

**MÀSTER UNIVERSITARIO EN BIOINFORMÀTICA**



VNIVERSITAT  
DE VALÈNCIA

**TRABAJO DE FIN DE MÁSTER**

**DETECCIÓN Y CARACTERIZACIÓN DE LAS  
DIFERENCIAS DE GÉNERO EN LA ENFERMEDAD DEL  
HÍGADO GRASO NO ALCOHÓLICO DESDE UNA  
APROXIMACIÓN *ÓMICA***

**AUTOR:**

**JOSÉ F. CATALÀ SENENT**

**TUTORES:**

**FRANCISCO GARCÍA GARCÍA**

**MARTA R. HIDALGO GARCÍA**

**VICENTE ARNAU LLOMART**

**ENERO, 2019**





VNIVERSITAT  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria **ETSE-UV**

## **MÁSTER UNIVERSITARIO EN BIOINFORMÁTICA**

### **TRABAJO DE FIN DE MÁSTER**

# **DETECCIÓN Y CARACTERIZACIÓN DE LAS DIFERENCIAS DE GÉNERO EN LA ENFERMEDAD DEL HÍGADO GRASO NO ALCOHÓLICO DESDE UNA APROXIMACIÓN *ÓMICA***

**AUTOR:**

**JOSÉ F. CATALÀ SENENT**

**TUTORES:**

**FRANCISCO GARCÍA GARCÍA**

**MARTA R. HIDALGO GARCÍA**

**VICENTE ARNAU LLOMART**

---

### **TRIBUNAL:**

PRESIDENTE/A:

VOCAL 1:

VOCAL 2:

**FECHA DE DEFENSA:**

**CALIFICACIÓN:**



## Resumen

La enfermedad del hígado graso no alcohólico (NAFLD) es una de las patologías hepáticas más graves, estimándose una prevalencia de alrededor del 25 %. Al igual que ocurre en otras enfermedades como el cáncer de pulmón, el NAFLD afecta de forma diferente a hombres y mujeres, siendo más frecuente en los primeros.

Con el fin de entender los motivos de estas diferencias y abrir la puerta a tratamientos más específicos, este trabajo tiene como objetivo identificar los mecanismos moleculares afectados de forma diferente entre hombres y mujeres en el NAFLD. Para ello, se ha llevado a cabo una aproximación *in silico* con el objetivo de realizar un metaanálisis combinando la información de una selección de estudios *ómicos*.

**Palabras clave:** NAFLD, género, *in silico*, metaanálisis.

---

## Abstract

Non-alcoholic fatty liver disease (NAFLD) is one of the most serious liver pathologies, with an estimated prevalence of 25 %. As also occurs in other diseases, such as lung cancer, NAFLD affects in different ways men and women, being more frequent in the former than in the latter.

In order to understand the grounds for these differences and open the door to more specific treatments, this work aims to identify the molecular mechanisms affected differentially between men and women in the NAFLD. For this, an *in silico* approach has been carried out, with the goal of performing a meta-analysis combining the information of a selection of omic studies.

**Keywords:** NAFLD, gender, *in silico*, meta-analysis.



---

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Enfermedad del hígado graso no alcohólico . . . . .	1
1.1.1. Epidemiología . . . . .	2
1.1.2. Diagnóstico . . . . .	3
1.1.3. Desarrollo de la enfermedad y factores de riesgo . . . . .	4
1.1.4. Genética de la enfermedad . . . . .	5
1.2. Tecnologías de alto rendimiento y datos <i>ómicos</i> . . . . .	6
1.3. Análisis de datos <i>ómicos</i> . . . . .	9
1.3.1. Estrategias de análisis . . . . .	9
1.3.2. Caracterización funcional . . . . .	9
1.3.3. Metaanálisis . . . . .	12
<b>2. Objetivos</b>	<b>15</b>
<b>3. Materiales y métodos</b>	<b>17</b>
3.1. Revisión sistemática y selección de los estudios . . . . .	18
3.2. Análisis individual de los estudios . . . . .	19
3.2.1. Procesamiento de los datos . . . . .	19
3.2.1.1. Normalización de la nomenclatura de la enfermedad	19
3.2.1.2. Anotación de los chips de ADN . . . . .	20
3.2.1.3. Análisis exploratorio . . . . .	21
3.2.2. Análisis de expresión diferencial . . . . .	22

---

3.2.3.	Análisis de enriquecimiento funcional . . . . .	23
3.2.4.	Análisis de rutas de señalización . . . . .	24
3.3.	Metaanálisis . . . . .	26
3.3.1.	Enriquecimiento funcional . . . . .	27
3.3.1.1.	Evaluación de la heterogeneidad . . . . .	28
3.3.1.2.	Representación gráfica de los resultados . . . . .	29
3.3.2.	Rutas de señalización . . . . .	32
<b>4.</b>	<b>Resultados</b>	<b>33</b>
4.1.	Revisión sistemática y selección de los estudios . . . . .	34
4.2.	Análisis individual de los estudios . . . . .	36
4.2.1.	Procesamiento de los datos y análisis exploratorio . . . . .	36
4.2.2.	Análisis de expresión diferencial . . . . .	41
4.2.3.	Análisis de enriquecimiento funcional . . . . .	43
4.2.4.	Análisis de rutas de señalización . . . . .	46
4.3.	Metaanálisis . . . . .	49
4.3.1.	Enriquecimiento funcional . . . . .	49
4.3.1.1.	Términos GO para procesos biológicos . . . . .	49
4.3.1.2.	Términos GO para funciones moleculares . . . . .	50
4.3.1.3.	Términos GO para componentes celulares . . . . .	50
4.3.1.4.	Rutas KEGG . . . . .	59
4.3.2.	Rutas de señalización . . . . .	59
<b>5.</b>	<b>Discusión</b>	<b>61</b>
<b>6.</b>	<b>Conclusiones</b>	<b>65</b>
	<b>Bibliografía</b>	<b>67</b>
	<b>Anexo I: Figuras</b>	<b>79</b>
	<b>Anexo II: Tablas</b>	<b>93</b>

---

# Índice de figuras

1.1. Estadios del NAFLD . . . . .	2
1.2. Dogma central de la biología molecular . . . . .	7
1.3. Ruta del NAFLD en KEGG . . . . .	11
3.1. Flujo de trabajo. . . . .	17
3.2. Diagrama de flujo de una revisión sistemática siguiendo <i>PRISMA</i> . .	18
3.3. Diagrama de flujo de trabajo con el paquete <i>Hipathia</i> . . . . .	25
3.4. Representación la heterogeneidad en mediante gráficos de embudo .	30
3.5. Ejemplo de un gráfico de bosque . . . . .	31
4.1. Diagrama de flujo <i>PRISMA</i> de la revisión sistemática . . . . .	35
4.2. Diagrama de barras indicando número de pacientes por estudio . . .	36
4.3. PCA de las muestras del estudio GSE83452 . . . . .	40
4.4. Diagramas de Venn de los resultados individuales del GSEA . . . . .	45
4.5. Ruta hsa04621 en mujeres y hombres con NASH vs NAFL . . . . .	48
4.6. Gráfico de bosque para GO:0005583. MA de GO para c. celulares . .	53
4.7. Gráfico de bosque para GO:0036038. MA de GO para c. celulares . .	54
4.8. Gráfico de bosque para GO:0070971. MA de GO para c. celulares . .	55
4.9. Gráfico de bosque para GO:0098643. MA de GO para c. celulares . .	56
4.10. C. celulares significativos con sobrerrepresentación en las mujeres . .	57
4.11. C. celulares significativos con sobrerrepresentación en los hombres .	58
4.12. Representación de la subruta P-hsa04930-8 en KEGG . . . . .	60

---

I.1.	<i>Clustering</i> exploratorio de las muestras del estudio GSE48452 . . . . .	79
I.2.	PCA de las muestras del estudio GSE48452 . . . . .	80
I.3.	<i>Clustering</i> exploratorio de las muestras del estudio GSE61260 . . . . .	81
I.4.	PCA de las muestras del estudio GSE61260 . . . . .	82
I.5.	<i>Clustering</i> exploratorio de las muestras del estudio GSE66676 . . . . .	83
I.6.	PCA de las muestras del estudio GSE66676 . . . . .	84
I.7.	<i>Clustering</i> exploratorio de las muestras del estudio GSE83452 . . . . .	85
I.8.	PCA de las muestras del estudio GSE83452 . . . . .	86
I.9.	<i>Clustering</i> exploratorio de las muestras del estudio GSE89632 . . . . .	87
I.10.	PCA de las muestras del estudio GSE89632 . . . . .	88
I.11.	Diagramas de embudo para GO:0005583. MA de GO para c. celulares.	89
I.12.	Diagramas de embudo para GO:0036038. MA de GO para c. celulares.	90
I.13.	Diagramas de embudo para GO:0070971. MA de GO para c. celulares.	91
I.14.	Diagramas de embudo para GO:0098643. MA de GO para c. celulares.	92

---

# Índice de tablas

3.1. Grupos de estudio para las comparaciones. . . . .	22
4.1. Estudios seleccionados . . . . .	34
4.2. Individuos incluidos en el estudio GSE48452 . . . . .	37
4.3. Individuos incluidos en el estudio GSE61260 . . . . .	38
4.4. Individuos incluidos en el estudio GSE66676 . . . . .	38
4.5. Individuos incluidos en el estudio GSE83452 . . . . .	39
4.6. Individuos incluidos en el estudio GSE89632 . . . . .	41
4.7. Número de genes diferencialmente expresados en cada estudio . . .	42
4.8. Funciones significativas en el análisis GSEA . . . . .	43
4.9. Número de resultados significativos según <i>Hipathia</i> . . . . .	46
4.10. Subrutas significativas según <i>Hipathia</i> en el estudio GSE83452. . . . .	47
4.11. Resultados del metanálisis GSEA para GO de procesos biológicos . .	49
4.12. Resultados del metanálisis GSEA para GO de funciones moleculares	50
4.13. Resultados del metanálisis GSEA para GO de componentes celulares	50
4.14. Estimadores del MA para términos GO de componentes celulares . .	51
4.15. Resultados del metanálisis GSEA para rutas KEGG . . . . .	59
4.16. Subrutas según su nivel de significación en el MA con <i>Hipathia</i> . . . .	60
II.1. Versiones del software utilizado en este trabajo. . . . .	94
II.2. Rutas KEGG de <i>H. sapiens</i> implementadas en el paquete <i>hipathia</i> . .	95



---

# Abreviaturas

**ADN** Ácido desoxirribonucleico.

**ARN** Ácido ribonucleico.

**BH** Benjamini & Hochberg (método de corrección del p-valor, también FDR).

**BY** Benjamini & Yekutieli (método de corrección del p-valor).

**DL** DerSimonian & Laird (modelo de metaanálisis).

**FDR** *False discovery rate* (método de corrección del p-valor, también BH).

**FE** Efectos fijos (modelo de metaanálisis).

**GO** *Gene Ontology* (base de datos).

**GSEA** *Gene Set Enrichment Analysis*.

**GWAS** Estudios de asociación del genoma completo.

**HCC** Hepatocarcinoma.

**HE** Hedges *et al.* (modelo de metaanálisis).

**HS** Schmidt & Hunter (modelo de metaanálisis).

**KEGG** *Kyoto Encyclopedia of Genes and Genomes* (base de datos).

**LI** Límite inferior (en un intervalo de confianza).

**LOR** Logaritmo de los *odds ratio*.

**LS** Límite superior (en un intervalo de confianza).

**MA** Metaanálisis.

**miARN** microARN.

**MPA** *Mechanistic pathway activity*.

**NAFL** Esteatosis, acumulación de grasa en el hígado > 5 %.

**NAFLD** Enfermedad del hígado graso no alcohólico.

**NASH** Esteatohepatitis.

**NCBI** *National Center for Biotechnology Information*, Estados Unidos.

**PCA** Análisis de componentes principales.

**SMet** Síndrome metabólico.

**SNP** Polimorfismo de nucleótido único.

---

---

# CAPÍTULO 1

---

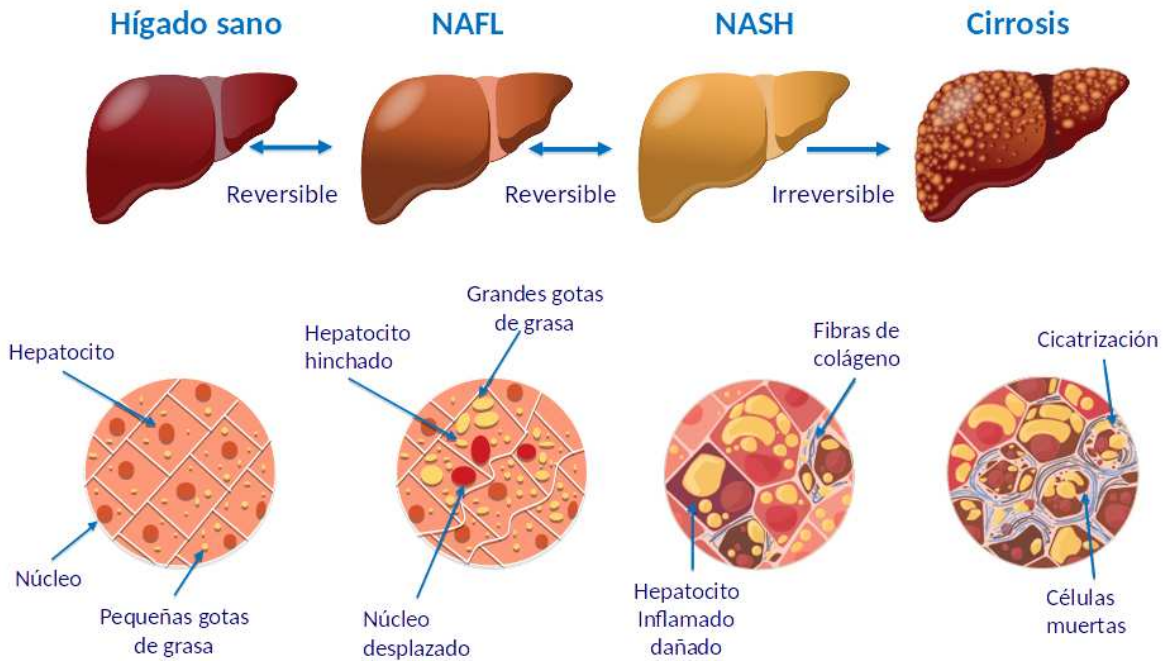
## Introducción

### 1.1 Enfermedad del hígado graso no alcohólico

---

La enfermedad del hígado graso no alcohólico (NAFLD, del inglés *Non-alcoholic fatty liver disease*) es un término genérico que abarca un amplio espectro de alteraciones histológicas hepáticas que varían en su gravedad y no se encuentran relacionadas con el consumo de alcohol. La más leve de estas condiciones es la esteatosis (NAFL, del inglés *Non-alcoholic fatty liver*), que consiste en una acumulación de grasa superior al 5% en el tejido hepático o los hepatocitos (células hepáticas) [1]. El siguiente estadio, la esteatohepatitis (NASH, del inglés *Non-alcoholic steatohepatitis*), implica la aparición de inflamación, daño en los hepatocitos y, normalmente, de fibrosis [2]. Con el avance de la enfermedad, los pacientes pueden desarrollar cirrosis y en los casos más graves un hepatocarcinoma (HCC) [3]. La Figura 1.1 muestra los estadios de la enfermedad.

En los últimos años ha aumentado el interés en esta enfermedad debido al crecimiento de su impacto en la salud mundial, llegando a ser considerada la enferme-



**Figura 1.1:** Estadios del hígado graso no alcohólico.

Modificado de <http://www.hcv-trials.com/nash/NAFLD-NASH.asp>

dad hepática más común en los países desarrollados [3], afectando a 1 de cada 4 individuos a nivel mundial.

### 1.1.1 Epidemiología

La prevalencia a nivel global del NAFLD se estima alrededor del 25%, aunque no es homogénea. Se calcula que la prevalencia es superior en América del Sur y en Oriente Medio (sobre un 31%), mientras que en África sólo es de un 13,48% [4]. La distribución entre sexos no está totalmente clara, existiendo contradicciones entre estudios [5, 6]. Sin embargo, los datos actuales indican una mayor prevalencia en hombres y mujeres postmenopáusicas que en mujeres premenopáusicas [7]. En el caso de las mujeres, algunos estudios indican que la prevalencia de la enfermedad puede aumentar a lo largo de la edad [8]. Múltiples estudios proponen el género como un factor a tener en cuenta en el abordaje de la enfermedad [5, 7, 9].

El hígado es el segundo órgano con mayor dimorfismo sexual en humanos. Tanto las funciones fisiológicas como las patogénicas del hígado difieren significativamente entre hombres y mujeres [10], encontrándose las principales diferencias en el metabolismo del colesterol, la detoxificación y la prevalencia de enfermedades como la hepatitis, el HCC y el propio NAFLD [11, 12, 13]. La expresión génica dimorfa afecta a cientos de genes [14] que podrían ser los responsables de la diferente prevalencia de estas enfermedades entre hombres y mujeres. Estas diferencias de expresión afectan a las propiedades tanto farmacocinéticas como farmacodinámicas del hígado de hombres y mujeres, teniendo un efecto en la progresión de la enfermedad en pacientes tratados con un medicamento y en la respuesta del hígado a dicho medicamento [15].

### 1.1.2 Diagnóstico

El método más ampliamente aceptado para el diagnóstico del NAFLD es la biopsia hepática. Esta técnica presenta inconvenientes, ya que es una prueba invasiva y presenta un riesgo de mortalidad relativamente elevado [16]. La gran mayoría de los pacientes con la enfermedad no presentan síntomas, lo que hace que el diagnóstico temprano sea complejo [6, 17]. La enfermedad se suele detectar a raíz de la aparición de indicios en pruebas clínicas rutinarias como análisis de sangre, evaluaciones médicas, pruebas de imagen médica [5, 18] o por la presencia de comorbilidades metabólicas [19].

Para el correcto diagnóstico de la enfermedad se requiere [16]:

- La confirmación de la esteatosis hepática por histología.
- La ausencia de un consumo significativo de alcohol.

- La ausencia de otras causas que provoquen esteatosis hepática como hepatitis C, enfermedad de Wilson o malnutrición.
- La ausencia de otras causas de enfermedades hepáticas crónicas como hemocromatosis, enfermedades hepáticas autoinmunes o hepatitis vírica crónica.

En los últimos años se han desarrollado técnicas de diagnóstico por imagen de la enfermedad, pero estas aún no son lo suficientemente sensibles como para detectar esteatosis en porcentajes inferiores al 30 % [6, 17], por lo que en este trabajo únicamente consideraremos la biopsia como método de diagnóstico definitivo [20].

### **1.1.3 Desarrollo de la enfermedad y factores de riesgo**

Tal y como se ha explicado previamente, el NAFLD abarca un espectro de alteraciones hepáticas de diferente gravedad (Figura 1.1) y, dependiendo del paciente, avanzará hasta un estadio u otro. El porcentaje de pacientes de NAFLD que llega a desarrollar NASH es de un 5-20 %, de los cuales un 10-20 % desarrolla fibrosis y menos de un 5 % acaba en un HCC [3]. Esto es de gran importancia, puesto que los pacientes que únicamente presentan NAFL tienen un bajo riesgo de complicaciones adversas mientras que aquellos que presentan NASH tienen mayor riesgo de padecer complicaciones [19].

El NAFLD puede progresar sin manifestaciones clínicas durante muchos años [21] y la llegada a los estadios finales de la enfermedad puede tardar décadas [22] o incluso no suceder. Sin embargo, esta enfermedad no solo supone un problema sanitario por sí misma, sino que se considera un factor de riesgo para otras enfermedades. Los pacientes con NAFLD presentan el doble de posibilidades de fallecer a causa de una enfermedad cardiovascular que de una enfermedad hepática, debido a la gran cantidad de factores de riesgos compartidos, como son la diabetes, la hipertensión y la obesidad [23]. De hecho, la muerte por enfermedad hepática es la tercera

causa de fallecimiento en pacientes con NAFLD, precedida por las enfermedades cardiovasculares y el cáncer [24].

El síndrome metabólico (SMet) hace referencia a la concurrencia de varias situaciones clínicas como la resistencia a la insulina, la obesidad, la dislipidemia aterogénica o la hipertensión [25]. Estas situaciones se consideran comorbilidades del NAFLD [16, 20] y uno de los mayores factores de riesgo en la enfermedad [19]. Se estima que la prevalencia del hígado graso no alcohólico en pacientes con SMet es aproximadamente el 83 %, siendo mayor en hombres que en mujeres [26].

#### 1.1.4 Genética de la enfermedad

El papel de la genética en la enfermedad no está totalmente claro, pero se ha demostrado la asociación de algunos genes con la misma. La predisposición genética a la enfermedad parece estar vinculada a mutaciones en genes que afectan al contenido en lípidos de los hepatocitos, al metabolismo y a la inflamación del hígado [27].

Se han realizado estudios de asociación del genoma completo (GWAS, del inglés *Genome Wide Association Studies*) con el objetivo de determinar mutaciones puntuales (SNPs, del inglés *Single Nucleotide Polymorphisms*) en genes que puedan ser responsables de una mayor prevalencia del NAFLD. Estos estudios han mostrado que polimorfismos en el gen *PNPLA3* están asociados con un mayor riesgo de padecer la enfermedad [28], ya que este gen codifica para una proteína involucrada en el metabolismo energético y que es responsable de la hidrólisis de los triacilglicérols [29]. La variante más común en el gen *PNPLA3* es una mutación en la posición 148 del gen (identificador en la base de datos *dbSNP* [30]: rs738409) que causa que la proteína pierda su actividad enzimática y acaba provocando una acumulación en los hepatocitos de triacilglicérols en forma de gotas de grasa [29, 31]. Otros

de los genes en los que se han encontrado SNPs relacionados con el riesgo a sufrir NAFLD son *CETP* [32], *SREBF* [33], *TMS6SF2* [34], *NCAN*, *GCKP*, *LYPLAY1* [35], *CPN1* y *HSD17B13* [36]; que forman parte del metabolismo de lípidos, inflamación, resistencia a la insulina, estrés oxidativo o apoptosis [29].

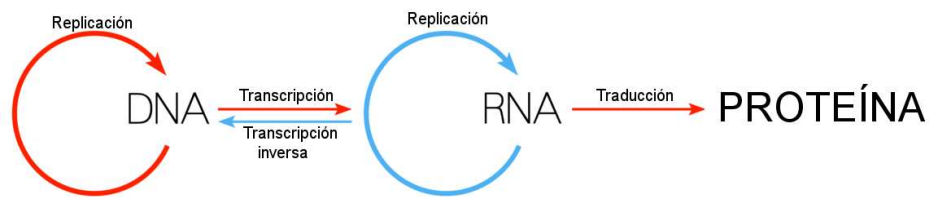
Existen pocos estudios de los efectos de la epigenética en el NAFLD. Algunos trabajos asocian la hipermetilación de algunas regiones promotoras de los genes *PPAR* con un mayor riesgo de sufrir fibrosis [37], pero son necesarios más estudios para entender este proceso.

Los microARNs (miARNs) también pueden jugar un papel importante en la enfermedad. Actualmente se asocian miR-122, miR-21 y miR-34a con la patogénesis del NAFLD [31, 38]. Se piensa que otros miARNs también están relacionados con el hígado graso no alcohólico, pero aún no hay evidencias que lo confirmen definitivamente [31].

## **1.2 Tecnologías de alto rendimiento y datos *ómicos***

---

Desde la formulación del dogma central de la biología molecular por Francis Crick en 1970 [39] (ver Figura 1.2), el interés por el estudio de todos los procesos implicados en el mismo no ha parado de aumentar. Este interés, junto a los grandes avances de la tecnologías existentes, ha llevado al desarrollo de las llamadas técnicas *ómicas*. La primera de estas técnicas, la genómica, fue definida en 1986 por Thomas Roderick como una "subdisciplina de la genética dedicada al estudio de la cartografía, secuenciación y análisis de las funciones de genomas completos". Posteriormente surgieron el resto de *ómicas* como la transcriptómica, la proteómica, la metagenómica o la epigenómica. Todas estas técnicas destacan por su carácter interdisciplinar, puesto que requieren de la biología, la estadística y la informática para el tratamiento de la enorme cantidad de datos que generan.



**Figura 1.2:** Dogma central de la biología molecular. En rojo se muestra la propuesta inicial de Francis Crick y en azul las modificaciones posteriores.

Modificada de: <https://commons.wikimedia.org/wiki/File:CentralDogma-es.png>

Esta gran cantidad de datos generados por estas nuevas técnicas ha supuesto un reto para el almacenamiento, ordenación y clasificación de los mismos que ha desembocado en la aparición de bases de datos, tanto públicas como privadas, en las que estos datos son depositados. La revista *Nucleic Acids Research* publica regularmente un número especial sobre bases de datos biológicas y, en su última revisión (enero de 2018), aparecen un total de 1737 bases de datos [40], lo que nos puede dar una idea de la magnitud de los datos generados.

La publicación de los datos generados por las investigaciones en bases de datos públicas no solo ayuda a la reproducibilidad de los experimentos, sino que también permite su uso en investigaciones posteriores por parte de la comunidad científica. Sin embargo, para que esto sea posible, es necesario el uso de formatos y estándares definidos que permitan compartir y utilizar los datos de forma eficiente [41, 42, 43]. En los últimos años han aparecido varias iniciativas que buscan este objetivo, siendo una de las más extendidas los principios FAIR (del inglés *Findable, Accessible, Interoperable, Re-usable*) [44] que ha sido adoptada por el *European Research Council* en los proyectos incluidos dentro del programa Horizonte 2020.

En este trabajo hemos utilizado datos disponibles en bases de datos públicas para realizar un estudio *in silico*, principalmente aquellos datos disponibles en la plataforma GEO (<https://www.ncbi.nlm.nih.gov/geo/>, *National Center for Biotechnology Information*) o en la plataforma ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>, *European Bioinformatics Institute*). Ambas

bases de datos son ampliamente utilizadas por la comunidad científica, se encuentran interconectadas y presentan la ventaja de que sus datos siguen los estándares MIAME (*Minimum Information About A Microarray Experiment*) o MINSEQE (*Minimum Information About a Sequencing Experiment*).

