replace = F,size=dim)), -(1:2)]
    x<-apply(x,2,mean)
    meanRandom<-rbind(meanRandom,x)
  }#for
  print("calculating the stats")
  meanRandom<-meanRandom[-1,]
  MeanOfMeans<-apply(meanRandom, 2, mean)
  print("calc mean of random values")
}

```

```

print(head(MeanOfMeans))
sdRandom<-apply(meanRandom, 2, sd)
print("SD or random values")
print(head(sdRandom))
zscore<-(meanLong-MeanOfMeans)/sdRandom
print("calc. z scores")
print(head(zscore))
pval<-pnorm(-abs(zscore))
print("calc. p vals")
print(head(pval))
temp<-rbind(meanLong,MeanOfMeans,sdRandom,zscore,pval)
print("Final table")
print(temp)
return(temp)
}

#base de datos real
FunZscore_r<-function(GENES, Rvalscol, PERMSnum){
  ##erase the id from the genes
  print("Reading variables...")
  GENES<- GENES
  Rvalscol <- Rvalscol
  TablaLong<-DataBS_r[DataBS_r[,1] %in% GENES[,Rvalscol],-(1:2)]
  dim<-dim(TablaLong)[1]
  print("Filtering brain life span data vs your PGLS or analysis selected genes with r values")
  print(dim)
  print("calculating mean of the data selected")
  meanLong<-apply(TablaLong, 2, mean)
  print(head(meanLong))
  meanRandom<-rep()
  print("Starting randomization", quote = F)
  print("Sampling the genes from DataBS_r by the number of your focal genes", quote =F)
  for (i in 1:PERMSnum){ ##change number
    #sample the genes and keep only the numeric info
    x<-DataBS_r[DataBS_r[,1] %in% (sample(DataBS_r[,1],replace = F,size=dim)), -(1:2)]
    x<-apply(x,2,mean)
    meanRandom<-rbind(meanRandom,x)
  }#for
  print("calculating the stats")
  meanRandom<-meanRandom[-1,]
  MeanOfMeans<-apply(meanRandom, 2, mean)
  print("calc mean of random values")
  print(head(MeanOfMeans))
  sdRandom<-apply(meanRandom, 2, sd)
  print("SD or random values")
  print(head(sdRandom))
  zscore<-(meanLong-MeanOfMeans)/sdRandom
  print("calc. z scores")
  print(head(zscore))
  pval<-pnorm(-abs(zscore))
  print("calc. p vals")
  print(head(pval))
  temp<-rbind(meanLong,MeanOfMeans,sdRandom,zscore,pval)
  print("Final table")
  print(temp)
  return(temp)
}

#Run the function for every list of genes
Castillo<-FunZscore(GENES = AGenes, Rvalscol = "Gene_stable_ID", PERMSnum = 10000) #base de datos z-score modificado
Castillo_r<-FunZscore_r(GENES = AGenes, Rvalscol = "Gene_stable_ID", PERMSnum = 10000) #base de datos real
write.csv(Castillo, paste("STATS.VALUES.",plotname, ".csv", sep = ""))
write.csv(AGenes, paste("CellExp.GENES.", plotname, ".csv", sep = ""))

##### PLOT #####

# Función para calcular R cuadrada
calcular_r2 <- function(x, y) {
  modelo <- lm(y ~ log10(x))
  summary(modelo)$r.squared

```

```

}

funPlot <- function(X, Y, COLOR, YLAB, MAIN) {
  # Calcular R2
  r2 <- calcular_r2(X, Y)
  # Calcular el rango completo de valores de Y
  range_Y <- diff(range(Y))
  # Calcular el paso para la secuencia de valores del eje y
  by_y <- range_Y / 5 # Ajusta el número de marcas según sea necesario
  # Crear la secuencia de valores del eje y
  seq_y <- seq(min(Y) - abs(min(Y) * 0.2), max(Y) * 1.0, by = by_y)
  seq_y_formatted <- sprintf("%.2f", seq_y)
  # Graficar
  plot(x = log10(X), y = Y, type = "p", pch = 21, cex = 1.2, col = "black",
       main = paste(MAIN, "\nR2 = ", round(r2, 3)), bg = COLOR, ylab = YLAB,
       xlab = "Time (years)", cex.lab = 1.2, axes = FALSE, cex.axis = 1.1,
       xlim = c(-1, 2), ylim = c(min(Y) - abs(min(Y) * 0.2), max(Y) * 1.0))
  # Agregar ejes
  axis(2, at = seq_y_formatted, cex.axis = 0.9, tick = TRUE)
  axis(side = 1, at = c(-1, 0, 1, 2), labels = c("0.1", "1", "10", "50"),
       tick = TRUE, las = 1, cex.axis = 1.1, mgp = c(3, 1, 0))
  # Agregar línea vertical
  abline(v = log10(0.75), col = "black", lty = "dashed", lwd = 1.2)
  # Agregar línea suavizada
  lines(smooth.spline(x = log10(X), y = Y, spar = 2), lwd = 2.5, col = COLOR)
}

