

MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA



VNIVERSITAT
ID VALÈNCIA

TRABAJO DE FIN DE MÁSTER

METAANÁLISIS FUNCIONAL DE LA LESIÓN MEDULAR A PARTIR DE ESTUDIOS TRANSCRIPTÓMICOS EN *RATTUS NORVEGICUS*

**AUTOR:
RUBÉN GRILLO RISCO**

**TUTORES:
FRANCISCO GARCÍA GARCÍA
ROSARIO GIL GARCÍA**

SEPTIEMBRE, 2019



MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA

TRABAJO DE FIN DE MÁSTER

METAANÁLISIS FUNCIONAL DE LA LESIÓN MEDULAR A PARTIR DE ESTUDIOS TRANSCRIPTÓMICOS EN *RATTUS NORVEGICUS*

**AUTOR:
RUBÉN GRILLO RISCO**

**TUTORES:
FRANCISCO GARCÍA GARCÍA
ROSARIO GIL GARCÍA**

TRIBUNAL:

PRESIDENTE/A:

VOCAL 1:

VOCAL 2:

FECHA DE DEFENSA:

CALIFICACIÓN:

Agradecimientos

A Paco, por darme la oportunidad de participar en este proyecto y ayudarme a abrirme camino.

A Vicky y Bea, por el asesoramiento y colaboración durante la investigación, y por lo que he podido aprender de ellas.

A mis amigos, con los que también comparto lugar de trabajo, por todo.

Resumen

La lesión medular es una patología con aproximadamente 12000 nuevos casos cada año tan solo en los Estados Unidos. Su etiología es variada, aunque la mayoría de los casos son debidos a causas traumáticas como accidentes de tráfico, caídas o heridas por armas, las cuales provocan una contusión en la médula espinal que desencadena un proceso patofisiológico con alteraciones en el sistema nervioso central y en un deterioro de los sistemas motores y sensoriales.

Actualmente no existen terapias capaces de restablecer las secuelas neurológicas asociadas a esta patología, por lo que es fundamental el uso de abordajes multidisciplinares que permitan una mejor caracterización de la lesión en sus diferentes fases.

Para llevar a cabo este objetivo, se realizó una revisión sistemática de todos los estudios transcriptómicos de rata disponibles en bases de datos públicas. Los estudios fueron seleccionados de acuerdo al tipo de tejido, fase de la enfermedad y gravedad de la lesión, incorporando grupos experimentales homogéneos para su evaluación conjunta. En el siguiente paso, se aplicó un pipeline de análisis bioinformático que incluía el procesamiento de los datos, análisis de expresión diferencial, caracterización funcional con GSA, análisis de rutas de señalización con Hipathia y finalmente, un metaanálisis de todos los resultados funcionales.

La estrategia seguida ha generado una gran cantidad de resultados, proporcionando una mejor caracterización de la lesión medular. Su evaluación en conjunto con expertos en la materia, permitirá la selección de candidatos de interés para la validación de nuevos biomarcadores en las diferentes fases de esta patología.

Palabras clave: lesión medular, GSA, transcriptómica, metaanálisis.

Abstract

Spinal Cord Injury (SCI) is a pathology with approximately 12000 new cases every year only in United States. The etiology of SCI is wide; however the most common causes are traumatic, such as motor vehicle accidents, falls or acts of violence. In all these accidents the contusion in the spinal cord triggers a cascade of pathophysiological events with alterations in the central nervous system and dysfunctions in motor and sensory systems.

Nowadays SCI is a pathology without effective therapeutic solutions, so it is essential to use multidisciplinary approaches that allow a better characterization of the molecular mechanisms in the different phases.

To achieve this objective, a systematic review of all available transcriptomic studies from public databases was carried out. Studies were selected according to tissue type, disease phase and severity of the lesion, incorporating homogeneous experimental groups in order to obtain a global evaluation. In the next step, to these selected datasets we applied a pipeline analysis that include differential gene expression analysis, functional enrichment analysis by GSA, signalling pathways analysis and finally, a meta-analysis of all functional results.

The strategy followed has generated a large amount of results, providing a better characterization of SCI. Their evaluation, discussed with experts in the field will allow the selection of candidates of interest for the validation of new biomarkers in the different stages of this pathology.

Índice general

Agradecimientos	v
Resumen	vii
Abstract	viii
1. Introducción	1
1.1. Marco contextual	1
1.2. Etiología y epidemiología	2
1.3. La médula espinal en el sistema nervioso	4
1.4. Anatomía de la médula espinal	5
1.5. Patofisiología de la lesión Medular traumática	6
1.6. Modelos animales	11
1.7. Transcriptómica	12
1.8. Caracterización funcional	12
1.9. Metaanálisis funcional	14
2. Objetivos	15
3. Materiales y métodos	16
3.1. Revisión sistemática y selección de estudios	17
3.2. Grupos experimentales y análisis exploratorio	18
3.2.1. Establecimiento de grupos experimentales	18
3.2.2. Análisis exploratorio	19
3.3. Análisis individual de estudios	19
3.3.1. Análisis de expresión diferencial	19
3.3.2. Creación de ficheros de anotación	20
3.3.3. Análisis de enriquecimiento funcional	21
3.3.4. Análisis de rutas de señalización	21
3.4. Metaanálisis de resultados con GSA	22
3.4.1. Determinación de la medida combinada del efecto	22
3.4.2. Evaluación de la heterogeneidad	23
3.4.3. Representación de resultados	24
3.5. Metaanálisis de resultados de Hipathia	25
4. Resultados	26
4.1. Revisión sistemática y selección de estudios	26
4.2. Grupos experimentales y análisis exploratorio	27

4.3.	Análisis individual de estudios	32
4.3.1.	Análisis de expresión diferencial	32
4.3.2.	Análisis de enriquecimiento funcional	33
4.3.3.	Análisis de rutas de señalización	35
4.4.	Metaanálisis	36
4.4.1.	Enriquecimiento funcional	36
4.4.1.1.	Términos GO para procesos biológicos	36
4.4.1.2.	Rutas KEGG	39
4.4.1.3.	Rutas Reactome	42
4.4.2.	Rutas de señalización	45
5.	Discusión	46
6.	Conclusiones	52
	Bibliografía	53
A.	Anexo A - Figuras	59
B.	Anexo B - Tablas	74

Índice de figuras

1.1. Organización del sistema nervioso	5
1.2. Anatomía de la médula espinal y la vértebra	6
1.3. Tipos de lesiones en la médula espinal	7
1.4. Resumen de las consecuencias patológicas tras una lesión medular	9
1.5. Diagrama esquemático que ilustra la composición estructural y celular de la médula espinal normal (arriba) y lesionada (abajo).	11
3.1. Flujo de trabajo seguido	16
3.2. Diagrama de flujo de una revisión sistemática según la declaración PRISMA.	17
3.3. Ejemplo de gráficos de embudo	24
3.4. Ejemplo de gráfico de bosque	25
4.1. Diagrama de flujo de la revisión sistemática, siguiendo la declaración PRISMA	27
4.2. Ejemplos de PCA	29
4.3. Ejemplos de clustering	29
4.4. Diagramas de Venn de los resultados del metaanálisis con términos GO	37
4.5. Resumen de términos GO con mayor LOR en valor absoluto para lesiones moderadas.	38
4.6. Diagramas de Venn de los resultados del metaanálisis con rutas KEGG	39
4.7. Resumen de rutas KEGG con mayor LOR en valor absoluto para lesiones moderadas.	41
4.8. Diagramas de Venn de los resultados del metaanálisis con rutas Reactome	42
4.9. Resumen de rutas Reactome con mayor LOR en valor absoluto para lesiones moderadas.	44
A.1. Análisis exploratorio del estudio GSE115067	59
A.2. Análisis exploratorio del estudio GSE102964	60
A.3. Análisis exploratorio del estudio GSE104317	61
A.4. Análisis exploratorio del estudio GSE93249	62
A.5. Análisis exploratorio del estudio GSE52763	63
A.6. Análisis exploratorio del estudio GSE45006	64
A.7. Análisis exploratorio del estudio GSE46988	65
A.8. Análisis exploratorio del estudio GSE29488	66
A.9. Análisis exploratorio del estudio GSE2599	67
A.10. Análisis exploratorio del estudio GSE464 (lesión moderada)	68
A.11. Análisis exploratorio del estudio GSE464 (lesión severa)	69
A.12. Análisis exploratorio del estudio VBB	70
A.13. Resumen de términos GO con mayor LOR en valor absoluto para lesiones severas.	71

A.14. Resumen de rutas KEGG con mayor LOR en valor absoluto para lesiones severas.	72
A.15. Resumen de rutas Reactome con mayor LOR en valor absoluto para lesiones severas.	73

Índice de tablas

4.1.	Estudios seleccionados tras la revisión sistemática	26
4.2.	Número de muestras seleccionadas en cada estudio	28
4.3.	Resumen de resultados del análisis de expresión diferencial.	32
4.4.	Resumen de resultados del análisis de enriquecimiento funcional para lesiones moderadas	33
4.5.	Resumen de resultados del análisis de enriquecimiento funcional para lesiones severas	34
4.6.	Resumen de resultados del análisis de rutas de señalización con Hipathia.	35
4.7.	Resultados del metaanálisis para procesos biológicos de la Gene Ontology.	36
4.8.	Selección de los 10 términos GO con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas	37
4.9.	Resultados del metaanálisis para rutas KEGG.	39
4.10.	Selección de las 10 rutas KEGG con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas	40
4.11.	Resultados del metaanálisis para rutas Reactome.	42
4.12.	Selección de las 10 rutas Reactome con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas	43
4.13.	Resultados del metaanálisis para rutas de señalización de Hipathia.	45
B.1.	Versiones del software empleado en este trabajo.	74

Acrónimos

ADN Ácido desoxirribonucleico.

ARN Ácido ribonucleico.

ATP Adenosín trifosfato.

CIPF Centro de Investigación Príncipe Felipe.

DAMP Patrones moleculares asociados a lesión.

DL DerSimonian & Laird.

FAIR *Findable, Accesible, Interoperable, Reusable.*

FDR False discovery rate.

GEO *Gene Expression Omnibus.*

GO *Gene Ontology.*

GSA Gene Set Analysis.

GWAS Estudios de asociación del genoma completo.

IFN Interferón.

IL Interleucina.

KEGG *Kyoto Encyclopedia of Genes and Genomes.*

LM Lesión medular.

logFC Logaritmo de la magnitud de cambio.

LOR Logaritmo de odds ratio.

ORA Over-representation analysis.

PAMP Patrones moleculares asociados a microorganismos patógenos.

PCA Análisis de componentes principales.

RNA-seq Secuenciación de ADN.

ROS Especies reactivas de oxígeno.

SNA Sistema nervioso autónomo.

SNC Sistema nervioso central.

SNP Sistema nervioso periférico.

SNS Sistema nervioso somático.

TNF Factor de necrosis tumoral.

1. Introducción

La Lesión Medular (LM) es una patología producida por un daño en la médula espinal, sea cual sea su etiología, que conduce a alteraciones en el sistema nervioso central (SNC) y al deterioro de los sistemas sensoriales y motores, tanto el autónomo como el somático.

Las consecuencias de estas alteraciones son devastadoras para la calidad de vida de los pacientes ya que se produce una parálisis de la movilidad voluntaria, que dependiendo del nivel donde se produzca puede derivar en paraplejia (parálisis de las extremidades inferiores) o tetraplejia (parálisis en las extremidades superiores e inferiores). Los fallos en el sistema nervioso autónomo (SNA) tienen múltiples consecuencias como pérdida de la sensibilidad y disfunción vesical, intestinal o sexual, entre otras [66].

Actualmente la LM no tiene cura, aunque existen terapias que permiten mejorar en cierta medida la calidad de vida de los pacientes. Por ello, es de vital importancia conocer los procesos moleculares que controlan la LM, y así detectar las funciones biológicas que se están activando o inhibiendo en las diferentes fases de esta enfermedad. Esto permitirá elaborar nuevas hipótesis dirigidas a encontrar dianas terapéuticas que podrán ser validadas experimentalmente en laboratorio.

1.1. Marco contextual

El presente trabajo no es un trabajo aislado, sino que se enmarca dentro de un proyecto mayor de corte multidisciplinar, impulsado por la asociación CORRE PER MI.

La asociación CORRE PER MI es una asociación sin ánimo de lucro nacida en Museros (Valencia), que tiene como objetivo recaudar fondos que ayuden a financiar investigaciones científicas enfocadas a encontrar una cura o terapias paliativas para la lesión medular.

El pasado 2 de diciembre tuvo lugar la II Marxa Solidària CORRE PER MI, organizada con el apoyo de distintas entidades. El dinero recaudado fue donado al Centro de Investigación Príncipe Felipe (CIPF), donde se está llevando a cabo un proyecto conjunto que une a la Unidad de Bioinformática y Bioestadística, liderada por el Dr. Francisco García y al Laboratorio de Regeneración Neuronal y Tisular, con la Dra. Victoria Moreno a la cabeza.

El proyecto se divide en dos partes. La primera implica el uso de abordajes bioinformáticos para realizar un metaanálisis de todos los estudios transcriptómicos disponibles para caracterizar funcionalmente la LM. Esta investigación es de carácter traslacional, pues el fin es ser

capaz de encontrar nuevas terapias que beneficien a las personas. Por ello, estos abordajes se realizarán en diferentes especies, incluyendo a *Homo sapiens*, ya que es importante conocer los mecanismos moleculares asociados a la LM que son compartidos entre distintos organismos modelos y humano. Esto permitirá testar tratamientos y trasladarlo a la clínica en caso de tener comportamientos similares. La segunda parte del proyecto es de vital importancia en esto, ya que los resultados obtenidos *in silico* podrán servir para generar nuevas hipótesis que puedan ser validadas de manera experimental en laboratorio, tanto *in vitro* como *in vivo*.

De esta manera, el transcurso de la investigación en un entorno multidisciplinar facilita la integración de distintos tipos de conocimiento. Así, los análisis bioinformáticos se guían gracias al criterio de investigadores con experiencia en la biología de la LM, y al mismo tiempo estos pueden realizar una mejor interpretación de los resultados y profundizar en ellos con ayuda de especialistas bioinformáticos y herramientas computacionales.

En este contexto, se decidió comenzar el metaanálisis funcional con los estudios transcriptómicos disponibles de *Rattus norvegicus*, dando lugar al presente trabajo.

1.2. Etiología y epidemiología

La LM es una patología con una etiología muy diversa, la cual se divide en dos grupos principales, la LM traumática y la LM no traumática. Las lesiones de origen traumático abarcan a todas las que se producen por golpes y contusiones en la médula espinal, mientras que las de origen no traumático suelen darse a raíz de enfermedades previas. Estas consideraciones son importantes a la hora de hablar de la incidencia de la LM, ya que esta va muy ligada al estilo de vida de cada país o región geográfica, por lo que la incidencia, la proporción de causas desencadenantes y el perfil sociodemográfico de los pacientes varía mucho en las diferentes partes del mundo. A ello hay que sumarle un añadido, y es el hecho de que no suelen existir registros a nivel nacional y hay que recurrir a estudios poblacionales para poder obtener cifras epidemiológicas [36].

Las estimaciones de incidencia fluctúan de los 40 casos por millón en Norteamérica, con unos 12000 nuevos casos cada año (siendo esta la región con mayor incidencia en el mundo), a los 15 casos por millón en regiones de Europa Occidental o Australia [1, 36]. Se calcula que la prevalencia en Norteamérica es de 853 casos por millón, mientras que en Europa Occidental es de 300 casos por millón [14].

En la literatura normalmente se indica que las lesiones traumáticas son las más frecuentes que las lesiones no traumáticas, siendo alrededor del 60% en países desarrollados y el 80% en países en desarrollo [42]. Principalmente son debidas a accidentes de tráfico, heridas con armas, caídas y accidentes deportivos. En EE.UU. el 38% de las lesiones traumáticas se deben

a los accidentes de tráfico, mientras que las caídas suponen el 31%, y los accidentes deportivos un 10-17% [1]. Un ejemplo que ilustra la variabilidad en la etiología en función del país se encuentra en un estudio realizado en Colombia en 2007, donde el 62% de los pacientes incluidos en la investigación habían sido lesionados por armas de fuego [45].

La naturaleza de las lesiones traumáticas influye en gran medida en el sexo de los pacientes que las sufren, ya que dependiendo del país, entre el 66 y el 92% son hombres, típicamente varones jóvenes de entre 15 y 30 años [1, 42].

Las lesiones medulares no traumáticas también tienen un amplio abanico de causas, habitualmente consecuencia de enfermedades previas. Aunque de nuevo los porcentajes varían mucho de unos estudios a otros, las principales causas son la presencia tumores, que comprimen la médula espinal, enfermedades degenerativas derivadas de desórdenes autoinmunes, infecciones o desórdenes vasculares. Además, también pueden ser de origen congénito como la espina bífida, sin embargo, existe un porcentaje de casos donde la causa es desconocida [31, 50]. Este tipo de circunstancias suele presentarse en pacientes de edad avanzada y con cuadros clínicos complicados, previos a la lesión, de manera que la mortalidad en las lesiones no traumáticas es mayor que en las lesiones traumáticas. Cabe destacar que en las lesiones no traumáticas no se encuentran diferencias de incidencia entre hombres y mujeres [15, 36].

En España se estima una incidencia de la LM de entre 12-20 casos por millón, con unos 800 nuevos pacientes cada año y una prevalencia de unas 30000 personas. Al no haber un sistema nacional de registro y la propia autogestión de las comunidades autónomas, entre otras razones, no se pueden obtener cifras epidemiológicas específicas. Al igual que en la mayoría de países desarrollados, el paciente típico sigue siendo un varón de edad media de 35 años, cuya causa predominante de lesión es la traumática con origen en accidentes de tráfico. Sin embargo, el perfil sociodemográfico de los pacientes está siguiendo otra tendencia en los últimos años. Está aumentando el número de pacientes cuya causa de la lesión también es traumática, pero debida a caídas en lugar de accidentes de tráfico. Por otro lado, también está aumentando el número de lesionados no traumáticos, debido a problemas de salud previos [36].

Dichas tendencias se dan gracias a múltiples factores. Los casos de lesionados por accidentes de tráfico vienen disminuyendo desde un conjunto de reformas que comenzaron a aplicarse a principios los 2000. Estas intervenciones en seguridad vial incluyen la mejora del estado de las carreteras, la implantación del carnet por puntos, un mayor número de radares y controles de alcoholemia, así como campañas de concienciación y educación vial, entre otros factores [56]. Por otra parte, el aumento en la longevidad de la ciudadanía unido a una mayor actividad en la tercera edad y la tipología de las mismas, en ciertos casos hace que haya más pacientes cuya causa de la lesión sean caídas. Pese a ello, como se mencionaba anteriormente, a partir de los 50 años la LM es de origen no traumático en un mayor porcentaje, surgiendo como consecuencia de otras patologías [36].

Desde el punto de vista económico, en EE.UU. se estima un coste aproximado durante el

primer año de unos 500.000\$ por paciente y de unos 70.000\$ en los años posteriores [17], con costes similares en Europa. A este gran impacto económico tanto para el sistema sanitario como para las familias de los pacientes, hay que añadir las enormes consecuencias que suponen estas lesiones a nivel personal y social, lo cual hace que resulte de una mayúscula importancia la investigación de la LM.

1.3. La médula espinal en el sistema nervioso

Para comprender qué es la médula espinal, la importancia de su correcto funcionamiento y las consecuencias que tienen para el organismo las patologías que le afectan, primero conviene entender cómo se estructura y cómo funciona el sistema nervioso.

El sistema nervioso puede ser dividido en dos partes, el sistema nervioso central (SNC) y el sistema nervioso periférico (SNP). El SNC lo forman el cerebro y la médula espinal, que se encargan de procesar e integrar toda la información que el individuo percibe del exterior y de los distintos órganos del cuerpo, respondiendo en consecuencia de los mismos. Por otro lado, el SNP lo forman ganglios, nervios y terminaciones nerviosas que conducen la información hasta el SNC y desde el SNC. Así pues, desde un punto de vista funcional el SNP tiene un componente sensitivo o aferente, que percibe los estímulos y transmite los impulsos hacia el SNC, y un componente motor o eferente, que lleva los impulsos generados por el cerebro hasta los órganos efectores. Este componente motor se subdivide en el sistema nervioso somático (SNS) y el sistema nervioso autónomo (SNA). El SNS lleva los impulsos generados en el cerebro hasta los músculos esqueléticos, por lo que se encarga de los movimientos voluntarios, mientras que el SNA lleva los impulsos a glándulas, músculo cardíaco o músculo liso, de manera que se encarga de todos los movimientos involuntarios del cuerpo, como el latido del corazón o la vasodilatación en respuesta a cambios de temperatura (figura 1.1) [55].

Como se decía anteriormente, la médula espinal forma parte del SNC, siendo el mediador entre el cerebro y el resto del cuerpo, ya que recibe en todo momento las señales percibidas, tanto del exterior como del interior, y envía las respuestas que deben realizarse en función de las mismas. Así, es fácil intuir que los daños en la médula espinal tendrán múltiples consecuencias en numerosas funciones del organismo.

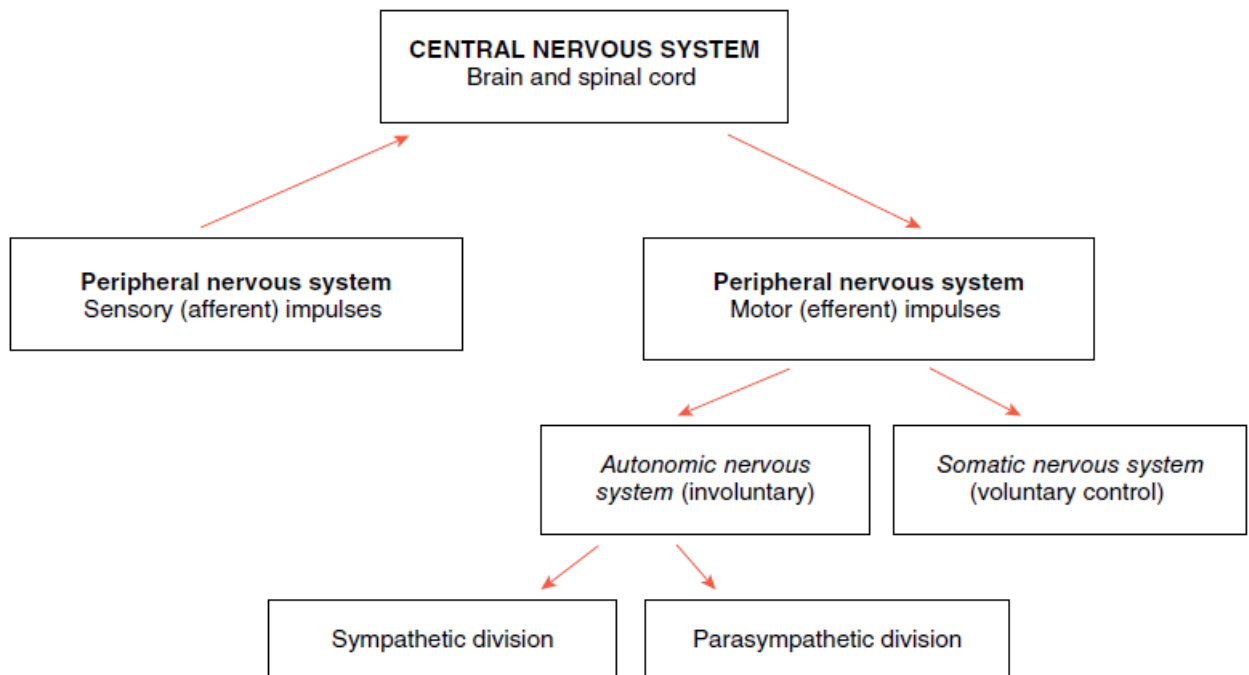


Figura 1.1: Organización del sistema nervioso. Fuente: Peate & Muralitharan (2016) [55].

1.4. Anatomía de la médula espinal

En humanos, la médula espinal en los adultos mide unos 45 cm de media. A lo largo de la médula espinal se identifican 31 pares de nervios espinales que se conectan con diferentes partes del cuerpo: 8 nervios cervicales, 12 torácicos, 5 lumbares, 5 sacros y un nervio coccígeo. Los 8 nervios cervicales controlan los músculos del cuello y la cabeza, parte de las extremidades superiores, la respiración o la frecuencia cardíaca. Los nervios torácicos controlan la musculatura del tronco y la regulación de la temperatura. Los nervios lumbares se asocian con el movimiento del tren inferior y el control de la función sexual, intestinal y vesical. Finalmente, el nervio coccígeo se asocia con musculatura del coxis y del recto [55].

Dado que el tejido nervioso es muy delicado, es importante para el organismo que esté bien protegido frente a daños. La primera capa de protección es la columna vertebral, que aporta una fuerte defensa rodeando la médula espinal con tejido óseo. Las meninges son la segunda capa de protección, formada por tres membranas que revisten la médula espinal, ubicadas entre la caja ósea y el tejido nervioso: la duramadre (más externa), la aracnoides y la piamadre (más interna). Existe una tercera defensa contra daños mecánicos en el espacio subaracnoideo, entre la aracnoides y la piamadre, el cual contiene líquido cefalorraquídeo, que suspende el tejido

nervioso central en un entorno ingrávido que absorbe los golpes (figura 1.2) [59, 68].

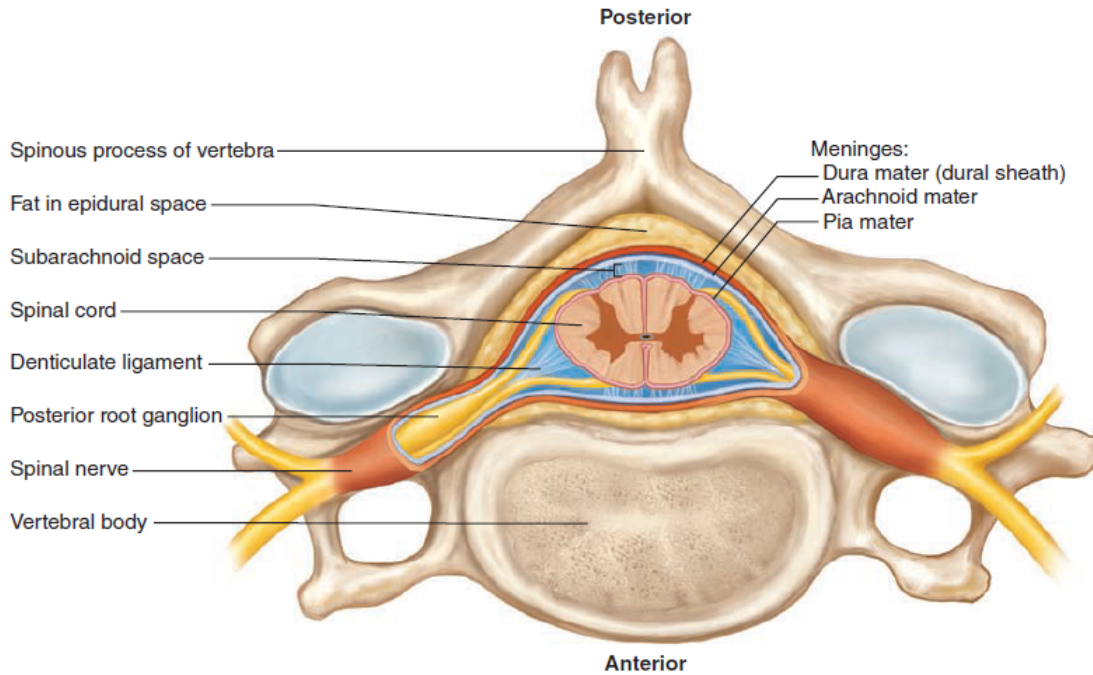


Figura 1.2: Anatomía de la médula espinal y la vértebra. Fuente: Saladin (2017) [59].