## **Transcriptómica y *microarrays***

La transcriptómica es el estudio del transcriptoma (el conjunto de transcritos de ARN producidos por el genoma bajo circunstancias específicas o en una célula determinada) utilizando métodos de alto rendimiento, como *microarrays* o *RNA-Seq*.

Los chips, arrays de ADN o *microarrays* son una serie de sondas de ADN unidas a un soporte sólido en una disposición regular y prefijada. El ácido nucleico diana que será detectado puede ser ADN o ARN y previamente a la hibridación debe ser marcado con una sustancia fluorescente o radiactiva. La principal ventaja con respecto a las técnicas de biología molecular, como la reacción en cadena de la polimerasa, es que pueden detectarse en un único procesamiento miles de genes [45]. La aparición de los arrays de ADN en 1996 [46, 47] supuso el primer abordaje global en estudios *ómicos* [48], permitiendo la identificación de genes que se expresan de forma diferencial en distintas poblaciones celulares o en respuesta a diferentes tratamientos.

El *RNA-Seq* es una técnica de análisis del transcriptoma mediante secuenciación masiva. Este método permite medir de forma más precisa los transcritos y obtener información de las diferentes isoformas del ARN mensajero.

En este trabajo, debido a los resultados de la revisión sistemática, nuestros datos de transcriptómica se obtienen de chips de ADN.

---

## 1.3 Análisis de datos *ómicos*

---

### 1.3.1 Estrategias de análisis

Existen múltiples estrategias para el análisis de datos *ómicos* que variarán en función del tipo de datos y de los objetivos del estudio.

En los estudios transcriptómicos es necesario realizar un preprocesamiento de los datos, que dependerá de la tecnología de alto rendimiento utilizada, de forma que obtengamos una matriz que contenga la cuantificación de cada unidad biológica de interés (gen, transcrito, etc.) para cada una de las muestras. Una vez obtenida esta matriz las estrategias más comunes son:

- ANÁLISIS DE EXPRESIÓN DIFERENCIAL. Permite detectar grupos de elementos con un diferente nivel de expresión en las distintas condiciones o poblaciones analizadas en el estudio.
- PREDICCIÓN DE CLASES. Mediante un entrenamiento a partir de muestras iniciales, este método es capaz de generar un predictor supervisado capaz de clasificar nuevas muestras [49].
- ANÁLISIS CLUSTER. Agrupa las muestras o las unidades biológicas utilizando un patrón de expresión común. Existen técnicas de *biclustering* que permiten agrupar de forma simultánea muestras y unidades biológicas [50].

### 1.3.2 Caracterización funcional

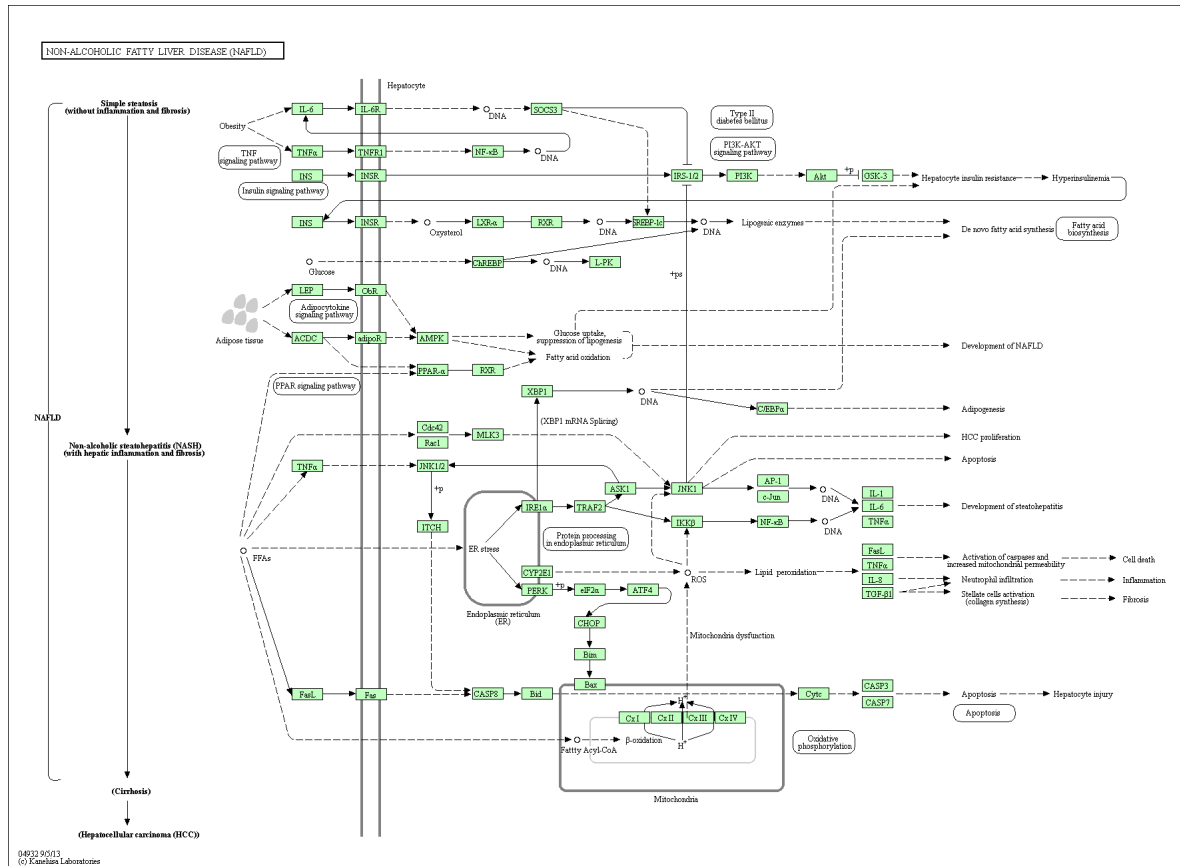
Las estrategias de análisis vistas en el Apartado 1.3.1 facilitan información sobre los elementos biológicos, ya sea su clasificación en grupos o su ordenación en ba-

se a criterios definidos, que puede ofrecer una respuesta a las hipótesis planteadas por los estudios. Sin embargo, una caracterización funcional proporcionará unos resultados más fácilmente interpretables. Para realizar esta caracterización funcional necesitamos bases de datos que relacionen nuestras unidades biológicas con funciones definidas. Estas anotaciones funcionales pueden ser obtenidas de diversas bases de datos:

- *Gene Ontology* (GO) [51] es una base de datos que contiene ontologías entre los genes y sus funciones. Estas ontologías se encuentran clasificadas en procesos biológicos, funciones moleculares y componentes celulares.
- *Kyoto Encyclopedia of Genes and Genomes* (KEGG) [52] es una base de conocimiento sobre las funciones de los genes. El componente principal de KEGG es la base de datos *PATHWAY*, que consta de diagramas de rutas bioquímicas que incluyen la mayoría de las rutas metabólicas y algunas de las rutas reguladoras conocidas. En la Figura 1.3 podemos ver, como ejemplo, la representación de la ruta del NAFLD en KEGG.

La integración de la información procedente de los datos experimentales y la información biológica disponible en las bases de datos, requiere de métodos para la caracterización o enriquecimiento funcional de los estudios transcriptómicos. Estos métodos se pueden agrupar en métodos de análisis de sobrerrepresentación y métodos de grupos de genes [53].

- Los métodos de ANÁLISIS DE ENRIQUECIMIENTO O ANÁLISIS DE SOBRRERPRESENTACIÓN [54, 55] caracterizan las funciones de un grupo de genes (u otra unidad biológica) de interés. Esto lo realizan en dos fases: primero se selecciona el grupo de genes de interés siguiendo un criterio (por ejemplo, el nivel de expresión diferencial) y, a continuación, las funciones anotadas de estos genes se comparan contra las funciones anotadas en un grupo de genes de referencia (por ejemplo, el genoma de referencia menos los genes iniciales u otra lista



**Figura 1.3:** Ruta "hsa04932" de KEGG que corresponde a la enfermedad del hígado graso no alcohólico.

de genes). Posteriormente, se evalúa estadísticamente la proporción de genes asociados a cada una de las funciones en cada una de las dos listas, pudiendo detectar así una sobrerrepresentación significativa en algunas de ellas [48]. Este análisis presenta limitaciones, como la decisión del criterio de selección: dependiendo de dónde se establezca el límite que marque que los genes se encuentren en uno u otro grupo los resultados obtenidos variarán. Este problema puede solventarse con los análisis de grupos de genes.

- Los métodos de ANÁLISIS DE GRUPOS DE GENES consideran todos los genes evaluados en el estudio, pero ordenados según un criterio relevante para el investigador [48]. El primero de estos análisis fue el *Gene Set Enrichment Analysis* (GSEA) [56, 57], que tiene como objetivo, una vez ordenados los genes, averi-

guar si los genes pertenecientes a un mismo grupo funcional tienden a agruparse en la parte superior o inferior de la lista, lo que se interpreta como una sobrerrepresentación significativa de una función.

- Los métodos MPA (del inglés, *Mechanistic pathway activity*) son una evolución de los métodos anteriores. Estas herramientas tienen en cuenta las topologías de las redes y subredes de activación para obtener una medida más precisa de la actividad de las funciones celulares, a partir de los niveles de expresión de los genes.

### 1.3.3 Metaanálisis

La investigación clínica tiene como objetivo la obtención de resultados claros y de calidad que puedan ser utilizados como base en la práctica clínica [58], sin embargo, los estudios clínicos no siempre consiguen este objetivo y, en ocasiones, contrastan entre ellos [59]. Una forma de abordar este problema es mediante el metaanálisis (MA), una metodología estadística que permite la obtención de un resultado combinado a partir de los resultados de estudios individuales.

El metaanálisis se puede definir, en términos generales, como la revisión sistemática y la síntesis de estudios relacionados pero independientes [60, 61, 62]. Las principales ventajas de los metaanálisis son que permiten obtener una medida del efecto combinada con una mayor precisión que la de los estudios individuales incluidos en la revisión sistemática y, por lo tanto, una mayor potencia estadística; cuantificar la variabilidad de los resultados de los estudios y aproximarse a preguntas no planteadas en los análisis individuales explicando las diferencias entre sus resultados [63]. Sin embargo, esta metodología presenta limitaciones como el uso de métodos estadísticos inadecuados o una interpretación errónea de los resultados [63, 58].

La utilización del metaanálisis comenzó en el campo de las ciencias sociales y la psicología como una solución a las irregularidades que surgían en los estudios individuales debidas a los diferentes contextos, tipos de sujetos y otros factores con los que se realizaban estos estudios [64]. Con el paso del tiempo el uso de esta metodología se extendió a diferentes áreas de la ciencia, incluyendo la medicina y la genómica [60].



---

---

## CAPÍTULO 2

---

### Objetivos

El NAFLD es una enfermedad que abarca un espectro de alteraciones hepáticas de diversa gravedad clínica. El paso de NAFL a NASH, es decir, de una excesiva acumulación de grasa en el hígado a una inflamación de las células hepáticas, es de gran importancia, ya que supone un incremento grave de los riesgos sanitarios para la salud de los pacientes. Adicionalmente, se conoce que la enfermedad afecta de forma diferente a ambos géneros, siendo mayor la prevalencia en hombres que en mujeres.

Por todo esto, y con el fin de entender los fundamentos subyacentes en las diferencias de género observadas en la enfermedad del hígado graso no alcohólico, este trabajo busca **identificar las bases moleculares afectadas de forma diferencial entre hombres y mujeres** que padecen la enfermedad, especialmente **en el paso de NAFL a NASH**, mediante un abordaje *in silico*.



---

## CAPÍTULO 3

---

# Materiales y métodos

A continuación se describirán los procedimientos realizados en este trabajo, basados en la revisión sistemática para la selección de estudios y su posterior metaanálisis. El flujo de trabajo seguido se muestra en la Figura 3.1.

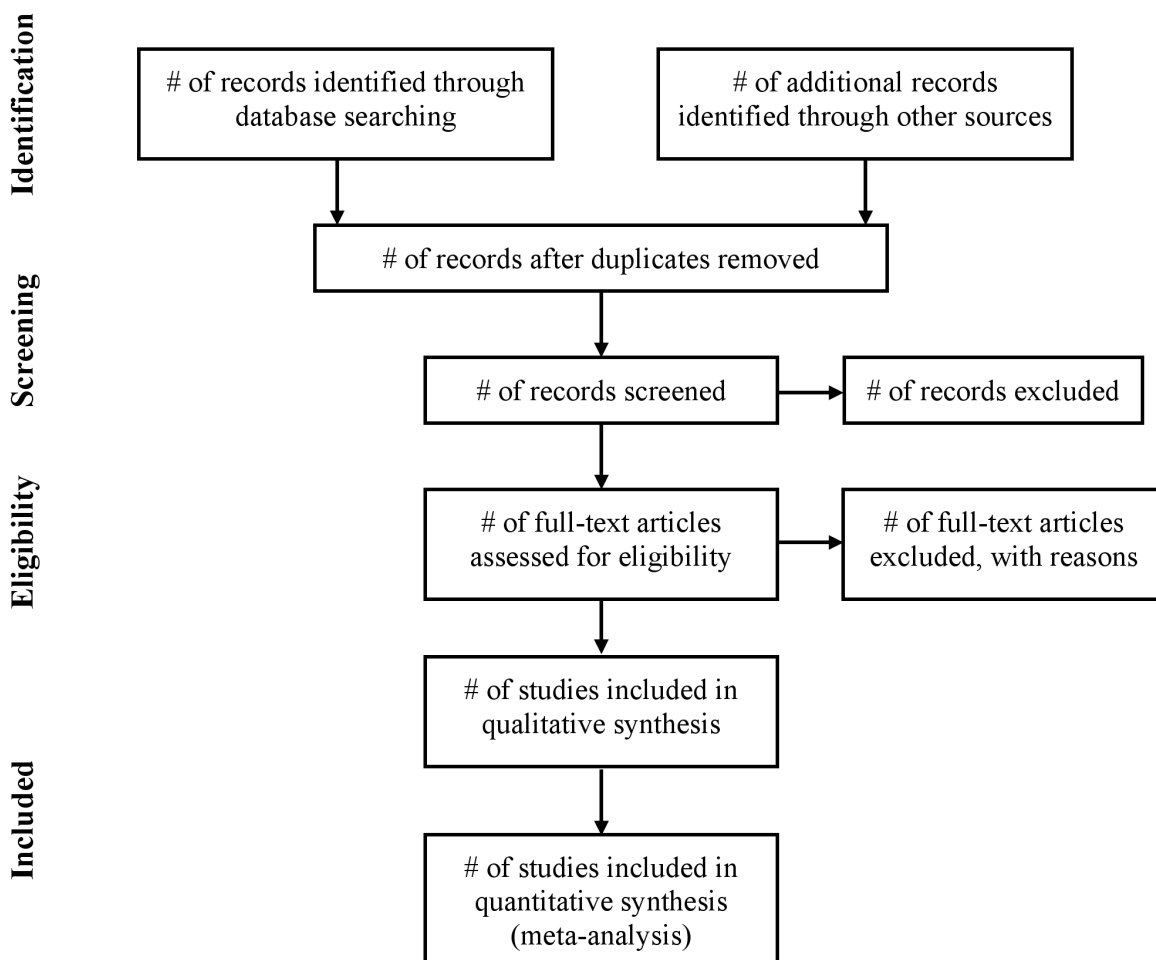


**Figura 3.1:** Flujo de trabajo.

Todo el análisis bioinformático fue realizado con el lenguaje de programación R [65]. Las librerías o paquetes indicados en cada uno de los pasos, que se encuentran detalladas en la Tabla II.1 del Anexo II.

### 3.1 Revisión sistemática y selección de los estudios

Para llevar a cabo un metaanálisis es necesaria, previamente, la realización de una revisión exhaustiva y sistemática de los datos disponibles. La revisión sistemática tiene como objetivo recolectar toda evidencia empírica que se ajuste a los de criterios de idoneidad, previamente establecidos, para responder a una pregunta de investigación [62]. Las directrices para la realización de una correcta revisión sistemática se encuentran publicadas en la *declaración PRISMA* [61, 62] que incluye un diagrama con los pasos necesarios a seguir (Figura 3.2).



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

<https://doi.org/10.1371/journal.pmed.1000100.g001>

El primer paso de nuestra revisión sistemática consistió en la búsqueda exhaustiva de estudios disponibles en bases de datos públicas sobre la enfermedad del hígado graso no alcohólico. Para ello nos centramos en las bases de datos GEO y ArrayExpress, buscando en ellas las palabras clave: *NAFLD*, *NAFL*, *steatosis*, *NASH* y *steatohepatitis*.

De los estudios resultantes de la búsqueda, se seleccionaron aquellos que estaban relacionados con la enfermedad, contenían información de género e incluían los diferentes estadios del NAFLD. Además, también se eliminaron aquellos realizados con especies diferentes de *Homo sapiens* o con un número de pacientes muy bajo.

La información y los datos de los estudios seleccionados se descargaron mediante el paquete GEOquery [66] para su posterior procesado y análisis.

## 3.2 Análisis individual de los estudios

---

### 3.2.1 Procesamiento de los datos

#### 3.2.1.1. Normalización de la nomenclatura de la enfermedad

Debido a que el NAFLD es una enfermedad cuyo estudio se ha intensificado en las últimas décadas, la nomenclatura de la misma ha sufrido ligeras modificaciones en los últimos años. Para evitar problemas al integrar los datos de los estudios en el metaanálisis se renombraron los estadios de la enfermedad de la siguiente manera (ver Figura 1.1):

1. **Control:** Pacientes sanos, incluye todos los pacientes que no presenten la enfermedad, sean obesos o no.

2. **NAFLD:** Pacientes con enfermedad, sin especificar el estadio de la misma que presentan.
3. **NAFL o *steatosis*:** Pacientes que presentan el estadio inicial de la enfermedad.
4. **NASH:** Pacientes que presentan un estadio mas avanzado de la enfermedad, caracterizado por la presencia de inflamación del tejido hepático.
5. **NA o indefinido:** Pacientes cuyo estadio no se incluía en los casos anteriores o no estaba claramente definido en los estudios.

#### 3.2.1.2. Anotación de los chips de ADN

La anotación es un paso fundamental, ya que los estudios seleccionados corresponden a chips de diferentes modelos y casas comerciales, por lo que es necesario convertir los códigos de las sondas de cada uno de los *microarrays* utilizados a un identificador común que pueda ser utilizado en los análisis posteriores. El identificador común seleccionado fue el nombre *Entrez* recogido para cada uno de los genes en las bases de datos del NCBI (*National Center for Biotechnology Information*, Estados Unidos).

Es habitual que varias sondas de un chip de ADN hibriden con secuencias de un mismo gen, por lo que un mismo gen puede aparecer varias veces en el mismo chip. Puesto que nuestros análisis se realizarán partiendo de la información a nivel de gen, necesitaremos obtener un único valor de expresión de cada gen por muestra en la que aparezca. Para realizar esto se calculó la media de los niveles de expresión de todas las sondas que apuntaban al mismo gen, en cada uno de los pacientes, utilizando la función *avereps* de la librería *limma* [67].

### 3.2.1.3. Análisis exploratorio

Una vez seleccionados y anotados los estudios, se procedió a realizar un análisis exploratorio de los mismos con el objetivo de detectar comportamientos anómalos de las muestras. Asimismo, se comprobó si los diferentes estadios de la enfermedad se agrupaban mostrando similitud de expresión según el tipo de muestra al que pertenecen.

Para evitar la introducción de nuevos sesgos se partió de los datos normalizados de los estudios seleccionados, por lo que el primer paso fue representar los niveles de expresión de los genes de cada una de las muestras en un diagrama de cajas para comprobar que no existiesen anomalías. En caso de encontrar alguna irregularidad en cualquiera de las muestras, se procedió a la descarga de los datos crudos del estudio y a su normalización utilizando la función RMA la librería `oligo` [68], en el caso de los chips de Affymetrix.

Tras obtener todos los datos correctamente normalizados se procedió al análisis exploratorio propiamente dicho. Se realizó un análisis de *clustering* utilizando la distancia euclídea y la distancia de correlación, con el objetivo de ver agrupaciones de los niveles de expresión de las muestras según el género o estadio de la enfermedad de los pacientes. El uso de dos tipos de distancias nos permitió ver las agrupaciones tanto a nivel de distancia “ordinaria” entre las muestras (distancia euclídea) como a nivel de las tendencias de expresión de las muestras (distancia de correlación). Además, se realizó un análisis de componentes principales (PCA, del inglés *Principal Component Analysis*) con el mismo objetivo.

En caso de encontrar algún tipo de sesgo se investigó el origen del mismo para ser tenido en cuenta a la hora de realizar los análisis posteriores.

### 3.2.2 Análisis de expresión diferencial

En estos análisis se realizaron comparaciones entre las diferentes fases de la enfermedad en hombres y mujeres, con el objetivo de ver los genes, funciones o rutas de señalización afectadas de forma diferencial en ambos géneros.

El primer paso para la realización de los análisis de expresión diferencial a nivel de gen fue definir los grupos de los estudios. Puesto que el objetivo del trabajo es identificar diferencias de género en la enfermedad, los grupos surgieron de la combinación del factor "género" y del factor "estadio de la enfermedad", tal y como se puede ver en la Tabla 3.1. Además, se definieron las posibles causas de sesgo encontradas en los análisis exploratorios para ser tenidas en cuenta en los siguientes pasos.

	<b>Control</b>	<b>NAFL</b>	<b>NASH</b>
<b>Hombres</b>	Control.Hombres	NAFL.Hombres	NASH.Hombres
<b>Mujeres</b>	Control.Mujeres	NAFL.Mujeres	NASH.Mujeres

**Tabla 3.1:** Grupos de estudio para las comparaciones.

Con los grupos y los posibles sesgos definidos y mediante funciones del paquete `limma` [67]; se realizó la matriz de contrastes y se ajustó, para cada gen, el modelo lineal con la función `lmFit`. Se calcularon los coeficientes estimados y los errores estándar a partir de la matriz de contrastes y el modelo lineal (`contrasts.fit`) y, a continuación, los estadísticos de la expresión diferencial con `eBayes` (logaritmos de los *odds ratios*, p-valores, etc.). Finalmente se ajustó el p-valor mediante el método de Benjamini & Hochberg (BH) [69].

### 3.2.3 Análisis de enriquecimiento funcional

Una vez obtenidos los datos de expresión diferencial de los genes, podemos realizar el análisis de enriquecimiento funcional mediante GSEA [56, 57].

#### Gene Ontology

El enriquecimiento funcional con los términos GO se realiza asociando los genes de la matriz de expresión a sus términos GO mediante el paquete `org.Hs.eg.db` [70], seleccionando aquellos que pertenecen a la categoría de interés (procesos biológicos, funciones moleculares o componentes celulares) y propagando la anotación (con la librería `mdgsa` [71]), de forma que los genes anotados con un término GO heredan la anotación de los términos ancestros. Por último, filtramos aquellas anotaciones excesivamente específicas o genéricas (bloques con tamaño inferior a 10 o superior a 500).

A partir de los estadísticos calculados previamente, se ordenaron los genes en función de sus p-valores y el signo del estadístico del contraste. Una vez ordenados, se transforma el ranking para que su distribución sea adecuada como variable independiente de un modelo de regresión logística univariante y se realiza un GSEA con los genes ordenados y la anotación previa de los términos GO. Todo este proceso se realiza con diversas funciones del paquete `mdgsa` [71]. Se consideran significativos los términos con un p-valor ajustado [69] inferior a 0.05.

#### KEGG

En el caso del enriquecimiento funcional con las rutas KEGG, únicamente se realiza la anotación de los genes con las rutas KEGG asociadas utilizando el paquete

org.Hs.eg.db [70]. Como las rutas KEGG no siguen una estructura jerárquica no es posible hacer una propagación de las mismas.

La ordenación de los genes y realización del GSEA se llevó a cabo de la misma manera que en el caso de los términos GO.

### 3.2.4 Análisis de rutas de señalización

*Hipathia* [72] es un método para el cálculo de la transducción de señales a lo largo de vías de señalización a partir de datos transcriptómicos. Este método, que pertenece al grupo de los *Mechanistic pathway activity* (MPA), se basa en un algoritmo iterativo que puede calcular la intensidad de la señal que pasa a través de los nodos de una red teniendo en cuenta el nivel de expresión de cada gen y la intensidad de la señal que llega a él. La aproximación utilizada por *Hipathia* proporciona un nuevo enfoque para el análisis funcional que permite calcular la señal que llega a las funciones anotadas en cada ruta.

Las rutas incorporadas en *Hipathia* están extraídas de la base de datos KEGG. Puesto que una misma ruta puede estar asociada a varias funciones a partir de los distintos nodos efectores de la misma, el método divide cada ruta en diferentes subrutas, analizándolas por separado. Cada subruta está definida por un nodo final (o efector) e incluye todos los nodos que llegan a él. Las subrutas se identifican por un identificador interno (ID) y un nombre formado por la ruta a la que pertenecen y su nodo final.

*Hipathia* está accesible mediante una herramienta web (<http://hipathia.babelomics.org/>) y mediante un paquete de R depositado en el repositorio Bioconductor (<https://doi.org/doi:10.18129/B9.bioc.hipathia>).