# Z scores
pdf(paste("3.Zscore_plot_age(years)_", plotname, ".pdf", sep = ""))
par(mfrow = c(2,2))
funPlot(X = Df.edades$Age_years, Y = as.numeric(Castillo[4,]), COLOR = "green4", YLAB = "Z score", MAIN = "Cortical Glia by ages") # Cambiar en MAIN nombre de
fenotipo
dev.off()

# Expression
pdf(paste("3.Expression_plot_age(years)_", plotname, ".pdf", sep = ""))
par(mfrow = c(2,2))
funPlot(X = Df.edades$Age_years, Y = as.numeric(Castillo_r[1,]), COLOR = "green4", YLAB = "Expression", MAIN = "Cortical Glia by ages") # Cambiar en MAIN nombre de
fenotipo
dev.off()

```

Script 10. Número de copias por familia

```

library(data.table)

# 1. Cargar la base correcta
setwd("C:/Users/James/OneDrive/Escritorio/match/385")
domains <- readRDS("domains_with_GO_FINAL.rds")

# 2. Convertir a data.table
setDT(domains)

# 3. Verificar columnas clave
stopifnot(
  "Species" %in% colnames(domains),
  "PANTHER_ID" %in% colnames(domains),
  "Gene_stable_ID" %in% colnames(domains)
)

# 4. Eliminar filas sin familia PANTHER
domains <- domains[!is.na(PANTHER_ID) & PANTHER_ID != ""]

# 5. Calcular número de copias por especie y familia
copy_number <- domains[
  ,
  .(Copy_number = uniqueN(Gene_stable_ID)),
  by = .(Species, PANTHER_ID)
]

# 6. Guardar resultados
fwrite(copy_number, "Copy_number_by_family_385.csv")
saveRDS(copy_number, "Copy_number_by_family_385.rds")

# 7. Vista rápida
head(copy_number)

```

```
#PLOT
library(ggplot2)
copies <- copy_number
setDT(copies)
ggplot(copies, aes(x = Species, y = Copy_number)) +
  geom_boxplot() +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, face = "italic")
  ) +
  labs(
    title = "Distribución del número de copias por especie (385)",
    x = "Especie",
    y = "Número de copias por familia"
  )
```

Script 11. Dominios proteicos por familia de genes

```
library(biomaRt)
library(dplyr)
###Cargar base de datos de interes###
setwd("C:/Users/James/OneDrive/Escritorio/match/385") #Direcccionar carpeta de trabajo #Address work folder
# Cambia el nombre del archivo si es necesario
data <- read.csv("NC_GF385_rat_match.csv", stringsAsFactors = FALSE)
# Verificar que la columna exista
stopifnot("Gene_stable_ID" %in% colnames(data))
###Conectar a Biomart y elegir la especie###
mart <- useMart(
  biomart = "ensembl",
  dataset = "rnorvegicus_gene_ensembl" # CAMBIA AQUÍ DATA SET BIOMART
)
###Obtener proteinas mas domios###
annotations <- getBM(
  attributes = c(
    "ensembl_gene_id",
    "ensembl_peptide_id",
    "interpro",
    "interpro_short_description"
  ),
  filters = "ensembl_gene_id",
  values = unique(data$Gene_stable_ID),
  mart = mart
)
###Limpiar y renombrar columnas###
annotations_clean <- annotations %>%
  filter(!is.na(ensembl_peptide_id), ensembl_peptide_id != "") %>%
  transmute(
    Gene_stable_ID = .data[[grep("gene", colnames(.), value = TRUE)[1]],
    Protein_stable_ID = ensembl_peptide_id,
    InterPro_ID = interpro,
    InterPro_domain = interpro_short_description
  )
head(annotations_clean)
###Unir con base original###
final_table <- data %>%
  left_join(annotations_clean, by = "Gene_stable_ID")
# Limpieza de datos
data1 <- final_table
# Eliminar filas sin PANTHER_ID
data_clean_1 <- data1 %>%
  filter(
    !is.na(PANTHER_ID) &
    PANTHER_ID != ""
  )
# 3. Eliminar filas sin InterPro ID
data_clean_2 <- data_clean_1 %>%
  filter(
    !is.na(InterPro_ID) &
    InterPro_ID != ""
  )
###Guardar###
write.csv(
  data_clean_2,
  "Gene_Protein_Domain_Table_rnorvegicus.csv",
  row.names = FALSE
)
```

Script 12. Comparación de dominios proteicos por familia de genes

```
## 0. Paquetes
library(data.table)
## 1. Cargar datos
setwd("C:/Users/James/OneDrive/Escritorio/match/385") #Direcccionar carpeta de trabajo #Address work folder
domains <- readRDS("all_species_domains.rds")
setDT(domains)
domains <- domains[!is.na(InterPro_ID) & InterPro_ID != ""]
## 2. Conteo x especie x dominio
```