En un corte transversal de la médula espinal se diferencian dos regiones. Internamente, la sustancia gris con forma de H, está formada por dendritas y cuerpos neuronales, axones amielínicos y neuroglia (células auxiliares de las neuronas). La función de la sustancia gris es el procesamiento e integración de los estímulos percibidos. Alrededor de la sustancia gris, se encuentra la sustancia blanca, formada principalmente por axones mielínicos de neuronas, a los que debe su nombre. La mielina es una capa que envuelve los axones para que el impulso nervioso sea más rápido, formada fundamentalmente por lípidos, de ahí que esta región se vea de color blanco. Por ello, la función de la sustancia blanca es enviar los impulsos nerviosos hasta y desde el cerebro [68].

1.5. Patofisiología de la lesión Medular traumática

De esta sección en adelante se tratará solamente de lesiones medulares traumáticas. En la literatura se distinguen cuatro tipos en función de su origen y anatomía: lesión por compresión, por contusión, por laceración o transección y compresión masiva [52].

- **Lesión por compresión:** es el tipo menos frecuente y se asocia con una apariencia normal de la médula espinal, ya que se genera por una compresión constante durante un largo periodo de tiempo, por ejemplo, por un desplazamiento de las vértebras. En la figura 1.3.B se puede apreciar la pérdida de la arquitectura normal en comparación con la figura 1.3.A.
- **Lesión por contusión:** es el tipo más frecuente, siendo causada por un impacto que provoca áreas con hemorragia, necrosis y cavitación, pero sin ruptura de la superficie de la médula espinal. La 1.3.C muestra como en la médula se ha formado un gran quiste.
- **Lesión por laceración o transección:** al contrario que en las lesiones por contusión, en las transecciones si se ve una ruptura de la superficie anatómica de la médula. Puede ser causada directamente por el impacto de proyectiles, o indirectamente por la rotura de las vértebras. En la figura 1.3.D se puede ver un claro corte en la superficie de la médula.
- **Lesión por compresión masiva:** en estos casos la médula queda fragmentada en un grado variable, por lo que se identifican cicatrices de tejido conjuntivo en estos espacios (1.3.E).

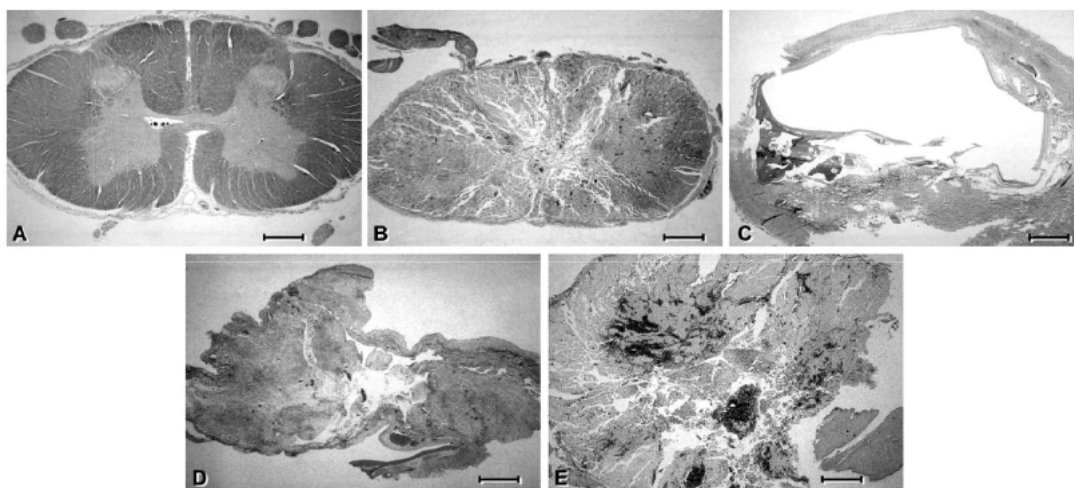


Figura 1.3: Corte transversal de la médula espinal tras diferentes tipos de lesión. (A). Corte transversal normal de médula espinal. **(B)** Lesión por compresión mostrando la pérdida de la arquitectura normal de la médula espinal. **(C)** Lesión por contusión en el que el tejido ha sido reemplazado por un quiste. **(D)** Lesión por laceración en la que se observa ausencia de tejido en la parte superior. **(E)** Lesión por compresión masiva mostrando fragmentación. Fuente: Norenberg et al. (2004) [52]

La patofisiología de la LM tiene dos fases, la lesión primaria y la lesión secundaria. La lesión primaria implica el daño mecánico inicial producido en la médula espinal. Este daño lleva a la rotura de los axones, vasos sanguíneos y membranas celulares, por lo que se interrumpe el impulso nervioso y el flujo sanguíneo, produciéndose hemorragias y un infarto local que afecta gravemente a las células de la materia gris, las cuales tienen un alto requerimiento metabólico. La lesión secundaria transcurre desde las horas posteriores al traumatismo inicial hasta meses o años después. En esta fase la lesión se extiende desde el epicentro hasta tejido sano que lo rodea, aumentando su tamaño. En función del tiempo, la fase secundaria puede ser dividida en cuatro fases: aguda, subaguda, fase intermedia y crónica tardía [23, 54].

A. Fase aguda (0-3 días): la fase aguda está marcada por un conjunto de alteraciones sistémicas y locales, que incluyen alteraciones de los mecanismos vasculares, desequilibrio iónico y una gran respuesta del sistema inmune.

Como se decía anteriormente, la médula espinal está ampliamente vascularizada, especialmente la materia gris. Al romperse los vasos hay una gran hemorragia que produce edema (retención de líquidos en los tejidos) e isquemia (disminución del flujo sanguíneo). Además, hay una entrada de células inflamatorias y citoquinas pro-inflamatorias, que raramente se ven en el SNC [46].

La combinación del daño mecánico celular con la isquemia hace que se pierda la homeostasis iónica, desregulándose las concentraciones de sodio, glutamato y otras moléculas, liberándose glutamato al medio extracelular. El glutamato es el principal neurotransmisor excitatorio del SNC, por lo que se produce una excitotoxicidad debido a la activación excesiva de sus receptores. La activación de estos receptores hace que entre Ca^{2+} a las células y se acumule en el citosol o en diferentes orgánulos, poniendo en marcha mecanismos de apoptosis mediante la activación de fosfolipasas, proteasas y fosforilasas [54].

En esta situación de estrés celular, se producen radicales libres que dañan las proteínas, ácidos nucleicos y las membranas celulares. Los radicales libres reaccionan con la membrana celular mediante peroxidación lipídica, alterando su estructura, por lo que hay un aumento de la permeabilidad de iones, favoreciendo la desregulación de la homeostasis iónica, que finalmente lleva a la inducción de la apoptosis [54, 62].

Dada la gravedad de la lesión y el conjunto de daños producido, el sistema inmune responderá de manera muy contundente para intentar contener el daño y restablecer la homeostasis. En esta respuesta inmune, la inflamación juega un papel clave, siendo un proceso de gran complejidad que incluye a múltiples poblaciones celulares y moléculas mediadoras intercelulares.

Los mecanismos de inflamación empiezan inmediatamente y algunos persisten hasta varios meses después. Tras el trauma, la microglía residente se activa mientras los linfocitos infiltrados por la rotura de los vasos sanguíneos comienzan a liberar citoquinas y especies reactivas de oxígeno (ROS). Tanto la microglía activada como los linfocitos liberan moléculas pro-inflamatorias

como citoquinas IL-6, IL-1, interferones o $\text{TFN-}\alpha$, permitiendo una mayor extravasación de linfocitos y mayor daño tisular. En respuesta a estas moléculas llegan neutrófilos, que aumentan la cantidad de tejido dañado al liberar proteasas y generar ROS, elevándose por tanto la liberación de señales pro-inflamatorias. Los macrófagos se ven atraídos por estas señales y comenzarán a fagocitar el tejido dañado y a liberar moléculas pro-inflamatorias [46, 62].

Como se ha visto, en la fase aguda hay una cascada de mecanismos moleculares que son consecuencia del traumatismo inicial y que se retroalimentan, agravando la lesión. La respuesta inflamatoria es muy importante para eliminar los desechos celulares y reparar los tejidos, sin embargo, una respuesta exagerada produce daño del tejido sano y exacerbación de la lesión [54]). La figura 1.4 muestra de manera esquemática los procesos patológicos que suceden tras una lesión traumática durante la lesión primaria y la fase aguda.

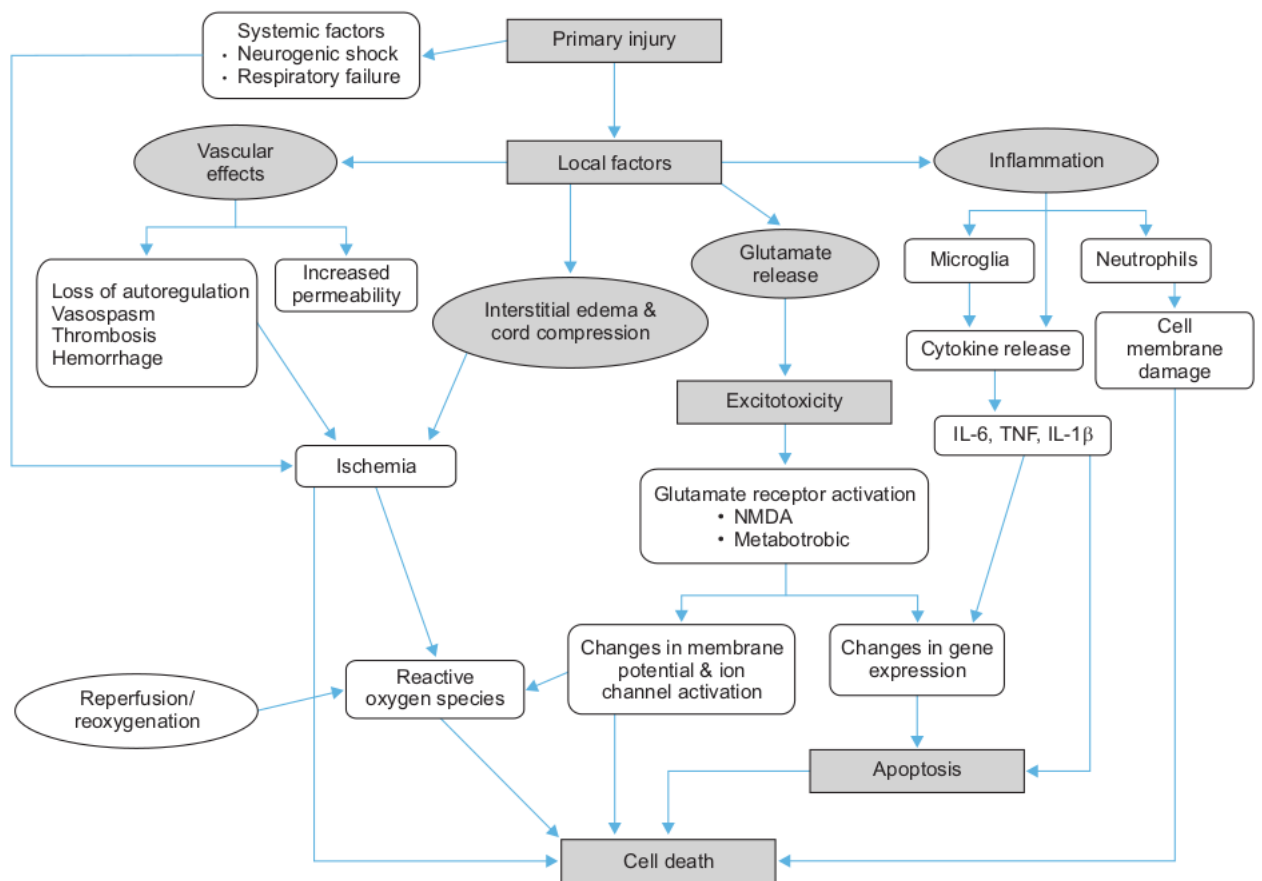


Figura 1.4: Resumen de las consecuencias patológicas tras una lesión medular. Fuente: Kim et al.(2017) [39].

B. Fase subaguda (3- 14 días): en este periodo hay un aumento de la actividad de los fagocitos para eliminar células y tejidos dañados, lo cual puede ser clave en la regeneración axonal. Por el contrario, rara vez se encuentran neutrófilos en la fase subaguda, por lo que su acción quedaría limitada a la fase aguda [4].

Los astrocitos son células con numerosas funciones importantes para las neuronas, como soporte, aporte de nutrientes o eliminación de nutrientes. En la fase subaguda los astrocitos comienzan a crecer tanto por hiperplasia como por hipertrofia, es decir, aumentan en número y en tamaño alrededor de la lesión para formar una cicatriz glial. Esto es beneficioso y perjudicial al mismo tiempo. La formación de la cicatriz permite separar la inflamación de los tejidos sanos para minimizar que se extiendan los daños secundarios descritos en la fase aguda. Sin embargo, en el futuro la cicatriz glial supondrá una barrera física que impedirá la regeneración axonal [49].

C. Fase intermedia o crónica temprana (14 días a 6 meses): se caracteriza por una disminución de la respuesta inflamatoria, que da lugar a un intento de regeneración axonal, regeneración vascular y reorganizaciones de la matriz extracelular y circuitos neuronales [1]. Por otro lado, la cicatriz glial sigue madurando, por lo que los intentos de regeneración y recuperación serán insuficientes en lesiones severas [54].

D. Fase crónica tardía (>6 meses): se caracteriza por la maduración y estabilización de la lesión. Puede darse cierta recuperación funcional derivada de mecanismos endógenos de remielinización, llevados a cabo por oligodendrocitos que pueden remielinizar algunos axones. Sin embargo, la regeneración resulta ineficaz ya que esta se inhibe por las moléculas que forman la matriz glial, la cual ha seguido formándose desde las primeras semanas [3, 57].

Al mismo tiempo ocurre una degradación Walleriana de los axones, que se da típicamente en fases crónicas de enfermedades neuronales. Es un proceso independiente de la apoptosis, donde los axones se degradan de manera activa mediante diferentes mecanismos como el sistema ubiquitin-proteasoma, que tiene al citoesqueleto como diana principal. El citoesqueleto es fundamental para la estructura y el metabolismo del axón, resultando su degradación en una efectiva destrucción del axón [26, 41].

Este proceso puede tardar años, por lo que la lesión no permanece estática y los pacientes pueden sufrir disfunciones neurológicas tardías, como dolor neuropático o disautonomía (fallo en la regulación de funciones involuntarias controladas por el SNA), entre otras múltiples condiciones patológicas [57].

En la figura 1.5 se representa de manera esquemática una comparación entre la médula espinal sana y la lesionada, en la cual se observan diferentes tipos celulares del sistema inmune, rotura y desmielinización de los axones y la formación de una cicatriz glial, entre otros procesos.

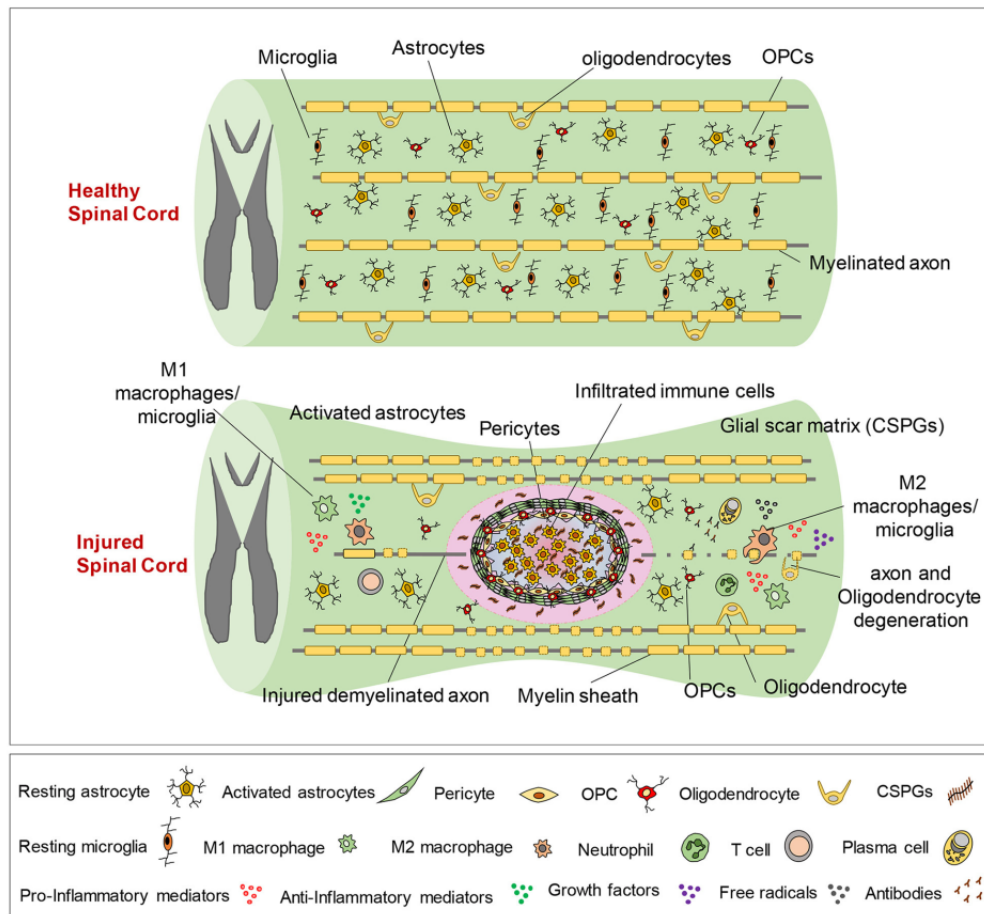


Figura 1.5: Diagrama esquemático que ilustra la composición estructural y celular de la médula espinal normal (arriba) y lesionada (abajo). En la parte inferior de la imagen se muestra la leyenda. Hay que señalar que algunos de los eventos mostrados pueden no coincidir al mismo tiempo. Fuente: Alizadeh et al.(2019) [4].

1.6. Modelos animales

El uso de modelos animales ha contribuido enormemente al entendimiento de la patofisiología de la LM y han sido muy útiles para probar nuevas terapias. La rata es el modelo animal más usado en investigaciones de LM debido a diversas ventajas. Son ampliamente conocidas, presentando un accesible mantenimiento controlado, lo que permite trabajar con múltiples individuos al mismo tiempo. Por otro lado, tienen una respuesta a la lesión parecida a la observada en humano, variando los tiempos de las fases de la lesión ya que tienen una esperanza de vida de unos 2 años. Sin embargo, existen diferencias que hacen que extrapolar resultados a humanos no siempre sea efectivo, como el tamaño corporal, la señalización molecular o la

anatomía, además de una mayor capacidad de recuperación. Por ello, los modelos animales de mayor tamaño como los primates no humanos, son importantes intermediarios para confirmar los resultados obtenidos en roedores [1].

1.7. Transcriptómica

El ADN codifica toda la información necesaria para determinar las propiedades y funciones de cada célula. Para ello la información contenida en el ADN se transcribe, sintetizándose el ARN mensajero (ARNm), que posteriormente será traducido para generar una proteína, las cuales son las ejecutoras de los procesos celulares. Al conjunto de ARNm de una célula o grupos de células en un momento dado se le denomina transcriptoma. Disponer de herramientas que permiten detectar y cuantificar al mismo tiempo todos los transcritos presentes en las células ha supuesto un enorme avance en el conocimiento de muchas enfermedades y de la biología en general.

Existen dos principales tecnologías de análisis de expresión génica, los microarrays y el RNA-seq. Los microarrays fueron los primeros en desarrollarse y se basan en la hibridación del ADN. Consisten en un chip con sondas fijadas de ADN de una sola hebra. Estas son complementarias a los transcritos, que son marcados con fluorocromos o con radioactividad, por lo que al hibridar con su sonda complementaria pueden detectarse y cuantificarse mediante análisis de imágenes. En este caso la cuantificación es relativa, ya que se comparan los niveles de expresión de los genes de dos muestras biológicas diferentes [11].

En los últimos años el RNA-seq está siendo más utilizado ya que se basa en la secuenciación masiva de los transcritos, lo que le confiere ciertas ventajas, como mayor resolución o la capacidad de identificar nuevos transcritos, ya que se conoce la secuencia nucleotídica de los mismos. Sin embargo, hay que destacar que los microarrays siguen siendo una metodología válida y de bajo coste, por lo que la elección de una u otra tecnología depende de los objetivos concretos del estudio [11].

1.8. Caracterización funcional

Los estudios de transcriptómica permiten comparar los perfiles de expresión entre varios grupos experimentales y detectar diferencias en el nivel de transcritos de los genes mediante análisis de expresión diferencial. Esto permite identificar qué genes se activan o inactivan en un grupo con respecto al otro, lo cual aporta información relevante a la hora de conocer los mecanismos moleculares subyacentes a cualquier estado celular. Pero aún se puede ir un paso más allá para comprender mejor qué procesos biológicos se están dando en una muestra de interés, y esto

es mediante los análisis de enriquecimiento funcional.

Los análisis de enriquecimiento funcional se nutren de los resultados obtenidos en la expresión diferencial para identificar los procesos celulares que se están llevando a cabo. De esta manera, si por ejemplo se compara un tejido enfermo con un tejido sano, puede determinarse qué funciones se han desregulado, lo que resulta de enorme utilidad a la hora de generar hipótesis que permitan testar nuevas dianas terapéuticas.

Estas funciones o módulos funcionales hacen referencia a grupos de genes relacionados por propiedades biológicas de interés, en este caso, genes que trabajan en una misma tarea biológica. Por tanto, son conjuntos de genes que comparten anotaciones funcionales extraídas de cualquier base de datos [21]. Una anotación funcional no es otra cosa que una etiqueta que describe las funciones biológicas asociadas a un gen o las rutas metabólicas donde participa. Cada base de datos se estructurará de una manera u otra y detallarán los procesos biológicos de manera diferente, por lo que normalmente se utilizan varias anotaciones para complementar la información y tener distintos puntos de vista, por ello en el presente trabajo se han usado tres bases de datos diferentes: la Gene Ontology (GO), KEGG y Reactome.

La ontología de términos GO es la base de datos por excelencia en el mundo del análisis funcional, probablemente la más usada en la literatura. Se compone de tres clases de términos GO, que describen las funciones moleculares, los componentes celulares y los procesos biológicos asociados a los genes. Tiene una estructura jerárquica donde los nodos pueden tener más de un padre, por lo que un nodo en concreto adoptará todas las anotaciones de los niveles superiores. En esta jerarquía las anotaciones de los niveles superiores hacen mención a funciones más generales, que se van especificando a medida que se baja de nivel [21].

Las rutas KEGG (Kyoto Encyclopedia of Genes and Genomes) también son ampliamente usadas, ya que contienen información sobre los genes y cómo se organizan las redes de reacciones enzimáticas. Esto es muy interesante, por ejemplo, a la hora de buscar dianas terapéuticas para tratar una enfermedad, donde hay que estudiar las rutas metabólicas y la actuación de los genes en ellas [21]. De igual manera, Reactome recoge información de rutas metabólicas, aunque con una organización diferente a la de KEGG, por lo que los resultados pueden complementarse.

Existen diferentes métodos de enriquecimiento funcional. Las primeras aproximaciones fueron los análisis de sobrerrepresentación (ORA, del inglés Over-Representation Analysis), donde se parte de un conjunto de genes de interés, como pudieran ser los genes expresados diferencialmente en un experimento. Después se compara la frecuencia de las funciones anotadas de los genes de interés con la frecuencia de anotaciones en el conjunto de referencia, es decir, las funciones asociadas al resto de genes del genoma correspondiente. Para cada función se evalúa estadísticamente si la proporción de genes en esa función es mayor de lo que cabría esperar por azar, en tal caso, se diría que esa función está sobrerrepresentada o enriquecida. Los métodos ORA tienen limitaciones. Por un lado, se comete un sesgo al escoger un grupo

de genes en base a un criterio arbitrario, y por otro, a todos los genes se les da el mismo peso a la hora de calcular si la función está enriquecida [20].

Los métodos GSA (Gene Set Analysis) van un paso más allá, ya que en estos métodos se utilizan todos los genes disponibles en el experimento (microarray o RNA-seq), y además se le da un peso diferente a cada uno de ellos en función de una “puntuación biológica”, como puede ser el p-valor del análisis de expresión diferencial. Esto permite hacer un ranking de los genes de acuerdo a su expresión diferencial (donde genes con un p-valor más cercano a 0 tendrían un mayor peso). Esto se acercaría más a la realidad ya que se esperaría que módulos de genes que trabajan en una misma función celular tuvieran una expresión coordinada [20]. Además, se tiene en cuenta la “dirección” de la expresión, es decir, si los genes están sobreexpresados en la condición A o en la B, por lo que se pueden identificar qué funciones se activan y qué funciones se pierden al pasar de una condición a la otra [29].

1.9. Metaanálisis funcional

Las revisiones sistemáticas y el metaanálisis son herramientas esenciales para resumir la evidencia de manera precisa y fiable. La revisión sistemática permite reunir todos los posibles estudios relacionados por un mismo tema y diseño, mientras que el metaanálisis es la metodología estadística que permite combinar los resultados de los diferentes estudios individuales [29]. Esto permite tener una medida combinada del efecto de interés con mayor precisión que en la obtenida en estudios individuales, por lo que se incrementa la potencia estadística, pudiendo cuantificar la variabilidad de los estudios individuales [53].

Aunque los metaanálisis se emplearon inicialmente en las ciencias sociales, a partir de la década de los 80 comenzaron a usarse más y más en investigaciones médicas, por lo que también se están aplicando en la genómica [29]. El metaanálisis en genómica se ha usado en estudios de expresión génica [32, 58], o en estudios de asociación de genoma completo (GWAS) [28, 73].

Un abordaje muy interesante es realizar el metaanálisis a partir de las funciones sobrerrepresentadas en cada uno de los estudios individuales, y así poder identificar las funciones que realmente están enriquecidas en los grupos de interés. Esto además facilita la comparación entre diferentes plataformas de microarrays o RNA-seq, ya que aunque genes concretos puedan estar ausentes entre unas plataformas y otras, la acción conjunta de módulos de genes permiten identificar las funciones sobrerrepresentadas.

2. Objetivos

El objetivo de este trabajo es identificar y comprender los mecanismos moleculares asociados a las lesiones de la médula espinal mediante un abordaje in silico en estudios ómicos. Para lograrlo, se han propuesto tres objetivos específicos a cumplir:

- Revisión sistemática y selección de estudios de lesión medular con datos de microarrays o RNA-seq en los repositorios Gene Expression Omnibus (GEO), ArrayExpress.
- Análisis individual de cada estudio seleccionado, incluyendo un análisis de expresión diferencial, análisis de enriquecimiento funcional y un análisis de rutas de señalización con Hipathia.
- Metaanálisis de los resultados funcionales obtenidos a nivel de enriquecimiento funcional y rutas de señalización.