El diagrama del flujo de trabajo y de las funciones de *Hipathia* está representado en la Figura 3.3 y se detalla a continuación:

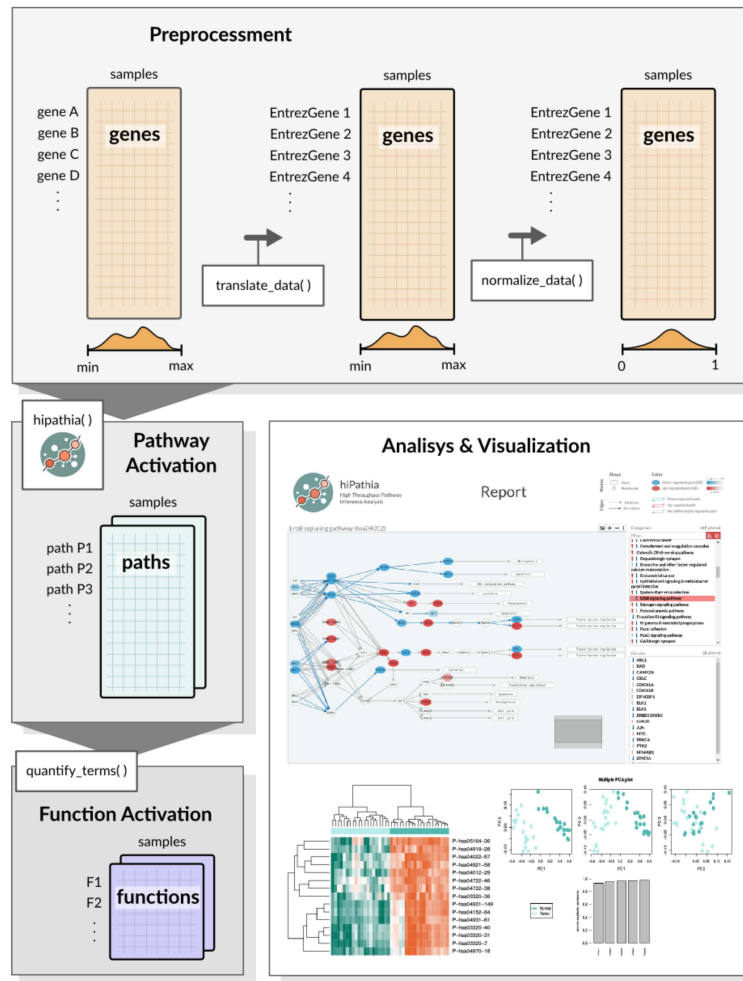


Figura 3.3: Diagrama de flujo de trabajo con el paquete *Hipathia*.

Imagen extraída de la viñeta del paquete.

1. **Traducción de los nombres de los genes** a su correspondiente nombre *Entrez*. En el caso de este trabajo, este paso no es imprescindible por estar previamente anotadas las matrices de expresión con los nombres *Entrez*.
2. **Normalización de los valores de expresión** a valores entre 0 y 1. Esta función permite incorporar, en caso de ser necesario, diferentes parámetros para la normalización.

3. **Cálculo del nivel de activación de cada subruta** en una ruta para cada una de las muestras del experimento. Puesto que el procesado de los grafos de las rutas es un proceso computacionalmente costoso, la versión utilizada de *Hipathia* (v1.3.1) incluye un módulo con 145 rutas KEGG listas para ser analizadas (ver Tabla II.2).
4. **Extracción de los valores de activación para cada ruta** y cuantificación de la activación para los términos GO y UniProt [73].

Una vez obtenidos todos los valores de activación, podemos proceder a las comparaciones entre los diferentes casos. El paquete *hipathia* tiene implementada la función `do_wilcoxon` que permite la comparación entre dos grupos utilizando el test de Wilcoxon, sin embargo, nosotros queremos realizar comparaciones entre varios grupos (ver Tabla 3.1), por lo que utilizaremos `limma` (al igual que en el apartado 3.2.2) para nuestras comparaciones a nivel de ruta (y subruta), término GO o término UniProt. Las anotaciones de estos términos, incluidas por defecto en *Hipathia*, corresponden a los términos GO pertenecientes a procesos biológicos (versión del 16-02-2016) y a los términos UniProt bajo la categoría "función molecular" (versión del 21-09-2015).

### 3.3 Metaanálisis

---

La naturaleza de los datos obtenidos en los dos tipos de análisis individuales es diferente: en el caso del enriquecimiento funcional con GSEA los resultados obtenidos son datos categóricos, mientras que en el caso del análisis de rutas de señalización con *Hipathia* obtenemos datos cuantitativos. Por este motivo, los métodos para la realización del metaanálisis serán diferentes para cada uno de métodos individuales aplicados.

### 3.3.1 Enriquecimiento funcional

El metaanálisis se realiza a partir de los datos de los análisis individuales del GSEA con los términos GO o las rutas KEGG. Por cada comparación de cada estudio de enriquecimiento funcional se realizará un metaanálisis, para lo que necesitamos crear dos matrices: una que contenga una medida del efecto de la comparación (el logaritmo de los *odds ratios*) y otra que contenga una medida de la variabilidad de la medida del efecto (el error estándar del logaritmo de los *odds ratios*). En cada una de estas matrices las filas representan las funciones o las rutas y las columnas hacen referencia a cada uno de los estudios. Los términos o rutas que únicamente aparecen en uno de los estudios primarios no se incluyeron en las matrices de los metaanálisis.

Previo al metaanálisis, es conveniente la exploración de estas matrices para un mayor conocimiento de los datos. Esto incluye la descripción de la magnitud del efecto y su variabilidad, la detección de valores atípicos y una cuantificación de los valores perdidos procedentes de la combinación de los distintos estudios [48].

En un metaanálisis, primero se debe evaluar la heterogeneidad de los estudios y, en función de esa información, decidir el modelo de estimación de la variabilidad del efecto. En nuestro caso, puesto que realizamos miles de metaanálisis de forma simultánea (uno por cada función), valoramos *a posteriori* la idoneidad del procedimiento seleccionado [48]. En la obtención de la medida resumen de cada función se ponderan los resultados de los estudios por la inversa de su varianza, de forma que aquellos con una menor variabilidad tendrán un mayor peso sobre el efecto conjunto. En caso de encontrar heterogeneidad, esta puede ser tenida en cuenta utilizando un modelo de efectos aleatorios o ser ignorada utilizando un modelo de efectos fijos [48, 60, 63]. El modelo de efectos fijos no considera la variabilidad entre los estudios incluidos, mientras que el modelo de efectos aleatorios sí lo hace.

Para la realización de los metaanálisis funcionales se utilizó el paquete *metafor* [74] de R. Todos los metaanálisis se realizaron con el modelos de efectos fijos (FE) y los siguientes modelos de efectos aleatorios: DerSimonian & Laird (DL) [75], Hedges *et al.* (HE) [76] y Schmidt & Hunter (HS) [77]; para su posterior evaluación.

### 3.3.1.1. Evaluación de la heterogeneidad

Después de la aplicación del proceso de metaanálisis, se obtienen un conjunto de estimadores de la medida del efecto combinado para cada una de las funciones/rutas estudiadas y una serie de indicadores de la heterogeneidad del metaanálisis. A continuación se comenta brevemente cada uno de ellos [48, 58]:

- *QE* y *QEp*. Estadístico de contraste y valor *p* del test Q de DerSimonian & Laird [75]. La hipótesis alternativa del test apunta a la presencia de heterogeneidad entre los estudios.
- *LOR* (*logaritmo de los odds ratio*). Es la estimación del efecto conjunto de todos los estudios para la función. La magnitud cuantifica el nivel de sobrerrepresentación y el signo indica la clase que contiene un mayor presencia de genes con un nivel elevado de expresión (positivo: mayor nivel el la primera con respecto a la segunda, negativo mayor nivel en la segunda con respecto a la primera).
- *Intervalo de confianza del LOR* (al 95 %) construido utilizando la variabilidad del modelo seleccionado. La significatividad del *LOR* se confirmaría por la ausencia del 0 en dicho intervalo.
- *P* y *FDR*. El valor de *p* (*P*) informa del nivel de significación de un logaritmo de los *odds ratio* combinado (para cada función por el conjunto de los estudios) nulo. Este valor de *p* no contempla el escenario de multiplicidad de que dis-

ponemos, por lo que es necesario ajustar el indicador por algún método de corrección. En este trabajo utilizaremos el método de la tasa de falsos descubrimientos (FDR) de Benjamini & Hochberg [69].

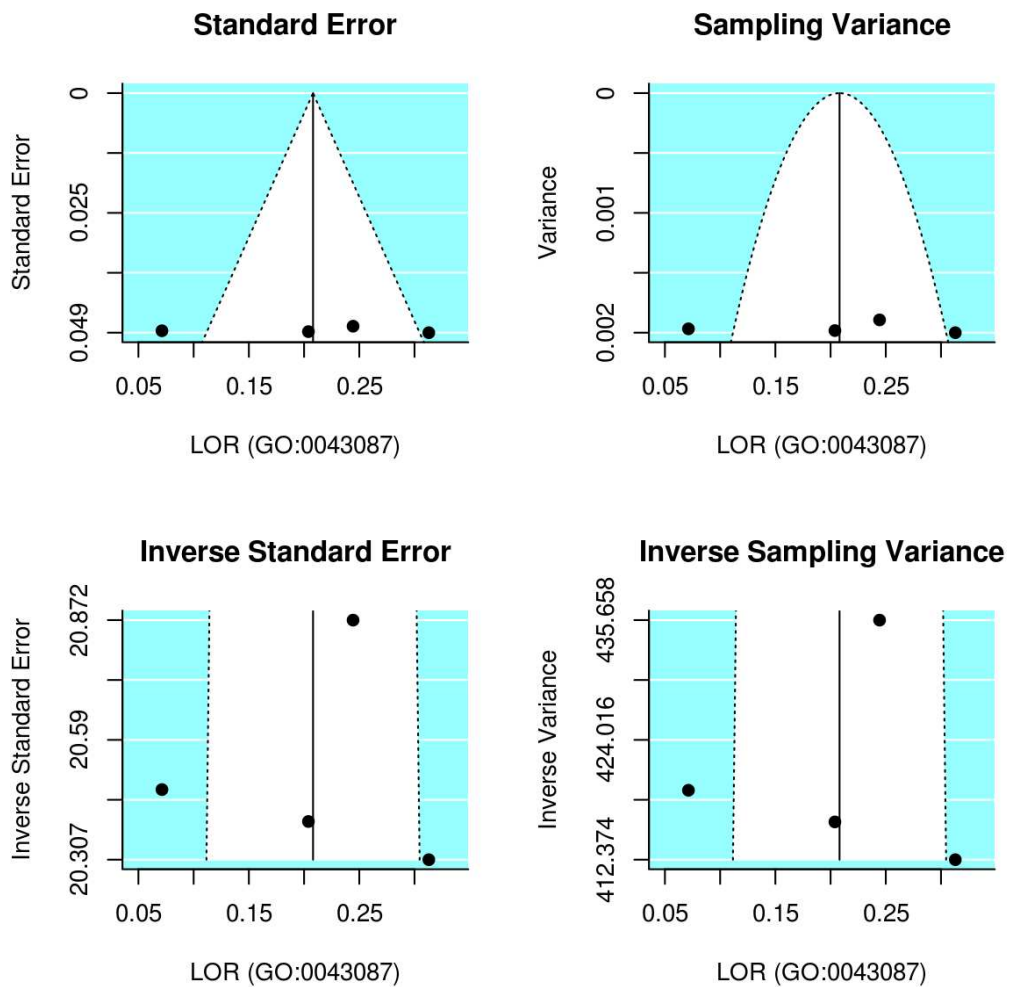
- $\tau^2$ . Estimación de la heterogeneidad entre estudios. En el caso del modelo de efectos fijos será 0.
- $I^2$ . Estimación, en forma porcentual, de la variabilidad entre estudios con respecto a la variabilidad global [78].
- $H^2$ . Cociente entre la variabilidad global y la variabilidad en el muestreo [78].

El test Q comentado anteriormente y propuesto por DerSimonian & Laird [75] presenta una baja potencia a la hora de trabajar con números pequeños de estudios primarios, por lo que en análisis de la heterogeneidad se complementa con métodos resumidos gráficamente [48, 58]:

- Los GRÁFICOS DE EMBUDO representan la variabilidad de los distintos estudios y la presencia de sesgos. Estos gráficos muestran la magnitud del efecto medido (eje X, en nuestro caso el logaritmo de los *odds ratio*) frente a una medida de la precisión (eje Y, en nuestro caso el error estándar y la varianza) [79]. Un ejemplo de estos gráficos se puede ver en la Figura 3.4.
- REPRESENTACIONES GRÁFICAS DE ESTIMADORES PREVIAMENTE DESCRITOS de forma que se puedan evaluar su efecto de manera global, por ejemplo, mediante diagramas de cajas.

### 3.3.1.2. Representación gráfica de los resultados

Además de los resultados crudos que nos aporta el metaanálisis, también se generan representaciones gráficas de los mismos. Uno de los más comunes, por su importan-



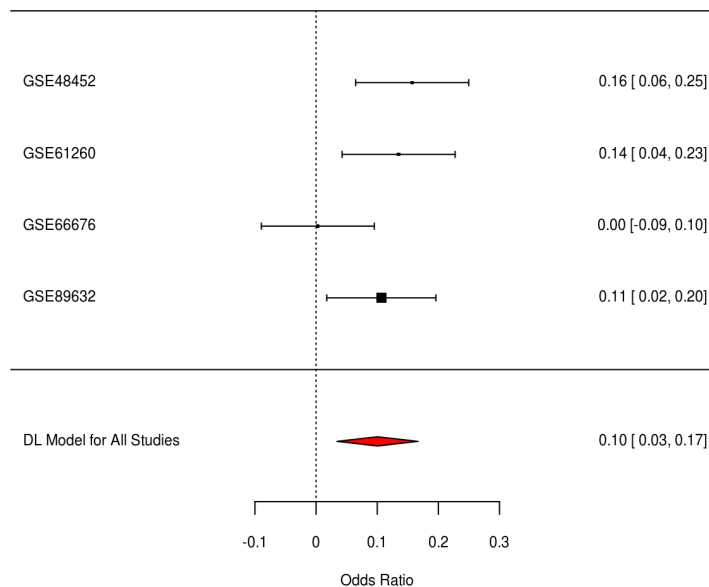
**Figura 3.4:** Ejemplo de gráficos de embudo para la representación de la heterogeneidad de un metaanálisis de 5 estudios para un término GO. Se puede ver como en todos los casos, 2 de los estudios se separan de la región de confianza (en color blanco).

cia y utilidad a la hora de evaluar los resultados globales y el peso en el resultado de cada estudio individual, son los GRÁFICOS DE BOSQUE. Vamos a explicar este tipo de representación con ayuda de la Figura 3.5:

- En la parte izquierda del gráfico se encuentran los estudios incluidos en el metaanálisis.

- A la derecha se incluye la medida resumen de cada estudio y su intervalo de confianza al 95 %.
- En el centro se observa un segmento y un cuadrado negro, que representan el intervalo de confianza y el valor de la medida del efecto de cada estudio. El tamaño del cuadro negro es proporcional a la precisión de las estimaciones.
- En la parte inferior izquierda se muestra el modelo de metaanálisis utilizado, a la derecha el resultado del efecto conjunto (*LOR*) y su intervalo de confianza (95 %). En la parte inferior central se representa un rombo centrado en el valor del *LOR* y cuya anchura corresponde al intervalo de confianza del mismo.

**GO:0044270 (cellular nitrogen compound catabolic process)**



**Figura 3.5:** Ejemplo de un gráfico de bosque para un término GO. Los *odds ratio* del eje X se encuentran en escala logarítmica y corresponden al *LOR*.

### 3.3.2 Rutas de señalización

Al igual que en el caso del metaanálisis de los datos del enriquecimiento funcional, a la hora de realizar el metaanálisis de *Hipathia* necesitamos dos matrices de datos. En ambas matrices las columnas corresponderán a los estudios individuales y las filas a los elementos biológicos (rutas, términos GO o términos UniProt). La primera matriz contendrá los p-valores crudos de la comparación realizada y la segunda los valores del estadístico  $t$  de *limma* (el signo de  $t$  nos indica el sentido del resultados de comparación). Una vez obtenidas estas matrices los pasos a seguir serán:

1. Realizar un filtrado de los elementos cuyo estadístico  $t$  mantenga el mismo signo a lo largo de todos los estudios.
2. Calcular el p-valor combinado para cada elemento biológico mediante la función `sump` del paquete `metap` [80].
3. Realizar una corrección de los p-valores combinados mediante el método de Benjamini & Hochberg (FDR) [69].

Con estos p-valores corregidos podremos evaluar el nivel de significación de cada elemento, de forma conjunta, en nuestra comparación.

---

---

## CAPÍTULO 4

---

# Resultados

En el presente capítulo se muestran los resultados obtenidos en este trabajo, siguiendo el orden establecido en el capítulo anterior. Para mejor lectura, en el capítulo se muestran las figuras de mayor relevancia, pudiéndose encontrar el resto en el Anexo I.

Los resultados aquí mostrados buscan responder los objetivos planteados en el Capítulo 2. Con este fin, y una vez definidos los grupos de pacientes por el género y el estadio de enfermedad de los pacientes (ver Tabla 3.1), nos centraremos en una de las comparaciones de mayor relevancia: (NASH.Mujeres - NAFL.Mujeres) - (NASH.Hombres - NAFL.Hombres). Esta comparación evalúa el paso de NAFL a NASH en ambos géneros y posteriormente compara las diferencias de este salto entre los dos géneros, permitiéndonos ver las bases moleculares diferencialmente alteradas entre los dos géneros en esta progresión de la enfermedad.

## 4.1 Revisión sistemática y selección de los estudios

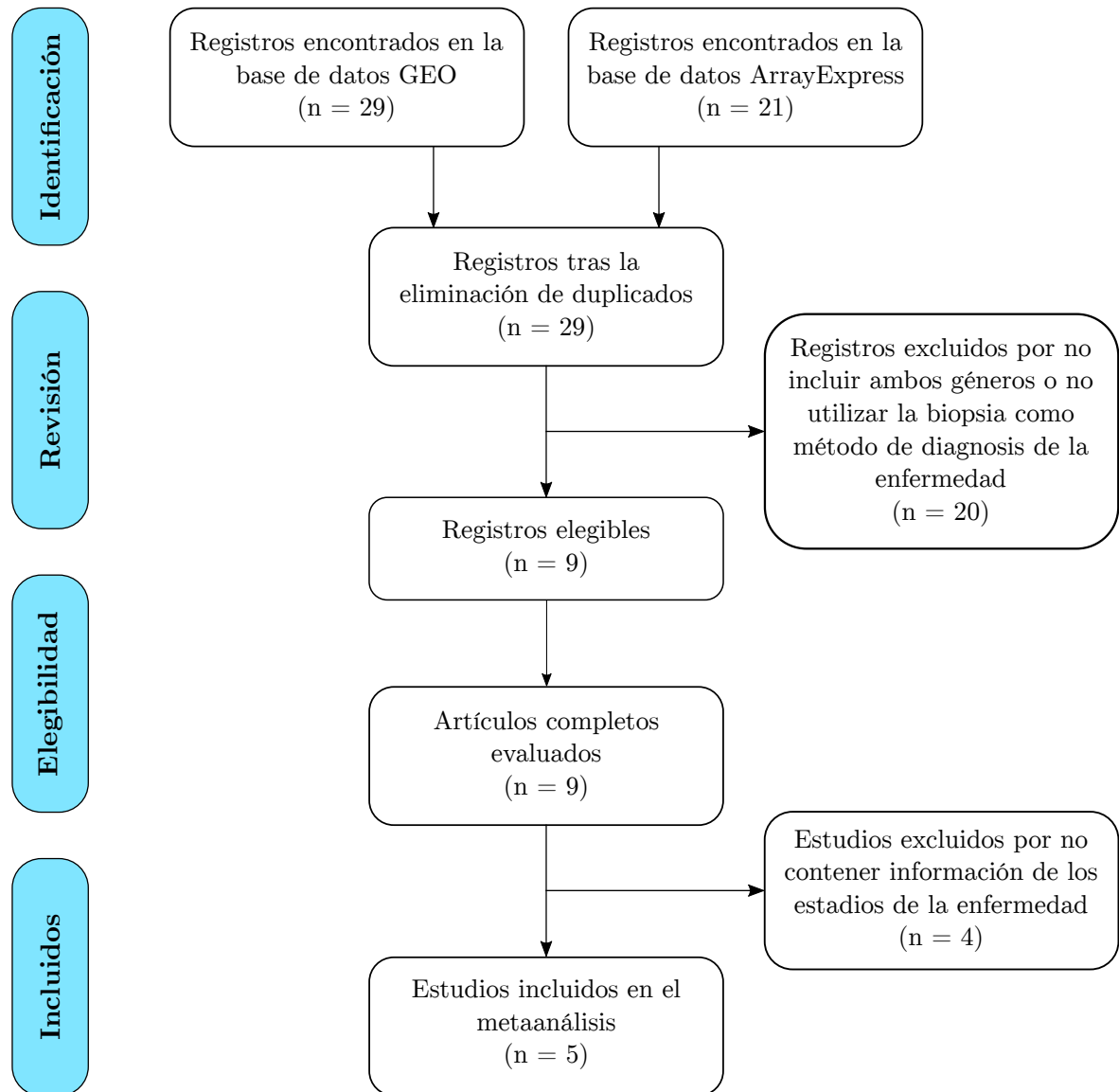
La revisión sistemática en bases de datos públicas encontró 50 estudios candidatos, de los cuales 5 fueron seleccionados para el posterior metaanálisis (Tabla 4.1). Las búsquedas se realizaron en GEO y ArrayExpress siguiendo las directrices de la *declaración PRISMA*, tal y como se muestra en la Figura 4.1.

Estudio (GEO Series)	Plataforma (y código GEO)	Artículo
GSE48452	Affymetrix Human Gene 1.1 ST Array (GPL11532)	[81]
GSE61260	Affymetrix Human Gene 1.1 ST Array (GPL11532)	[82]
GSE66676	Affymetrix Human Gene 1.0 ST Array (GPL6244)	[83]
GSE83452	Affymetrix Human Gene 2.0 ST Array (GPL16686)	[2]
GSE89632	Illumina HumanHT-12 WG-DASL V4.0 R2 (GPL14951)	[84]

**Tabla 4.1:** Estudios seleccionados tras la revisión sistemática junto con la plataforma con la que se realizaron y la publicación asociada.

En aquellos estudios en los que la información del género de los individuos no se encontraba disponible se intentó obtener la misma poniéndose en contacto con los responsables de los estudios para incorporarla a los mismos. La biopsia hepática fue utilizada como único método válido de diagnóstico del NAFLD, debido a las limitaciones de los métodos no invasivos descritas en el Apartado 1.1.2.

Otra de las limitaciones a la hora de la selección, fue la existencia de información sobre el estadio de la enfermedad en los estudios, ya que algunos de los registros de las bases de datos hacían referencia a ensayos en los que se trataban a pacientes de la enfermedad con algún medicamento, sin indicar ni tener en cuenta la etapa de la enfermedad de los pacientes.



**Figura 4.1:** Diagrama de flujo de la revisión sistemática llevada a cabo en este trabajo, según la *declaración PRISMA*.

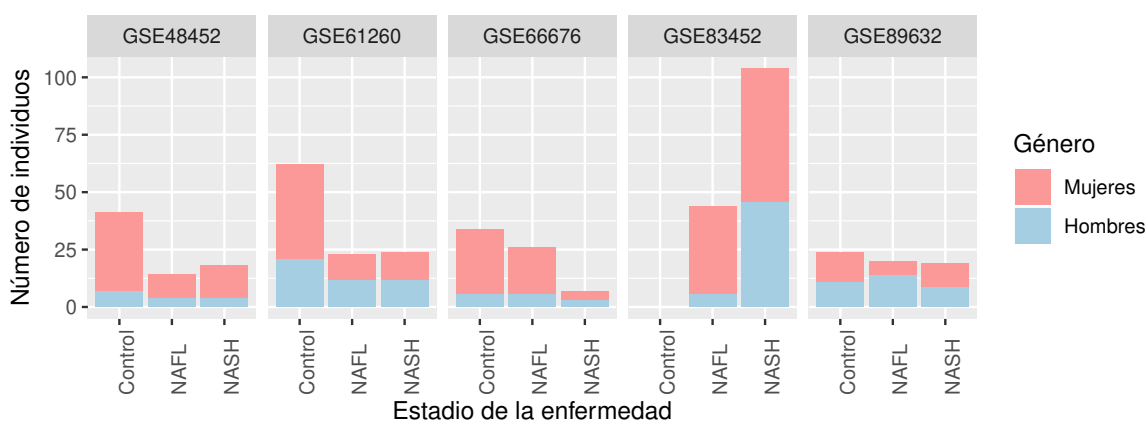
## 4.2 Análisis individual de los estudios

### 4.2.1 Procesamiento de los datos y análisis exploratorio

Una vez descargados los datos de cada estudio y armonizada la nomenclatura (ver Apartado 3.2.1.1), se comprobó la normalización las matrices de expresión de los arrays mediante un diagrama de cajas. En el caso de los estudios GSE61260 y GSE66676 se descargaron los datos crudos y se normalizaron con RMA.

Con los datos ya normalizados, se describió la cantidad de individuos por género y estadio de la enfermedad en cada estudio, tal y como se puede ver en la Figura 4.2 y en las Tablas 4.2 a 4.6. Se puede observar, en general, una mayor presencia de mujeres en los estudios.

El ensayo GSE83452 no posee pacientes control, por lo que únicamente podrá ser utilizado cuando se comparen individuos con los estadios NAFL y NASH.



**Figura 4.2:** Diagrama de barras indicando el número de pacientes de cada género en cada estudio, separados por el estadio de la enfermedad que presentan.

A continuación se muestran los resultados específicos del análisis exploratorio de cada estudio.

**GSE48452**

Este estudio [81] analizó muestras de hígado de pacientes con obesidad mórbida, que abarcaban todo el espectro del NAFLD, y pacientes control. Tal y como se explica en el Apartado 3.2.1.1, los controles con peso normal y obesos se agrupan como "Control".

La Tabla 4.2 muestra los individuos participantes en el estudio por género y estadio de la enfermedad.

	Control	NAFL	NASH	Total
Hombres	7	4	4	15
Mujeres	34	10	14	58
Total	41	14	18	73

**Tabla 4.2:** Individuos incluidos en el estudio GSE48452 separados por género y estadio de la enfermedad.

Los análisis de *clustering* (Figura I.1) y de componentes principales (Figura I.2) no mostraron ningún agrupamiento o sesgo a tener en cuenta.

**GSE61260**

En este estudio [82] se realiza un análisis de expresión en tejido hepático de pacientes alemanes (tanto sanos como con la enfermedad) con el objetivo de correlacionar los datos de expresión génica con el índice de masa corporal de los pacientes.

Pacientes del estudio con enfermedades diferentes al NAFLD, como colangitis esclerosante primaria o cirrosis biliar primaria, se reclasificaron como "NA" y no fueron tenidos en cuenta en posteriores análisis (ver Tabla 4.3).

	Control	NAFL	NASH	Total
Hombres	21	12	12	45
Mujeres	41	11	12	64
Total	62	23	24	109

**Tabla 4.3:** Individuos incluidos en el estudio GSE61260 separados por género y estadio de la enfermedad.

Las representaciones del PCA y del *clustering* de este experimento no muestran sesgos o agrupamientos (ver Figuras I.3 y I.4).