```

interpro_species <- domains[
  , .N,
  by = .(Species, InterPro_ID, InterPro_domain)
]
## 3. Normalización
species_totals <- interpro_species[
  , .(Total_domains = sum(N)),
  by = Species
]

interpro_species <- merge(
  interpro_species,
  species_totals,
  by = "Species"
)

interpro_species[
  , Relative_abundance := N / Total_domains
]
## 4. Guardar resultados
saveRDS(interpro_species, "InterPro_species_normalized.rds")
fwrite(interpro_species, "InterPro_species_normalized.csv")
## 5. Visualización de datos
#### Top global de dominios
top_domains <- domains[
  , .N,
  by = InterPro_domain
][order(-N)][1:15, InterPro_domain]
plot_data <- interpro_species[InterPro_domain %in% top_domains]
library(ggplot2)
ggplot(plot_data,
  aes(x = Species,
      y = Relative_abundance,
      fill = InterPro_domain)) +
  geom_col() +
  theme_minimal() +
  labs(
    title = "Perfil relativo de dominios por especie (385)",
    y = "Abundancia relativa",
    x = "Especie"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, face = "italic"))

```

Script 13. Dominios diferencialmente abundantes

```

####DOMINIOS DIFERENCIALMENTE ABUNDANTES###
library(data.table)
setwd("C:/Users/James/OneDrive/Escritorio/match/385")
# 1. CARGAR DATOS
domains <- readRDS("domains_with_GO_FINAL.rds")
setDT(domains)
stopifnot(
  "Species" %in% colnames(domains),
  "InterPro_ID" %in% colnames(domains)
)
# 2. DEFINIR GRUPOS (USANDO NOMBRES REALES)
primates <- c(
  "Homo sapiens",
  "Gorilla",
  "Macaca mulatta",
  "Macaca fascicularis",
  "Papio anubis",
  "Callithrix jacchus",
  "Microcebus murinus",
  "Otolemur garnettii"
)

domains[, Group := ifelse(Species %in% primates, "Primates", "Non_primates")]

# Verificación
domains[, .N, by = Group]
# 3. CONTAR DOMINIOS
counts <- domains[
  , .N,
  by = .(Group, InterPro_ID)
]
counts[, freq := N / sum(N), by = Group]
# 4. REFORMATAR (PRIMATES vs NO)
counts <- dcast(
  counts,
  InterPro_ID ~ Group,
  value.var = "freq",
  fill = 0
)
setDT(counts)
print(colnames(counts)) # DEBE mostrar Primates y Non_primates
# 5. ANÁLISIS DIFERENCIAL
results <- counts[
  ,
  {
    logFC <- log2((Primates + 1e-6) / (Non_primates + 1e-6))

    p <- wilcox.test(
      x = c(Primates),
      y = c(Non_primates),
      exact = FALSE
    )$p.value
  }
]

```

```

    (
      logFC = logFC,
      p_value = p
    )
  },
  by = InterPro_ID
]
# 6. AJUSTE POR MÚLTIPLES TESTS
results[, padj := p.adjust(p_value, method = "BH")]
# 7. GUARDAR
fwrite(results, "Differential_InterPro_domains.csv")
saveRDS(results, "Differential_InterPro_domains.rds")
# 8. HEATMAP
library(pheatmap)
domains <- results
# Conteos normalizados
heat_data <- domains[
  , N,
  by = .(Species, InterPro_ID)
][
  , freq := N / sum(N),
  by = Species
]
# Pasar a matriz
heat_mat <- dcast(
  heat_data,
  InterPro_ID ~ Species,
  value.var = "freq",
  fill = 0
)
# Convertir a matriz
mat <- as.matrix(heat_mat[, -1])
rownames(mat) <- heat_mat$InterPro_ID
# Filtrar dominios variables (evita ruido)
mat <- mat[apply(mat, 1, sd) > 0.0001, ]
# Crear etiquetas en cursiva para especies
species_labels <- colnames(mat)
species_labels_italic <- parse(
  text = paste0("italic(", species_labels, ")")
)

pheatmap(
  mat,
  scale = "row",
  clustering_distance_rows = "correlation",
  clustering_distance_cols = "correlation",
  show_rownames = FALSE,
  labels_col = species_labels_italic,
  fontsize_col = 12,
  main = "Distribución comparativa de dominios (385)"
)

```



Constancia de participación

En el **1er Congreso Mexicano de Evolución: Ciencia y Comunidad en Biología Evolutiva** celebrado del 18 al 21 de noviembre de 2025, en C. U., Universidad Nacional Autónoma de México, CDMX, México.

En la modalidad:

Con el trabajo Expansión de familias de genes relacionado con el sistema inmune titulado: asociada a un mayor número de neuronas en el cerebro de mamíferos

A:

Jaime Tenopala Rodríguez



Dra. Nancy Claudia Saavedra Sotelo
Presidenta
Red Mexicana de Biología Evolutiva A. C. (ReMBE)



Dr. Ricardo Noguera Solano
Seminario Universitario de Evolución
Secretario Técnico
Seminario Universitario de Evolución (SUE)