3. Materiales y métodos

A continuación se indican los pasos seguidos en el flujo de trabajo para llevar a cabo el metaanálisis funcional (figura 3.1):

1. Revisión sistemática y selección de estudios.
2. Análisis exploratorio.
3. Análisis de estudios individuales:
 - a) Análisis de expresión diferencial.
 - b) Análisis de enriquecimiento funcional (GSA).
 - c) Análisis de pathways (HiPathia).
4. Metaanálisis a nivel de función.

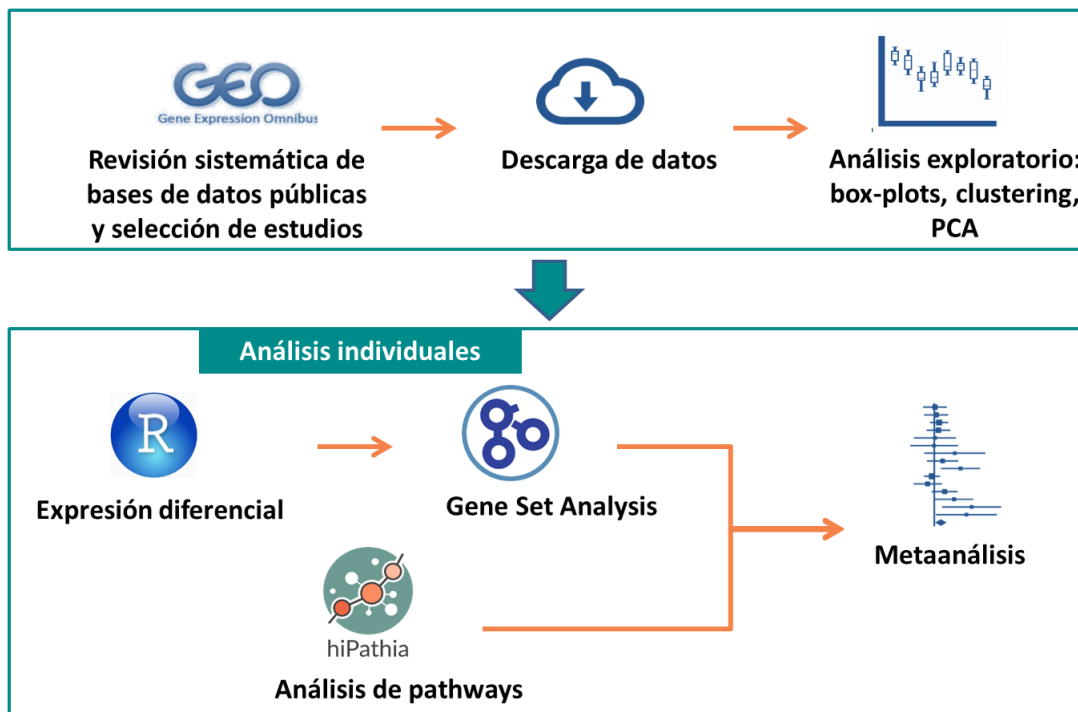


Figura 3.1: Flujo de trabajo seguido.

Todo el análisis bioinformático fue realizado mediante scripts propios con el lenguaje de programación R [65]. Las librerías o paquetes indicados en cada uno de los pasos se encuentran detalladas en la tabla B.1 del Anexo B.

3.1. Revisión sistemática y selección de estudios

Las revisiones sistemáticas permiten cotejar toda la evidencia científica que cumplen unos ciertos criterios para responder a preguntas concretas. Previamente es preciso establecer los criterios de elegibilidad y evaluar la calidad de los estudios para obtener resultados robustos. En el desarrollo del presente trabajo se han seguido las directrices y el flujo de trabajo consensuados en PRISMA (figura 3.2) [43].

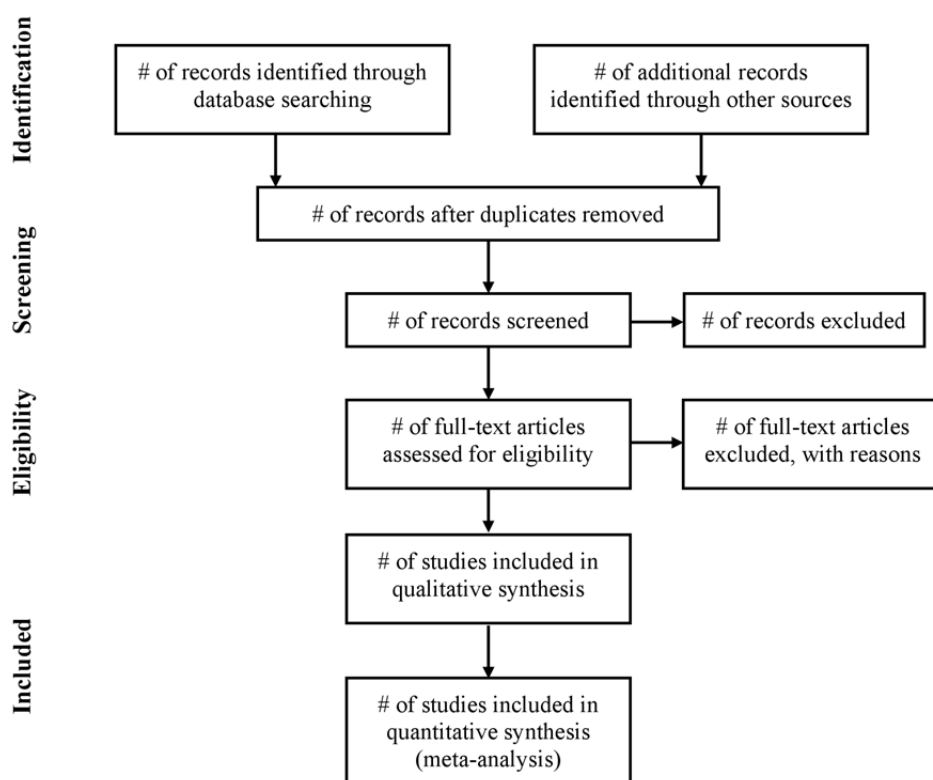


Figura 3.2: Diagrama de flujo de una revisión sistemática según la declaración PRISMA.

Los datos pueden obtenerse de diversas fuentes, como bases de datos públicas. Existen dos principales repositorios de datos transcriptómicos, GEO [8] y ArrayExpress [40], donde se realizó una búsqueda de estudios con las palabras clave “spinal cord injury”. Tras su revisión los estudios aparecían duplicados en ambas bases de datos, por lo que se trabajó con GEO

dada su mayor accesibilidad. Por otro lado, también se dispone de datos propios generados en el Laboratorio de Regeneración Tisular y Neuronal, liderado por la Dra. Victoria Moreno (datos no publicados).

De los estudios obtenidos en esta búsqueda, se descartaron aquellos donde:

- El tejido no era médula espinal.
- La metodología de análisis de expresión génica no era microarray o RNA-seq.
- No había muestras control.
- No se pudo obtener ficheros de anotación, necesarios para hacer el análisis de enriquecimiento biológico.

3.2. Grupos experimentales y análisis exploratorio

3.2.1. Establecimiento de grupos experimentales

La revisión sistemática también sirvió para poner en evidencia la diversidad de grupos experimentales que existe en los estudios, debido a la propia naturaleza de la LM y a los intereses concretos de los investigadores. Para establecer los grupos experimentales se siguieron estas consideraciones bajo la supervisión de un investigador especializado en la materia:

1. **Severidad de lesión:** se distingue entre lesiones moderadas y severas. Se consideran lesiones severas las provocadas por una contusión de una fuerza de 200 kdyn o mayor, las provocadas por el impacto de un peso de 10 g a 50 mm de altura o la retirada completa de una sección de médula espinal.
 2. **Fase de la lesión:** se establecen 4 grupos en función del tiempo que pasa desde que se provoca la lesión hasta que se recogen las muestras.
 - a) Fase aguda: de 0 a 3 días.
 - b) Fase subaguda: de 4 a 14 días.
 - c) Fase intermedia: de 15 a 30 días.
 - d) Fase crónica: más de 30 días.
-

3.2.2. Análisis exploratorio

Tras la selección de estudios que cumplieran las características de interés se procedió a la descarga de los datos, ya normalizados, para evitar la introducción de nuevos sesgos. Se realizó un análisis exploratorio con el objetivo de evaluar la distribución de las muestras en los grupos de interés y detectar comportamientos anómalos de las mismas.

Dado que en algunos de los estudios seleccionados se investiga el efecto de diferentes tratamientos en la capacidad de recuperación tras la lesión, las muestras que no son de interés para la caracterización funcional no fueron incluidas en la realización del análisis exploratorio ni en los posteriores análisis.

El análisis exploratorio consiste en la representación de los valores de expresión de cada muestra mediante diagramas de caja y análisis de agrupamiento o clustering, utilizando la distancia de correlación y la distancia euclídea, para observar posibles agrupaciones de los niveles de expresión de las muestras según el grupo experimental. Con el mismo objetivo también se realizó un análisis de componentes principales (PCA).

3.3. Análisis individual de estudios

Antes de realizar el metaanálisis se llevó a cabo una estrategia de análisis aplicada a cada estudio, que consta de análisis de expresión diferencial, análisis de enriquecimiento funcional y análisis de rutas de señalización.

3.3.1. Análisis de expresión diferencial

Dado que el objetivo es caracterizar la LM en cada una de sus fases y analizar la influencia de la severidad de la lesión, en la expresión diferencial se compararon las muestras de las diferentes fases con los controles, tanto para lesiones moderadas como para lesiones severas, de tal manera que existen 9 grupos experimentales con los que se realizaron 8 contrastes.

1. Lesiones moderadas:

1. Fase aguda vs control
2. Fase subaguda vs control
3. Fase intermedia vs control
4. Fase crónica vs control

2. Lesiones severas:

1. Fase aguda vs control
2. Fase subaguda vs control
3. Fase intermedia vs control
4. Fase crónica vs control

La expresión diferencial en microarrays se realizó con las funciones del paquete *limma*. Las matrices de diseño se construyeron con la función *model.matrix*, ajustándose un modelo lineal para cada gen con la función *lmFit*. De la matriz de contrastes obtenida con la función *makeContrasts* se calcularon los coeficientes estimados y los errores estándar con la función *contrasts.fit*, para finalmente obtener los diferentes estadísticos de la expresión diferencial con la función *eBayes*. El p-valor se ajustó mediante el método de Benjamini & Hochberg [9].

En los estudios de RNA-seq la expresión diferencial se realizó con el paquete *edgeR*. Como se partía de matrices de conteos ya normalizadas, primero se calculó el factor de normalización para hacer comparable el tamaño de las librerías con la función *calcNormFactors*, para posteriormente estimar la dispersión común a todos los genes de la muestra (*estimateCommonDisp*) y la dispersión por gen (*estimateCommonDisp*), para finalmente realizar un test exacto con la función *exactTest*, que permite comprobar si existen diferencias en la media de expresión entre dos grupos.

3.3.2. Creación de ficheros de anotación

Crear los ficheros de anotación es un paso fundamental a la hora de realizar análisis de enriquecimiento funcional, ya que estos son los que asocian los identificadores de los genes, en cualquiera de sus formas, con los identificadores de las funciones de las distintas bases de datos.

Dado que se priorizó disponer de la información del mayor número de genes con una anotación común, para cada una de las plataformas de microarray o RNA-seq se creó un fichero de anotación que asociara los identificadores de partida con los identificadores de procesos biológicos de la Gene Ontology, KEGG y Reactome.

Así, para las diferentes plataformas de microarrays se utilizaron los paquetes de anotación específicos de cada una, mientras que para RNA-seq se utilizaron los identificadores de ENSEMBL para asociar los genes con las funciones.

3.3.3. Análisis de enriquecimiento funcional

A partir de los resultados de expresión diferencial se realizó el análisis de enriquecimiento funcional con el método Gene-Set Analysis (GSA) incluido en el paquete *mdGSA* [48]. Todas las funciones necesarias en este análisis pertenecen a este paquete. Los análisis de enriquecimiento funcional se realizaron para cada contraste con las anotaciones de los procesos biológicos de la Gene Ontology [6] y las rutas de KEGG [37] y Reactome [70].

Para ello, con la función *pval2index* primero se crea un índice que combina el p-valor con el signo del logFC o el estadístico t de cada gen del estudio, obteniéndose un índice que combina la fuerza de la expresión diferencial con el sentido o dirección. Por tanto, se clasifican los genes en base a su expresión diferencial, desde los más sobreexpresados a los más infraexpresados. A continuación, con la función *indexTransform* se transforma este índice para que su distribución sea adecuada como variable independiente de un modelo de regresión logística univariante. Finalmente, la función *uvGsa* realiza el GSA, tomando como input el ranking de genes y un fichero de anotación. Este fichero de anotación previamente se convierte en una lista que asocia cada función con los genes, y se eliminan aquellos módulos formados por menos 10 genes o más de 500 para evitar la selección de módulos poco informativos, bien por ser demasiado pequeños como para considerarlos módulos de genes, o bien por ser demasiado grandes como para ser específicos de algún proceso biológico [48]. Se consideraron significativos los resultados con un p-valor ajustado con el método de Benjamini-Yekutieli [10] menor a 0.05.

3.3.4. Análisis de rutas de señalización

Hipathia es un método que permite calcular la señal de transducción a lo largo de rutas metabólicas a partir de datos transcriptómicos. Se basa en un algoritmo iterativo que calcula la intensidad de la señal que pasa a través de los nodos (genes) de una red, teniendo en cuenta el nivel de expresión de cada gen y la intensidad de la señal que le llega. Esta aproximación permite estimar la señal final que llega a cada nodo efector de una ruta e identificar si las funciones asociadas a los nodos efectores se han activado o se han inhibido [35].

Las rutas que se incluyen en Hipathia provienen de la base de datos de KEGG. Como una misma ruta puede estar asociada a varios nodos efectores con diferentes funciones, en Hipathia se analiza cada nodo efector como una subruta independiente. Las subrutas se componen de un nodo efector final y todos los nodos que llevan a él.

El análisis con Hipathia parte de una matriz de expresión donde las filas son genes cuyo identificador debe traducirse a Entrez (identificadores para genes en la base de datos del NCBI). Posteriormente hay que escalar los valores de expresión para que estén entre 0 y 1 y se calcula el nivel de activación de cada subruta para cada muestra con la función *hiapathia*. Finalmente, se hace un contraste con *limma* para comparar los niveles de activación de cada subruta entre dos grupos experimentales de interés, y comprobar así la ruta se ha activado o inhibido.

3.4. Metaanálisis de resultados con GSA

A partir de los resultados de los análisis de enriquecimiento funcional se realizó un metaanálisis para cada una de las funciones obtenidas en cada estudio individual, tanto para términos GO como para rutas de KEGG y Reactome. Para ello se utilizan los logaritmos de odds ratios (LOR) de cada función y la varianza de estos. Los LOR aportan una medida del efecto entre dos condiciones. Con ello se construyen los datos de entrada del metaanálisis, habiendo que transformarlos en una matriz de LOR y otra matriz con las varianzas de LOR.

3.4.1. Determinación de la medida combinada del efecto

Para cada una de las funciones se realiza un metaanálisis que combina el efecto medido de todos los estudios analizados. Generalmente, en los metaanálisis primero se evalúa la heterogeneidad de los estudios y en función de ello se decide el modelo de estimación de la variabilidad del efecto. Sin embargo, al trabajar con miles de metaanálisis simultáneamente (uno por función), primero se computa la estimación de la variabilidad del efecto con diferentes métodos, y después se valoran los resultados del estudio de heterogeneidad para seleccionar el mejor método [29]. La heterogeneidad hace referencia a la variabilidad en los efectos evaluados, debida a la propia diversidad de los datos y a la de las metodologías empleadas en los estudios.

Los metaanálisis se realizaron con las funciones necesarias del paquete metafor [71]. Se seleccionó el modelo DerSimonian y Laird (DL) [16] para evaluar la medida del efecto, ya que es un modelo de efectos aleatorios que permite incorporar la variabilidad interestudio en el estimador combinado [33]. Esto es necesario por las diferencias que existen entre los estudios, como las diferentes tecnologías y plataformas utilizadas para analizar la expresión génica.

El resultado del metaanálisis es un conjunto de estimadores de la medida del efecto combinado para cada función e indicadores de la heterogeneidad del metaanálisis [29]:

- QE y QEp son el estadístico de contraste y p-valor, respectivamente, del método DL utilizado para detectar la heterogeneidad entre estudios. La hipótesis nula indica la no heterogeneidad entre los estudios.
- LOR es la estimación del efecto combinado de los estudios. El signo indica la sobrerrepresentación de una determinada función en uno de los dos grupos comparados. Si el símbolo es negativo la función estará sobrerrepresentada en el grupo control. El valor de LOR es un indicador cuantitativo de dicha sobrerrepresentación.
- SE es el intervalo de confianza del LOR (al 95%), construido a partir de la variabilidad estimada en el modelo seleccionado. Si no incluye el 0, el LOR se considera significativo.
- El p-valor indica el nivel de significación del LOR. Dado que el p-valor no contempla el escenario de multiplicidad del análisis (múltiples funciones) es necesario una corrección del mismo, que se realizó con el método de Benjamini-Hochberg.
- τ^2 describe la heterogeneidad entre los estudios. Será 0 en modelos de efectos fijos.
- I^2 indica el porcentaje de variabilidad entre estudios y el total de la variabilidad.
- H^2 es el cociente entre la variabilidad total y la variabilidad en el muestreo.

3.4.2. Evaluación de la heterogeneidad

El test Q propuesto por DerSimonian y Laird presenta una baja potencia cuando el número de estudios es pequeño, por lo que el análisis de heterogeneidad se complementa con varios métodos que se resumen gráficamente en gráficos de embudo. Los gráficos de embudo evalúan la variabilidad de los diferentes estudios y la presencia de sesgos. La figura 3.3 muestra un ejemplo de gráficos de embudo. Se representa la magnitud del efecto medido (LOR) en el eje X frente a una medida de precisión (eje Y), como la desviación estándar o el inverso de la varianza [29].

3.4.3. Representación de resultados

Para cada una de las funciones del metaanálisis se generaron gráficos de bosque. Estos permiten evaluar los resultados globales y el peso que aporta cada estudio individual en una función determinada.

En los gráficos de bosques se enumeran los estudios individuales incluidos en el metaanálisis a la izquierda, mientras que a la derecha se indica la estimación de la medida resumen de cada estudio con el intervalo de confianza al 95%. En el centro de gráfico se representa con cuadrados negros el valor de LOR con el intervalo de confianza a sus extremos. El tamaño del cuadro es proporcional a la precisión de las estimaciones, por lo que una mayor variabilidad se visualiza implica un menor tamaño del cuadro. En la parte inferior se indica en modelo de análisis empleado a la izquierda, y a la derecha el resultado del efecto conjunto (LOR) con el intervalo de confianza. El rombo de la parte inferior central representa este valor y la anchura se corresponde con el intervalo de confianza (figura 3.4).

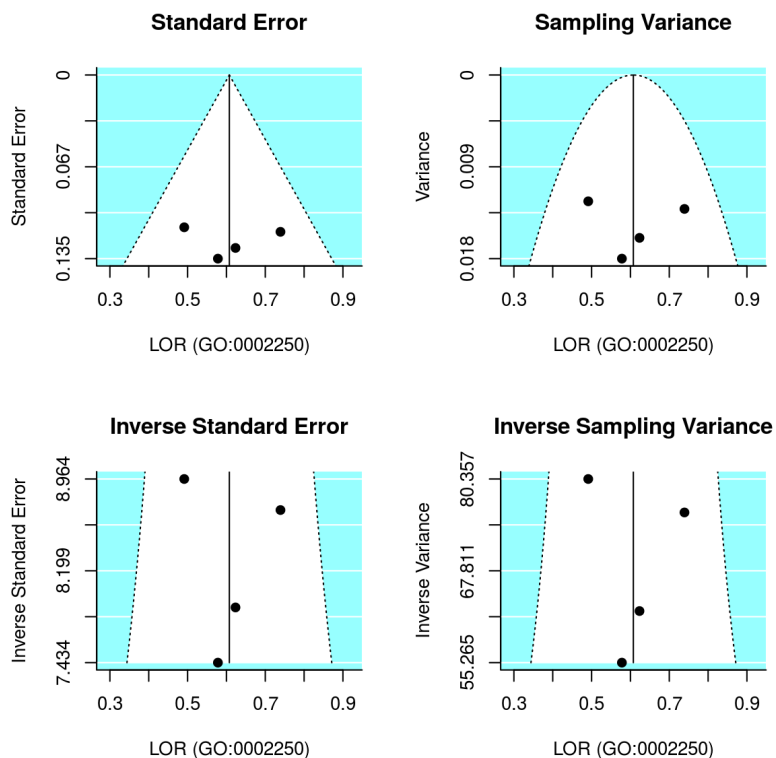


Figura 3.3: Ejemplo de gráficos de embudo para la representación de la heterogeneidad de un metaanálisis. En este ejemplo, para la función GO:0002250 hay 4 estudios, todos ellos dentro del intervalo de confianza.

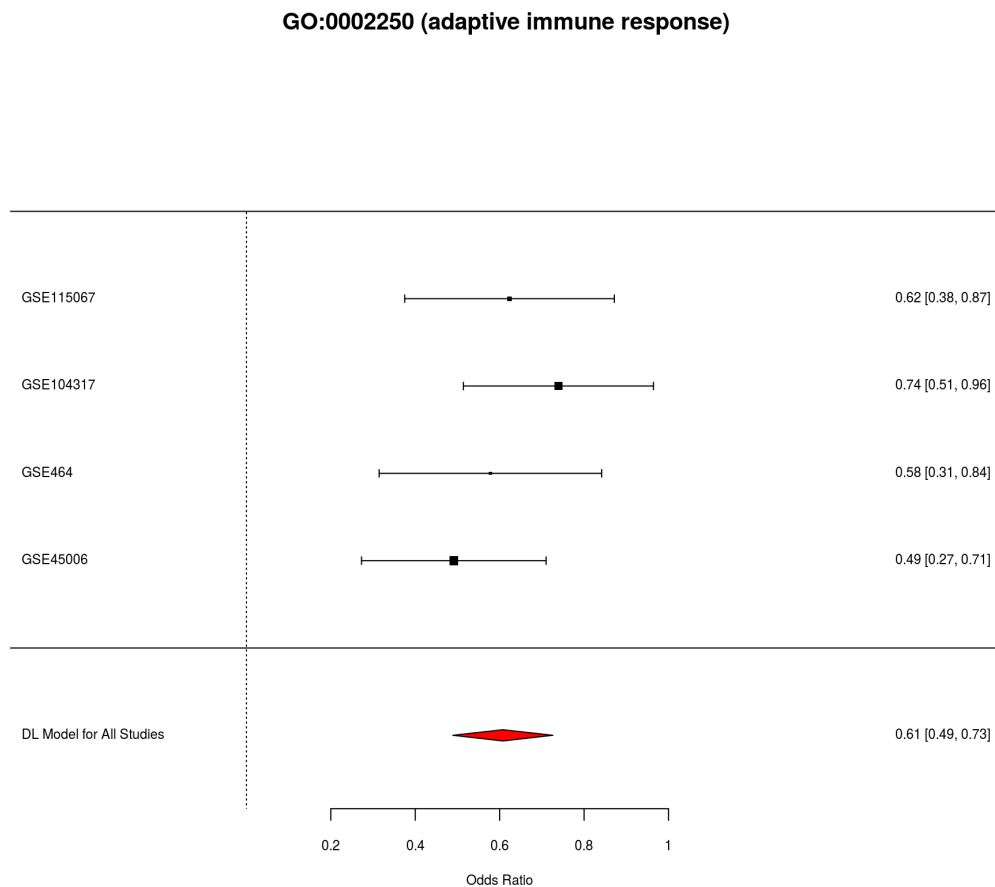


Figura 3.4: Ejemplo de un gráfico de bosque. Gráfico de bosque para la función GO:0002250 hay 4 estudios. Los odds ratio del eje X se encuentran en escala logarítmica.

3.5. Metaanálisis de resultados de Hipathia

Dado que la naturaleza de los resultados obtenidos en el análisis de rutas de señalización con Hipathia (datos cuantitativos) son diferentes a los que se obtienen en el enriquecimiento funcional (datos categóricos), el metaanálisis se realiza de manera distinta.

Al igual que en el caso anterior, se necesitan dos matrices. En lugar de generar una matriz con LOR y otra con las varianzas de estos, se genera una matriz con el valor del estadístico t de *limma* y otra con los p-valores ajustados de los mismos. Una vez obtenidas las matrices se realiza un filtro, en el que se exige que una determinada ruta efectora tenga el mismo signo en el estadístico t en todos los estudios. Finalmente, en las subrutinas que queden se calcula un p-valor combinado mediante la función *sump* del paquete *metap* [18], considerándose significativas aquellas que tengan un p-valor combinado menor a 0.05.

4. Resultados

En este apartado se muestran los resultados obtenidos, siguiendo el orden establecido en Materiales y métodos.

4.1. Revisión sistemática y selección de estudios

Introduciendo las palabras clave “*spinal cord injury*” en GEO se identificaron 25 estudios candidatos para *R. norvegicus*. De estos se excluyeron 10 por ser un tejido diferente a médula espinal, 3 por ser plataformas diferentes a microarrays de expresión o RNA-seq, uno se descartó por no tener grupos control y otro por no poder obtenerse ficheros de anotación. Finalmente, a estos 10 estudios restantes se añadieron los datos propios obtenidos en el laboratorio de la Dra. V. Moreno, los cuales reunían los requisitos de elegibilidad, incluyéndose finalmente 11 estudios en el metaanálisis (tabla 4.1 y figura 4.1).

Tabla 4.1: Estudios seleccionados tras la revisión sistemática. Se indica la plataforma utilizada en cada estudio y las publicaciones asociadas.