### GSE66676

El objetivo de este estudio [83] fue el análisis de datos *ómicos* de pacientes adolescentes con diferentes estadios de la enfermedad para estudiar las funciones y rutas desreguladas por el NAFLD. Los pacientes del estudio se muestran en la Tabla 4.4.

	Control	NAFL	NASH	Total
Hombres	6	6	3	15
Mujeres	28	20	4	52
Total	34	26	7	67

**Tabla 4.4:** Individuos incluidos en el estudio GSE66676 separados por género y estadio de la enfermedad.

El *clustering* de los datos no muestra una agrupación de los mismos (Figura I.5), mientras que el PCA (Figura I.6) muestra dos posibles grupos cuya causa no ha podido ser identificada y, por lo tanto, no ha sido incorporada a los posteriores análisis.

**GSE83452**

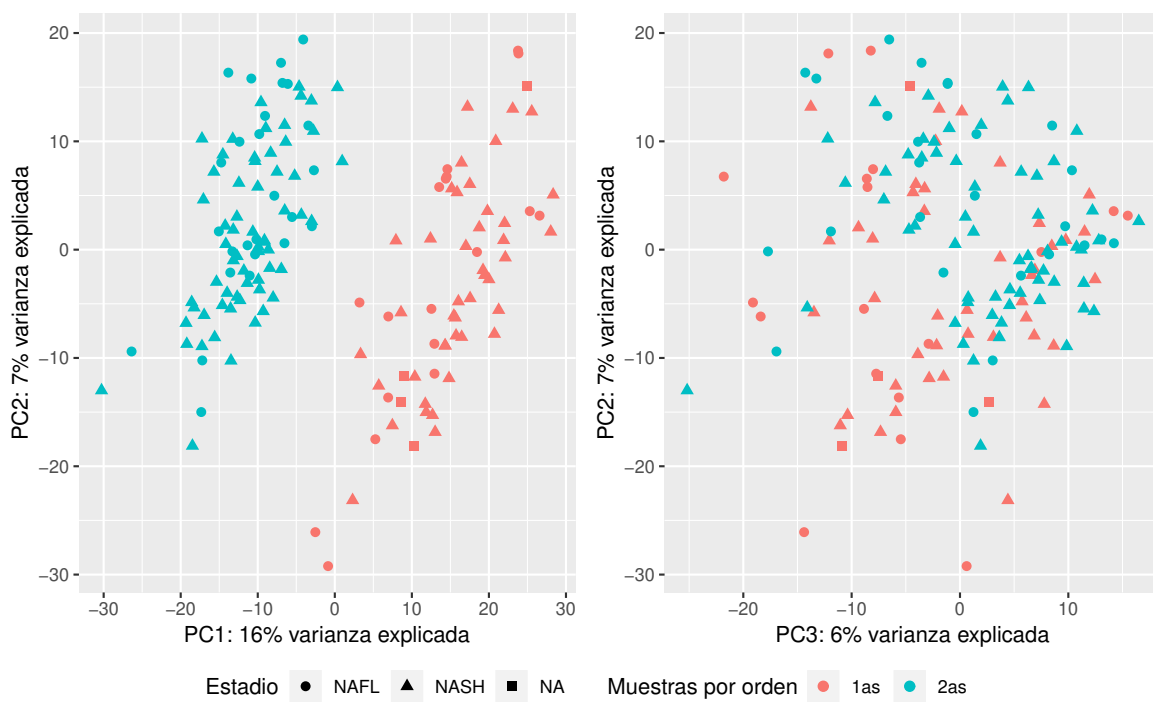
El fin de este estudio [2] fue el análisis de datos transcriptómicos de muestras hepáticas de pacientes con NAFLD. Algunos de estos pacientes fueron sometidos a cirugía bariátrica para perder peso y analizados de nuevo un año después de la intervención.

Como consecuencia de su diseño experimental, este estudio no posee controles sanos. Además, para nuestros análisis únicamente son útiles los datos de los pacientes obtenidos antes de la cirugía de reducción de peso, ya que los análisis tras un tratamiento no forman parte de nuestro estudio. Las muestras de este estudio utilizadas en el presente trabajo se detallan en la Tabla 4.5 y corresponden a aquellas que fueron tomadas de forma previa al tratamiento.

	<b>Control</b>	<b>NAFL</b>	<b>NASH</b>	<b>Total</b>
<b>Hombres</b>	0	38	58	96
<b>Mujeres</b>	0	6	46	52
<b>Total</b>	0	44	104	<b>148</b>

**Tabla 4.5:** Individuos incluidos en el estudio GSE83452 separados por género y estadio de la enfermedad.

En análisis del PCA del estudio GSE83452 (Figura I.8) se observa un agrupamiento que no está provocado por el género o los estadios de las muestras. Tras evaluar varios factores como la edad o el índice de masa corporal, que podrían estar provocando esta separación, se detectó que estaba motivado por el orden de los identificadores de las muestras: las 64 primeras en un grupo y las 88 siguientes en el otro grupo (Figura 4.3). Este efecto será tenido en cuenta para los posteriores análisis.



**Figura 4.3:** Análisis de componentes principales de las muestras del estudio GSE83452. Los colores representan las muestras ordenadas (*1as*, las 64 primeras y *2as*, las 88 siguientes) y las formas geométricas los géneros.

**GSE89632**

Este trabajo [84] analizó la expresión génica de pacientes sanos y con diferentes estadios de la enfermedad con el objetivo de relacionar los datos de expresión con la composición de los ácidos grasos del hígado. La Tabla 4.6 muestra los individuos de este estudio.

	<b>Control</b>	<b>NAFL</b>	<b>NASH</b>	<b>Total</b>
<b>Hombres</b>	11	14	9	34
<b>Mujeres</b>	13	6	10	29
<b>Total</b>	24	20	19	63

**Tabla 4.6:** Individuos incluidos en el estudio GSE89632 separados por género y estadio de la enfermedad.

Los análisis exploratorios no mostraron agrupamiento de las muestras (Figuras I.9 y I.10).

### 4.2.2 Análisis de expresión diferencial

El análisis de expresión diferencial de cada uno de los estudios individuales proporciona la detección de genes más expresados en cada una de los grupos de interés. Estos resultados serán utilizados en las siguientes fases de la estrategia de análisis.

La Tabla 4.7 muestra el número de genes expresados de forma diferencial en cada estudio ( $p$ -valor ajustado (BH) [69]  $< 0.05$ ) para diferentes comparaciones. Aunque la comparación de interés es (NASH.Mujeres - NAFL.Mujeres) - (NASH.Hombres - NAFL.Hombres) (número 10 en la Tabla 4.7), en la tabla se muestran, con fin exploratorio, los resultados de múltiples comparaciones posibles con los datos disponibles.

Comparación	GSE48452		GSE61260		GSE66676		GSE83482		GSE89632	
	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up
1 Control.M - Control.H	949	765	410	792	14	7	-	-	68	22
2 NAFL.M - NAFL.H	15	1	13	10	14	7	579	1297	31	9
3 NASH.M - NASH.H	13	0	13	14	12	3	32	51	43	11
4 NASH.M - NAFL.M	0	0	0	2	0	3	0	0	1	8
5 NASH.H - NAFL.H	0	0	0	0	0	1	290	584	1	1
6 NASH.M - Control.M	198	332	14	40	294	550	-	-	1150	1437
7 NASH.H - Control.H	7	6	283	539	0	0	-	-	1835	2671
8 NAFL.M - Control.M	0	0	24	6	0	0	-	-	1466	2342
9 NAFL.H - Control.H	31	34	227	292	0	0	-	-	1163	2007
<b>10 (NASH.M - NAFL.M) - (NASH.H - NAFL.H)</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>365</b>	<b>161</b>	<b>0</b>	<b>0</b>
11 (NASH.M - Control.M) - (NASH.H - Control.H)	0	0	0	0	0	0	-	-	0	0
12 (NAFL.M - Control.M) - (NAFL.H - Control.H)	0	1	1	2	0	0	-	-	0	0
13 (NASH.M - NASH.H) - (NAFL.M - NAFL.H)	0	0	0	0	0	0	365	161	0	0

**Tabla 4.7:** Número de genes diferencialmente expresados en cada estudio individual. La letra M corresponde a Mujeres y la letra H a Hombres. Los términos *Down* y *Up* hacen referencia al término de la comparación en el que los genes están sobreexpresados (*Up* al primero y *Down* segundo). Con un guión (-) se indican las comparaciones no realizadas por falta de datos.

### 4.2.3 Análisis de enriquecimiento funcional

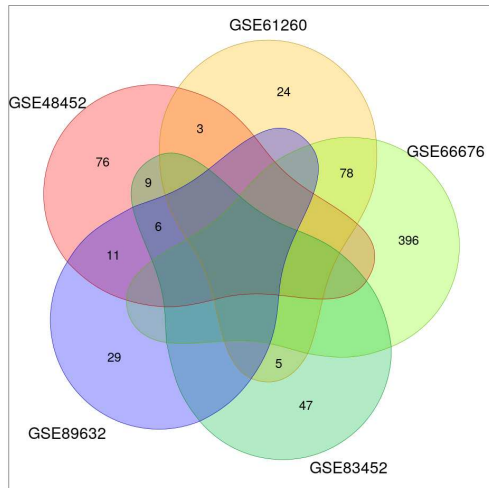
Los análisis del GSEA se realizaron tanto para términos GO (Procesos biológicos, funciones moleculares, componentes celulares) como para ruta KEGG, comparando los grupos (NASH.Mujeres - NAFL.Mujeres) - (NASH.Hombres - NAFL.Hombres).

Los resultados se resumen en la Tabla 4.8. En ella se muestran las funciones significativas (p-valor ajustado según Benjamini & Yekutieli [85]  $< 0.05$ ) agrupadas en *Down* ( $LOR < 0$ ) si están sobrerrepresentadas en hombres o *Up* ( $LOR > 0$ ) si lo están en mujeres.

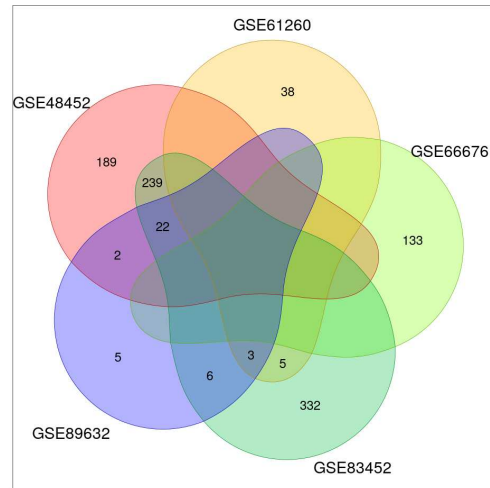
Estudio	Términos GO (bp) significativos			Términos GO (mf) significativos			Términos GO (cc) significativos			Rutas KEGG significativas		
	Down	Up	Total	Down	Up	Total	Down	Up	Total	Down	Up	Total
GSE48452	452	105	557	114	56	170	141	16	157	21	5	26
GSE61260	46	110	156	17	13	30	3	23	16	6	3	9
GSE66676	113	474	607	65	102	167	23	179	202	2	30	32
GSE83452	607	67	674	142	18	160	172	5	177	26	1	27
GSE89632	38	46	84	15	35	50	27	36	63	9	11	20

**Tabla 4.8:** Funciones (términos GO o rutas KEGG) significativos tras el análisis GSEA. Los términos *Down* y *Up* hacen referencia al término de la comparación en el que los genes están sobreexpresados (*Up* al primero (mujeres) y *Down* segundo (hombres)).

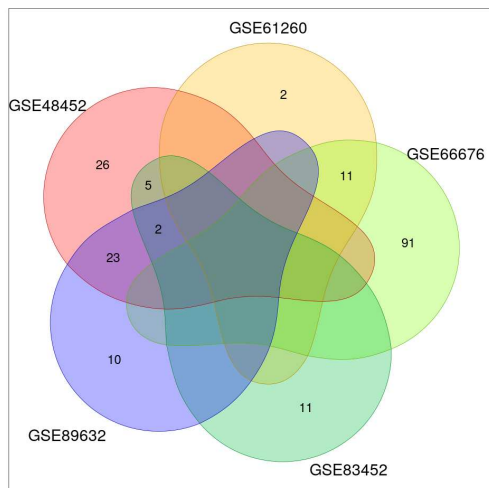
Para cada uno de los grupos funcionales (procesos biológicos, funciones moleculares, componentes celulares y rutas KEGG) se realizó un diagrama de Venn [86] mostrando las funciones significativamente sobreexpresadas en mujeres (*Up*) y en hombres (*Down*), para ver aquellas funciones que estaban compartidas por varios de los estudios (Figura 4.4). Tal y como se puede observar en la figura, ninguno de los términos o funciones era significativo (en el mismo sentido) en más de 3 ensayos.



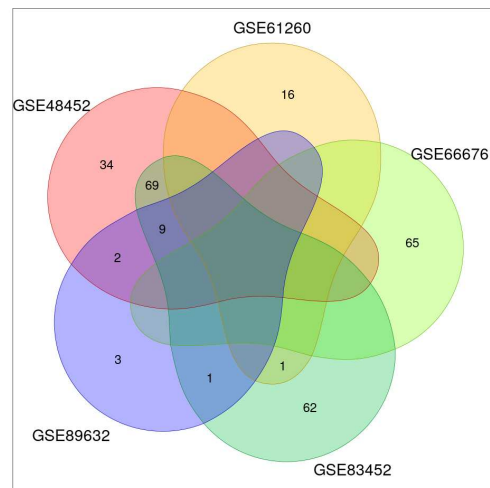
**(a)** Términos GO de procesos biológicos (bp) sobreexpresados en mujeres.



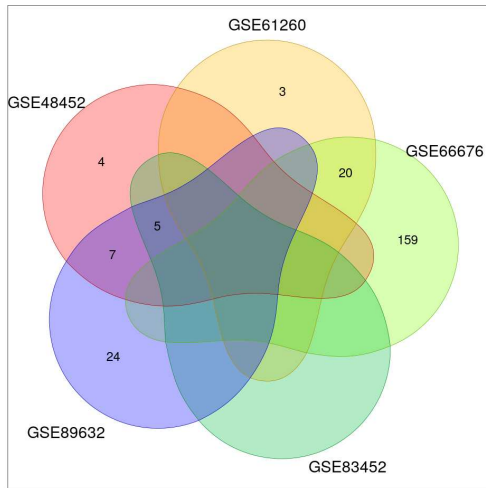
**(b)** Términos GO de procesos biológicos (bp) sobreexpresados en hombres.



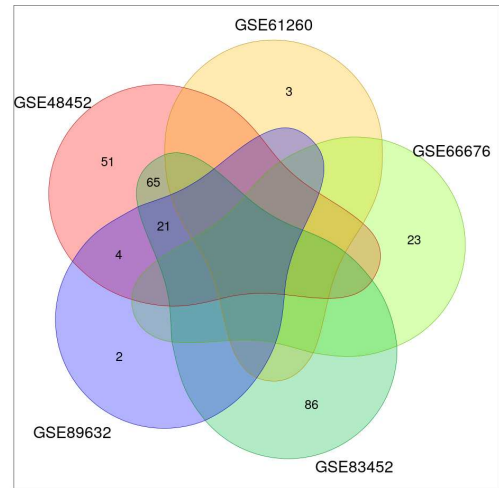
**(c)** Términos GO de funciones moleculares (mf) sobreexpresadas en mujeres.



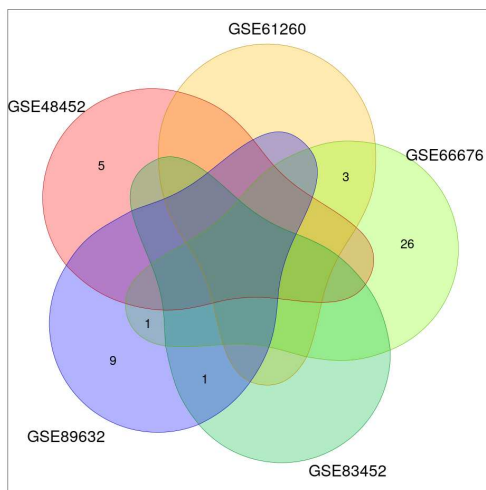
**(d)** Términos GO de funciones moleculares (mf) sobreexpresadas en hombres.



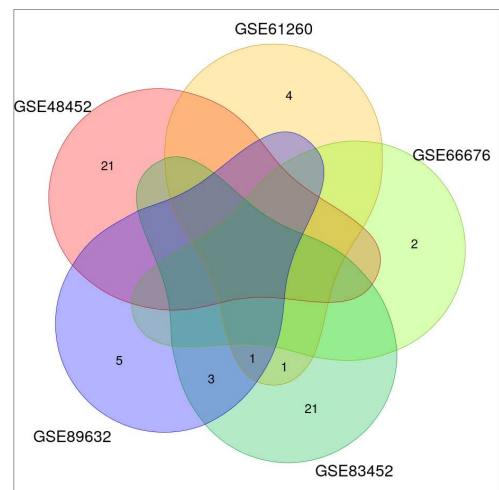
(e) Términos GO de componentes celulares (cc) sobreexpresados en mujeres.



(f) Términos GO de componentes celulares (cc) sobreexpresados en hombres.



(g) Rutas KEGG sobreexpresadas en mujeres.



(h) Rutas KEGG sobreexpresadas en hombres.

Figura 4.4: Diagramas de Venn de los resultados individuales del GSEA.

#### 4.2.4 Análisis de rutas de señalización

*Hipathia* permite el cálculo de las señales de activación de rutas de señalización. A partir de estas señales de activación es posible cuantificar la actividad de las funciones utilizando una anotación que relaciona las rutas con funciones. La versión de *Hipathia* utilizada en este trabajo implementa los términos GO y UniProt.

Al igual que en el caso de los GSEA, todos los análisis corresponden a la comparación (NASH.Mujeres - NAFL.Mujeres) - (NASH.Hombres - NAFL.Hombres). En la Tabla 4.9 se muestran los resultados de los análisis utilizando *Hipathia* (p-valor ajustado (BH) [69] < 0.05).

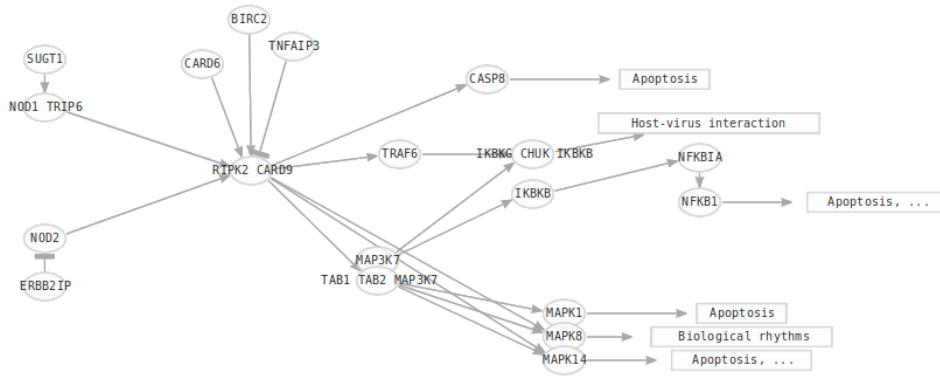
Estudio	Subrutas KEGG			Términos GO			Términos UniProt		
	Down	Up	Total	Down	Up	Total	Down	Up	Total
GSE48452	0	0	0	0	0	0	0	0	0
GSE61260	0	0	0	0	0	0	0	0	0
GSE66676	0	0	0	0	0	0	0	0	0
GSE83452	10	1	11	57	2	59	0	0	0
GSE89632	0	0	0	0	0	0	0	0	0

**Tabla 4.9:** Términos y subrutas significativas para los estudios individuales según *Hipathia*. Los términos *Down* y *Up* hacen referencia al término de la comparación en el que los genes están sobreexpresados (*Up* al primero (mujeres) y *Down* segundo (hombres)).

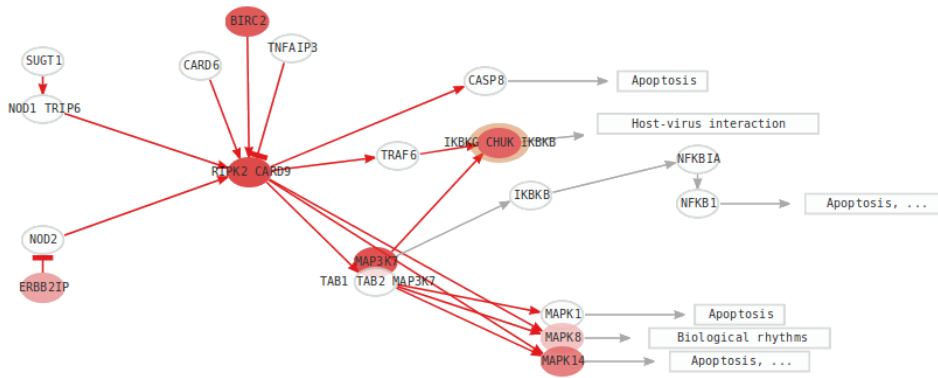
En la Tabla 4.10 se muestran las subrutas significativas, la mayor parte de las cuales pertenecen a la ruta *NOD-like receptor signaling pathway* (hsa04621). Este resultado concuerda con otros trabajos publicados que muestran diferencias significativas en esta ruta entre pacientes con NAFL y pacientes con NASH [87, 88].

ID	Nombre	p-valor	FDR	Sentido
P-hsa04621-28	<i>NOD-like receptor signaling pathway: CASP8</i>	$4.570125^{-7}$	0.00053	<i>Down</i>
P-hsa04621-44	<i>NOD-like receptor signaling pathway: MAPK14</i>	$5.775443^{-7}$	0.00053	<i>Down</i>
P-hsa04621-42	<i>NOD-like receptor signaling pathway: MAPK8</i>	$2.109054^{-6}$	0.00131	<i>Down</i>
P-hsa04621-43	<i>NOD-like receptor signaling pathway: MAPK1</i>	$1.993235^{-5}$	0.00930	<i>Down</i>
P-hsa04621-478	<i>NOD-like receptor signaling pathway: IKBKG CHUK IKBKB</i>	$9.824984^{-5}$	0.02899	<i>Down</i>
P-hsa05168-138	<i>Herpes simplex infection: CHUK</i>	$1.128176^{-4}$	0.02899	<i>Down</i>
P-hsa04014-42	<i>Ras signaling pathway: PLA2G4B</i>	$1.241647^{-4}$	0.02899	<i>Down</i>
P-hsa04014-43	<i>Ras signaling pathway: ELK1</i>	$1.241647^{-4}$	0.02899	<i>Down</i>
P-hsa04910-17	<i>Insulin signaling pathway: D-Glucose</i>	$1.665169^{-4}$	0.03456	<i>Down</i>
P-hsa05014-51	<i>Amyotrophic lateral sclerosis (ALS): DERL1</i>	$2.745322^{-4}$	0.04792	<i>Down</i>
P-hsa05231-21	<i>Choline metabolism in cancer: PCYT1A</i>	$2.822176^{-4}$	0.04792	<i>Up</i>

**Tabla 4.10:** Subrutas significativas según *Hipathia* en el estudio GSE83452. Los términos *Down* y *Up* hacen referencia al término de la comparación en el que las rutas están significativamente activadas de forma diferencial (*Up* al primero (mujeres) y *Down* al segundo (hombres)).



(a) Mujeres.



(b) Hombres.

**Figura 4.5:** Ruta hsa04621 (*NOD-like receptor signaling pathway*) entre pacientes con NAFL y pacientes con NASH en ambos géneros. El color rojo indica activación de los nodos o las rutas.

En la Figura 4.5 se muestra ruta la hsa04621 (*NOD-like receptor signaling pathway*) en mujeres (a) y en hombres (b). El color hace referencia a la expresión diferencial entre pacientes con NAFL y NASH dentro de cada género. Se puede ver cómo existe una activación de ciertos genes y rutas en hombres (b) que no se da en mujeres (a).

## 4.3 Metaanálisis

### 4.3.1 Enriquecimiento funcional

Siguiendo con la estructura de los análisis individuales, el metaanálisis se realizó estudiando por separado los términos GO de las diferentes categorías (procesos biológicos, funciones moleculares y componentes celulares) y las rutas KEGG. En los siguientes apartados se muestran cada uno de los resultados mediante tablas que indican el número de funciones significativas según los distintos métodos de metaanálisis utilizados (Sig. *Down* o Sig. *Up*) y aquellas que, además, presentan una magnitud del  $LOR > 0.05$  (Sig. *LOR-Down* o Sig. *LOR-Up*).

#### 4.3.1.1. Términos GO para procesos biológicos (bp)

Método	<i>Down</i>	<i>Up</i>	Sig. <i>Down</i>	Sig. <i>Up</i>	Sig. <i>LOR-Down</i>	Sig. <i>LOR-Up</i>
DL	3367	2727	0	0	0	0
HE	3366	2730	0	0	0	0
HS	3357	2729	0	0	0	0
FE	3308	2785	271	136	3	2

**Tabla 4.11:** Resultados del metaanálisis de los estudios individuales para los términos GO pertenecientes a la categoría de procesos biológicos.

El metaanálisis de términos GO no mostró ningún procesos biológico significativo entre los grupos de nuestra comparación para los modelos de efectos aleatorios. En el caso del modelo de efectos fijos sí que se detectaron términos significativos, pero sabemos que nuestros estudios primarios tienen distinta procedencia y son variables entre ellos, por lo que el modelo FE no es adecuado.

#### 4.3.1.2. Términos GO para funciones moleculares (mf)

Método	<i>Down</i>	<i>Up</i>	Sig. <i>Down</i>	Sig. <i>Up</i>	Sig. LOR- <i>Down</i>	Sig. LOR- <i>Up</i>
DL	718	460	0	0	0	0
HE	718	460	0	0	0	0
HS	716	461	0	0	0	0
FE	702	472	84	36	1	1

**Tabla 4.12:** Resultados del metaanálisis de los estudios individuales para los términos GO pertenecientes a la categoría de funciones moleculares.

En el caso de las funciones moleculares, el metaanálisis utilizando modelos aleatorios tampoco encontró diferencias significativas entre (NASH.Mujeres - NAFL.Mujeres) y (NASH.Hombres - NAFL.Hombres).