Estudio (GEO ID)	Plataforma	Artículo
GSE115067	Illumina NextSeq 550	[64]
GSE102964	Affymetrix Rat Gene 2.0 ST Array	[63]
GSE104317	Affymetrix Rat Genome 230 2.0	[7]
GSE93249	Illumina NextSeq 2000	[24]
GSE52763	Affymetrix Rat Genome 230 2.0	[61]
GSE45006	Affymetrix Rat Genome 230 2.0	[13]
GSE46988	Affymetrix Rat Gene 1.1 ST	[67]
GSE29488	Affymetrix Rat Genome 230 2.0	[22]
GSE2599	Affymetrix Rat Genome 230 2.0	[2]
GSE464	Affymetrix Rat Genome U34	[19]
VMM	Agilent-011868 Whole Rat Genome Microarray (G2519F)	No publicado

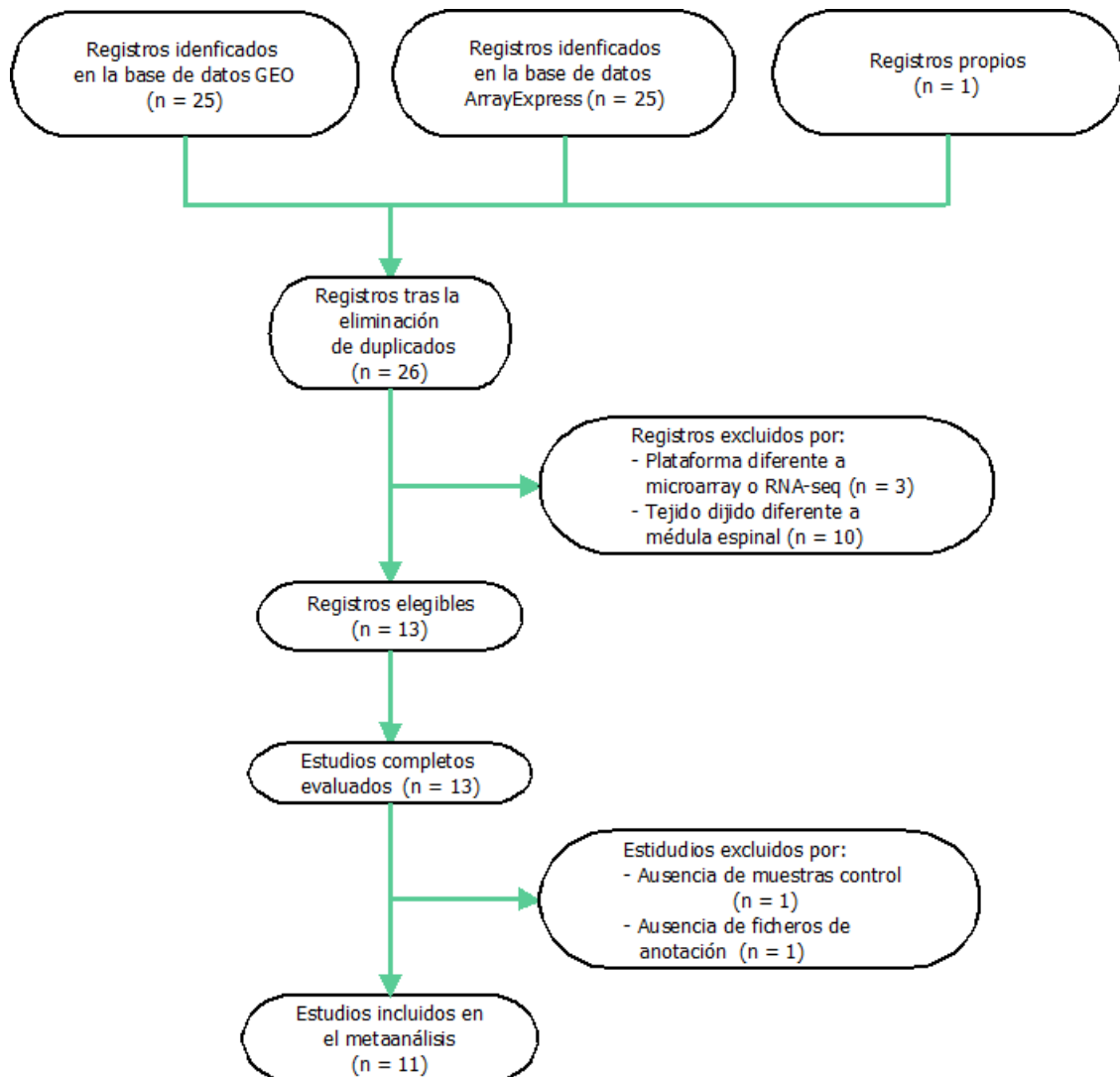


Figura 4.1: Diagrama de flujo de la revisión sistemática, siguiendo la declaración PRISMA.

4.2. Grupos experimentales y análisis exploratorio

Tras establecer los grupos experimentales según los criterios de interés en los diferentes estudios, se procedió al análisis exploratorio, donde se evalúa la normalización de las matrices de expresión mediante diagramas de cajas y el agrupamiento de las muestras mediante análisis de clustering y PCA.

En general se espera una distribución más o menos uniforme al visualizar los niveles de expresión de las diferentes muestras de un estudio con diagramas de caja. En cuanto al análisis de clustering y PCA, se espera que las muestras de un mismo grupo experimental se agrupen entre sí, y que las pertenecientes a fases más tempranas de la lesión se diferencien más de los controles que las pertenecientes a fases más tardías. En algunos casos se da este comportamiento (figura 4.3a), mientras que en otros se observan muestras de diferentes grupos experimentales que se agrupan entre sí (figura 4.3b). Esto puede ser debido a la calidad de las muestras, o bien a que el paso de una fase a otra de la LM es gradual, por lo que siguen compartiendo patrones de expresión entre sí. Los gráficos con el análisis de clustering y PCA de cada estudio pueden visualizarse en el anexo A. Para una mejor visualización, todos los gráficos, incluyendo diagramas de cajas, pueden encontrarse en el material suplementario alojado en el siguiente enlace: <http://bioinfo.cipf.es/ubb/wiki/start/metafun/sci/start>.

A continuación se describe brevemente el objetivo de cada estudio y se detalla el número de muestras y grupos experimentales identificados en cada estudio, que se resumen en la tabla 4.2.

Tabla 4.2: Número de muestras seleccionadas en cada estudio. Se presentan distribuidas por severidad y fase de la lesión. **Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica.

Estudio (GEO ID)	Control	Lesión moderada				Lesión severa			
		Agu	Sub	Int	Cro	Agu	Sub	Int	Cro
GSE115067	5	-	5	-	-	-	5	-	-
GSE102964	4	-	-	-	4	-	-	-	-
GSE104317	4	4	4	-	-	-	-	-	-
GSE93249	3	-	-	3	5	-	-	-	-
GSE52763	6	-	-	-	-	-	4	8	-
GSE45006	4	8	8	-	4	-	-	-	-
GSE46988	4	-	-	-	-	4	12	-	-
GSE29488	3	-	-	-	-	15	9	18	18
GSE2599	3	-	-	3	-	-	-	-	-
GSE464	17	24	19	9	-	8	7	4	-
VMM	4	-	-	-	-	-	8	4	4
Total	57	36	36	15	13	27	45	34	22

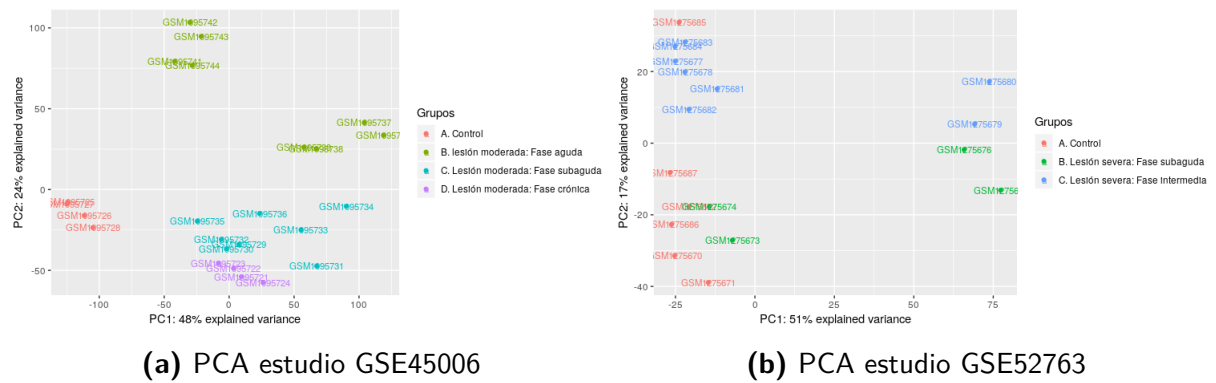


Figura 4.2: Ejemplos de PCA Las figuras muestran PCAs donde el color indica el grupo experimental. Se compara la componente principal I con la componente principal II. En la imagen (a) los grupos experimentales se separan siguiendo el comportamiento esperado, mientras que en la imagen (b) hay mezcla de muestras entre diferentes grupos.

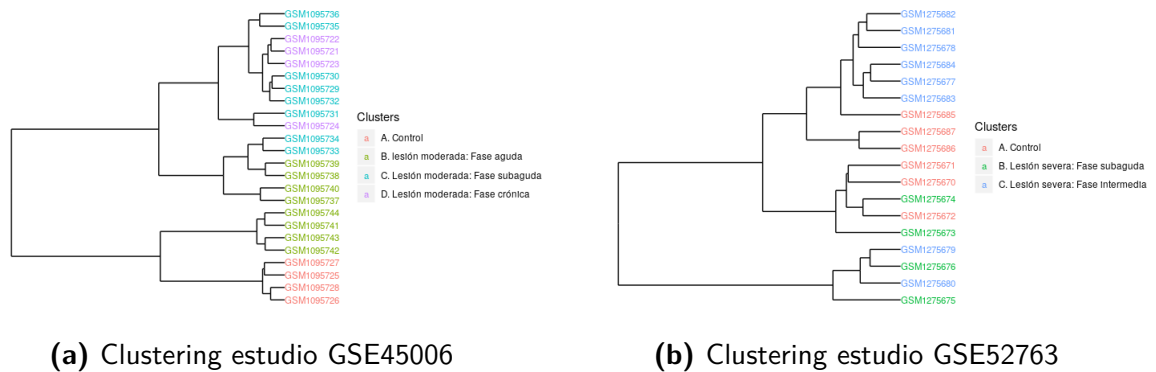


Figura 4.3: Ejemplos de clustering exploratorio. El color indica el grupo experimental. En la imagen (a) los grupos experimentales se agrupan siguiendo el comportamiento esperado, mientras que en la imagen (b) hay mezcla de muestras entre diferentes grupos.

▪ **GSE115067**

Este estudio se realiza un análisis integrativo, reuniendo información procedente de transcriptómica, proteómica y genes asociados a la respuesta de la LM descritos en la literatura. Además generan datos propios, que incluyen controles (n=5) y muestras pertenecientes a la fase subaguda, tanto para lesión moderada (n=5) como para lesión severa (n=5).

- **GSE102964**

Este estudio analiza el impacto de una dieta alta en grasas en la LM. Dado que esto no influye en muchos de los mecanismos que ocurren tras la lesión, se incluyen en el metaanálisis muestras pertenecientes a la fase crónica con una lesión moderada (n=4) y controles (n=4).

- **GSE104317**

En este estudio se examina el efecto del tratamiento con el complejo albúmina-ácido hidroxioleico en la recuperación de la lesión. Dado que hay muestras que solo han sido tratadas con un vehículo salino, pueden ser incorporadas en el metaanálisis. Así, se han seleccionado las muestras control (n=4) y muestras pertenecientes a la fase aguda (n=4) y subaguda (n=4), con lesiones moderadas.

- **GSE93249**

Este estudio analiza la expresión génica en fases tardías de la LM. Se incluyen controles (n=3) y muestras de la fase intermedia (n=3) y crónica (n=5), con lesiones moderadas.

- **GSE52763**

Este estudio analiza la expresión génica en distintas fases de la lesión y cómo afecta el entrenamiento en cintas de correr en la recuperación de la lesión. Dado que el entrenamiento no afecta a algunos de los procesos que ocurren tras la lesión, como la respuesta inflamatoria y que suele ser una práctica habitual en la investigación de la LM en ratas las muestras de ratas con entrenamiento también fueron seleccionadas. Así, se incluyeron muestras control (n=6) y muestras de la fase subaguda (n=4) e intermedia (n=8), con lesiones severas.

- **GSE45006**

Este estudio busca caracterizar la LM en sus diferentes fases mediante los perfiles de expresión génica. Se incluyen controles (n=4) y muestras de la fase aguda (n=8), subaguda (n=4) y crónica (n=4), con lesiones moderadas.

- **GSE46988**

Este estudio analiza cómo afecta trasplantes de células mesenquimales estromales y células olfativas envainadas en la recuperación tras la lesión. Se seleccionaron las muestras control (n=4) y las que no recibieron ningún trasplante, correspondientes a la fase aguda (n=4) y subaguda (n=12), todas ellas con lesiones severas.

- **GSE29488**

Este estudio analiza la expresión génica en ratas a las que se les ha sustituido una parte de la médula espinal por un tubo de quitosano (un biomaterial) con neurotrofina-3 (una proteína que activa el crecimiento y diferenciación de neuronas). Se han incluido las muestras control (n=3), y las muestras provenientes de ratas a las que se les ha retirado una parte de la médula espinal y no han tenido ningún tratamiento, considerándose una lesión severa. Estas pertenecen a la fase aguda (n=15), subaguda (n=9), intermedia (n=18) y crónica (n=18).

- **GSE2599**

Este estudio analiza la expresión génica en diferentes fases de la lesión. Sin embargo solo están disponibles las muestras control (n=3) y muestras pertenecientes a la fase intermedia (n=3) con una lesión moderada.

- **GSE464**

Este estudio analiza cómo cambian los perfiles de expresión génica a lo largo del tiempo, tras provocar lesiones de diferentes severidades, tomando muestras en distintas regiones en función del punto donde se provoca la lesión. Se han seleccionado las muestras control y las pertenecientes a la fase aguda, subaguda e intermedia, tanto para lesiones moderadas como para lesiones severas, provocadas en el epicentro de la lesión. El número exacto de muestras de cada uno de los grupos experimentales se detalla en la tabla 4.2. Aunque se han usado tres plataformas diferentes de microarrays, que conforman el genoma completo de rata (RG-U34A, RG-U34B, RG-U34C), solo se ha trabajado con la plataforma RG-U34A. Esto se debe por un lado a que es la que contiene la mayoría de genes conocidos y anotados, y por otro, a que el número de muestras disponibles difería entre las diferentes plataformas. Dado que los metadatos no contenían información que permitiera asociar las plataformas con el origen de las muestras, no fue posible unificar los valores de expresión en una única matriz.

- **VMM**

En este estudio se analiza la expresión génica en diferentes fases de la lesión y cómo afecta el trasplante de células endimarias en la recuperación. Se seleccionaron las muestras control (n=4) y las que no recibieron el trasplante, correspondientes a la fase subaguda (n=8), fase intermedia (n=4) y fase crónica (n=4), todas ellas con lesiones severas.

4.3. Análisis individual de estudios

4.3.1. Análisis de expresión diferencial

A continuación se muestra el número de genes expresados diferencialmente (p -valor ajustado menor a 0.05) en las diferentes comparaciones de cada estudio. Los resultados se muestran agrupados en Up si el logFC es mayor a 0, y en Down si es menor a 0 (tabla 4.3).

Tabla 4.3: Resumen de resultados del análisis de expresión diferencial. Se realizaron los contrastes para lesiones moderadas y severas comparando las diferentes fases de la lesión con los controles. **C**, control; **Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica. Los resultados se muestran agrupados en *Up* si el logFC es mayor a 0, y en *Down* si es menor a 0.

Estudio	Sig	Lesión moderada				Lesión severa			
		Agu-C	Sub-C	Int-C	Cro-C	Agu-C	Sub-C	Int-C	Cro-C
GSE115067	Up	-	1799	-	-	-	916	-	-
	Down	-	3242	-	-	-	1961	-	-
GSE102964	Up	-	-	-	1028	-	-	-	-
	Down	-	-	-	838	-	-	-	-
GSE104317	Up	5288	5101	-	-	-	-	-	-
	Down	5556	6094	-	-	-	-	-	-
GSE93249	Up	-	-	811	1024	-	-	-	-
	Down	-	-	1628	1628	-	-	-	-
GSE52763	Up	-	-	-	-	-	326	900	-
	Down	-	-	-	-	-	72	560	-
GSE45006	Up	4999	3630	-	2747	-	-	-	-
	Down	5136	4296	-	3157	-	-	-	-
GSE46988	Up	-	-	-	-	3928	4927-	-	-
	Down	-	-	-	-	4890	6330	-	-
GSE29488	Up	-	-	-	-	4367	3782	4757	3732
	Down	-	-	-	-	5033	5297	6196	5098
GSE2599	Up	-	-	375	-	-	-	-	-
	Down	-	-	569	-	-	-	-	-
GSE464	Up	589	804	311	-	344	725	532	-
	Down	364	1100	2301	-	106	637	1068	-
VMM	Up	-	-	-	-	-	8050	7083	7694
	Down	-	-	-	-	-	10796	8758	9337

4.3.2. Análisis de enriquecimiento funcional

A continuación se muestran los resultados del análisis de enriquecimiento funcional con GSA para los procesos biológicos de la Gene Ontology, rutas KEGG y rutas Reactome, para cada una de las comparaciones realizadas en los diferentes estudios. Se indica el número de funciones significativas (p-valor ajustado menor a 0.05), agrupadas según su LOR. Se etiquetan con *Enriched* las funciones que están enriquecidas o sobrerrepresentadas en los distintos grupos con lesión (LOR mayor a 0), mientras que *Depleted* hace referencia a las funciones empobrecidas, es decir, funciones que se han perdido tras la lesión o que están sobrerrepresentadas en los controles (LOR menor a 0). En la tabla 4.4 y 4.5 se resumen los resultados para lesiones moderadas y severas, respectivamente.

Tabla 4.4: Resumen de resultados del análisis de enriquecimiento funcional para lesiones moderadas Los resultados se agrupan por fase de la lesión. Los términos *Enriched* (**Enrich**) y *Depleted* (**Deple**) hacen referencia al grupo de la comparación en el que las funciones están sobrerrepresentadas, *Enriched* en casos y *Depleted* en controles.

	Lesión moderada								
	Fase aguda vs control								
	GO			KEGG			Reactome		
	Enrich	Deple	Total	Enrich	Deple	Total	Enrich	Deple	Total
GSE464	254	10	264	53	15	68	139	15	154
GSE45006	407	165	572	77	18	95	358	71	429
GSE104317	429	41	470	78	6	84	359	24	383
	Fase subaguda vs control								
GSE464	158	81	239	45	20	65	141	31	172
GSE45006	361	159	520	70	22	92	274	106	380
GSE104317	531	71	602	80	13	93	341	41	382
GSE115067	357	87	444	58	17	75	152	60	212
	Fase intermedia vs control								
GSE464	72	51	123	27	6	33	83	24	107
GSE2599	133	58	192	32	21	53	62	66	128
GSE93249	211	15	226	51	13	64	79	35	114
	Fase crónica vs control								
GSE45006	265	147	412	59	20	79	151	97	248
GSE93249	265	33	298	44	9	53	71	56	127
GSE102964	298	103	401	51	17	68	128	90	218

Tabla 4.5: Resumen de resultados del análisis de enriquecimiento funcional para lesiones severas Los resultados se agrupan por fase de la lesión. Los términos *Enriched (Enrich)* y *Depleted (Deple)* hacen referencia al grupo de la comparación en el que las funciones están sobrerrepresentadas, *Enriched* en casos y *Depleted* en controles.

Lesión severa									
Fase aguda vs control									
	GO			KEGG			Reactome		
	Enrich	Deple	Total	Enrich	Deple	Total	Enrich	Deple	Total
GSE464	160	31	191	26	21	47	30	27	57
GSE29488	194	42	236	84	5	89	368	23	391
GSE46988	274	181	455	52	33	85	260	70	330
Fase subaguda vs control									
GSE464	147	88	235	47	17	64	125	34	159
GSE29488	549	56	605	76	8	84	202	38	240
GSE46988	296	174	470	57	24	81	260	70	330
GSE52763	24	49	73	14	15	29	67	47	114
GSE115067	344	141	485	61	28	89	198	90	288
VBB	373	73	446	68	12	80	214	50	264
Fase intermedia vs control									
GSE464	16	57	73	8	7	15	30	18	48
GSE29488	592	74	666	80	7	87	286	31	317
GSE52763	63	64	127	29	18	47	41	138	179
VBB	335	63	398	67	10	77	133	44	177
Fase crónica vs control									
GSE29488	564	47	611	73	6	79	195	27	222
VBB	375	67	442	66	9	75	123	43	166

4.3.3. Análisis de rutas de señalización

A continuación se muestran los resultados del análisis de rutas de señalización con Hipathia, indicándose el número de subrutas significativas (FDR menor a 0.05) que hay en los grupos experimentales correspondientes de cada estudio. Los resultados se muestran agrupados en *Up* si las subrutas están diferencialmente activadas en los casos y *Down* si están diferencialmente activadas en controles (tabla 4.6).

Tabla 4.6: Resumen de resultados del análisis de rutas de señalización con Hipathia. Se realizaron los contrastes para lesiones moderadas y severas comparando las diferentes fases de la lesión con los controles. **C**, control; **Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica. Los resultados se muestran agrupados en *Up* si las subrutas están diferencialmente activadas en los casos y *Down* si están diferencialmente activadas en controles.

Estudio	Sig	Lesión moderada				Lesión severa			
		Agu-C	Sub-C	Int-C	Cro-C	Agu-C	Sub-C	Int-C	Cro-C
GSE115067	Up	-	5	-	-	-	0	-	-
	Down	-	0	-	-	-	0	-	-
GSE102964	Up	-	-	-	490	-	-	-	-
	Down	-	-	-	182	-	-	-	-
GSE104317	Up	243	187	-	-	-	-	-	-
	Down	124	638	-	-	-	-	-	-
GSE93249	Up	-	-	94	137	-	-	-	-
	Down	-	-	76	598	-	-	-	-
GSE52763	Up	-	-	-	-	-	65	80	-
	Down	-	-	-	-	-	18	42	-
GSE45006	Up	163	161	-	332	-	-	-	-
	Down	245	216	-	695	-	-	-	-
GSE46988	Up	-	-	-	-	400	321	-	-
	Down	-	-	-	-	693	273	-	-
GSE29488	Up	-	-	-	-	436	87	365	192
	Down	-	-	-	-	164	104	822	306
GSE2599	Up	-	-	121	-	-	-	-	-
	Down	-	-	116	-	-	-	-	-
GSE464	Up	556	465	5	-	48	279	20	-
	Down	90	75	5	-	26	104	0	-
VMM	Up	-	-	-	-	-	704	693	730
	Down	-	-	-	-	-	280	281	352

4.4. Metaanálisis

Todos los ficheros con los resultados detallados del metaanálisis pueden consultarse en el [material suplementario](#). Igualmente, las figuras que se muestran a continuación están también disponibles para una mejor visualización.

4.4.1. Enriquecimiento funcional

El metaanálisis funcional se realizó de manera independiente con los términos GO para procesos biológicos, rutas KEGG y rutas Reactome. En los próximos apartados se muestran en tablas el número de funciones significativas obtenidas para cada base de datos utilizada (p-valor ajustado menor a 0.05). También se indica el número de funciones enriquecidas o sobrerrepresentadas en los grupos lesionados (*Enriched*), si el valor del LOR es mayor a 0, y el número de funciones empobrecidas o sobrerrepresentadas en los controles (*Depleted*), si el valor del LOR es menor a 0. Por otra parte, se incluyen diagramas de Venn que indican el número de funciones comunes y específicas de las diferentes fases de la lesión, así como gráficos que permiten visualizar de manera conjunta las funciones más sobrerrepresentadas de las diferentes fases de la lesión.

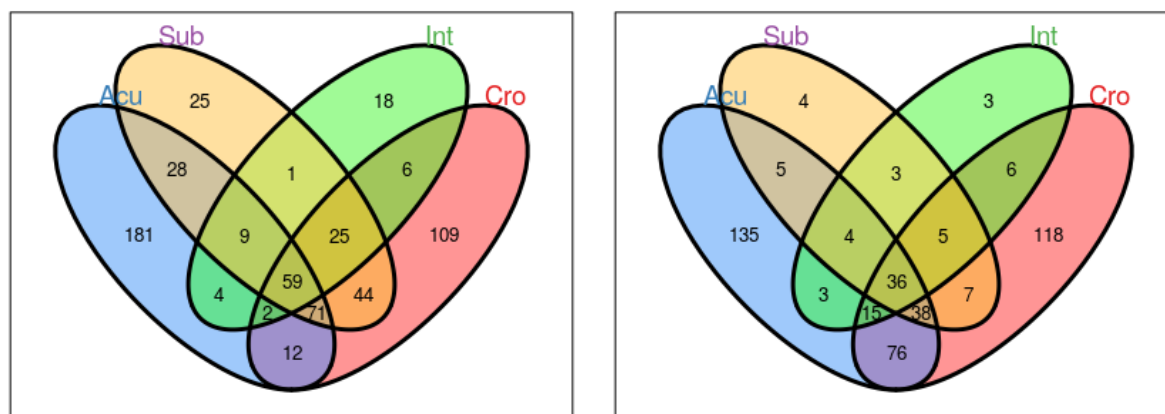
4.4.1.1. Términos GO para procesos biológicos

En la tabla 4.7 se resumen los resultados obtenidos en el metaanálisis de términos GO con procesos biológicos, para cada uno de los contrastes realizados. La figura 4.4, muestra en diagramas de Venn la intersección de funciones entre las diferentes fases, tanto para lesiones moderadas como para lesiones severas.

La tabla 4.8 es un ejemplo de los resultados brutos obtenidos en el metaanálisis. Se detalla el valor de los diferentes estimadores de la medida del efecto y heterogeneidad para las 10 funciones con mayor LOR en valor absoluto obtenidas en el metaanálisis de la fase aguda con lesión moderada.

Tabla 4.7: Resultados del metaanálisis para procesos biológicos de la Gene Ontology. *Enriched* y *Depleted* hacen referencia a las funciones que presentan un p-valor ajustado menor a 0.05, y con un LOR mayor o menor a 0, respectivamente.

	Lesión moderada			Lesión severa		
	Enriched	Depleted	Total	Enriched	Depleted	Total
Fase aguda vs control	326	40	366	263	49	312
Fase subaguda vs control	211	51	262	77	25	102
Fase intermedia vs control	99	25	124	51	24	75
Fase crónica vs control	248	80	328	274	27	301



(a) Lesión moderada

(b) Lesión severa

Figura 4.4: Diagramas de Venn de los resultados del metaanálisis con términos GO. Se muestra la intersección de las funciones obtenidas en las cuatro fases de la LM (**Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica), a la izquierda en lesiones moderadas **(a)**, y a la derecha en lesiones severas **(b)**.

Tabla 4.8: Selección de los 10 términos GO con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas. Se muestran los estimadores de la medida de efecto y los indicadores de la heterogeneidad.

ID	Nombre	LI 95%	LOR	LR 95%	FDR
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	1.309	1.617	1.924	0
GO:1900454	positive regulation of long-term synaptic depression	-1.711	-1.438	-1.165	0
GO:1900264	positive regulation of DNA-directed DNA polymerase activity	1.082	1.434	1.786	0
GO:1904851	positive regulation of establishment of protein localization to telomere	1.145	1.413	1.681	0
GO:0042254	ribosome biogenesis	1.098	1.41	1.721	0
GO:0000028	ribosomal small subunit assembly	1.116	1.397	1.678	0
GO:0043279	response to alkaloid	1.085	1.384	1.682	0
GO:0071223	cellular response to lipoteichoic acid	1.045	1.377	1.709	0
GO:0000027	ribosomal large subunit assembly	1.174	1.376	1.578	0
GO:0000463	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	1.117	1.365	1.613	0

ID	QE	QE _p	SE	tau2	I2	H2
GO:0002755	0.444	0.505	0.157	0	0	1
GO:1900454	0.419	0.517	0.139	0	0	1
GO:1900264	0.223	0.637	0.18	0	0	1
GO:1904851	0.521	0.471	0.137	0	0	1
GO:0042254	1.728	0.189	0.159	0.022	42.145	1.728
GO:0000028	1.175	0.556	0.143	0	0	1
GO:0043279	1.062	0.588	0.152	0	0	1
GO:0071223	3.268	0.195	0.169	0.033	38.809	1.634
GO:0000027	0.25	0.617	0.103	0	0	1
GO:0000463	0.919	0.338	0.126	0	0	1

Desde el punto de vista biológico, resulta de interés a la hora de interpretar los resultados visualizar de manera resumida las funciones más significativas de cada fase. Por ello se han realizado diagramas de puntos en los que se representan las funciones con un mayor LOR, en valor absoluto. Así, para cada contraste puede identificarse al mismo tiempo las funciones comunes y específicas, además de las funciones enriquecidas y empobrecidas. En la figura 4.5 se visualiza el diagrama de puntos para términos GO enriquecidos en lesiones moderadas. El diagrama de puntos para lesiones severas puede encontrarse en el anexo A.13.