#### 4.3.1.3. Términos GO para componentes celulares (cc)

Método	<i>Down</i>	<i>Up</i>	Sig. <i>Down</i>	Sig. <i>Up</i>	Sig. LOR- <i>Down</i>	Sig. LOR- <i>Up</i>
DL	444	274	2	2	1	2
HE	444	274	2	2	1	2
HS	443	275	2	3	1	2
FE	428	292	84	34	1	4

**Tabla 4.13:** Resultados del metaanálisis de los estudios individuales para los términos GO pertenecientes a la categoría de componentes celulares.

El metaanálisis sí que mostró términos GO significativos en el caso de los componentes celulares utilizando los modelos aleatorios. La Tabla 4.13 muestra los términos significativos según los modelos de efectos aleatorios.

La Tabla 4.14 muestra 3 funciones significativamente sobreexpresadas en mujeres (GO:0005581\*, GO:0005583 y GO:0098643) y 2 funciones significativamente sobre-

expresadas en hombres (GO:0036038 y GO:0070971). El término GO:0005581 (marcado con un asterisco) únicamente fue indicado como significativo por el modelo HS.

Algunos de estos componentes celulares se presentaron diferencialmente expresados en estudios previos, lo que aporta solidez a nuestros resultados. Los términos significativos relacionados con los trímeros colágeno (*collagen trimer*) (GO:0005581\* y GO:0005583) habían sido encontrados como significativamente diferenciales en un estudio previo que comparaba pacientes con estadios medios del NAFLD contra pacientes con estadios avanzados [89]. Otros estudios [88] indican rutas metabólicas relacionadas con el retículo endoplasmático (GO:0070971) afectadas en pacientes con la enfermedad.

ID	Nombre	LI 95 %	LOR	LS 95 %
<b>GO:0005581*</b>	<i>Collagen trimer</i>	0.122	0.261	0.4
<b>GO:0005583</b>	<i>Fibrillar collagen trimer</i>	0.351	0.641	0.931
<b>GO:0036038</b>	<i>MKS complex</i>	-0.748	-0.501	-0.254
<b>GO:0070971</b>	<i>Endoplasmic reticulum exit site</i>	-0.702	-0.475	-0.249
<b>GO:0098643</b>	<i>Banded collagen fibril</i>	0.351	0.641	0.931

ID	P	FDR	QE	QEp	SE	$\tau^2$	$I^2$	$H^2$
<b>GO:0005581*</b>	0	0.034	8.884	0.064	0.071	0.011	43.279	1.763
<b>GO:0005583</b>	0	0.005	1.768	0.622	0.148	0	0	1
<b>GO:0036038</b>	0	0.013	1.516	0.824	0.126	0	0	1
<b>GO:0070971</b>	0	0.009	2.359	0.67	0.116	0	0	1
<b>GO:0098643</b>	0	0.005	1.768	0.622	0.148	0	0	1

**Tabla 4.14:** Estimadores de la medida del efector e indicadores de la heterogeneidad en el metaanálisis funcional con componentes celulares. El término GO:0005581 (marcado con un asterisco) únicamente fue significativo por el método HS.

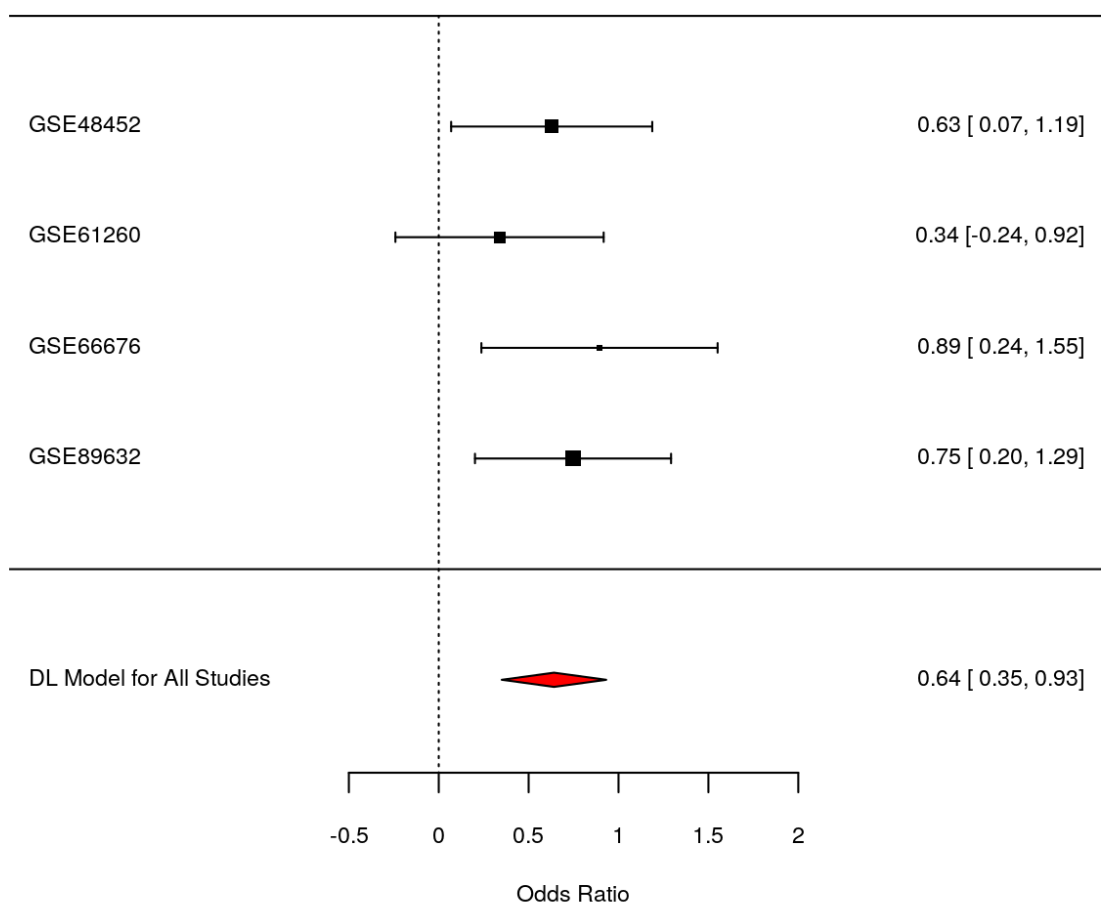
En las Figuras 4.6 a 4.9 se muestran los gráficos de bosque de los términos significativos. Tal y como se indica en el Apartado 3.3.1.2, estas representaciones muestran el *LOR* y el intervalo de confianza para cada estudio individual y para el metaanálisis. Los resultados del metaanálisis se muestran en forma de rombo rojo y el signo indica en que elemento de la comparación está sobreexpresada la función (positivo: mujeres, negativo: hombres).

Se puede observar cómo las Figuras 4.6 y 4.9 no muestran información sobre el estudio GSE83452. Esto es así debido a que ninguno de los genes estudiados en dicho análisis está asociados al término GO analizado y, por lo tanto, el estudio no puede aportar ningún valor al metaanálisis.

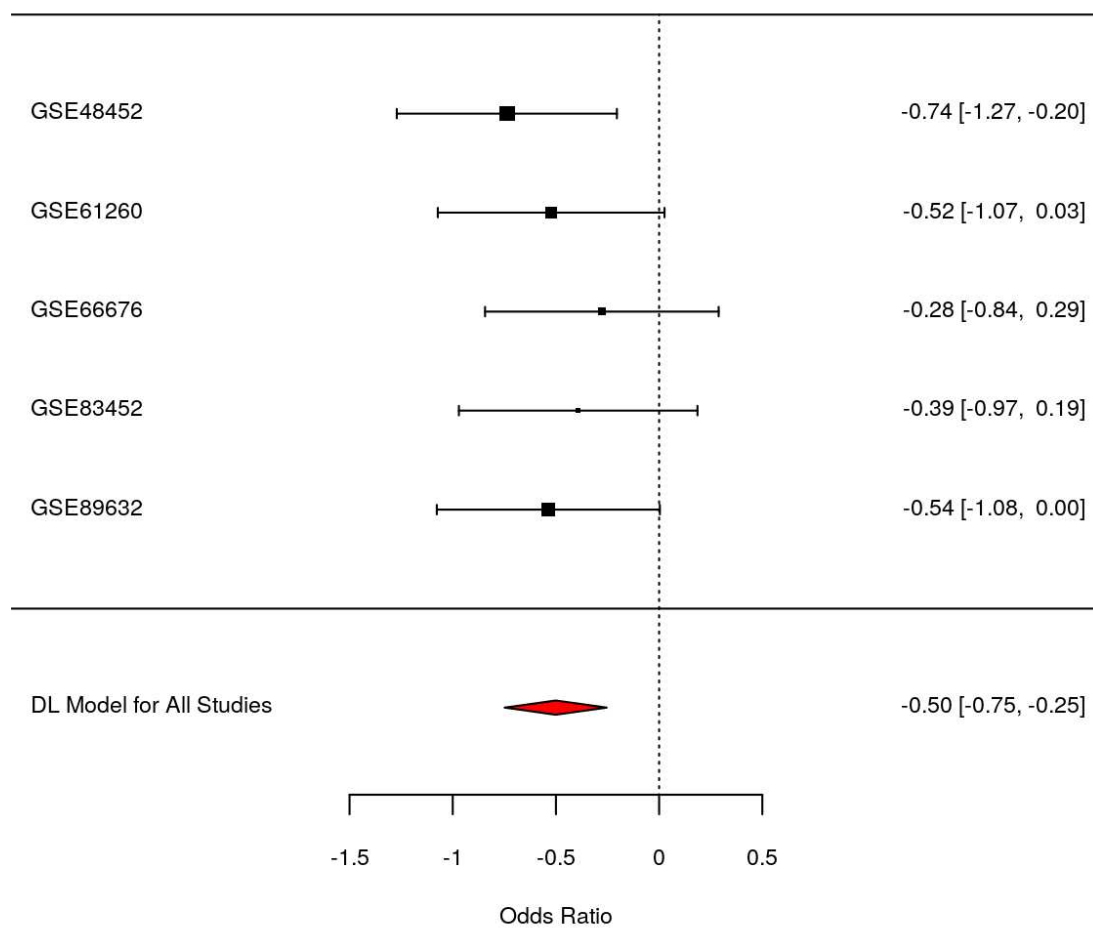
Tras los gráficos de bosque se muestran dos árboles que ubican en su contexto biológico los términos GO de componentes celulares significativos en mujeres (Figura 4.10) y en hombres (Figura 4.11).

Por otro lado, los diagramas de embudo muestran que todos los estudios se encuentran en la región de confianza (delimitada por la zona blanca), lo cual da soporte a los resultados del metaanálisis.

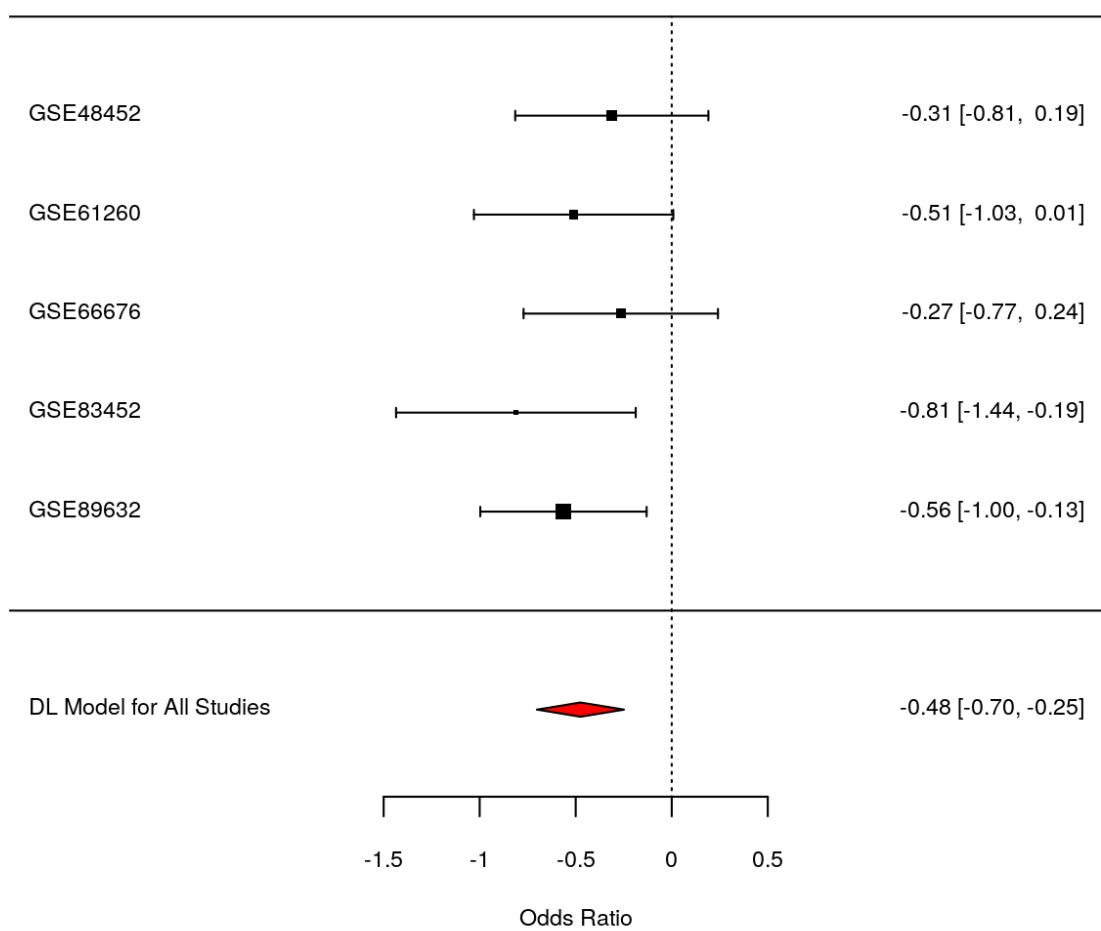
Finalmente, en el Anexo de figuras, se muestran los diagramas de embudo de los términos significativos (Figuras I.11 a I.14). Este tipo de representación nos permite ver la variabilidad y los sesgos para cada término estudiado. En el caso de nuestros 4 componentes celulares significativos, todos los estudios se encuentran en la región de confianza (delimitada por la zona blanca), confirmando la robustez de los resultados del metaanálisis.

**GO:0005583 (fibrillar collagen trimer)**

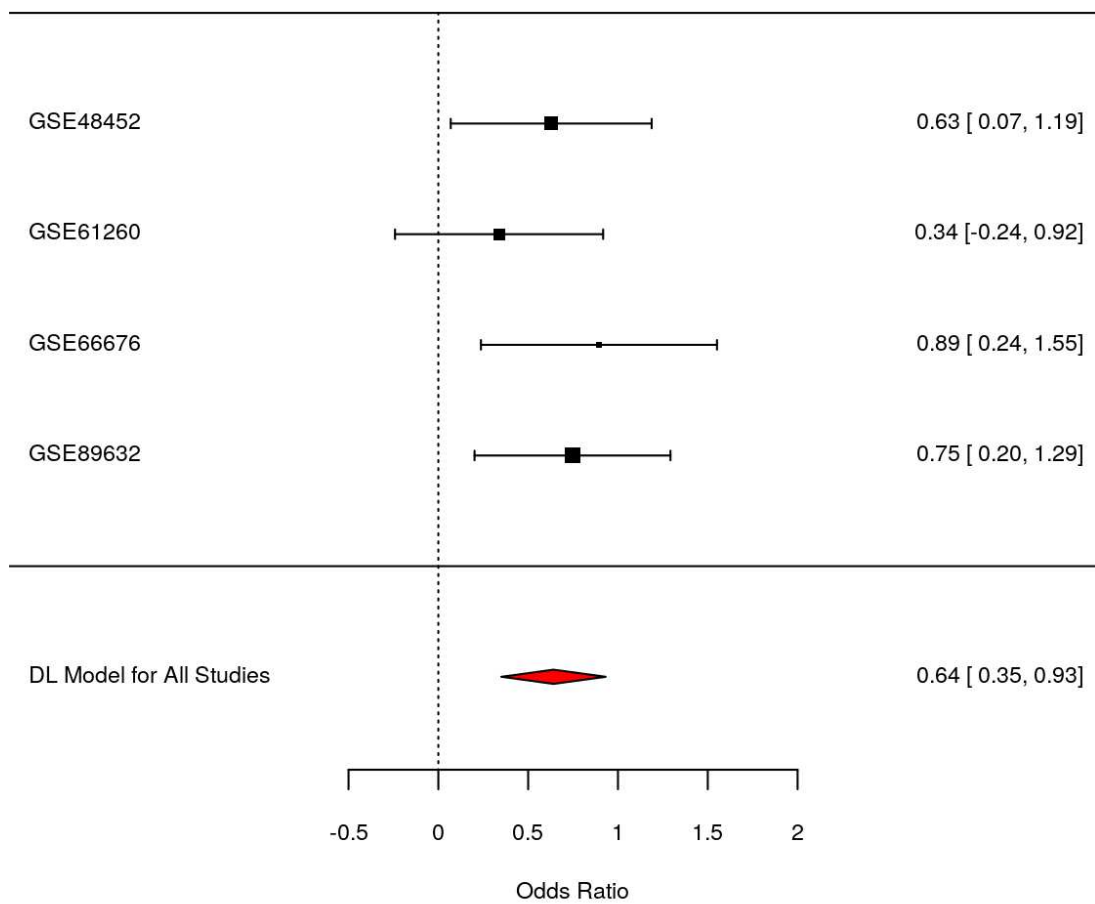
**Figura 4.6:** Gráfico de bosque para el término GO:0005583. Metaanálisis para componentes celulares.

**GO:0036038 (MKS complex)**

**Figura 4.7:** Gráfico de bosque para el término GO:0036038. Metaanálisis para componentes celulares.

**GO:0070971 (endoplasmic reticulum exit site)**

**Figura 4.8:** Gráfico de bosque para el término GO:0070971. Metaanálisis para componentes celulares.

**GO:0098643 (banded collagen fibril)**

**Figura 4.9:** Gráfico de bosque para el término GO:0098643. Metaanálisis para componentes celulares.

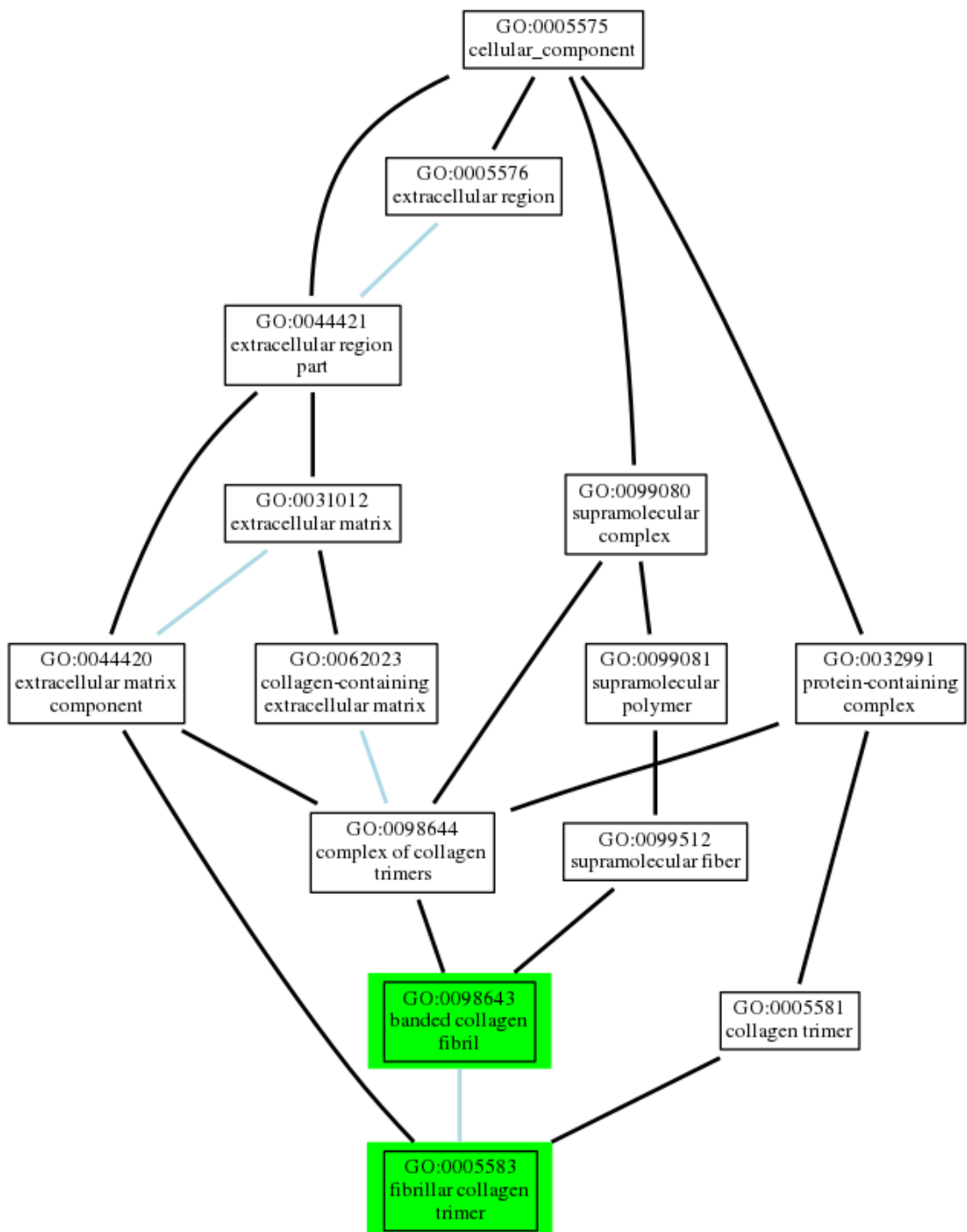


Figura 4.10: Componentes celulares significativos con sobrerrepresentación en las mujeres.

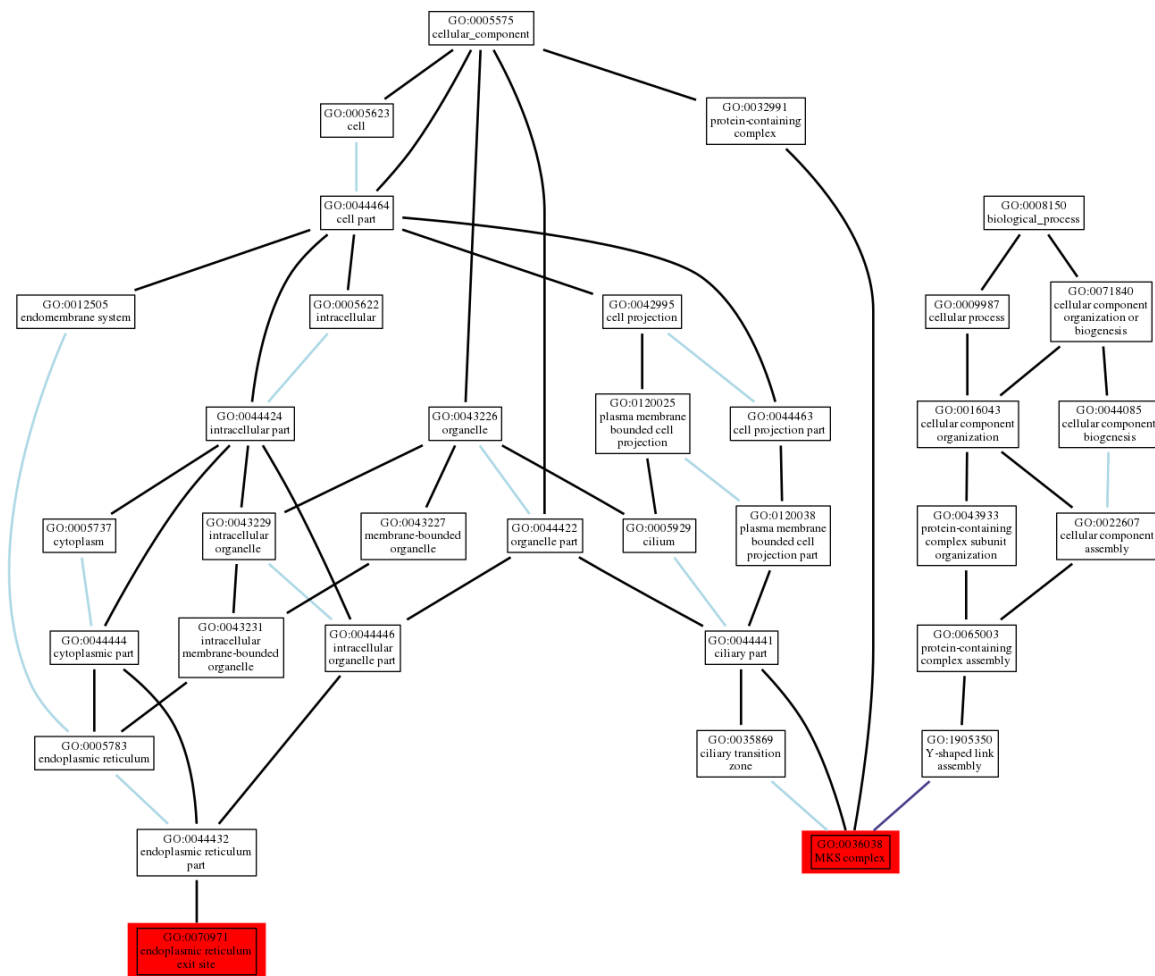


Figura 4.11: Componentes celulares significativos con sobrerrepresentación en los hombres.

#### 4.3.1.4. Rutas KEGG

Método	<i>Down</i>	<i>Up</i>	Sig. <i>Down</i>	Sig. <i>Up</i>	Sig. <i>LOR-Down</i>	Sig. <i>LOR-Up</i>
DL	153	76	0	0	0	0
HE	154	75	0	0	0	0
HS	151	78	0	0	0	0
FE	141	88	22	15	3	0

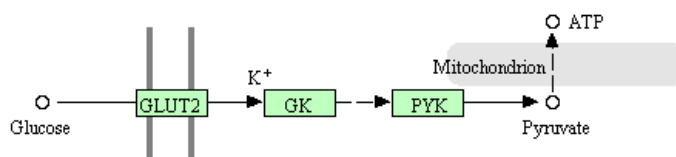
**Tabla 4.15:** Resultados del metaanálisis de los estudios individuales para rutas KEGG.