Figura 4.5: Resumen de términos GO con mayor LOR en lesiones moderadas. Para cada fase se muestran funciones enriquecidas (*Enriched*) y funciones empobrecidas (*Depleted*). El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos.

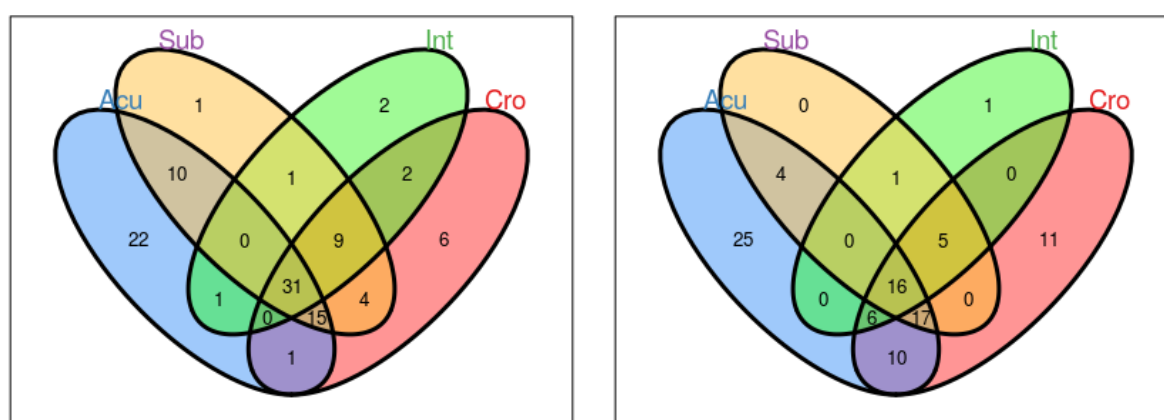
4.4.1.2. Rutas KEGG

En la tabla 4.9 se resumen los resultados obtenidos en el metaanálisis con rutas KEGG, para cada uno de los contrastes realizados.

Tabla 4.9: Resultados del metaanálisis para rutas KEGG. *Enriched* y *Depleted* hacen referencia a las funciones que presentan un p-valor ajustado menor a 0.05, y con un LOR mayor o menor a 0, respectivamente.

	Lesión moderada			Lesión severa		
	Enriched	Depleted	Total	Enriched	Depleted	Total
Fase aguda vs control	70	10	80	56	22	78
Fase subaguda vs control	58	13	71	36	7	43
Fase intermedia vs control	34	12	46	25	4	29
Fase crónica vs control	52	16	68	60	5	65

La figura 4.6, muestra en diagramas de Venn la intersección de funciones entre las diferentes fases, tanto para lesiones moderadas como para lesiones severas.



(a) Lesión moderada

(b) Lesión severa

Figura 4.6: Diagramas de Venn de los resultados del metaanálisis con rutas KEGG Se muestra la intersección de las funciones obtenidas en las cuatro fases de la LM (**Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica), a la izquierda en lesiones moderadas (a), y a la derecha en lesiones severas (b).

La tabla 4.10 detalla el valor de los diferentes estimadores de la medida del efecto y heterogeneidad para las 10 rutas con mayor LOR en valor absoluto obtenidas en el metaanálisis de la fase aguda con lesión moderada.

Tabla 4.10: Selección de las 10 rutas KEGG con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas. Se muestran los estimadores de la medida de efecto y los indicadores de la heterogeneidad.

ID	Nombre	LI 95%	LOR	LR 95%	FDR
03030	DNA replication	1.178	1.477	1.777	0
03010	Ribosome	0.552	1.208	1.864	0
03430	Mismatch repair	0.617	1.073	1.53	0
03410	Base excision repair	0.783	1.014	1.246	0
03008	Ribosome biogenesis in eukaryotes	0.672	0.998	1.325	0
03050	Proteasome	0.82	0.997	1.174	0
03440	Homologous recombination	0.686	0.975	1.264	0
00740	Riboflavin metabolism	0.588	0.956	1.324	0
04623	Cytosolic DNA-sensing pathway	0.675	0.883	1.09	0
03420	Nucleotide excision repair	0.564	0.828	1.091	0

ID	QE	QEp	SE	tau2	I2	H2
03030	2.247	0.134	0.153	0.026	55.492	2.247
03010	61.201	0	0.335	0.325	96.732	30.6
03430	3.116	0.078	0.233	0.074	67.91	3.116
03410	0.751	0.386	0.118	0	0	1
03008	5.691	0.017	0.167	0.046	82.428	5.691
03050	0.527	0.468	0.09	0	0	1
03440	1.375	0.241	0.147	0.012	27.262	1.375
00740	0.016	0.901	0.188	0	0	1
04623	2.165	0.339	0.106	0.003	7.626	1.083
03420	2.114	0.146	0.134	0.019	52.697	2.114

En la figura 4.7 se visualiza el diagrama de puntos para rutas KEGG enriquecidas en lesiones moderadas. El diagrama de puntos con las rutas KEGG significativas en lesiones severas puede encontrarse en el anexo A.14.

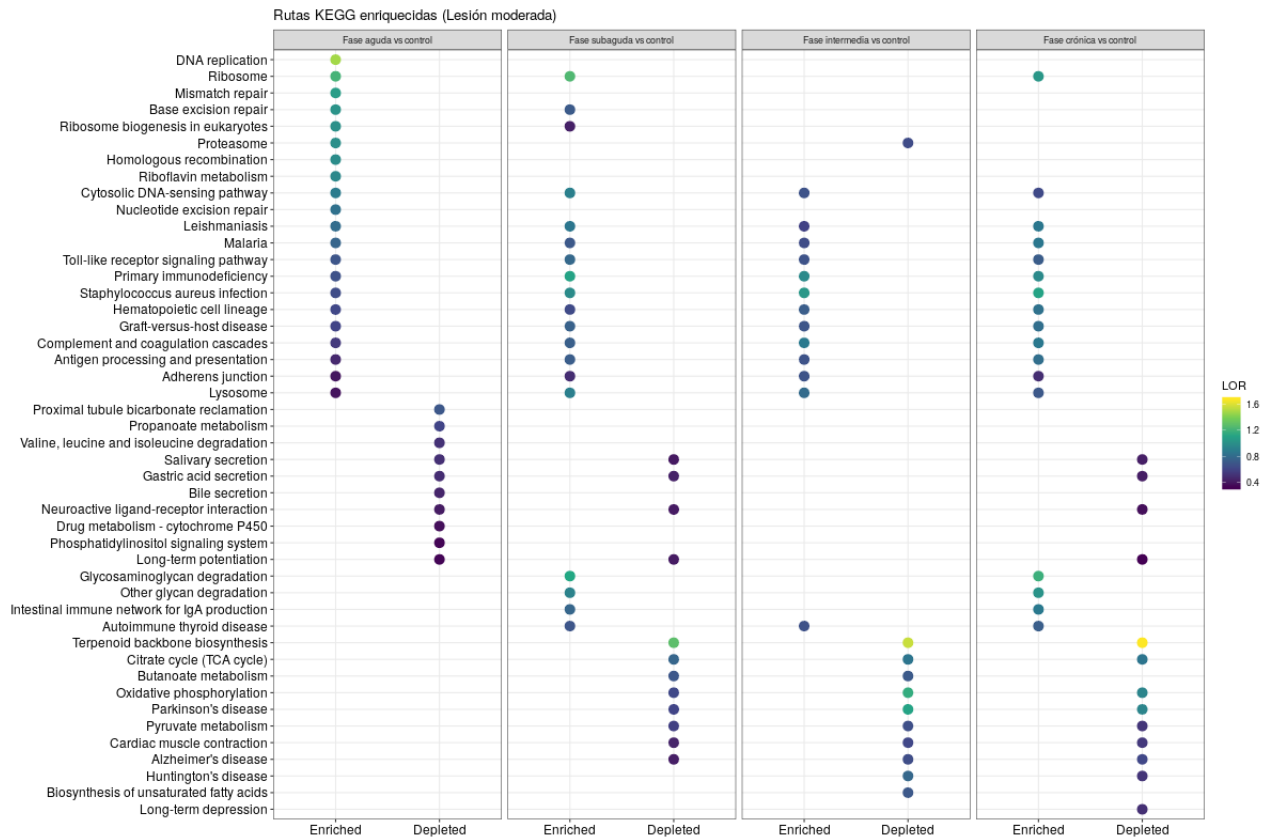


Figura 4.7: Resumen de rutas KEGG con mayor LOR en lesiones moderadas. Para cada fase se muestran funciones enriquecidas (*Enriched*) y funciones empobrecidas (*Depleted*). El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos.

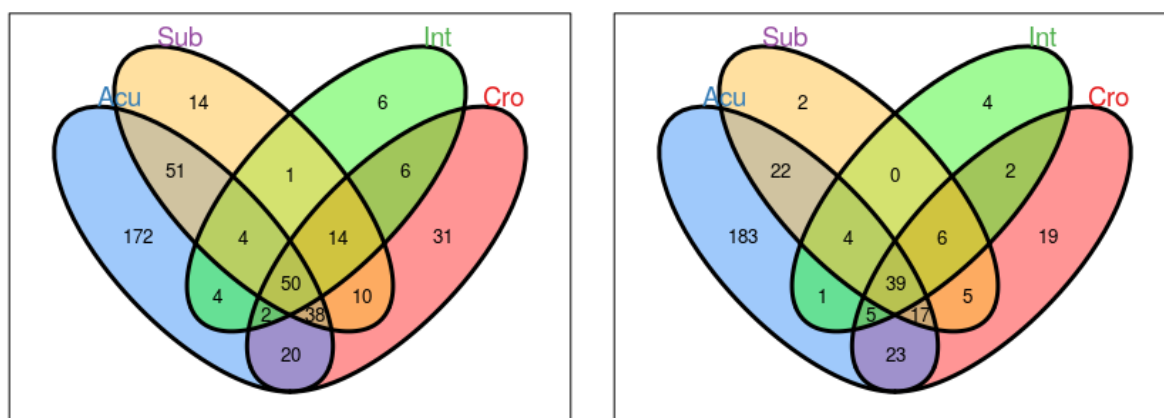
4.4.1.3. Rutas Reactome

En la tabla 4.11 se resumen los resultados obtenidos en el metaanálisis con rutas Reactome, para cada uno de los contrastes realizados.

Tabla 4.11: Resultados del metaanálisis para rutas Reactome. *Enriched* y *Depleted* hacen referencia a las funciones que presentan un p-valor ajustado menor a 0.05, y con un LOR mayor o menor a 0, respectivamente.

	Lesión moderada			Lesión severa		
	Enriched	Depleted	Total	Enriched	Depleted	Total
Fase aguda vs control	315	26	341	257	37	294
Fase subaguda vs control	146	36	182	71	24	95
Fase intermedia vs control	57	30	87	34	27	61
Fase crónica vs control	105	61	171	93	23	116

La figura 4.8, muestra en diagramas de Venn la intersección de funciones entre las diferentes fases, tanto para lesiones moderadas como para lesiones severas.



(a) Lesión moderada

(b) Lesión severa

Figura 4.8: Diagramas de Venn de los resultados del metaanálisis con rutas Reactome Se muestra la intersección de las funciones obtenidas en las cuatro fases de la LM (**Agu**, fase aguda; **Sub**, fase subaguda; **Int**, fase intermedia; **Cro**, fase crónica), a la izquierda en lesiones moderadas (a), y a la derecha en lesiones severas (b).

La tabla 4.12 detalla el valor de los diferentes estimadores de la medida del efecto y heterogeneidad para las 10 rutas con mayor LOR en valor absoluto obtenidas en el metaanálisis de la fase aguda con lesión moderada.

Tabla 4.12: Selección de los 10 rutas Reactome con mayor LOR en valor absoluto del metaanálisis para la fase aguda de lesiones moderadas. Se muestran los estimadores de la medida de efecto y los indicadores de la heterogeneidad.

ID	Nombre	LI 95%	LOR	LR 95%	FDR
R-RNO-446353	Cell-extracellular matrix interactions	1.238	1.454	1.67	0
R-RNO-72731	Recycling of eIF2:GDP	1.119	1.413	1.707	0
R-RNO-5651801	PCNA-Dependent Long Patch Base Excision Repair	1.1	1.403	1.706	0
R-RNO-110373	Resolution of AP sites via the multiple-nucleotide patch replacement pathway	1.073	1.366	1.659	0
R-RNO-111367	SLBP independent Processing of Histone Pre-mRNAs	1.025	1.339	1.652	0
R-RNO-69190	DNA strand elongation	1.09	1.338	1.587	0
R-RNO-174414	Processive synthesis on the C-strand of the telomere	0.919	1.334	1.749	0
R-RNO-174437	Removal of the Flap Intermediate from the C-strand	0.919	1.334	1.749	0
R-RNO-174411	Polymerase switching on the C-strand of the telomere	1.046	1.328	1.61	0
R-RNO-69091	Polymerase switching	1.046	1.328	1.61	0

ID	QE	QEp	SE	tau2	I2	H2
R-RNO-446353	1.144	0.564	0.11	0	0	1
R-RNO-72731	0.856	0.355	0.15	0	0	1
R-RNO-5651801	1.637	0.201	0.155	0.019	38.895	1.637
R-RNO-110373	1.868	0.172	0.149	0.021	46.476	1.868
R-RNO-111367	0.173	0.677	0.16	0	0	1
R-RNO-69190	0.83	0.362	0.127	0	0	1
R-RNO-174414	0.081	0.776	0.212	0	0	1
R-RNO-174437	0.081	0.776	0.212	0	0	1
R-RNO-174411	0.98	0.322	0.144	0	0	1
R-RNO-69091	0.98	0.322	0.144	0	0	1

En la figura 4.9 se visualiza el diagrama de puntos para rutas Reactome enriquecidas en lesiones moderadas. El diagrama de puntos con las rutas Reactome significativas en lesiones severas puede encontrarse en el anexo A.15.

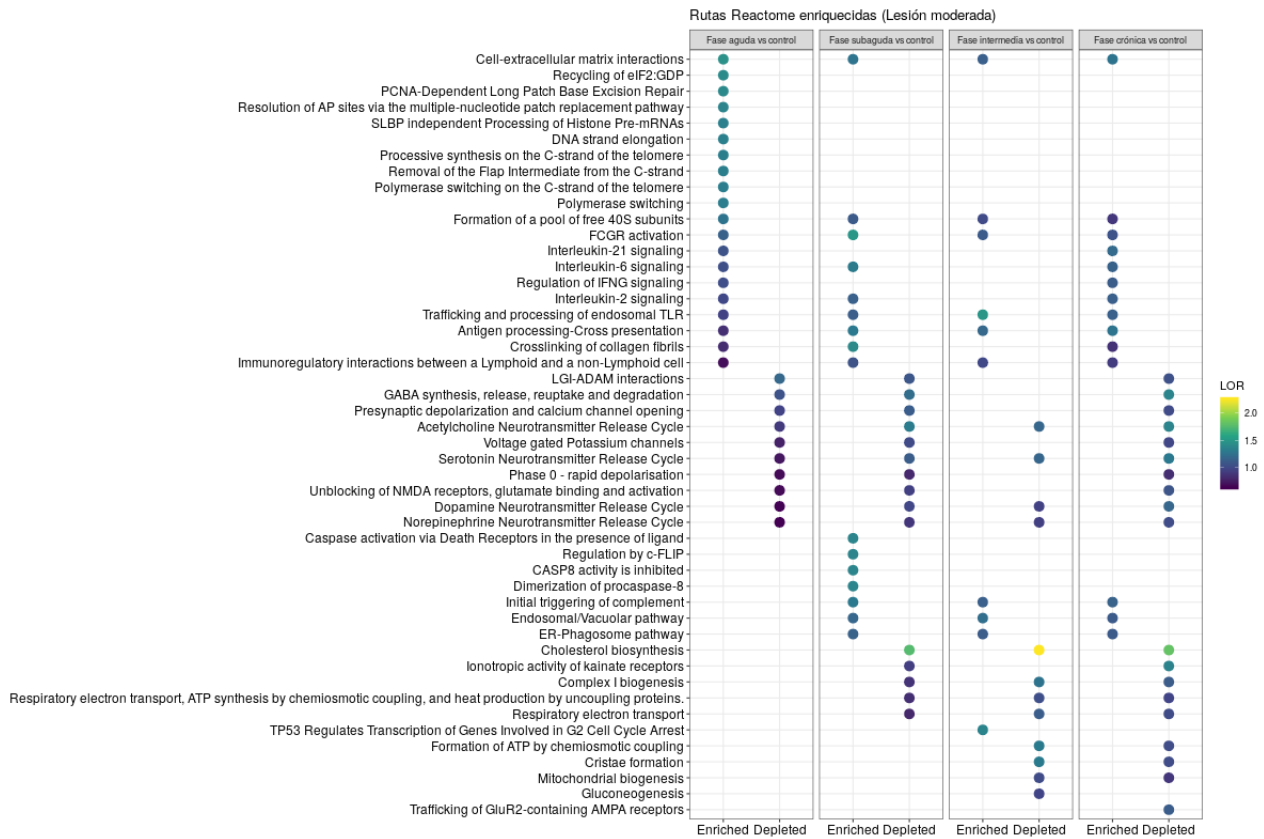


Figura 4.9: Resumen de rutas Reactome con mayor LOR en lesiones moderadas. Para cada fase se muestran funciones enriquecidas (*Enriched*) y funciones empobrecidas (*Depleted*). El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos.

4.4.2. Rutas de señalización

A continuación se muestra el número de subrutas efectoras significativas (FDR menor a 0.05), tras realizar el metaanálisis con los resultados del análisis de rutas de señalización con Hipathia (4.13). Se indica para cada uno de los 8 contrastes realizados el número de subrutas efectoras que se activan (*Up*) e inhiben (*Down*), al comparar una condición frente a los controles.

Tabla 4.13: Resultados del metaanálisis para rutas de señalización de Hipathia. Se indica el número de subrutas efectoras significativas, FDR menor a 0.05, activadas *Up*, e inhibidas *Down*, en cada uno de los contrastes realizados.

	Lesión moderada			Lesión severa		
	Up	Down	Total	Up	Down	Total
Fase aguda vs control	65	7	72	44	34	78
Fase subaguda vs control	52	0	52	0	0	0
Fase intermedia vs control	6	6	12	5	1	6
Fase crónica vs control	24	50	74	137	91	228

5. Discusión

La LM tiene una patofisiología altamente compleja que supone un gran obstáculo a la hora de desarrollar nuevas estrategias terapéuticas que permitan una recuperación funcional. Esta complejidad se debe a la multitud de rutas moleculares que actúan en una amplia variedad de tipos celulares, de manera coordinada, a lo largo de las diferentes fases de la lesión [64].

Para caracterizar funcionalmente la lesión e identificar los procesos celulares y rutas metabólicas que subyacen en cada una de las fases, en este trabajo se propone una metodología no usada hasta la fecha en la comprensión de la LM. Esta metodología resulta novedosa, primeramente, por incluir una revisión sistemática y metaanálisis de todos los estudios transcriptómicos relacionados con LM disponibles en bases de datos públicas. Además, también son novedosos los abordajes ómicos utilizados en los análisis individuales: análisis de enriquecimiento biológico con el método implementado en *mdGSA* y análisis de rutas de señalización con Hipathia.

El metaanálisis es un método robusto para el análisis conjunto de datos, que incrementa la precisión ofrecida por los estudios individuales, y por tanto, el poder estadístico de los hallazgos. Esta técnica permite combinar los estudios seleccionados y generar información conjunta que de otra forma no podría obtenerse, ni por la intersección ni por la suma de resultado individuales.

En un metaanálisis la revisión sistemática y selección de estudios es una parte fundamental del flujo de trabajo, siendo de gran importancia contar con el asesoramiento de expertos en la materia que indiquen las características de interés a tener en cuenta, los criterios de inclusión y exclusión de estudios o los tipos de muestras que pueden combinarse entre sí. Dada la naturaleza de la LM, uno de los puntos claves en este trabajo ha sido el establecimiento de los grupos experimentales, solventándose con la cooperación del Laboratorio de Regeneración Neuronal y Tisular del CIPF. Al contrario que en otros tipos de enfermedades, que podrían caracterizarse con una simple comparación caso y control, en la LM se distinguen cuatro fases en función del tiempo que ha transcurrido desde el traumatismo. Pese a que algunos autores realizan una separación en tres fases, uniendo la fase intermedia y la fase crónica [46], resulta evidente pensar, y está descrito, que los perfiles de expresión génica no van a ser iguales un mes después de la lesión, que 6 meses después [24]. Por otro lado, en rata se han creado diferentes modelos de lesiones con el objetivo de emular distintos tipos de lesiones humanas o estudiar diferentes aspectos de la patofisiología [47]. En estos modelos la severidad de la lesión que se provoca es un parámetro importante a tener en cuenta, pues dependiendo de si las lesiones son moderadas o severas, cambiará la cascada de eventos moleculares que ocurren en las diferentes fases. Por ejemplo, en lesiones moderadas, al haber menos tejido dañado, se produce una menor activación del sistema inmune o menor degeneración axonal, lo que conduce a una recuperación funcional más rápida [30].

Teniendo en cuenta estos aspectos, se decidió establecer los 9 grupos experimentales con los que se ha trabajado, ya que posibilita la identificación de potenciales biomarcadores tanto de fase como de severidad de la lesión. Sin embargo, pese a seleccionarse datos de 11 estudios independientes, la presencia de tantos grupos experimentales causa que no todos estén igual de representados en el metaanálisis. Aunque en la mayoría de los contrastes realizados se reúnen muestras procedentes de al menos 3 estudios independientes (con un máximo de 6 estudios, en la fase subaguda de lesiones severas), en la fase crónica de lesiones severas, tan solo se pudieron combinar muestras de dos estudios (tabla 4.2). Pese a que evidentemente siempre lo ideal es contar con el mayor número de estudios posible, el número mínimo de estudios que debe incluir un metaanálisis es dos [69].

Cabe destacar que en el estudio con identificador GSE2599 solo hay disponibles 6 muestras en GEO, correspondientes a dos grupos experimentales, mientras que en el artículo asociado indican que también se tomaron muestras de otros tres grupos experimentales diferentes. Al ser de gran interés para el presente trabajo, se contactó con el responsable de los datos para solicitarlos, sin embargo, el estado actual de los datos es “not found”. Por otro lado, también existen artículos que generan datos de gran valor para la comunidad científica pero no los hacen públicos [60]. Por ello hay que hacer constancia de la importancia de hacer públicos los datos, y que esto se haga bajo los principios FAIR. Los principios FAIR (*Findable, Accessible, Interoperable, Reusable*) [72], implican que los datos puedan encontrarse e identificarse apropiadamente, de manera que es necesaria la existencia de metadatos descriptivos y estandarizados, así como que sean accesibles y descargables para que puedan ser reutilizados sin problemas por otros investigadores. La adopción de estos principios son en beneficio de la comunidad científica, pues son clave para mejorar y ampliar el conocimiento generado, y ya son requisito, por ejemplo, en el programa H2020 de la Unión Europea [27].

En los resultados individuales del análisis de enriquecimiento funcional con GSA, se obtuvo una gran cantidad de funciones enriquecidas con las tres bases de datos usadas. Generalmente el número de funciones es de varios cientos para términos GO y rutas Reactome, mientras que en las rutas de KEGG nunca se obtuvieron más de cien rutas enriquecidas, debido a que esta base de datos se compone tan solo de 330 rutas en su versión actual. Pese a ello, el número de rutas KEGG enriquecidas es alto en proporción con el número de rutas de la anotación. En cuanto a la proporción de funciones enriquecidas (sobrerrepresentadas en lesionados) o empobrecidas (sobrerrepresentadas en controles) en cada contraste, existe un mayor número de funciones enriquecidas en casi todos los casos. Esto es de esperar, ya que a raíz del traumatismo, se desencadena una cascada de eventos en la que intervienen múltiples procesos y se ponen en marcha numerosos mecanismos para contener la lesión.

Los resultados del metaanálisis mantienen el comportamiento descrito en los análisis individuales con GSA, con un mayor número de funciones sobrerrepresentadas en los grupos lesionados. Por otro lado, cabría esperar que el número de funciones significativas fuera disminuyendo en etapas más tardías de la lesión, tras la contención y recuperación del daño inicial, y además,

que el número total de funciones identificadas fuera mayor en lesiones severas que en lesiones moderadas, dada la gravedad del daño y la necesidad de una respuesta más contundente. Sin embargo, el número de funciones significativas en el metaanálisis es algo menor en las lesiones severas, y aunque el número se reduce en las etapas más tardías, en la fase crónica se han identificado más funciones que en la fase subaguda e intermedia (tablas 4.7, 4.9 y 4.11).

Esto se explica por el hecho de que el metaanálisis se ve influenciado, primeramente, por los resultados individuales del análisis de expresión diferencial. Así, con un mayor número de genes expresados diferencialmente se obtiene un mayor número de funciones enriquecidas al usar una misma anotación. En consecuencia, el metaanálisis se ve afectado por el número de funciones obtenidas en los estudios individuales, por lo que la presencia de estudios con pocos resultados limita las posibles comparaciones, causando que el número de funciones significativas en el metaanálisis sea menor. Un ejemplo de ello es el metaanálisis para la fase intermedia de lesiones severas, donde hay 4 estudios, en los que dos de ellos tienen 73 y 127 términos GO enriquecidos, mientras que los otros dos tienen 398 y 666, obteniéndose finalmente 75 funciones significativas en el metaanálisis.

De igual manera, los resultados del metaanálisis con rutas de señalización se ven muy influenciados por los resultados de los análisis individuales, y dependerá mucho del número de subrutas efectoras significativas que se obtiene en cada uno de ellos y el sentido de su activación. El metaanálisis en este caso es especialmente exigente, ya que el sentido de activación de una misma subruta en el conjunto de estudios analizados debe ser el mismo, en caso contrario no pueden pasar ese filtro. Esto perjudica especialmente el metaanálisis con muchos estudios, como es el caso de la fase subaguda en lesiones severas, donde no se pudo identificar ninguna ruta significativa. Además, es de carácter prácticamente indispensable que una subruta dada esté diferencialmente activada ($FDR < 0.05$) en todos los estudios, pues se emplea un método de combinación de valores de probabilidades para experimentos independientes [25]. Así, pese a que existen 1853 subrutas efectoras en total, en el metaanálisis el número general de subrutas significativas es bajo (tabla 4.13).