El metaanálisis de las rutas KEGG no mostró, en el caso de los modelos de efectos aleatorios, ninguna ruta significativa en nuestra comparación (Tabla 4.15).

### 4.3.2 Rutas de señalización

Los resultados del metaanálisis de los datos individuales obtenidos mediante *Hi-pathia* no mostraron ningún resultado significativo ( $FDR < 0.05$ ) a nivel de subruta, término GO o término UniProt. La Tabla 4.16 muestra los resultados de las subrutas ( $FDR < 1$ ), mientras que los resultados de los términos GO y UniProt no se muestran por dar en todos los casos un  $FDR > 0.20$ .

En el caso de las subrutas, se observó una subruta con un p-valor ajustado de 0.0572 (P-hsa04930-8, Figura 4.12) que tendremos en cuenta, ya que se encuentra muy próximo a nuestro punto de corte y su nivel de significación está muy alejado de la subruta mas próxima. Estudios en ratas y ratones publicados recientemente [90, 91], sugieren la implicación del piruvato (nodo efector de la subruta P-hsa04930-8) en la patogénesis del NAFLD.



**Figura 4.12:** Representación de la subruta P-hsa04930-8 en KEGG.

Modificada de: [https://www.genome.jp/kegg-bin/show\\_pathway?hsa04930](https://www.genome.jp/kegg-bin/show_pathway?hsa04930)

ID	Nombre	p-valor	FDR	Sentido
P-hsa04930-8	<i>Type II diabetes mellitus: Pyruvate</i>	0.0008	0.0572	Up
P-hsa04623-9	<i>Cytosolic DNA-sensing pathway: IKBKE</i>	0.0307	0.7184	Up
P-hsa04950-5	<i>Maturity onset diabetes of the young: GCK</i>	0.0220	0.7184	Up
P-hsa04210-64	<i>Apoptosis: TNFRSF10D*</i>	0.0597	0.8367	Up
P-hsa05215-9	<i>Prostate cancer: CDKN1B</i>	0.0574	0.8367	Up
P-hsa05205-113	<i>Proteoglycans in cancer: TP53</i>	0.0811	0.9462	Up
P-hsa04728-36	<i>Dopaminergic synapse: TH</i>	0.0972	0.9730	Up

**Tabla 4.16:** Rutas según su nivel de significación en el metaanálisis con *Hipathia*. Los términos *Down* y *Up* hacen referencia al término de la comparación en el que las rutas están significativamente activadas de forma diferencial (*Up* al primero (mujeres) y *Down* al segundo (hombres)).

---

---

## CAPÍTULO 5

---

### Discusión

El metaanálisis es un método robusto para el análisis conjunto de datos. En este trabajo hemos confirmado cómo esta estrategia nos ha permitido generar información conjunta de los estudios que no se podía obtener de la intersección o de la suma de los resultados individuales. Esto lo hemos conseguido para las dos aproximaciones planteadas: enriquecimiento funcional y rutas de señalización.

Desde el punto de vista de las diferencias de género, presenta especial interés la búsqueda de las bases moleculares que marcan el paso del estadio inicial de la enfermedad (NAFL) al estadio más avanzado (NASH), y que lo hacen de forma diferencial entre mujeres y hombres. Por este motivo, de entre todas las posibles comparaciones, nos concentramos en (NASH.Mujeres - NAFL.Mujeres) - (NASH.Hombres - NAFL.Hombres).

El metaanálisis de los estudios primarios aportó resultados interesantes, algunos de los cuales ya habían sido previamente descritos como relacionados con el NAFLD [88, 89, 90, 91], sin embargo, nunca en el contexto de las diferencias de género de esta patología.

Este trabajo ha mostrado la validez del metanálisis para el estudio de diferencias de género en el NAFLD a partir de datos disponibles en repositorios públicos. Sin embargo, nuestro trabajo también ha evidenciado algunas de las limitaciones de esta metodología. El bajo número de individuos por grupo, en cada uno de los estudios individuales, condiciona desde el principio la potencia estadística de nuestro trabajo. Un mayor número de pacientes por grupo (y un mayor número de estudios) podrían aportar una mayor cantidad de resultados significativos. En la actualidad, algunos organismos como la Unión Europea están empezando a exigir que los resultados obtenidos de estudios bajo su financiación sean puestos a disposición de la comunidad científica bajo los principios FAIR [44], permitiendo su acceso y reutilización posterior.

En la caracterización funcional mediante GSEA, los resultados globales mostraron 4 componentes celulares significativos para todos los métodos con efectos aleatorios. En mujeres, la sobreexpresión se observó en los términos GO:0005583 y GO:0098643 (Figura 4.10), ambos relacionados con el colágeno. Alteraciones asociadas al colágeno en pacientes con NAFLD ya habían sido descritas en estudios previos [88, 89] y se había detectado que el gen *COL1A1* (*Collagen type I alpha 1 chain*) se encuentra sobreexpresado en pacientes con NASH frente a pacientes control [89]. Además, se conoce que dicho gen se encuentra activo durante el proceso de fibrosis en el hígado [92, 93], lo que nos puede llevar a pensar que las alteraciones del colágeno en la enfermedad se desarrollan a nivel de la fibrogénesis hepática que pueden sufrir los pacientes en estadio NASH. Sin embargo, algunos autores van más allá y proponen que el gen *COL1A1*, junto con el gen *COL1A2*, podría también jugar un papel importante en regulación de genes asociados con la adhesión celular y la inflamación durante la enfermedad [89]. Aunque ya existía bibliografía previa que relacionaba el colágeno con el hígado graso no alcohólico, esta es la primera ocasión en que se aporta una perspectiva de género a estos resultados. En hombres, los componentes celulares sobreexpresados correspondían a los términos GO:0036038 (*MKS complex*)

---

y GO:0070971 (*endoplasmatic reticulum exit site*) (Figura 4.11), que se relacionan por primera vez con la enfermedad y abren una nueva vía al tratamiento de la misma.

Por otro lado, en la caracterización proporcionada por el metaanálisis de los resultados de rutas de señalización mediante *Hipathia*, la subruta P-hsa04930-8 de la ruta *Type II diabetes mellitus* y su metabolito: el piruvato (Figura 4.12) fueron destacados como diferencialmente activados en el avance de la enfermedad en mujeres. Estudios recientes en ratones, indican que alteraciones en enzimas asociadas al metabolismo del piruvato, como la piruvato deshidrogenasa quinasa 4, afectan a la formación de lípidos [94] y a la progresión del NAFLD [90, 91, 94]. De hecho, en los últimos meses, han aparecido publicaciones que sugieren la modulación del metabolismo del piruvato (a nivel mitocondrial) como tratamiento para la enfermedad [94, 95]. Estas investigaciones, junto con la perspectiva de género aportada por este trabajo, podrían llevar a un tratamiento más personalizado de la enfermedad.

Del mismo modo que en la revisión sistemática y en la selección de estudios es clave el asesoramiento de especialistas hepatólogos, también en la interpretación de resultados resulta esencial la participación de expertos en este área clínica que valoren la relevancia de los mecanismos moleculares obtenidos y planifiquen la adecuada validación experimental de los resultados alcanzados.

En las estrategias de caracterización funcional utilizadas se evalúan miles de funciones biológicas simultáneamente, proporcionando numerosos resultados. Para evitar falsos positivos y detectar aquellas funciones con mayor robustez en el conjunto de los estudios evaluados, se han seleccionado métodos con criterios estrictos. Por ello el número final de elementos significativos en ambos metaanálisis no es elevado. Este enfoque permite conocer y evaluar aquellos cambios de mayor intensidad en NAFLD con una perspectiva diferencial de género.

La relajación de algunos de estos criterios de evaluación, como un incremento del nivel de significación, 0.10 en lugar de 0.05, o la selección de un método de efectos fijos en el metaanálisis (en lugar de un método de efectos aleatorios) proporcionaría un mayor número de resultados significativos (Tablas 4.11, 4.12, 4.13 y 4.15).

---

---

## CAPÍTULO 6

---

### Conclusiones

1. El metaanálisis es un método robusto con capacidad de obtener resultados globales a partir de la evaluación e integración de datos procedentes de una selección de estudios *ómicos*.
2. Por primera vez se ha utilizado esta metodología para identificar bases moleculares de la enfermedad del hígado graso no alcohólico, alteradas de forma diferente entre géneros, a partir de datos disponibles en repositorios públicos.
3. Se han detectado dos componentes celulares relacionados con el colágeno (GO:0005583 y GO:0098643) significativamente sobreexpresados en mujeres, que habían sido asociados a la enfermedad previamente.
4. Los componentes celulares *MKS complex* (GO:0036038) y *endoplasmatic reticulum exit site* (GO:0070971) se encuentran sobreexpresados en hombres, lo cual no había sido previamente descrito.
5. La subruta P-hsa04930-8 de la ruta *Type II diabetes mellitus* y su metabolito: el piruvato, se encuentran sobreexpresados en las mujeres con respecto a los hombres. La validación experimental de este resultado podría llevar a un tra-

tamiento más personalizado de la enfermedad mediante la modulación del metabolismo mitocondrial del piruvato.

6. El diseño y ejecución de la estrategia bioinformática desarrollada abre una nueva vía para detectar y caracterizar las diferencias de género en salud mediante abordajes *ómicos*, incorporando esta necesaria perspectiva diferencial en los estudios biomédicos que permita potenciar una Medicina Personalizada y de Precisión.

---

## Bibliografía

- [1] Green, C. J. & Hodson, L. The influence of dietary fat on liver fat accumulation. *Nutrients* **6**, 5018–5033 (2014). URL <http://www.mdpi.com/2072-6643/6/11/5018>.
- [2] Lefebvre, P. *et al.* Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPAR $\alpha$ -regulated dermatopontin. *JCI Insight* **2** (2017). URL <https://insight.jci.org/articles/view/92264>.
- [3] Pappachan, J. M., Babu, S., Krishnan, B. & Ravindran, N. C. Non-alcoholic Fatty Liver Disease: A Clinical Update. *Journal of Clinical and Translational Hepatology* **5**, 384–393 (2017). URL <https://dx.doi.org/10.14218/JCTH.2017.000136>.
- [4] Younossi, Z. M. *et al.* Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* **64**, 73–84 (2016). URL <http://doi.wiley.com/10.1002/hep.28431>.
- [5] Skubic, C., Drakulić, Ž. & Rozman, D. Personalized therapy when tackling nonalcoholic fatty liver disease: a focus on sex, genes, and drugs. *Expert Opinion on Drug Metabolism & Toxicology* **14**, 831–841 (2018). URL <https://www.tandfonline.com/doi/full/10.1080/17425255.2018.1492552>.
- [6] Benedict, M. & Zhang, X. Non-alcoholic fatty liver disease: An expanded review. *World Journal of Hepatology* **9**, 715 (2017). URL <http://www.wjgnet.com/1948-5182/full/v9/i16/715.htm>.

- [7] Ballestri, S. *et al.* NAFLD as a Sexual Dimorphic Disease: Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk. *Advances in Therapy* **34**, 1291–1326 (2017). URL <http://link.springer.com/10.1007/s12325-017-0556-1>.
- [8] Hamaguchi, M. Aging is a risk factor of nonalcoholic fatty liver disease in premenopausal women. *World Journal of Gastroenterology* **18**, 237 (2012). URL <http://www.wjgnet.com/1007-9327/full/v18/i3/237.htm>.
- [9] Li, M., Reynolds, C. M., Segovia, S. A., Gray, C. & Vickers, M. H. Developmental Programming of Nonalcoholic Fatty Liver Disease: The Effect of Early Life Nutrition on Susceptibility and Disease Severity in Later Life. *BioMed Research International* **2015**, 1–12 (2015). URL <http://www.hindawi.com/journals/bmri/2015/437107/>.
- [10] Zheng, D., Wang, X., Antonson, P., Gustafsson, J.-Å. & Li, Z. Genomics of sex hormone receptor signaling in hepatic sexual dimorphism. *Molecular and Cellular Endocrinology* **471**, 33–41 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0303720717302940>.
- [11] Lee, Y.-H. *et al.* Sex-specific metabolic interactions between liver and adipose tissue in MCD diet-induced non-alcoholic fatty liver disease. *Oncotarget* **7**, 46959–46971 (2016). URL <http://www.oncotarget.com/fulltext/10506>.
- [12] Wang, S.-H. *et al.* Identification of androgen response elements in the enhancer I of hepatitis B virus: A mechanism for sex disparity in chronic hepatitis B. *Hepatology* **50**, 1392–1402 (2009). URL <http://doi.wiley.com/10.1002/hep.23163>.
- [13] Jiang, Z. *et al.* Sex-Related Differences of Lipid Metabolism Induced by Triptolide: The Possible Role of the LXR $\alpha$ /SREBP-1 Signaling Pathway. *Frontiers in Pharmacology* **7**, 87 (2016). URL <https://www.frontiersin.org/articles/10.3389/fphar.2016.00087>.

- [14] Wauthier, V. & Waxman, D. J. Sex-Specific Early Growth Hormone Response Genes in Rat Liver. *Molecular Endocrinology* **22**, 1962–1974 (2008). URL <https://doi.org/10.1210/me.2007-0549>.
- [15] Ling, G., Sugathan, A., Mazor, T., Fraenkel, E. & Waxman, D. J. Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression. *Molecular and Cellular Biology* **30**, 5531–5544 (2010). URL <http://mcb.asm.org/cgi/doi/10.1128/MCB.00601-10>.
- [16] Chalasani, N. *et al.* The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* **67**, 328–357 (2018). URL <http://doi.wiley.com/10.1002/hep.29367>.
- [17] Dyson, J. K., Anstee, Q. M. & McPherson, S. Non-alcoholic fatty liver disease: A practical approach to treatment. *Frontline Gastroenterology* **5**, 277–286 (2014). URL <http://doi.org/10.1136/flgastro-2013-100404>.
- [18] Rinella, M. E. Nonalcoholic Fatty Liver Disease: A Systematic Review. *JAMA* **313**, 2263 (2015). URL <https://doi.org/10.1001/jama.2015.5370>.
- [19] Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M. & Sanyal, A. J. Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine* **24**, 908–922 (2018). URL <http://www.nature.com/articles/s41591-018-0104-9>.
- [20] Aller, R. *et al.* Documento de consenso. Manejo de la enfermedad hepática grasa no alcohólica (EHGNA). Guía de práctica clínica. *Gastroenterología y Hepatología* **41**, 328–349 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0210570518300037>.
- [21] Ryaboshapkina, M. & Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Scientific Reports* **7**, 12361 (2017). URL <http://www.nature.com/articles/s41598-017-10930-w>.

- [22] Singh, S. *et al.* Fibrosis Progression in Nonalcoholic Fatty Liver vs Nonalcoholic Steatohepatitis: A Systematic Review and Meta-analysis of Paired-Biopsy Studies. *Clinical Gastroenterology and Hepatology* **13**, 643–654.e9 (2015). URL <http://linkinghub.elsevier.com/retrieve/pii/S1542356514006028>.
- [23] Lindenmeyer, C. C. & McCullough, A. J. The Natural History of Nonalcoholic Fatty Liver Disease—An Evolving View. *Clinics in Liver Disease* **22**, 11–21 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S1089326117300648>.
- [24] Adams, L. A., Sanderson, S., Lindor, K. D. & Angulo, P. The histological course of nonalcoholic fatty liver disease: a longitudinal study of 103 patients with sequential liver biopsies. *Journal of Hepatology* **42**, 132–138 (2005). URL <http://linkinghub.elsevier.com/retrieve/pii/S0168827804004350>.
- [25] Huang, P. L. A comprehensive definition for metabolic syndrome. *Disease Models & Mechanisms* **2**, 231–237 (2009). URL <http://dmm.biologists.org/cgi/doi/10.1242/dmm.001180>.
- [26] Castro-Martínez, M. G., Banderas-Lares, D. Z., Ramírez-Martínez, J. C. & Escobedo-de la Peña, J. Prevalence of nonalcoholic fatty liver disease in subjects with metabolic syndrome. *Cirugía y cirujanos* **80**, 128–33 (2012). URL <http://www.ncbi.nlm.nih.gov/pubmed/22644007>.
- [27] Buzzetti, E., Pinzani, M. & Tsochatzis, E. A. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metabolism* **65**, 1038–1048 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0026049515003832>.
- [28] Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics* **40**, 1461–1465 (2008). URL <http://www.nature.com/articles/ng.257>.
- [29] Kovac, U. & Rozman, D. Genetics of Non-alcoholic Fatty Liver Disease. In *eLS*, 1–8 (John Wiley & Sons, Ltd, Chichester, UK, 2015). URL <http://doi.wiley.com/10.1002/9780470015902.a0025335>.

- [30] Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308–311 (2001). URL <http://dx.doi.org/10.1093/nar/29.1.308>.
- [31] Eslam, M., Valenti, L. & Romeo, S. Genetics and epigenetics of NAFLD and NASH: Clinical impact. *Journal of Hepatology* **68**, 268–279 (2018). URL <http://linkinghub.elsevier.com/retrieve/pii/S0168827817322821>.
- [32] Adams, L. A. *et al.* Cholesteryl ester transfer protein gene polymorphisms increase the risk of fatty liver in females independent of adiposity. *Journal of Gastroenterology and Hepatology* **27**, 1520–1527 (2012). URL <http://doi.wiley.com/10.1111/j.1440-1746.2012.07120.x>.
- [33] Musso, G., Cassader, M., Bo, S., De Michieli, F. & Gambino, R. Sterol Regulatory Element-Binding Factor 2 (SREBF-2) Predicts 7-Year NAFLD Incidence and Severity of Liver Disease and Lipoprotein and Glucose Dysmetabolism. *Diabetes* **62**, 1109–1120 (2013). URL <http://diabetes.diabetesjournals.org/cgi/doi/10.2337/db12-0858>.
- [34] Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nature Genetics* **46**, 352–356 (2014). URL <http://www.nature.com/articles/ng.2901>.
- [35] Speliotes, E. K. *et al.* Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits. *PLoS Genetics* **7**, e1001324 (2011). URL <http://dx.plos.org/10.1371/journal.pgen.1001324>.
- [36] Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature Genetics* **43**, 1131–1138 (2011). URL <http://www.nature.com/articles/ng.970>.
- [37] Zeybel, M. *et al.* Differential DNA methylation of genes involved in fibrosis progression in non-alcoholic fatty liver disease and alcoholic liver disease. *Clinical Epigenetics* **7**, 25 (2015). URL <http://www.clinicalepigeneticsjournal.com/content/7/1/25>.

- [38] Hsu, S.-H. *et al.* Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. *Journal of Clinical Investigation* **122**, 2871–2883 (2012). URL <http://www.jci.org/articles/view/63539>.
- [39] Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970). URL <http://www.nature.com/articles/227561a0>.
- [40] Rigden, D. J. & Fernández, X. M. The 2018 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Research* **46**, D1–D7 (2018). URL <http://academic.oup.com/nar/article/46/D1/D1/4781210>.
- [41] Chervitz, S. A. *et al.* *Bioinformatics for Omics Data*, vol. 719 of *Methods in Molecular Biology* (Humana Press, Totowa, NJ, 2011). URL <http://link.springer.com/10.1007/978-1-61779-027-0>.
- [42] Tenenbaum, J. D., Sansone, S.-A. & Haendel, M. A sea of standards for omics data: sink or swim? *Journal of the American Medical Informatics Association* **21**, 200–203 (2014). URL <https://doi.org/10.1136/amiajnl-2013-002066>.
- [43] Lapatas, V., Stefanidakis, M., Jimenez, R. C., Via, A. & Schneider, M. V. Data integration in biological research: an overview. *Journal of Biological Research-Thessaloniki* **22**, 9 (2015). URL <http://www.jbiolres.com/content/22/1/9>.
- [44] Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016). URL <http://www.nature.com/articles/sdata201618>.
- [45] Doménech-Sáncheza, A. & Vilab, J. Fundamento, tipos y aplicaciones de los arrays de ADN en la microbiología médica. *Enfermedades Infecciosas y Microbiología Clínica* **22**, 46–54 (2004). URL <https://linkinghub.elsevier.com/retrieve/pii/S0213005X0473030X>.
- [46] Lockhart, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* **14**, 1675–1680 (1996). URL <http://doi.org/10.1038/nbt1296-1675>.

- [47] Schena, M. Genome analysis with gene expression microarrays. *BioEssays* **18**, 427–431 (1996). URL <http://doi.wiley.com/10.1002/bies.950180513>.
- [48] García-García, F. *Métodos de análisis de enriquecimiento funcional en estudios genómicos*. Ph.D. thesis, Universidad de Valencia (2016). URL <http://roderic.uv.es/handle/10550/55208>.
- [49] Lee, J. W., Lee, J. B., Park, M. & Song, S. H. An extensive comparison of recent classification tools applied to microarray data. *Computational Statistics & Data Analysis* **48**, 869–885 (2005). URL <http://linkinghub.elsevier.com/retrieve/pii/S016794730400101X>.
- [50] Madeira, S. & Oliveira, A. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **1**, 24–45 (2004). URL <http://ieeexplore.ieee.org/document/1324618/>.
- [51] Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000). URL <http://www.nature.com/doifinder/10.1038/75556>.
- [52] Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000). URL <https://doi.org/10.1093/nar/28.1.27>.
- [53] Alandes, S. *Abordajes bioinformáticos para la integración funcional de datos transcriptómicos en biomedicina*. Master's thesis, Universidad Politécnica de Valencia (2017). URL <http://hdl.handle.net/10251/86291>.
- [54] Al-Shahrour, F., Diaz-Uriarte, R. & Dopazo, J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics* **20**, 578–580 (2004). URL <https://doi.org/10.1093/bioinformatics/btg455>.
- [55] Al-Shahrour, F. *et al.* FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Research* **35**, W91–W96 (2007). URL

<https://doi.org/10.1093/nar/gkm260>.

- [56] Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267–273 (2003). URL <http://www.nature.com/articles/ng1180>.
- [57] Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.0506580102>.
- [58] Bolaños Díaz, R. & Calderón Cahua, M. Introducción al meta-análisis tradicional. *Gastroenterol Peru* **34**, 45–51 (2014). URL <http://www.scielo.org.pe/pdf/rgp/v34n1/a07v34n1.pdf>.
- [59] Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. & Chalmers, T. C. Meta-Analyses of Randomized Controlled Trials. *New England Journal of Medicine* **316**, 450–455 (1987). URL <https://doi.org/10.1056/NEJM198702193160806>.
- [60] Normand, S.-L. T. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* **18**, 321–359 (1999). URL <http://go.uv.es/9BRketz>.
- [61] Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* **6**, e1000097 (2009). URL <https://dx.plos.org/10.1371/journal.pmed.1000097>.
- [62] Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology* **62**, e1–e34 (2009). URL <http://linkinghub.elsevier.com/retrieve/pii/S0895435609001802>.
- [63] Catalá-López, F. & Tobías, A. Metaanálisis de ensayos clínicos aleatorizados, heterogeneidad e intervalos de predicción. *Medicina Clínica* **142**, 270–274 (2014). URL <http://doi.org/10.1016/j.medcli.2013.06.013>.

- [64] Glass, G. V. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* **5**, 3 (1976). URL <http://doi.org/10.2307/1174772>.
- [65] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2018). URL <https://www.R-project.org/>.
- [66] Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007). URL <https://doi.org/10.1093/bioinformatics/btm254>.
- [67] Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47–e47 (2015). URL <https://doi.org/10.1093/nar/gkv007>.
- [68] Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010). URL <http://doi.org/10.1093/bioinformatics/btq431>.
- [69] Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995). URL <http://www.jstor.org/stable/2346101>.
- [70] Carlson, M. *org.Hs.eg.db: Genome wide annotation for Human* (2018). URL <https://doi.org/doi:10.18129/B9.bioc.org.Hs.eg.db>. R package version 3.6.0.
- [71] Montaner, D. & Dopazo, J. Multidimensional gene set analysis of genomic data. *PLoS One* **5**, e10348 (2010). URL <https://doi.org/10.1371/journal.pone.0010348>.
- [72] Hidalgo, M. R. *et al.* High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. *Oncotarget* **8** (2017). URL <https://doi.org/10.18632/oncotarget.14107>.

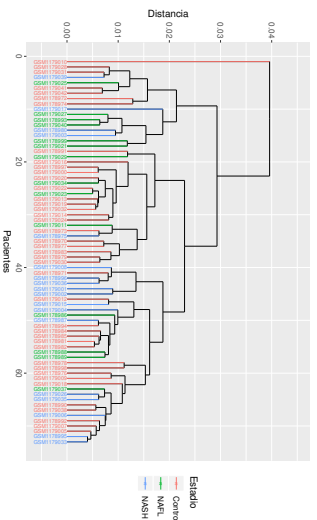
- [73] Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**, D158–D169 (2017). URL <https://doi.org/10.1093/nar/gkw1099>.
- [74] Viechtbauer, W. Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software, Articles* **36**, 1–48 (2010). URL <https://www.jstatsoft.org/v036/i03>.
- [75] DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188 (1986). URL [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
- [76] Hedges, L. V., Gurevitch, J. & Curtis, P. S. The Meta-Analysis of Response Ratios in Experimental Ecology. *Ecological Society of America* **80**, 1150–1156 (1999). URL <https://www.jstor.org/stable/177062>.
- [77] Schmidt, F. L. & Hunter, J. E. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, vol. 39 (SAGE Publications, Ltd, 1 Oliver’s Yard, 55 City Road London EC1Y 1SP, 2015). URL <http://dx.doi.org/10.4135/9781483398105>.
- [78] Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539–1558 (2002). URL <http://doi.wiley.com/10.1002/sim.1186>.
- [79] Sterne, J. A. & Egger, M. Funnel plots for detecting bias in meta-analysis. *Journal of Clinical Epidemiology* **54**, 1046–1055 (2001). URL <http://www.ncbi.nlm.nih.gov/pubmed/11576817>.
- [80] Dewey, M. *metap: meta-analysis of significance values* (2018). URL <https://CRAN.R-project.org/package=metap>. R package version 1.0.
- [81] Ahrens, M. *et al.* DNA Methylation Analysis in Nonalcoholic Fatty Liver Disease Suggests Distinct Disease-Specific and Remodeling Signatures after Bariatric Surgery. *Cell Metabolism* **18**, 296–302 (2013). URL <https://doi.org/10.1016/j.cmet.2013.07.004>.