La interpretación de los resultados desde un punto de vista funcional es una etapa del proyecto costosa en cuanto a tiempo, ya que el metaanálisis del análisis de enriquecimiento funcional ha permitido generar una gran cantidad de información. El gran número de resultados obtenidos en cada una de las comparaciones propuestas, con las tres bases de datos empleadas, unido a la naturaleza ómica de los mismos, hace que relacionar toda la información entre sí no sea una actividad trivial. Por ello, esta fase requiere de la cooperación de investigadores de “*wet lab*” y “*dry lab*”, con el fin de poder organizar y sintetizar la información de manera que permita seleccionar los procesos biológicos y rutas de señalización de interés para su validación *ex vivo* e *in vivo*.

Desde una perspectiva general, puede afirmarse que los resultados del metaanálisis son representativos de los procesos descritos en la literatura. Atendiendo a los procesos biológicos de la Gene Ontology, la base de datos por excelencia en los análisis funcionales, se observa que tras

la lesión se activan numerosos procesos de respuesta a estímulos, como respuesta a hipoxia (GO:0001666), respuesta inflamatoria aguda (GO:0002526), respuesta celular a ADN dañado (GO:0006974), respuesta a estrés oxidativo (GO:0006979) o respuesta a estímulos mecánicos (GO:0009612). Todos estos términos fueron mencionados en la descripción de eventos que suceden tras el traumatismo (sección 1.5).

En lo relativo al sistema vascular, la respuesta a hipoxia se explica por la rotura y hemorragias de los vasos sanguíneos, que impiden la llegada tanto de oxígeno como de nutrientes. También se han identificado en todas las fases varios términos implicados en la activación de la angiogénesis, remodelación y morfogénesis de vasos o coagulación (GO:0045766, GO:0001974, GO:0048514, GO:0007596), los cuales son necesarios para restablecer el flujo sanguíneo y la perfusión en la médula espinal [51]. Por otro lado, la isquemia hace que llegue menos oxígeno a las células, viéndose reducida la capacidad de sintetizar ATP (GO:0006120, GO:0015986), que conlleva a la pérdida de las funciones celulares dependientes de ATP [13].

La mayoría de funciones que se han perdido tras la lesión tienen que ver con el sistema nervioso y la transmisión de la señal nerviosa. Los términos GO:0035725, GO:0070588, GO:0071805 hacen referencia al transporte transmembrana de sodio, calcio y potasio, respectivamente, pudiendo identificarse en las diferentes fases de la lesión. La rotura de los axones junto con la isquemia hace que los diferentes mecanismos de homeostasis iónica se desregulen, afectando a la transmisión sináptica (GO:0007268) y secreción de neurotransmisores (GO:0007269), perdiéndose en general la funcionalidad del sistema nervioso (GO:0050877).

Todos estos eventos conducen inevitablemente a la apoptosis mediante diferentes vías de señalización (GO:0097190), tanto por la vía extrínseca, mediante receptores del factor de necrosis tumoral (TNF) (GO:0034612) o caspasas (R-RNO-5357769), como por la vía intrínseca, bajo el control de p53 como respuesta al daño del ADN (GO:0042771) [5].

El conjunto de daños producidos tras la lesión genera una gran cantidad de moléculas y señales pro-inflamatorias, que son reconocidas por las diferentes poblaciones celulares del sistema inmune, las cuales pondrán en marcha una rápida respuesta inflamatoria con el objetivo de contener la lesión. Estas señales, a su vez, permiten activar la transcripción de multitud de genes, sintetizándose y liberándose citoquinas y quimioquinas, favoreciendo el reclutamiento de células del sistema inmune [13]. Se han identificado numerosas funciones que hacen referencia a ello, como respuesta a citoquinas (GO:0034097), regulación positiva de la producción de quimioquinas (GO:0032722), regulación positiva de la secreción de quimioquinas (GO:0090197), respuesta celular a interferon alfa (GO:0035457), respuesta celular a interferon beta (GO:0035458) o respuesta celular a interferon gamma (GO:0071346). También se identifican términos de respuesta a las principales citoquinas pro-inflamatorias del sistema inmune innato, secretadas por las células tisulares frente a daño tisular o infección, como TNF (GO:0034612), IL-1 (GO:0070555) o IL-6 (GO:0071354). En respuesta a estas moléculas se ven atraídas diferentes poblaciones celulares como neutrófilos (GO:0030593), monocitos (GO:0002548), macrófagos (GO:0010759) o linfocitos (GO:0030595) [44].

Los receptores de tipo Toll son una familia de receptores que pueden reconocer tanto patrones moleculares asociados a microorganismos patógenos (PAMP), como patrones moleculares asociados a lesión (DAMP). Cuando estos reconocen una molécula activan la transducción de señales y promueven funciones antimicrobianas o pro-inflamatorias [44]. Se han identificado términos que indican una activación de la ruta de señalización mediante receptores tipo Toll (GO:0002224) en todos los contrastes realizados, en la mayoría de ellos mediante la vía dependiente de MyD88 (GO:0002755). Se ha descrito que las personas con lesión medular crónica tienen una mayor expresión de receptores tipo Toll que las personas sin lesión, lo que conlleva a un estado de inflamación permanente, siendo una línea de investigación a tener en cuenta de cara a posibles terapias [34].

El sistema del complemento es un conjunto de proteínas con diversas funciones en el sistema inmune, que forma parte tanto de la respuesta innata, al tener actividad proteolítica y pro-inflamatoria, como de la respuesta adaptativa al interactuar con linfocitos B y T [12]. El sistema del complemento también se activa tras traumatismos mediante la vía clásica, esto es, mediante la interacción con anticuerpos unidos a antígenos generados por la lesión tisular. Existen otras dos vías de activación del complemento, la vía alternativa y la vía de la lectina, que requieren de la unión de las proteínas del complemento con microorganismos [44]. Se han identificado términos referentes a la activación de la vía clásica del complemento (GO:0006958) en todas las fases salvo en la fase aguda. Resultados semejantes fueron obtenidos por Chamankhah et al. (2013) [13], donde describen una infraexpresión de genes relacionados con la activación de la vía clásica del complemento en los primeros tres días (fase aguda), que posteriormente aumentan su expresión hasta la fase crónica.

Los linfocitos B y linfocitos T tienen un papel vital en la regulación del sistema inmune, al igual que las distintas fases de la LM. Se han identificado términos referentes a linfocitos B en todas las fases, relacionados con la regulación positiva de la proliferación (GO:0030890), regulación positiva de la diferenciación (GO:0045579) y las rutas de señalización iniciadas tras la unión de antígenos a receptores de células B (GO:0045579). Por otro lado, se identificó un número mayor de términos que implican a los linfocitos T: regulación positiva de la migración de células T (GO:2000406), coestimulación de células T (GO:0031295), activación de células T (GO:0042110), regulación positiva de la proliferación de células T activadas (GO:0042104), regulación positiva de la diferenciación (GO:0030217) y regulación positiva de la citotoxicidad mediada por células T (GO:0001916). Algunos de estos términos ya se encuentran desde la fase aguda, pese a que los linfocitos suelen estar más implicados en la respuesta inmune adaptativa. La rotura de los vasos sanguíneos hace que se infiltren linfocitos en la zona de la lesión, aunque su extravasación es mayor en etapas posteriores [4].

Así, tras esta descripción general de los resultados obtenidos en el metaanálisis con términos GO, se demuestra la validez de la metodología propuesta, pues es capaz de identificar los procesos que ocurren en la LM. También hay que poner en valor la capacidad de reanalizar datos publicados previamente, pues brinda la oportunidad de utilizar nuevas estrategias que no existieran con anterioridad y sacar más provecho de ellos. El uso de aproximaciones que

permiten combinar el efecto de la totalidad de genes de un experimento y asociarlos funcionalmente, posibilita detectar procesos moleculares que no serían obvios relacionar si se trabajara a nivel de gen. La base de estos métodos es que genes con grandes cambios en su expresión aportarán más a la biología del proceso en estudio, pero a veces pequeños cambios en el nivel de expresión de un grupo genes es suficiente para dar una respuesta a nivel celular. Al mismo tiempo, el reanálisis de datos con esta metodología permite integrar mediante un metaanálisis la información funcional que aportan diferentes tecnologías como la secuenciación y los microarrays, procedentes a su vez de diferentes estudios.

Sin embargo, para poder iniciar nuevas vías de investigación en busca de terapias, es necesario una exploración y comprensión más profunda de los resultados. Para ello se requerirá complementar los resultados de las diferentes bases de datos. En el caso de las rutas de Reactome, el vocabulario es más específico que los términos GO, posibilitando poner la atención en rutas metabólicas concretas, reforzando la información obtenida con términos GO. Por ejemplo, se ha identificado que las interacciones LGI-ADAM (R-RNO-5682910) se pierden tras la lesión. LGI y ADAM son receptores implicados en la sinapsis de las neuronas, con un importante papel en el desarrollo y funcionamiento del sistema nervioso [38]. Aún no está descrito su papel en la LM, por lo que podría ser un candidato de estudio. Por otro lado, la ventaja de KEGG es que aporta conocimiento sobre rutas metabólicas y la interacción entre los genes que las forman. Por ello, resulta interesante complementar los resultados obtenidos en el análisis de enriquecimiento biológico con KEGG, cuya información es más general, con el análisis de rutas de señalización con Hipathia, que son más específicos, al indicar qué subrutas están diferencialmente activadas.

De cara a líneas futuras de trabajo, se aplicará la metodología expuesta en el presente trabajo en diferentes especies animales, incluyendo a ratón, pez cebra y humano. Es de interés analizar las diferencias y similitudes existentes entre las diferentes especies de cara a poder trasladar y reproducir los mecanismos que impliquen una mejora en la recuperación de la lesión en humanos. Por otro lado, la visualización de datos ómicos es una de las asignaturas pendientes en la bioinformática, de manera que se trabajará en el desarrollo herramientas que ayuden a la organización y comprensión de los resultados producidos.

6. Conclusiones

1. El metaanálisis es un método robusto para la evaluación e integración de datos provenientes de una revisión sistemática de estudios ómicos.
2. Se han identificado un gran número de funciones en todos los contrastes realizados con los grupos experimentales propuestos, que son representativas de los eventos descritos en la literatura.
3. Las diferentes fases de la lesión comparten numerosas funciones entre sí, por lo que existen mecanismos de respuesta a la lesión que se activan en el inicio y se mantienen a lo largo del tiempo.
4. Hay una alta respuesta del sistema inmune que persiste en fases crónicas, sosteniendo un permanente estado de inflamación.
5. Los resultados suponen una base para el inicio de nuevas vías de investigación que permitan desarrollar terapias eficaces, siendo necesario un exhaustivo análisis de los mismos.
6. Los estudios *in silico* aportan información valiosa para la comprensión de las enfermedades y el avance en biomedicina, viéndose su potencial aumentado con la adopción de los principios FAIR.

Bibliografía

- [1] Ahuja, C. S., Wilson, J. R., Nori, S., Kotter, M. R., Druschel, C., Curt, A., and Fehlings, M. G. (2017). Traumatic spinal cord injury. *Nature reviews Disease primers*, 3:17018.
- [2] Aimone, J. B., Leasure, J. L., Perreau, V. M., Thallmair, M., Consortium, C. R. P. F. R., et al. (2004). Spatial and temporal gene expression profiling of the contused rat spinal cord. *Experimental neurology*, 189(2):204–221.
- [3] Albayar, A. A., Roche, A., Swiatkowski, P., Antar, S., Ouda, N., Emara, E., Smith, D. H., Ozturk, A. K., and Awad, B. I. (2019). Biomarkers in spinal cord injury: Prognostic insights and future potentials. *Frontiers in neurology*, 10.
- [4] Alizadeh, A., Dyck, S. M., and Karimi-Abdolrezaee, S. (2019). Traumatic spinal cord injury: an overview of pathophysiology, models and acute injury mechanisms. *Frontiers in neurology*, 10:282.
- [5] Angosto, M. C. (2003). Bases moleculares de la apoptosis. In *Anales de la Real Academia Nacional de Farmacia*, volume 69.
- [6] 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., et al. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25.
- [7] Avila-Martin, G., Mata-Roig, M., Galan-Arriero, I., Taylor, J. S., Busquets, X., and Escriba, P. V. (2017). Treatment with albumin-hydroxyoleic acid complex restores sensorimotor function in rats with spinal cord injury: Efficacy and gene expression regulation. *PloS one*, 12(12):e0189151.
- [8] Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., et al. (2012). Ncbi geo: archive for functional genomics data sets—update. *Nucleic acids research*, 41(D1):D991–D995.
- [9] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.
- [10] Benjamini, Y., Yekutieli, D., et al. (2001). The control of the false discovery rate in multiple testing under dependency. *The annals of statistics*, 29(4):1165–1188.
- [11] Castillo, D., Gálvez, J. M., Herrera, L. J., San Román, B., Rojas, F., and Rojas, I. (2017). Integration of rna-seq data with heterogeneous microarray data for breast cancer profiling. *BMC bioinformatics*, 18(1):506.

-
- [12] Chakraborty, S., Karasu, E., and Huber-Lang, M. (2018). Complement after trauma: suturing innate and adaptive immunity. *Frontiers in immunology*, 9.
- [13] Chamankhah, M., Eftekharpour, E., Karimi-Abdolrezaee, S., Boutros, P. C., San-Marina, S., and Fehlings, M. G. (2013). Genome-wide gene expression profiling of stress response in a spinal cord clip compression injury model. *BMC genomics*, 14(1):583.
- [14] Chehensse, C., Bahrami, S., Denys, P., Clement, P., Bernabe, J., and Giuliano, F. (2013). The spinal control of ejaculation revisited: a systematic review and meta-analysis of anejaculation in spinal cord injured patients. *Human reproduction update*, 19(5):507–526.
- [15] Cosar, S., Yemisci, O., Oztop, P., Çetin, N., Sarifakioglu, B., Yalbuздag, S., Ustaomer, K., and Karatas, M. (2010). Demographic characteristics after traumatic and non-traumatic spinal cord injury: a retrospective comparison study. *Spinal cord*, 48(12):862.
- [16] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188.
- [17] DeVivo, M., Chen, Y., Menemeyer, S., and Deutsch, A. (2011). Costs of care following spinal cord injury. *Topics in spinal cord injury rehabilitation*, 16(4):1–9.
- [18] Dewey, M. (2019). *metap: meta-analysis of significance values*. R package version 1.1.
- [19] Di Giovanni, S., Knobloch, S. M., Brandoli, C., Aden, S. A., Hoffman, E. P., and Faden, A. I. (2003). Gene profiling in spinal cord injury shows role of cell cycle in neuronal death. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 53(4):454–468.
- [20] Dopazo, J. (2009a). Formulating and testing hypotheses in functional genomics. *Artificial intelligence in medicine*, 45(2-3):97–107.
- [21] Dopazo, J. (2009b). Functional profiling methods in cancer. In *Cancer Gene Profiling*, pages 363–374. Springer.
- [22] Duan, H., Ge, W., Zhang, A., Xi, Y., Chen, Z., Luo, D., Cheng, Y., Fan, K. S., Horvath, S., Sofroniew, M. V., et al. (2015). Transcriptome analyses reveal molecular mechanisms underlying functional recovery after spinal cord injury. *Proceedings of the National Academy of Sciences*, 112(43):13360–13365.
- [23] Dumont, R. J., Okonkwo, D. O., Verma, S., Hurlbert, R. J., Boulos, P. T., Ellegala, D. B., and Dumont, A. S. (2001). Acute spinal cord injury, part i: pathophysiologic mechanisms. *Clinical neuropharmacology*, 24(5):254–264.
- [24] Duran, R. C.-D., Yan, H., Zheng, Y., Huang, X., Grill, R., Kim, D. H., Cao, Q., and Wu, J. Q. (2017). The systematic analysis of coding and long non-coding rnas in the sub-chronic and chronic stages of spinal cord injury. *Scientific reports*, 7:41008.
-

-
- [25] Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *The Journal of Psychology*, 80(2):351–363.
- [26] Ehlers, M. D. (2004). Deconstructing the axon: Wallerian degeneration and the ubiquitin–proteasome system. *Trends in neurosciences*, 27(1):3–6.
- [27] European Commission (2016). Guidelines on fair data management in horizon 2020.
- [28] Evangelou, E. and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, 14(6):379.
- [29] García-García, F. (2016). *Methods of functional enrichment analysis in genomic studies*. PhD thesis.
- [30] Ghorbani, M., Shahabi, P., Ebrahimi-kalan, A., Soltani-Zangbar, H., Mahmoudi, J., Bani, S., Sadeghzadeh-Oskouei, B., Rafiee-Byraami, Y., and Salimi, O. (2018). Induction of traumatic brain and spinal cord injury models in rat using a modified impactor device. *Physiology and Pharmacology*, 22(4):228–239.
- [31] Grassner, L., Marschallinger, J., Dünser, M. W., Novak, H. F., Zerbs, A., Aigner, L., Trinka, E., and Sellner, J. (2016). Nontraumatic spinal cord injury at the neurological intensive care unit: spectrum, causes of admission and predictors of mortality. *Therapeutic advances in neurological disorders*, 9(2):85–94.
- [32] Grützmann, R., Boriss, H., Ammerpohl, O., Lüttges, J., Kalthoff, H., Schackert, H. K., Klöppel, G., Saeger, H. D., and Pilarsky, C. (2005). Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. *Oncogene*, 24(32):5079.
- [33] Guallar, E., Damián, J., and Martín-Moreno, J. M. (1997). Metaanálisis y revisiones sistemáticas en cardiología. *Revista Española de Cardiología*, 50(5):345–354.
- [34] Herman, P., Stein, A., Gibbs, K., Korsunsky, I., Gregersen, P., and Bloom, O. (2018). Persons with chronic spinal cord injury have decreased natural killer cell and increased toll-like receptor/inflammatory gene expression. *Journal of neurotrauma*, 35(15):1819–1829.
- [35] Hidalgo, M. R., Cubuk, C., Amadoz, A., Salavert, F., Carbonell-Caballero, J., and Dopazo, J. (2017). High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget*, 8(3):5160.
- [36] Huete, A., Díaz, E., Ortega, E., and Esclarín, A. (2012). Análisis sobre la lesión medular en España. *Federación Nacional de ASPAYM. Toledo: Intersocial*.
- [37] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic acids research*, 32(suppl_1):D277–D280.
- [38] Kegel, L., Aunin, E., Meijer, D., and Bermingham Jr, J. R. (2013). Lgi proteins in the nervous system. *ASN neuro*, 5(3):AN20120095.
-

-
- [39] Kim, Y.-H., Ha, K.-Y., and Kim, S.-I. (2017). Spinal cord injury and related clinical trials. *Clinics in orthopedic surgery*, 9(1):1–9.
- [40] Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2014). Arrayexpress update—simplifying data submissions. *Nucleic acids research*, 43(D1):D1113–D1116.
- [41] Korhonen, L. and Lindholm, D. (2004). The ubiquitin proteasome system in synaptic and axonal degeneration: a new twist to an old cycle. *J Cell Biol*, 165(1):27–30.
- [42] Lema, C. P. H. and Parra, J. E. P. (2010). Lesiones medulares y discapacidad: revisión bibliográfica. *Aquichan*, 10(2):157–172.
- [43] Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., and Moher, D. (2009). The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine*, 6(7):e1000100.
- [44] Lichtman, A. H., Pillai, S., and Abbas, A. K. (2011). *Inmunología celular y molecular* (7^a ed.). Elsevier.
- [45] Lugo, L. H., Salinas, F., and García, H. I. (2007). Out-patient rehabilitation programme for spinal cord injured patients: Evaluation of the results on motor firm score. *Disability and Rehabilitation*, 29(11-12):873–881.
- [46] Mataliotakis, G. I. and Tsirikos, A. I. (2016). Spinal cord trauma: pathophysiology, classification of spinal cord injury syndromes, treatment principles and controversies. *Orthopaedics and Trauma*, 30(5):440–449.
- [47] Minakov, A., Chernov, A., Asutin, D., Konovalov, N., and Telegin, G. (2018). Experimental models of spinal cord injury in laboratory rats. *Acta naturae*, 10(3):4.
- [48] Montaner, D. and Dopazo, J. (2010). Multidimensional gene set analysis of genomic data. *PloS one*, 5(4):e10348.
- [49] Nathan, F. M. and Li, S. (2017). Environmental cues determine the fate of astrocytes after spinal cord injury. *Neural regeneration research*, 12(12):1964.
- [50] New, P. W., Rawicki, H. B., and Bailey, M. J. (2002). Nontraumatic spinal cord injury: demographic characteristics and complications. *Archives of physical medicine and rehabilitation*, 83(7):996–1001.
- [51] Ng, M. T., Stammers, A. T., and Kwon, B. K. (2011). Vascular disruption and the role of angiogenic proteins after spinal cord injury. *Translational stroke research*, 2(4):474–491.
- [52] Norenberg, M. D., Smith, J., and Marcillo, A. (2004). The pathology of human spinal cord injury: defining the problems.
-

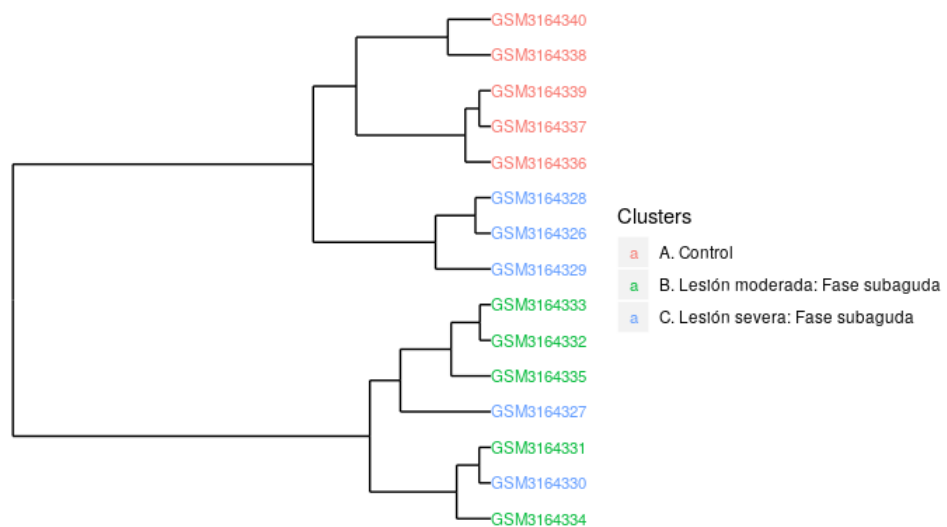
- [53] Normand, S.-L. T. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in medicine*, 18(3):321–359.
- [54] Padilla-Zambrano, H., Ramos-Villegas, Y., Alvis-Miranda, H. R., Joaquin, A. F., and Moscote-Salazar, L. R. (2017). Fisiopatología del trauma raquimedular. *Revista Mexicana de Neurociencia*, 18(5):45–53.
- [55] Peate, I. and Nair, M. (2016). *Fundamentals of anatomy and physiology: For nursing and healthcare students (2nd ed.)*. John Wiley & Sons.
- [56] Pérez, K., Novoa, A. M., Santamariña-Rubio, E., Narvaez, Y., Arrufat, V., Borrell, C., Cabeza, E., Cirera, E., Ferrando, J., García-Altés, A., et al. (2012). Incidence trends of traumatic spinal cord injury and traumatic brain injury in Spain, 2000–2009. *Accident Analysis & Prevention*, 46:37–44.
- [57] Plaza, V. B., Pacheco, B. M., Aguilar, C. M., Valenzuela, J. F., and Pérez, J. J. Z. (2012). Lesión de la médula espinal. actualización bibliográfica: fisiopatología y tratamiento inicial. *Coluna/Columna*, 11(1):73–76.
- [58] Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D., and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer research*, 62(15):4427–4433.
- [59] Saladin, K. (2017). *Human Anatomy (5th ed.)*. McGraw-Hill Education.
- [60] Shi, L.-L., Zhang, N., Xie, X.-M., Chen, Y.-J., Wang, R., Shen, L., Zhou, J.-S., Hu, J.-G., and Lü, H.-Z. (2017). Transcriptome profile of rat genes in injured spinal cord at different stages by RNA-seq. *BMC genomics*, 18(1):173.
- [61] Shin, H. Y., Kim, H., Kwon, M. J., Hwang, D. H., Lee, K., and Kim, B. G. (2014). Molecular and cellular changes in the lumbar spinal cord following thoracic injury: regulation by treadmill locomotor training. *PLoS One*, 9(2):e88215.
- [62] Silva, N. A., Sousa, N., Reis, R. L., and Salgado, A. J. (2014). From basics to clinical: a comprehensive review on spinal cord injury. *Progress in neurobiology*, 114:25–57.
- [63] Spann, R. A., Lawson, W. J., Grill, R. J., Garrett, M. R., and Grayson, B. E. (2017). Chronic spinal cord changes in a high-fat diet-fed male rat model of thoracic spinal contusion. *Physiological genomics*, 49(9):519–529.
- [64] Squair, J. W., Tigchelaar, S., Moon, K.-M., Liu, J., Tetzlaff, W., Kwon, B. K., Krassioukov, A. V., West, C. R., Foster, L. J., and Skinnider, M. A. (2018). Integrated systems analysis reveals conserved gene networks underlying response to spinal cord injury. *Elife*, 7:e39188.
-

-
- [65] Team, R. C. (2019). A language and environment for statistical computing. vienna, austria: R foundation for statistical computing; 2019. URL <https://www.R-project.org>.
- [66] Torres Alaminos, M. A. (2018). Aspectos epidemiológicos de la lesión medular en el hospital nacional de paraplégicos. *Ene*, 12(2).
- [67] Torres-Espín, A., Hernández, J., and Navarro, X. (2013). Gene expression changes in the injured spinal cord following transplantation of mesenchymal stem cells or olfactory ensheathing cells. *PLoS One*, 8(10):e76141.
- [68] Tortora, G. J. and Derrickson, B. (2013). *Principios de anatomía y fisiología (13th ed.)*. Médica Panamericana.
- [69] Valentine, J. C., Pigott, T. D., and Rothstein, H. R. (2010). How many studies do you need? a primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2):215–247.
- [70] Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., et al. (2007). Reactome: a knowledge base of biologic pathways and processes. *Genome biology*, 8(3):R39.
- [71] Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of statistical software*, 36(3):1–48.
- [72] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3.
- [73] Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., de Bakker, P. I., Abecasis, G. R., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature genetics*, 40(5):638.
-