- [82] Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences* **111**, 15538–15543 (2014). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1412759111>.
- [83] Xanthakos, S. A. *et al.* High Prevalence of Nonalcoholic Fatty Liver Disease in Adolescents Undergoing Bariatric Surgery. *Gastroenterology* **149**, 623–634.e8 (2015). URL <https://doi.org/10.1053/j.gastro.2015.05.039>.
- [84] Arendt, B. M. *et al.* Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. *Hepatology* **61**, 1565–1578 (2015). URL <http://doi.wiley.com/10.1002/hep.27695>.
- [85] Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188 (2001). URL <https://doi.org/10.1214/aos/1013699998>.
- [86] Dusa, A. *venn: Draw Venn Diagrams* (2018). URL <https://CRAN.R-project.org/package=venn>. R package version 1.7.
- [87] du Plessis, J. *et al.* Association of Adipose Tissue Inflammation With Histologic Severity of Nonalcoholic Fatty Liver Disease. *Gastroenterology* **149**, 635–648.e14 (2015). URL <https://doi.org/10.1053/j.gastro.2015.05.044>.
- [88] Hoyles, L. *et al.* Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nature Medicine* **24**, 1070–1080 (2018). URL <http://www.nature.com/articles/s41591-018-0061-3>.
- [89] Qi, S., Wang, C., Li, C., Wang, P. & Liu, M. Candidate genes investigation for severe nonalcoholic fatty liver disease based on bioinformatics analysis. *Medicine* **96**, e7743 (2017). URL <http://insights.ovid.com/crossref?an=00005792-201708110-00033>.
- [90] Moon, C.-M. *et al.* Metabolic biomarkers for non-alcoholic fatty liver disease induced by high-fat diet: In vivo magnetic resonance spectroscopy of hyperpolarized [1-<sup>13</sup>C] pyruvate. *Biochemical and Biophysical Research Communications*

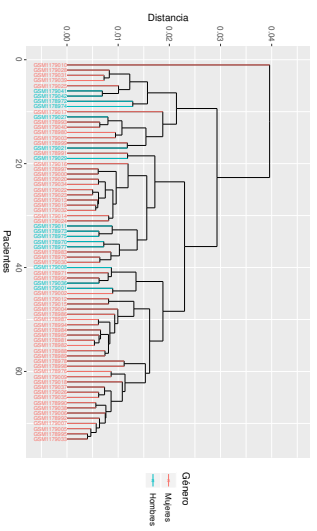
- 482, 112–119 (2017). URL <https://doi.org/10.1016/j.bbrc.2016.08.118>.
- [91] Zhang, M., Zhao, Y., Li, Z. & Wang, C. Pyruvate dehydrogenase kinase 4 mediates lipogenesis and contributes to the pathogenesis of nonalcoholic steatohepatitis. *Biochemical and Biophysical Research Communications* **495**, 582–586 (2018). URL <https://doi.org/10.1016/j.bbrc.2017.11.054>.
- [92] Asselah, T. *et al.* Liver Gene Expression Signature of Mild Fibrosis in Patients With Chronic Hepatitis C. *Gastroenterology* **129**, 2064–2075 (2005). URL <https://doi.org/10.1053/j.gastro.2005.09.010>.
- [93] Iizuka, M. *et al.* Induction of microRNA-214-5p in human and rodent liver fibrosis. *Fibrogenesis & Tissue Repair* **5**, 12 (2012). URL <https://doi.org/10.1186/1755-1536-5-12>.
- [94] McCommis, K. S. & Finck, B. N. Treating Hepatic Steatosis and Fibrosis by Modulating Mitochondrial Pyruvate Metabolism. *Cellular and Molecular Gastroenterology and Hepatology* (2018). URL <https://doi.org/10.1016/j.jcmgh.2018.09.017>.
- [95] Finck, B. N. Targeting Metabolism, Insulin Resistance, and Diabetes to Treat Nonalcoholic Steatohepatitis. *Diabetes* **67**, 2485–2493 (2018). URL <https://doi.org/10.2337/dbi18-0024>.

# Anexo I: Figuras

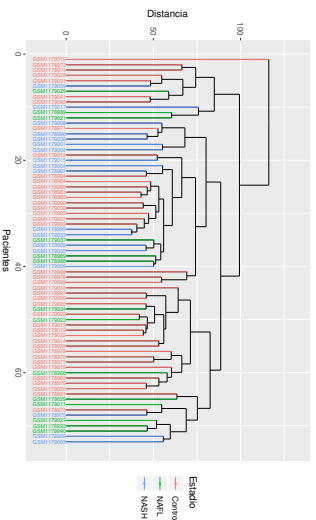
---



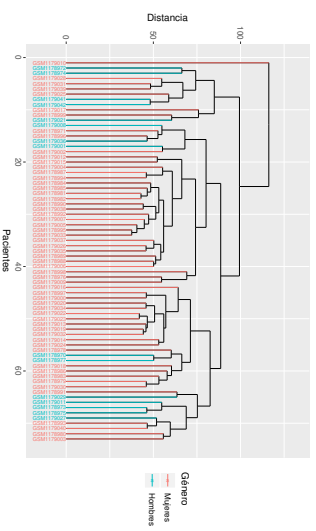
(a) Distancia de correlación (estadío)



(b) Distancia de correlación (género)

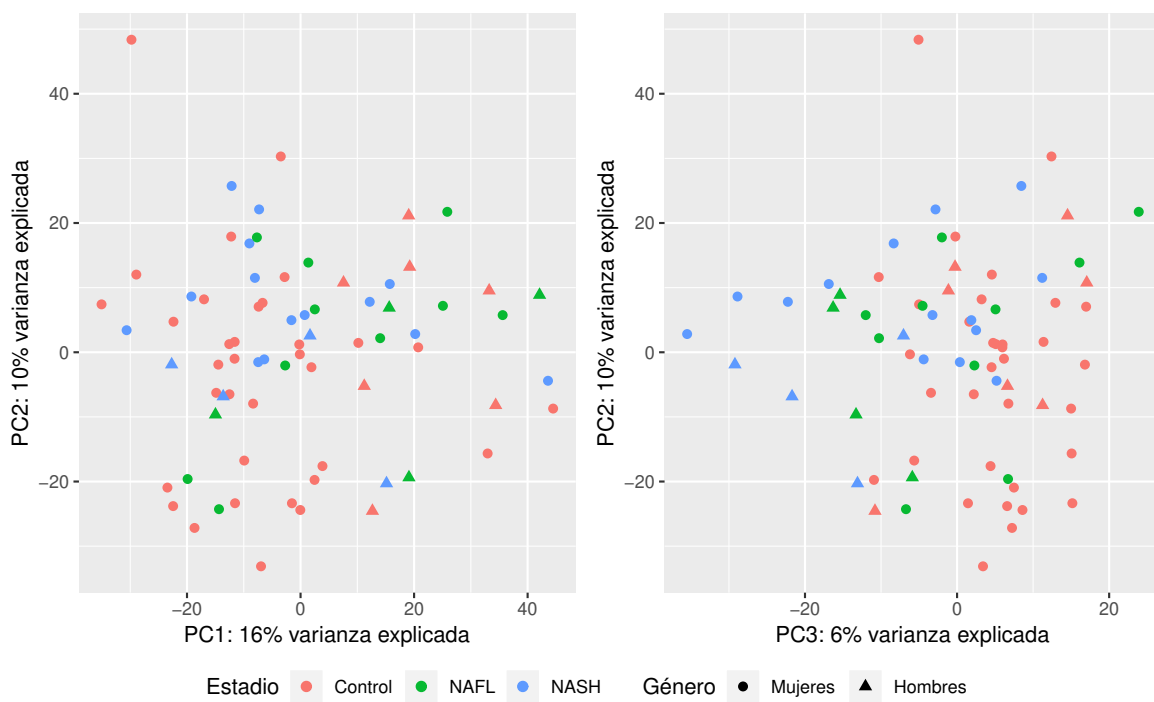


(c) Distancia de euclídea (estadío)

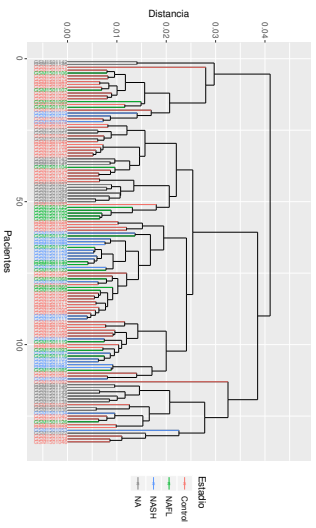


(d) Distancia de euclídea (género)

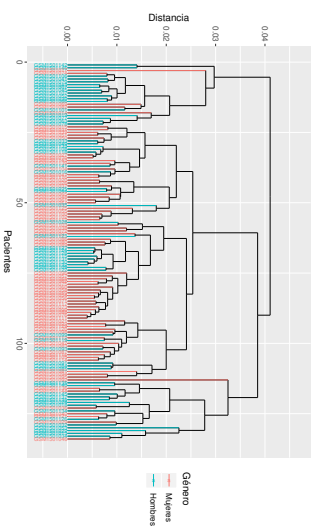
**Figura I.1:** *Clustering* exploratorio de las muestras del estudio GSE48452 mediante distancia de correlación (a y b) o distancia euclídea (c y d). Las muestras se encuentran coloreadas por el estadio de la enfermedad (a y c) o por el género de los individuos (b y d).



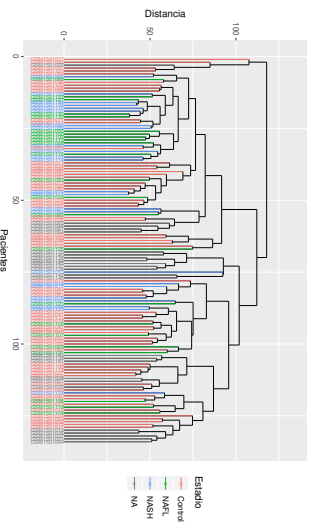
**Figura I.2:** Análisis de componentes principales de las muestras del estudio GSE48452. Los colores representan los estadios y las formas geométricas los géneros.



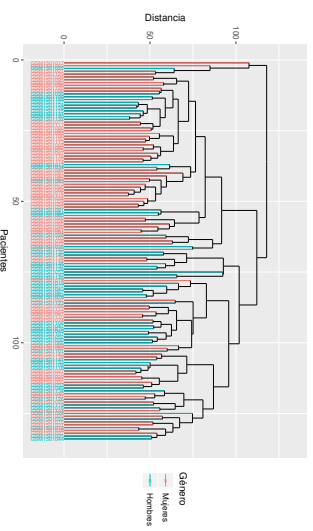
**(a)** Distancia de correlación (estadío)



**(b)** Distancia de correlación (género)

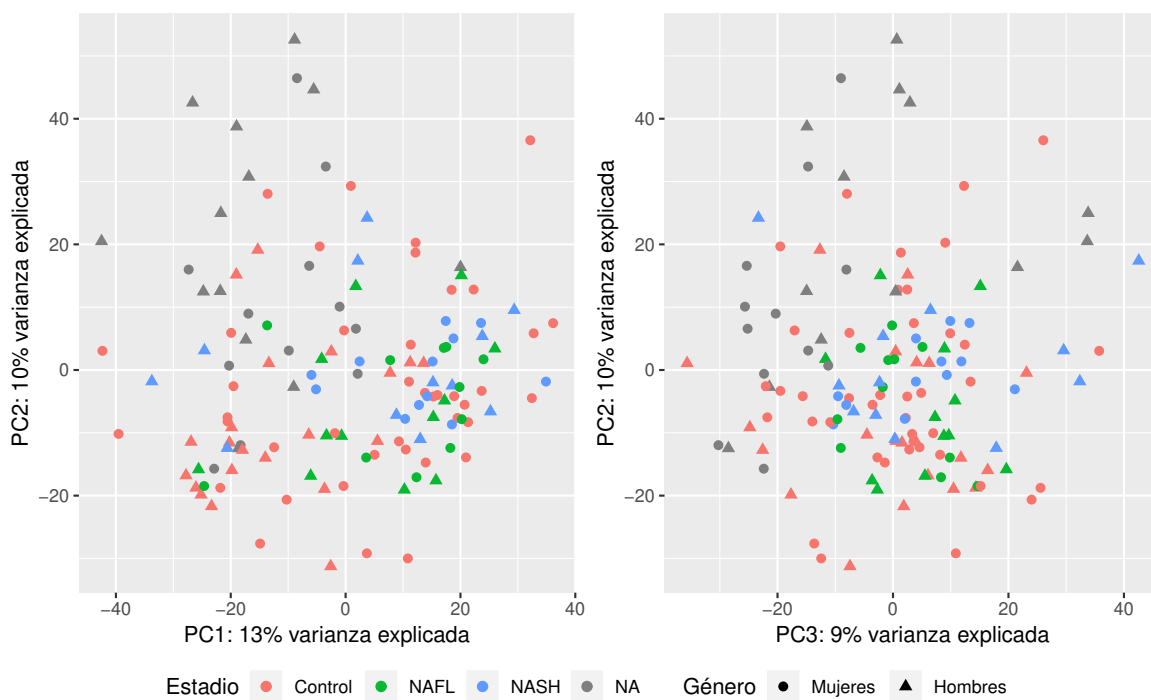


**(c)** Distancia de euclídea (estadío)

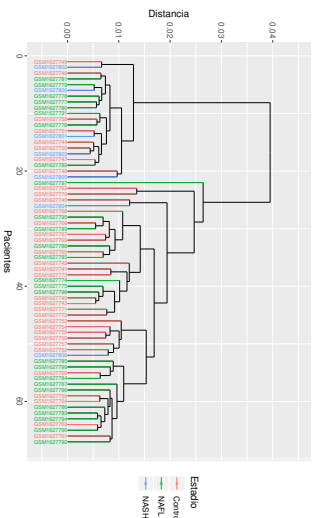


**(d)** Distancia de euclídea (género)

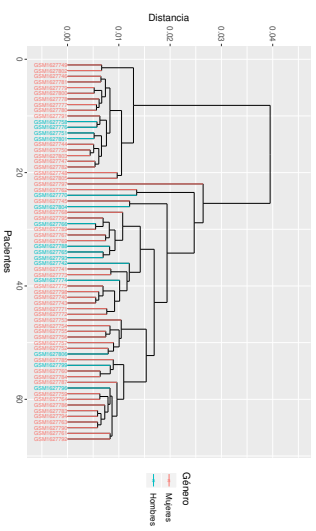
**Figura I.3:** *Clustering* exploratorio de las muestras del estudio GSE61260 mediante distancia de correlación (a y b) o distancia euclídea (c y d). Las muestras se encuentran coloreadas por el estadio de la enfermedad (a y c) o por el género de los individuos (b y d).



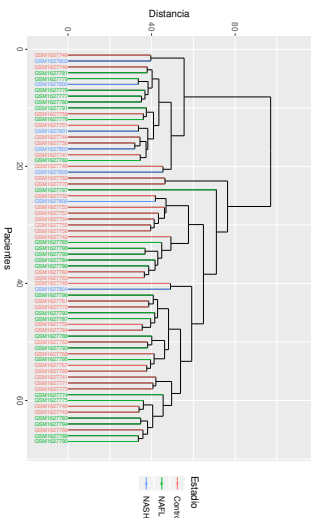
**Figura I.4:** Análisis de componentes principales de las muestras del estudio GSE61260. Los colores representan los estadios y las formas geométricas los géneros.



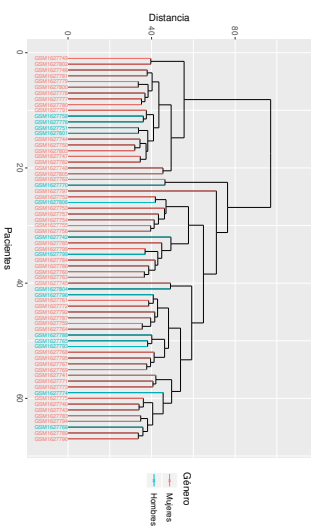
**(a)** Distancia de correlación (estadío)



**(b)** Distancia de correlación (género)



**(c)** Distancia de euclídea (estadío)

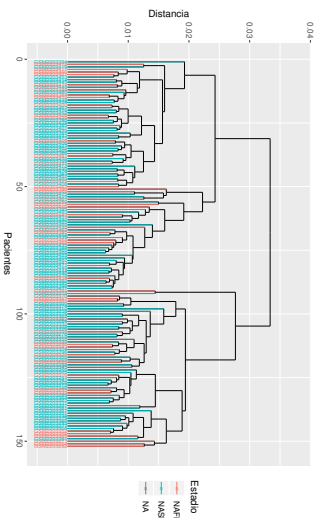


**(d)** Distancia de euclídea (género)

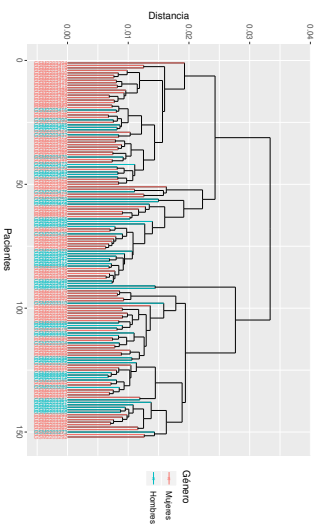
**Figura I.5:** *Clustering* exploratorio de las muestras del estudio GSE66676 mediante distancia de correlación (a y b) o distancia euclídea (c y d). Las muestras se encuentran coloreadas por el estadio de la enfermedad (a y c) o por el género de los individuos (b y d).



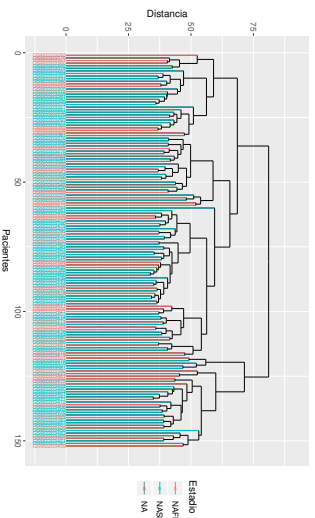
**Figura I.6:** Análisis de componentes principales de las muestras del estudio GSE66676. Los colores representan los estadios y las formas geométricas los géneros.



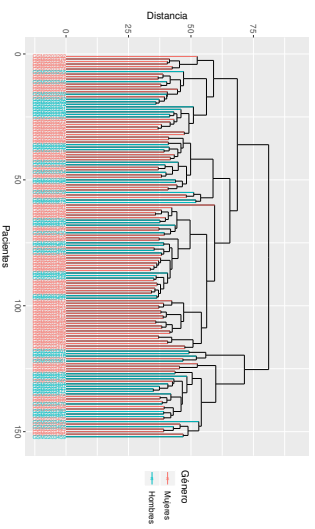
**(a)** Distancia de correlación (estado)



**(b)** Distancia de correlación (género)

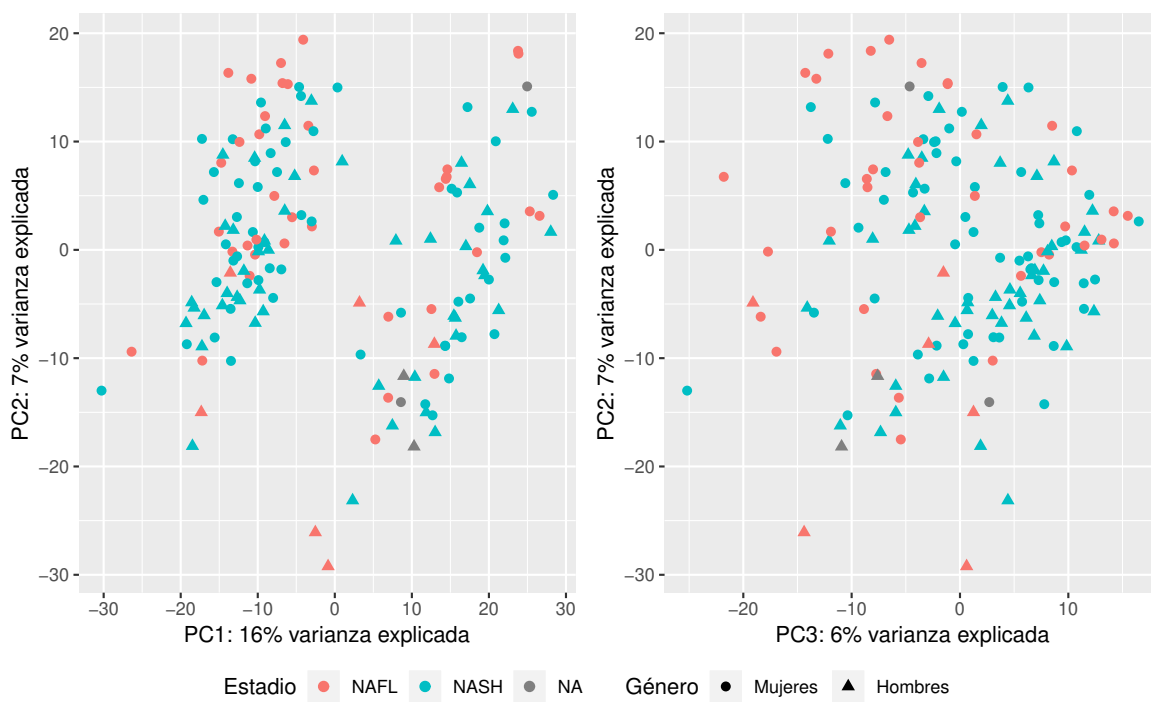


**(c)** Distancia de euclídea (estado)

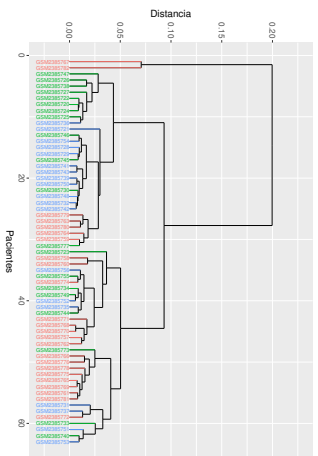


**(d)** Distancia de euclídea (género)

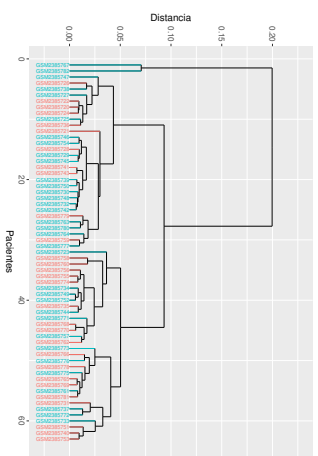
**Figura I.7:** *Clustering* exploratorio de las muestras del estudio GSE83452 mediante distancia de correlación (a y b) o distancia euclídea (c y d). Las muestras se encuentran coloreadas por el estadio de la enfermedad (a y c) o por el género de los individuos (b y d).



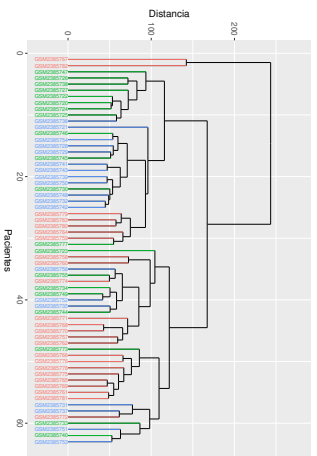
**Figura I.8:** Análisis de componentes principales de las muestras del estudio GSE83452. Los colores representan los estadios y las formas geométricas los géneros.