A. Anexo A - Figuras

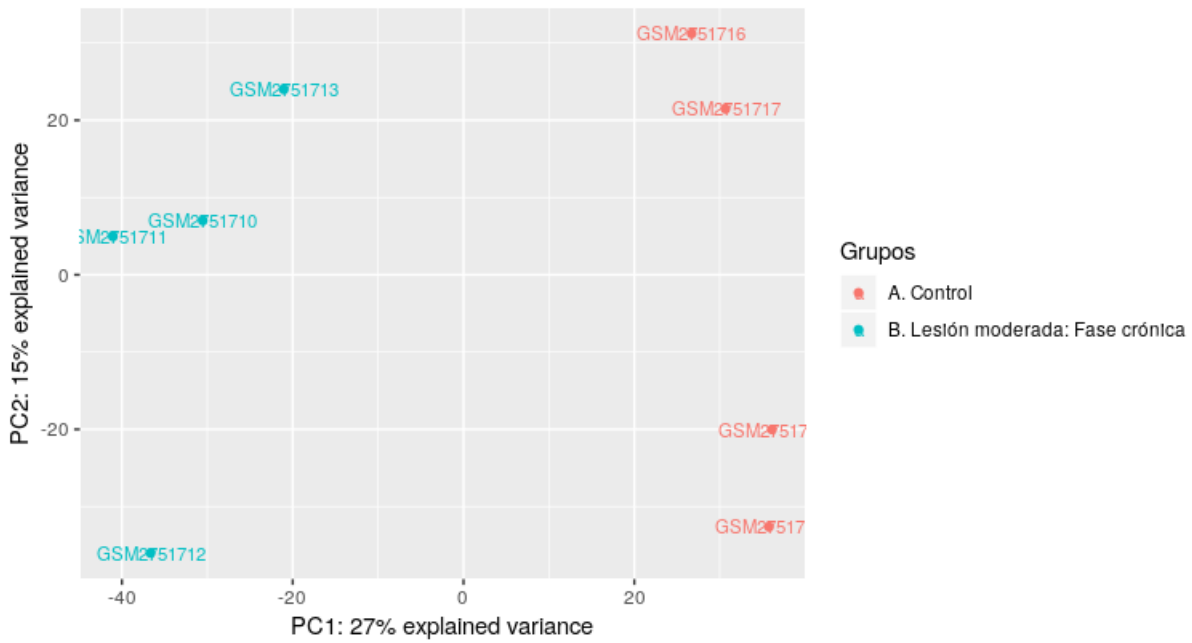


(a) PCA estudio GSE115067

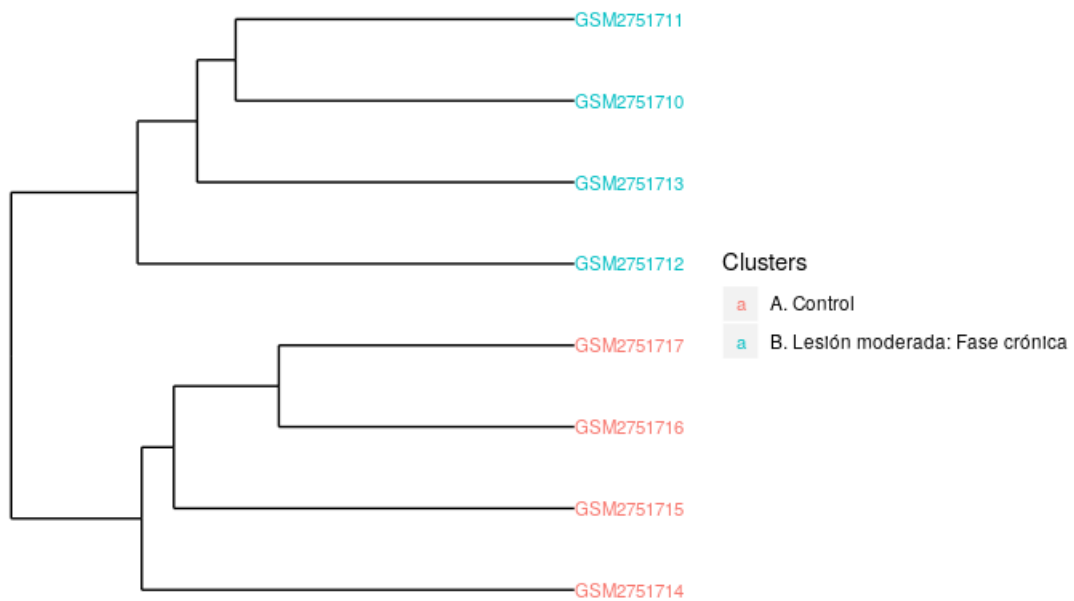


(b) Clustering estudio GSE115067

Figura A.1: Análisis exploratorio de las muestras del estudio GSE115067. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

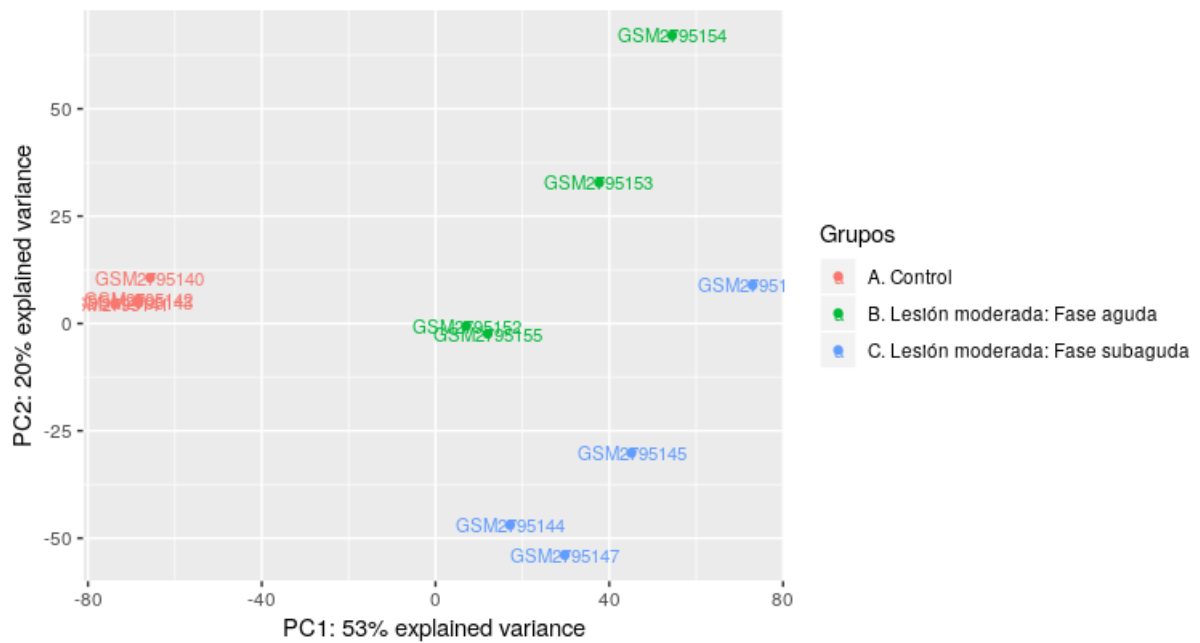


(a) PCA estudio GSE102964

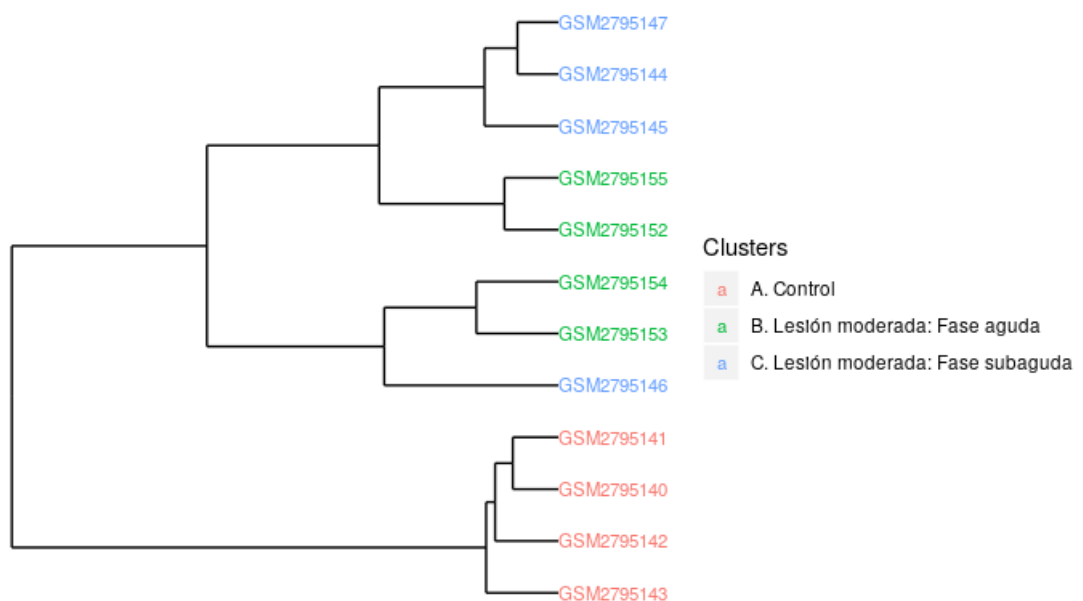


(b) Clustering estudio GSE102964

Figura A.2: Análisis exploratorio de las muestras del estudio GSE102964. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.



(a) PCA estudio GSE104317

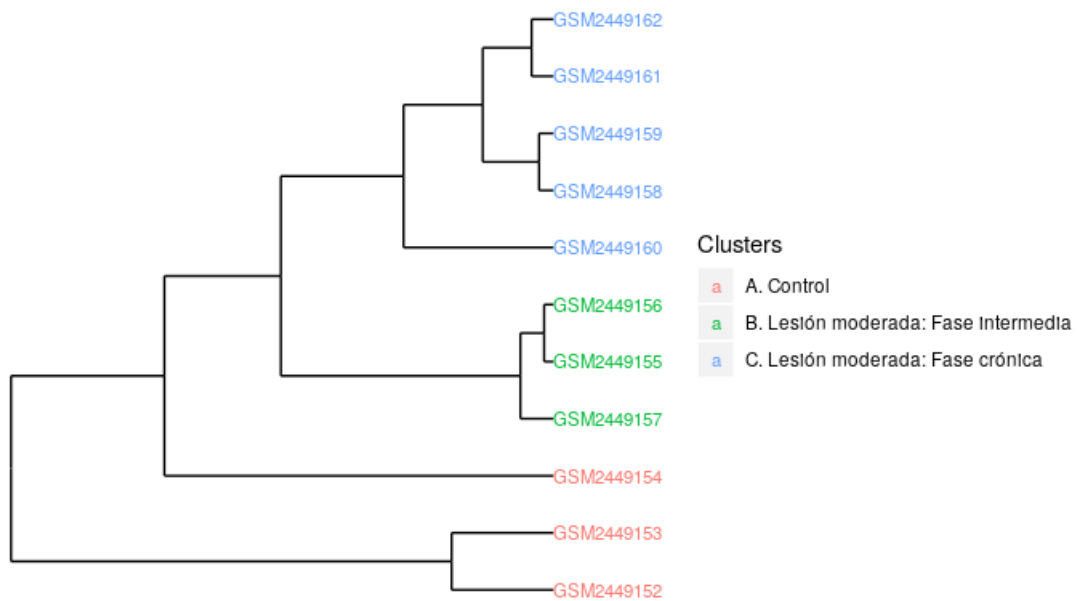


(b) Clustering estudio GSE104317

Figura A.3: Análisis exploratorio de las muestras del estudio GSE104317. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

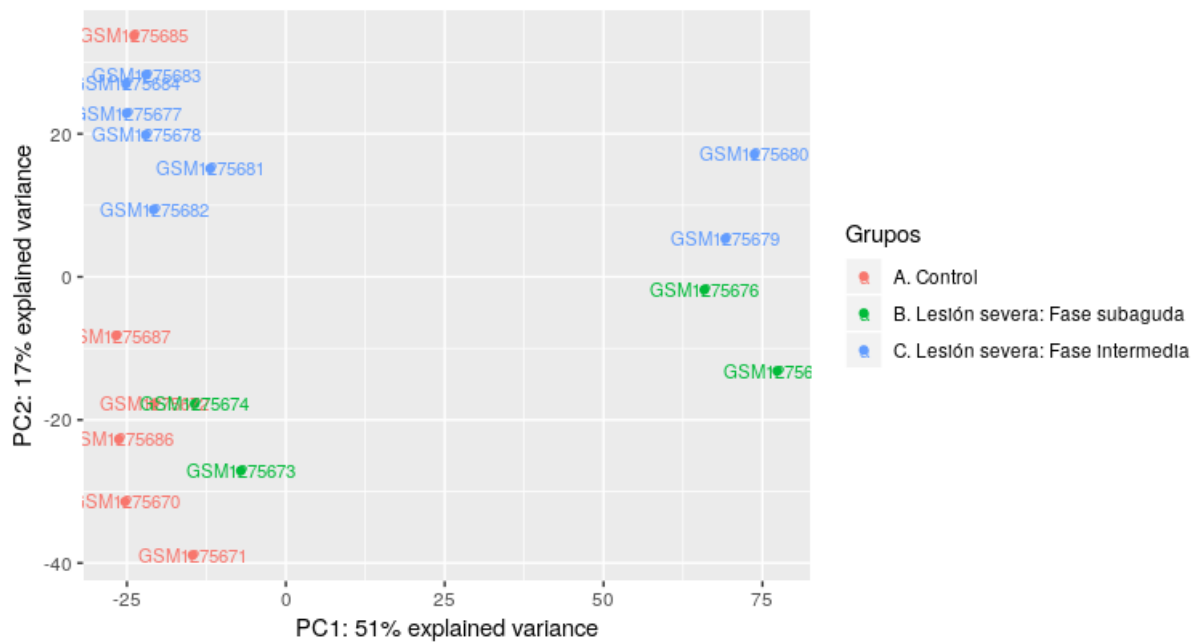


(a) PCA estudio GSE93249

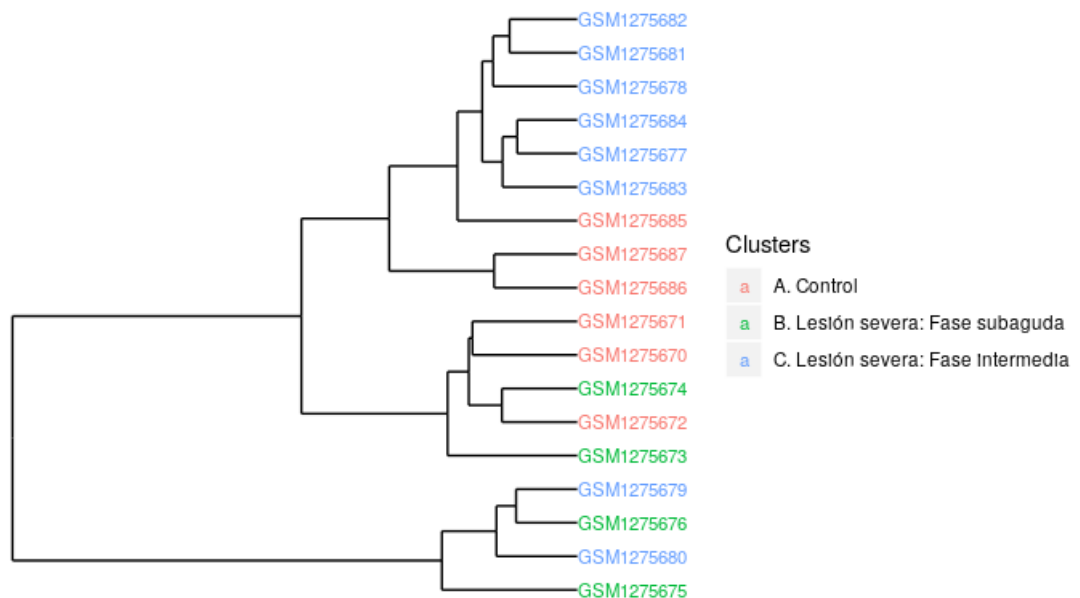


(b) Clustering estudio GSE93249

Figura A.4: Análisis exploratorio de las muestras del estudio GSE93249. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

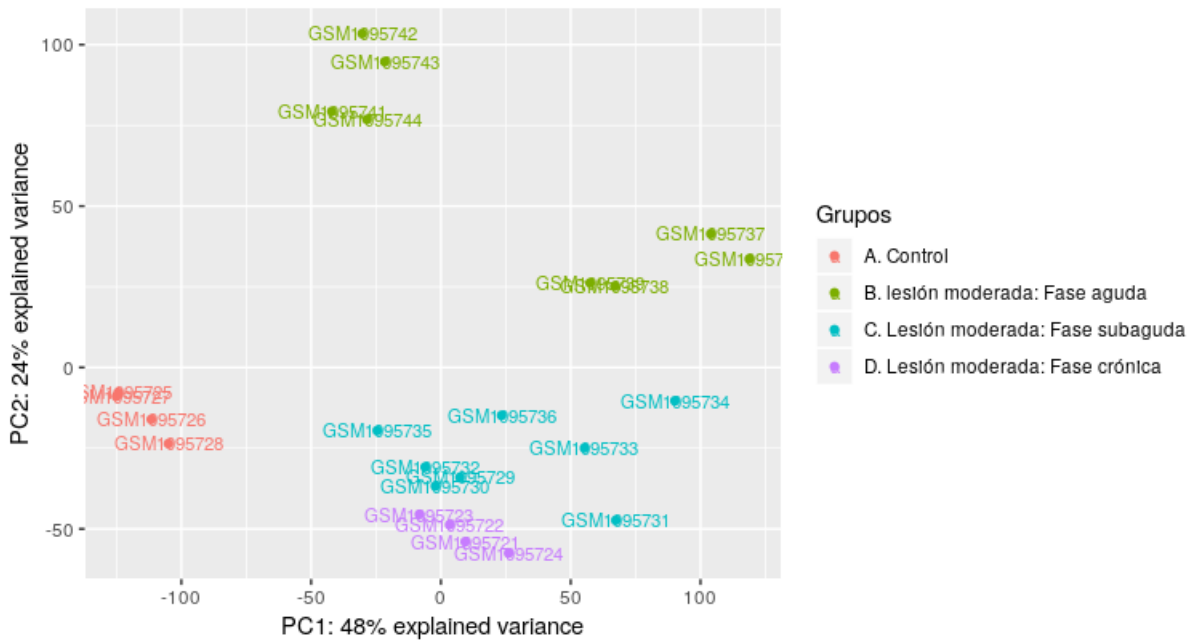


(a) PCA estudio GSE52763

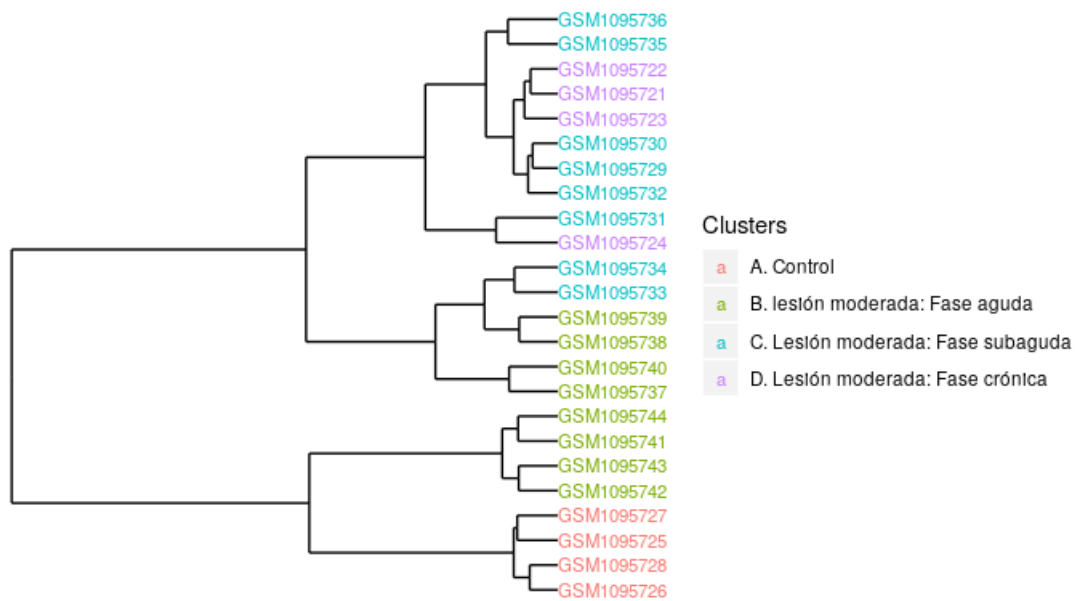


(b) Clustering estudio GSE52763

Figura A.5: Análisis exploratorio de las muestras del estudio GSE52763. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

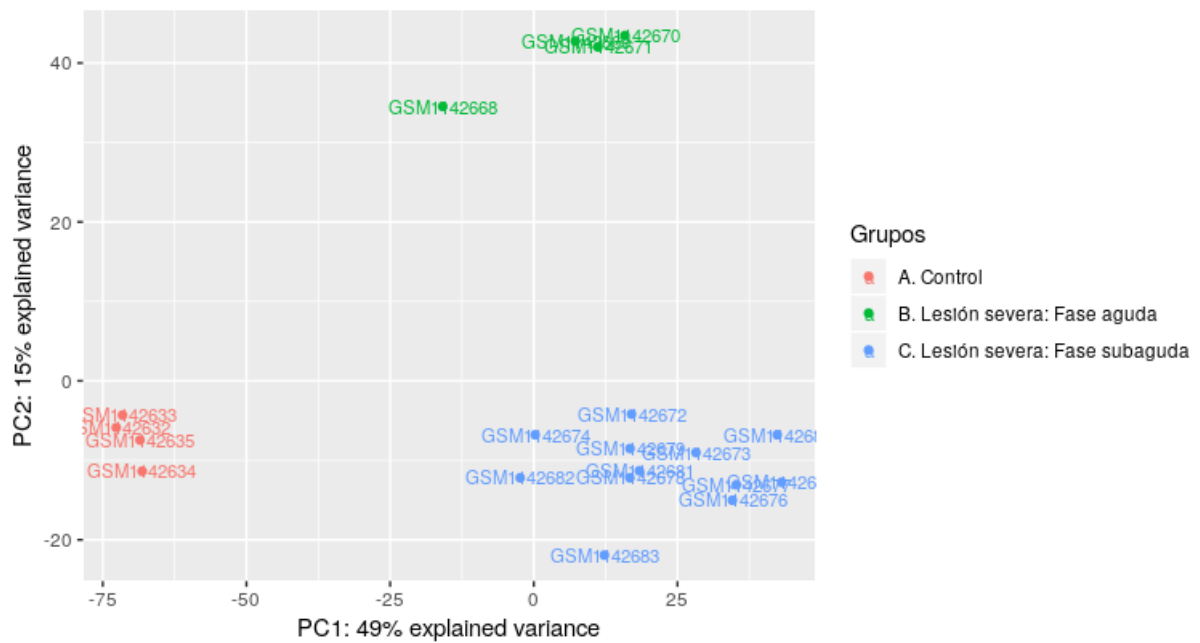


(a) PCA estudio GSE45006

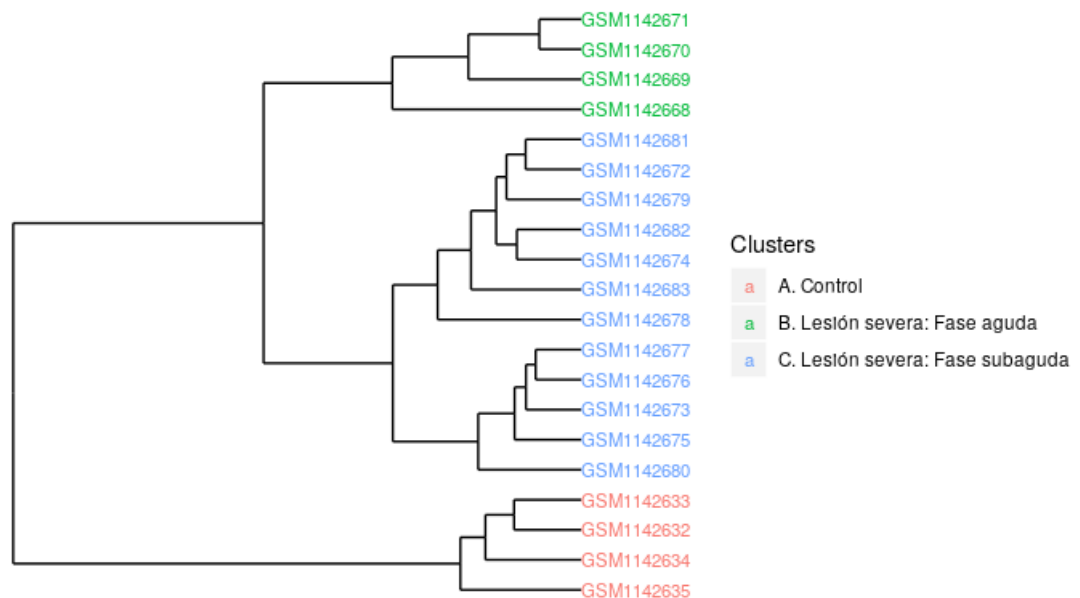


(b) Clustering estudio GSE45006

Figura A.6: Análisis exploratorio de las muestras del estudio GSE45006. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

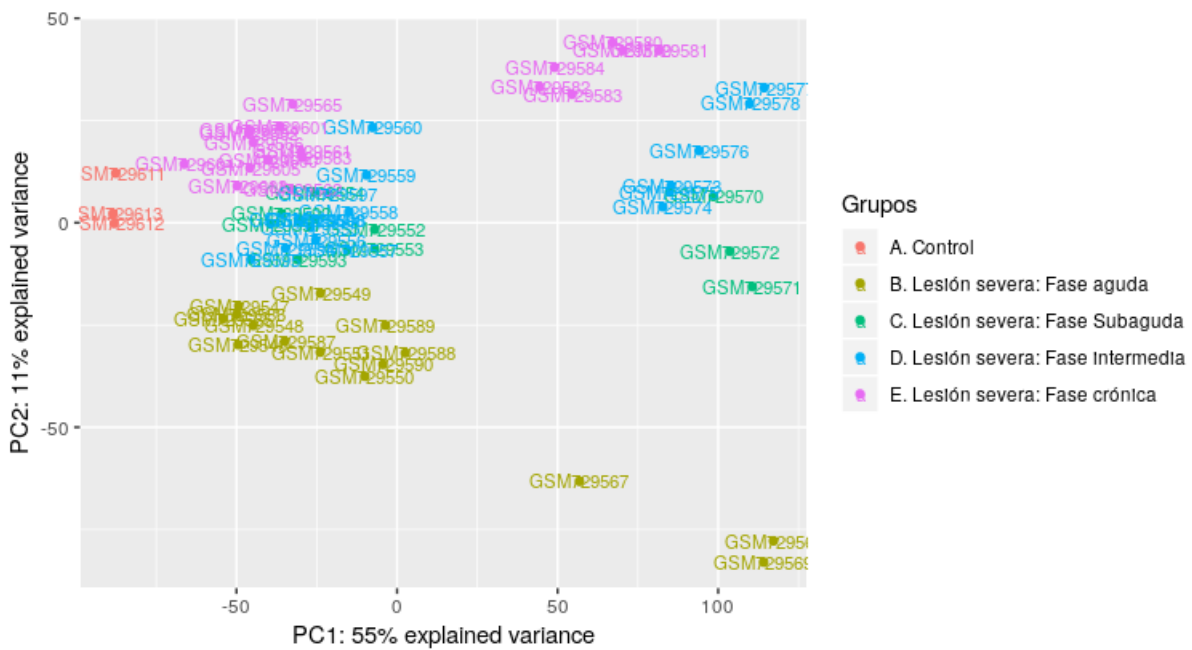


(a) PCA estudio GSE46988

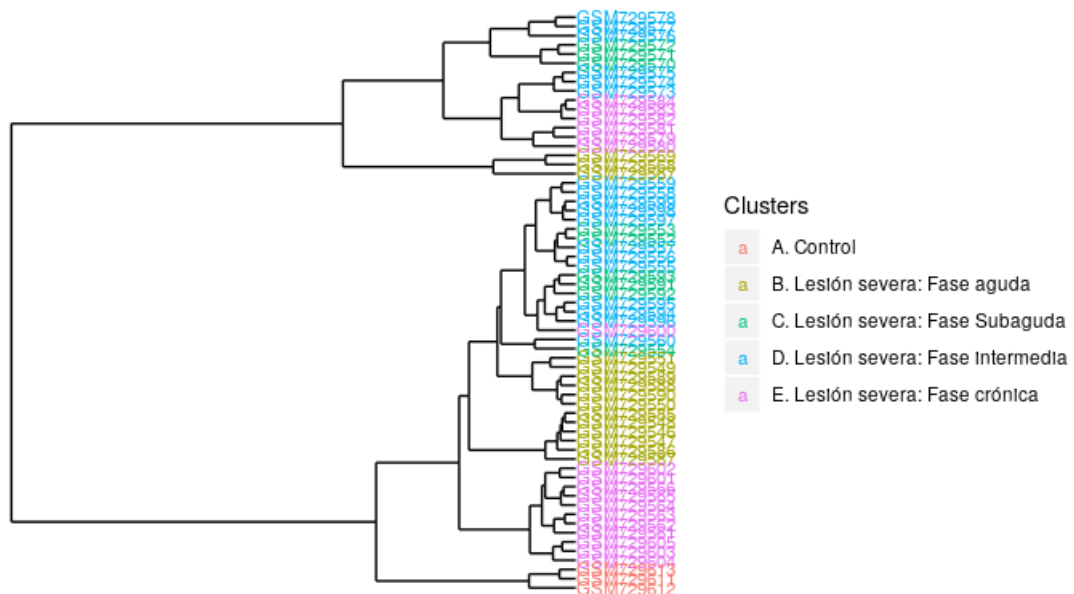


(b) Clustering estudio GSE46988

Figura A.7: Análisis exploratorio de las muestras del estudio GSE46988. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