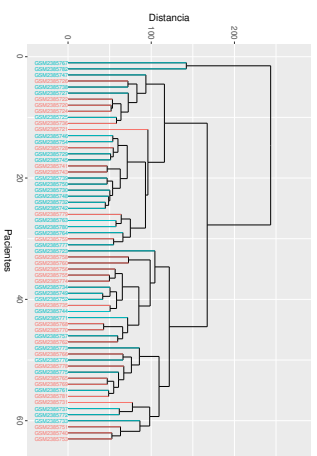
(a) Distancia de correlación (estado)



(b) Distancia de correlación (género)

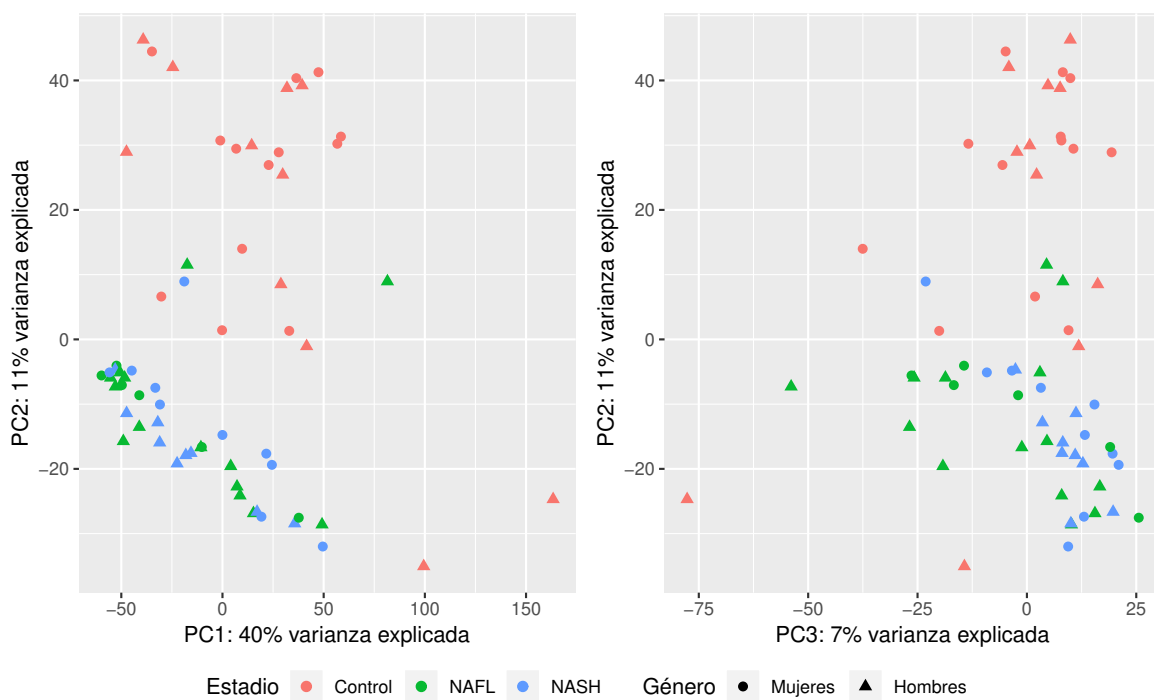


(c) Distancia de euclídea (estado)



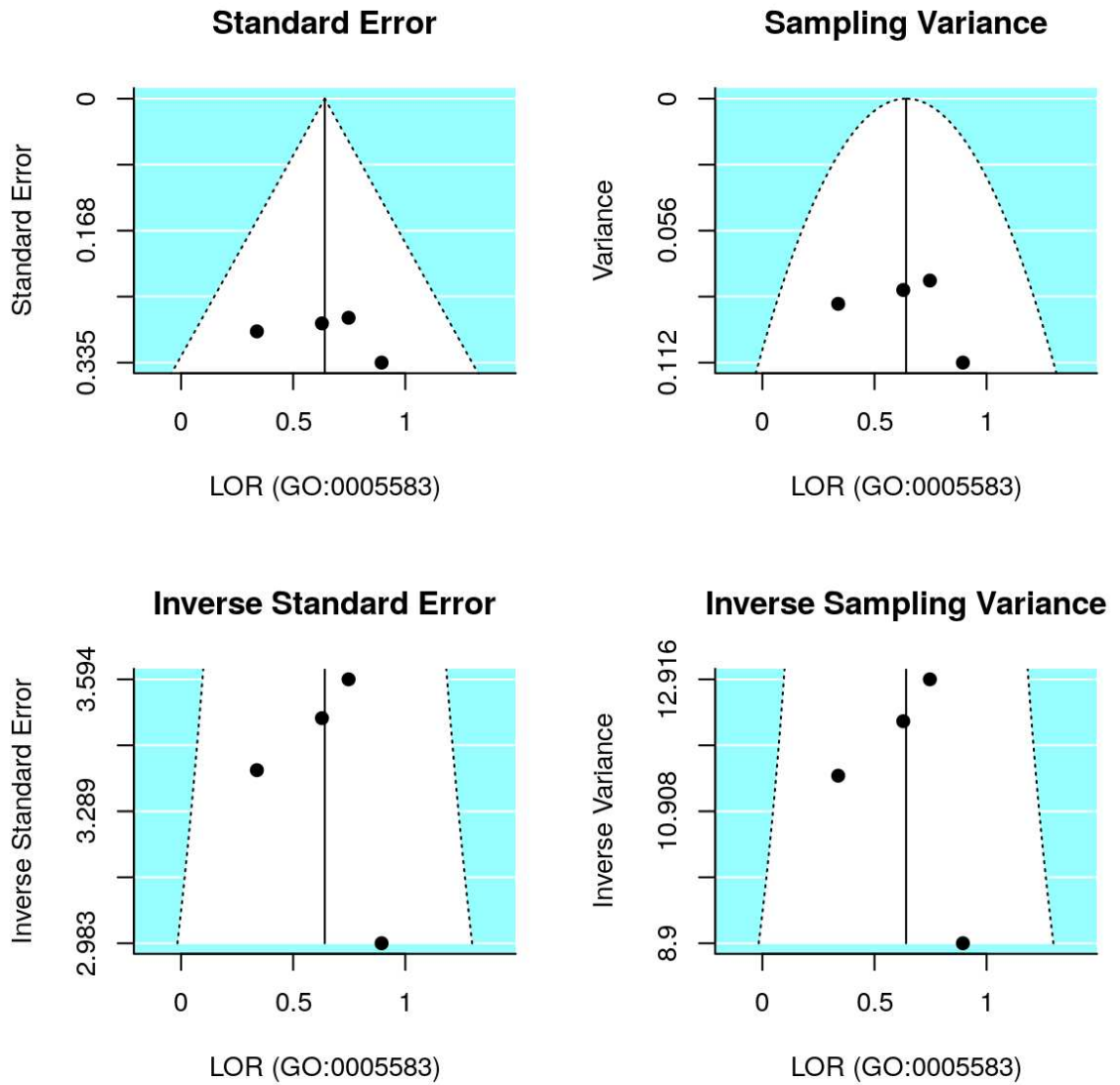
(d) Distancia de euclídea (género)

**Figura I.9:** *Clustering* exploratorio de las muestras del estudio GSE89632 mediante distancia de correlación (a y b) o distancia euclídea (c y d). Las muestras se encuentran coloreadas por el estadio de la enfermedad (a y c) o por el género de los individuos (b y d).

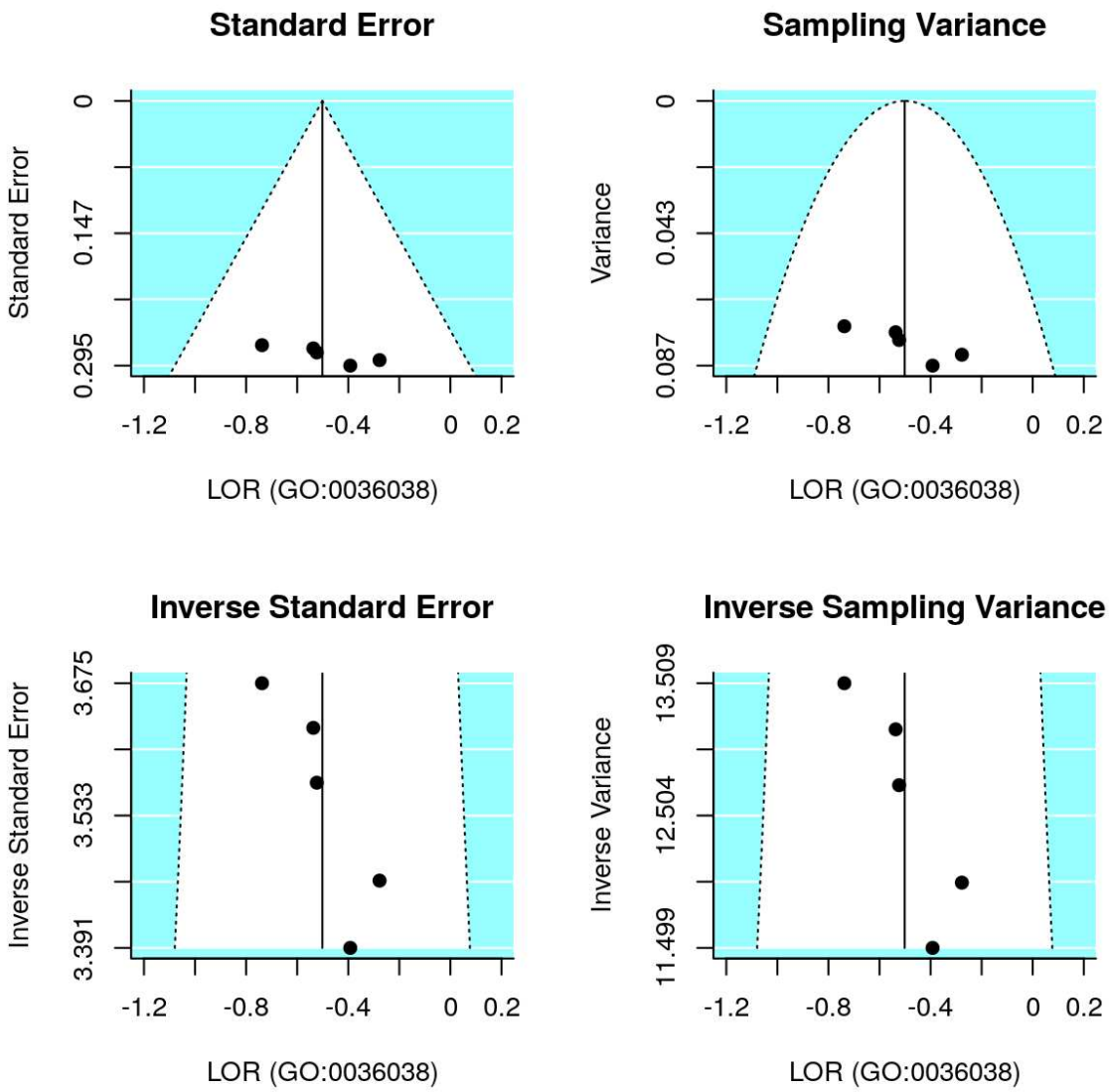


**Figura I.10:** Análisis de componentes principales de las muestras del estudio GSE89632.

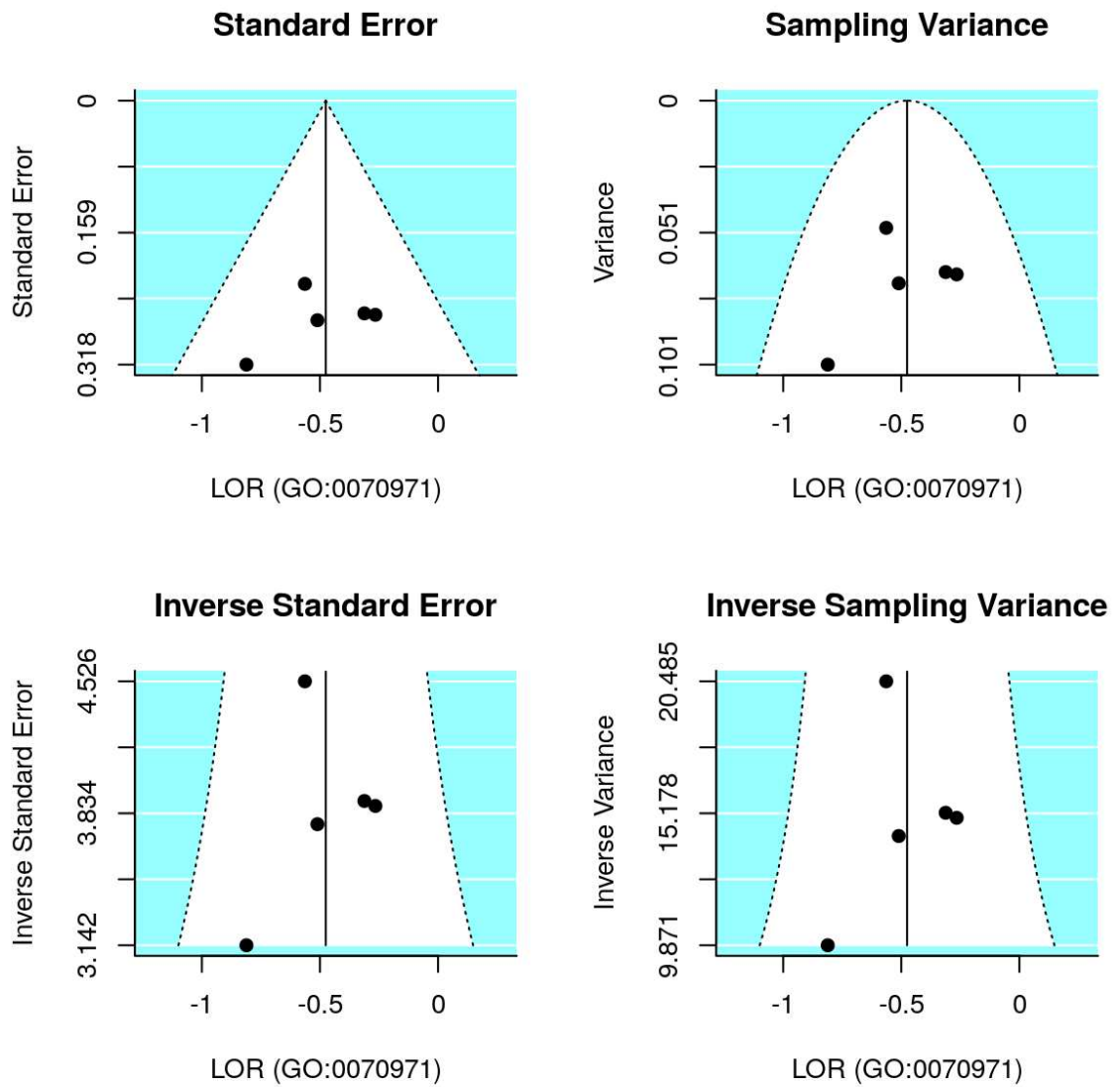
Los colores representan los estadios y las formas geométricas los géneros.



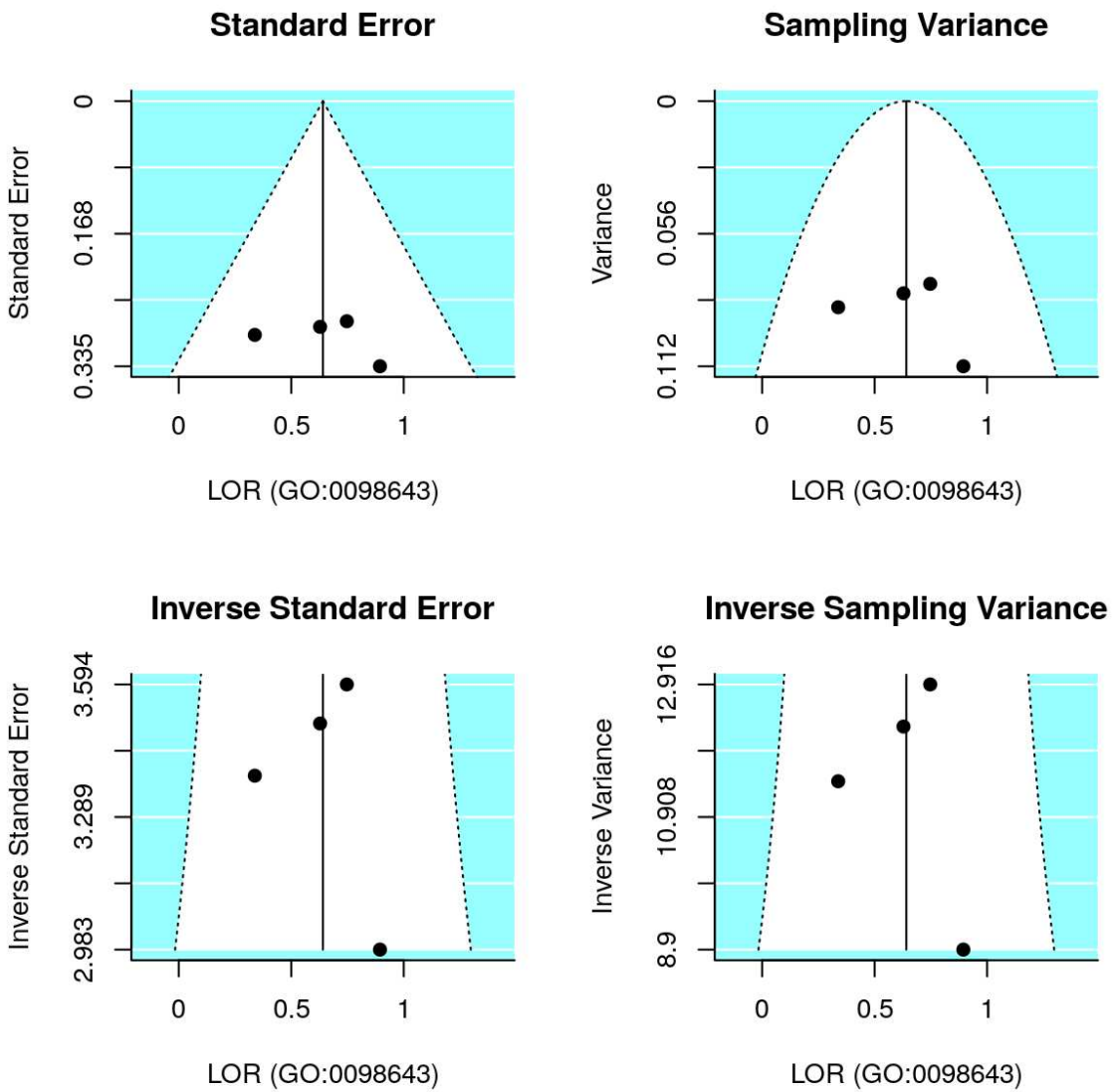
**Figura I.11:** Diagramas de embudo para GO:0005583 en el metaanálisis de términos GO de componentes celulares



**Figura I.12:** Diagramas de embudo para GO:0036038 en el metaanálisis de términos GO de componentes celulares



**Figura I.13:** Diagramas de embudo para GO:0070971 en el metaanálisis de términos GO de componentes celulares



**Figura I.14:** Diagramas de embudo para GO:0098643 en el metaanálisis de términos GO de componentes celulares

---

## **Anexo II: Tablas**

**Tabla II.1:** Versiones del software utilizado en este trabajo.

<b>Software</b>	<b>Versión</b>
R	3.5.1

<b>Paquete de R</b>	<b>Versión</b>
AnnotationDbi	1.42.1
Biobase	2.42.0
dplyr	0.7.8
GEOquery	2.50.4
ggplot2	3.1.0
ggpubr	0.2
hipathia	1.3.1
hugene10sttranscriptcluster.db	8.7.0
hugene11sttranscriptcluster.db	8.7.0
limma	3.38.3
mdgsa	1.14.0
metafor	2.0-0
metap	1.0
org.Hs.eg.db	3.6.0
purrr	0.2.5
RamiGO	1.28.0
reshape	0.8.8
stringr	1.3.1
tidyr	0.8.2
venn	1.7
xtable	1.8-3

**Tabla II.2:** Listado de rutas KEGG, de la especie *Homo sapiens*, implementadas en el paquete *hipathia*.

Ruta	Nombre de la ruta
hsa03320	<i>PPAR signaling pathway</i>
hsa03460	<i>Fanconi anemia pathway</i>
hsa04010	<i>MAPK signaling pathway</i>
hsa04012	<i>ErbB signaling pathway</i>
hsa04014	<i>Ras signaling pathway</i>
hsa04015	<i>Rap1 signaling pathway</i>
hsa04020	<i>Calcium signaling pathway</i>
hsa04022	<i>cGMP-PKG signaling pathway</i>
hsa04024	<i>cAMP signaling pathway</i>
hsa04062	<i>Chemokine signaling pathway</i>
hsa04064	<i>NF-kappa B signaling pathway</i>
hsa04066	<i>HIF-1 signaling pathway</i>
hsa04068	<i>FoxO signaling pathway</i>
hsa04071	<i>Sphingolipid signaling pathway</i>
hsa04072	<i>Phospholipase D signaling pathway</i>
hsa04110	<i>Cell cycle</i>
hsa04114	<i>Oocyte meiosis</i>
hsa04115	<i>p53 signaling pathway</i>
hsa04150	<i>mTOR signaling pathway</i>
hsa04151	<i>PI3K-Akt signaling pathway</i>
hsa04152	<i>AMPK signaling pathway</i>
hsa04210	<i>Apoptosis</i>
hsa04211	<i>Longevity regulating pathway - mammal</i>
hsa04213	<i>Longevity regulating pathway - multiple species</i>
hsa04261	<i>Adrenergic signaling in cardiomyocytes</i>

Continúa en la siguiente página

**Tabla II.2 – viene de la página anterior**

ID	Nombre
hsa04270	<i>Vascular smooth muscle contraction</i>
hsa04310	<i>Wnt signaling pathway</i>
hsa04330	<i>Notch signaling pathway</i>
hsa04340	<i>Hedgehog signaling pathway</i>
hsa04350	<i>TGF-beta signaling pathway</i>
hsa04360	<i>Axon guidance</i>
hsa04370	<i>VEGF signaling pathway</i>
hsa04380	<i>Osteoclast differentiation</i>
hsa04390	<i>Hippo signaling pathway</i>
hsa04510	<i>Focal adhesion</i>
hsa04520	<i>Adherens junction</i>
hsa04530	<i>Tight junction</i>
hsa04540	<i>Gap junction</i>
hsa04550	<i>Signaling pathways regulating pluripotency of stem cells</i>
hsa04610	<i>Complement and coagulation cascades</i>
hsa04611	<i>Platelet activation</i>
hsa04612	<i>Antigen processing and presentation</i>
hsa04620	<i>Toll-like receptor signaling pathway</i>
hsa04621	<i>NOD-like receptor signaling pathway</i>
hsa04622	<i>RIG-I-like receptor signaling pathway</i>
hsa04623	<i>Cytosolic DNA-sensing pathway</i>
hsa04630	<i>Jak-STAT signaling pathway</i>
hsa04650	<i>Natural killer cell mediated cytotoxicity</i>
hsa04660	<i>T cell receptor signaling pathway</i>
hsa04662	<i>B cell receptor signaling pathway</i>
hsa04664	<i>Fc epsilon RI signaling pathway</i>
hsa04666	<i>Fc gamma R-mediated phagocytosis</i>

**Continúa en la siguiente página**

**Tabla II.2 – viene de la página anterior**

ID	Nombre
hsa04668	<i>TNF signaling pathway</i>
hsa04670	<i>Leukocyte transendothelial migration</i>
hsa04710	<i>Circadian rhythm</i>
hsa04713	<i>Circadian entrainment</i>
hsa04720	<i>Long-term potentiation</i>
hsa04722	<i>Neurotrophin signaling pathway</i>
hsa04723	<i>Retrograde endocannabinoid signaling</i>
hsa04724	<i>Glutamatergic synapse</i>
hsa04725	<i>Cholinergic synapse</i>
hsa04726	<i>Serotonergic synapse</i>
hsa04727	<i>GABAergic synapse</i>
hsa04728	<i>Dopaminergic synapse</i>
hsa04730	<i>Long-term depression</i>
hsa04740	<i>Olfactory transduction</i>
hsa04742	<i>Taste transduction</i>
hsa04750	<i>Inflammatory mediator regulation of TRP channels</i>
hsa04810	<i>Regulation of actin cytoskeleton</i>
hsa04910	<i>Insulin signaling pathway</i>
hsa04911	<i>Insulin secretion</i>
hsa04912	<i>GnRH signaling pathway</i>
hsa04913	<i>Ovarian steroidogenesis</i>
hsa04914	<i>Progesterone-mediated oocyte maturation</i>
hsa04915	<i>Estrogen signaling pathway</i>
hsa04916	<i>Melanogenesis</i>
hsa04917	<i>Prolactin signaling pathway</i>
hsa04918	<i>Thyroid hormone synthesis</i>
hsa04919	<i>Thyroid hormone signaling pathway</i>

**Continúa en la siguiente página**

**Tabla II.2 – viene de la página anterior**

ID	Nombre
hsa04920	<i>Adipocytokine signaling pathway</i>
hsa04921	<i>Oxytocin signaling pathway</i>
hsa04922	<i>Glucagon signaling pathway</i>
hsa04923	<i>Regulation of lipolysis in adipocytes</i>
hsa04924	<i>Renin secretion</i>
hsa04925	<i>Aldosterone synthesis and secretion</i>
hsa04930	<i>Type II diabetes mellitus</i>
hsa04931	<i>Insulin resistance</i>
hsa04932	<i>Non-alcoholic fatty liver disease (NAFLD)</i>
hsa04933	<i>AGE-RAGE signaling pathway in diabetic complications</i>
hsa04950	<i>Maturity onset diabetes of the young</i>
hsa04960	<i>Aldosterone-regulated sodium reabsorption</i>
hsa04961	<i>Endocrine and other factor-regulated calcium reabsorption</i>
hsa04962	<i>Vasopressin-regulated water reabsorption</i>
hsa04970	<i>Salivary secretion</i>
hsa04971	<i>Gastric acid secretion</i>
hsa04972	<i>Pancreatic secretion</i>
hsa04973	<i>Carbohydrate digestion and absorption</i>
hsa04976	<i>Bile secretion</i>
hsa05010	<i>Alzheimer's disease</i>
hsa05012	<i>Parkinson's disease</i>
hsa05014	<i>Amyotrophic lateral sclerosis (ALS)</i>
hsa05016	<i>Huntington's disease</i>
hsa05020	<i>Prion diseases</i>
hsa05030	<i>Cocaine addiction</i>
hsa05031	<i>Amphetamine addiction</i>
hsa05032	<i>Morphine addiction</i>

Continúa en la siguiente página

**Tabla II.2 – viene de la página anterior**

ID	Nombre
hsa05034	<i>Alcoholism</i>
hsa05100	<i>Bacterial invasion of epithelial cells</i>
hsa05110	<i>Vibrio cholerae infection</i>
hsa05120	<i>Epithelial cell signaling in Helicobacter pylori infection</i>
hsa05130	<i>Pathogenic Escherichia coli infection</i>
hsa05131	<i>Shigellosis</i>
hsa05132	<i>Salmonella infection</i>
hsa05133	<i>Pertussis</i>
hsa05134	<i>Legionellosis</i>
hsa05140	<i>Leishmaniasis</i>
hsa05142	<i>Chagas disease (American trypanosomiasis)</i>
hsa05145	<i>Toxoplasmosis</i>
hsa05150	<i>Staphylococcus aureus infection</i>
hsa05152	<i>Tuberculosis</i>
hsa05160	<i>Hepatitis C</i>
hsa05161	<i>Hepatitis B</i>
hsa05162	<i>Measles</i>
hsa05164	<i>Influenza A</i>
hsa05166	<i>HTLV-I infection</i>
hsa05168	<i>Herpes simplex infection</i>
hsa05169	<i>Epstein-Barr virus infection</i>
hsa05200	<i>Pathways in cancer</i>
hsa05205	<i>Proteoglycans in cancer</i>
hsa05210	<i>Colorectal cancer</i>
hsa05211	<i>Renal cell carcinoma</i>
hsa05212	<i>Pancreatic cancer</i>
hsa05213	<i>Endometrial cancer</i>

Continúa en la siguiente página

**Tabla II.2 – viene de la página anterior**

ID	Nombre
hsa05214	<i>Glioma</i>
hsa05215	<i>Prostate cancer</i>
hsa05216	<i>Thyroid cancer</i>
hsa05217	<i>Basal cell carcinoma</i>
hsa05218	<i>Melanoma</i>
hsa05219	<i>Bladder cancer</i>
hsa05220	<i>Chronic myeloid leukemia</i>
hsa05221	<i>Acute myeloid leukemia</i>
hsa05222	<i>Small cell lung cancer</i>
hsa05223	<i>Non-small cell lung cancer</i>
hsa05231	<i>Choline metabolism in cancer</i>
hsa05321	<i>Inflammatory bowel disease (IBD)</i>