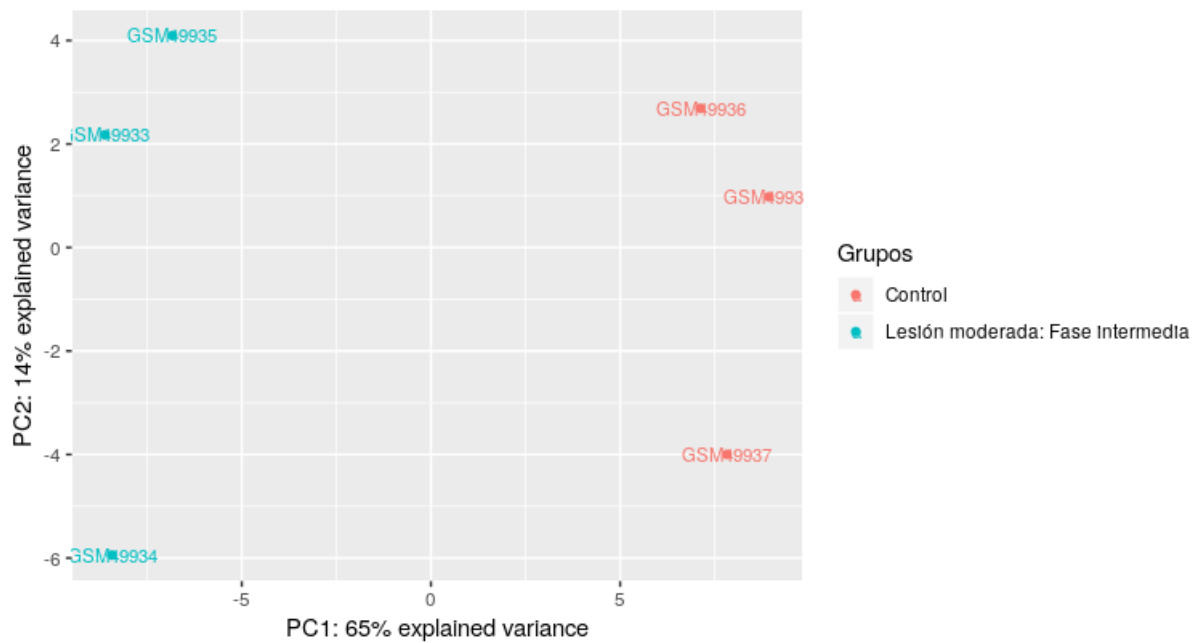


(a) PCA estudio GSE29488

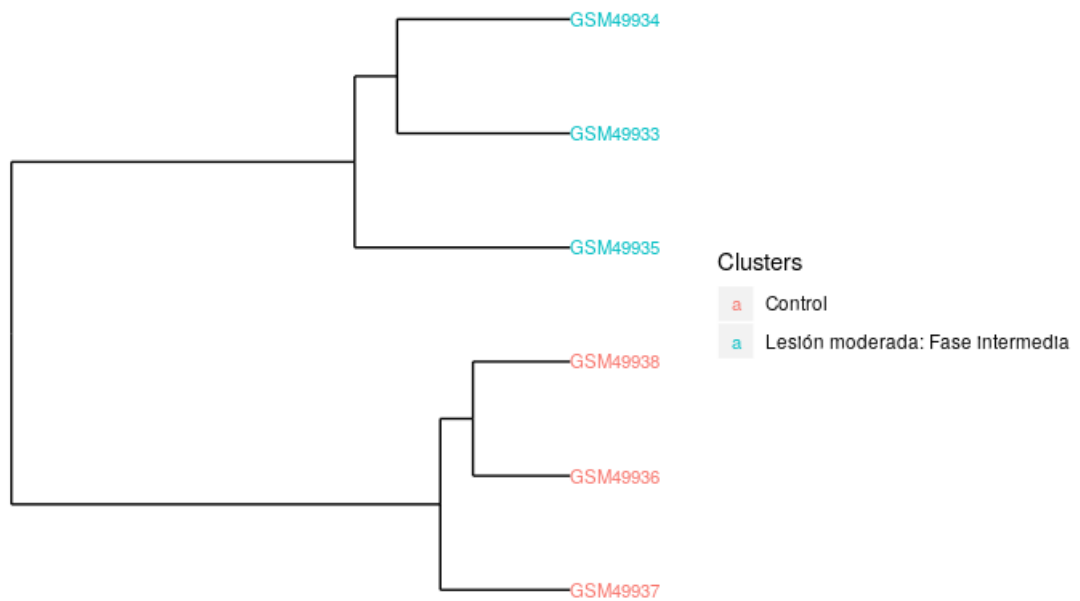


(b) Clustering estudio GSE29488

Figura A.8: Análisis exploratorio de las muestras del estudio GSE29488. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

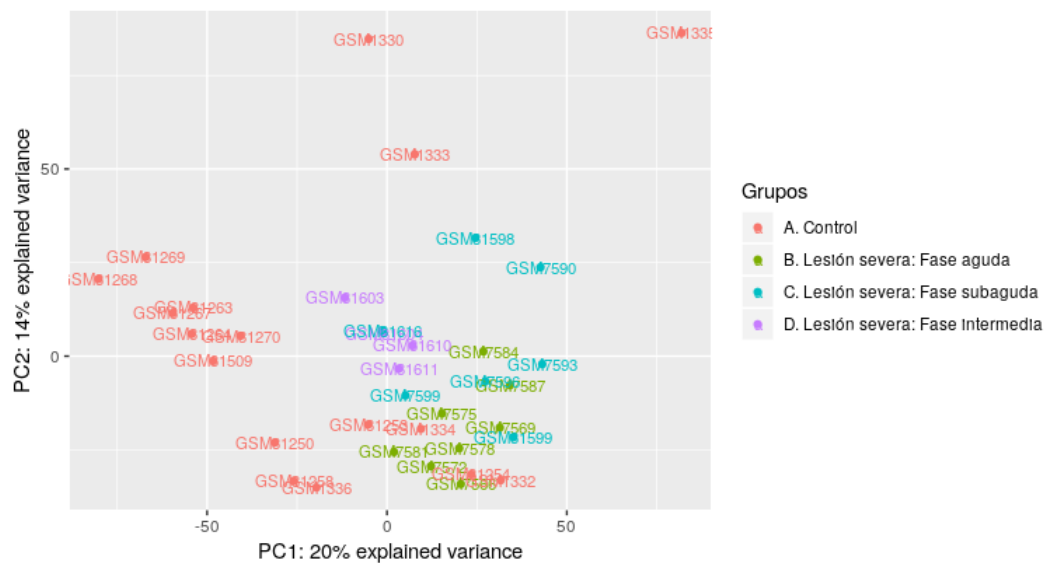


(a) PCA estudio GSE2599

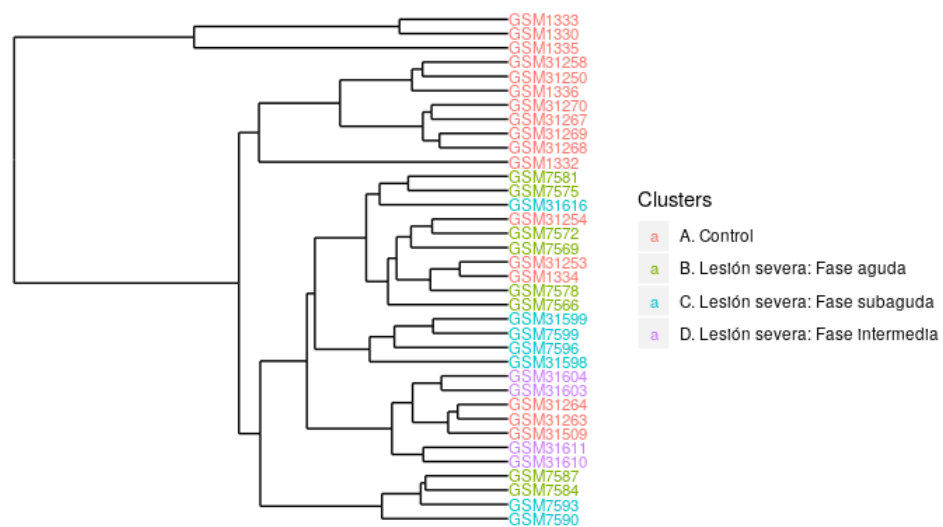


(b) Clustering estudio GSE2599

Figura A.9: Análisis exploratorio de las muestras del estudio GSE2599. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

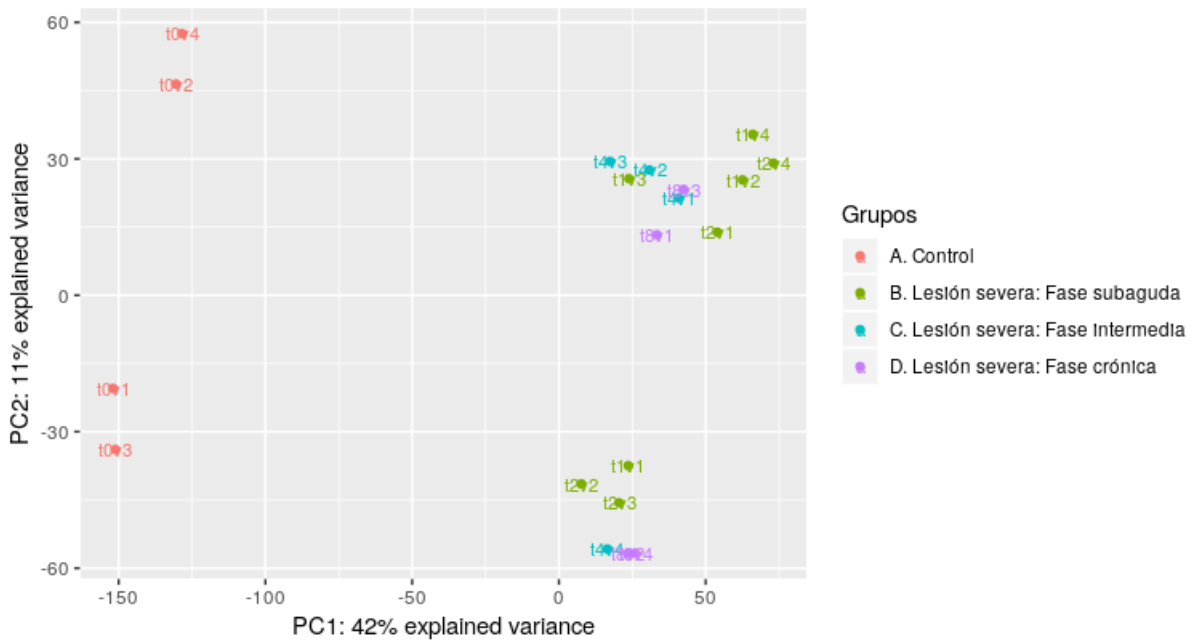


(a) PCA estudio GSE464 (lesión severa)

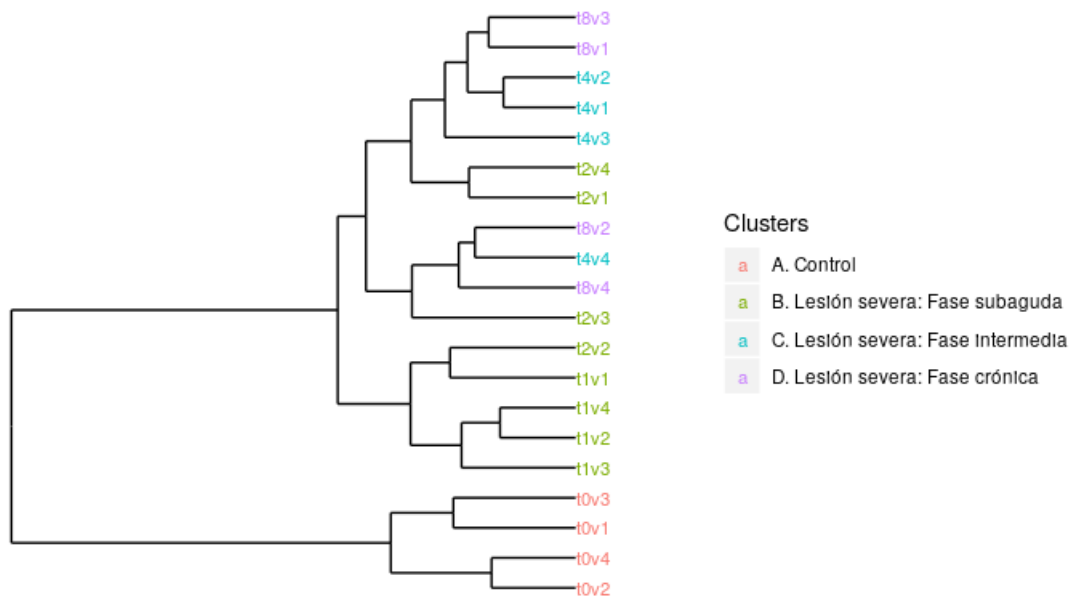


(b) Clustering estudio GSE464 (lesión severa)

Figura A.11: Análisis exploratorio de las muestras del estudio (lesión severa). El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.



(a) PCA estudio VBB



(b) Clustering estudio VBB

Figura A.12: Análisis exploratorio de las muestras del estudio VBB. El color indica el grupo experimental. **(a)** Análisis de componentes principales. Se compara la componente principal I con la componente principal II. **(b)** Clustering mediante distancias de correlación.

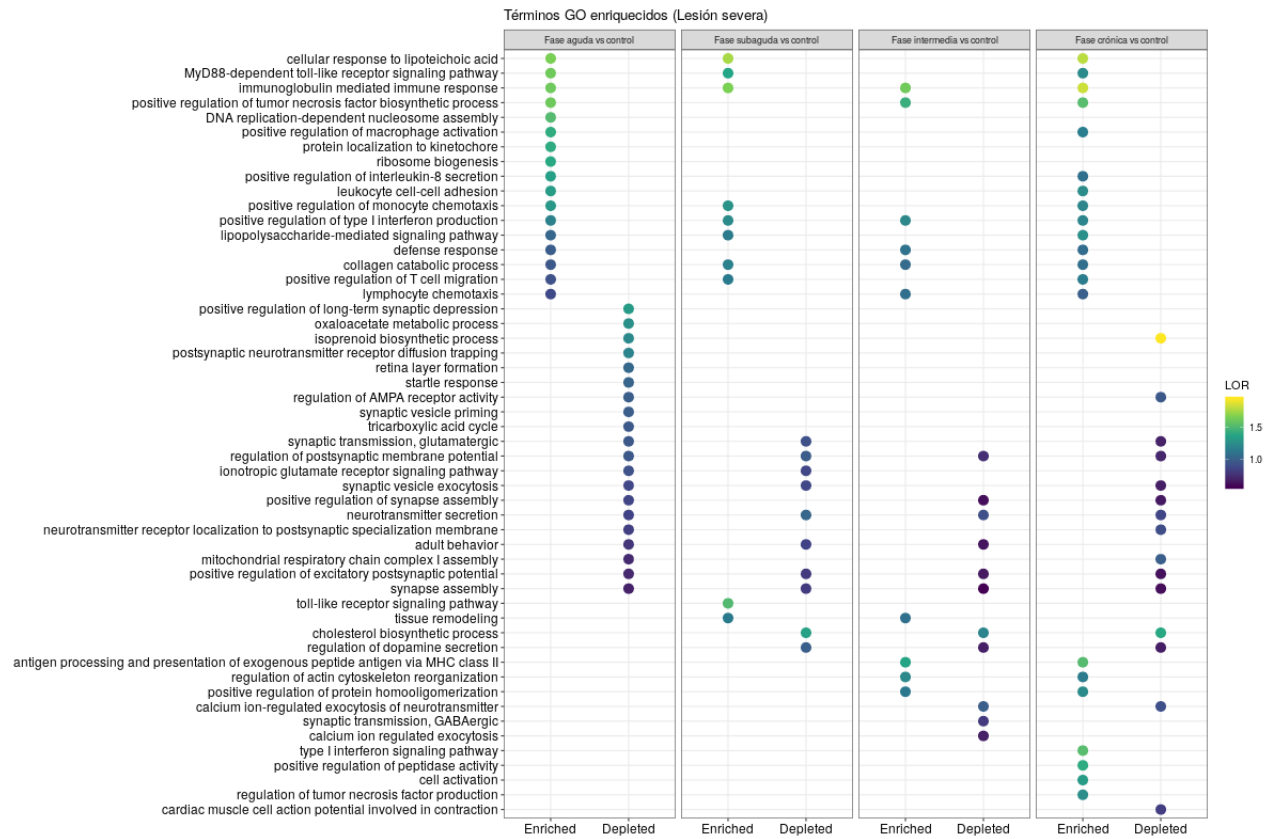


Figura A.13: Resumen de términos GO con mayor LOR en valor absoluto para lesiones severas. El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos. Para cada fase se muestran funciones ganadas tras la lesión (*Enriched*) y funciones que se pierden (*Depleted*).

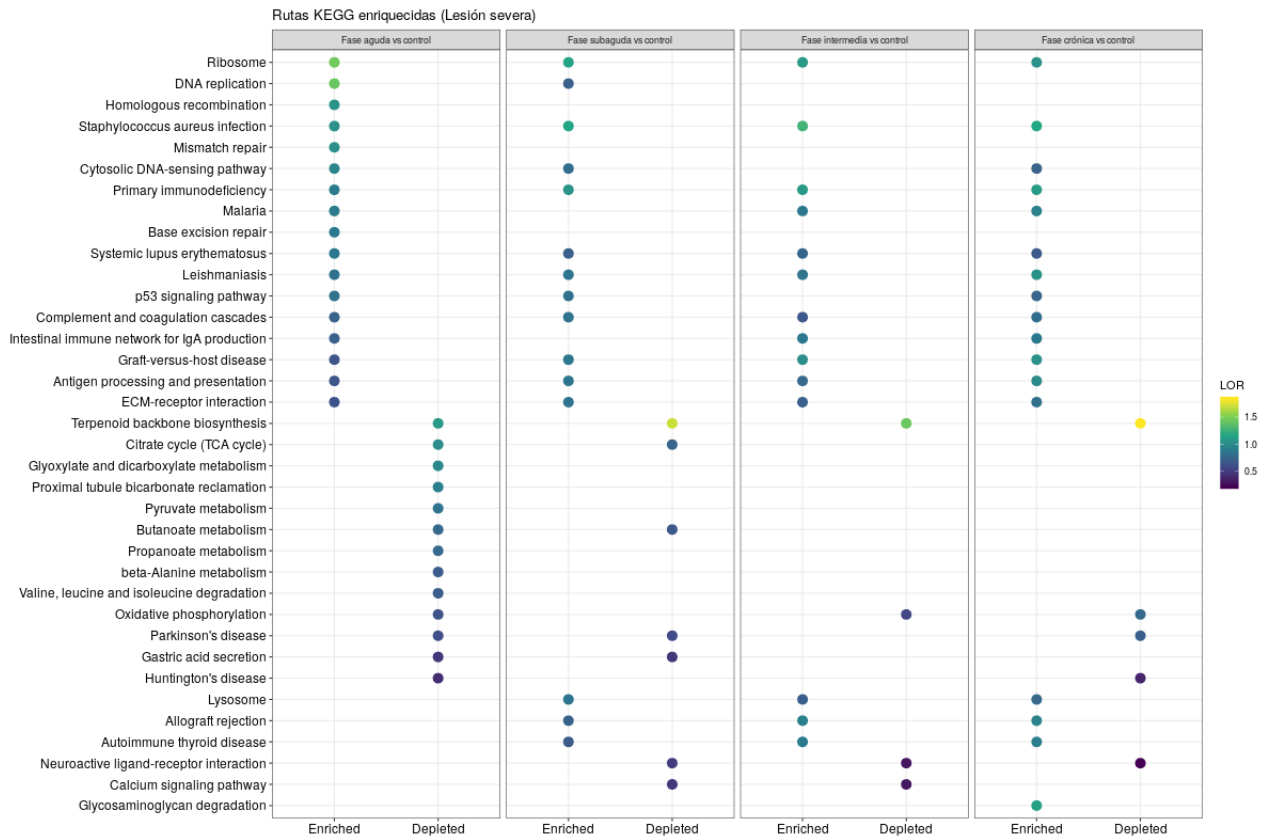


Figura A.14: Resumen de rutas KEGG con mayor LOR en valor absoluto para lesiones severas. El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos. Para cada fase se muestran funciones ganadas tras la lesión (**Enriched**) y funciones que se pierden (**Depleted**).

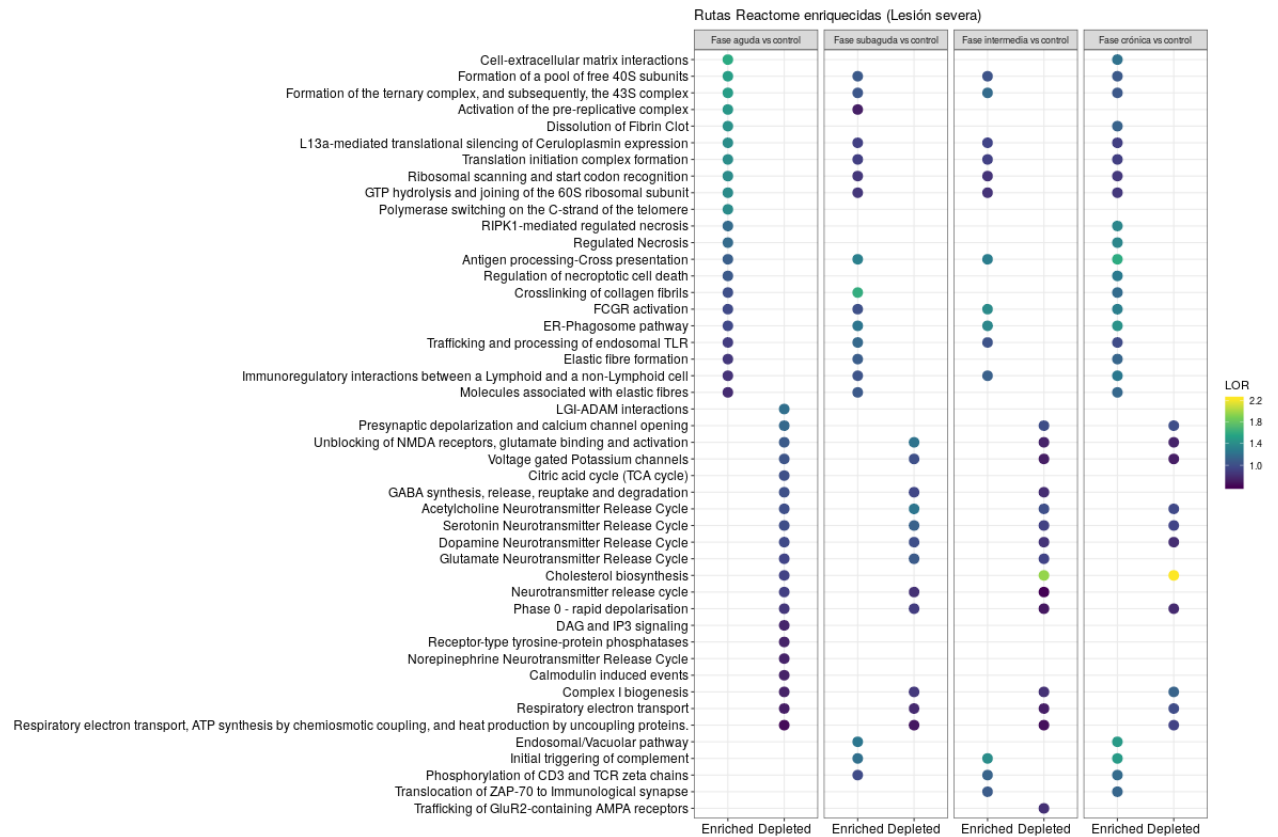


Figura A.15: Resumen de rutas Reactome con mayor LOR en valor absoluto para lesiones severas. El color de los puntos indica el valor del LOR, siendo los tonos amarillos los valores más altos, mientras que los tonos más oscuros representan valores más bajos. Para cada fase se muestran funciones ganadas tras la lesión (**Enriched**) y funciones que se pierden (**Depleted**).

B. Anexo B - Tablas

Tabla B.1: Versiones del software empleado en este trabajo.

Software	Versión
R	3.6.1

Paquete	Versión
edgeR	3.26.7
hipathia	2.0.0
AnnotationHub	2.16.0
metafor	2.1-0
rgu34a.db	3.2.3
ragene11sttranscriptcluster.db	8.7.0
ragene20sttranscriptcluster.db	8.7.0
rat2302.db	3.2.3
affy	1.62.0
biomaRt	2.40.0
reactome.db	1.68.0
mdgsa	1.16.0
metap	1.1
limma	3.40.2
ggplot2	3.1.1
org.Rn.eg.db	3.8.2
AnnotationDbi	1.46.0
Biobase	2.44.0
clusterProfiler	3.12.0